Упоређивање перформанси аналитичких пословних информационих система

By Славиша Илић

WORD COUNT 35499
TIME SUBMITTED 22-MAY-2024 11:44AM
PAPER ID 109136822

УНИВЕРЗИТЕТ СИНГИДУНУМ
БЕОГРАД
ДЕПАРТМАН ЗА ПОСЛЕДИПЛОМСКЕ СТУДИЈЕ

ДОКТОРСКА ДИСЕРТАЦИЈА

Упоређивање перформанси аналитичких пословних информационих система

МЕНТОР:

Проф. др Петар Спалевић

СТУДЕНТ: Славиша Илић

БРОЈ ИНДЕКСА: 465018/2018

Београд, 2024. год.

Ментор:

Проф. др Петар Спалевић, ментор, редовни професор

Универзитета Сингидунум

Чланови комисије:

1. Проф. др Петар Спалевић, ментор, редовни професор Универзитета Сингидунум
2. Проф. др Младен Веиновић, редовни професор Универзитета Сингидунум
3. Проф. др Стефан Панић, редовни професор Природно-математичког факултета Универзитета у Приштини са привременим седиштем у Косовској Митровици

Датум одбране дисертације:

_____

Захвалница

Захваљујем се ментору проф. др Петру Спалевићу као и коментторима проф. др Младену Веиновићу и проф. др Стефану Панићу на драгоценим саветима током израде ове докторске дисертације.

Такође, захваљујем се проф. др Синиши Илићу на експертским саветима из области база података и континуираној подршци.

Посвећено породици

Резиме

У данашњој ери велике количине података (*Биг Дата*), брзина доношења пословних одлука које су базиране на подацима је пресудна за опстанак и развој предузећа. Нове технологије су допринеле низу промена, почевши од промене начина истраживања тржишта, до новог приступа у дизајну нових производа и њиховом пласирању на тржишту. Развој Интернета, социјалне мреже и нови начини комуникације су довели до генерисања енормне количине података које могу да помогну предузећима да из њих извуку корисне информације и донесу праве одлуке за своју будућност. Један од највећих изазова данашњих предузећа је како да паметно инвестирају у одговарајућа ИТ решења која ће им омогућити генерисање правих пословних информација из великих количина података. Подаци се могу налазити на бројним изворима, како у оквиру информационог система предузећа, тако и на Интернету. Решења која се данас нуде на тржишту су различита. Једни од најбитнијих параметара за доношење одлука предузећа о избору одговарајућег решења су карактеристике, перформансе и цена.

Модеран информациони систем за пословно извештавање се састоји од више решења која раде заједно. Мерење и упоређивање перформанси различитих решења која укључују системе за управљање базама података, *Дата Варехоусе* решења, алате за манипулацију подацима (ЕТЛ) и алате за приказ информација добијених из података (БИ) су предмет истраживања и закључака добијених на основу спроведених експеримената. Сваки закључак укључује и одговарајуће препоруке за пословна предузећа у смислу смерница за избор одговарајућег решења.

Кључне речи: Перформансе база података, перформансе ЕТЛ алата, перформансе БИ алата, системи за управљање базама података, ЕТЛ, *Бусинесс Интеллигенце*.

Научна област: Базе података

Садржај

РЕЗИМЕ **5**

САДРЖАЈ **7**

ЛИСТА СКРАЋЕНИЦА **9**

1. УВОД **11**

1.1. Идентификација проблема и предлог решења 13
1.2. Допринос дисертације 15

2. СТРУКТУРА ДИСЕРТАЦИЈЕ **16**

3. БАЗЕ ПОДАТАКА **17**

3.1. Поређење перформанси релационих система за управљање базама података 20
3.2. Поређење перформанси НоСКЛ и релационих система за управљање базама података 31
3.3. Поређење перформанси различитих НоСКЛ система за управљање базама података 41
3.4. Преглед, могућности и перформансе ЕТЛ алата 47
3.5. Преглед, примена и карактеристике БИ алата 54

4. ПОРЕЂЕЊЕ БРЗИНА ИЗВРШЕЊА УПИТА НАД ВЕЛИКИМ КОЛИЧИНАМА ПОДАТАКА КОРИСТЕЋИ РАЗЛИЧИТЕ СИСТЕМЕ ЗА УПРАВЉАЊЕ БАЗАМА ПОДАТАКА **59**

4.1. Платформе база података коришћене у експериментима 61
4.1.1. ХП Вертица база података 63
4.1.2. Орацле 11г Р2 са Дата Варехоусе шаблоном и 12ц са Ин-Мемору Цолумн-Сторе 64
4.1.3. МС СКЛ са ЦолумнСторе Индексом 65
4.1.4. МариаДБ ЦолумнСторе 66
4.1.5. МуСКЛ са ИнноДБ и ИнфоБригхт енгине 67
4.2. Припрема тест окружења 68
4.3. Резултати истраживања 72
4.4. Закључак 78

5. ПОРЕЂЕЊЕ БРЗИНА ИЗВРШЕЊА ЕТЛ УПИТА ПРИ КОРИШЋЕЊУ СКЛ КОДА, ССИС УГРАЂЕНОГ ЕТЛ АЛАТА И ПЕНТАХО БЕСПЛАТНОГ ЕТЛ АЛАТА **80**

5.1. Окружење за извођење експеримената 82
5.1.1. ЕТЛ решење 1 – коришћење СКЛ упита 85
5.1.2. ЕТЛ Решење 2 - коришћење ССИС функционалних компоненти 86

5.1.3. ЕТЛ Решење 3 - коришћење Пентахо ЕТЛ алата ... 89
5.2. Резултати експеримената ... 95
5.3. Закључак ... 97

6. ПЕРФОРМАНСЕ БИ АЛАТА ... 98

6.1. Окружење за извођење експеримената ... 100
6.1.1. БИ решење 1 – Повер БИ Десктоп ... 103
6.1.2. БИ Решење 2 – Пентахо Аналутицс ... 107
6.2. Резултати експеримената ... 112
6.3. Закључак ... 114

7. ЗАКЉУЧАК ... 115

8. ЛИТЕРАТУРА ... 119

СПИСАК ТАБЕЛА ... 133

СПИСАК СЛИКА ... 134

СПИСАК ГРАФИКОНА ... 135

Листа скраћеница

ДБМС (енг. Датабасе Манагемент Систем)	Систем за управљање базама података (срб. СУБП)
ОЛТП - (енг. Онлине Трансацтион Процессинг)	Трансакционо процесирање у реалном времену
ОЛАП - (енг. Онлине Аналутицал Процессинг)	Аналитичко процесирање у реалном времену
ЕТЛ - (енг. Ехтрацт – Трансформ – Лоад)	Екстракција – трансформација - учитавање
ДW - (енг. Дата Wарехоусе)	Складиште података
Дата Март	Једноставна форма складишта података (ДW) која се фокусира само на једну област, као што је продаја или маркетинг.
Дата Мининг	Процес откривања шаблона у великим скуповима података, коришћењем метода машинског учења, статистике и система база података.
БИ - (енг. Бусинесс Интеллигенце)	Пословна интелигенција/пословно информисање
НоСQЛ Датабасе	База података која није заснована на СQЛ језику
ЦПУ - (енг. Централ Процессинг Унит)	Централна процесорска јединица
ГПУ - (енг. Графхицс Процессинг Унит).	Графичка процесорска јединица
РАМ - (енг. Рандом Аццесс Мемору)	Меморија са случајним приступом
ДРТДБС - (енг. Дистрибутед Реал-тиме Датабасе Сустемс)	Дистрибуирани системи база података који функционишу у реалном времену
ДМЛ - (енг. Дата Манипулатион Лангуаге)	Програмски језик за манипулацију података
Хадооп / Апацхе Хадооп	Хадуп представља софтверску библиотеку која омогућава дистрибуирано процесирање великих сетова података користећи кластере рачунара
ХМЛ - (енг. Ехтенсибле Маркуп Лангуаге)	Програмски језик који дефинише правила за опис података
JСОН	Отворени стандард за дефинисање формата фајлова и података који користи читљив текст
УМЛ - (енг. *Унифиед Моделинг Лангуаге*)	Језик за визуално моделирање

ЦРУД - (*ЦРЕАТЕ, РЕАД/РЕТРИЕВЕ, УПДАТЕ, ДЕЛЕТЕ*)	Четири основне операције у базама података: креирај, читај/претражи, ажурирај, бриши.
ЦСВ - (енг. цомма-сепаратед валуес)	Текстуални фајл формат који користи зарезе за одвајање вредности
ОС - (енг. Оператинг Сустем)	Оперативни систем
ОЛЕ ДБ	Апликативни програмски интерфејс развијен од стране компаније Мајкрософт, који омогућава приступ подацима на униформан начин.
ЈДБЦ (енг. Јава ДатаБасе Цоннецтивиту)	Представља апликативни програмски интерфејс који омогућава приступ бази података у јава програмском окружењу.
ССД (енг. Солид Стате Дриве)	Полупроводнички диск за смештање информација који не користи покретне делове.
КПИ – Кеу Перформанце Индицаторс	Кључни индикатори перформанси
АИ - (енг. Артифициал Интеллигенце)	Вештачка интелигенција

1. Увод

У савременом пословању, брза еволуција технологије и свеприсутна дигитализација трансформишу начин на који организације послују и доносе кључне одлуке. Коришћење информација из података постаје основна претпоставка успешног пословања, отварајући врата новим могућностима, али истовремено постављајући изазове пред организације које желе да остану конкурентне. Овај рад је резултат истраживања перформанси система за управљање базама података и алата за трансформацију и визуализацију података у контексту коришћења информација у доношењу одлука.

Дигитална трансформација доноси са собом огромне количине података, али кључно питање је како организације могу искористити ту енормну количину информација ради остваривања конкурентске предности. У ери *Биг Дата*, аналитика података постаје императив. Пословни лидери схватају да није довољно само прикупити податке, већ и развити способност анализе и извлачења значајних сазнања из тих података. Кроз правилну интерпретацију података организације могу идентификовати трендове, препознати прилике за иновације и антиципирати изазове, што им омогућава доношење правих пословних одлука.

Међутим, упркос огромном потенцијалу постоје и озбиљни изазови у коришћењу информација и база података. Питања приватности података, етичких норми и сигурносних претњи постају све релевантнија. Организације морају пажљиво балансирати између употребе података и заштите приватности својих клијената, чиме се ствара комплексан оквир у ком послују. Осим тога, потребно је осигурати интегритет података, како би се избегле погрешне интерпретације које могу довести до доношења погрешних одлука.

Данашње пословање је неодвојиво повезано са способношћу ефикасног коришћења информација и база података. Организације које успешно савладају изазове дигиталне трансформације и постану агилне у коришћењу података, имају прилику да обликују будућност своје индустрије. Међутим, да би се постигао овај циљ, потребно је не само инвестирати у технологију, већ и развити јасне политике, етичке смернице и безбедносне стандарде. На тај начин би се очувао интегритет и поверење у коришћење информација и база података.

Доносиоцима пословних одлука није само важно да из података добију одговарајуће пословне показатеље, већ и колико се брзо може доћи до њих, и по којој цени. Управо су

ова питања чинила полазну тачку за истраживања различитих технолошких решења која су представљена у овој дисертацији.

1.1. Идентификација проблема и предлог решења

Дигитална трансформација је присутна у свим гранама индустрије са циљем унапређења подршке пословним процесима и пружања подршке доношењу одлука. Од појаве персоналних рачунара настала су и софтверска решења за евидентирање различитих догађаја. Такви системи су на почетку били оријентисани на чување информација о појединачним догађајима. За креирање и генерисање агрегационих извештаја са приказом трендова је у то време било потребно напредно информатичко знање, а сам процес се мерио сатима а понекада и данима.

Већина пословних организација која базира своје одлуке на подацима, поседује два информациона система за управљање подацима: *Онлине Трансацтион Процессинг* (ОЛТП) и *Онлине Аналутицал Процессинг (*ОЛАП*)*. ОЛТП систем је намењен евидентирању пословних догађаја. Он омогућава великом броју конкурентних корисника система да приступе, евидентирају нове догађаје, врше претрагу и ажурирају постојеће евиденције. Овај систем је дизајниран за брз унос и ажурирање података, као и за извештавање о појединачним догађајима. Уколико би било потребно креирати сложенији агрегациони упит над већим бројем табела и великом количином података ради генерисања извештаја у овом систему, таква активност би произвела значајан пад перформанси система који би свакако осетили остали корисници система. За разлику од ОЛТП система који је оптимизован за унос података, ОЛАП систем је дизајниран за анализу и припрему података за генерисање комплексних агрегационих извештаја. Један од првих корака у пројектима напредног извештавања је идентификација и трансформација података из ОЛТП система, као и њихово учитавање у ОЛАП систем користећи одговарајући *ЕТЛ (Ехтрацт – Трансформ – Лоад)* алат. Овај процес се изводи периодично како би ОЛАП систем користио најсвежије податке. ОЛАП систем је унапред припремљен да складишти податке користећи одговарајућу архитектуру која омогућава извршавање потребних анализа по различитим параметрима (димензијама), великом брзином. Ови системи могу да обезбеде различите анализе пословања као што су на пример: вредности продаје, набавке по регионима, груписане по пословницама, по годишњим кварталима, по данима у недељи и слично. Такође, они омогућавају и проучавање скривених законитости у пословању. Оптимизовани су за брзо претраживање и анализу података. Подаци се у ОЛАП системима чувају у тзв. складиштима података (*Дата Варехоусе - ДВ*), а корисницима (најчешће руководиоцима фирме) се анализа података и резултати приказују преко софтвера/алата пословне интелигенције (*Бусинесс Интеллигенце - БИ*). Ови алати нуде не само табеларни, већ и графички приказ резултата са динамичким опцијама филтрирања. Савремени БИ алати су прилично једноставни и не захтевају познавање програмских језика.

Пословним организацијама су информације из ОЛАП система података потребне чим је пре могуће, како би остариле одговарајуће предности у односу на конкуренцију. Из тога разлога перформансе система имају једну од кључних улога у избору одговарајућег решења. Није свеједно да ли ће време потребно за добијање резултата једног упита за анализу трајати неколико сати, сат времена, неколико минута или неколико секунди.

Предмет и резултати истраживања који су представљени у овој дисертацији се односе на анализу перформанси различитих доступних решења који су саставни део пројеката имплементације напредних решења за пословно извештавање и то:

- анализа перформанси *Опен Соурце/бесплатних* и комерцијалних *Дата Варехоусе* система који се користе за ОЛАП информационе системе
- анализа перформанси *Опен Соурце/бесплатних* и комерцијалних ЕТЛ алата којим се врши издвајање података из ОЛТП базе, трансформација података и пуњење ОЛАП база,
- анализа перформанси *Опен Соурце/бесплатних* и комерцијалних Бусинесс Интеллигенце система за повлачење података из *Дата Варехоусе* система и за приказивање резултата анализе.

На основу извршених истраживања, за сваки од наведених компоненти су изведени одговарајући закључци и дате препоруке у вези избора одговарајућих решења, имајући у виду различите аспекте и захтеве данашњег пословања.

1.2. Допринос дисертације

У овом раду су представљени следећи научни доприноси:

- У дисертацији је детаљно наведен преглед досадашњих истраживања. Анализиране су 122 референце у којима су наведена истраживања примене, перформанси, упоређивања перформанси различитих информационих система на различитим платформама, било да су они у облаку (*Цлоуд Басед*), на виртуалним машинама, на одређеном оперативном систему, на комбинацији оперативних система, на комерцијалним или *Опен Соурце/бесплатним* софтверим и то:
  - ✓ поређење перформанси релационих система за управљање базама података,
  - ✓ поређење перформанси НоСQЛ и релационих система за управљање базама података,
  - ✓ поређење перформанси различитих НоСQЛ система за управљање базама података,
  - ✓ преглед, могућности и перформансе ЕТЛ алата,
  - ✓ преглед, примена и карактеристике БИ алата.

  На основу анализе постојеће литературе добијена је полазна основа за даље истраживање.

- Извршена је анализа карактеристика често коришћених *Опен Соурце/бесплатних* и комерцијалних *Дата Варехоусе* система, са посебним акцентом на конфигурисање одређених подешавања система (меморија, *ров-ориентед, цолумн-ориентед, ин-мемору*) и показани су резултати мерења перформанси тих система уз различита подешавања.
- Наведени су често коришћени *Опен Соурце/бесплатни* и комерцијални ЕТЛ алати, показани су добијени резултати мерења перформанси тих алата и извршена је анализа добијених резултата.
- Извршена је анализа карактеристика *Опен Соурце/бесплатних* и комерцијалних софтвера за пословну интелигенцију (*Бусинесс Интеллигенце*), мерење њихових перформанси и представљена је анализа добијених резултата.

На основу свих добијених резултата дате су препоруке за пословне фирме, које алате треба изабрати у односу на карактеристике као што су: перформансе, могућност

интеграције са другим системима итд. Најзначајније препоруке су дате у закључцима појединих секција дисертације као и у крајњем закључку ове дисертације.

2. Структура дисертације

У уводном делу дисертације у поглављу 1 је представљен проблем који је предмет истраживања, као и начин методолошког приступа. У дисертацији је изложен основни поглед на системе за управљање базама података, од њиховог настанка до данас. Преглед досадашњих истраживања се заснива на прикупљеним подацима о перформансама и упоређивању перформанси различитих информационих система на различитим платформама.

Досадашња истраживања су представљена у поглављу 3 и подељена су у пет секција према областима истраживања:

- Поређење перформанси релационих система за управљање базама података,
- Поређење перформанси НоСQЛ и релационих система за управљање базама података,
- Поређење перформанси различитих НоСQЛ система за управљање базама података,
- Преглед, могућности и перформансе ЕТЛ алата,
- Преглед, примена и перформансе БИ алата.

Наведена истраживања су послужила за дефинисање полазних основа за даља истраживања представљена у овој дисертацији.

Поређење брзина извршења упита над великим количинама података користећи различите системе за управљање базама података је представљено у поглављу 4. Извршена је анализа и мерење карактеристика често коришћених комерцијалних и *Опен Соурце/бесплатних Дата Варехоусе* система са посебним акцентом на конфигурисање и резултате одређених подешавања, као што су меморија, *row-ориентед, цолумн-ориентед, ин-мемору* и слично.

Истраживање поређења перформанси најчешће коришћених комерцијалних и *Опен Соурце/бесплатних* ЕТЛ алата као и анализа добијених резултата је представљена у поглављу 5.

У поглављу 6 је представљена анализа перформанси комерцијалног БИ алата Мицрософт ПоwерБИ Десктоп и бесплатног БИ алата Пентахо Аналутицс, на основу извршених експеримената.

У поглављу 7 је дат свеобухватни закључак који обухвата наведена истраживања из области *Дата Wарехоусе* система, ЕТЛ и БИ алата, као и правци даљих истраживања. Закључак обухвата и препоруке за избор одговарајућих решења намењених пословним предузећима.

Коришћена литература је наведена у поглављу 8.

3. Базе података

Типови база података се често називају и моделима база података и представљају одговарајуће структуре које се користе за организовање података у оквиру Система за управљање базама података (СУБП) или на енглеском *Датабасе Манагемент Сустем (ДБМС)*. Постоје различити типови база података. Неке базе података можемо да посматрамо из историјске перспективе јер данас више нису у употреби, док су остале које су у употреби, спремне да реше изазове новијих захтева и очекивања од стране крајњих корисника и различитих индустрија.

Старији типови база података као што су *Флат Филе* базе (чувају податке у оквиру једног фајла у формату чистог текста), хијерархијске базе података (користе релације родитељ-дете да би повезале податке у структуру стабла), или пак мрежне базе података (омогућавају креирање комплекснијих релација података), представљају неку врсту еволуције ка развоју система за управљање подацима, које имамо данас у савременој употреби.

Прекретницу у развоју база података су представљале релационе базе података које користе табеле са одговарајућим везама за смештање и управљање добро структуираним подацима. Своју примену су нашле у скоро свим индустријама и могу бити веома робусне. Неки од представника ових СУБП су: МуСQЛ, Орацле, ИБМ ДБ2, Мицрософт СQЛ Сервер, МариаДБ, ПостгреСQЛ и друге.

За потребе решавања изазова за смештање и манипулацију огромних количина неструктуираних и полуструктуираних података на интернету, почетком 2000-их година су се појавиле такозване НоСQЛ базе података, често називане "*нон-СQЛ*" или "*нот онлу СQЛ*". Оне представљају алтернативу за управљање подацима које је тешко сместити у традиционални релациони систем. Постоји неколико типова НоСQЛ база података које имају своју примену у решавању савремених изазова, од којих су најчешће коришћени:

- *Кеу-валуе* базе података, које користе једноставне складишне ћелије са означеним кључем, којима се може приступати од стране дистрибуираних система или серверских кластера. Најчешћа примена ових база података је у сценаријима када је потребна брза манипулација над подацима, као што су конфигурациони подаци, информације о статусним стањима и слично. Представници ових база података су: Азуре Табле Стораге, МЕМБАСЕ, Риак, Редис, Мемцацхед, Етцд.
- *Доцумент* базе података које се често називају и ЈСОН (ЈаваСцрипт Објецт Нотатион) базе, смештају податке у флексибилне и само-описујуће структуре у форми докумената. Подаци се смештају у колекције докумената. Сваки документ представља засебан ентитет и садржи податке и асоциране метаподатке. Представници ових база података су: ЦоуцхДБ, МонгоДБ, Доцумент-басед МуСQЛ, Терраторе, РетхинкДБ.
- *Граπх* базе података су намењене складиштењу огромних количина података на интернету. Подаци се смештају у нодове који представљају ентитете или објекте, и ивице (едгес) које представљају релације између нодова. Дизајниране су за складиштење података са комплексним релацијама и користе се за смештање података социјалних мрежа, система за детекцију превара и друго. Представници су: Нео4Ј, Инфините Грапх, Сонес, ИнфоГрид, ЈанусГрапх, Дграпх.
- *Цолумн* базе података или *Wиде Цолумн-сторе / Цолумн Фамилиес* су дизајниране да реше изазове који су настали у празном простору функционалности између релационих и *доцумент* база података. У релационим базама података, шема дефинише колоне у табели, тип података сваке колоне, као и друге критеријуме. Сви редови у табели морају да одговарају дефинисаној шеми. *Цолумн* базе података, користе структуру фамилија колона (*цолумн фамилиес*) чији редови користе своје засебне формате. Сваки ред се састоји од од јединственог идентификатора реда и скупа имена колона и њихових вредности. Користећи овакав дизајн, сваки ред у фамилији колона има своју шему. Шема се може лако модификовати уколико се за тим укаже потреба. Представници ових база података су: Хадооп / ХБасе, Цассандра, Хупертабле, Цлоудата, Амазон СимплеДБ, СциДБ.

Такође, треба напоменути и неке додатне типове база података које због својих специфичних карактеристика могу представљати одличан избор:

- *Цолумн-ориентед* базе су врло сличне релационим базама података, само што се код њих подаци смештају на диск по колонама, уместо по редовима. То у пракси значи да су сви подаци једне колоне у табели смештени заједно,

чиме се омогућава брза агрегација већих скупова података. Овај тип базе података представља одлично решење за аналитичке потребе. Ради избегавања конфузије, треба потпуно раздвојити *Колумн-ориентед* базе података представљене у овом пасусу и претходно наведене НоСQЛ *Колумн* базе.

- *Објектно оријентисане* базе података, код којих су подаци организовани у објекте. Објекти садрже како саме податке, тако и методе за извршавање операција над тим подацима. Дизајниране су да подрже веома комплексне структуре података и њихова најшира употреба је од стране апликација које захтевају приступ великој количини структуираних или полуструктуираних података.

У данашњем пословању једна од битних карактеристика за избор одговарајуће платформе база података су њене перформансе. Имајући у виду присутност различитих типова база података које могу бити коришћене као најбољи избор за специфичне изазове и захтеве, бројни истраживачи су покушавали да дају одговор у виду одговарајућих смерница на основу извршених експеримената.

У наредним поглављима је изложен преглед досадашњих истраживања везаних за испитивање перформанси и упоређивање перформанси различитих информационих система (инсталираних на различитим платформама), било да су они у облаку (*Цлоуд-Басед*), на виртуалним машинама, на одређеним оперативним системима, на комбинацији оперативних система, на комерцијалним или *Опен Соурце/бесплатним* софтверима.

3.1. Поређење перформанси релационих система за управљање базама података

Често коришћен термин у експериментима поређења перформанси различитих система је "Бенчмарк". Он у свом општем значењу представља процес упоређивања перформанси, процеса или пракси организације са онима који се сматрају најбољима у индустрији или сличним организацијама. Бенчмарк база података представља стандардни скуп извршних инструкција које се користе за мерење и упоређивање перформанси различитих система база података, извођењем контролисаних експеримената.

Аутори су у раду [1] представили синтетички прототип метод који поред стандардних бенчмарк експеримената користи и додатни сет од више генерализованих корисничких експеримената. Под стандардним бенчмарк тестовима се подразумевају: ТП1 (дизајниран за мерење комплетне трансакције), ТПЦ-А (дизајниран за мерење броја трансакција у секунди које систем може да изврши када се покреће са више терминала), ТПЦ-Б (дизајниран за мерење пропустности, нпр. колико трансакција у секунди дати систем може да изврши), ТПЦ-Ц (дизајниран за он-лине обраду трансакција која укључује скуп од пет истовремених трансакција различитих типова и сложености, а које се могу извршити на мрежи или у реду за одложено извршење), ТПЦ-Д (дизајниран за подршку у одлучивању који моделира неколико опште применљивих аспеката система за подршку одлучивању, а укључује упите и одржавање података), ТПЦ-Х (дизајниран за обраду трансакција базе података, специфичан за подршку у одлучивању), ТПЦ-Р (дизајниран за подршку у одлучивању - сличан ТПЦ-Х, који омогућава додатне оптимизације на основу коришћења напреднијих упита и истовремених модификација података), ТПЦ-ВМС (дизајниран за тестирање база података које раде у режиму виртуализације), ТПЦ-W (дизајниран за трансакциони веб бенчмарк, код кога се радно оптерећење обавља у контролисаном окружењу интернет трговине). Прототип је креиран са корисничким интерфејсом ради уноса додатних захтева. Аутори закључују да тест резултати стандардних бенчмарка представљају процене могућих перформанси система за одређене предефинисане типове апликација и да се перформансе система база података на актуалном домену базе могу значајно разликовати од оних који су предвиђени стандардним бенчмарком. Резултати група стандардних ТПЦ бенчмарк тестова постају валидни само ако се тестирање обавља под прописаним условима, као што су: количина података у бази мора бити већа од прописане, оптерећење базе мора бити веће од прописаног, број редова у табелама над којима се врши тестирање мора бити већи од дефинисаног броја, итд. Из овога се јасно види да се ТПЦ тестови не могу извршити над сваким системом.

Информациони системи база података су се раније традиционално покретали (у време писања овог рада) на једном хост компјутеру, на коме се налазио оперативни систем, апликациони сервер и сервер базе података [2]. Код овакве архитектуре, постојање сервера базе података заједно са апликационим сервером на истом рачунару може значајно да деградира перформансе хост рачунара. Архитектура са базом података на засебном хосту (*бацкенд датабасе мацхине арцхитецтуре*) је предложена ради премештања оптерећења са заједничког хоста на засебан специјализован компјутерски систем. На овај начин се укупно оптерећење дели на два рачунарска система, апликативни који генерише захтеве ка систему база података и систем база података који процесира захтеве над базом података. Анализа предности и мана ове архитектуре укључује разматрања параметара као што су перформансе, поузданост, дељење базе, безбедност и модуларност. Иако је повећање перформанси система кључни приоритет, основни изазов се огледа у прихватљивом односу „цена/перформансе“. Истраживачи на Универзитету у Висконсину су развили свеобухватну стратегију за тестирање перформанси система за управљање базама података (СУБП) у архитектури где је СУБП инсталиран на истом хосту са апликативним сервером и архитектури где је СУБП инсталиран на засебном рачунару. Како би се тестирале разлике у перформансама ове две архитектуре, креирана је посебна синтетичка база података и извршено је *сингле-усер* и *мултипле-усер* тестирање на системима који укључују Орацле, Ингрес, СQЛ-ДС и ИДМ-500. Резултати тестирања су показали да архитектура са СУБП на засебном хосту за краће време приказује резултате упита.

У савременим информационим системима, базе података се могу користити као ОЛТП базе, ОЛАП базе, сервисне архитектуре и *е-Сциенце* апликације [3]. Због велике количине података које се у њима налазе и великог броја корисника који приступају апликацијама и сервисима, често су базе података организоване у дистрибуиране системе. Ти системи се налазе на великом броју мрежних чворова и приликом извршавања упита морају да размењују податке преко рачунарске мреже. Комуникација која се дешава између чворова укључује процесе повезане са контролом конкурентности, подешавањем репликације, алгоритмима дистрибуираног ажурирања и слично. У тим системима базе су инсталиране на различитим архитектурама и рачунарима са различитим физичким перформансама, као што су број и брзина ЦПУ-ова, количина РАМ меморије, величина и брзина дискова итд. Над базама се извршавају и трансакције којима се ажурира одређени број табела сагласно алгоритмима за контролу конкурентности. Како би се на неки начин мериле перформансе таквих система, одређена група аутора је пробала да постави моделе и дефинише параметре тих модела како би се прво теоријски, а касније и симулацијама, на неки начин потврдила ваљаност модела.

Један од коришћених модела за анализу перформанси софтвера је тзв. *Qуеуинг Нетwоркс (QН)* [4]. То је модел који решава захтеве корисника (*сервице реqуестс*) који долазе у

сервисне центре (сервери) и формирају ред (*qyeye*), како би били услужени. У овим моделима, као параметри модела се користе:

- брзина којом корисници долазе на ред (*worклоад интенситу*), нпр. сваке две секунде,
- средње време потражње услуга од стране корисника (*авераге сервице реqуиремент оф а цустомер*), нпр. 1.25 секунди,
- оптерећење система (*утилизатион*),
- средње време потребно да се корисник потпуно услужи (*ресиденце тиме*),
- број корисника који чекају на ред (*qyeye ленгтх*) и
- пропустност (*тхроугхпут*).

Користећи наведену методу, аутори у раду [5] повезују перформансе компоненти дистрибуираних система база података (брзина и број ЦПУ-ова, брзина дискова, капацитет рачунарске мреже, време између закључавања табела, број редова који се уписују у табеле и др.) у горе наведене параметре. Они су такође извршили и анализу великог броја објављених радова у којима су на различите начине моделовани системи са базама података користећи QH методу и покушали да групишу те моделе у зависности од нивоа детаља којима је интерни дизајн трансакција постављен у моделу. Тако су сви анализирани модели из 30-ак радова подељени у: модел црне кутије (где трансакциони дизајн није представљен у моделу), модел трансакционе обраде (где је трансакција представљена као одговор на захтев за услугу), модел величине трансакције (где је модел заснован на броју објеката на које трансакција делује) и модел трансакционе фазе (где је модел заснован на броју фаза кроз које трансакција пролази приликом извршавања захтева). За сваки анализирани рад је дато објашњење како је моделован QH. Закључак аутора је да на перформансе система база података утичу комплексност и међузависне функционалности, као и референтне локације извршења трансакција и конкурентног приступа меморији и диску.

Систем за управљање базама података или скраћено СУБП има значајну улогу у данашњем пословању у многим индустријама и пројектима. Ови системи се користе за складиштење података, њихову манипулацију и извештавање. Постоји низ компанија које производе СУБП системе и они се на тржишту могу наћи по различитим ценама, од могућности бесплатног коришћења, у случају *Опен Соурце/бесплатних* решења, до решења у износу од више десетина хиљада долара у случају *Проприетару* СУБП решења. Потенцијални корисници ових решења доносе одлуку о избору адекватног решења на основу различитих карактеристика које решење нуди (перформансе, поузданост, безбедност и др.). Поред тога битну улогу у избору решења може да има и његова цена.

Аутори су у раду [6] извршили мерење перформанси пет различитих СУБП доступних на тржишту: МС СQЛ Сервер 2008, Орацле 11г, ИБМ ДБ2, МуСQЛ 5.5 и МС Аццесс 2010. За потребе тестирања коришћени су СQЛ упити различите комплексности над свим наведеним системима. *Перформанце* бенчмарк је коришћен за мерење времена извршења упита, ЦПУ оптерећења, коришћење меморије, коришћење виртуелне меморије, као и број извршених процесорских нити (*тхреадс*). Тестирања су изведена на компјутерској платформи са Дуал-Процесором, Интел Хеон Е5649, 6к2 кора, процесор клок 2.53ГХз са 32ГБ РАМ-а и 2ТБ секундарног складишног капацитета, користећи МС Виндовс Сервер 2008, 64-битне верзије. Аутори су развили апликацију за тестирање користећи Ц#.НЕТ под НЕТ Фрамеворк 4.0. Апликација је извршавала два задатка: аутоматско попуњавање табела база са 1,000,000 редова пре извршења самог тестирања и извршавање СQЛ упита различитих комплексности. Поред претходно наведеног, у апликацију је уграђен и тајмер за мерење извршења процеса. За мерење коришћења меморије и процесорског времена је коришћен МС Виндовс Таск Манагер (WTM) алат који долази уз МС Виндовс оперативни систем. На свим СУБП платформама аутори су користили исто окружење са примером базе података која се користи у *ретаил* пословању. Ово окружење је укључивало *Фронт Енд* систем за генерисање фактура, рачуна и поруџбеница и *Бацк Енд* за управљање подацима у складишту. Табеле базе података су пре теста попуњене са по 1,000,000 редова, насумичним вредностима. За потребе теста извршено је 10 различитих типова упита и нивоа комплексности. Резултати тестирања су показали да је за дате примере тестова ИБМ ДБ2 показао највећу брзину извршења упита, док је МС Аццесс показао најмање оптерећење процесора у поређењу са осталим СУБП системима. Просечно коришћење меморије је најниже код МС Аццесс, а највише код ИБМ ДБ2.

Системи база података се могу класификовати као централизовани и дистрибуирани. Док се процеси код централизованих система извршавају над једном компјутерском јединицом, дистрибуирани системи база података укључују компјутерске јединице на више локација повезане комуникационом мрежом, при чему систем на свакој локацији има скуп своја правила, а системи на свим локацијама раде заједно. Корисник на било којој локацији може приступати и радити са подацима који се физички налазе на другим локацијама, на начин као да се сви подаци налазе на локацији корисника. Данашње пословање захтева дистрибуирано процесирање података из више разлога, а највише ради обезбеђења конкурентности на тржишту. Алгоритми који се примењују у системима база података за приступ и рад са подацима морају да задовоље временске рокове подржаних апликација. Аутори у раду [7] разматрају проблеме са перформансама које су значајне за *дистрибутед реал-тиме датабасе сустемс (ДРТДБС)* и истражују досадашња сазнања и истраживања над параметрима као што су системи доделе приоритета процеса, *реал-тиме* извршни протоколи, као и оптимизација коришћења меморије у репликационим, односно не-репликационим окружењима, који се односе на процесирање дистрибуираних *реал-тиме* трансакција. Аутори на основу свог

истраживања закључују да полисе доделе приоритета имају свој значај само када се кохорти трансакције (скуп процеса који се извршавају у име иницијалне трансакције, на другим сајтовима) извршавају секвенцијално. Највећи део *Реал Тиме Цоммит* протокола покушава да побољша перформансе, омогућавајући трансакцијама да користе податке закључане од стране других трансакција. Тренутно, постоји озбиљан недостатак протокола који би могли боље и ефикасније да управљају коришћењем меморије, креирањем мањег броја привремених објеката.

У току пројекта развоја решења "Репосе" за *Ресорт Манагемент Сустем (РМС)* [8] који је имао за циљ да складишти информације о хотелима, њиховим корисницима, сервисима који се нуде, резервацијама итд., аутори су увидели прилику за анализу и поређење перформанси два СУБП и то МуСQЛ и СQЛ Сервер. Они су виђени као најужи избор решења за управљање и складиштење података. МуСQЛ је једна од најпопуларнијих *Опен Соурце/бесплатних* релационих СУБП, која поред особина као што су подршка различитим типовима података, подршка записима са фиксном и променљивом дужином, брзо извршење процеса итд., нуди и независност од платформе оперативног система. Са друге стране, СQЛ Сервер је СУБП производ компаније Мицрософт, дизајниран за рад у *Ентерприсе* окружењима са веома напредним карактеристикама као што су оптимизација меморије у раду са табелама и ускладиштеним процедурама, обезбеђивање алата за миграцију података са најкоришћенијих система база података, подршка за кластер и репликационе сервисе итд. "Репосе" РМС је коришћен као алат за прикупљање података користећи дневну интеракцију са корисницима који су уносили податке у систем. За потребе тестирања и мерења перформанси коришћени су СQЛ Сервер 2008 и МуСQЛ 5.6.17. Подаци су прво прикупљени у МуСQЛ а потом мигрирани на СQЛ сервер користећи алат Миграцион Ассистант фор МуСQЛ. Рачунар који је коришћен за упоредно тестирање је 64-битни Виндовс 8 са Интел® Цоре™ и5-2430М ЦПУ, са клок фреквенцијом 2.4 ГХз. и 4ГБ РАМ. Аутори су за потребе тестирања дизајнирали неколико различитих ДМЛ упита (*СЕЛЕЦТ, ИНСЕРТ, УПДАТЕ* и *ДЕЛЕТЕ*) који су извршени на оба *Бацк-Енд* СУБП система, при чему је мерено време њиховог извршења. Резултати мерења перформанси су показали да СQЛ Сервер нуди боље перформансе у односу на МуСQЛ када се посматра параметар "време извршења". У свим тест случајевима, изузев Инсерт упита, СQЛ Серверу је било потребно мање времена за извршење. МуСQЛ је такође показао слабије резултате у смислу скалирања, скоро два пута више времена је било потребно за извршење упита у случајевима повећања броја редова. Најубедљивија разлика у перформансама наведена два СУБП се види код примене *СЕЛЕЦТ* упита над табелом са 3,000 редова, где је МуСQЛ-у било потребно двоструко више времена за извршење упита у односу на СQЛ Сервер. Иако резултати мерења перформанси недвосмислено указују да је СQЛ Сервер очигледан избор за *Бацк-Енд*, његова цена представља велики изазов за мала предузећа. *Опен Соурце/бесплатна* природа МуСQЛ-а уједно значи и да ће цена имплементације бити минимална, те из тог разлога мала

предузећа често бирају овај производ који нуди задовољавајуће перформансе на мањим базама података које се користе у web окружењима.

У *Ентерприсе* окружењима за рад са великом количином података избор одговарајуће релационе СУБП платформе има велики значај. Базе података представљају језгро апликативних решења у многим индустријама. Платформе релационих база података које се нуде на тржишту могу бити *Проприетару* за које је потребно купити лиценцу (Мицрософт, Орацле, ИБМ) док остале могу бити *Опен Соурце/бесплатне*. Избор релационе базе података може бити добар уколико корисник поседује ограничену количину података. Ради постизања најбољих перформанси на изабраној платформи, дизајнери базе података имају кључну улогу, јер ће од дизајна базе зависити и брзина извршавања процеса над подацима у бази. Аутори су у раду [9] извршили компаративну анализу резултата добијених мерењем перформанси три најчешће коришћене серверске базе података: СQЛ Сервер, Орацле и МуСQЛ. Различити типови упита са различитим нивоима комплексности су коришћени за потребе мерења времена извршења. Сви експерименти су извршени на Wиндowс оперативном систему. Коришћена је методологија тестирања времена извршења упита на СQЛ Сервер 2016, Орацле 12ц и МуСQЛ 5.7, покретањем различитих типова упита. За потребе експеримента аутори су лимитирали број редова у табелама између 300,000 и 400,000. Тачније 324,500 редова је убачено у сваку од наведене три базе података користећи софтвер који је специјално направљен за потребе овог експеримента. Резултати тестирања над овом количином података су показали да МуСQЛ показује најбоље перформансе код упита које немају постављене услове у упиту. Код једноставних кондиционалних упита са *WXEPE* клаузулом, МуСQЛ и СQЛ Сервер су показали добре и идентичне перформансе за разлику од Орацле, коме је потребно више времена. СQЛ Сервер показује најбоље перформансе у односу на остала два решења код наредби за ажурирање података у једној колони над свих 324,500 поља у табели. Орацле показује најбоље резултате у односу МуСQЛ и СQЛ Сервер код експеримента копирања комплетној садржаја предметне табеле у другу. Избор МуСQЛ базе података може бити добар у случају малих предузећа када количина података није велика, јер у тим случајевима МуСQЛ показује добре карактеристике перформанси. Код избора базе података треба имати у виду и њихову цену. МС СQЛ Сервер и Орацле база имају цене од неколико десетина хиљада долара док је МуСQЛ бесплатан.

Поређењем перформанси релационих база података, Мицрософт СQЛ Сервера и и Орацле су се бавили и аутори у раду [10], користећи различите упите над табелама од 100,000 и 500,000 редова као и функције креирања, ажурирања и брисања записа из табела. Резултати истраживања су показала да је СQЛ сервер 2000 показао генерално боље резултате од Орацле 9и код већине упита.

У последњих неколико декада релационе базе података су доминирале у скоро свим индустријама и представљале окосницу платформе за складиштење и управљање подацима. Релационе базе најчешће уписују скупове података континуирано, у редовима на физичком диску. Овакав приступ се обично назива "*Н-ару стораге модел (НСМ)*". За разлику од претходно наведеног, поједине имплементације релационих база користе такозвани *Цолумн-Сторе* приступ уписа података на диск, односно "*децомпосед стораге модел (ДСМ)*". Аутори су у раду [11] извршили низ експеримената испитивања перформанси користећи представнике имплементација ова два приступа, Орацле и Вертица. Истраживање су започели анализом дата модела и оптимизацијом перформанси оба система. Потом су развили бенчмарк перформанси како би квантификовали перформансе система са *row* и *цолумн* оријентацијом података и на крају на основу резултата експеримената извели одговарајуће закључке за различите типове упита и оптерећења. Свака имплементација релационих база података дефинише како се подаци физички смештају на диск. Код Орацле 11г дата фајлови се се смештају редно у тзв. табеларна поља ("*таблеспацес*"), при чему се сваки дата фајл може налазити у само једном табеларном пољу, а свако табеларно поље може да садржи више дата фајлова. Друге значајне структурне компоненте Орацле базе укључују контролне фајлове, архивиране логове, фајлове за праћење промена итд. Величине дата блокова код Орацле су од 2КБ до 32КБ. Вертица користи *Цолумн-сторе* имплементацију уписа података. Она користи два типа складишта, *Реад Оптимизед Сторе (РОС)* и *Wрите Оптимизед Сторе (WOC)* за смештање података и свако складиште има свој тип контејнера. РОС контејнер садржи скупове комплетних података сортираних по одређеном редоследу, као пар фајлова по колони. Свака колона је смештена у два засебна фајла, један са актуелним подацима, а други са позиционим индексом. Позициони индекс је важан податак који значајно утиче на перформансе. WOC контејнер садржи податке који се чувају искључиво у меморији, при чему њихова оријентација (у редовима или колонама) има врло мали значај. Аутори су експерименте извели на серверу са 4хИнтел Хеон процесорима са по 4 кора, колокованим на 2.4 ГХз, 128 ГБ РАМ и 4 ТБ чврстим диском. Базе података су инсталиране на виртуалним машинама користећи ВМwаре ЕСХи софтвер. Свакој виртуелној машини је додељено 2 ЦПУ кора, 6 ГБ РАМ-а и 16 ГБ складиштеног простора. За представнике база података коришћене су Орацле 11г СЕ као Орацле инстанца и ХП Вертица Цоммуниту Едитион као Вертица инстанца. Иста шема база података је импортована на обе инстанце, користећи продукциону базу података путничке агенције као извор података. Иако се за бенчмарк код мерења перформанси *Цолумн-сторе* база обично користи *ТПЦ-Х бенцхмарк* или *Вертица Стар Сцхема Бенцхмарк* (ССБ) који је изведен из ТПЦ-Х, аутори овог рада су одлучили да користе свој бенчмарк. Разлог за такав избор се огледа у значајној разлици карактеристика шеме базе података која је коришћена за експерименте у односу на ону која је прописана горе наведеним бенчмарком. Моделирање сопственог бенчмарка је омогућило ауторима овог истраживања да укључе упите које имају другачију намену у односу на СББ и ТПЦ-Х

бенчмарке. Дизајнирани бенчмарк се састојао од два скупа упита, тзв. генералних упита који мере перформансе/време извршења једноставних операција (*ИНСЕРТ, УПДАТЕ, ДЕЛЕТЕ* и упита за не-аналитичку намену) и аналитичких упита који се користе за мерење различитих типова аналитичких захтева. Резултати мерења су показали да Орацле поседује значајно боље перформансе у односу на Вертица код операција манипулације подацима у бази и упита који су не-аналитичког типа. Међутим, када се разматрају перформансе система при извршењу аналитичких упита, Вертица је показала знатно већу брзину извршења упита од Орацле, тј. перформансе Вертица су биле од 3 до 32 пута боље од Орацле у зависности од типа упита. Имајући у виду резултате истраживања, закључак је да је са аспекта перформанси Вертица бољи избор за *Дата Wарехоусе* решења, док је Орацле бољи избор за традиционалне трансакционе типове оптерећења.

Огромно повећање количине података које је потребно складиштити и анализирати доноси потребу за одговарајућим алатима, техникама и окружењима за *Биг Дата* аналитику. Хадооп и Апацхе Хиве су алати који дизајнирани за рад са *Биг Дата* [12]. Апацхе Хиве представља *Опен Соурце* апликацију за анализирање великих скупова података. Можемо рећи да је овај алат у ствари *Дата Wарехоусе* софтвер који омогућава поједностављене упите над подацима, управља скуповима података у дистрибуираним складиштима и користи Хадооп као основни слој. Хадооп представља *Опен Соурце* имплементацију *МапРедуце* програмског модела који је написан у Јави и омогућава дистрибуирано процесирање великих база података које се налазе на различитим кластерима. Дизајниран је да буде скалабилан и лако се проширује од једног до неколико хиљада сервера, при чему сваки од њих нуди локално процесирање и складиште података. Орацле СQЛ представља традиционалну релациону базу података. На основу информација из литературе, аутори рада су закључили да је више истраживача покушало да изведе *Биг Дата* аналитику користећи традиционалне методе. Ови покушаји су резултирали прилично лошим резултатима са аспекта перформанси, због ограничења меморије. Из тог разлога су одлучили истраже и упореде перформансе Апацхе Хиве дистрибуиране базе и популарног Орацле СУБП. Укупно четири параметра су коришћена за одређивање финалног резултата и то: број редова, разматрање типова упита, величина дата-сетова и поређење перформанси извршења упита између Апацхе Хиве и Орацле СQЛ. Четири дата сета који су коришћена у експерименту су преузета су са *УЦИ Мацхине Леарнинг Репоситору*. Прва два дата сета су садржала мању количину података, за разлику од друга два дата сета, како би се установиле карактеристике обе платформе при обради различите количине података. Аутори су дефинисали 5 различитих упита и тестирали их над сва четири дата сета на Апацхе Хиве и Орацле окружењу и то: упит који враћа јединствене записе, упит сортирања, упит инверзног сортирања, упит са груписањем и упит који враћа максималну вредност. Закључак аутора на основу резултата експеримената је да за процесирање мање количине података Апацхе Хиве

захтева више времена од Орацле базе, док је код великих дата сетова резултат обрнут, у корист Апацхе Хиве.

Истраживање за избор одговарајуће платформе база података на основу анализе перформанси различитих типова база података у контексту података здравственог система је био предмет истраживања у раду [13]. Различити стандарди за електронске здравствене записе као што су ХЛ7, ОпенЕХР и ЦЕН ЕН 13606 су у истраживању били коришћени за креирање објектно оријентисане базе података. До тренутка писања овог рада ниједан од наведених стандарда није коришћен за креирање објектно оријентисане базе података. Аутори су прво извршили анализу трендова у здравственом систему Индије и упоредили перформансе система МуСQЛ базе као релационе базе и Дб4о објектно оријентисане базе за исти узорак података за 100 пацијената. Резултати изведених експеримената су показали да су објектно оријентисане базе боље у поређењу са релационим базама када се разматра време потребно за процесирање података, док са друге стране објектне базе користе већу количину складишног простора у поређењу са релационим базама.

Избором адекватне платформе база података за рад са подацима из области здравствене заштите се баве и аутори у раду [14]. После низа година коришћења релационих база података у здравству, за њихово трајно складиштење размотрене су могуће алтернативе као што су *триплесторес* системи (код којих се ентитети података састоје од дефиниција субјеката, предиката и објектних информација). Резултати истраживања су показали да *триплесторес* системи могу имати веома добре перформансе када се посматра просечно време складиштења, односно читања података из система.

Постоје два начина за мапирање табела дводимензионалне релационе базе података у складиште и то: упис података ред по ред (*пов ориентед*) или упис података колона по колона (*цолумн ориентед*). Традиционалне базе података врше уписе по систему ред по ред. Међутим за неке апликације које су аналитичког типа упис података по систему ред по ред није најбоље решење са аспекта перформанси. Аутори су у раду [15] извршили мерење перформанси ових система и дошли до закључка да је *цолумн ориентед* приступ показао боље резултате перформанси код *Дата Варехоусе* система, док је *пов ориентед* приступ имао боље перформансе код процеса уписа и ажурирања базе података.

На избор базе података утиче и цена, и то не само цена платформе база података већ и цена потребног хардвера. Велике количине података захтевају доста простора на диску. Аутори се у раду [16] баве основним разликама традиционалних *пов ориентед* и *цолумн ориентед* база података и закључују да су за многе апликације које захтевају аналитичке способности базе података, *цолумн ориентед* системи бољи избор јер имају боље карактеристике.

Перформансама *row oriented* и *column oriented* складишта се баве и аутори у раду [17] упоређивањем перформанси комерцијалне *row oriented* базе са различитим конфигурацијама и *column oriented* базе. Утврђено је су перформансе *row oriented* база значајно слабије користећи *Star Schema Benchmark (SSBM)*, као и да постоје значајне разлике у перформансама између ова два система, како на нивоу слоја за извршење упита, тако и на слоју за складиштење података. Наведени бенчмарк представља *Data Warehouse* бенчмарк, који је изведен из ТПЦ-Х бенчмарка.

Иако је је генерално присутно мишљење да је за *Data Warehouse* и ОЛАП процесе најбоље користити *column-store* систем уписа података, аутори су у раду [18] на основу истраживања закључили да се овај тип база може користити и код других апликација које захтевају широке и слабо попуњене табеле, без деградације перформанси. Имајући то у виду, *column-store* базе се могу користити за потребе *Семантиц* Web података, ХМЛ, као и података са ГЕМ шемама.

Релационе базе података су оптимизоване за И/О операције, а приступ хијерархији кеша на систему може имати велики утицај на перформансе базе података. Комерцијалне базе података користе НСМ (*Н-ару Стораге Модулар Стораге Модел*) уместо ДСМ (*Децомпоситион Стораге Модел*) као основни метод за складиштење података из разлога ниже цене за реконструкцију података. Аутору су у раду [19] на основу извршених експеримената утврдили да НСМ модел за складиштење податка може имати негативне ефекте на перформансе кеш меморије.

Када је потребно извршити аналитичке задатке, базе података које користе *Column* оријентисано складиштење података показују боље карактеристике перформанси од традиционалних *Row* оријентисаних база података. Разлог што *Column* оријентисане базе имају много боље перформансе код операција читања лежи у томе што оне приступају само подацима колона које се користе у упитима. Аутори су у раду [20] извршили експерименте над базом ХП Вертица која користи *column-store* записе и МуСQЛ базе података која користи *row store* записе и представља најкоришћенију *Опен Соурце* базу података у употреби. За потребе мерења перформанси оба система коришћени су насумично генерисани подаци од неколико милиона редова. Резултати изведених експеримената су показали да је ХП Вертица имала знатно боље перформансе од трансакционе МуСQЛ базе када се аналитички упити извршавају над великом количином података. Разлика у перформансама се смањује са смањењем количине података.

Енорман раст података у данашњим системима коришћен од стране различитих апликација и социјалних мрежа уводи *цлоуд цомпутинг* као незаобилазан фактор у обезбеђивању како складишног простора, тако и обезбеђивању добрих перформанси система који су задужени за управљање подацима чија структура постаје све комплекснија. Када год разматрамо такве системе, процеси који раде с подацима морају

бити високо паралелни и да се извршавају над различитим системима. Аутори се у раду [21] баве укупним прегледом ове проблематике укључујући анализу захтева за данашње дата генераторе и појашњавају досадашња решења, као и проблеме који до тренутка писања рада нису решени.

Оптимизација CQЛ упита има значајну улогу за перформансе система традиционалних релационих база података. Да би се разумела улога процеса оптимизације упита, аутори су се у раду [22] осврнули како на статичке тако и на динамичке процесе оптимизације упита. Такође, разматрали су и извесна побољшања користећи *Волцано* оптимајзер који се користи код објектно оријентисаних база података.

Једна од метода за побољшање перформанси система за управљање базама података може бити и коришћење метода денормализације. Аутори у раду [23] предлажу коришћење денормализације као међукорака између логичког и физичког моделирања код креирања аналитичких процедура за дизајн критеријума апликативних захтева. Нормализација представља процес груписања атрибута у предефинисаним структурама података. Денормализација се може представити као процес за смањење степена нормализације ради повећања перформанси извршења упита. Једна од најчешћих области примене техника денормализације је у *Дата Варехоусе* имплементацијама за *дата мининг* трансакције. Дате смернице у овом раду су генералне, тако да се могу применити на највећи део база података. Побољшања у перформансама се могу добити само уколико се смернице примене под условом потпуног разумевања апликативних захтева.

Код база података постоји низ параметара који могу утицати на свеукупне перформансе система. Један од тих параметара је *ин-мемору*. *Ин-мемору* базе података чувају све податке у физичкој меморији и због тога имају веома добре карактеристике перформанси и пропусног опсега трансакција у односу на базе које уписују податке на чврсти диск. Група аутора је у раду [24] представила резултате истраживања мерења перформанси ове две врсте система са одговарајућом анализом, користећи ТПЦ-Ц бенчмарк.

Анализом перформанси *Опен Соурце* ОЛТП база МуСQЛ и МариаДБ се бавила група аутора у раду [25]. Резултати истраживања су показали да је у окружењу где су обе базе података инсталиране на Виртуал Манагер Виртуал Мацхине Хен 4.4 и тестиране користећи две врсте различитих оптерећења, МуСQЛ приказао боље перформансе у односу на МариаДБ. Из угла коришћења ресурса обе базе података су користиле исти ниво оптерећења процесора и меморије.

3.2. Поређење перформанси НоСQЛ и релационих система за управљање базама података

Имајући у виду све нижу цену складиштења података као и потребу данашњих апликација да складиште и врше упите над великом количином података, примећује се све већа употреба такозваних НоСQЛ база података, које се користе у концепту *Биг Дата*. НоСQЛ базе су креиране да пруже високе перформансе, у смислу брзине и складиштеног простора, као и високе доступности [26] [27] [28]. Може се рећи да су оне нус-производ web 2.0 ере потребне дизајнерима web сервиса са великим бројем корисника. Данас постоји више од 120 оваквих решења у понуди.

Аутори су у раду [29] извршили поређење НоСQЛ решења са МуСQЛ производом који је представник релационе базе података, како би детектовали шта се добија, а шта губи коришћењем НоСQЛ решења у односу на традиционално решење. Први предузети корак је избор/класификација производа који имају сличну сврху и сличне особине. Селектовани НоСQЛ производи које су аутори изабрали за поређење са МуСQЛ су Цассандра и Хбасе. Из квалитативног угла, аутори су описно приказали поређење следећих карактеристика различитих производа: постојаност, репликација, висока доступност, трансакције, свесност о локацији података, подржани програмски језик, подршка различитих добављача и тип потребне лиценце. Из квантитавног угла, аутори су разматрали поређење карактеристика које се односе на аспекте количине складиштења података и перформанси.

Аспект количине складиштења података: Како су НоСQЛ производи дизајнирани за огромне количине података не постоји смисао њиховог поређења са МуСQЛ базом која у просеку може да садржи 1 милион записа без меморијског кеширања. Преко овог лимита резултат упита постаје сувише спор.

Аспект перформанси: Коришћењем лабораторијског бенчмарка који користи *ЈСЦБ (Yahoo! Цлоуд Сервинг Бенцхмарк)* мерење је извршено на 120 милиона записа мале величине (1кБ), на 6 нода, и 0.12 ТБ еквивалентних инсталација сва три производа, при чему су мерени следећи параметри: *Реад латенцу ин а врите интенсиве енвиронмент*, *Врите латенцу ин а врите интенсиве енвиронмент*, *Реад латенцу ин а реад интенсиве енвиронмент* и *Врите латенцу ин а реад интенсиве енвиронмент*. Иако СQЛ и НоСQЛ базе података деле неке заједничке могућности, оне се ипак не понашају слично у датим окружењима. То указује да се оне не могу користити као замењиве за решавање било које врсте проблема, већ је потребно направити одговарајући избор.

С ‘обзиром на све већу популарност и коришћење НоСQЛ база података, веома често се поставља питање њиховог поређења у смислу карактеристика и перформанси са стандардним релационим базама података. Велики произвођачи релационих база података на тржишту као што су Мицрософт, ИБМ и Орацле су дуги низ година улагали значајна средства у њихов развој, а њихова експлоатација је присутна у свим индустријама. Са друге стране, НоСQЛ базе података су другачије, нису релационе, не морају да користе *qyepy* језике, немају шему, подржавају једноставну и лаку репликацију и омогућавају складиштење огромних количина података. Главна предност НоСQЛ база се огледа у једноставном дизајну, хоризонталној скалабилности и финој контроли доступности података. Цассандра је једна од најпопуларнијих НоСQЛ база података и користи се дуже време за смештање података Фацебоок-а (развијена је од стране Фацебоок-а). Аутори су у раду [30] имали за циљ да упореде перформансе СQЛ Сервера и Цассандра решења, како би установили у ком тренутку би корисници могли да пређу са релационе базе на НоСQЛ базу и обрнуто. За поређење перформанси обе базе коришћене су табеле са по 1,000,000 редова на једном ноду, а експеримент је обухватао три типа операција: *Реад* над пољем *Кеy* у СQЛ Серверу и *Кеy Валуе Паир* у Цассандра бази; *Врите* над табелама у обе базе; као и операцију *Селецт*. Цассандра подржава једноставни *Селецт* упит, али има ограничење да не може да врати више од 15,000 записа. Експеримент је изведен на x64 рачунару са И3 процесором, 4ГБ РАМ и Мицрософт Виндовс 10 Професионал оперативном систему. За Цассандра тест је коришћена ДатаСтах инсталација Цассандра верзије 3.4. За СQЛ сервер тест је коришћен СQЛ Сервер 2012. Тест подаци су генерисани користећи АдвентуреВоркс2012 базу у СQЛ Серверу 2012, тачније креиране су табеле у обе базе користећи СQЛ скрипту за пуњење 1,000,000 редова са 26 колона. Резултати експеримената су показали да је приликом увоза података у базу, СQЛ Сервер показао значајно боље перформансе у односу на Цассандра базу. Цассандра бази је било потребно 321 секунду док је СQЛ Сервер завршио тај процес за само 34.42 секунде. Код извоза података резултат је био сличан претходном, Цассандра је процес извршила за 322 секунде док је СQЛ серверу требало само 20.3 секунди. Сличан однос у резултатима су показали и тестови са *Селецт* упитима, при чему Цассандра има лимит у максималном броју од 15,000 редова које може да прикаже. Ограничење извршених експеримената се огледа у коришћењу само једног ПЦ рачунара у тест окружењу.

Упоредна студија између не-релационе НоСQЛ МонгоДБ базе и релационе Мицрософт СQЛ базе података при обради неструктуираних подата у ХМЛ и ЈСОН формату је била предмет истраживања групе аутора у раду [31]. Фокус аутора је био у истраживању могућности које пружају обе базе података за случај складиштења података који нису нормализовани. То је најчешћи пример који се користи код неструктуираних података социјалних мрежа. Истраживање је базирано на креирању бенчмарк апликације коју су аутори развили у Ц#, користећи Висуал Студио 2013. Имајући у виду рад са подацима

који не могу бити нормализовани, али су и даље повезани на неки начин са подацима у релационој бази података, аутори су изабрали хибридни модел за СQЛ Сервер насупрот модела за складиштењу података у МонгоДБ. У СQЛ серверу су складиштени *ИД* и *хмл* типови података, док су у МонгоДБ подаци складиштени као документи. Циљ рада је да се одговори на питање у којим случајевима је боље користити релационе, а у којим НоСQЛ базе података. Закључак аутора, на основу извршених експеримената је да у случајевима када није могуће извршити нормализацију података и у случају обраде велике количине података, МонгоДБ је боље решење која пружа значајно боље перформансе.

Истраживање поређења перформанси релационих и не-релационих база података, конкретно Орацле и МонгоДБ код извршавања комплексних упита над великом количином података, који су били доступни у документ формату и као такви конвертовани у Орацле табеле, су били предмет рада аутора у раду [32]. За тест податке су коришћени подаци са адресе инфоцхимпс.цом. Преко 500,000 редова из већег броја табела у ЦСВ формату је увезено у Орацле и МонгоДБ. За потребе тестирања аутори су припремили неколико различитих типова упита при чему је коришћен рачунар са Интел® Цоре™2 Дуо ЦПУ Т5750 @ 2.00ГХз, 4 ГБ РАМ, 40 ГБ партицијом и софтверима: Убунту 12.04 64бит, Орацле Датабасе 11г Релеасе 2, Монго ДБ схелл версион: 2.0.9, Убунту 12.10 64-бит. Резултати тестирања су показали да је МонгоДБ приказао боље перформансе код упита враћања резултата, док је подбацио у односу на Орацле код извршења агрегационих функција. То доказује да НоСQЛ базе не могу бити замена за релационе базе, већ да у зависности од захтева које треба испунити, треба изабрати одговарајуће решење.

Група аутора се у раду [33] такође бави поређењем перформанси НоСQЛ базе података МонгоДБ и релационе базе података Орацле. Ради најбољег поређења перформанси обе базе података аутори су разматрали време извршења операција над базом. Први тест је укључивао операцију убацивања података у обе базе података у секвенцама од по 10; 100; 1,000; 10,000; 100,000 и 1,000,000 редова. МонгоДБ-у је било потребно знатно више времена да убаци мање сетове података у односу на Орацле, али је за веће сетове података то радио знатно брже од Орацле. Слични тестови су урађени и за операцију ажурирања података у бази, где је МонгоДБ показао знатно боље карактеристике од Орацле, без обзира на величину сетова података. Сличан резултат у корист МонгоДБ је добијен и код операција брисања података из базе.

Аутори су у раду [34] представили своје истраживање мерењем перформанси над релационом базом података МуСQЛ и НоСQЛ базом МонгоДБ. Експерименти су вршени користећи физички рачунар са Цоре и5, 8ГБ РАМ и ХД 500 ГБ. За софтверско окружење су коришћени Линух Минт 17.1 «Ребецца», Јава «1.8.0_121», Апацхе Мавен 3.2.5, ЧЦСБ 0.8.0, МуСQЛ Сервер 5.5.54, и МонгоДБ 3.2.8. Неколико сетова података од 1,000 до 600,000 редова су коришћени за различите типове и тежине упита, као и ажурирања

података у табелама. Аутори су применили *Yahoo! Цлоуд Сервинг Бенцхмарк* за потребе арбитраже перформанси ова два система. Време извршење упита је био главни параметар који су аутори користили за упоредно мерење перформанси. Код операција читања података се показало да је MyCQЛ добар избор за мање до средње велике базе података. Са порастом количине података перформансе MyCQЛ система брзо опадају. МонгоДБ је показао много боље карактеристике код операција читања у случајевима велике количине података. Код операција уписивања, МонгоДБ је показао значајно боље карактеристике од MyCQЛ. Један од разлога је и то што МонгоДБ не примењује процесе верификација код уписивања података, за разлику од MyCQЛ. Ако се посматра потреба за проширењем база података током експлоатације и уколико се не очекује велики пораст количине података, MyCQЛ може бити добро решење. Међутим, ако су потребе за значајним проширењем база података видљиве, а уједно се разматра *Биг Дата* и *Цлоуд Цомпутинг* окружење са потребом одзива у реалном времену, МонгоДБ представља бољи избор.

Слично истраживање поређења перформанси MyCQЛ и МонгоДБ су се бавили и аутори у раду [35] при чему су за вршење експеримената користили реалну базу података Супермаркет управљачког система. Закључак аутора је да се обе базе података понашају слично у смислу времена извршења упита када се операције врше над мањим сетом података. Међутим, како се количина података над којом треба извршити операције повећава, МонгоДБ показује боље карактеристике перформанси.

Аутори се у раду [36] такође баве поређењем карактеристика MyCQЛ и МонгоДБ. Спроводећи низ експеримената долазе до сличних закључака као у претходном раду. Аутори представљају и могућност интеграције ове две базе користећи *миддлеваре* компоненте између њих, у случајевима када овај приступ може представљати најбоље решење.

Студија поређења перформанси MyCQЛ и МонгоДБ, фокусирана на мерење параметара *тхреадинг* и системских ресурса (процесора и меморије) приликом извршења операција уписивања и читања комплексних типова података је представљена у раду [37]. МонгоДБ је у експериментима приказао већу брзину код комплексних типова и веће количине података, при чему је користио више системских ресурса у односу на MyCQЛ. Овај закључак представља битан параметар у одлучивању о избору базе података, имајући у виду цену изнајмљивања *Цлоуд* ресурса потребних за МонгоДБ, док је MyCQЛ бесплатан. Са друге стране, МонгоДБ је лакши за коришћење од стране програмера у односу на MyCQЛ, јер не захтева знање CQЛ језика, нити коришћење нормализације података. Поред тога, учитавање и извоз података је једноставнији код МонгоДБ.

Група аутора је у раду [38] представила поређење перформанси MyCQЛ и МонгоДБ на основу експеримената извршених над мањом количином података које садрже до 30,000 редова. ПХП *Опен Соурце* програмски језик је коришћен за спајање података из две

табеле користећи МуСQЛ *јоин* упит и идентичну агрегацијску функцију *иннер јоин* у МонгоДБ. Мерени су параметри брзине одговора (*респонсе тиме*) и времена враћања података (*ретриевал тиме*). Експерименти су вршени на једном физичком рачунару са Интел И5 процесором. Резултати експеримената су показали да је у датим случајевима МуСQЛ показао боље перформансе у односу на МонгоДБ, што показује да је код мање количине података МуСQЛ бољи избор, не само по питању перформанси, већ и цене.

Поређењем перформанси МонгоДБ и МуСQЛ базе података се бавио и аутор у раду [39]. Он је вршио испитивања метрика над процесима учитавања мање и веће количине података код обе базе података. При томе је користио *Yahoo! Цлоуд Сервинг Бенцхмарк (ЦСБ)* и три типа теста, од укупно шест која овај бенчмарк дефинише и то: А- *Хеаву упдате* (50% операција читања, 50% операција ажурирања), Ц- *Реад онлу* (100% операција читања) и Ф – *Реад-Модифу-Wритес* (50% операција читања, 50% операције читања, уписа, ажурирања). Тестови су вршени над виртуалним машинама креираним на Амазон *Цлоуд* платформи. На основу извршених експеримената аутор закључује да је МонгоДБ показао боље перформансе од МуСQЛ за сва три типа теста која укључују рад са већом количином података.

Графх НоСQЛ базе података све више добијају на популарности због брзине обраде комплексних упита. Ове базе података користе *траверсал басед* језик упита познат као *Цупхер,* који омогућава корисницима да претражују шаблоне и везе међу неструктуираним подацима. *Графх* базе користе *графх* структуре, нодове и ивице (*едгес*), док релационе базе података користе табеле и релације међу њима. Аутори су у раду [40] представили резултате мерења перформанси *Графх* базе Нео4ј и релационе базе података МуСQЛ. За потребе експеримената су коришћени различити типови упита које се могу сврстати у 4 категорије: селекција-претрага, рекурзија, агрегација и проналажење шаблона у подацима над реалном базом података „Цареер Виллаге“. Резултати тестирања су показали да Нео4ј има боље перформансе у односу на МуСQЛ у највећем броју случајева, а посебно у случајевима проналажења шаблона у подацима и рекурзивним упитима код операција са великом количином података.

Студијом мерења перформанси документ НоСQЛ базе ЦоуцхДБ и релационе базе података МуСQЛ се бавила група аутора у раду [41]. Процена перформанси је базирана на коришћењу ЦРУД операција (*ЦРЕАТЕ, РЕАД, УПДАТЕ, ДЕЛЕТЕ*) над обе базе података. Главни циљ студије се огледао у изради компаративне анализе утицаја обе база података на перформансе апликације која покреће ЦРУД операције. За потребе тестирања аутори су креирали две сличне апликације развијене у Јави, при чему је свака користила одговарајућу базу података. Две структуре података су креиране за ЦоуцхДБ, прва у којој је сваки документ садржао референцу ка другом документу и другу структуру у коме је сваки документ садржао све податке о сваком ентитету. Да би се добили релевантни резултати сваки упит/операција је примењена на опсегу од 1,000 до 1,000,000 елемената.

Прва структура у ЦоуцхДБ је показала мању ефикасност у односу на другу, без обзира на број елемената. МуСQЛ база се показала као ефикаснија у случајевима мањег броја елемената и поред тога што индекси нису креирани над било којом табелом. Са аспекта перформанси, друга структура ЦоуцхДБ је показала боље перформансе у односу на МуСQЛ код *инсерт, селецт* и *делете* операција, поготово код операција са великом количином података.

Када дође до појаве нових база података, креирају се нови бенчмарк тестови и врше тестирања перформанси са ранијим базама података. Најчешће се дешава да резултати мерења перформанси показују да су базе новијег датума брже од традиционалних. Група аутора је у раду [42], на основу истраживања утврдила да ова поређења перформанси не морају увек бити тачна. Они су поново размотрили нека ранија истраживања перформанси Нео4j графх базе података и релационих база података и дошли до могућих узрока који су довели до не-фер закључака о много лошијим перформансама тестираних релационих база података. Аутори су открили да су резултати неких тестова који су извршени на ПостгреСQЛ бази показали велику зависност од фактора као што су конфигурација, шема базе података, фино подешавање и слично. Штавише, аутори су на основу тестирања добили добре резултате перформанси за таква сценарија тестирања која су ранијим истраживањима и тестирањима приказана као лоша за релационе базе података. Препорука аутора је да сваки систем који се тестира треба фино подесити за одговарајуће тестове, па чак и креирати сопствене бенчмарке како би се добили фер резултати.

Данашње потребе за складиштењем велике количине података и фреквентним операцијама над великим сетовима података су све веће. Група аутора је у раду [43] приказала студију поређења перформанси две НоСQЛ базе података, Цассандра и Хбасе и релационе СQЛ базе података Цитус/ПостгреСQЛ. Све три базе података су тестиране у окружењу кластера *Сингле Боард* компјутера (СБЦ). *Сингле Боард* компјутер је компјутер смештен на једној матичној плочи са микропроцесором, меморијом и другим хардверским компонентама које има традиционални рачунар. Кластер се састојао од 15 *Распберру Пи* 3 нодова са *Доцкер Сварм* оркестрационим алатом за примену сервиса и балансером оптерећења над *Сингле Боард* компјутерима. Претпоставка аутора је да овакав кластер може репрезентовати *цлоуд* сервис окружење укључујући скалабилност, еластичност и високу доступност. Аутори су користили *Yахоо Цлоуд Сервинг Бенцхмарк (YЦСБ)* који представља бенчмарк за *цлоуд* базе података и састоји се од неколико различитих тестова оптерећења. Резултати тестирања су показали да постоји компромис између параметара перформанси и репликације. Оба параметра су значајна када се узима у обзир конфигурација дистрибуираних система користећи СБЦ. Цассандра је постигла боље резултате користећи нивое издржљивости подесиве од стране корисника, док су Цитус и Хбасе показале добру издржљивост, али са падом перформанси у случају повећања броја реплика.

Поређењем параметара перформанси НоСQЛ и СQЛ база података за операције читања, уписивања и брисања на *кеу-валуе* типу складишта НоСQЛ база и традиционалних СQЛ база био је предмет рада аутора у раду [44]. Генерално гледано НоСQЛ базе података су оптимизоване за складиштење података које користе *кеу-валуе* систем, док традиционалне СQЛ базе података то нису.

Најбољим избором базе података која би складиштила податке Географског Информационог система (ГИС) се бавила група аутора у раду [45]. Количине података у којој корисници на интернету означавају гео податке стално расту. Са друге стране апликације које користе гео податке имају потребе за већим пропусним опсезима. У оквиру овог истраживања аутори су извршили експерименте над релационом базом ПостгреСQЛ и две НоСQЛ базе МонгоДБ документ базе и Нео4ј графх базе. За потребна тестирања коришћени су ОпенСтреетМап подаци. Све три базе су инсталиране на 64-бит Виндовс окружењу са Интел® цоре ™ и7-7700 ХQ процесором на 2.80 ГХз и 16 ГБ РАМ меморије. Аутори су креирали два једноставна тест упита за селекцију свих корисника који су означили неко место на мапи и за селекцију свих нодова креираних од стране корисника. Резултати тестирања су показали да је МонгоДБ показао најбоље карактеристике перформанси у датој поставци.

Традиционалне релационе базе података се све чешће пореде са новијим системима база података које могу да управљају огромним количинама података и које се генерално могу поделити на НоСQЛ, НеwСQЛ и системе базиране на претрази (*Сеарцх басед сустемс*). НеwСQЛ системи представљају иновиране релационе база података које могу да понуде сличне карактеристике перформанси као НоСQЛ базе за *Онлине Трансацтион Процессинг (*ОЛТП*)* операције читања и писања, при чему обезбеђују главне карактеристике специфичне традиционалним релационим базама података као што су АЦИД (*атомицитy, цонсистенцy, исолатион, дурабилитy*). АЦИД представљају својства трансакција база података која обезбеђују валидност података без обзира на грешке и аномалије које се могу појавити. Група аутора је у раду [46] извршила анализу архитектуре, карактеристика и класификација НеwСQЛ база података за ОЛТП трансакције. Такође је у раду дат и преглед поредивих метрика традиционалних, НоСQЛ и Неw СQЛ база. Свако решење има својих предности и мана. Ако посматрамо метрике перформанси, НоСQЛ и НеwСQЛ имају боље карактеристике, али и веће пропусте у безбедносном аспекту у односу на традиционалне релационе базе података.

Коришћење виртуализационих решења наставља да се увећава у информационим системима. Везом између перформанси база података и система виртуализације која се може користити, тачније утицајем виртуализације типа И на перформансе НеwСQЛ базе података се бави аутор у раду [47]. Разлика између Типа И и ИИ виртуализације је у томе што се тип И извршава директно на хардверу (*баре-метал*) док се код типа ИИ виртуализација извршава на оперативном систему. Исти хардвер је коришћен за

тестирање перформанси окружења без и са виртуализационим софтвером и то: Хewлетт Пацкард Ентерприсе (ХПЕ) Аполло р2600 шасија са 3 ХПЕ дуал-процессор нода. Сваки процесор се састојао од два Интел Хеон Броадwелл Е5-2698в4 2.4ГХз ЦПУ, при чему је сваки имао 14 корова и 50МБ кеша, 128ГБ ДДР4-2400 меморије и један 480МБ САТА *солид-стате дриве* (ССД). Комуникација између нодова се одвијала помоћу 1 ГбЕ *Етхернет* портова. За окружење са виртуализацијом је коришћен ВМwаре ЕСХи виртуализациони софтвер. Мерене су следеће метрике: коришћење процесора, меморије, диск и мрежни улаз/излаз као и број трансакција у секунди иницираних од стране базе података. Сваки нод је имао исти оперативни систем ЦентОС 7.7.1908, *Опен Соурце* верзију Ред Хат Ентерприсе Линух. Два нода су коришћена за кластер базе података НуоДБ НewCQЛ релационе базе података, на којима се налазио и софтвер за мерење параметара перформанси. Трећи нод је коришћен за покретање скрипти оптерећења. Треба напоменути и да је НуоДБ база података дизајнирана за рад на *баре-метал* конфигурацији, виртуализационом и *цлоуд* окружењу. Резултати тестирања су показали да је степен оптерећења процесора увећан за преко 60% код окружења са виртуализацијом. Коришћење меморије је дискретно увећано код окружења са виртуализацијом у опсегу између 1.4% и 2.7%. Диск И/О операције су показале увећане вредности код виртуализационог окружења за више од 30%. Наведени резултати директно указују да виртуализационо окружење проузрокује додатно оптерећење на систем.

Група аутора је у раду [48] је представила подршку ЈСОН *дата манагемент* операцијама у верзији Орацле базе 12цР1 која омогућава корисницима да складиште, индексирају податке и врше упите користећи ЈСОН дата модел. На тај начин је омогућен рад са подацима у релационој бази података без претходног дефинисања шеме, што је карактеристика НоСQЛ база података. Аутори су такође представили бинарни формат упита за декодирање ЈСОН дата модела који смањује разлику у перформансама између шема кодираних релационих података и ЈСОН текстуалних података без шеме, које се користе у НоСQЛ базама. Са додавањем ових нових могућности Орацле релациона база података је позиционирана као добро решење, како за случајеве који захтевају фиксну СQЛ шему, тако и за флексибилну шему која се користи код НоСQЛ приступа.

Ако посматрамо *Дата Wарехоусе* решења за потребе извештавања користећи све веће количине података којим данашње пословање располаже, можемо закључити да она могу бити веома комплексна и да им је за рад потребно све више времена. Аутори су у раду [49] извршили поређење и анализу перформанси база података за *Биг Дата Wарехоусе* потребе. Тестиране су две базе података, МуСQЛ кластер који припада традиционалној СQЛ бази података и Хиве над Хадооп-ом који припада НоСQЛ бази података. За потребе тестирања аутори су припремили неколико различитих сценарија оптерећења. За процену перформанси је коришћен *Стар Сцхема Бенцхмарк (ССБ)*. Резултати експеримената су показали слабе перформансе МуСQЛ кластера за велике сетове података и поред

могућности коришћења додатних слободних ресурса, што доводи до закључка да је ово решење најбоље користити за трансакционе системе. Хиве је са друге стране показао скалабилне резултате, поготово када је у могућности да користи додатне ресурсе, што га позиционира као добро решење за огромне количине података.

Користећи постојећа истраживања других аутора која су била доступна до тренутка објаве рада [50], група аутора је извршила поређење карактеристика СQЛ и НоСQЛ база података, поређењем 10 различитих карактеристика, почевши од цене и скалабилности до коегзистентности и безбедности. Закључак аутора је да традиционалне релационе базе података представљају најбољи избор у случајевима коришћења лимитиране количине структуираних података за разлику од НоСQЛ база података, које карактерише велика скалабилност и добре перформансе код операција над великом количином података, са освртом на ограничене могућности са аспекта безбедности.

Избором одговарајућих решења користећи истраживања других аутора у сегменту поређења карактеристика СQЛ и НоСQЛ база података као и поређењем *Цлоуд датабасе* система се баве аутори у раду [51]. На основу представљених анализа аутори препоручују данашњем пословању следећа решења:

- НоСQЛ базе података су добар избор за највећи део најзахтевнијих операција које оперишу са огромним количинама података. Оне су брзе, скалабилне и поуздане.
- Амазон РДС (МуСQЛ) нуди више конфигурационих опција, добре перформансе и добра су препорука за већину организација.
- Мање фирме би требало да користе Амазон СимплеДБ због добрих карактеристика са мањом количином података и повољне цене
- *Фиребасе Реалтиме* базе података (НоСQЛ) су најбољи избор за личну употребу или за групу корисника, поготово када није потребно приступити бази података по принципу 24/7 и када је потребно користити једноставно складиште података.

У данашњем пословању ефикасност информационих система представља основни предуслов и најчешће се мери временом враћања резултата упита, бројем конкурентских корисничких сесија које може да обради, као и нивоом енергетске ефикасности система. На оптимизацију информационог система се веома често гледа као на комплексан и неопходан задатак. Порастом количине података може доћи до пада перформанси система и обично се на систем база података гледа као на систем чије карактеристике перформанси треба измерити, а на основу резултата исте унапредити. Аутор у раду [52] даје систематичан преглед анализе 117 различитих истраживања које су се бавиле перформансама различитих традиционалних релационих и НоСQЛ база података.

Детаљном анализом издате литературе које се баве поређењем софтверске архитектуре, карактеристика и перформанси СQЛ и НоСQЛ база података се бави и група аутора у раду [53]. Закључак аутора на основу истраживања је да нема потребе за транзицијом традиционалних релационих података на НоСQЛ базе података. Свака база података има својих предности и мана, те да треба направити одговарајући избор на основу захтева које треба испунити. Генерално посматрано, ако организација треба да задовољи захтеве за стандардизацијом и коегзистенцијом података онда је најбољи избор релациона база података. НоСQЛ базе су најбољи избор код велике количине неструктуираних података.

У тренутку објаве рада [54], преко 200 различитих НоСQЛ база података је било расположиво за коришћење обезбеђујући различите карактеристике и функционалности. Аутори су у овом раду представили истраживање мерења перформанси неколико најчешће коришћених *ин-мемору* НоСQЛ база података: МонгоДБ, Редис, Мемцасхед, Цассандра и Х2 релационе базе података. Х2 се разликује од осталих релационих база података по томе што чува податке у меморији уместо на хард диску. За потребе тестирања перформанси наведених база аутори су развили специјализовани софтвер у Јави за различите експерименте и то: уписивање парова кључева, читање вредности за одговарајући кључ, брисање парова кључева и учитавање свих података. Сви експерименти су вршени на истом физичком рачунару са Интел и7 процесором и 16ГБ РАМ. Тестови су вршени над базама података са различитим бројевима записа, од 1,000 до 1,000,000. Закључак аутора је да СQЛ и НоСQЛ базе података поседују различите карактеристике и да оне нису замена једне за другу. Уколико је захтев за коегзистентност података у бази нефлексибилан, онда је релациона база података најбољи избор. У супротном, НоСQЛ базе података су бољи избор због обезбеђивања високе доступности, скалабилности и високих перформанси.

3.3. Поређење перформанси различитих НоСQЛ система за управљање базама података

Повећањем количине података на Интернету као и на основу захтева web 2.0 базираних апликација НоСQЛ базе података постају опште прихватљив избор за складиштење и управљање подацима. Аутори у раду [55] представљају дата моделе НоСQЛ база података, могућности упита који се могу над њима постављати, механизме за контролу конкурентног приступа подацима, начине партиционисања података који се дистрибуирају на вишеструке нодове, као и репликационе карактеристике. Имајући у виду све већи избор НоСQЛ база података које нуде различите карактеристике и могућности, увек је потребно извршити одговарајућу процену најбољег решења сходно захтевима који се требају испунити.

Група аутора се у раду [56] бави разлозима за транзицију са традиционалних релационих података на НоСQЛ базе података и анализира перформансе три најпопуларнија НоСQЛ решења: Цассандра, МонгоДБ и ЦоуцхБасе. Цассандра је представник *Кеy-Валуе* базе, оптимизована је за операције уписивања података, нема централни нод и подаци се могу читати или уписивати на било који нод у кластеру. Резистентна је на пад било ког нода у кластеру тиме што ће други нод преузети оптерећење. Карактерише је хоризонтална скалабилност. Цассандра је Апацхе 2.0 лиценцирано решење. МонгоДБ је документ оријентисана НоСQЛ база података која не користи шему и податке уписује у БСОН/ЈСОН бинарном формату, који пружа могућност једноставне интеграције података са најчешће коришћеним апликацијама. Карактерише је хоризонтална скалабилност и отпорна је на пад било ког нода у кластеру. МонгоДБ поседује централни (*мастер*) нод и вишеструке потчињене (*славе*) нодове. Ова база података је *Опен Соурце* решење лиценцирано од стране компаније 10.ген. ЦоуцхБасе је *Опен Соурце* документ оријентисана НоСQЛ база података. Намењена је раду са интерактивним web и мобилним апликацијама. Подаци се складиште у ЈСОН формату и база је отпорна на пад појединачних нодова. Карактерише је хоризонтална скалабилност и лиценцирана је од стране Апацхе 2.0. На основу различитих тестова оптерећења над све три базе, резултати експеримената су показали да МонгоДБ чита податке из базе брже од Цассандра базе, али спорије од ЦоуцхБасе. Цассандра и ЦоуцхБасе брже уписују податке у односу на МонгоДБ. Општи закључак је да избор једног решења није увек једноставан процес, имајући у виду да је свако од наведених решења оптимизовано за различите типове оптерећења и захтева.

Поређењем перформанси 5 НоСQЛ база података: Редис, *Кеy-валуе сторе*, верзије 3.2.1.; МонгоДБ *Доцумент Сторе* верзије 3.2.7; ЦоуцхБасе *Доцумент Сторе*, верзије 4.0.0;

Цассандра *Цолумн фамилу Сторе*, верзије 2.2.7 и ХБасе *Цолумн фамилу Сторе*, верзије 0.98.20 су се бавили аутори у раду [57]. Коришћен је *Yahoo! Цлоуд Сервинг Бенцхмарк (ЦСБ)* који ради у две фазе, учитавање података и тестирање трансакција. Приликом тестирања се може фино подешавати тежина оптерећења користећи однос писања и читања, као и дефинисање опсега података у фази тестирања. Аутори су изабрали три стандардна теста оптерећења овог бенчмарка, означена словним ознакама А,Ц и Х над случајно генерисаним подацима у свим базама података. Закључак изведен на основу резултата тестова указује да је Редис најбоље позициониран код учитавања података и извршења операција над подацима, мада су видљива ограничења у раду са екстремно великом количином података. *Доцумент* и *цолумн-фамилу* НоСQЛ базе су показале добре просечне перформансе.

Детаљном студијом поређења карактеристика и перформанси НоСQЛ база података се бавила група аутора у раду [58]. Тачније, три НоСQЛ базе података су биле предмет истраживања и то ЦоуцхБасе, МонгоДБ и РетхинкДБ користећи два сценарија, једноструких и вишеструких нити (тхреадс). Резултати експеримената су показали да ЦоуцхБасе има боље перформансе у скоро свим извршеним операцијама, изузев код операција учитавања и враћања скупа докумената са вишеструким нитима, где је МонгоДБ показао боље резултате.

Постоје различити примери у многобројним индустријама као и на интернету који захтевају складиштење и рад са сликама, као што су анализа саобраћаја, медицински записи, социјалне мреже и др. Група аутора је у раду [59] представила истраживање поређења перформанси документ оријентисаних НоСQЛ база, МонгоДБ и ЦоуцхБасе у случајевима уписивања и читања слика различитих величина из базе података. Код једноставних операција читања и писања слика мањих дата сетова није утврђена велика разлика. У случајевима операција уписа веће количине слика, МонгоДБ је захтевао више времена у односу на ЦоуцхБасе, док је код операција читања слика резултат био супротан, Монго ДБ је радио брже у односу на ЦоуцхБасе.

НоСQЛ базе података обезбеђују отпорност на пад нодова у кластеру као и високу доступност. Ради обезбеђивања ових особина користе се репликациони механизми. Аутори су у раду [60] представили студију утицаја репликације на перформансе НоСQЛ база података. НоСQЛ базе података користе одговарајуће протоколе за испуњење коегзистентности/доследности (*цонсистенцу*) података у процесу репликације на друге нодове у кластеру. Ови протоколи контролишу како се подаци преносе између нодова, како се рукује операцијама читања и писања, када се копије података на другим нодовима ажурирају, као и који је задовољавајући ниво коегзистенције података. Код механизма репликације користе се следећи параметри: репликациони фактор, параметар који дефинише број нодова на који се један податак реплицира и ниво коегзистенције, који дефинише колики број реплика/нодова мора да одговори на долазећи упит како би се

сматрао валидним. Ови параметри могу значајно да утичу на перформансе рада базе података. Нижи нивои коегзистентности доводи до бољих перформанси и обрнуто. Две НоСQЛ базе су биле у фокусу ове студије, Цассандра и МонгоДБ. Аутори су користили ЧЦСБ бенчмарк за извођење експеримената над окружењем које је постављено на *Привате Цлоуд* инфраструктури. Коришћени су тестови високог оптерећења читања и писања. Резултати експеримената су показали да је са аспекта пропустности података код операција интензивног читања МонгоДБ показао за нијансу боље резултате од Цассандра базе, док је код операција интензивног писања Цассандра надјачала перформансе Монго ДБ за више од 70 процената. На тај начин је потврђено да је Цассандра оптимизована за операције интензивног уписа података. Са аспекта утицаја репликације на перформансе, утврђено је да репликациони модел *Мастер-Славе* МонгоДБ има мањи утицај на перформансе пропустности података у односу на *Мулти-Мастер* репликациони модел Цассандра базе.

Код НоСQЛ база података, параметри перформанси, доступности и коегзистентности података су веома повезани. Да би утврдили ниво повезаности ових параметара аутори су у раду [61] представили истраживање утицаја различитих подешавања коегзистентности на скалабилност, латентност (кашњење) и пропусни опсег (*тхроугхпут*) три реплицирана Цассандра кластер окружења на Амазон Веб Сервису (AWC), подешена према различитим сценаријима. Резултати истраживања су потврдили претходна очекивања аутора, да повећана коегзистентност података резултира смањењем пропусног опсега базе и повећањем кашњења операција читања и писања. Најбоље перформансе за сва постављена подешавања коегзистентности се добијају када се Цассандра кластер налази у једном дата центру.

Иста група аутора је истраживала перформансе операција читања и писања на Цассандра кластеру који је постављен на Амазон ЕЦ2 Цлоуд-у [62]. Један од најзначајнијих сазнања добијен експериментима је да је могуће смањити кашњења у операцијама Цассандра базе оптималним подешавањима параметра коегзистентности за обе операције, читања и писања. Високи параметри коегзистентности података утичу на пад перформанси за око 25%, док најбоља подешавања коегзистентности зависе од односа операција читања и писања. Аутори у свом раду предлажу методологију бенчмарка за оптимизацију подешавања коегзистентности, како би се добиле максималне перформансе Цассандра базе, а уједно и обезбедила добра коегзистентност података код обе врсте операција.

Група аутора је у раду [63] представила истраживање утицаја улоге апликационог *тимеоут-а* (временског ограничења) као главне детерминанте у релацији између системске доступности и перформанси које се огледају у одзиву сервиса. Апликациони *тимеоут* представља главни показатељ грешака или преоптерећености које се могу јавити у раду сервиса и управо је он иницијатор операција за опоравак у случају отказа. Архитекте система би требале да имају могућност подешавања жељеног времена

одговора система, водећи рачуна о одговарајућем односу између коегзистентности података и њихове доступности. Међутим, многи дистрибуирани системи којима припада и Цассандра користе статичка *тимеоут* подешавања која често могу бити превише велика, доводећи у питање ефикасну детекцију грешака и иницијацију механизама за управљање отказима. Препорука аутора на основу извршених експеримената је да оптимални *тимеоут* треба да буде динамички и да зависи од структуре базе, величине и односа операција читања и писања. Такође, потребно је укључити и факторе као што су тренутно оптерећење система, број реплика као и параметре коегзистентности.

Истраживање НоСQЛ база података, МонгоДБ (*доцумент сторе*), Цассандра (*цолумн-сторе*) и Редис (*кеу-валуе сторе*) са фокусом на поређење њихових перформанси користећи ЧЦСБ бенчмарк је био предмет студије аутора у раду [64]. Извршено је тестирање наведених база података користећи шест различитих типова оптерећења над сетовима података различите величине. На основу извршених експеримената Редис је приказао најбоље перформансе читања у односу на остале две базе. Разлог оваквих перформанси Редис-а се огледа у могућности коришћења променљиве меморије за смештање и враћање података. МонгоДБ је био бржи у операцијама читања од Цассандра базе. Са друге стране, МонгоДБ је показао боље перформансе у операцијама скенирања од Редис-а и Цассандра базе.

Са порастом коришћења различитих сервиса и апликација на Интернету видљив је и пораст мултимедијалне комуникације корисника користећи мобилне уређаје. Потреба за мултимедијалном комуникацијом је видљива у свим индустријама као и на апликацијама социјалних мрежа. За обезбеђивање и контролу мрежне мултимедијалне комуникације се користи *ИП Мултимедиа Субсустем* (ИМС) архитектура, а у оквиру ње *Сессион Инитиатион Протоцол* (СИП) који управља успостављањем, одржавањем и прекидањем мултимедијалних позива између два уређаја. Мноштво параметара мултимедијалних сесија СИП протокол смешта у базу података која може бити локална, на истом серверу где се налази и СИП сервис. Како је мултимедијална комуникација сервис у реалном времену, СИП захтева високе перформансе свих зависних компоненти, како не би дошло до деградације самог сервиса. Из тог разлога су *Ин-мемору* НоСQЛ базе података алтернативно решење за бржи рад са подацима, који се могу чувати у меморији. Група аутора је у раду [65] представила студију поређења перформанси СИП сервера (ОпенСИПС) користећи три *кеу-валуе* НоСQЛ базе података, локалну базу на ОпенСИПС серверу, Мемцацхед и Редис. У тестовима су вршена мерења параметара меморије, пропусног опсега и време одговора на позив. Резултати тестирања су показали да локална база на ОпенСИПС серверу користи мање меморије у односу на друге две базе. Са аспекта пропусног опсега најбољи резултат је приказала такође локална база. Закључак аутора је да је локална база на ОпенСИПс серверу најбољи избор из перспективе перформанси, уколико се не захтева да се подаци чувају у меморији дуже

времена након затварање сесије која их је креирала. Уколико се ова карактеристика захтева, онда је Мемцацхед база података одличан избор.

Годинама уназад су релационе базе података биле избор за смештање података клиничких истраживања у области здравства. Међутим, због веће количине података и све већом потребом за бољим перформансама, скалабилношћу, високом доступношћу НоСQЛ базе података су постале предмет разматрања као потенцијалног решења. Група аутора је у раду [66] представила Орацле НоСQЛ базу података као избор решења у својој студији.

Група аутора је у раду [67] представила истраживање мерења перформанси Хбасе и Цассандра НоСQЛ база података, користећи Јава базирани алат ЏЦСБ за покретање различитих тестова оптерећења на сетовима података од 100,000 до 250,000 записа. За потребе тестирања и анализе је коришћен *Yahoo Цлоуд Сервицинг Бенцхмарк*. Окружење обе базе података је постављено на виртуелној инстанци са 2 ВЦПУ и 8ГБ меморије. За изабране тестове читања и ажурирања података у бази, Хбасе је показао боље перформансе у односу на Цассандра.

На избор одговарајуће НоСQЛ базе података за неку апликацију утичу многи параметри, од којих један од најзначајнијих могу бити перформансе код извођења такозваних ЦРУД операција (*ЦРЕАТЕ, РЕАД, УПДАТЕ, ДЕЛЕТЕ*). Група аутора је у раду [68] представила истраживање две најчешће коришћене документ НоСQЛ базе података, МонгоДБ и документ МуСQЛ у смислу измерених перформанси при извођења ЦРУД операција. Обе базе података су добар избор за *Биг Дата* апликације са одличним временом извршења операција, без обзира на њихову комплексност и величину обухваћених сетова података. Предност документ МуСQЛ-а се огледа у једноставности и лакшем коришћењу, поготово за оне кориснике који имају знања из области релационих база података.

Мерење и поређење перформанси МонгДБ и ЦоуцхДБ база података је било предмет истраживања аутора презентованог у раду [69]. Закључак студије је да уколико база података треба испунити услов веће ефикасности и брзине рада, онда је МонгоДБ бољи избор у односу на ЦоуцхДБ. Међутим, у случају да је потребно обезбедити мулти мастер репликацију базе, као и подршку за мобилне уређаје онда је ЦоуцхДБ бољи избор.

Упоредна анализа поређења перформанси шест НоСQЛ база података је представљена у раду [70]. Аутори су користили *Yahoo! Цлоуд Сервинг Бенцхмарк* ради изведбе експеримената над МонгоДБ, Цассандра, ХБасе, Редис, ЦоуцхБасе и ОриентДБ базама података. Тестирања су вршена користећи шест различитих типова оптерећења дефинисана наведеним бенчмарком и различитим сетовима података на физичком рачунару са Цоре и5, 8ГБ РАМ, ХД 1 ТБ и следећим софтверским окружењима: Убунту 16.04, 64-бит, МонгоДБ 3.2.8, ЦоуцхБасе 2.5.2, Цассандра 2.2.5, ХБасе 0.94.8, Редис 3.0.2 и ОриентДБ 2.1.3. Неки од кључних закључака студије су да је ЦоуцхБасе показао највећу брзину учитавања великих сетова података у односу на друге базе података. Хбасе је

демонстрирао најбоље перформансе код уписивања нових записа у базу, док су Цассандра и Хбасе демонстрирали најбоље перформансе ажурирања постојећих записа у односу на остала решења.

3.4. Преглед, могућности и перформансе ЕТЛ алата

Дата Варехоусе или ДВ представља аналитичку базу података (ОЛАП - *Онлине Аналутицал Процессинг*) која се користи за потребе пословне аналитике, утврђивање трендова и извештавања, користећи различите изворе података. Главни циљ ДВ је пружање подршке пословним процесима, анализама и процесима доношења пословних одлука на основу консолидованих историјских података [71]. Дизајнирање ДВ система представља комплексан пројекат и захтева добро познавање како пословних процеса и података везаних за те процесе, тако и техничка знања потребна за израду решења које ће испунити очекивања корисника и оправдати инвестицију [72]. *Бусинесс Интеллигенце* или БИ представља скуп технологија, алата, процеса и пракси које помажу организацијама да анализирају и претворе податке у корисне информације. Циљ пословне интелигенције је омогућавање доносиоцима одлука у организацијама да стекну дубљи увид у пословање и донесу најбоље одлуке на основу анализе историјских података и добро дизајнираног БИ решења [73] [74] [75].

Један од важних процеса при креирању напредног аналитичког решења је пуњење ДВ система подацима из различитих извора података. Извори података најчешће имају различите структуре и различите шеме, те их је потребно трансформисати како би одговарали потребној шеми и структури ДВ система. Подаци на извору могу бити упитни са аспекта квалитета, те се често дешава да неки атрибути недостају, да неки нису коегзистентни и слично, те је пре пуњења ДВ базе потребно применити технике и процесе “чишћења” података. Поред тога, подаци се у продукционој трансакционој бази које представља извор података константно ажурирају и допуњују, те је потребно константно податке у ДВ систему синхронизовати са подацима у бази извора. Наведени изазови се решавају адекватним процесима и алатима који се покрећу у одговарајућим интервалима и називају се ЕТЛ (*Ехтрацтион-Трансформатион-Лоадинг*) процесима и алатима. ЕТЛ процес у основи представља скуп задатака (*таскс*) који врше екстракцију података из изворне базе, трансформацију, чишћење и учитавање података у ДВ базу [76].

Иако се ДВ решења користе од почетка 1990-их година, дизајн и моделирање ЕТЛ процеса се и даље извршава у алатима који су зависни од платформе и шеме у којој се налази дестинациона база података. Аутор је у раду [77] представио први моделинг приступ за спецификацију ЕТЛ процеса на начин који није зависан од добављача ДВ базе и који аутоматизује генерисање одговарајућег програмског кода на комерцијалним ЕТЛ платформама.

Од квалитета података зависи и квалитет аналитичког решења. Група аутора је у раду [78] представила истраживање најчешће коришћених алата за ”чишћење” података. Они су све изазове везане за квалитет података поделили на тзв. проблеме на нивоу шеме података и на нивоу инстанце. Проблеми који постоје на нивоу шеме се могу решити побољшањем дизајна шеме и интеграцијом података, док проблеми на нивоу инстанце могу бити повезани са засебним записима или вишеструким записима. Најчешће се то дешава у случајевима као што су недостатак валидног податка у *Нот Нулл* пољу (случај када се поље попуњава насумично, како би се избегло појављивање грешака приликом уписа записа у базу) или синтаксне грешке оператера који уноси податак у поље и слично. У зависности од типа задатака које расположиви алати обављају, аутори су их поделили у шест група: алати за профилисање података, алати за анализу података, алати за трансформацију података, алати за чишћење података, алати за елиминисање дупликата и алати за обогаћивање података. Анализом и прегледом ЕТЛ алата се бавила и група аутора у раду [79], представљајући упоредну студију комерцијалних алата и њихове карактеристике у односу на димензије квалитета података, као што су дупликација, комплетност, тачност, концизност, доследност. Најчешће коришћени комерцијални алати представљени овој студији су: Таленд Дата Интегратион, Пентахо Дата Интегратион, Мицрософт СQЛ Сервер Интегратион Сервицес и Информатица Дата Интегратион.

Постоји значајан број апликација које су развијене за извршење ЕТЛ процеса. Избор најбољег алата за одговарајуће захтеве је увек изазов и избор зависи од његове ефикасности, перформанси, цене и других карактеристика. Аутори су у раду [80] представили студију упоредне анализе ЕТЛ процеса користећи два ЕТЛ алата, Мицрософт СQЛ Сервер 2008 Интегратион Сервице (ССИС) и Пентахо Дата Интегратион (ПДИ). За извор података је коришћена Нортхwинд база. Сценарио за извођење експеримената на СQЛ серверу се састојао од:

- екстракције података из Нортхwинд базе постављене на СQЛ серверу,
- процеса трансформације података, како би се испунили дати захтеви за учитавање података у ДW базу на СQЛ серверу и
- учитавање података у ДW користећи дата импорт функционалност СQЛ сервера.

Сценарио за извођење експеримената користећи Пентахо се састојао од:

- екстракције података из Нортхwинд базе постављене на СQЛ серверу,
- процеса трансформације података у МуСQЛ формат података, како би се испунили дати захтеви за учитавање података у ДW базу на СQЛ серверу и
- учитавање података у ДW користећи дата импорт функционалност Пентахо Дата Интегратион.

Резултати експеримената су показали да је интеграција података користећи ССИС компликованија и временски захтевнија у односу на ПДИ, те да је у ПДИ бољи избор.

Још једна група аутора се бавила поређењем карактеристика Мицрософт СQЛ Сервер Интегратион Сервице (ССИС) и Пентахо Дата Интегратион (ПДИ) ЕТЛ алата [81]. Оба алата представљају робусна решења. ССИС нуди више могућности кроз конфигурације објеката мада је код комплекснијих операција потребно извршити кодирање. Предност и снага ССИС лежи у једноставним контролним токовима, токовима података и архитектури која је базирана на догађајима, што програмерима нуди високу флексибилност дизајна ЕТЛ процеса. ПДИ са друге стране нуди више опција за повезивање и увоз података из Гоогле Аналутицс, као и додатне опције за приступ веб сервисима. Може се користи како на Виндовс, тако и на Линух оперативним системима. Генерално посматрано, ССИС задовољава потребе великих организација и ради са великим количинама података. Уколико посматрамо *Опен Соурце/бесплатне* алате, ПДИ се може користити код мањих организација и у тим случајевима може показати добре перформансе.

Најчешћи је случај да све модерне базе података имају уграђене алате које омогућавају брзи извоз и увоз података у базу. Група аутора је у раду [82] извршила истраживање поређења перформанси тих алата код неколико база података које се користе на тржишту и то: ХП Вертица, МариаДБ Целумнсторе, Орацле и МС СQЛ Сервер. За потребе извођења експеримената коришћене су табеле са подацима Вмарт шеме. Вмарт је *Дата Варехоусе Дата Март* који се користи као пример шеме у Вертица бази података. Коришћено је укупно 7 табела са подацима у распону од 10,000 до 5,000,000 редова у табели. Експерименти су изведени на Виндовс 10 хост рачунару са Интел И7-6700ХQ процесором на 2.60 ГХз и ВМваре виртуелним машинама (једна виртуална машина по бази података). Свакој ВМ је додељено по 6ГБ РАМ. За потребе креирања табела и релација Вмарт шеме, на свакој виртуелној машини је коришћења СQЛ скрипта са одговарајућим командама. За сваку базу података у наведеној студији припремљене су и извршене скрипте за уклањање и успостављање ограничења података (*дата цонстраинтс*), као што су "*Цхецк*", "*Фореиген Кеу*", "*Примару Кеу*". Циљ употреба ових скрипти је да мерено време за извршење тестираних операција не укључи и извођење операције ограничења података. Такође, за потребе тестирања перформанси на нивоу сваке базе података су припремљене скрипте за тестове које извозе податке из табела у ЦСВ фајлове, као и скрипте које увозе податке из ЦСВ скрипти у базу података. На основу резултата извршених тестова на нивоу различитих платформи, утврђена је врло мала разлика у брзини извођења операција извоза података из базе у фајл и увоза података из фајла у базу података користећи уграђене алате. Ипак, најбоље време пуњења базе са подацима из највећих табела је постигао Орацле СQЛ Лоадер, док је најбоље време извоза података постигнут наредбом "*Цопу*" у ХП Вертица.

ДW концепт се огледа у складиштењу екстракованих, агрегираних и филтрираних података у мета- базама података које пружају могућност рада са вишеслојним, вишедимензионалним скуповима података неопходним за извршење одговарајућих анализа и доношење одлука. За потребе студије, група аутора је у раду [83] користила ОЛТП базу података "Дата Леарнинг Сустем" (ДЛС) оригинално развијену од стране аутора за потребе средње техничке школе, како би припремила одговарајуће ЕТЛ процесе за учитавање података у ДW базу података. Аутори су дизајнирали ЕТЛ процес који се састојао од припреме (брисање ограничења спољних кључева у ДW бази, брисање свих података ДW базе), попуњавања ДW димензионих табела подацима из ДЛС базе и других извора, попуњавања ДW табела чињеница подацима из ДЛС базе и других извора, и на крају враћања ограничења спољних кључева у ДW табеле. За извођење ЕТЛ процеса коришћен је СQЛ Сервер Интегратион Сервицес. Закључак аутора је да се представљено ЕТЛ решење може применити код ОЛТП база података које не садрже велику количину података и када постоји потреба да се прво уклоне сви подаци из ДW базе података, а затим у ДW изврши учитавање нових података.

Група аутора је у раду [84] извршила поређење различитих карактеристика популарних ЕТЛ алата и то Информатица, Датастаге, АбИнитио, Орацле Дата Интегратор и Мицрософт ССИС и извршила анализу њихових предности и мана. Такође, аутори су анализирали и неке додатне карактеристике као што су оптимизација перформанси, анализа података у реалном времену и представили их у оквиру своје студије. Слична анализа карактеристика различитих ЕТЛ алата, укључујући поред неких од претходно наведених и ИБМ Информатион Сервер, Таленд Опен Студио, САС Дата Интегратион Студио, Пентахо и Цловер ЕТЛ објављена је у раду [85].

Поглед на одговарајући избор и имплементацију ЕТЛ процеса при изради ДW решења, користећи реалну студију случаја за агенцију за пољопривреду Индонезије су представили аутори у раду [86].

Иако је ЕТЛ незаобилазан алат у пројектима израде ДW, потребно је имати на уму да нису сви ЕТЛ софтверски алати исти. Најбољи ЕТЛ алат треба изабрати на основу компоненти окружења које учествују у пројекту и примерима случаја за које ће бити коришћени. Група аутора је у раду новијег датума [87] представила преглед и компаративну анализу доступних ЕТЛ алата. ЕТЛ алати су подељени по категоријама и заједничким карактеристикама, што их чини лакшим за одабир на основу захтева које требају испунити. ЕТЛ алати су представљени из угла индустрије или типа организације за које се могу примењивати, као и за најоптималнију величину организације. На крају, на избор одговарајућег алата утиче и његова цена те су аутори у своју анализу укључили и тај аспект.

У *Биг Дата* ери задатак ЕТЛ алата је да процесирају велике количине неструктуираних и структуираних података ради комплексних анализа пословања. За стандардне ЕТЛ алате је то велики изазов, те је разматрање њиховог унапређења неопходно ради оправдања инвестиције пословања. Једна од идеја је дизајнирање уређаја са ниско буџетним компонентама које карактеришу адекватни компјутерски и складишни параметри за извођење ЕТЛ процеса. Група аутора у раду [88] предлаже Хадооп на иСЦСИ преко *Етхернет адаптед Нетwорк Аттацхед Стораге* (НАС) као решење за извођење ЕТЛ операција на основу студије поређења перформанси екстракције података, перформанси трансформације и перформанси учитавања података са стандардним ХДФС фајл системом (*Хадооп Дистрибутед Филе Сустем*). На основу резултата експеримената, предложена архитектура са ниско буџетним компонентама је показала скалабилне перформансе за операције са подацима у *Биг Дата* сегменту.

ЕТЛ представља најзахтевнију компоненту аналитичког решења и према ауторима у публикацији [89] на ЕТЛ одлази од 60% до 80% времена комплетног ДW и БИ пројекта. Са порастом броја извора података и количине података стандардни приступ у раду са ЕТЛ алатима који су се заснивали на додавању хардвера, како би се добиле боље перформансе, више није опција. Група аутора је у у својој студији [90] представила поређење перформанси комерцијалних ЕТЛ алата и Хадооп алата. Хадооп је *Опен Соурце* Апацхе пројекат. Анализа добијених резултата перформанси је показала да су Хадооп решења боља у поређењу са комерцијалним ЕТЛ алатима за рад са великом количином података.

Поређење ЕТЛ техника представља веома захтеван задатак код кога постоје бројни изазови које треба имати у виду, како би се добили што тачнији резултати. Аутори у раду [91] представљају упоредну студију различитих ЕТЛ процеса са аспекта познатих ЕТЛ изазова, као и изазова ЕТЛ техника тестирања.

Традиционални ЕТЛ алати се најчешће извршавају на једном серверу и у случају гео просторних података такав приступ није у стању да обезбеди добре перформансе, поготово када су у опсег укључене велике количине гео просторних података. За унапређење перформанси ЕТЛ процеса се може користити *МапРедуце* базирана паралелизација [92]. Група аутора је у раду [93] представила истраживање перформанси ЕТЛ процеса велике количине гео просторних података прикупљених од стране *Интернет оф Тхингс* (ИоТ) сензора и предложила Д_ЕЛТ (*делауед ехтрацтинг–лоадинг –трансформинг*) процес коришћењем *МапРедуце* паралелизације. ИоТ сензори генеришу велике количине података који се морају брзо анализирати, јер врло често старији подаци немају велику употребну вредност. Брзина припреме података је у овим случајевима кључна за добијање добрих резултата анализе. Коришћени експерименти су послужили за мерење и поређење перформанси стандардних ЕТЛ/ЕЛТ (*ехтрацтинг–лоадинг–трансформинг*) приступа и Д_ЕЛТ приступа, користећи различите количине података и

нивоа комплексности. Резултати експеримената су показали да је Д_ЕЛТ приступ имао боље перформансе у односу на ЕТЛ или ЕЛТ, поготово код обраде веће количине и комплексности података, због већег степена паралелизације.

У данашњем *Биг Дата* свету који се карактерише са 4В (*Вариету, Велоцитy, Волуме анд Верацитy*), односно карактеристикама разноликости, брзине, обима и веродостојности, брзина рада са великим количинама података представља основни императив сваког решења. ЕТЛ решење мора да има способност рада у *Цлоуд* инфраструктури, коју одликује енормна снага процесирања и огроман складишни простор. Највећи изазов у оваквом окружењу је цена, обзиром на пословни модел плаћања по потрошњи. Аутори су у раду [94] представили истраживања карактеристика перформанси различитих алата и предложили смернице за избор одговарајућих решења, имајући у виду однос перформанси и цена.

Ради унапређења перформанси ЕТЛ процеса се могу користити одговарајуће стратегије које укључују оптимизацију процеса екстракције података, смањење редундантности података и коришћење паралелизације [95]. За побољшање перформанси ЕТЛ процеса за потребе *Дата Wарехоусе,* група аутора је у раду [96] предложила технику кеширања упита и његових резултата, како би се у случају постављања сличног упита од стране других корисника користили кеширани резултати.

Познато је да се за потребе аналитичких функција над подацима, поред ЦПУ користи и ГПУ (*грапхицс процессинг унит*). ЦПУ извршава све захтеве инсталираног софтвера на рачунару. ГПУ може подржати ЦПУ извршавајући конкурентне калкулације и може поједине задатке поделити у неколико мањих и извршити их у паралели. Разматрање укључивања ГПУ у ЕТЛ процесе ради повећања перформанси било је предмет студије [97], у којој је група аутора на основу извршених експеримената процесирања сетова података различитих величина, различитих типова упита и различитих типова оптерећења, закључила да се најбоље карактеристике перформанси могу добити адекватним дистрибуирањем оптерећења на комбиновано ЦПУ-ГПУ окружење.

Имајући у виду улогу ЕТЛ процеса у изради аналитичких решења, предмет истраживања групе аутора објављеног у раду [98] је проналажење модела који би на једноставнији начин представио ЕТЛ процесе. Они су поједностављени приступ представили користећи три категорије: моделирање на основу мапирања одговарајућих израза, моделирање концептуалних конструкција и моделирање на основу УМЛ (*Унифиед Моделинг Лангуаге*) окружења. Друга група аутора, чији је фокус такође био на моделирању ЕТЛ процеса, је у раду [99] анализирала различите ЕТЛ приступе моделирања, почевши од традиционалних до оних који се користе у садашњем *Биг Дата* окружењу, поредећи њихове карактеристике. Студија приказује основне изазове везане за

дизајн *Биг Дата Warehouse* решења, као и недостатак генеричког дизајн модела за прикупљање података, њихово смештање, процесирање, генерисање упита и анализу.

Комплетан преглед основних принципа, ЕТЛ техника и најбољих пракси је дат у раду [100]. Овај рад укључује више области, почевши од провере квалитета података, тестирања, оптимизације па до безбедности и перформанси ЕТЛ-а. Најзначајније најбоље праксе укључују битност мапирања података и њихову валидацију за време процеса интеграције, контролу квалитета података током процеса трансформације, партиционисање и индексирање података ради ефикасног пуњења базе, дефинисање и управљање распоредом активности, као и руковање грешкама. Квалитет дефинисаних процеса утиче између осталог и на повећање перформанси комплетног решења.

3.5. Преглед, примена и карактеристике БИ алата

Бусинесс Интеллигенце (БИ) је скуп технологија, алата, процеса и пракси које помажу организацијама да прикупе, обраде, анализирају и претворе податке у корисне информације. Циљ БИ је омогућити доносиоцима одлука у организацијама да стекну дубљи увид у пословање путем анализе велике количине података, ради доношења најбољих одлука за пословање. Кључни елементи БИ укључују складиштење података, алате за анализу података, алате за извештавање и визуализацију, као и технике предиктивне анализе.

Значај и процес доношења одлука у данашњем пословању представља комплексан процес и ослања се на избор једне од више доступних алтернатива. Највећи изазов је избор најоптималније алтернативе која има за циљ унапређење пословних резултата. Аутори у раду [101] користе три димензије модерног приступа доношења пословних одлука: квалитативан аспект, квантитаван аспект као и информативно-комуникациони аспект. Експерименти су извршени над три различита типа података из различитих извора користећи БИ методе и технике. Резултати су представљени у табеларном облику за приказ квантитавног аспекта (кроз опис резултата коришћењем атрибута и под-атрибута), за приказ квалитативног аспекта и визуелним представљањем добијених података за приказ информационо-комуникационог аспекта. Аутори су у свом истраживању користили АХП методу анализе (*аналутицал-хиерарцхицал процесс*) доношења одлука, која је базирана на концепту равнотеже за одређивање укупног релативног значаја групе атрибута, активности или критеријума.

БИ системи су дуги низ година представљали један од најбитнијих приоритета пословног руководства. Доносиоци одлука у БИ пројектима раде заједно са дизајнерима решења како би решење пружило релевантне информације за праве пословне одлуке. Већина предузећа данас поседује неко БИ решење. Након инвестиције и имплементације БИ решења, доносиоци пословних одлука често врше анализу резултата својих одлука, а на основу информација добијених од стране имплементираних решења. Аутор у раду [102] представља истраживање директних и индиректних ефеката управљања БИ решењем на квалитет донесених одлука од стране пословних руководиоца. Он користи модел парцијалних најмањих квадрата (ПЛС - Партиал Леаст Cqуаре Моделлинг) за анализу одговора које су виши ИТ менаџери у Аустралији доставили у анкети. Резултати утврђују укупан ефекат, а такође откривају посредне ефекте квалитета података и опсега БИ решења.

Група аутора у раду [103] проучава однос између БИ функција, директне користи подршке у одлучивању и користи организације у контексту окружења одлучивања. Истраживање користи квантитативну студију која је заснована на анкети, а моделом

парцијалних најмањих квадрата анализира одговоре 228 фирми које припадају различитим индустријама у региону блиског истока. Резултати студије потврђују очекивану претпоставку да постоји смислени однос између функција пословне интелигенције, подршке у пословном одлучивању и користи за пословну организацију.

Квалитет БИ решења увелико зависи од квалитета подршке доносиоца одлука у пословном предузећу. Група аутора је у свом истраживању [104] верификовала значај улоге доносиоца одлука у имплементацији БИ решења, на практичном примеру глобалне компаније која се бави логистиком. У овом примеру је разматрана подршка добијена од стране различитих менаџмент нивоа и успешност БИ решења, која су у директној спрези. Резултат подршке менаџмента БИ иницијативама доводи до бољег искоришћења података ради остваривања конкурентске предности.

БИ алати и решења проналазе примену у свакој индустрији. Планирање потреба и ефикасност ланца снабдевања је од кључног значаја у фармацеутској производњи. Група аутора у свом раду [105] представља оквир за предвиђање пословања. Он се састоји од специјализованих алата и различитих корака који могу да помогну доносиоцима одлука да тачно предвиде продају и задовољство корисника, а на тај начин и перформансе фирме. Предложени оквир се састоји од низа компоненти и укључује коришћење одговарајућег софтвера за генерисање оптималног предвиђања, као и Power БИ алата за праћење тачности добијених предвиђања продаје.

Анализа БИ технологија у ери *Биг Дата* на примеру Алибаба групе, пионира у области е-трговине [106] истражује трансформациону улогу пословне интелигенције. Опсег овог истраживања укључује БИ технологије последње генерације које су у примени, при чему је један од главних фокуса на интелигентне системе препорука, персонализоване маркетинг стратегије као и ефикасно управљање ланцем снабдевања. Анализом доступних података се креирају прилагођене препоруке за понуду производа корисницима платформе, које имају за циљ како повећање продаје, тако и веће задовољство корисника. Маркетиншке стратегије које се базирају на подацима, омогућавају Алибаби да креира персонализоване продајне промоције и лојалност корисника. Правовременом испоруком производа управља комплексан систем интелигентног управљања ланцем снабдевања, који у реалном времену прати одговарајуће активности оптимизације транспорта. Студија случаја "Синглес' Дау Глобал Схоппинг Фестивал" приказује снагу пословне интелигенције у обезбеђивању јединственог динамичног доживљаја куповине за крајњег корисника.

Истраживање примене пословне интелигенције у малим и средњим предузећима у областима управљања ланцима снабдевања је представљено у раду [107]. Изазови у овој области настају због ограничења ресурса, као и због оперативних ограничења. Група

аутора у својој студији предлаже одговарајући модел за повећање ефикасности управљања ланцима снабдевања у малим и средњим предузећима, стратешком применом пословне интелигенције. Основни подаци су добијени кроз полуструктуиране интервјуе који су затим анализирани одговарајућим тематским анализама. Резултати анализе указују на оквир који се састоји од 98 почетних тема које су анализом подељене на 55 основних тема, 15 тема које се односе на организацију и 4 глобалне теме које обухватају неколико различитих аналитика, почевши од аналитике мреже снабдевања, аналитике пословног учинка, до оптималног управљања путањом и анализом задржавања лојалности.

Пословна интелигенција налази велику примену у образовању. Пример области примене пословне интелигенције у високошколским установама је предмет студије групе аутора у раду [108]. Ове установе доносе одлуке у неколико различитих домена, веома често без коришћења података које би подржале потребне одлуке, доводећи у питање исправност таквих одлука. На тај начин се може угрозити успех и ефикасност образовне установе. Ови изазови се могу решити коришћењем пословне интелигенције која се базира на одговарајућим подацима. Представљена студија презентује систематичан преглед литературе која предлаже одговарајућу мапу пута за имплементацију система пословне интелигенције у високошколским установама.

Ради одговарајућег усвајања и примене БИ у пословном предузећу, као и остваривања постављених пословних циљева, потребно је претходно препознати факторе који на то утичу [109]. Група аутора у овој студији користи карактеристике организације, њену комплексност, корисничку перцепцију, технологију, знање и друштвени оквир за идентификацију кључних аспеката. Користећи одговарајуће анализе примарних података, пословна предузећа могу добити користан увид за решавање изазова примене пословне интелигенције.

Група аутора у свом истраживању [110] истражује факторе који утичу на коришћење БИ у банкарској индустрији. Посебан фокус представљеног истраживања се односи на утицај радног искуства запослених на коришћење БИ у комерцијалним банкама у Јордану. Модел анализе је развијен коришћењем ТОЕ (Тецхнологу Организатион Енвиронмент) оквира ради процене утицаја различитих фактора на коришћење БИ у банкама. Креирана је емпиријска анализа и дизајниран упитник за прикупљање података од стране запослених. Група аутора је анализирала 277 одговора коришћењем парцијалног моделирања најмањих квадрата, ради тестирања студије. Резултати истраживања указују позитиван утицај фактора ТОЕ на употребу пословне интелигенције у банкама.

Преглед трендова у алатима за пословну интелигенцију у финансијском сектору са посебним освртом на њихову употребу у САД и Африци [111], показује како се

финансијска индустрија развија. Коришћење података и пословне интелигенције представља основни предуслов за одрживи раст. Ова студија се бави испитивањем кључних индикатора развоја и изазова који обликују имплементацију БИ решења на оба континента. У САД је приметан пораст коришћења БИ у финансијском сектору, вођен све већом потребом за аналитиком података у реалном времену, као и усклађености са регулативама. Кроз интеграцију вештачке интелигенције и машинског учења у БИ алате обезбеђује се напредна предиктивна аналитика која омогућава пословном руководству да брзо доноси одговарајуће одлуке. Могућност коришћења БИ у облаку доноси многе предности као што су доступност, скалабилност и приступачност. У афричком финансијском сектору је видљив напредак у коришћењу БИ решења. Међутим, континент који карактеришу различити нивои развијености економија се суочава се јединственим изазовима. Владе земаља на афричком континенту као и њихови финансијски сектори улажу у одговарајућу технолошку инфраструктуру која треба да испуни одговарајуће предуслове за даља унапређења на овом пољу.

Имајући у виду данашња пословања која су богата подацима, способност извлачења одговарајућих увида из огромне количине текстуалних и структуираних података представља један од најбитнијих циљева БИ. Група аутора у својој студији [112] представља истраживање интеграције аналитике текста, вештачке интелигенције и решења за планирање ресурса ради побољшања резултата пословне интелигенције. Користећи алгоритме вештачке интелигенције над текстуалним подацима додатно се олакшава њихова анализа, при чему се могу детектовати одговарајући обрасци и трендови који доприносе бољем доношењу одлука. Кроз интеграцију са системом за планирање ресурса се може добити холистички поглед на пословање пословног предузећа.

Развој технологија машинског учења и интеграција са алатима пословне интелигенције представља додатни потенцијал за доношење правих пословних одлука [113]. Разумевање и предвиђање учинка запослених у пословним предузећима представља кључан фактор успеха у одељењима задуженим за ефикасно управљање људским ресурсима. Користећи предиктивно моделирање и анализу историјских података, данашње предузећа могу да идентификују обрасце и могуће исходе, како би на најефикаснији начин управљали талентима у својој организацији.

Истраживање утицаја вештачке интелигенције, аналитике и велике количине података у области здравства, конкретно на квалитет неге пацијената и избор дестинације у области медицинског туризма је био предмет истраживања групе аутора [114]. Презентована студија се базира на вишеструким параметрима које обухвата пословна интелигенција као што су задовољство пацијената, избор дестинације, исходима лечења, конкурентност, исплативост и друго.

Студија примене система доношења одлука на комерцијалној берзи користећи пословну интелигенцију и управљање на примеру конволуционих неуронских мрежа (*Цонволутионал Неурал Нетворкс* - ЦНН) је представила група аутора у свом раду [115]. Истраживање представљено у овој студији се базира на трансформационом утицају машинског учења са посебним фокусом на конволуционе неуронске мреже, као што су ВГГ16, РесНет50 и ИнцептионВ3, на пољима организационог менаџмента и пословне интелигенције. Ово истраживање укључује коришћење свеобухватне методологије, при чему се посебно наглашава коришћење машинског учења ради бољих резултата у доношењу пословних одлука. Резултати студије су показали веома добре перформансе ЦНН модела у односу на традиционалне алгоритме, достижући стопу тачности од чак 89.45%.

Претходни примери показују широку употребу БИ решења у различитим индустријама. На тржишту су доступни многобројни БИ алати и често постављено питање је који алат представља најбољи избор за конкретне потребе које дефинише пословање. Поред функционалности и перформанси које БИ алати пружају треба размотрити и њихову цену, посебно имајућу у виду да су неки од њих бесплатно доступни, а неки су подложни лиценцирању. Група аутора у својој студији представља и пореди БИ алате који се најчешће користе у компанијама у Турској [116]. Студија представља листу БИ алата који се највише рекламирају, при чему Мицрософт Повер БИ и САП заузимају прва два места. Дата је и детаљна анализа функционалности и карактеристика, како различитих верзија БИ алата истог произвођача, тако и упоредна анализа многобројних карактеристика различитих БИ алата и цена. Компаративном анализом карактеристика различитих БИ алата се баве и групе аутора у радовима [117], [118], [119], [120] и [121].

Бенчмарк анализа БИ алата отвореног кода који се најчешће користе у здравству је предмет студије групе аутора [122]. Основни циљ студије је анализа бесплатних алата и могућности примене у здравственим клиникама, имајући у виду све специфичности и захтеве клиничког окружења. Аутори су анализирали карактеристике шест БИ алата: КликВиев, Пало БИ Суите, Јасперсофт БИ, Таблеау Публиц, Спаго БИ и Пентахо БИ Суите. На основу свеобухватне процене карактеристика свих алата према дефинисаним захтевима, закључено је да је алат Спаго БИ најбољи избор.

4. Поређење брзина извршења упита над великим количинама података користећи различите системе за управљање базама података

Количине података које се генеришу у данашњој ери социјалних мрежа, интернет и мобилних апликација, као и свим индустријама су енормне. Ради унапређења пословања, компаније врло често користе базе података инсталиране у својој интерној пословној мрежи, као и код различитих *цлоуд* провајдера на интернету, како би из њих добили пословно вредне информације ради доношења пословних одлука. Традиционалне базе података које се користе за смештање велике количине података, а које карактеришу велике брзине конкурентних процеса уписивања у складиште се називају ОЛТП базе података или Онлине Трансацтионал Процессинг СУБП. Међутим, овај тип база података не поседује добре карактеристике перформанси када је потребно из складишта података генерисати резултате сложених упита. За потребе брзог генерисања *ад-хоц* извештаја користећи велике количине података, базу података треба да карактерише велика брзина читања. Поред тога, да би се добиле боље карактеристике перформанси извршења упита, потребно је претходно извршити денормализацију табела из базе. Такође, треба креирати и мањи број табела које ће садржати све потребне информације за оне типове упита који су потребни за генерисање извештаја које очекује пословање.

Највећи број система за управљање базама података (СУБП) користи архитектуру складишта података у којима се подаци уписују по редовима у сваком циклусу уписа (*ров-ориентед дата стораге*). На тај начин се добијају високе брзине писања. Поред ове архитектуре, постоји и архитектура складишта података која врши уписивање података по колонама, која за резултат има значајно боље перформансе читања. Обзиром да се упити заснивају на селекцији садржаја колона на основу филтера постављених у самом упиту, база са редним уписом података мора да прочита све колоне из свих селектованих табела, те да из њих филтрира одговарајуће податке из тражених колона. Код база података са уписом по колонама је тај процес много бржи, јер су подаци већ уписани по колонама и чита се само садржај потребних колона.

Да би се добиле најбоље перформансе система за пословну аналитику и извештавање, код организација са великом количином података најбољи приступ је користити базе података које су оптимизоване за аналитичко процесирање или Онлине Аналутицал Процессинг (ОЛАП) системе. Таква архитектура користи ОЛТП базе података као изворе података, који се затим из њих читају и трансформишу на одговарајуће начине и у одговарајуће формате користећи ЕТЛ алате, а ради пуњења ОЛАП база података. Пројекти напредног извештавања који користе наведену архитектуру могу бити веома

комплексни и захтевају пажљиву припрему дизајна ОЛАП система и пажљиво планирање ЕТЛ процеса. ЕТЛ процеси ће неминовно између осталог укључити чишћење и трансформацију података, који се могу налазити на већим бројем ОЛТП система и других извора.

Студија објављена у овом раду је базирана на мерењу *онлине процессинг* перформанси различитих комерцијалних СУБП, укључујући архитектуре које подржавају како складишта са редним уписом података, тако и складишта са уписом података по колонама, дизајнирана за рад са великим количинама података.

Експерименти су изведени на неколико доступних СУБП система и то ХП Вертица (Цоммуниту Едитион) која не захтева лиценцу и нема временских ограничења, Мицрософт СQЛ Сервер са Мицрософт Имагине лиценцом. Орацле са Девелопер Лиценцом и на два *Опен Соурце* СУБП, МуСQЛ са ИнфоБригхт Енгине и МариаДБ ЦолумнСторе.

4.1. Платформе база података коришћене у експериментима

Да би се добили фер резултати перформанси изабраних база података и то: ХП Вертица, Мицрософт СQЛ Сервер, Орацле, МариаДБ ЦолумнСторе и МуСQЛ, свака од њих је инсталирана на засебној виртуалној машини (ВМ), која се покретала користећи ВМware Workстатион Плауер. Свака од виртуалних машина је подешена тако да користи исти ниво хардверских ресурса као што су РАМ меморија и број процесора. На нивоу подешавања БИОС-а, на сва три физичка рачунара која су коришћена за извођење експеримената је омогућена виртуализација. Виртуалне машине су смештене на интерним хард дисковима на сваком рачунару.

Wиндоwс 10 Профессионал са 16ГБ РАМ-а је коришћен за хост оперативни систем за све виртуалне машине. Експерименти су изведени на два типа софтверских окружења:

- Старије софтверско окружење које се састојало од следећих верзија оперативних система и база података и то:
    - ХП Вертица Аналутиц Датабасе Цоммуниту Едитион 6.0.0 *релатионал датабасе манагемент сустем* (РДБМС) на ЦентОС 5.5 x86 64 оперативном систему (ОС), (виртуална машина је преузета са адресе му.вертица.цом);
    - МуСQЛ 5.6 РДБМС са две одвојене инстанце, једна са ИнноДБ и друга са ИнфоБригхт *стораге енгине* на ЦентОС 6.3 x86 64 Оперативном систему;
    - Орацле 11г Релеасе 2 *Релатионал Датабасе Манагемент Сустем* (РДБМС) са Орацле *Дата Wарехоусе* шаблоном базе, на ЦентОС 6.5 x86 64 Оперативном систему;
    - Мицрософт СQЛ Сервер 2014 Ентерприсе едитион РДБМС на Мицрософт Wиндоwс Сервер 2008 СП1 x64 Оперативном систему.
- Новије софтверско окружење које се састојало од следећих верзија оперативних система и база података и то:
    - ХП Вертица Аналутиц Датабасе 9.0.1 Цоммуниту едитион преинсталирана на Центос 7 x86 64 оперативном систему (виртуална машина је преузета са адресе му.вертица.цом);
    - МариаДБ ЦолумнСторе база података верзије 10.2.13 инсталирана на Центос 7 x86 64 оперативном систему;
    - Орацле ДБ Сервер 12ц, едиција 12.1.0 Ентерприсе инсталирана на Центос 7 x86 64 оперативном систему;
    - Мицрософт СQЛ 2017 Сервер (верзије 14.0) Стандард едитион инсталиран на МС Wиндоwс Сервер 2016 оперативном систему.

За потребе извођења експеримената су коришћени следећи рачунари:

1. Тосхиба Qосмио X775 са И7-2630QM процесором и 2 ГХз *цлоцк* брзином и 16ГБ РАМ,
2. ХП Енву 17-р104нн са И7 6700XQ процесором и 2.6 ГХз *цлоцк* брзином и 16ГБ РАМ,
3. Не-брендирани десктоп рачунар са И7-2600К процесором и 3.4 ГХз *цлоцк* брзином и 16ГБ РАМ.

Виртуелне машине постављене на сваком физичком рачунару су током експеримената конфигурисане да користе 4ГБ и 8ГБ РАМ меморије.

Пре извршења експеримената сви системи за управљање базама података су само инсталирани без додатног конфигурисања. То је за последицу имало да се брзине извршавања разликују и за ред величине. Зато се приступило проучавању подешавања параметара сваког СУБП-а понаособ (величина меморије потребне за сортирајуће бафере, величина меморије потребне за мерге-јоин бафере итд.), па су након подешавања брзине извршења биле уједначеније. У наставку је за сваки СУБП описан начин подешавања тих конфигурација.

Први тест који је извршен на сваком физичком рачунару је провера перформанси сваког од њих. За тај тест је коришћен УсерБенцхМарк алат (хттпс://www.усербенцхмарк.цом] који је извршен на сваком рачунару. Резултати теста су дати у табели 1.

Тест \ Процессор	И7-2630QM 2.0 ГХз	И7-6700XQ 2.6 ГХз	И7-2600К 3.4 ГХз
ЦПУ бенцхмарк (тотал тхроугхпут)			
4Цоре Интегер тхроугхпут	240.3	264.0	315.4
4Цоре Флоат тхроугхпут	222.0	285.5	322.6
4Цоре Михед тхроугхпут	234.4	262.7	341.3
РАМ бенцхмарк (мему латенцу)			
32МБ мем. латенцу (нс)	72.13	74.77	59.38
64МБ мем. латенцу (нс)	73.83	74.77	60.05
128МБ мем. латенцу (нс)	75.56	77.35	61.41
ХДД бенцхмарк (ХДД Реад)			
Сеqуент. Реад (МБ/с)	76.7	65.3	143.6
4к Реад (МБ/с)	0.8	2.0	0.7

Табела 1. Бенчмарк тест параметри за рачунаре коришћене у експериментима

На неким базама података инсталираним у старијем окружењу као што су: МуСQЛ 5.6 са ИнноДБ, Орацле 11г Р2 и МС СQЛ Сервер 2014 (ДатаБасе 1 која користи РowСторе принцип складиштења података) креирани су стандардни индекси над спољним кључевима у *фацт* табелама. У ХП Вертица 6.0 бази су креирана ограничења (*цонстраинтс*) над спољним кључевима у *фацт* табелама. Наведене активности обезбеђују пројекције одговарајућих веза у току извршавања упита, што ће за резултат имати значајно боље перформансе. На МС СQЛ Серверу 2014 (ДатаБасе 2 која користи ЦолумнСторе принцип складиштења података) креирани су не-кластеровани индекси над комплетним табелама, док код МуСQЛ базе података која користи ИнфоБригхт *енгине*, индекси нису креирани. У новијем окружењу над Орацле 12ц, ХП Вертица 9.0 и МариаДБ, ЦолумнСторе индекси нису креирани. На МС СQЛ Сервер 2017, креирани су не-кластеровани индекси над свим табелама.

Од додатних подешавања треба напоменути да су конфигурациони фајлови му.цнф у МуСQЛ, му-иб.цнф у МариаДБ и сп-филе у Орацле подешени на вредност РАМ меморије која је додељена виртуалној машини. Такође, код Орацле базе, за опцију од 4ГБ РАМ меморије додељене виртуалној машини, подешени су додатни параметри: сга_таргет и пга_аггрегате_таргет на 1600МБ, дб_цацхе_сизе на 312МБ, мемору_таргет на 3136МБ, сга_мах_сизе на 1920МБ и инмемору_сизе (за *Ин-мемору*) на 768МБ. За опцију од 8ГБ РАМ меморије додељене виртуалној машини, сви претходно наведени параметри су дуплирани. Слична подешавања су постављена у фајловима му.цнф и му-иб.цнф где је 70% од укупне количине РАМ меморије додељене виртуалној машини (2867МБ / 5734МБ) додељено иннодб_буффер_поол_сизе (2867МБ за 4ГБ и 5734МБ зар 8ГБ), кеу_буффер_сизе је подешен на 1024МБ/2048МБ и муисам_сорт_буффер_сизе је постављен на 256МБ/512МБ. За МС СQЛ и ХП Вертица сервере коришћен је *аутоматиц мемору манагемент* за додељивање означене меморије виртуалној машини.

4.1.1. ХП Вертица база података

ХП Вертица представља аналитичку базу података и намењена је за складиштење и рад са великом количином података и нуди одличне перформансе за коришћење код *Дата Варехоусе* пројеката, као и када је потребно обезбедити најбоље перформансе читања података. Имајући у виду да је ова база дизајнирана за аналитичке процесе, њене перформансе за трансакционе операције су слабије. Бесплатне верзије ХП Вертица (ХП Вертица Цоммуниту Едитион) које су коришћене у овој студији су Ентерприсе Едитион (вер. 6.0) и Аналутицс Платформ (вер. 9.1). Оне су доступне за преузимање у облику виртуалних машина (хттпс://www.вертица.цом/лог-ин) са преинсталираним оперативним системом ЦентОС (64бит верзија) и ХП Вертица базом на 1-3 нода, ради анализе до 1 ТБ

података, односно до 500ГБ за верзију 6.0. Дозвољено је професионалним корисницима да ажурирају лиценцу ради повећања кластера за масивно паралелно процесирање или МПП (*Массивелу Параллел Процессинг*), које обезбеђује добре перформансе за рад са егзабајт величинама података. Ова база података бележи велики број процесираних редова података у секунди извршавајући аналитичке захтеве. Исте перформансе важе не само код извршења упита, већ и код процеса учитавања података у базу. Процеси учитавања велике количине података одједном (*Булк лоад*) морају бити брзи и не смеју да спречавају или неоправдано успоравају извршавање упита, који се дешавају у исто време када и учитавање података. Вертица користи систем организације табела у пројекције, које су у ствари сортирани подскупови атрибута табеле. Свака пројекција поседује специфичан редослед сортирања података. За потребе складиштења података могу се користити два приступа и то: *Реад Оптимизед Сторе* (РОС) и *Врите Оптимизед Сторе* (WOC). РОС смешта податке на стандардни фајл систем на диску користећи вишеструке РОС контејнере. Сваки РОС контејнер, логички гледано садржи неки број комплетних скупова података сортираних по редоследу сортирања дате пројекције, складиштен као пар фајлова по колони. Један фајл садржи актуелне податке а други фајл садржи позиције индекса. ХП Вертица је су суштини *Цолумн-сторе* база, јер се фајлови који носе садржаје колона могу учитавати и из физички различитих складишта. За потребе оптимизације оптерећења при извршавању постављених упита Вертица користи Датабасе Десигнер (ДБД) алат који аутоматски проналази одговарајуће скупове пројекција. Такође, треба напоменути и да ову базу карактеришу и високе перформансе операција сортирања и компресије.

4.1.2. Орацле 11г Р2 са Дата Варехоусе шаблоном и 12ц са Ин-Мемору Цолумн-Сторе

Орацле 11г Р2 база података може да се инсталира користећи један од следећих шаблона: Трансацтион Процессинг, Генерал Пурпосе, Дата Варехоусе и Цустом Датабасе уз помоћ уграђеног алата Датабасе Цонфигуратион Ассистант (ДБЦА). Једна од специфичности Дата Варехоусе шаблона који је коришћен при извођењу експеримената је подешавање системског параметра стар_трансформатион_енаблед на ”труе”. Укључивање овог параметра је веома важно када се планирају извршавања бројних комплексних упита над великим количинама података, ради добијања најбољих перформанси система. *Стар трансформатион оптимизер* је први пут представљен у Орацле 8и ради процесирања тзв. *Стар* упита који се најчешће користе у *Дата Варехоусе* апликацијама. У суштини, основни циљ стар трансформације је ефикасан приступ редукованом сету података из *фацт* табела када је упит филтриран по предикатима *фацт* или *димензионс* табела. Орацле 11г Р2 Дата Варехоусе шаблон омогућава *ин-мемору* паралелно извршавање

процеса које даје свој допринос коришћењем *буффер* кеша базе података, чија се вредност може тачно подесити преко системског параметра дб_цацхе_сизе. Коришћењем паралелног извршавања и приступајући подацима у бафер кешу, подаци се могу читати и десет пута брже него када се подацима мора приступити са диска. Овај шаблон такође омогућује коришћење *Битмап* индекса као и *Битмап Јоин* индекса, чиме се значајно добија на перформансама система.

Традиционалне базе података смештају податке на диск у редовима. Аналитичке *ин-мемору* базе података смештају податке по колонама. Орацле је у верзији 12ц имплементирао оба наведена приступа, која су до тада била доступна кроз засебне производе. *Ин-мемору Цолумн-сторе* даје могућност корисницима да имају трансакциону и аналитичку базу података у оквиру једног производа. За трајно складиштење података се користи *pow* формат на диску. Када се захтевају операције читања/писања, подаци се учитавају у традиционални *Pow Сторе* бафер кеш, док у случајевима захтева само за операцијама читања, подаци се учитавају у *Ин-мемору Цолумн-сторе*. Када је потребно извршити трансакцију која укључује уписивање, ажурирање или брисање, нови подаци ће се аутоматски појавити у оба складишта.

Администратор Орацле 12ц може изабрати које ће специфичне групе колона, комплетне табеле, погледе или партиције табела да чува у *Ин-мемору Цолумн-сторе*. Из тог разлога је ово складиште добар избор за процесе скенирања података који захтевају условно филтрирање, упите који враћају мањи број колона из табела које поседују већи број колона, упите који спајају мање табеле у велике табеле и упите који агрегирају податке. *Ин-мемору Цолумн-сторе* опција је у потпуности компатибилна са другим својствима Орацле база података, као што су компресија, енкрипција и партиционисање табела. Количина меморије која се додељује за *Ин-мемору Цолумн-сторе* се подешава користећи системски параметар инмемору_сизе, док се објекти као што су табеле, колоне и сл. убацују у складиште користећи клаузулу *ИНМЕМОРУ* у ДДЛ наредбама као што су *ЦРЕАТЕ ТАБЛЕ, АЛТЕР ТАБЛЕ* итд.

4.1.3. МС СQЛ са ЦолумнСторе Индексом

МС СQЛ сервер има могућност креирања *Цолумн-сторе* индекса и представља не-хибридну имплементацију, због тога што складишти све податке сваке колоне на засебним страницама. Резултат овог приступа се огледа у бољим перформансама читања и ефикаснијем коришћењу меморије. *Цолумн-сторе* индекси чувају податке колона у компресионом формату који је дизајниран за брзо читање комплетног садржаја колона. Скуп редова које је потребно складиштити се прво дели у групе редова, при чему се свака

група састоји од фиксног броја колона. Свака *row* група се се кодира и компресује независно. Резултат оваквог приступа је један *цолумн* компресовани сегмент за сваку укључену колону. Сваки *цолумн* сегмент се складишти као засебан БЛОБ објекат (*Бинару Ларге Објецт*).

Мицрософт је проширио својства постојећег *qуеру процессинг енгине* за *Дата Варехоусе*. Такође, у оквиру овог решења се налази и оптимајзер упита који може да изабере најбољи тип оператора (између опција *row-at-a-time* и *батцх-ат-а-тиме*) за дате процесе. Коришћењем наведеног оптимајзера се добија већа брзина извршења упита.

Мицрософт СQЛ Сервер *Цолумнсторе индех* је дизајниран да ради са *Дата Варехоусе* имплементацијама, посебно код оних са великим *фацт* табелама. Одличне перформансе карактеришу процесе учитавање велике количине података одједном (*булк лоадс*) и упитима који анализирају велике количине података. *Цолумн-сторе* индекси нуде одличне перформансе приликом скенирања комплетних табела, при чему нису дизајнирани за претрагу специфичних вредности.

Могу се креирати два типа индекса, кластеровани и некластеровани. Кластеровани *Цолумн-сторе* индекс се не може комбиновати са другим индексима. Он је примарни и једини за целу табелу и обухвата све колоне. Са друге стране, не-кластеровани индекс је секундарни индекс који може да индексира подскуп колона у кластерованом индексу и захтева додатни складишни простор за копије колона у индексу. За разлику од кластерованих индекса, некластеровани се могу комбиновати са другим индексима у табели.

4.1.4. МариаДБ ЦолумнСторе

МариаДБ је дистрибуирана база података која користи масивну паралелизацију и складишти податке по колонама. Њена основа потиче од ИнфиниДБ 4.6.7 базе података, која је примењена на МариаДБ 10.1 систему. Карактеристике ове базе података су брзо читање података из складишта, као и брзо извршавање упита над великим скупом података у бази. Такође, МариаДБ пружа могућност извршавања паралелних упита, што у пракси значи да се упити могу извршавати истовремено на више нодова, чиме се убрзава добијање резултата упита. Због колумнарног складиштења података, пружа се могућност једноставног и брзог додавања или укидања колона, чак и у *онлине* режиму. Подржано је скалирање за *Биг Дата* на нивоу извршавања процеса обраде података величине петабајта. Такође, подржана је линеарна скалабилност, компресија података као и вертикало партиционисање.

4.1.5. МуСQЛ са ИнноДБ и ИнфоБригхт енгине

МуСQЛ база података је представник традиционалних база података која се масивно користи како у пословним окружењима, тако и за складиштење података за интернет апликације и сервисе. Она је дизајнирана за ОЛТП трансакције и у основној конфигурацији користи ИнноДБ трансакциони *енгине* и НДБЦЛУСТЕР *енгине* за складиштење података.

На основу препорука из референтног водича за МуСQЛ, за потребе *Дата Варехоусе* имплементација, ИнфоБригхт *енгине* је најбољи избор на основу упоредне анализе доступних *енгине*. Неке од карактеристика ИнфоБригхт *енгине* су ниска цена складиштења података, коришћење *кноwледге грид* технике, брзо учитавање података користећи *хигх-спеед дата лоадер*, колумнарни упис података и високи нивои компресије података, ради пружања одличних *Дата Варехоусе* перформанси.

4.2. Припрема тест окружења

За потребе извођења експеримената све коришћене базе података су припремљене на сличан начин. Иста шема база података као и исти скуп табела и садржаја је постављен на свакој бази података. Коришћена је Вмарт база података које је пример базе података постављен од стране произвођача у Вертица виртуалној машини. Ова уграђена база података се најчешће користи за тестирање карактеристика и перформанси Вертица *енгине*. Ради дефинисања шеме Вмарт базе података се може користити уграђена скрипта вмарт_дефине_сцхема.цql, која се налази у фолдеру /опт/вертица/ехамплес/ВМарт_Сцхема/ Вертица виртуалне машине. Имајући у виду да се у експериментима користе различите базе података, ова иницијална скрипта је посебно прилагођена за сваку базу података. Вмарт база података се састоји од три шеме: Публиц, Сторе и Онлине_Салес. За извођење експеримената четири *фацт* табеле су посебно значајне, због укупног броја редова које садрже и то: инвентору_фацт и сторе_ордерс_фацт са 300.000 редова и сторе_салес_фацт и онлине_салес_фацт са 5.000.000 редова.

Постављање одговарајућег окружења је започето инсталирањем примера Вмарт базе података на Вертица систему, извршавањем одговарајуће Вмарт инсталационе скрипте. Затим су све табеле извезене у фајлове са ЦСВ форматом. Оригиналне скрипте за креирање табела користећи ДДЛ команде су посебно прилагођене користећи синтаксе других база података, које се користе у експерименту.

Вмарт база података је затим инсталирана на све СУБП извршавајући прво скрипте за креирање шема, а затим су на свакој од њих покренути одговарајући алати за групно учитавање података из ЦСВ датотека.

У Вертица 9.0 је коришћен алат Вертица Манагемент Цонсоле и Наредба „*ЦОПУ*“ за учитавање података. Сличан приступ за учитавање података из ЦСВ датотеке је коришћен у Вертица 6.0 бази, коришћењем исте наредбе у РазорСQЛ клијентској апликацији. Код МуСQЛ и МариаДБ је коришћена наредба „*Лоад дата*“ из МуСQЛ Воркбенцх алата. За пуњење Орацле базе је коришћена “*Сql Лоадер*” команда из линух терминалског схелл-а са Орацле корисничким акредитивима, док је код МС СQЛ Сервера коришћен БЦП програм извршен из командне линије.

За потребе повезивања на одговарајућу базу података и извршавање упита употребљена су четири различита алата постављена на одговарајућим виртуалним машинама. За Вертица 6.0 базу је коришћен алат РазорСQЛ, за Орацле базу алат Орацле СQЛ Девелопер, за МуСQЛ и МариаДБ Цолумнсторе базу МуСQЛ Воркбенцх алат, док је СQЛ Сервер Манагемент Студио коришћен за повезивање са МС СQЛ Сервером. Алати

РазорСQЛ, СQЛ Девелопер и МуСQЛ Воркбенцх поседују уграђену функционалност која има могућност приказа времена извршења упита. Ради приказивања времена извршења операција у СQЛ Манагемент Студиу, коришћена је опција '*Инцлуде Цлиент Статистицс*' која приказује „укупно време извршења", заједно са другим релевантним статистичким параметрима повезаним са процесима извршења упита.

Експерименти су конципирани на извршавању различитих *ад/хоц* упита и мерењу њиховог времена извршења на свакој од припремљених база података. Креирана су укупно четири упита која су имала за циљ да учитају и процесирају све редове из *фацт* табела, без коришћења филтера (преко *WXEPE* клаузуле), а ради максималног оптерећења система. Сви дефинисани упити су дизајнирани користећи ПЛ/СQЛ синтаксу, при чему су изведене одговарајућа подешавања у СQЛ синтаксама, због извршења на различитим системима.

Постављени циљ **Упита 1** (**Q1**) је тестирање перформанси агрегације у једној табели чињеница од 5 милиона редова, спојеној *иннер-јоин* везом са једном колоном у димензионој табели *промотион_дименсион,* која се састоји од 1,000 редова и *иннер-јоин* везама са две колоне у димензионој табели *продуцт_дименсион,* која се састоји од 60.000 редова. Поље вредности валуте је коришћено за агрегацију, а груписање је изведено по две колоне из димензионих табела. Дефинисани упит треба да прикаже резултате остваренog профита, који је груписан по категоријама производа и врстама снижења цена.

```
СЕЛЕЦТ пром.прице_редуцтион_тупе, прод.цатегору_десцриптион,
сум(о.гросс_профит_доллар_амоунт) АС сум_профит
ФРОМ онлине_салес.онлине_салес_фацт о
ИННЕР ЈОИН публиц.промотион_дименсион пром ОН
        о.промотион_кеу = пром.промотион_кеу
ИННЕР ЈОИН публиц.продуцт_дименсион прод ОН
        о.продуцт_кеу = прод.продуцт_кеу АНД о.продуцт_версион =
прод.продуцт_версион
ГРОУП БУ пром.прице_редуцтион_тупе, прод.цатегору_десцриптион;
```

Упит 2 (**Q2**) је дизајниран на начин да изврши тестирање перформанси агрегације у једној табели чињеница од 300.000 редова, спојеној *иннер-јоин* везом са једном колоном у димензионој табели *варехоусе_дименсион,* која садржи 100 редова и *иннер-јоин* везама са две колоне у димензионој табели *продуцт_дименсион* са 60.000 редова. Поље количине је коришћено за агрегацију, а груписање је изведено по две колоне из димензионих табела. Дефинисани упит треба да прикаже број производа који се чувају у складиштима, груписан по регионима и категоријама производа.

```
СЕЛЕЦТ w.wарехоусе_регион, п.цатегору_десцриптион,
сум(и.qту_ин_стоцк) ас сум_qту
ФРОМ публиц.инвентору_фацт и
ИННЕР ЈОИН публиц.wарехоусе_дименсион w ОН
    и.wарехоусе_кеу = w.wарехоусе_кеу
ИННЕР ЈОИН публиц.продуцт_дименсион п ОН
    и.продуцт_кеу = п.продуцт_кеу АНД и.продуцт_версион =
п.продуцт_версион
ГРОУП БУ w.wарехоусе_регион, п.цатегору_десцриптион;
```

Упит 3 (**Q3**) је припремљен за тестирање перформансе агрегације у једној табели чињеница са 5 милиона редова, спојеној *иннер-јоин* везама са три димензионе табеле и то: датум_дименсион са 1.826 редова, сторе_дименсион са 250 редова и цустомер_дименсион са 50.000 редова. Овај упит садржи две функције агрегације са једном *ЦАСЕ* клаузулом. Поред тога, коришћена је и *РОУНД* функција у оквиру функције агрегације. Креиране су везе (*јоинс*) по једној колони над све три димензионе табеле. Дефинисани упит треба да прикаже профите ланца супермаркета Вмарт и проценте оствареног профита по регионима купаца, груписане по календарским годинама

```
СЕЛЕЦТ д.цалендар_yеар, РОУНД((СУМ(ЦАСЕ WХЕН
цд.цустомер_регион = сд.сторе_регион ТХЕН 1 ЕЛСЕ 0 ЕНД) /
ЦОУНТ(*)) * 100, 2) АС перцентаге_саме_регион,
СУМ(гросс_профит_доллар_амоунт) АС сум_профит
ФРОМ сторе.сторе_салес_фацт сф
ИННЕР ЈОИН публиц.дате_дименсион д ОН
    сф.дате_кеу = д.дате_кеу
ИННЕР ЈОИН сторе.сторе_дименсион сд ОН
    сф.сторе_кеу = сд.сторе_кеу
ИННЕР ЈОИН публиц.цустомер_дименсион цд ОН
    сф.цустомер_кеу = цд.цустомер_кеу
ГРОУП БУ д.цалендар_yеар;
```

Упит 4 (**Q4**) је дизајниран за тестирање перформанси агрегације у једној табели чињеница са 300.000 редова без постављање било каквих релација са табелама димензија, али са вишеструким датумским и математичким функцијама као што су: *ЕКТРАЦТ, ДАТЕДИФФ, РОУНД, СУМ, АВГ*. Овај упит је припремљен са циљем мерења брзине процеса *онлине* агрегације неагрегираних података. Дефинисани упит треба да прикаже просечно време у данима, које је потребно транспортним фирмама да испоруче робу Вмарт магацину, као и проценат кашњења испоруке. Извештај треба да буде груписан према именима пошиљаоца и календарским годинама извршене испоруке.

```
SELECT shipper_name, EXTRACT(year FROM date_ordered) AS
calendar_year,
ROUND((SUM(CASE WHEN date_delivered > expected_delivery_date
THEN 1 ELSE 0 END)/COUNT(*))*100, 2) AS percentage_latency,
ROUND(AVG(DATEDIFF(date_delivered, expected_delivery_date)),
2) AS delivering_days,
ROUND(AVG(DATEDIFF(date_ordered, date_shipped)), 2) AS
shipping_days
FROM store.store_orders_fact
GROUP BY shipper_name, calendar_year;
```

4.3. Резултати истраживања

Посебна пажња је посвећена припреми окружења и извршавању тестова да би се добили фер резултати. Приликом прелиминарног тестирања уочено је да постоји утицај кешираних података без обзира да ли се у периоду између два иста теста рестартују сервиси базе података. Такође, уочено је да се поновљени упит покренут после рестарта сервиса база података брже извршава у односу на извршавање истог упита после рестартовања виртуалне машине. Из тог разлога, приликом вршења експеримената сваки упит је извршаван након поновног покретања виртуалне машине. Такође, код прелиминарног тестирања је приликом покретања упита на МС СQЛ серверу уочено да је време извршења поновљеног упита било брже у односу на прво, без обзира што је виртуална машина у међувремену рестартована. Разлог оваквог понашања је у томе што је МС СQЛ након извршеног првог упита сачувао статистичке објекте у табелама чињеница и димензија, те користио те податке приликом следећег извршења упита. Да би се добили фер резултати мерења, након сваког извршеног упита су генерисани статистички подаци обрисани.

Упити су вршени након почетног смањења оптерећења система (испод 0,2) након покретања виртуалне машине. Принцип мерења резултата је постављен на начин да се сваки тест/упит извршавао три пута (поштујући претходно наведено правило извршења једног упита по покретању виртуалне машине), а затим је за измерену вредност узиман просечан резултат сва три мерења. У оквиру првог експеримента се налазе упити покренути на „старијим“ ОС и СУБП, а у оквиру другог експеримента упити покренути на „новијим“ ОС и СУБП.

Приликом вршења експеримената, утврђено је да су брзине извршавања упита Q1 и Q3 на МуСQЛ-а са ИнноДБ *енгине* и МС СQЛ Сервер-у са стандардним *PowCtope* индексима на најспоријем рачунару биле веома мале, тако да се упити/експерименти на овим базама података у другим хардверским конфигурацијама нису покретали. Добијени резултати брзина извршених упита по експериментима су приказани у табелама 2 и 3.

Табеле са резултатима мерења брзине упита су форматиране за приказивање више димензија и то:

- Архитектура рачунара приказана кроз перформансе процесора и РАМ меморије,
- Количина РАМ меморије која је додељена свакој виртуалној машини,
- Коришћени систем база података и
- Број/ознака упита (Q1-Q4).

Процессор / РАМ	**Qуериес**	**МС СQЛ 2008 (цолумнсторе)**		**ХП Вертица 6.0**		**МуСQЛ 5.6 (ИБ)**		**Орацле 11г датаwарехоусе**		**МС СQЛ (powсторе)**	**МуСQЛ (ИнноДБ)**
		4ГБ	**8ГБ**	**4ГБ**	**8ГБ**	**4ГБ**	**8ГБ**	**4ГБ**	**8ГБ**	**4ГБ**	**4ГБ**
И7-2630QМ 2.0ГХз ДДР3 РАМ 1333МХз ЦЛ9	Q1	3182	3088	4423	2750	6023	5018	9531	8026	79633	214399
	Q2	2090	2037	456	272	1018	1015	1699	1056	710	6433
	Q3	3416	3244	3899	3329	6012^{-}	6019^{+}	9526	9522	12320	87719
	Q4	1014	858	318	318	1010^{-}	1012^{+}	1009^{-}	1012^{+}	103	1321
И7- 6700XQ 2.6ГХз ДДР3 РАМ 1600МХз ЦЛ11	Q1	2303^{-}	2414^{+}	2873	1991	5151	4527	7037	6520		
	Q2	2086	2187	550	432	508	508	507^{-}	509^{+}		
	Q3	2144	2109	2210*	1575	5013^{-}	5030^{+}	8022	7600		
	Q4	952	1123	303	280	1008	1007	1009^{-}	1011^{+}		
И7-2600К 3.4ГХз ДДР3 РАМ 1600МХз ЦЛ11	Q1	1981	1775	2463	1539	4017^{-}	4021^{+}	5016	5015		
	Q2	998	951	415	376	571	543	506	506		
	Q3	1887	1703	1579	1075	4666	4513	5521	5514		
	Q4	530	530	203^{-}	207^{+}	1007	507	1011	1011		

Табела 2. Време (у мс) потребно за извршавање упита за различите „старије“ СУБП на различитим рачунарима и различитој меморији додељеној у Експерименту 1.

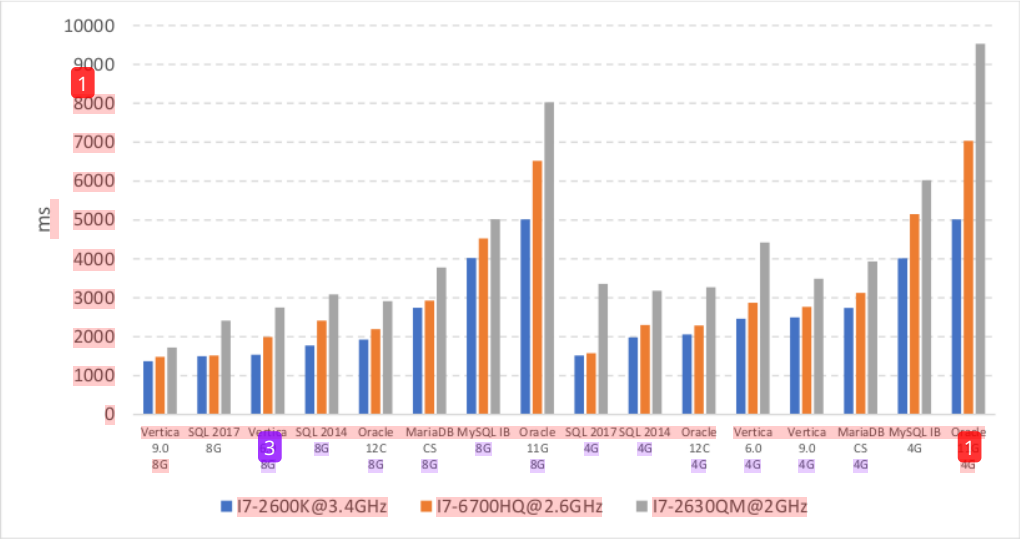
Процессор / Цлоцк спеед	Qуериес	МС СQЛ 2017 (цолумнсторе)		ХП Вертица 9.0.1		МариаДБ цолумнсторе 10.2.13		Орацле 12ц Ин-Мемору Цолумн Сторе	
		4ГБ	8ГБ	4ГБ	8ГБ	4ГБ	8ГБ	4ГБ	8ГБ
И7-2630QМ 2.0ГХз ДДР3 РАМ 1333МХз ЦЛ9	Q1	3358	2414	3490	1720	3936	3779	3272	2915
	Q2	844	836	357	323	1305	1262	498	382
	Q3	2594	2875	2075	1059	4452	4282	4312	4148
	Q4	875	795	315	311	1042	1022	922	610
И7- 6700XQ 2.6ГХз ДДР3 РАМ 1600МХз ЦЛ11	Q1	1575	1515	2769	1483	3124	2927	2289	2193
	Q2	907	440	413	282	542	489	473	252
	Q3	2008	1915	2265*	936	3889	3285	3844	3671
	Q4	718^{-}	796^{+}	200^{-}	222^{+}	732	716	646	584
И7-2600К 3.4ГХз ДДР3 РАМ 1600МХз ЦЛ11	Q1	1516	1500	2498	1371	2741^{-}	2745^{+}	2057	1928
	Q2	392	461	236	214	466^{-}	483^{+}	253	222
	Q3	1606	1579	1482	803	2347	2308	2930	2778
	Q4	516	497	292	163	344^{-}	404^{+}	433	363

Табела 3. Време (у мс) потребно за извршавање упита за различите „новије“ СУБП на различитим рачунарима и различитој меморији додељеној у експерименту 2

Ради приказивања изузетака у резултатима у обе табеле су коришћени симболи. Симболи „+“ и „-“ у суперскрипту означавају „изузетке“, да се исти упит извршава брже користећи мање меморије на истом рачунару. Симбол „*“ у суперскрипту такође означава другу врсту „изузетка“, да се исти упит брже извршава на „старијој“ верзији СУБП на истом рачунару.

Измерени резултати су оправдали очекивања. У већини случајева упити су се извршавали брже на бољим рачунарима, користећи више РАМ-а и на „новијим“ верзијама СУБП. Време извршења упита Q1 и Q3 над табелама чињеница са 5М редова је било најбрже на Вертица 9.0 СУБП користећи 8Г РАМ-а на свим рачунарима. У опцији са виртуалним машинама код којих је додељено 4ГБ меморије, упит Q1 се брже извршавао на МС СQЛ 2017, а упит Q3 је био бржи на Вертица 9.0. Наведени упити су се најспорије извршавали у Орацле 11г, што је и очекивано, јер он нема имплементиране *ЦолумнСторе* индексе. Може се закључити да без обзира на најспорије извршење упита у поређењу са системима које подржавају *ЦолумнСторе* индексе, брзина извршења упита у Орацле 11Г је значајно боља у поређењу са временом извршавања истих упита код МС СQЛ 2014 и МуСQЛ (ИнноДБ) база података које користе *PowCторе* приступ.

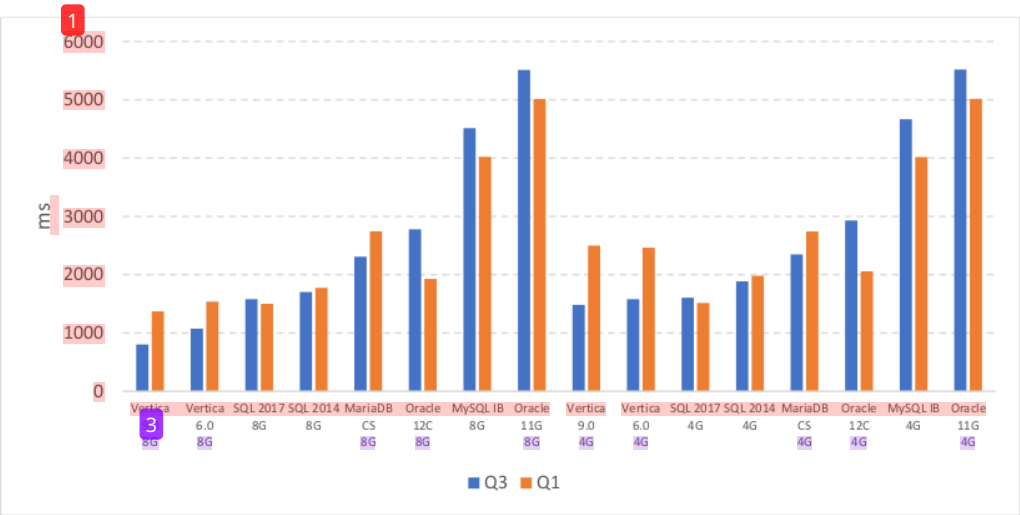
Поређење времена извршавања упита Q1 у односу на архитектуре система, количине РАМ-а додељене виртуалним машинама и системе база података приказано је на графикону 1. Јасно се може уочити да је време извршења упита у складу са перформансама архитектуре рачунара и да се упит Q1 углавном брже извршава на виртуалним машинама са већом додељеном РАМ меморијом. Редослед приказа система база података је сортиран са лева на десно, сходно брзини извршења упита на сваком рачунару и подељен у две групе на основу додељене количине меморије.



Графикон 1. Брзина извршавања упита Q1

На основу приказаних резултата у обе табеле се не може лако видети који од упита Q1 и Q3 су били брже извршени на различитим платформама. Упит Q1 се извршавао брже од упита Q3 у МС СQЛ 2017 у свим случајевима (осим на рачунару са И7-2630QM процесором и 4ГБ РАМ меморије додељене виртуалној машини), у Орацле 12Ц у свим случајевима, у Орацле 11ГР2 у свим случајевима (осим на најспоријем рачунару са 4ГБ РАМ меморије додељене виртуалној машини), у МариаДБ ЦС на споријим рачунарима, у МуСQЛ ИБ на најбржем рачунару, као и на споријим рачунарима са 8ГБ РАМ меморије додељене виртуалној машини. Упит Q3 се извршавао брже у другим случајевима. Такође се може уочити, да је слична ситуација је и са упитима Q2 и Q4. Код неких случаја упит Q2 се извршавао брже (у Орацле 12ц на свим рачунарима, на бржим рачунарима у МуСQЛ ИБ и Орацле 11Г, итд.), док је код неких других упит Q4 био бржи. Упити Q2 и Q4 су били извршени са приближним брзинама на различитим платформама база података без обзира на тип индекса, због релативно малог броја редова у табели чињеница („само“ 300.000 редова у табели).

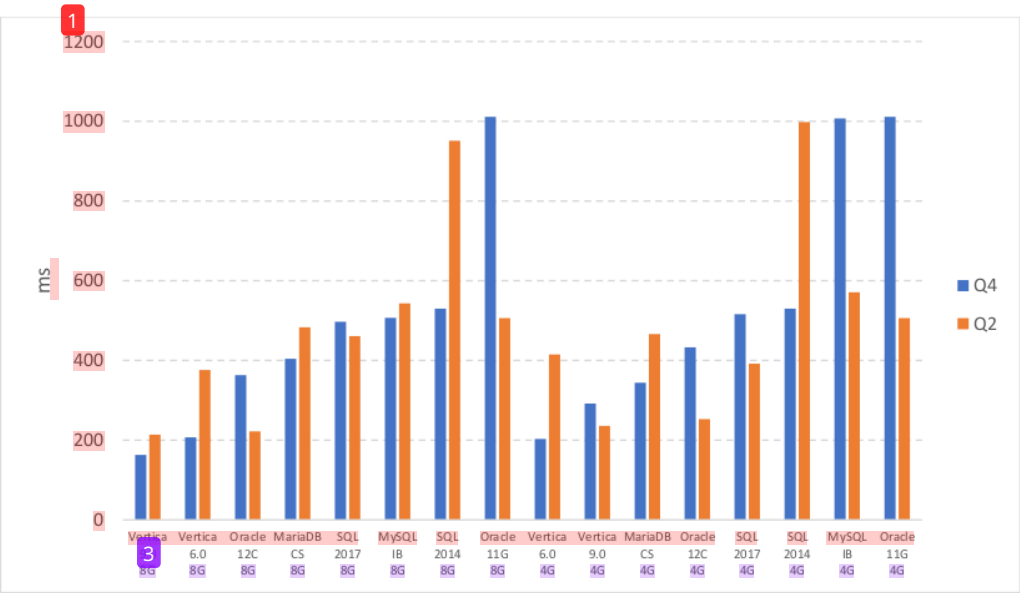
Пошто ће се у продукционим кружењима користити снажнији рачунари, репрезентативни пример који се може изабрати је пример поређења времена извршавања упита Q1 и Q3 са *фацт* табелом од 5М редова на најбржем рачунару. Резултати овог примера су приказани на графикону 2.



Графикон 2. Брзина извршавања упита Q1 и Q3 са већим Фацт табелама на најбржем рачунару који се користи у експериментима

На графикону 2 су платформе база података сортиране са лева на десно према времену извршења упита Q3. Може се видети да је време извршења упита Q1 и Q3 мање од 5,6 секунди на свим платформама база података на најбржем рачунару. Најбоља времена извршења су измерена базе Вертица и MC CQЛ.

Анализа перформанси при извршењу упита Q2 и Q4 се може извршити на основу приказаних резултата на графикону 3.



Графикон 3. Брзина извршавања упита Q2 и Q4 са мањим Фацт табелама на најбржем рачунару који се користи у експериментима

Упити Q2 и Q4 су се најбрже извршавали на Вертица платформи, а затим следе Орацле 12Ц и МариаДБ ЦолумнСторе. Најбоље време извршавања за ова два упита у Вертица бази података је износило око 200 мс.

Као што се и очекивало, генерално гледано, упити на табелама са већим бројем редова су се извршавали спорије од оних са мањим бројем редова. Са коришћењем снажнијих рачунара добијена су боља времена извршавања упита. Међутим, са повећањем РАМ меморије додељене виртуалним машинама, нису увек добијене веће брзине извршавања упита. Платформе база података које користе *ЦолумнСторе* индексе над табелама, као што се и очекивало, извршавају упите знатно брже него оне са стандардним *РowСторе* индексима.

4.4. Закључак

У оквиру ове студије је приказано поређење перформанси извршења упита над великим скупом неагрегираних података користећи различите платформе база података. Процеси извршења агрегационих упита извршени су у *он-лине* режиму. Захваљујући подешавању конфигурација СУБП на администраторском нивоу, добијена времена извршења упита су за поједине СУБП била знатно краћа у односу на иницијално инсталиране базе без подешавања конфигурационих фајлова. На основу извршених експеримената доказана је општа претпоставка да ће колумнарне платформе надмашити стандардне базе података које уписују податке у складиште по редовима. Све изабране платформе база података над којима су вршени експерименти: МС СQЛ Сервер, ХП Вертица, Орацле, МариаДБ ЦолумнСторе и МуСQЛ су постављене на виртуалним машинама користећи ВМware виртуализационо окружење на неколико рачунара са различитим хардверским перформансама и са 4ГБ и 8ГБ РАМ меморије додељене виртуелним машинама. Тест база података која се састојала од четири табеле чињеница и неколико димензионих табела је инсталирана на свакој изабраној платформи база података. Свака база података је тестирана су четири различита упита над табелама чињеница које су садржале по 5,000,000 и 300,000 редова.

Највећа брзина извршења постигнута је очекивано на најбржем рачунару коришћеном у експерименту, на комерцијалним СУБП: ХП Вертица и МС СQЛ Сервер, код којих су табеле у бази података индексиране у ЦолумнСторе формату. Резултати експеримената су показали да је извршење упита (Q1 и Q3) над већим табелама чињеница (са 5,000,000 редова) било брже у Вертица бази када је 8ГБ РАМ меморије додељено виртуалној машини, док је упит Q1 брже извршен на МС СQЛ Серверу са 4ГБ РАМ меморије додељене виртуалној машини. ХП Вертица је показала брже извршавање упита (Q2 и Q4) са мањим табелама чињеница, без обзира на количину РАМ меморије додељене виртуалним машинама. Некомерцијалне базе података МариаДБ ЦолумнСторе и МуСQЛ ДБ са ИнфоБригхт механизмом за складиштење које користе ЦолумнСторе индексирање су показале прилично добре перформансе. Орацле 12ц је упите Q1 и Q2 брже извршавао него МариаДБ ЦС, али су упити Q3 и Q4 брже извршавани на МариаДБ ЦС. Генерално посматрано, сви упити су извршавани на свим рачунарима за мање од 10 секунди (осим у случајевима када је база података користила индексе *powcmope* у МС СQЛ и МуСQЛ). На основу измерених резултата препорука је да се колумнарне платформе база података користе за аналитичке потребе.

На избор одговарајућег решења за пословање може утицати и цена самог решења. У тренутку писања ове дисертације, на основу расположивих информација на web

сајтовима произвођача чији производи су коришћени у експериментима, а који су подложни лиценцирању је дата у следећим тачкама:

- Мицрософт СQЛ Сервер се може лиценцирати по Пер Цоре и Сервер + ЦАЛ (Цлиент Аццесс Лиценсе) моделима, користећи различите програме лиценцирања. Цена лиценце СQЛ Сервер Ентерприсе едиције која покрива 2 процесорска језгра кроз програм количинског лиценцирања износи 15,123 УСД.
- ХП Вертица Цоммуниту Едитион је бесплатна. Уколико је потребно користити комерцијалну верзију, ХП Вертица се лиценцира по ноду који може имати до 16 језгара. Вертица Премиум Лиценца са укљученом 24x7 једногодишњом подршком износи 13,930 УСД
- Орацле База података се лиценцира по процесору. Процесорска Лиценца Ентерприсе Едиције износи 47,500 УСД а на цену лиценце треба додати и Софтwаре Упдате и подршку која износи 10,450 УСД.
- МариаДБ Цоммуниту Сервер је бесплатан, док је МариаДБ Ентерприсе подложан лиценцирању и цена зависи од региона коришћења, те се за конкретне случајеве треба обратити произвођачу.
- МуСQЛ је *Опен Соурце/бесплатан* производ. МуСQЛ Ентерприсе производ је подложан лиценцирању а почетна цена МуСQЛ Ентерприсе Едитион Сусбцриптион (Он премисе) лиценце која покрива сервер са 1-4 сокета износи 5,350 УСД.

Имајући у виду висок ниво интеграције са осталим системима као и одличне перформансе бесплатне едиције ХП Вертица (ХП Вертица Цоммуниту Едитион), пословне организације би требало да размотре њено коришћење за потребе извештавања. Уколико ограничење бесплатне едиција не утичу негативно на перформансе решења, овај СУБП може да представља одличан избор за мала и средња предузећа.

.

5. Поређење брзина извршења ЕТЛ упита при коришћењу СQЛ кода, ССИС уграђеног ЕТЛ алата и Пентахо бесплатног ЕТЛ алата

Квалитетни подаци на основу којих се генеришу адекватни извештаји и обезбеђују одговарајуће аналитичке функције представљају први предуслов за доношење квалитетних пословних одлука. Подаци се врло често налазе у бројним изворима који могу бити структуирани, полуструктуирани и неструктуирани, смештени у базама података као и фајловима различитих формата. Пројекти напредног пословног извештавања су једни од најкомплекснијих и захтевају адекватна знања и технике, почевши од чишћења и трансформације изворних података, преко дизајна аналитичке базе података (ОЛАП) у коју се смештају припремљени подаци, па до БИ алата задужених за дизајн извештаја и приказивање

пословању да детектује и означи најпотребније информације из огромне количине расположивих података, ради доношења адекватних пословних одлука.

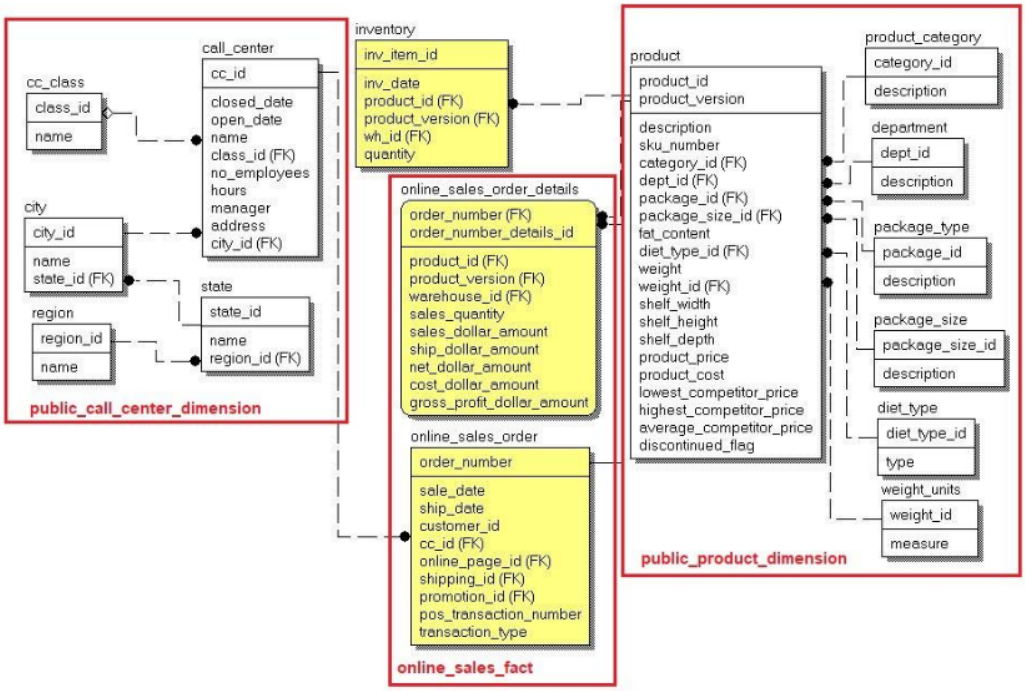
Што је количина података на њиховим изворима већа то су и ЕТЛ процеси дужи, те је веома чест пример да се они извршавају ван радног времена. То је разлог зашто подаци у ОЛАП бази могу бити старији и више од једног дана у односу на оне у ОЛТП бази. Имајући у виду да пословање нема времена за кашњење и да пословне одлуке треба доносити брзо и правовремено, веома је значајно да ОЛТП и ОЛАП системи буду фреквентно синхронизовани, без кашњења.

Постављени циљ ове студије је мерење времена ЕТЛ процеса ради пуњења ОЛТП базе великом количином података, користећи савремене ЕТЛ алате. Имајући у виду да брзина извршавања ЕТЛ процеса зависи од перформанси рачунара на којима се они извршавају, експерименти су вршени на неколико рачунара различитих хардверских карактеристика.

Експерименти су изведени на Мицрософт СQЛ Сервер СУПБ са Мицрософт Имагине лиценцом.

5.1. Окружење за извођење експеримената

За потребе извођења експеримената креирана је Вмарт ОЛТП база података из постојеће Вмарт ОЛАП базе података. Наведена ОЛТП база се састојала од 48 табела од којих су 4 табеле од 5,000,000 редова и 3 табеле од 300,000 редова биле највеће. Све остале табеле су имале 25,000 и мање записа. Имајући у виду да је шема Вмарт ОЛТП базе превелика, на слици 1 је приказан део шеме са 15 табела. Неке табеле су ꝛруписане ради трансформације кроз ЕТЛ процес.



Слика 1. Део шеме ВМарт ОЛТП базе

МС СQЛ 2017 стандард едиција (ДреамСпарк Премиум | Ацадемиц Софтwаре Дисцоунтс) је коришћена као платформа на којој је инсталирана ВМарт ОЛТП база података. Виртуална машина базе података је инсталирана на Wиндоwс Сервер 2016 оперативном систему коришћењем ВМwаре виртуализационе платформе. Виртуалној машини је је додељено 80ГБ простора на диску и 8ГБ РАМ меморије. Табеле на слици 1 које су означене жутом тј. тамнијом позадинском бојом су трансформисане у табеле чињеница (*фацт* табеле), док су табеле са белом позадинском бојом трансформисане у табеле димензија (*дименсионс* табеле). Табеле чињеница су садржале догађаје, тј.

половне процесе у облику следећих поља*: пурцхасе, процуремент, лендинг, паументс* итд. Табеле димензија су садржале атрибуте повезане са местима, временима и другим табелама као што су: *продуцтс, вендорс, схипперс* итд.

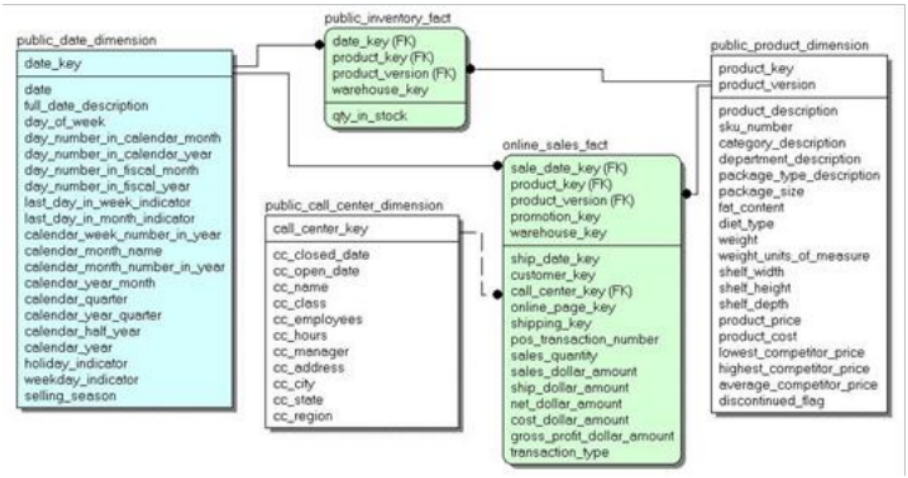
Креирана је и база података ВМартД као дестинациона ОЛАП база, са одговарајућим табелама чињеница и димензија. У окружењу за вршење експеримената, ОЛТП Вмарт база података је коришћена као изворна база из којих су се подаци екстраковали и трансформисали ради пуњења ВМартД ОЛАП базе података коришћењем Мицрософт ЕТЛ апликације - МС СQЛ Сервер Интегратион Сервицес (ССИС) и Пентахо ЕТЛ алата.

Први експеримент је заснован на извршавању ЕТЛ процеса коришћењем Решења 1, код кога су процеси креирани коришћењем скрипти са СQЛ кодом. Други експеримент се састојао од извршавања ЕТЛ процеса употребом Решења 2, које је засновано на коришћењу расположивих ССИС ЕТЛ алата. Трећи експеримент је подразумевао извршење ЕТЛ процеса користећи Решење 3 базирано на бесплатном Пентахо алату (*цоммуниту едитион* фирме Хитацхи Вантара). За сваки од наведених експеримената је вршено мерење брзине ЕТЛ процеса.

Подаци из пет табела ОЛТП базе, као што је приказано на Слици 1 (групација приказана на левој страни), су груписани користећи релације са табелом ***цалл_центер*** (***цц_цласс, цити, стате, регион***) и спојени (трансформисани) у једну димензиону табелу ***публиц_цалл_центер_дименсион*** у ОЛАП бази података. Подаци из шест табела ОЛТП базе приказане на Слици 1 (групација приказана на десној страни) повезани релацијама са табелом ***продуцт*** (***продуцт_цатегору, департмент, пацкаге_типе, пацкаге_сизе, диет_типе, веигхт_унитс***) су спојени у једну димензиону табелу ***публиц_продуцт_дименсион*** у ОЛАП бази података. Подаци две ОЛТП табеле ***онлине_салес_ордер*** и ***онлине_салес_ордер_детаилс*** ће бити трансформисане у ОЛАП табелу чињеница ***онлине_салес_фацт***. Поред тога, подаци из четири групе табела које су сличне табелама са жутом (тамнијом) позадином на Слици 1, ће бити трансформисане и учитане у четири табеле чињеница.

Циљ ЕТЛ процеса се огледа у екстракцији свих података горе наведених табела, извођењем *јоин* операција и учитавањем трансформисаних података у одговарајуће табеле димензија ОЛАП базе података. Сумарно посматрано, подаци из укупно 41 табеле ОЛТП базе ће бити пресликане у 9 ОЛАП димензионих табела.

Дестинационе ОЛАП табеле које су добијене ЕТЛ трансформацијом изворних табела ОЛТП базе података су приказане на слици 2.



Слика 2. ВМарт ОЛТП табеле добијене трансформацијом ОЛТП табела са Слике 1

Анализом табела се може видети да су подаци у атрибутима: ***цц_цласс, цц_циту, цц_стате*** и ***цц_регион*** у табели ***публиц_цалл_центер_дименсион*** екстарковани из атрибута "*наме*" табела: ***цц_цласс, циту, стате*** и ***регион***. На исти начин су изведене и трансформације података из атрибута "*десцриптион*", "*типе*" и "*меасуре*" из ОЛТП табела у ОЛАП табелу ***публиц_продуцт_дименсион***. Такође се може видети да су подаци скоро свих атрибута из ОЛТП табела ***онлине_салес_ордер*** и ***онлине_салес_ордер_детаилс*** спојени у једну ОЛАП табелу чињеница ***онлине_салес_фацт***.

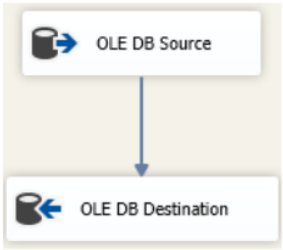
Уколико посматрамо само табеле приказане на горе наведеним сликама, јасно се може видети да су подаци 15 ОЛТП табела трансформисане у 5 ОЛАП табела са истом количином података али другачијом организацијом.

За потребе извршења експеримената изабране су три хардверске платформе са Мицрософт Виндовс 10 оперативним системом на којима су постављене виртуалне машине:

- Рачунар са И7-11800Х, 2.3ГХз процесором и ССД диском
- Рачунар са и7-9750Х, 2.6ГХз процесором и ССД диском
- Рачунар са и7-2600К, 3.7 ГХз процесором и ХДД диском

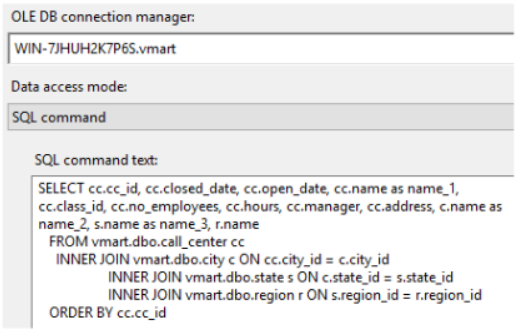
5.1.1. ЕТЛ решење 1 – коришћење СQЛ упита

СQЛ Сервер Интегратион Сервицес (ССИС) конзола поседује уграђене компоненте за креирање ЕТЛ послова (јобс). Они се могу дизајнирати коришћењем две компоненте, једне за упис СQЛ кода за екстракцију и трансформацију изворних података, и друге за мапирање колона *Селецт* упита над изворним табелама са колонама у дестинационим табелама, ради правилног пуњења табела *Дата Варехоусе* базе. На слици 3 су приказане су приказане обе компоненте *ОЛЕ ДБ Соурце* и *ОЛЕ ДБ Дестинатион* повезане одговарајућом конекцијом.



Слика 3. ОЛЕ ДБ Соурце и ОЛЕ ДБ Дестинатион компоненте

Пример *Селецт* СQЛ кода за екстракцију и трансформацију изворних података који се налази на ОЛЕ ДБ Соурце компоненти је приказан на слици 4.

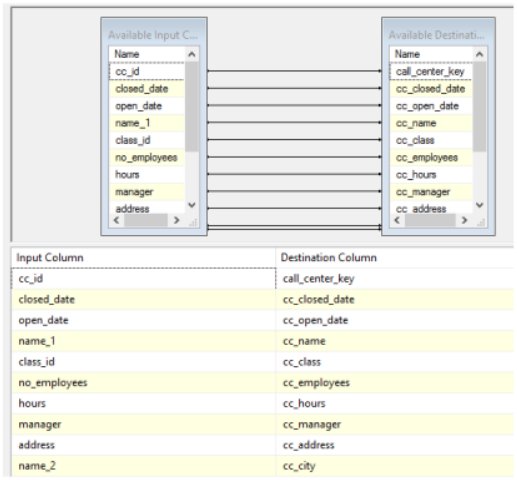


Слика 4. ОЛЕ ДБ Соурце едитор

У оквиру процеса конфигурације се прво бира одговарајући извор базе података, што је у овом случају ОЛТП база. Приказани упит селектује податке из табела повезане са

цалл_центер табелом, а резултујуће колоне су изабране из табела које учествују у *Join* процесу.

У другој компоненти, ОЛЕ ДБ Дестинатион, прво се врши избор дестинационе базе тј. ОЛАП базе као и избор табеле у коју ће се подаци који су екстарковани и трансформисани, учитати. Имајући у виду да се називи колона у упиту над изворном табелом могу разликовати од назива колона у дестинационој табели, могуће је тачно дефинисати која се изворна колона упарује са одговарајућом дестинационом колоном. На слици 5 је приказан конфигурациони прозор за мапирање колона у *публиц_цалл_центер_димензион* табели.



Input Column	Destination Column
cc_id	call_center_key
closed_date	cc_closed_date
open_date	cc_open_date
name_1	cc_name
class_id	cc_class
no_employees	cc_employees
hours	cc_hours
manager	cc_manager
address	cc_address
name_2	cc_city

Слика 5. Мапирање колона за табелу публиц_цалл_центер_димензион

За сваку дестинациону табелу у припремљеној *Дата Wарехоусе* бази је дизајниран по један *јоб*.

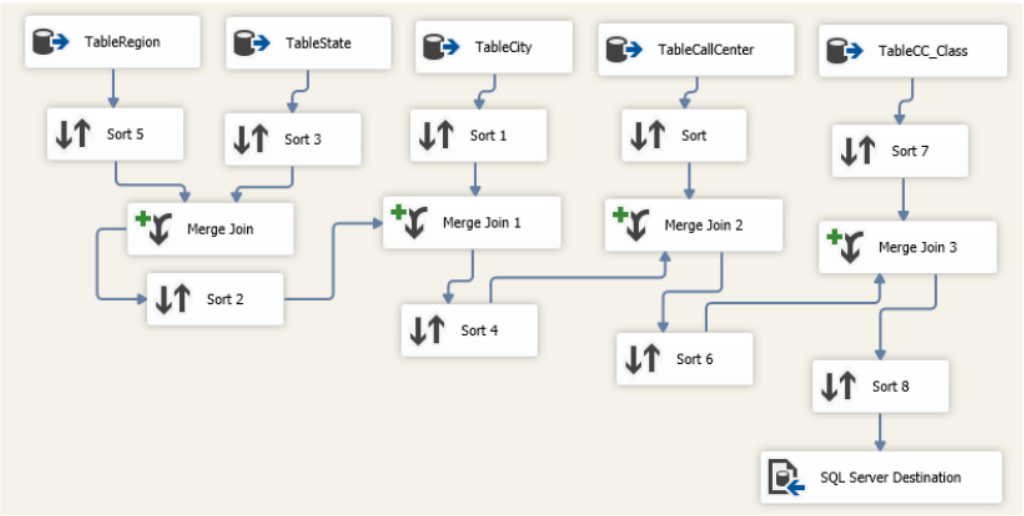
5.1.2. ЕТЛ Решење 2 - коришћење ССИС функционалних компоненти

За креирање ЕТЛ послова се могу користити уграђене функционалности које се налазе у оквиру кутије алата у ССИС менаџмент конзоли. Једноставним постављањем и конфигурисањем одговарајућих доступних алата, без потребе познавања СQЛ

програмског језика се може дизајнирати сваки од потребних послова. Када поредимо овај приступ са претходно описаним, можемо закључити да се ЕТЛ решење 1 извршава на самој бази података док се ЕТЛ решење 2 извршава у оквиру ЕТЛ апликације. Можемо такође да уочимо да се у овом решењу ОЛЕ ДБ Соурце компонента у ствари табела а не упит, као што је случај у ЕТЛ решењу 1.

Коришћењем ЕТЛ апликације процес спајања табела се изводи коришћењем две ОЛЕ ДБ изворне компоненте, које у суштини екстракују податке из две табеле одвојено ради сортирања компоненти. Свака изворна табела мора да буде сортирана по колонама које учествују у процесу спајања да би се резултат сортирајућих компоненти спојио *Мерге Јоин* компонентом. Када постоји потреба да се изврши неколико операција спајања табела, потребно је користити две *Сорт* и једну *Мерге Јоин* операцију за свако спајање. За имплементацију овог решења је коришћено неколико ЕТЛ компоненти из великог броја доступних компоненти које нуди ССИС.

На слици 6 је приказана шема компоненти ЕТЛ апликације које су коришћене за екстракцију, трансформацију и учитавање података у публиц_цалл_центер_димензион табелу.

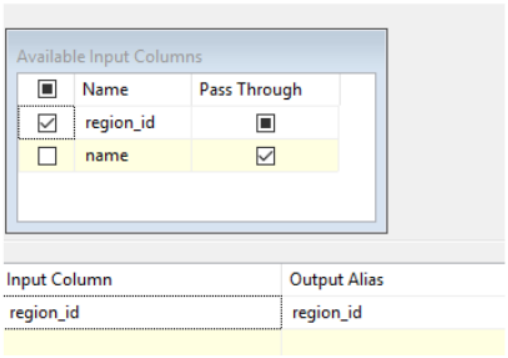


Слика 6. Пример ССИС посла у Дата Флоw Едитору

Анализом приказане шеме на слици 6 се може видети да је пет ОЛЕ ДБ Соурце компоненти коришћено за пет изворних табела, као и четири *Мерге Јоин* компоненти за извршавање *Иннер-јоин* операција. Да би се подаци спојили *Мерге Јоин* компонентом, неопходно је претходно извршити сортирање података изворних табела. Како ССИС не

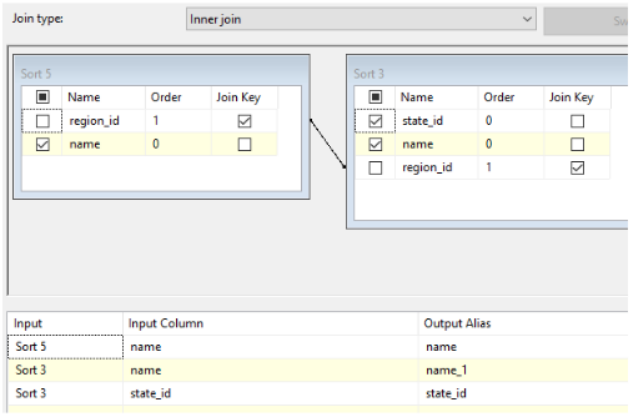
користи индексе колона, можемо очекивати слабије перформансе извођења сортирања у овом случају.

На слици 7 је представљен конфигурациони прозор *Сорт* компоненте, на којој се могу видети доступне опције, т.ј. може се изабрати колона, или колоне по којима ће се извршити сортирање. Потребно је водити рачуна при селектовању колона у овом прозору, како би биле изабране само неопходне колоне, да не би дошло до учитавања непотребних података и тиме повећало оптерећење и време за извршење процеса.



Слика 7. Сорт Трансформатион Едитор

По подешавању компоненти сортирања, потребно је у оквиру *Мерге Јоин* компоненте изабрати и све оне колоне чији ће садржај бити укључен у излазу ове компоненте. Овај избор треба направити у Мерге Јоин Трансформатион Едитор прозору који је приказана на слици 8, избором одговарајућих поља са леве стране изворних табела.



Слика 8. Мерге Јоин Трансформатион Едитор

Последња компонента приказана у шеми ЕТЛ апликације

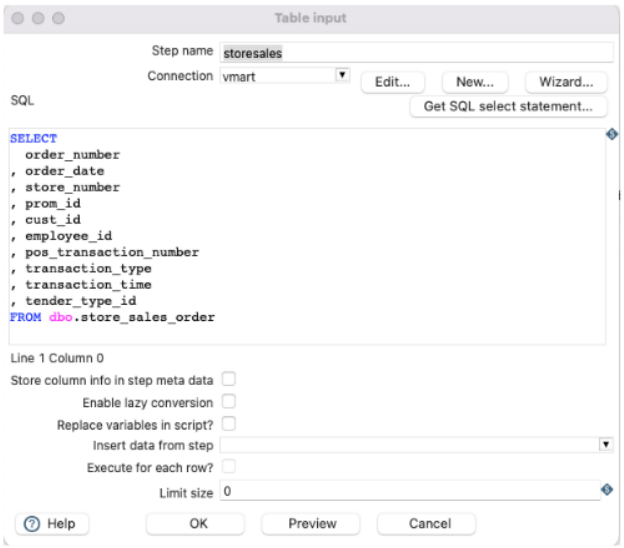
Пентахо Дата Интегратион (ПДИ) Цоммуниту Едитион је бесплатан ЕТЛ алат. Овај алат садржи компоненте за визуелни дизајн процесних токова као и конекторе за повезивање са различитим изворима података. У конкретном случају, коришћен је ПДИ верзије 9.4 за чији рад је препоручен Јава СДК верзије 18. У циљу извођења експеримента, на засебној виртуалној машини инсталиран је оперативни систем Виндовс 10, Орацле Јава СДК верзије 18 за Виндовс, ПДИ верзије 9.4 (хттпс://www.хитацхивантара.цом/ен-ус/продуцтс/пентахо-плус-платформ/дата-интегратион-аналутицс/пентахо-цоммуниту-едитион.хтмл), као и Нативе МС СQЛ Сервер ЈДБЦ драјвер (јтдс-1.3.1.јар) кога је потребно сачувати у под-директоријуму либ овог алата.

Пентахо Дата Интегратион (ПДИ) нуди две основне могућности у свом окружењу:

1. **Јоб** – овај део подразумева спајање више трансформација како би оне формирале једну целину и извршиле комплетан посао вишеструких трансформација и
2. **Трансформатион** – у њима се извршавају конкретне трансформације над подацима (сортирање, спајање, додавање и уклањање колона, пуњење дестинационих табела, итд.).

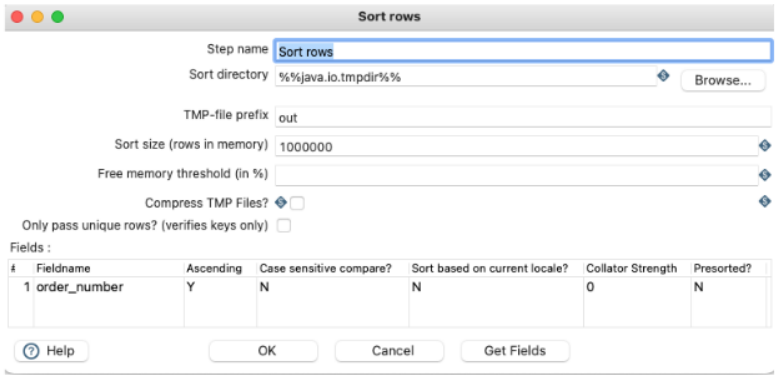
Слично као у решењу 2, у решењу 3 се користе следеће трансформације (кораци):

- **Табле инпут** – овде се бира конекција ка бази (у нашем случају Вмарт) и конкретна табела из које се читају изворни подаци. На слици 10 је показан едитор подешавања ове трансформације.



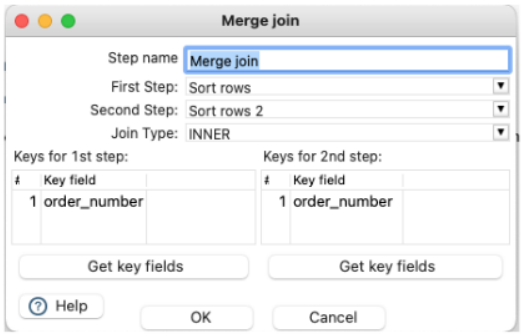
Слика 10. Табле инпут Едитор

- **Сорт powc** – ова трансформација служи да редове из скупа података који се налазе на њеном улазу поређа по растућем / опадајућем редоследу једне или више колона, при чему се у едитору трансформације може бирати да ли се сортирање врши водећи рачуна о великим / малим словима. На излазу трансформације се налази исти број колона као и на њеном улазу, без потребе да се то децидивно наведе. Едитор трансформације је показан на слици 11, где се види да податке треба поређати у растућем редоследу (Асцендинг = Y), не водећи рачуна о великим и малим словима (Цасе сенситиве цомпаре = Н).



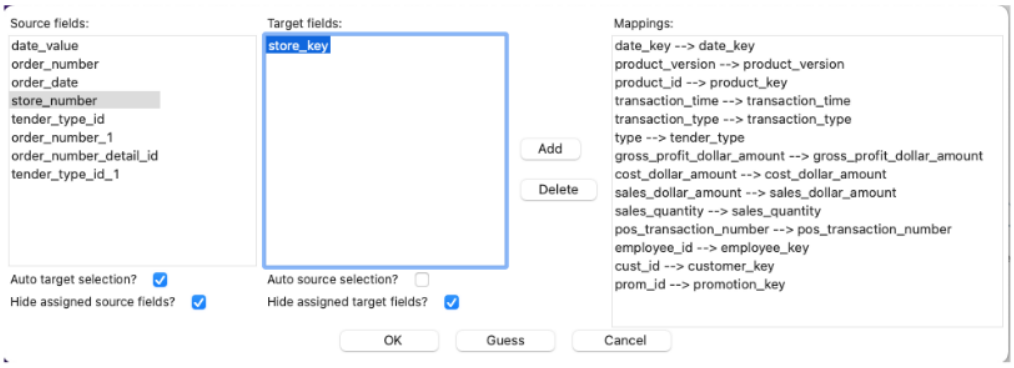
Слика 11. Сорт powc Едитор

- **Мерге јоин** – ова трансформација омогућава повезивање података из два извора по услову који се задаје у едитору трансформације. Најчешћи услов је да су вредности у једној или више колона из једног извора једнаки вредностима из једне или више колона другог извора. Трансформација може сама да "открије" по којим колонама треба повезати податке (опција Гет кеу фиелдс), мада и корисник може сам да дефинише те колоне. Едитор ове трансформације је показан на слици 12, где се види да се на улазу налазе два извора: Сорт powc и Сорт powc 2 и да податке из ових извора треба спојити по колони ордер_нумбер.



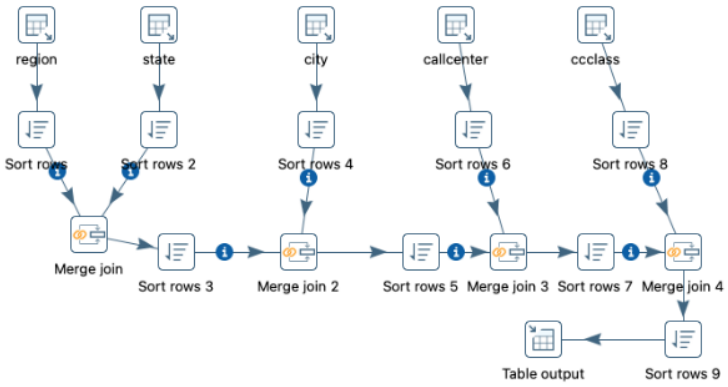
Слика 12. Мерге јоин Едитор

- **Табле оутпут** – ова трансформација омогућава учитавање трансформисаних података у дестинациону табелу ОЛАП базе. Најчешће се приликом *мерге-јоин* трансформација у претходним корацима креира већи број колона од потребних за дестинациону табелу, тако да је у овој трансформацији потребно мапирати које колоне из трансформисаних података треба учитати у одговарајуће колоне дестинационе табеле. На слици 13 је показан едитор мапирања података. На слици се јасно виде колоне трансформисаних података у листи Соурце фиелдс и колоне из дестинационе табеле у листи Таргет фиелдс. Након избора колоне из једне листе и колоне из друге листе и кликом на дугме Адд у листи Маппингс се појављује изабрано мапирање.



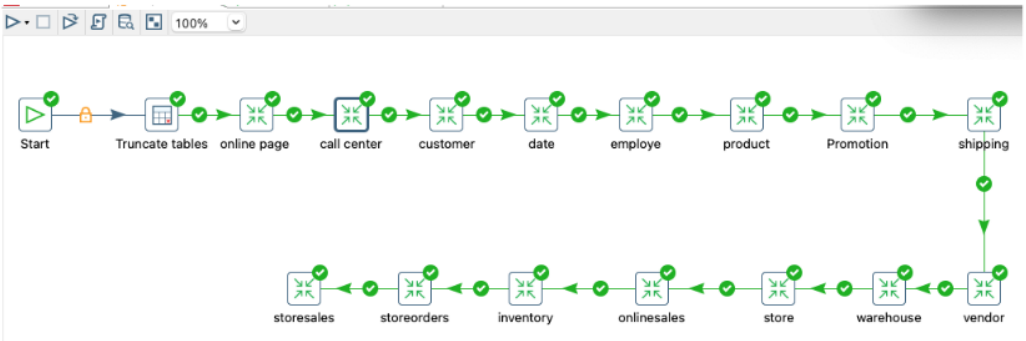
Слика 13. Едитор мапирања колона података након трансформација и дестинационе табеле

Као и у решењу 2, у овом решењу се трансформација података неколико табела ОЛТП базе у једну димензиону табелу обавља истим бројем корака. Тако је на слици 14 приказана трансформација еквивалентна трансформацији у ССИС ЕТЛ алату на слици 6.



Слика 14. Пример ПДИ трансформације 5 ОЛТП у једну ОЛАП табелу

Како је циљ ЕТЛ посла да одговарајући скуп ОЛТП табела трансформише у листу *димензион* и *фацт* ОЛАП табела, то је задатак да се појединачне трансформације (као што је раније приказано показана на слици 9) поређају и изврше једна за другом. Најчешће се прво обављају трансформације пуњења *Димензион* табела, а затим трансформације пуњења *фацт* табела. Комплетан посао (*јоб*) креиран у Пентахо ЕТЛ алату је показан на слици 15. Иконица са називом ***цалл центар*** на слици 15 представља трансформацију показану на слици 14.



Слика 15. Извршење ЕТЛ посла трансформација ОЛТП табела у листу Димензион и Фацт ОЛАП табела

5.2. Резултати експеримената

Експерименти су конципирани на ЕТЛ процесу који се састојао од неколико задатака (таскс) и то:

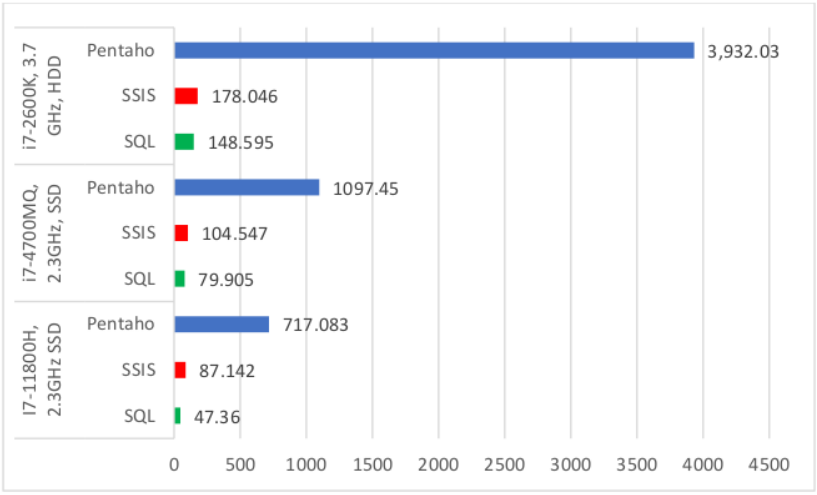
- Уклањање свих података из табела (трунцатинг)
- Трансформација и учитавање података у Димензионе табеле (два задатка)
- Трансформација и учитавање података у табеле чињеница (два задатка)

Посматрано са аспекта количине података у табелама, девет од укупно једанаест димензионих табела су садржале мање од 2,000 записа (цалл_центер_дименсион, онлине_паге_дименсион, дате_дименсион, емплоуее_дименсион, промотион_дименсион, схиппинг_дименсион, вендор_дименсион, варехоусе_дименсион, сторе_дименсион) и ова група табела је означена под називом ДимТ1. Преостале две димензионе табеле су садржале 50,000 и више записа (цустомер_дименсион, продуцт_дименсион) и означене су под називом ДимТ2. Табеле чињеница су подељене у две групе, ФацтТ1 у коју спадају две табеле (инвентору_фацт и сторе_ордерс_фацт) са по 300,000 записа и ФацтТ2 у коју спадају две табеле (онлине_салес_фацт и сторе_салес_фацт) са по 5,000,000 записа. У табели 4 су приказана времена извршења ЕТЛ послова: Трунцате, ДимТ1, ДимТ2, ФацтТ1, ФацтТ2 користећи СQЛ код, ССИС ЕТЛ и Пентахо ЕТЛ алате . Сваки посао је извршен на три различите хардверске платформе.

	И7-11800Х, 2.3ГХз ССД			**и7-4700МQ, 2.3ГХз, ССД**			**и7-2600К, 3.7 ГХз, ХДД**		
	СQЛ	ССИС	Пентахо	СQЛ	ССИС	Пентахо	СQЛ	ССИС	Пентахо
трунцате	0.079	0.047	0.198	0.109	0.094	0.547	0.156	0.156	3.653
ДимТ1	1.344	7.345	13.118	1.843	10.219	13.256	4.516	9.484	29.031
ДимТ2	0.89	3.609	8.347	1.344	5.156	10.587	2.907	8.39	25.819
ФацтТ1	1.907	3.625	14.198	3.75	4.812	19.718	4.922	9.421	45.496
ФацтТ2	43.14	72.516	681.222	72.859	84.266	1053.342	136.094	150.595	3,828.027
Тотал	**47.36**	**87.142**	**717.083**	**79.905**	**104.547**	**1097.45**	**148.595**	**178.046**	**3,932.026**

Табела 4. Време извршења ЕТЛ послова у секундама

На графикону 1 су приказана укупна времена извршења свих ЕТЛ послова за сваки од алата и за све три три различите хардверске платформе.



Графикон 4. укупна времена извршења свих ЕТЛ послова за сваки од алата и за све три три различите хардверске платформе у секундама

Генерално посматрано, као што се види из добијених резултата експеримената, може се закључити да укупно време извршења ЕТЛ послова на датом примеру табела није превелико, свега неколико минута код СQЛ кода и ССИС ЕТЛ алата, док је код Пентахо алата у рангу од 10 до 66 минута. Такође, резултати показују да је време извршења ЕТЛ послова користећи СQЛ код и до 40% краће у односу на ЕТЛ послове који се заснивају на ССИС апликацији са адекватним сортирањима и спајањима података. Разлика у брзини извршења ЕТЛ процеса између ССИС са једне и Пентахо ЕТЛ алата са друге је и преко 8 пута већа у корист ССИС. ЕТЛ послови су се најбрже извршавали на рачунару са најбољим процесором, као што је и било очекивано.

5.3. Закључак

Време потребно за трансформацију и учитавање података из ОЛТП у ОЛАП табеле које садрже по неколико милиона редова је на нивоу неколико минута код СQЛ кода и ССИС, док је код Пентахо ЕТЛ алата између 10 и 66 минута. То указује да се најбољи резултати добијају када се користи иста платформа за базу података, ЕТЛ алат и оперативни систем, због компатибилности интерфејса. Могуће је да резултати добијени наведеним експериментима могу да се разликују у случају када су платформе база података и ЕТЛ решења постављена на различитим рачунарима, због кашњења у мрежном сегменту.

Имајући у виду да је време потребно за извршавање ЕТЛ послова у случајевима са вишемилионским записима у табелама прилично кратко, препорука је да се ЕТЛ процеси покрећу чешће ради извршавања пословне аналитике над свежим подацима у ОЛАП бази података.

На избор одговарајућег решења за пословање може утицати и цена самог решења. Мицрософт СQЛ Сервер је подложан лиценцирању. ССИС компонента је саставни део СQЛ Сервера. У тренутку писања ове дисертације, на основу расположивих информација на web сајту произвођача, Стандард и Ентерприсе едиције СQЛ сервера се лиценцирају по *Пер Цоре* и Сервер + ЦАЛ (*Цлиент Аццесс Лиценсе*) моделима користећи различите програме лиценцирања. Цена лиценце СQЛ Сервер Ентерприсе едиције која покрива 2 процесорска језгра кроз Волуме Лиценсинг канал износи 15,123 УСД. Са друге стране Пентахо Дата Интегратион (ПДИ) Цоммунити Едитион је бесплатан ЕТЛ алат.

У зависности од расположивог буџета и прихватљивих перформанси различитих ЕТЛ алата, пословна предузећа могу донети одлуку о избору одговарајућег алата. Мала и средња предузећа, уколико немају потребу за трансформацијом огромне количине података, а ажурирање ОЛАП базе се не мора вршити фреквентно у току дана, могу изабрати бесплатно ЕТЛ решење.

6. Перформансе БИ алата

Пословна интелигенција (на енглеском *Бусинесс Интеллигенце* или БИ) представља један од кључних сегмената савременог пословног окружења. Она представља скуп различитих методологија и софтверских решења за анализу података. Код извођења БИ пројеката је потребно прво извршити идентификацију потребних историјских података. Они могу доћи из различитих извора, како из интерних трансакционих база података предузећа, тако и екстерних података који могу доћи са Интернет мреже. Њихово "чишћење" кроз процес трансформације и учитавање у ОЛАП базу података се врши уз помоћ ЕТЛ алата. ОЛАП база података се назива и *Дата Варехоусе* база. Она је оптимизована за брзо читање података и подаци су у њој складиштени по унапред дизајнираној архитектури. БИ алати се најчешће користе од стране пословних аналитичара и пружају могућност различитих анализа над подацима из *Дата Варехоусе* базе података. Ови алати су дизајнирани на начин да не захтевају техничка знања крајњих корисника. Крајњи корисници могу на једноставан начин да изаберу формате приказа упита из базе, упаре податке који су међусобно зависни, подесе одговарајуће филтере, изврше различите типове анализа и креирају одговарајуће извештаје. Унапред припремљени формати извештаја потребни за доношење одлука могу бити смештени на интерне портал локације у предузећу. На тај начин им доносиоци одлука могу лако приступити и уз коришћење додатних опција у оквиру самих припремљених извештаја, извршити додатна филтрирања и подешавања ради увида у информације из различитих перспектива.

БИ такође укључује процесе попут анализе пословних перформанси, извештавања и управљања перформансама. Кроз ове процесе, предузећа могу идентификовати шаблоне, анализирати кључне показатеље перформанси (КПИ – *Кеу Перформанце Индицаторс*) и оптимизовати операције.

Предности коришћења БИ су вишеструке. Пословна предузећа на овај начин стичу способност брзе реакције на промене у пословном окружењу, могу проактивно идентификовати нове прилике за раст и повећање профита, као и смањити трошкове и унапредити укупну ефикасност пословања. Због могућности предвиђања будућих трендова на основу предиктивне анализе историјских података, предузећа могу пре своје конкуренције покренути одговарајуће активности ради стицања компаративне предности на тржишту.

Када се разматра будућност пословне интелигенције, она ће наставити да еволуира. Очекује се да ће технологије вештачке интелигенције (*Артифициал Интеллигенце* - АИ) и машинског учења (*Мацхине Леарнинг*) постати саставни део БИ, пружајући још софистицираније анализе и аутоматско откривање узорака. Такође, мобилни приступ БИ

6.1. Окружење за извођење експеримената

Data Warehouse база података ВмартД која је коришћена у представљеној ЕТЛ студији у поглављу 5 овог рада, је коришћена и за потребе експеримената поређења перформанси БИ алата.

ВмартД база је постављена на ВМware виртуалној машини са инсталираним Мицрософт Windows 2016 Сервер Стандард оперативним системом у оквиру Мицрософт СQЛ Сервера 2017. При инсталацији окружења је коришћена лиценца за образовне установе. Виртуелној машини је додељено 6 ГБ РАМ и 120ГБ простора на диску.

Посебне виртуалне машине са одговарајућим БИ алатима су припремљене за потребе експеримената покретања упита ка ВмартД бази података и то:

- Виртуална машина са инсталираним Windows 10 Про оперативним системом и Мицрософт ПоверБИ десктоп апликацијом, којој је додељено 4 ГБ РАМа и 60 ГБ простора на диску.
- Виртуална машина са инсталираним Windows 10 Про оперативним системом и бесплатним Пентахо Бусинесс Аналутицс Цоммуниту Едитион алатом, којој је додељено 4 ГБ РАМ-а и 60 ГБ простора на диску.

Обе виртуалне машине са БИ алатима су повезане са ВмартД базом података користећи ВМware виртуалне мрежне интерфејсе са НАТ подешавањем.

Дизајнирана су 4 различита упита ка ВмартД бази за извођење експеримената и имплементирани су користећи одговарајуће табеларне и графичке интерфејсе у оквиру оба БИ алата. Структура упита је представљена у даљем тексту, користећи СQЛ језик.

Упит 1

Постављени циљ Упита 1 (Q1) је тестирање перформанси упита и приказа резултата упита у једној табели чињеница *онлине_салес_фацт* од 5 милиона редова, спојеној *иннер-јоин* везом са једном колоном у димензионој табели *публиц_дате_дименсион* која се састоји од 1,826 редова. Поље вредности профита УСД валуте је коришћено за сумирање укупне вредности оствареног профита, груписане по колони годишњих квартала из наведене димензионе табеле и растућим редоследом временске димензије. Дефинисани упит треба да прикаже резултате оствареног профита онлајн продаје по годишњим кварталима. Формат приказа је табеларни у оба БИ алата.

СЕЛЕЦТ сум(о.гросс_профит_доллар_амоунт)

```
СЕЛЕЦТ сум(ссф.гросс_профит_доллар_амоунт) АС сум_профит,
с.сторе_наме, с.Сторе_Циту, с.сторе_регион, д.цалендар_уеар_qуартер
ФРОМ сторе_салес_фацт ссф
ИННЕР ЈОИН сторе_димензион с ОН
        ссф.сторе_кеу = с.сторе_кеу
ИННЕР ЈОИН публиц_дате_димензион д ОН
        ссф.дате_кеу = д.дате_кеу
ГРОУП БУ с.сторе_наме, с.Сторе_Циту, с.сторе_регион,
д.цалендар_уеар_qуартер
ОРДЕР БУ с.сторе_наме, д.цалендар_уеар_qуартер;
```

Упит 4

Упит 4, слично као и претходни, има за циљ враћање већег сета резултантних података.

Извршење упита је над табелом чињеница *сторе_ордерс_фацт* која садржи 300,000 записа трансакција поруџбина у радњама и која је спојена *иннер-јоин* везама са две димензионе табеле, *публиц_емплоуее_димензион* и *публиц_вендор_димензион* преко колна *емплоуее_кеу* и *вендор_кеу*, респективно. Циљ упита је да прикаже укупну вредност остварених поруџбина, груписану по титулама запослених који су остварили продају и произвођачима поручене робе.. Редослед приказа резултата је опадајући по укупној вредности поруџбина и растући по титулама запослених и произвођачима поручене робе. Резултати упита ће бити приказани у табеларном формату у оба БИ алата.

```
СЕЛЕЦТ сум(соф.тотал_ордер_цост) ас Тотал_Ордер_Валуе,
е.јоб_титле, в.вендор_наме
ФРОМ сторе_ордерс_фацт соф
ИННЕР ЈОИН публиц_емплоуее_димензион е ОН
        соф.емплоуее_кеу = е.емплоуее_кеу
ИННЕР ЈОИН публиц_вендор_димензион в ОН
        соф.вендор_кеу = в.вендор_кеу
ГРОУП БУ е.јоб_титле, в.вендор_наме
ОРДЕР БУ Тотал_Ордер_Валуе ДЕСЦ, е.јоб_титле, вендор_наме;
```

За сваки упит ће бити извршена по три мерења у сваком БИ алату и просечна вредност резултата ће бити приказана као резултантна.

Ради успостављања конекције између обе виртуалне машине конфигурисан је Виндовс фиревалл на СQЛ сервер виртуалној машини да пропусти сву комуникацију ка и од виртуалних машина са БИ алатима.

За потребе извршења експеримената изабране су три хардверске платформе на којима су постављене виртуалне машине:

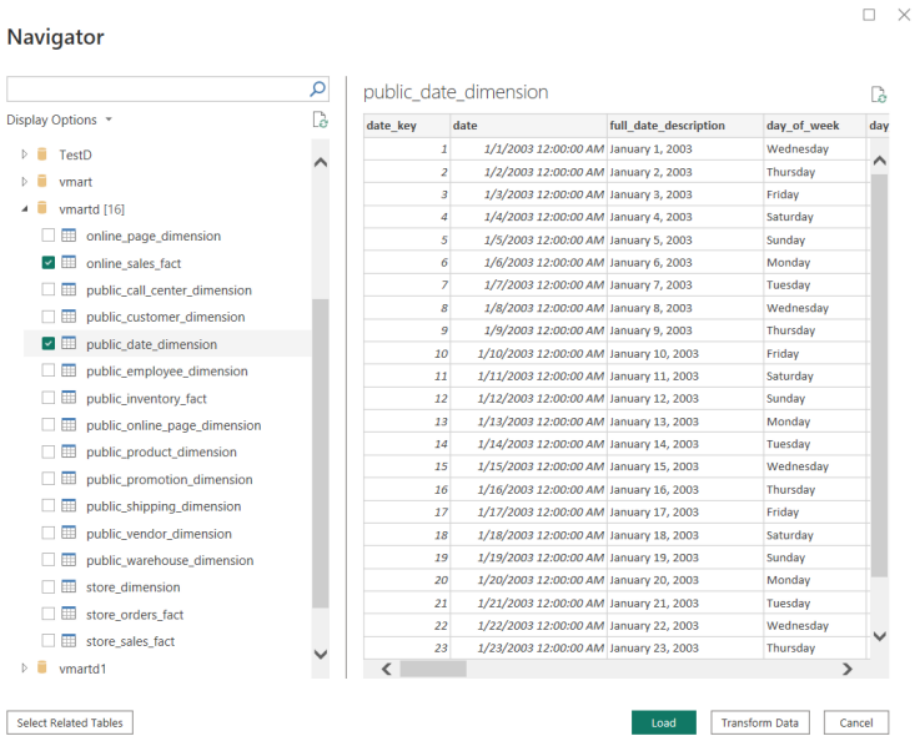
- Рачунар са И7-11800Х, 2.3ГХз процесором и ССД диском,
- Рачунар са и7-9750Х, 2.6ГХз процесором и ССД диском,
- иМац рачунар (21.5-инцх, 2017) са и5-7360У 2.3ГХз процесором.

6.1.1. БИ решење 1 – Повер БИ Десктоп

Повер БИ Десктоп представља Мицрософт БИ алат који омогућава анализу и визуализацију података добијених из различитих извора. Он се може бесплатно преузети и инсталирати на рачунару под Виндовс оперативним системом. Повер БИ Десктоп омогућава повезивање на више различитих извора података који се могу комбиновати, односно моделовати у одговарајући дата модел. Такав дата модел се може користити за креирање различитих визуализација података или колекција визуализација које се могу поделити са другим корисницима у организацији. Најчешће се извештаји креирају од стране пословних аналитичара, а потом се деле са другим корисницима користећи Повер БИ сервис. Повер БИ сервис подлеже правилима лиценцирања. За потребе вршења експеримената у овој студији је коришћена само бесплатна Повер БИ десктоп апликација.

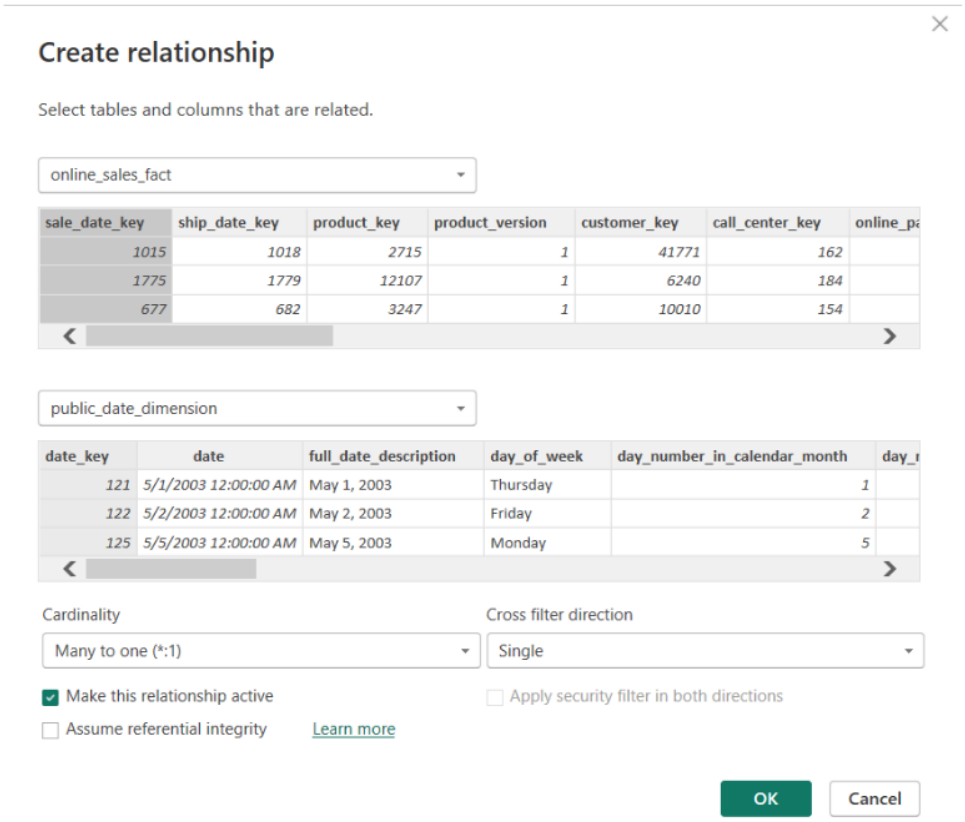
У Повер БИ Десктоп конзоли су креирана сва четири претходно дефинисана упита.

Упит 1 је креиран на следећи начин. Избором *Гет Дата* опције је изабран СQЛ Сервер као извор података, уписана је ИП адреса сервера и изабран је *Дирецт Qуеру* режим повезивања. У новоотвореном прозору Навигатор су изабране *онлине_салес_фацт* и *публиц_дате_дименсион* табеле ВмартД базе података, као што је приказано на слици 16.



Слика 16. Навигатор прозор за избор табела из базе података

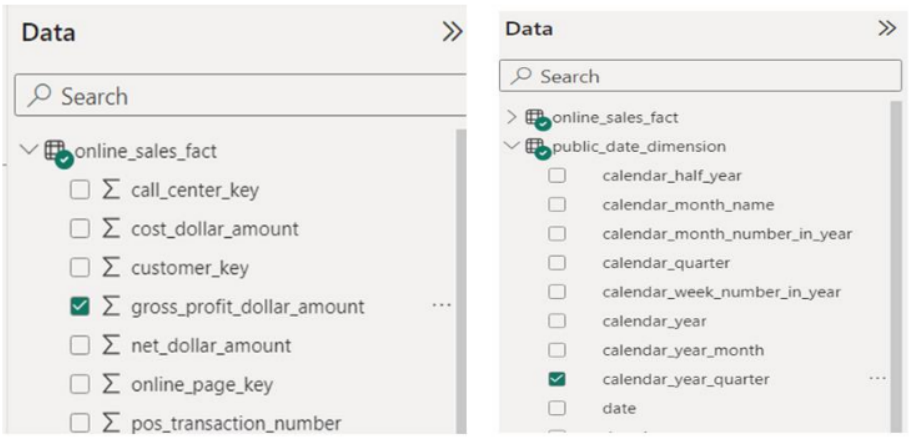
Ради креирања релација између наведених табела коришћена је опција *Манаге релатионсхипс* са *Моделинг* рибона. Табеле су повезане релацијом *Ману то оне* користећи *Сале_Дате_Кеу* и *Дате_Кеу колоне*, као што је приказано на слици 17.



Слика 17. Креирање релација између табела у прозору Цреате Релатионсхип

Из секције са визуализацијама је изабран табеларан приказ извештаја.

У секцији Дата су изабрана поља/колоне извештаја, *∑гросс_профит_доллар_амоунт* из табеле *онлине_салес_фацт* и *цалендар_yеар_qуартер* из табеле *публиц_дате_дименсион,* као што је приказано на слици 18.



Слика 18. Избор колона извештаја

Упит 2 је дефинисан на сличан начин, избором истих табела као у Упиту 1, при чему је у секцији са визуализацијама изабран *Клустеред Цолумн Цхарт* приказ извештаја. У секцији Дата су изабрана поља/колоне извештаја, *∑гросс_профит_доллар_амоунт* из табеле *онлине_салес_фацт* и *цалендар_yеар* из табеле *публиц_дате_димензион*.

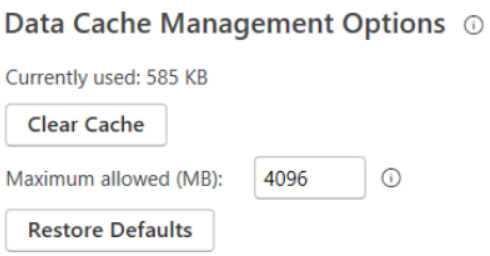
За конфигурацију упита 3 су изабране табеле *сторе_салес_фацт*, *сторе_димензион* и *публиц_дате_димензион*. Креиране су две релације између табела. Прва релација, *Ману то оне* је дефинисана између табела *сторе_салес_фацт* и *сторе_димензион* користећи колоне *сторе_кеy* у обе табеле. Друга релација, *Ману то оне* је дефинисана између табела *сторе_салес_фацт* и *публиц_дате_димензион* користећи колоне *дате_кеy* у обе табеле. Из секције са визуализацијама је изабран табеларан приказ извештаја. Селектоване су следеће колоне извештаја: *∑гросс_профит_доллар_амоунт* из табеле *сторе_салес_фацт; сторе_наме, сторе_циту* и *сторе_регион* из табеле *сторе_димензион; цалендар_yеар_qуартер* из табеле *публиц_дате_димензион.*

Приликом конфигурације упита 4 су изабране табеле *сторе_ордерс_фацт*, *публиц_емплоyее_димензион* и *публиц_вендор_димензион*.

Креиране су две релације између наведених табела. Прва релација, *Ману то оне* је дефинисана између табела *сторе_ордерс_фацт* и *публиц_емплоyее_димензион* користећи колоне *емплоyее_кеy* у обе табеле. Друга релација, *Ману то оне* је дефинисана између табела *сторе_ордерс_фацт* и *публиц_вендор_димензион* користећи колоне *вендор_кеy* у обе табеле. Из секције са визуализацијама је изабран табеларан приказ извештаја. За колоне извештаја су изабране *∑тотал_ордер_цост* из табеле *сторе_салес_фацт, јоб_титле* из табеле *публиц_емплоyее_димензион* и *вендор_наме* из табеле *публиц_вендор_димензион.*

Сваки од наведених упита је сачуван у оквиру засебног фајла на хард диску.

Повер БИ Десктоп поседује уграђену компоненту Перформанце Аналузер, који се користити за мерење времена извршења и приказа упита. Избором опције Старт Рецординг и потом Рефресх се може видети време извршења упита у мс. Да би се избегло коришћење кешираних података из ранијих упита, пре сваког мерења из менија Филе – Оптионс анд Сеттингс – Оптионс, у секцији Дата Лоад, ће се испразнити сачувани кеш подаци, као што је приказано на слици 19.



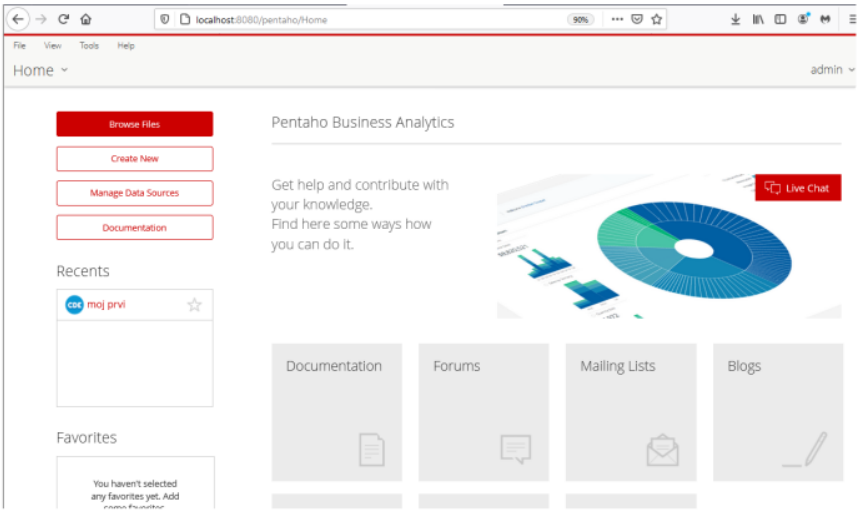
Слика 19. Пражњење кеша из Дата Цацхе манагемент опција

6.1.2. БИ Решење 2 – Пентахо Аналутицс

Пентахо представља софтверску платформу која се користи за рад са подацима, почевши од приступа, интеграције, манипулације до визуализације и анализе података. Подаци могу бити смештени на различитим изворима укључујући поред традиционалних база података и изворе као што су Хадооп кластери, НоСКЛ базе и складишта у *цлоуд*-у. Пентахо Аналутицс је компонента Пентахо платформе која се користи за анализу и визуализацију података. За припрему једноставнијих извештаја нису потребне програмерске вештине. Уколико је потребно дизајнирати комплексније извештаје, Пентахо нуди могућност коришћења апликативних програмских интерфејса, мада су у том случају потребне извесне програмерске вештине.

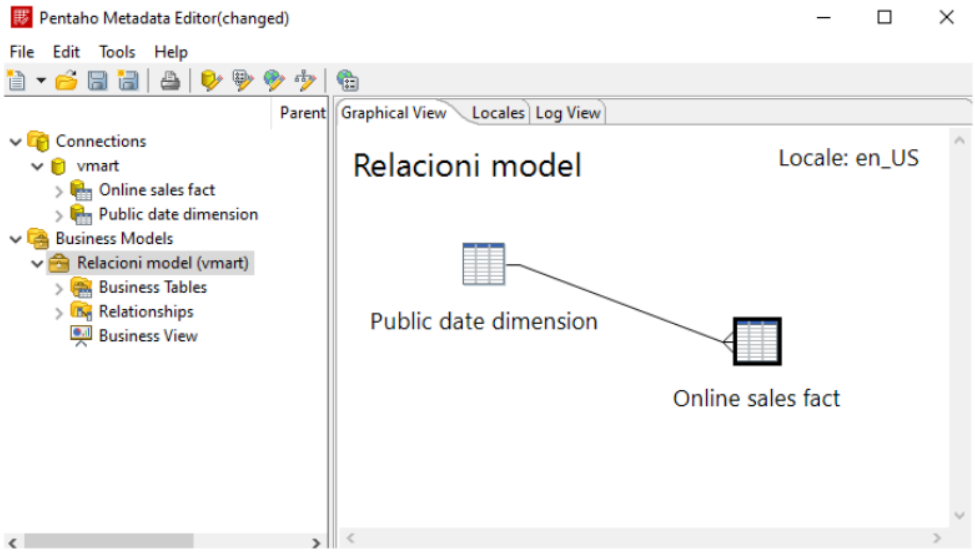
Пентахо софтверска платформа долази у две едиције, бесплатна Пентахо Цоммуниту едиција и Пентахо Ентерприсе едиција базирана на претплати. За потребе извршења експеримената је коришћена Пентахо Цоммуниту едиција инсталирана на Виндовс 10 Про оперативном систему.

На виртуалној машини са Виндовс 10 оперативним системом је извршена претходна припрема која укључује покретање Веб апликационог сервера Томцат. Почетна web страна Пентахо Бусинесс Аналутицс је приказана на слици 20.



Слика 20. Почетна web страна Пентахо Бусинесс Аналутицс

Алат Пентахо Метадата едитор је коришћен за креирање конекције ка ВмартД бази података као и дефинисање релационих модела за сваки од 4 предметна извештаја. Поглед на Пентахо Метадата едитор са креираним релационим моделом за Упит 1 је приказан на слици 21, док је креирање релација између табела користећи Пентахо Метадата Едитор приказано на слици 22.

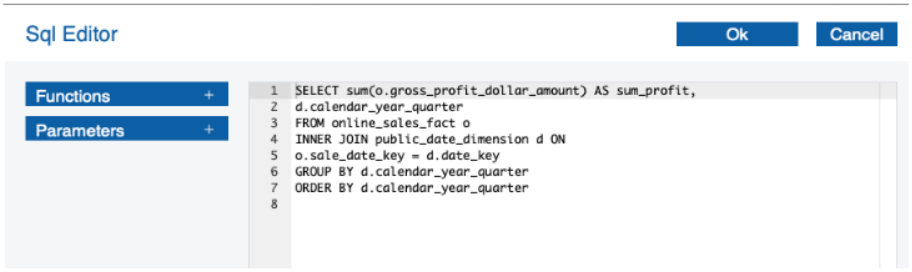


Слика 21. Пентахо Метадата Едитор за дефинисање Метадата Соурце у Пентахо БИ



Слика 22. Креирање релација између табела у прозору Цреате Релатионсхип

Упити су на Пентахо БИ серверу креирани директно, уписом СQЛ наредби над већ снимљеном конекцијом (*датасоурце* конекција креирана у метадата едитору) у пољу *Qуеру* у *Датасоурце* Панелу БИ сервера. Пример уписа упита 1 је приказан на слици 23.



Слика 23. Упис упита у пољу Куереу у Датасоурце Панелу

Пентахо Бусинесс Аналутицс сервер не нуди уграђену могућност приказивање времена извршења процеса, од тренутка покретања упита ка бази података, до приказа извештаја на web страници. Да би се измерило време трајања овог процеса коришћен је следећи приступ: потребно време да Пентахо Бусинесс Аналутицс прикаже web страницу са агрегираним подацима на основу задатог упита се добија одузимањем системског времена у тренутку у када је извештај приказан и системског времена у тренутку када је покренут упит ка серверу база података. Ради мерења времена извршења комплетног процеса креирана је Јава скрипта. ЈавaСцрипт који показује системско време почетка процеса у пасусу под називом "датетиме1" је приказан на слици 24.

```
<h2>upit 1</h2>
<p id="datetime1"></p>
<p id="datetime2"></p>

<script>
    var now1=new Date();
    var dt1=now1.toISOString();
    document.getElementByID("datetime1").innerHTML="vreme pocetka: "+dt1;
</script>
```

Слика 24. ЈавaСцрипт који приказује системско време започетог процеса

ЈавaСцрипт који приказује системско време завршетка креирања извештаја у пасусу под називом "датетиме2" је приказан на слици 25.

```
1  function F2()
2  {
3      var now2=new Date();
4      var dt2=now2.toISOString();
5      document.getElementById("datetime2").innerHTML="vreme zavrsetka: "+dt2;
6  }
7
```

Слика 25. ЈаваСцрипт који приказује системско време завршетка извештаја

6.2. Резултати експеримената

Имајући у виду да Повер БИ Десктоп представља десктоп апликацију која се користи за приказ извештаја, а Пентахо Бусинесс Аналутицс користи wеб серверску компоненту базирану на Јави и инсталирану на нивоу оперативног система за генерисање извештаја у оквиру wеб странице, иницијална претпоставка је да ће Пентаху бити потребно више времена за извођење операција упита и приказа извештаја у односу на Повер БИ десктоп. Непознаница пре извођења експеримената је да ли оперативни системи који се користе на Хост рачунарима, а који су базирани на Униху, као што је МацОС могу да утичу на брже извршење процеса на виртуелној машини са Пентахо решењем у односу на Хост рачунаре са Виндовс оперативним системом и виртуалном машином са ПоверБИ десктоп решењем.

Поред примарног циља истраживања који се огледа у мерењу и поређењу перформанси оба алата за који се претпоставља да ће Повер БИ десктоп бити бржи, секундарни циљ је утврђивање нивоа евентуалног временског кашњења због коришћења wеб сервиса на Пентахо Виртуалној машини.

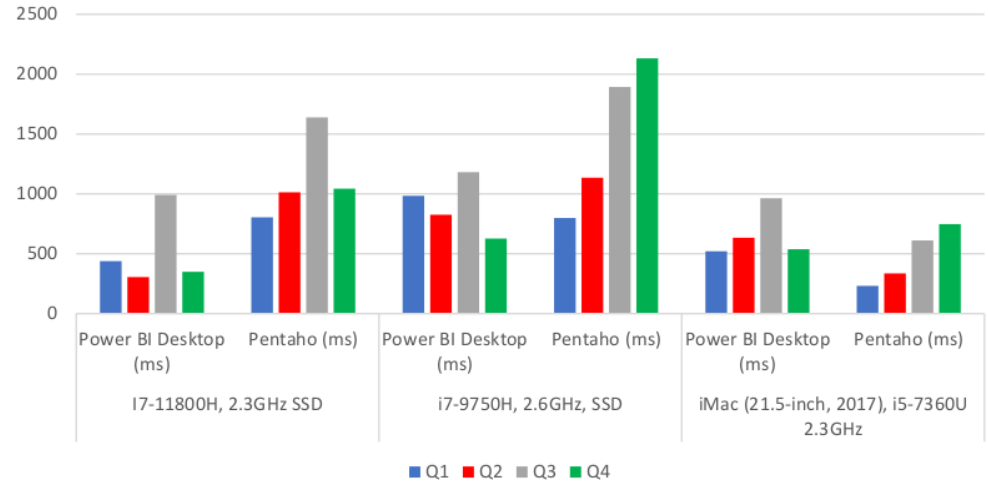
У табели 5 су приказани резултати мерења брзине упита. Табела је форматирана за приказивање више димензија и то:

- Архитектура рачунара,
- Коришћени БИ алат и
- Број/ознака упита (Q1-Q4)

	И7-11800Х, 2.3ГХз ССД		и7-9750Х, 2.6ГХз, ССД		иМац (21.5-инцх, 2017), и5-7360У 2.3ГХз	
	Повер БИ Десктоп (мс)	Пентахо (мс)	Повер БИ Десктоп (мс)	Пентахо (мс)	Повер БИ Десктоп (мс)	Пентахо (мс)
Q1	438	804.33	984	798.33	521	231.33
Q2	304.67	1012.67	825.67	1134.67	634	335.67
Q3	991.33	1639	1181.33	1892.67	963	610.67
Q4	348.67	1044	626.33	2131.33	537.33	747.33

Табела 5. Резултати експеримената БИ алата приказани у милисекундама

На графикону 5 су представљени резултати експеримената БИ алата.



Графикон 5. Резултати експеримената БИ алата приказани у милисекундама

На основу добијених резултата се може видети да су на И7-11800Х хардверској конфигурацији сва четири упита извршена на ПоwерБИ десктопу била бржа у односу на Пентахо за 1.65 пута (код упита Q3) до 3.3 пута (код упита Q2). Предност у брзини извршења упита користећи ПоwерБИ десктоп на хардверској конфигурацији и7-9750Х је била у сличном рангу у односу на Пентахо и кретала се у односу 1.37 (код упита Q2) до 3.40 (код упита Q4), са изузетком брзине извршења упита Q1 где је Пентахо био бржи у односу на ПоwерБИ десктоп за 1.23 пута. Обе претходно наведене хардверске платформе су користиле Виндовс 10 оперативни систем на хост рачунару.

Резултати на МацОС хардверској платформи су показали брже извршење упита Пентахо решења у односу на ПоwерБИ Десктоп за чак 2.5 пута код упита Q1, 1.88 пута код упита Q2 и 1.58 пута код упита Q3, док је упит Q4 био бржи код ПоwерБи десктоп решења за 1.39 пута.

6.3. Закључак

Посебно треба имати на уму да је база података постављена на Мицрософт платформи оперативног система (Виндовс Сервер 2016) и Мицрософт платформи базе података (СQЛ Сервер 2017), као и да су све виртуалне машине са БИ алатима инсталиране на Мицрософт оперативном систему, Виндовс 10. На основу резултата експеримената можемо извести неколико закључака:

1. На најснажнијој хардверској конфигурацији на којој је хост рачунар Виндовс 10, ПоверБИ Десктоп је показао значајно боље резултате од Пентахо алата.
2. На слабијој хардверској конфигурацији на којој је хост рачунар Виндовс 10, ПоверБИ Десктоп је показао боље резултате код комплекснијих упита док је код најмање комплексног упита Пентахо имао бољи резултат.
3. На МацОС хардверској платформи је Пентахо генерално показао бољи резултат у односу на ПоверБИ Десктоп. Овакав резултат се може објаснити бољом компатибилности МацОС-а и Пентахо Веб сервера, који у позадини користе Уних систем. Пентахо користи веб сервер заснован на Јави, која се брже извршава на Уних-базираном систему.

Посматрајући све резултате користећи оба БИ алата, највеће одступање резултата је мање од 1 секунде. Самим тим се може закључити да та разлика у брзини, чак и код најкомплекснијих упита не представља пресудни утицај на брзину доношења пословних одлука. Већи утицај на избор одговарајућег БИ решења може бити цена сваког од решења.

Мицрософт Повер БИ Десктоп се може бесплатно преузети и користити на рачунару под Виндовс оперативним системом. Уколико је потребно да се извештаји креирани од стране пословних аналитичара деле са другим корисницима, потребно је купити претплатничку лиценцу за Повер БИ сервис. У зависности од потребних функционалности, у понуди су Повер БИ Про претплата која износи 10 УСД по кориснику месечно и Повер БИ Премиум која износи 20 УСД по кориснику месечно.

Пентахо платформа долази у две едиције, бесплатна Пентахо Цоммуниту едиција и Пентахо Ентерприсе едиција базирана на претплати.

Имајући у виду висок ниво интеграције тестираних алата са бројним СУБП и да је разлика у перформансама тестираних алата на нивоу мањем од 1 секунде, препорука за пословна предузећа је да користе БИ решења која омогућују дистрибуцију извештаја за пословне кориснике и која су бесплатна.

7. Закључак

У циљу анализе перформанси различитих система за управљање базама података, различитих алата за извођење ЕТЛ процеса и различитих БИ алата уложено је много труда и рада у инсталирање и подешавања различитих комбинација оперативних система, система за управљање базама података, ЕТЛ и БИ алата, у учитавање података у инсталиране базе користећи различите СQЛ дијалекте и алате за учитавање података, као и извођење самих експеримената.

Инсталирање једне комбинације оперативног система и одговарајућих СУБП, ЕТЛ и БИ алата на неколико рачунара ради извођења експеримената, брисање постојећих и инсталирање других оперативних система и других СУБП, ЕТЛ и БИ алата на тим истим рачунарима ради извођења другог експеримента, па понављање поступка ради извођења трећег, четвртог и наредних експеримента изгледало је као мукотрпан и недостижан процес. Зато је одлучено да се све ове инсталације изведу на виртуалним машинама, како би се исте могле копирати на различите рачунаре у циљу мерења перформанси. Имајући у виду да су коришћена и старија и новија решења која раде на различитим оперативним системима (Wиндowc десктоп, Wиндowc Сервер, Линух Центос 6, Центос 7 итд.) било је потребно креирати већи број виртуалних машина. Исто тако, у зависности од инсталираног алата, на виртуалним оперативним системима се морало водити рачуна о инсталираним верзијама драјвера за базе података, као и инсталираним клијентским софтверима за приступ базама података.

Након инсталирања појединачног СУБП на ОС виртуалне машине (МС СQЛ на Wиндowc Серверу, ОРАЦЛЕ на Центосу, МуСQЛ на Центосу, МариаДБ на Центосу, ХП Вертица на Центосу итд.) било је потребно учитати податке из ОЛТП и ОЛАП табела у одговарајуће базе података користећи одговарајуће дијалекте СQЛ-а. У ОЛАП окружењу учитано је око 10-ак димензионих табела величине од 100-тинак до 60,000 редова, као и четири *фацт* табеле величине од 300,000 и 5,000,000 редова. Повезивање различитих машина је остварено уз помоћ ВМwаре Нетwорк Едитор-а, где су морале да се усагласе ИП адресе виртуалних машина у одговарајућим НАТ конфигурацијама. Наведене виртуалне машине су копиране на неколико рачунара различитих конфигурација, како би се измерио и утицај самог хардвера на резултате експеримената.

За поједине СУБП брзина извршења упита иницијално је била спора из разлога што су конфигурациони параметри имали подразумеване вредности након инсталације. Након што су сви конфигурациони фајлови код СУБП (који су дозвољавали да Администратор може да промени наведене параметре) подешени на оптималне вредности, а на основу неколико итерација тестирања, перформансе су се значајно побољшале. Исто тако су за

појединe СУПБ тестирани различити начини смештања и читања података (*pow-store, цолумн-сторе, ин-мемору датабасе* итд.)

Перформансе аналитичких база података тестиране су за он-лине режим, који подразумева да агрегациони подаци нису унапред припремани преко ОЛАП коцки, већ се агрегација извршавала директно по постављању упита. Очекивано је доказана претпоставка да *цолумн-сторе* платформе по перформансама надмашују стандардне базе података које уписују податке у складиште по редовима. Резултати су добијени извршењем упита над СУБП у виртуалним окружењима са по 4ГБ и 8ГБ додељене меморије.

За тестирање перформанси различитих база инсталираних на различитим оперативним системима коришћена су четири комплексна СQЛ агрегациона упита (Q1-Q4). Највећа брзина извршења експеримената постигнута је на најбржем рачунару коришћеном у експерименту, на комерцијалним СУБП: ХП Вертица и МС СQЛ Сервер код којих су табеле у бази података индексиране у *Цолумн-сторе* формату. Некомерцијалне базе података МариаДБ ЦолумнСторе и МуСQЛ ДБ са ИнфоБригхт механизмом за складиштење која користи *цолумн-сторе* индексирање су показале прилично добре перформансе. Орацле 12ц и МариаДБ ЦС су исто тако показали солидне резултате. Генерално посматрано, сви упити над *цолумн-сторе* табелама које садрже 5,000,000 редова су извршавани на свим рачунарима за мање од 10 секунди, док су у експериментима где је база података користила индексе *pow сторе* у МС СQЛ и МуСQЛ, та времена била већа. На основу измерених резултата, препорука је да се *цолумн-сторе* платформе база података користе за аналитичке потребе. Имајући у виду висок ниво интеграције са осталим системима као и одличне перформансе бесплатне едиције ХП Вертица (ХП Вертица Цоммуниту Едитион), пословне организације би требало да размотре њено коришћење за потребе извештавања. Уколико ограничење бесплатне едиције не утичу негативно на перформансе решења, овај СУБП може да представља одличан избор за мала и средња предузећа.

Код поређења перформанси ЕТЛ алата добијени резултати су показали да се најбоље перформансе добијају када се користи иста платформа за базу података, ЕТЛ алат и оперативни систем, јер се тада остварује највећа компатибилност интерфејса. Интерфејси су међусобно оптимизовани. Тако, на Мајкрософт платформи (комбинација Виндовс Сервер ОС – МС СQЛ Сервер – ССИС) су постигнути најбољи резултати. За комплетну трансформацију података из ОЛТП у ОЛАП базу горе наведених табела, ЕТЛ процес је извршен у опсегу од око једног минута на најбржем рачунару до око три минута на најспоријем рачунару. На основу измерених резултата, препорука је да се за најбоље перформансе изабере ЕТЛ алат компатибилан са базом података и оперативним системом. Имајући у виду могућност коришћења комерцијалних и бесплатних ЕТЛ

8. Литература

[1] Јиа-Ланг Сенг, С. Бинг Yао, Алан Р. Хевнер “Рeqуиремeнтс-дривен датабасе сустемс бенцхмарк метход “, Децисион Суппорт Сустемс, 38, п 629-648, 2005

[2] С. БИНГ YАО, АЛАН Р. ХЕВНЕР, ХЕЛЕНЕ YОУНГ-МУЕРС “Аналусис оф Датабасе Систем Арцхитецтурес Усинг Бенцхмаркс “, ИЕЕЕ ТРАНСАЦТИОНС ОН СОФТWАРЕ ЕНГИНЕЕРИНГ, ВОЛ. СЕ-13, НО. 6, пп 709-725, 1987

[3] Пхилип А. Бернстеин, Ериц Hewцомер “Принциплес оф Трансацтион Процессинг “, Морган Кауфманн Публисхерс, 2009,

[4] Едwард Д. Лазowска “Qуантитативе Сустем Перформанце, Цомпутер Сустем Аналусис Усинг Qуеуинг Нетwорк Моделс “, Прентице Халл, Инц. 1984

[5] Расха Осман, Wиллиам Ј. Кнottенбелт “Датабасе сустем перформанце евалуатион моделс: А сурвеу “, Перформанце Евалуатион 69, пп 471–493, 2012

[6] Yоуссеф Бассил “А Цомпаративе Студу он тхе Перформанце оф тхе Топ ДБМС Сустемс “, Јоурнал оф Цомпутер Сциенце & Ресеарцх (ЈЦСЦР) - ИССН 2227-328Х, Вол. 1, Но. 1, Пагес. 20-31, 2012

[7] Удаи Сханкер, Маној Мисра, Анил К. Сарје “Дистрибутед реал тиме датабасе сустемс: бацкгроунд анд литературе ревиеw “, Дистриб Параллел Датабасес 23, пп 127–149, 2008

[8] Амлањуоти Саикиа, Схерин Јоу, Дхондуп Долма, Роселине Мару “Цомпаративе Перформанце Аналусис оф МуСQЛ анд СQЛ Сервер Релатионал Датабасе Манагемент Сустемс ин Wиндоwс Енвиронмент “, Интернатионал Јоурнал оф Адванцед Ресеарцх ин Цомпутер анд Цоммуницатион Енгинееринг Вол. 4, Иссуе 3, пп 160-164, Марцх 2015

[9] Кхawар Ислам, Камран Ахсан, С.А.К. Бари, Мухаммад Саеед, Суед Асим Али “Хуге анд Реал-Тиме Датабасе Сустемс: А Цомпаративе Студу анд Ревиеw фор СQЛ Сервер 2016, Орацле 12ц & МуСQЛ 5.7 фор Персонал Цомпутер “, Јоурнал оф Басиц & Апплиед Сциенцес, 3, пп 481-490, 2017

[10] Цătălин Тудосе, Цармен Одубăştеану “Датабасе Манагемент Сустемс Цомпаративе Студу: Перформанцес оф Мицрософт СQЛ Сервер Версус Орацле”, 3рд Романиан-Хунгариан Јоинт Сумпосиум он Апплиед Цомпутатионал Интеллигенце (САЦИ), Маy 25-26, 2006, Тимисоара, Романиа

[11] Христо Кууркцхиев, Калинка Калоуанова “Перформанце Студу оф Аналутицал Qуериес оф Орацле анд Вертица “, Цонференце: Информатион Сустемс & Грид Тецхнологиес Софиа, ДОИ:10.13140/2.1.3667.0726, 2013

[12] Ротснарани Сетху, Сантосх Кумар Дасх, Мрутуњауа Панда “Перформанце Цомпарисон Бетween Апацхе Хиве анд Орацле СQЛ фор Биг Дата Аналутицс “, Процеедингс оф тхе Еигхтх Интернатионал Цонференце он Софт Цомпутинг анд Паттерн Рецогнитион (СоЦПаР 2016) пп 130–141, Веллоре, Индиа, Децембер 19–21, 2016

[13] Судхансху Кулсхрестха, Схеллу Сацхдева “Перформанце Цомпарисон фор Дата Стораге - Дб4о анд МуСQЛ Датабасес”, ИЕЕЕ, 2014, ДОИ: 10.1109/ИЦ3.2014.6897167

[14] Вассилис Килинтзис, Николаос Бередимас, Иоанна Цхоуварда “Евалуатион оф тхе перформанце оф опен-соурце РДБМС анд триплесторес фор сторинг медицал дата овер а wеб сервице”, ИЕЕЕ, 2014, ДОИ: 10.1109/ЕМБЦ.2014.6944623

[15] Амит Кумар Дwиведи, Ц.С. Ламба, Схweта Схукла “Перформанце Аналусис оф Цолумн Ориентед Датабасе версус Pow Ориентед Датабасе”, Интернатионал Јоурнал оф Цомпутер Апплицатионс (0975 – 8887) Волуме 50 – Но.14, Јулу 2012, пп 31- 34

[16] Пунам Бајај Симрањит Каур Дхиндса “А Цомпаративе Студу оф Датабасе Сустемс”, Интернатионал Јоурнал оф Енгинееринг анд Инновативе Тецхнологу Волуме 1, Иссуе 6, Јуне 2012, пп 267-269

[17] Даниел Ј. Абади, Самуел Р. Мадден, Набил Хацхем “ЦолумнСторес вс. PowСторес: Хow Дифферент Аре Тхеу Реаллу?”, СИГМОД '08: Процеедингс оф тхе 2008 АЦМ СИГМОД интернатионал цонференце он Манагемент оф дата, Јуне 2008, Ванцоувер, БЦ, Цанада, Пагес 967–980

[18] Даниел Ј. Абади “Цолумн-Сторес фор Wиде анд Спарсе Дата”, Цонференце: ЦИДР 2007, Тхирд Биенниал Цонференце он Инновативе Дата Сустемс Ресеарцх, Асиломар, ЦА, УСА, 2007

[19] Анастассиа Аиламаки, Давид Ј. ДеWитт, Марк Д. Хилл, Мариос Скоунакис “Wеавинг Релатионс фор Цацхе Перформанце”, ВЛДБ '01: Процеедингс оф тхе 27тх Интернатионал Цонференце он Веру Ларге Дата Басес, Септембер 2001, Рома, Италу, Пагес 169–180

[20] Синиша Илић, Драгиша Миљковић, Слободан Обрадовић, Алемпије Вељовић, Бојан Јовановић “ЦОМПАРИСОН ОФ ПЕРФОРМАНЦЕ ИН ДАТА АНАЛУСИС ИН ДЕДИЦАТЕД АНД ТРАДИТИОНАЛ ДБМС”, ИНТЕРНАТИОНАЛ СЦИЕНТИФИЦ ЦОНФЕРЕНЦЕ, 18 – 19 Новембер 2016, ГАБРОВО, Булгариа, пп ИИ-219 - 225

[21] Тилманн Рабл, Меикел Поесс “Параллел Дата Генератион фор Перформанце Аналусис оф Ларге, Цомплех РДБМС”, Цонференце: Процеедингс оф тхе Фоуртх Интернатионал Wорксхоп он Тестинг Датабасе Сустемс, ДБТест 2011, Атхенс, Греece, Јуне 13, 2011, ДОИ:10.1145/1988842.1988847

[22] Cwati Jain, Parag Nath Barwal “Performance Analysis of Optimization Techniques for CQL Multi Query Expressions Over Text Database in RDBMS”, International Journal of Information & Computation Technology. ISSN 0974-2239 Volume 4, Number 8 (2014), pp. 841-852

[23] G. Lawrence Sanders, Seungkyoon Shin “Denormalization Effects on Performance of RDBMS”, IEEE, 2001, DOI: 10.1109/HICSS.2001.926306

[24] Aayub M. B., Ali N. “Performance comparison of in-memory and disk-based database using transaction processing performance council (TPC) benchmarking”, Academic Journals, Vol. 8(1), pp. 1-8, August 2018, DOI 10.5897/JIIS2018.0106

[25] Sasalak Tongkaw, Aumnat Tongkaw “A Comparison of Database Performance of MariaDB and MyCQL with OLTP Workload”, IEEE, 2016, DOI: 10.1109/ICOS.2016.7881999

[26] Jing Han, Haihong E, Guan Le, Jian Du “Survey on NoCQL Database”, IEEE, 2011, DOI: 10.1109/ICPCA.2011.6106531

[27] Miral Borad, Tejas Chauhan, Harsh Mehta “A Review on NoCQL Database”, International Journal of Science and Research (IJSR), Paper ID: 6(11), ART20177781, pp 26-28, 2015

[28] Md. Razu Ahmed, Mst. Arifa Khatun, Md. Asraf Ali, Kennetx Sundaraj “A literature review on NoCQL database for big data processing”, International Journal of Engineering & Technology, 7 (2), pp 902-906, 2018

[29] Bogdan George Tudorica, Christian Bucur “A comparison between several NoCQL databases with comments and notes”, IEEE, 2011, DOI: 10.1109/RoEduNet.2011.5993686

[30] Khalid Mahmood “Performance Comparison of NOSQL Database Cassandra and SQL Server for Large Database”, Journal of Independent Studies and Research – Computing Volume 14 Issue 2, pp 21-25, 2016

[31] Cornelia Győrödi, Robert Győrödi, Alexandra Ştefan, Livia Bandici, “A Comparative Study of Database with Different Methods of Internal Data Management”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 4, pp 266-271, 2016

[32] Azhi Faraj, Bilal Rashid, Twana Shareef “COMPARATIVE STUDY OF RELATIONAL AND NONRELATIONS DATABASE PERFORMANCES USING ORACLE AND MONGODB SYSTEMS”, INTERNATIONAL JOURNAL OF

ЦОМПУТЕР ЕНГИНЕЕРИНГ & ТЕЦХНОЛОГҮ (ИЈЦЕТ), Волуме 5, Иссуе 11, пп. 11-22, 2014

[33] Алехандру Боицеа, Флорин Радулесцу, Лаура Иоана Агапин “МонгоДБ вс Орацле - датабасе цомпарисон”, 2012 Тхирд Интернатионал Цонференце он Емергинг Интеллигент Дата анд Wеб Тецхнологиес, ИЕЕЕ, 2012, ДОИ 10.1109/ЕИДWТ.2012.32

[34] Хоуцине Маталлах, Гхалем Белалем, Алгериа Карим Боуамране “Цомпаративе Студу Бетwеен тхе МуCQЛ Релатионал Датабасе анд тхе МонгоДБ НоCQЛ Датабасе”, Интернатионал Јоурнал оф Софтwаре Сциенце анд Цомпутатионал Интеллигенце, 2021, ДОИ: 10.4018/ИЈССЦИ.2021070104

[35] Дипина Дамодаран Б, Схирин Салим, Сурекха Мариум Варгесе “МОНГОДБ ВС МҮCQЛ: А ЦОМПАРАТИВЕ СТУДҮ ОФ ПЕРФОРМАНЦЕ ИН СУПЕР МАРКЕТ МАНАГЕМЕНТ СҮСТЕМ”, Интернатионал Јоурнал оф Цомпутатионал Сциенце анд Информатион Тецхнологу (ИЈЦСИТҮ) Вол.4, Но.2, пп 31- 38, 2016

[36] Санобар Кхан, Проф.Ванита Мане “CQЛ Суппорт овер МонгоДБ усинг Метадата”, Интернатионал Јоурнал оф Сциентифиц анд Ресеарцх Публицатионс, Волуме 3, Иссуе 10, 2013, ИССН 2250-3153

[37] Самеера Аббас ФАДХЕЛ, Енас Али ЈАМЕЕЛ “А ЦОМПАРИСОН БЕТWЕЕН НОCQЛ АНД РДБМС: СТОРАГЕ АНД РЕТРИЕВАЛ”, Интернатионал Јоурнал оф Апплиед Сциенцес анд Тецхнологу, Волуме 4, Иссуе 3, пп 172- 184, 2022

[38] Мухаммад Зохаиб Кхан, Фахим Уз Заман, Мухаммад Аднан, Аисха Имроз, Махира Абдул Рауф, Зухаиб Пхул “Цомпаративе Цасе Студу: Ан Евалуатион оф Перформанце Цомпутатион Бетwеен CQЛ Анд НоCQЛ Датабасе”, Синдх Јоурнал оф Хеадwаус ин Софтwаре Енгинееринг, Вол 1 Но 2 пп 14-23, 2023

[39] Рацхит Пандеу “Перформанце Бенцхмаркинг анд Цомпарисон оф Цлоуд-Басед Датабасес МонгоДБ (НоCQЛ) Вс МуCQЛ (Релатионал) усинг ҮЦСБ”, 2020, ДОИ: 10.13140/РГ.2.2.10789.32484

[40] Мит Јаин, Асхисх Кханцхандани, Цајетан Родригуес “Перформанце Цомпарисон оф Грапх Датабасе анд Релатионал Датабасе”, Маy 2023, ДОИ: 10.13140/РГ.2.2.27380.32641

[41] Цорнелиа А. Гуорӧди, Диана В. Думсе-Буресцу, Доина Р. Змаранда, Роберт С. Гуорӧди , Гианина А. Габор, Георге Д. Пецхерле “Перформанце Аналусис оф НоCQЛ анд Релатионал Датабасес wитх ЦоуцхДБ анд МуCQЛ фор Апплицатион’с Дата Стораге”, Апплиед Сциенцес Волуме 10 Иссуе 23, 2020, дои:10.3390/апп10238524

[42] Ywe Хохенстеин анд Мартин Јерглер “Датабасе Перформанце Цомпарисонс: Ан Инспецтион оф Фаирнесс”, Процеедингс оф тхе 8тх Интернатионал Цонференце он Дата Сциенце, Тецхнологу анд Апплицатионс ДАТА - Волуме 1, 243-250, 2019, ДОИ: 10.5220/0007926602430250

[43] Луцас Ферреира да Силва, João B. Ф. Лима “Ан евалуатион оф релатионал анд НоСQЛ дистрибутед датабасес он а лов-повер цлустер ”, Тхе Јоурнал оф Суперцомпутинг 79, пп 13402–13420, 2023

[44] Yисхан Ли анд Сатхиамоортху Манохаран “А перформанце цомпарисон оф CQЛ анд НоСQЛ датабасес”, ИЕЕЕ, 2013, ДОИ: 10.1109/ПАЦРИМ.2013.6625441

[45] Моника Схарма, Висхал Деер Схарма, Махесх М. Бунделе “Перформанце Аналусис оф РДБМС анд Но CQЛДатабасес: ПостгреСQЛ, МонгоДБ анд Нео4ј”, ИЕЕЕ, 2018, ДОИ: 10.1109/ИЦРАИЕ.2018.8710439

[46] А Б М Монируззаман “НewCQЛ: Товардс Нехт-Генератион Сцалабле РДБМС фор Онлине Трансацтион Процессинг (ОЛТП) фор Биг Дата Манагемент”, Интернатионал Јоурнал оф Датабасе Тхеору анд Апплицатион Вол.7, Но.6, пп.121-130, 2014, ДОИ:10.14257/ијдта.2014.7.6.11

[47] Ј. Бруан Осборне “Перформанце Импацт оф Тупе-И Виртуализатион он а НewCQЛ Релатионал Датабасе Манагемент Систем”, (ИЈАЦСА) Интернатионал Јоурнал оф Адванцед Цомпутер Сциенце анд Апплицатионс, Вол. 13, Но. 1, 2022, ДОИ: 10.14569/ИЈАЦСА.2022.0130101

[48] Зхен Хуа Лиу, Беда Хаммерсцхмидт, Доуг МцМахон, Yинг Лу, Хуи Јое Цханг “Цлосинг тхе фунцтионал анд Перформанце Гап бетween CQЛ анд НоCQЛ”, СИГМОД '16: Процеедингс оф тхе 2016 Интернатионал Цонференце он Манагемент оф Дата, Пагес 227–238, 2016, ДОИ:10.1145/2882903.2903731

[49] Рафаел Алмеида, Јорге Бернардино “Тестинг CQЛ анд НоCQЛ аппроацхес фор биг дата wарехоусе сустемс”, Инт. Ј. Бусинесс Процесс Интегратион анд Манагемент, Вол. 7, Но. 4, пп 322-334, 2015, ДОИ:10.1504/ИЈБПИМ.2015.073656

[50] Доуглас Кунда, Хазаел Пхири “А Цомпаративе Студу оф НоCQЛ анд Релатионал Датабасе”, Замбиа ИЦТ Јоурнал 1(1): пп 1-4, 2017, хттпс://дои.орг/10.33260/зицтјоурнал.в1и1.8

[51] Twана Хуссеин Схарееф, Карзан Хуссеин Схарееф, Билал Најмаддин Расхид “А Сурвеу оф Цомпаринг Дифферент Цлоуд Датабасе Перформанце: CQЛ анд НоCQЛ”, Пассер Јоурнал, Пассер 4, пп 45- 57, 2022, ДОИ:10.24271/псп.2022.301247.1104

[52] Тони Таипалус, “Датабасе манагемент систем перформанце цомпарисонс: А систематиц литературе ревиеw”, Оцтобер 2023, Јоурнал оф Системс анд Софтwаре 208(2):111872, ДОИ:10.1016/j.јсс.2023.111872

[53] Wисал Кхан, Теератх Кумар, Цхенг Зханг, Кислау Рај, Арунабха М. Роу, Бин Луо “СQЛ анд НоСQЛ Датабасе Софтwаре Арцхитецтуре Перформанце Аналусис анд Ассессментс—А Систематиц Литературе Ревиеw”, Биг Дата Цогн. Цомпут. 2023, 7(2), 97; хттпс://дои.орг/10.3390/бдцц7020097

[54] Абдуллах Талха Кабакус, Ресул Кара “А перформанце евалуатион оф ин-мемору датабасес”, Јоурнал оф Кинг Сауд Университу – Цомпутер анд Информатион Сциенцес (2017) 29, пп 520–525, ДОИ:10.1016/j.јксуци.2016.06.007

[55] Робин Хецхт, Стефан Јаблонски “НоСQЛ евалуатион: А усе цасе ориентед сурвеу”, 2011 Интернатионал Цонференце он Цлоуд анд Сервице Цомпутинг, ДОИ: 10.1109/ЦСЦ.2011.6138544

[56] Цристина БĂЗĂР, Цосмин Себастиан ИОСИФ “Тхе Транситион фром РДБМС то НоСQЛ. А Цомпаративе Аналусис оф Тхрее Популар Нон-Релатионал Солутионс: Цассандра, МонгоДБ анд Цоуцхбасе”, Датабасе Системс Јоурнал, Ацадему оф Ецономиц Студиес - Буцхарест, Романиа, вол. 5(2), пагес 49-59, Септембер.

[57] ЕнqингТанг, Yусхун Фан “Перформанце Цомпарисон бетwеен Фиве НоСQЛ Датабасес”, ИЕЕЕ, 7тх Интернатионал Цонференце он Цлоуд Цомпутинг анд Биг Дата (ЦЦБД), 2016, ДОИ: 10.1109/ЦЦБД.2016.030

[58] Диого Аугусто Переира, Wагнер Оуриqуе де Мораис & Едисон Пигнатон де Фреитас “НоСQЛ реал-тиме датабасе перформанце цомпарисон”, Интернатионал Јоурнал оф Параллел, Емергент анд Дистрибутед Системс, 2017, ДОИ: 10.1080/17445760.2017.1307367

[59] Мрс. Рупали М. Цхопаде, Мр. Никхил С. Дхавасе “МонгоДБ, ЦоуцхБасе: Перформанце Цомпарисон фор Имаге Датасет”, ИЕЕ, 2нд Интернатионал Цонференце фор Цонвергенце ин Тецхнологу (И2ЦТ), 2017, пп 255-258, ДОИ: 10.1109/И2ЦТ.2017.8226131

[60] Герард Хаугхиан, Расха Осман, Wиллиам Ј. Кнottенбелт, Бенцхмаркинг Реплицатион ин Цассандра анд МонгоДБ НоСQЛ Датасторес, Интернатионал Цонференце он Датабасе анд Ехперт Системс Апплицатионс, 2016, ДОИ:10.1007/978-3-319-44406-2_12

[61] А. ГОРБЕНКО, А. КАРПЕНКО, О. ТАРАСҮУК “ПЕРФОРМАНЦЕ ЕВАЛУАТИОН ОФ ВАРИОУС ДЕПЛОҮМЕНТ СЦЕНАРИОС ОФ ТХЕ 3-РЕПЛИЦАТЕД ЦАССАНДРА

НОСQЛ ЦЛУСТЕР ОН АWС", Радиоелецтрониц анд Цомпутер Сустемс (4). пп. 157-165. ИССН 2663-2012 ДОИ: хттпс://дои.орг/10.32620/рекс.2021.4.13

[62] Анатолиу Горбенко, Алехандер Романовску, Олга Тарасуук "Интерплауинг Цассандра НоСQЛ Цонсистенцу анд Перформанце: а Бенцхмаркинг Аппроацх", Депендабле Цомпутинг - ЕДЦЦ 2020 Worксхопс (пп.168-184), 2020, ДОИ:10.1007/978-3-030-58462-7_14

[63] А. ГОРБЕНКО, О. ТАРАСҮУК "ЕХПЛОРИНГ ТИМЕОУТ АС А ПЕРФОРМАНЦЕ АНД АВАИЛАБИЛИТҮ ФАЦТОР ОФ ДИСТРИБУТЕД РЕПЛИЦАТЕД ДАТАБАСЕ СҮСТЕМС", РАДИОЕЛЕЦТРОНИЦ АНД ЦОМПУТЕР СҮСТЕМС (4). пп. 98-105, 2020, ИССН 1814-4225 ДОИ: хттпс://дои.орг/10.32620/рекс.2020.4.09

[64] Мохаммад Абу Каусар, Мохаммад Насар, Ароцкиасаму Соосаиманицкам "А Студу оф Перформанце анд Цомпарисон оф НоСQЛ Датабасес: МонгоДБ, Цассандра, анд Редис Усинг ҮЦСБ", ИНДИАН ЈОУРНАЛ ОФ СЦИЕНЦЕ АНД ТЕЦХНОЛОГҮ, 2020, пп 1532-1540, ДОИ: 10.17485/ИЈСТ/в15и31.1352

[65] Али Ал-Аллаwее, Пасцал Лоренз, Абделхафид Абоуаисса, Мослех Абуалхај "А Перформанце Евалуатион оф Ин-Мемору Датабасес Операционс ин Сессион Инитиатион Протоцол", Нетwорк 2023, 3(1), 1-14; хттпс://дои.орг/10.3390/нетwорк3010001

[66] Францини Хак, Тиаго Гуимарãес, Антóнио Абелха, Мануел Сантос "Ан Ехплоратору Студу оф а НоСQЛ Датабасе фор а Цлиницал Дата Репоситору", WорлдЦИСТ 2020: Трендс анд Инноватионс ин Информатион Сустемс анд Тецхнологиес пп 476–483, ДОИ:10.1007/978-3-030-45697-9_46

[67] Прасхантх Јаккула "ХБасе ор Цассандра? А Цомпаративе студу оф НоСQЛ Датабасе Перформанце", Интернатионал Јоурнал оф Сциентифиц анд Ресеарцх Публицатионс, Волуме 10, Иссуе 3, Марцх 2020, пп 808-820, ДОИ:10.29322/ИЈСРП.10.03.2020.п9999

[68] Цорнелиа А. Гуорöди, Диана В. Дум¸се-Буресцу, Доина Р. Змаранда анд Роберт¸ С. Гуорöди "А Цомпаративе Студу оф МонгоДБ анд Доцумент-Басед МуСQЛ фор Биг Дата Апплицатион Дата Манагемент", Биг Дата Цогн. Цомпут. 2022, 6(2), 49; хттпс://дои.орг/10.3390/бдцц6020049

[69] Рупали Каур, Јаспреет Каур Сахиwал "А ревиеw оф цомпарисон бетweен НоСQЛ Датабасес: МонгоДБ анд ЦоуцхДБ", Интернатионал Јоурнал оф Рецент Тецхнологу анд Енгинееринг (ИЈРТЕ), ИССН: 2277-3878, Волуме-7, Иссуе-6С, Марцх 2019, пп 892-898

[70] Хоуцине Маталлах, Гхалем Белалем "Евалуатион оф НоСQЛ Датабасес: МонгоДБ, Цассандра, ХБасе, Редис, Цоуцхбасе, ОриентДБ", Интернатионал Јоурнал оф Софтwаре

Сциенце анд Цомпутатионал Интеллигенце, Волуме 12, Иссуе 4, 2020, пп 71- 91, ДОИ:10.4018/ИЈССЦИ.2020100105

[71] Сацхин Цхаудхару, Девендра Прасад Мурала, В. К. Сривастав “А Цритицал Ревиew оф Дата Wарехоусе”, Глобал Јоурнал оф Бусинесс Манагемент анд Информатион Тецхнологу, Волуме 1, Нумбер 2 (2011), пп. 95-103, хттпс://www.рипублицатион.цом/Волуме/гјбмит1н2.хтм

[72] Мр. Дисхек Манкад, Мр. Преуасх Дхолакиа “Тхе Студу он Дата Wарехоусе Десигн анд Усаге”, Интернатионал Јоурнал оф Сциентифиц анд Ресеарцх Публицатионс, Волуме 3, Иссуе 3, Марцх 2013, ИССН 2250-3153, хттпс://www.ијсрп.орг/ресеарцх-папер-0313/ијсрп-п1597.пдф

[73] Мутаз М. Ал-Дебеи “Дата Wарехоусе ас а Бацкбоне фор Бусинесс Интеллигенце: Иссуес анд Цхалленгес”, Еуропеан Јоурнал оф Ецономицс, Финанце анд Административе Сциенцес, ИССН 1450-2275 Иссуе 33 (2011), пп153-166

[74] Саеед Роухани, Сара Асгари, Сеуед Вахид Мирхоссеини “Ревиew Студу: Бусинесс Интеллигенце Цонцептс анд Аппроацхес”, Америцан Јоурнал оф Сциентифиц Ресеарцх, ИССН 1450-223Х Иссуе 50 (2012), пп. 62-75

[75] ОУКУ ИСИК, МАРУ Ц. ЈОНЕС, АННА СИДОРОВА “Бусинесс интеллигенце (БИ) суццесс анд тхе роле оф БИ цапабилитиес”, ИНТЕЛЛИГЕНТ СУСТЕМС ИН АЦЦОУНТИНГ, ФИНАНЦЕ АНД МАНАГЕМЕНТ, Интелл. Сус. Ацц. Фин. Мгмт. 18, 161–176 (2011), ДОИ:10.1002/исаф.329

[76] Панос Вассилиадис “А Сурвеу оф Ехтрацт–Трансформ–Лоад Тецхнологу”, Септембер 2011, Интернатионал Јоурнал оф Дата Wарехоусинг анд Мининг 5(3):1-27, ДОИ:10.4018/јдwм.2009070101

[77] Нитин Ананд “Апплицатион оф ЕТЛ Тоолс ин Бусинесс Интеллигенце”, Интернатионал Јоурнал оф Сциентифиц анд Ресеарцх Публицатионс, Волуме 2, Иссуе 11, Новембер 2012, ИССН 2250-3153, хттпс://www.ијсрп.орг/ресеарцх-папер-1112.пхп?рп=П11356

[78] Јосе Баратеиро, Хелена Галхардас “А Сурвеу оф Дата Qуалиту Тоолс”, Датенбанк-Спектрум 14(5):15-21, 2005, хттпс://www.ресеарцхгате.нет/публицатион/220102796_А_Сурвеу_оф_Дата_Qуалиту_Тоолс

[79] Манел Соуибгуиа,б, Фатен Атигуиб, Салоуа Заммалиа, Самира Цхерфиб, Садок Бен Yахиаа “Дата qуалиту ин ЕТЛ процесс: А прелиминару студу”, 23рд Интернатионал Цонференце он Кноwледге-Басед анд Интеллигент Информатион &

ЕнгинеерингСустемс, СциенцеДирецт, Процедиа Цомпутер Сциенце 159 (2019) пп 676–687

[80] Ј П А Рунтуwене, И Р Х Т Тангкаwарow, Ц Т М Маноппо, Р Ј Салаки “А Цомпаративе Аналусис оф Ехтрацт, Трансформатион анд Лоадинг (ЕТЛ) Процесс”, ИОП Цонф. Сериес: Материалс Сциенце анд Енгинеering 306 (2018) 012066 дои:10.1088/1757-899Х/306/1/012066

[81] Абделлах Амине, Рацхид Аит Даоуд, Белаид Боуикхалене “Еффициенцу Цомпараисон анд Евалуатион бетwеен Тwо ЕТЛ Ехтрацтион Тоолс”, Индонесиан Јоурнал оф Елецтрицал Енгинеering анд Цомпутер Сциенце, Вол. 3, Но. 1, Јулу 2016, пп. 174 ~ 181, ДОИ:10.11591/ијеецс.в3.и1.пп174-181

[82] Синиса С. Илиц, Драгиса Миљковиц, Алемпије Вељовиц, Јасмина Новаковиц, Владимир Вељовиц “ЦОМПАРИСОН ОФ ПЕРФОРМАНЦЕС ОФ БУИЛТ-ИН ТООЛС ФОР ЛОАД АНД ЕХПОРТ ОФ ДАТА ТО АНД ФРОМ ПОПУЛАР ДАТАБАСЕС”, ЦОМПУТЕР СЦИЕНЦЕ АНД ТЕЦХНОЛОГИЕС, 2018, Нумбер 1/2018 пп 32-38

[83] Синиша С. Илић, Данијела Милентијевић, Алемпије Вељовић, Слободан Обрадовић, Александар Жорић, Небојша Арсић “ТХЕ ИМПЛЕМЕНТАТИОН ОФ ЕТЛ ПРОЦЕССЕС ОН ЕХАМПЛЕ ОФ АУТХОРИСЕД ДЛС”, ИНТЕРНАТИОНАЛ СЦИЕНТИФИЦ ЦОНФЕРЕНЦЕ “УНИТЕЦХ 2015”, ГАБРОВО, Булгариа, пп ИИ-183 – ИИ-188

[84] Рајендрани Мукхерјее, Прагма Кар “А Цомпаративе Ревиеw Оф Дата Wарехоусинг ЕТЛ Тоолс Witх Неw Трендс Анд Индустру Инсигхт”, 2017 ИЕЕЕ 7тх Интернатионал Адванце Цомпутинг Цонференце, пп 943-948, ДОИ: 10.1109/ИАЦЦ.2017.0192

[85] Аманпартап Сингх ПАЛЛ, Др. Јаитег Сингх КХАИРА “А цомпаративе Ревиеw оф Ехтрацтион, Трансформатион анд Лоадинг Тоолс”, Датабасе Сустемс Јоурнал, Ацадему оф Ецономиц Студиес - Буцхарест, Романиа, вол. 4(2), пагес 42-51, Јулу. хттпс://идеас.репец.орг/а/аес/дбјоур/в4у2013и2п42-51.хтмл

[86] Рахмади Wијаја, С.Си., МТ, Бамбанг Пудјоатмодјо, С.Си., МТ. “Ан Овервиеw анд Имплементатион оф Ехтрацтион-Трансформатион-Лоадинг (ЕТЛ) Процесс Ин Дата Wарехоусе (Цасе Студу: Департмент оф Агрицултуре)”, 2015 3рд Интернатионал Цонференце он Информатион анд Цоммуницатион Тецхнологу (ИЦоИЦТ), 2015, ДОИ:10.1109/ИЦоИЦТ.2015.7231399

[87] Асма Qаисер, Мухаммад Умер Фарооq, Суед Мухаммад Набеел Мустафа, Назиа Абрар, Цомпаративе Аналусис оф ЕТЛ Тоолс ин Биг Дата Аналутицс, Пакистан Јоурнал оф Енгинеering анд Тецхнологу, ПакЈЕТ, ИССН (е): 2664-2050, Волуме: 6, Нумбер: 1, Пагес: 7- 12, Yеар: 2023,

https://www.researchgate.net/publication/369094822_Comparative_Analysis_of_ETL_Tools_in_Big_Data_Analytics

[88] Adnan, Amil Ahmad Ilham, Suahrul Usman "Performance Analysis of Extract, Transform, Load (ETL) in Apache Hadoop atop NAS Storage using ISCSI", IEEE, 2017 4th International Conference on Computer Application and Information Processing Technology (CAIPT), 2017, DOI: 10.1109/CAIPT.2017.8320716

[89] Eckerson, W., White, C.: Evaluating ETL and Data Integration Platforms. THE DATA WAREHOUSING INSTITUTE, A 101Communications Publication, http://download.101com.com/tdwi/research_report/2003ETLReport.pdf

[90] Sumit Misra, Sanjoy Kumar Saha, and Chandan Mazumdar "Performance Comparison of Hadoop Based Tools with Commercial ETL Tools – A Case Study", BDA 2013, LNCS 8302, pp. 176–184, 2013, DOI:10.1007/978-3-319-03689-2_12

[91] Sonali Vyas, Pragya Vaishnav "A comparative study of various ETL process and their testing techniques in data warehouse", Journal of Statistics and Management Systems, 20:4, 753-763, DOI: 10.1080/09720510.2017.1395194, 2017

[92] Seema Maitreya, C.K. Jha "MapReduce: Simplified Data Analysis of Big Data", ScienceDirect, Procedia Computer Science 57 (2015) 563 – 571, doi: 10.1016/j.procs.2015.07.392

[93] Jungkee Jo, Kang-Woo Lee "MapReduce-Based D_ELT Framework to Address the Challenges of Geospatial Big Data", International Journal of Geo-Information, 2019, 8, 475; doi:10.3390/ijgi8110475

[94] Papa Sengkane Diouf, Aliou Bolu, Samba Ndiaye "Performance of the ETL processes in terms of volume and velocity in the cloud: state of the art", IEEE, 2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS), 2017, DOI: 10.1109/ICETAS.2017.8277875

[95] Dhamotharan Seenivasan "Improving the Performance of the ETL Jobs", International Journal of Computer Trends and Technology Volume 71 Issue 3, 27-33, March 2023, DOI:10.14445/22312803/IJCTT-V71I3P105

[96] Vishal Gour, Dr. C.C.Sarangdevot, Govind Singh Tanwar, Anand Sharma "Improve Performance of Extract, Transform and Load (ETL) in Data Warehouse", (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 03, 2010, 786-789,
https://www.researchgate.net/publication/49618608_Improve_Performance_of_Extract_Transform_and_Load_ETL_in_Data_Warehouse

[97] Сyyеон Лее, Сунгуонг Парк “Перформанце Аналусис оф Биг Дата ЕТЛ Процесс овер ЦПУ-ГПУ Хетерогенеоус Арцхитецтурес”, 2021 ИЕЕЕ 37тх Интернатионал Цонференце он Дата Енгинееринг Воркскопс (ИЦДЕW), пп 42- 47, ДОИ: 10.1109/ИЦДЕW53142.2021.00015

[98] Схакер Х. Али Ел-Саппагх а, Абделтаwаб М. Ахмед Хендаwи, Али Хамед Ел Бастаwиссу “А пропосед модел фор дата wарехоусе ЕТЛ процессес”, Јоурнал оф Кинг Сауд Университу – Цомпутер анд Информатион Сциенцес (2011) 23, 91–104, ДОИ:10.1016/ј.јксуци.2011.05.005

[99] Асма Дхаоуади, Кхадија Боусселми, Мохамед Мохсен Гаммоуди, Сébастиен Моннет, Слимане Хаммоуди “ДатаWарехоусинг Процесс Моделинг фром Цлассицал Аппроацхес то Неw Трендс: Маин Феатурес анд Цомпарисонс”, Дата 2022, 7(8), 113; хттпс://дои.орг/10.3390/дата7080113

[100] Дхамотхаран Сеенивасан “ЕТЛ (Ехтрацт, Трансформ, Лоад) Бест Працтицес”, Интернатионал Јоурнал оф Цомпутер Трендс анд Тецхнологу Волуме 71 Иссуе 1, 40-44, Јануару 2023, ДОИ:10.14445/22312803/ИЈЦТТ-В71И1П106

[101] Вишња Истрат, Сања Станисављев, Бранко Маркоски “ТХЕ РОЛЕ ОФ БУСИНЕСС ИНТЕЛЛИГЕНЦЕ ИН ДЕЦИСИОН ПРОЦЕСС МОДЕЛИНГ “, Тхе Еуропеан Јоурнал оф Апплиед Ецономицс, 12(2), 44-52, 2015

[102] Бернхард Wиедер, Мариа-Луисе Оссимитз “Тхе импацт оф Бусинесс Интеллигенце он тхе qуалиту оф децисион макинг – а медиатион модел“, Цонференце он ЕНТЕРприсе Информатион Сустемс / ЦЕНТЕРИС 2015 Оцтобер 7-9, 2015, Процедиа Цомпутер Сциенце 64 (2015) 1163 – 1171

[103] Саеед Роухани Амир Асхрафи Ахад Заре Равасан Самира Афсхари “Тхе импацт модел оф бусинесс интеллигенце он децисион суппорт анд организатионал бенефитс “, Јоурнал оф Ентерприсе Информатион Манагемент, Вол. 29 Исс 1 пп. 19 – 50, хттп://дх.дои.орг/10.1108/ЈЕИМ-12-2014-0126

[104] Јакуб Андар, Петра Каспарова “Импацт оф Манагемент Суппорт он Бусинесс Интеллигенце Адоптион: Иллустративе Цасе Студу Тестинг Дифферент Манагериал Стратегиес“, Ацта Информатица Прагенсиа 2024, Волуме 13, Иссуе 1, пп. х–х, хттпс://дои.орг/10.18267/ј.аип.230

[105] Марием Белгхитха, Ханен Бен Аммарб, Абделкарим Еллоумиа, Wафик Хацхицхац “А неw роллинг форецастинг фрамеwорк усингМицрософт Поwер БИ фор дата висуализатион: Ацасе студу ин а пхармацеутицал индустру “, Анналес Пхармацеутиqуес Францаисес, Волуме 82, Иссуе 3, Маy 2024, Пагес 493-506, хттпс://дои.орг/10.1016/ј.пхарма.2023.10.013

[106] Схенwеи Зханг “Аналусис оф Бусинесс Интеллигенце Тецхнологу ин тхе Биг Дата Ера - А Цасе Студу оф Алибаба Гроуп“, Процеедингс оф тхе 3рд Интернатионал Цонференце он Бусинесс анд Полицу Студиес, ДОИ: 10.54254/2754-1169/74/20241520

[107] Мохаммад Тагхи Садегхи, Ибаа Ал хасан “Енханцинг Супплу Цхаин Манагемент ин Смалл-Медиум Ентерприсес тхроугх Бусинесс Интеллигенце: А Цонцептуал Модел Аппроацх “, Ресеарцх Cqуаре, хттпс://дои.орг/10.21203/рс.3.рс-4249032/в1

[108] Ромеу Сеqуеира, Арсénио Реис, Пауло Алвес анд Фредерицо Бранцо “Роадмап фор Имплементинг Бусинесс Интеллигенце Системс ин Хигхер Едуцатион Институтионс: Систематиц Литературе Ревиеw“, 20тх Интернатионал Цонференце он Смарт Бусинесс Тецхнологиес (ИЦСБТ 2023). Роме, Италу, 11–13 Јулу 2023, Информатион 2024, 15, 208, хттпс://дои.орг/10.3390/инфо15040208

[109] Радхакрисхнан Субраманиам, Прасхобхан Палакеел, Манимутху Арунмозхи, Маникандан Сридхаран, Утхауакумар Маримутху “Фацторс дривинг бусинесс интеллигенце адоптион: ан ехтендед тецхнологу-организатион-енвиронмент фрамеwорк “, Индонесиан Јоурнал оф Елецтрицал Енгинееринг анд Цомпутер Сциенце, Вол. 34, Но. 3, Јуне 2024, пп. 1893~1903, ИССН: 2502-4752, ДОИ: 10.11591/ијеецс.в34.и3.пп1893-1903

[110] Асхраф Бану Мохаммед, Манаф Ал-Окаилу, Дхиа Qасим, Мохаммад Кхалаф Ал-Мајали “Товардс ан ундерстандинг оф бусинесс интеллигенце анд аналутицс усаге: Евиденце фром тхе банкинг индустру “, Интернатионал Јоурнал оф Информатион Манагемент Дата Инсигхтс 4 (2024) 100215, хттпс://дои.орг/10.1016/ј.јјимеи.2024.100215

[111] Адетуми Адеwуми, Есе Еигбадон Осхиосте, Онуека Франца Асузу, Ндубуиси Леонард Ндубуиси, Кехинде Феранми Аwоннуга анд Онуека Хенру Дараојимба “Бусинесс интеллигенце тоолс ин финанце: А ревиеw оф трендс ин тхе УСА анд Африца “, Wорлд Јоурнал оф Адванцед Ресеарцх анд Ревиеwс, 2024, 21(03), 608–616, хттпс://дои.орг/10.30574/wјапп.2024.21.3.0333

[112] Сади Бади “Енханцинг Бусинесс Интеллигенце: Харнессинг Тект Аналутицс, АИ, анд ЕРП фор Дата-Дривен Децисион Макинг “, Марцх 2024, ДОИ:10.13140/РГ.2.2.31479.43681

[113] Асад Аббас “Трансформинг Емплоуее Перформанце Предицтион: Харнессинг Мацхине Леарнинг анд Аналутицс фор Бусинесс Интеллигенце Адванцемент “, Марцх 2024, ДОИ: 10.13140/РГ.2.2.10770.06087

[114] Мд. Асхрафуззаман, Исрат Султана, Рабита Рахман Ера, Арјун Тхиммаиах, Амит Кумар Схарма, Саира Бинте Аwал, Јитендра Схарма “Трансформинг Медицал Тоурисм Тхроугх ан Интегратед Бусинесс Интеллигенце Платформ: Дата-Дривен Инсигхтс анд

Хеалтхцаре Индустру Импацт “, Фебруару 2024, ИГИ Глобал, ДОИ: 10.4018/979-8-3693-1210-0.цх010

[115] Малау саркар, Расел Махмуд Jewел, Мд Салим Цхowдхуру, Мд Ал-Имран, Румана Схахид, Аисхаруја Роу Пуја, Рејон Кумар Рау анд Сандип Кумар Гхосх “Револутионизинг Организатионал Децисион-Макинг фор Стоцк Маркет: А Мацхине Леарнинг Аппроацх witх ЦННс ин Бусинесс Интеллигенце анд Манагемент “, Јоурнал оф Бусинесс анд Манагемент Студиес, ПУБЛИСХЕД: 13 Фебруару 2024, ДОИ: 10.32996/јбмс.2024.6.1.16

[116] Мурат Оздемир, Еууп Емре Улку анд Казим Yилдиз “Аналусис анд Цомпарисон оф Бусинесс Интеллигенце Тоолс Мост Преферред бу Цомпаниес ин Туркеу “, БАЛКАН ЈОУРНАЛ ОФ ЕЛЕЦТРИЦАЛ & ЦОМПУТЕР ЕНГИНЕЕРИНГ, Вол. 11, Но. 2, Април 2023, п.144-155

[117] Царло ДЕЛЛ’АQУИЛА, Францесцо ДИ ТРИА, Езио ЛЕФОНС, анд Филиппо ТАНГОРРА “Бусинесс Интеллигенце Сустемс: а Цомпаративе Аналусис“, WCEAC ТРАНСАЦТИОНС он ИНФОРМАТИОН СЦИЕНЦЕ & АППЛИЦАТИОНС, ИССН: 1790-0832, Иссуе 5, Волуме 5, 2008, п.612-621

[118] Випин Кумар Сараоги, Јитендер Кумар, Др. А.К. Малик “ДАТА АНАЛYТИЦС БУСИНЕСС ИНТЕЛЛИГЕНЦЕ ТООЛС: А ЦОМПАРАТИВЕ СТУДY“, ИЈЕСР/НЦСЕТ-2020/ Цонф. Иссуе/ Артицле Но-1/1-3, ИССН 2277-2685

[119] Фехми Скендер, Виолета Маневска, “Дата Висуализатион Тоолс - Превиеw анд Цомпарисон “, Превиеw анд Цомпарисон. Јоурнал оф Емергинг Цомпутер Тецхнологиес, 2(1), 30-35. Волуме:2, Но:1, Yеар: 2022, Пагес: 30-35, Јуне 2022, Јоурнал оф Емергинг Цомпутер Тецхнологиес

[120] Данелла Патрицк; Абдуллах Ауб Кхан; Фаууаз Али; Ирфан Али Кандхро; Махноор Анwар; Асадуллах Кехар; Анwар Али Сањрани, “А новел цомпарисон оф дата аналутицс анд бусинесс интеллигенце тоолс: ан информатион пресерватион анд ледгер манагемент солутион “, Интернатионал Јоурнал оф Елецтрониц Сецуриту анд Дигитал Форенсицс, 2023 Вол.15 Но.4, пп.387 – 412, ДОИ: 10.1504/ИЈЕСДФ.2023.131957

[121] Вицтор М. Парра, Азеем Мохаммад, Али Суед, Малка Н. Халгамуге “Пентахо анд Јасперсофт: А Цомпаративе Студу оф Бусинесс Интеллигенце Опен Соурце Тоолс Процессинг Биг Дата то Евалуате Перформанцес “, (ИЈАЦСА) Интернатионал Јоурнал оф Адванцед Цомпутер Сциенце анд Апплицатионс, Вол. 7, Но. 10, 2016, п.20-29

[122] Андреиа Брандão, Елиана Переира, Мариса Естевес, Филипе Портела, Мануел Филипе Сантос, Антóнио Абелха анд Јосé Мацхадо “А Бенцхмаркинг Аналусис оф Опен-Соурце Бусинесс Интеллигенце Тоолс ин Хеалтхцаре Енвиронментс“, Информатион 2016, 7, 57; дои:10.3390/инфо7040057

Списак табела

Табела 1. Бенчмарк тест параметри за рачунаре коришћене у експериментима 62
Табела 2. Време (у мс) потребно за извршавање упита за различите „старије“ СУБП на различитим рачунарима и различитој меморији додељеној у Експерименту 1. 73
Табела 3. Време (у мс) потребно за извршавање упита за различите „новије“ СУБП на различитим рачунарима и различитој меморији додељеној у експерименту 2.....74
Табела 4. Време извршења ЕТЛ послова у секундама 95
Табела 5. Резултати експеримената БИ алата приказани у милисекундама.....112

Списак слика

Слика 1. Део шеме ВМарт ОЛТП базе....82
Слика 2. ВМарт ОЛТП табеле добијене трансформацијом ОЛТП табела са Слике 1....84
Слика 3. ОЛЕ ДБ Соурце и ОЛЕ ДБ Дестинатион компоненте....85
Слика 4. ОЛЕ ДБ Соурце едитор....85
Слика 5. Мапирање колона за табелу публиц_цалл_центер_дименсион....86
Слика 6. Пример ССИС посла у Дата Флов Едитору....87
Слика 7. Сорт Трансформатион Едитор....88
Слика 8. Мерге Јоин Трансформатион Едитор....88
Слика 9. Цонтрол Флов Едитор за цео ЕТЛ посао....89
Слика 10. Табле инпут Едитор....90
Слика 11. Сорт повс Едитор....91
Слика 12. Мерге јоин Едитор....92
Слика 13. Едитор мапирања колона података након трансформација и дестинационе табеле....92
Слика 14. Пример ПДИ трансформације 5 ОЛТП у једну ОЛАП табелу....93
Слика 15. Извршење ЕТЛ посла трансформација ОЛТП табела у листу Дименсион и Фацт ОЛАП табела....93
Слика 16. Навигатор прозор за избор табела из базе података....104
Слика 17. Креирање релација између табела у прозору Цреате Релатионсхип....105
Слика 18. Избор колона извештаја....106
Слика 19. Пражњење кеша из Дата Цацхе манагемент опција....107
Слика 20. Почетна wеб страна Пентахо Бусинесс Аналутицс....108
Слика 21. Пентахо Метадата Едитор за дефинисање Метадата Соурце у Пентахо БИ....109
Слика 22. Креирање релација између табела у прозору Цреате Релатионсхип....109
Слика 23. Упис упита у пољу Qуеру у Датасоурце Панелу....110
Слика 24. ЈаваСцрипт који приказује системско време започетог процеса....110
Слика 25. ЈаваСцрипт који приказује системско време завршетка извештаја....111

Списак графикона

Графикон 1. Брзина извршавања упита Q1 75
Графикон 2. Брзина извршавања упита Q1 и Q3 са већим Фацт табелама на најбржем рачунару који се користи у експериментима 76
Графикон 3. Брзина извршавања упита Q2 и Q4 са мањим Фацт табелама на најбржем рачунару који се користи у експериментима 77
Графикон 4. укупна времена извршења свих ЕТЛ послова за сваки од алата и за све три три различите хардверске платформе у секундама.....96
Графикон 5. Резултати експеримената БИ алата приказани у милисекундама.....113

Упоређивање перформанси аналитичких пословних информационих система

ORIGINALITY REPORT

1 %
SIMILARITY INDEX

PRIMARY SOURCES

1	hrcak.srce.hr Internet	100 words — < 1%
2	nardus.mpn.gov.rs Internet	70 words — < 1%
3	forums.parallax.com Internet	38 words — < 1%
4	www.cris.uns.ac.rs Internet	36 words — < 1%
5	sr.wikipedia.org Internet	24 words — < 1%
6	eteze.kg.ac.rs Internet	11 words — < 1%
7	fedorabg.bg.ac.rs Internet	9 words — < 1%
8	fedorakg.kg.ac.rs Internet	9 words — < 1%
9	www.singipedia.singidunum.ac.rs Internet	6 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF