Usporedba Data Warehouse i Big Data principa te njihova primjenjivost na sustave daljinskog očitanja potrošnje prirodnog plina i električne energije

Kluk, Vedran

Master's thesis / Diplomski rad

2018

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:200:247857

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-28**

Repository / Repozitorij:

Faculty of Electrical Engineering, Computer Science and Information Technology Osijek

zir.nsk.hr

Image not found or type unknown

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I INFORMACIJSKIH TEHNOLOGIJA

Sveučilišni studij

USPOREDBA DATA WAREHOUSE I BIG DATA PRINCIPA TE NJIHOVA PRIMJENJIVOST NA SUSTAVE DALJINSKOG OČITANJA POTROŠNJE PRIRODNOG PLINA I ELEKTRIČNE ENERGIJE

Diplomski rad

Vedran Kluk

Osijek, 2018.

SADRŽAJ:

1. UVOD 1
1.1. ZADATAK DIPLOMSKOG RADA 1
2. TEHNOLOGIJE ZA POHRANU I OBRADU VELIKE KOLIČINE PODATAKA 1
2.1. SUSTAV DATA WAREHOUSE 2
2.1.1. Karakteristike sustava Data Warehouse 2
2.1.2. Struktura sustava Data Warehouse 3
2.1.3. Primjena i ponuditelji sustava Data Warehouse 5
2.2. PRINCIP BIG DATA 5
2.2.1. Karakteristike principa Big Data 5
2.2.2. Struktura principa Big Data 6
2.2.3. Hadoop i Map Reduce 7
2.3. PRIMJENA I PONUDITELJI PRINCIPA BIG DATA 10
3. USPOREDBA SUSTAVA DATA WAREHOUSE I PRINCIPA BIG DATA 11
3.1. ODRŽIVOST IMPLEMENTACIJE SUSTAVA DATA WAREHOUSE I PRINCIPA BIG DATA 11
3.2. DOSTUPNOST SUSTAVA DATA WAREHOUSE I PRINCIPA BIG DATA 12
3.3. KORISNOST SUSTAVA DATA WAREHOUSE I PRINCIPA BIG DATA U POSLOVANJU 13
4. PRIMJENA SUSTAVA I PRINCIPA ZA POHRANU I OBRADU VELIKE KOLIČINE PODATAKA U DALJINSKOM OČITAVANJU POTROŠNJE PRIRODNOG PLINA 15
4.1. OPIS PILOT PROJEKTA 15
4.2. OPREMA I PODACI KOJI SE SKUPLJAJU PRILIKOM DALJINSKOG OČITANJA 16
4.3. ZAHTJEVI PILOT PROJEKTA 19
4.4. POTENCIJALNII PONUDITELJI RJEŠENJA PILOT PROJEKTA 21
4.4.1. Microsoft 21
4.4.2. Qlik 23
4.4.3. Statistical Analysis System (SAS) 25
4.5. USPOREDBA POTENCIJALNIH PONUDITELJA 26
4.6. ISPLATIVOST PROJEKTA 28
4.7. ODABIR NAJBOLJEG POTENCIJALNOG PONUDITELJA PILOT PROJEKTA POHRANE I OBRADE VELIKE KOLIČINE PODATAKA 29
5. ZAKLJUČAK 32
SAŽETAK 33
ABSTRACT 34
LITERATURA 36

1. UVOD

Porastom količine podataka koja se stvara svakog dana došlo je do razvoja sustava pohrane podataka u vidu modernih baza podataka. Porast količine podataka postao je očit prilikom pojave i intenzivnog korištenja Interneta i društvenih mreža. 90% ukupnih podataka koji postoje u svijetu stvoreno je u protekle dvije godine. Zbog toga su se u poslovnim sustavima počela primjenjivati Data Warehouse rješenja koja su odlična za strukturirane podatke kao i Big Data rješenja koja, uz strukturirane podržavaju i nestrukturirane podatke, koji se često ne uzimaju kao relevantni prilikom odlučivanja u poslovnom svijetu. Ovim diplomskim radom dana je usporedba sustava Data Warehouse i Big Data principa te njihove primjenjivosti na sustave daljinskog očitanja potrošnje prirodnog plina. U prvom poglavlju rada opisani su sustav Data Warehouse i Big Data princip na način da su prikazane karakteristike i strukture te su navedene primjene i potencijalni ponuditelji navedenih tehnologija. U drugom poglavlju uspoređene su navedene tehnologije za pohranu i obradu velike količine podataka kroz održivost rješenja, dostupnost te korisnost u poslovanju. U trećem poglavlju opisan je budući pilot projekt pohrane i obrade velike količine podataka koja se skuplja putem sustava daljinskog očitanja potrošnje prirodnog plina u suradnji s tvrtkom distributerom prirodnog plina te su utvrđeni zahtjevi pilot projekta koje je potrebno zadovoljiti od strane potencijalnih isporučitelja rješenja pilot projekta. Opisani su potencijalni ponuditelji pilot projekta, uspoređeni te je dano mišljenje o tome koji od navedenih isporučitelja, na temelju javno dostupnih podataka o mogućnosti ispunjenja zahtjeva pilot projekta, nudi optimalno rješenje za tvrtku distributera prirodnog plina.

1.1. ZADATAK DIPLOMSKOG RADA

Zadatak diplomskog rada je istražiti razliku između Data Warehouse i Big Data principa te izraditi konceptualno DW/BD rješenje za potrebe prikupljanja, spremanja i analize podataka prikupljenih iz sustava daljinskog očitanja potrošnje prirodnog plina, s fokusom na održivost rješenja (dugotrajno spremanje podataka) te poslovnu korisnost (validacija podataka, statičko izvještavanje, ad-hoc analiza podataka, primjenjivost za traženje odstupanja i predviđanje).

2. TEHNOLOGIJE ZA POHRANU I OBRADU VELIKE KOLIČINE PODATAKA

Napretkom informatizacije sve većih poslovnih sustava došlo je i do razvoja sustava pohrane sve većih količina podataka u vidu modernih baza podataka. U poslovnim sustavima su se

počela primjenjivati različita rješenja kao što su skladištenje podataka (eng. *Data Warehouse*, DW) koji je prisutan duže vrijeme te princip Big Data koji je relativno nov način rješavanja problema obrade velike količine podataka.

2.1. SUSTAV DATA WAREHOUSE

Sustav DW predstavlja veliku bazu podataka u koju se prikupljaju podaci iz različitih izvora, s ciljem obrade i vizualizacije podataka koja korisniku sustava pruža mogućnost izvlačenja skrivene vrijednosti iz ovisnosti obrađenih podataka.

2.1.1. Karakteristike sustava Data Warehouse

Ono što sustav DW izdvaja u odnosu na druge sustave je:

- jedinstvena verzija istine (eng. *single source of truth*, SSOT),
- performanse,
- jednostavnost,
- dosljednost podataka.

SSOT predstavlja praćenje jednakih podataka koji su prikupljeni putem sustava DW te svođenje istih na jedinstven oblik (lista kupaca sortirana prema prezimenu abecedno i sl.).
Performanse sustava DW predstavljaju vremenski interval u kojem će prikupljeni podaci iz različitih izvora biti obrađeni i dostupni u obliku izvještaja krajnjem korisniku sustava DW.
Ako ne postoji sustav DW, pronalazak potrebnih podataka korisnika koji koristi sustav svodi se na klasično pretraživanje raznih tablica i baza podataka. Navedena metoda može se primijeniti kada se radi o jednostavnim i malim količinama podataka koji ne zahtijevaju veliki vremenski period obrade podataka.
Međutim, ako se radi o većoj količini podataka čiju obradu nije moguće provesti klasičnim načinom pretraživanja podatak po podatak, potrebno je koristiti sustav DW koji je danas rasprostranjen i široko primjenjiv na različite vrste problema. Način obrade podataka odabire se shodno veličini podataka; količina podataka definira se kao velika ukoliko se obrada podataka ne može provesti u razumnom roku.
Jednostavnost kao karakteristika sustava DW odnosi se na sistematičniji prikaz podataka koji se nalaze u različitim bazama podataka i izvorima te ih korisnik sustava DW prikuplja iz određenih razloga.

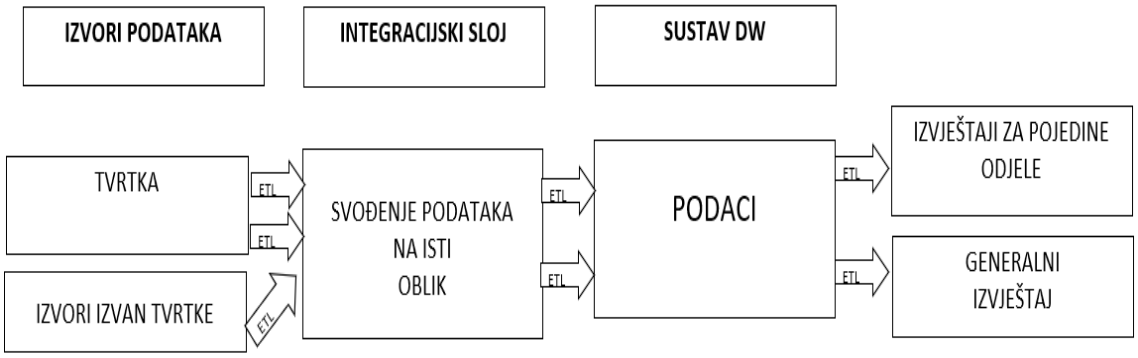
Dosljednost podataka odnosi se na praćenje podataka o kategoriji koja se promatra. Ukoliko se promatra baza podataka kupaca neke tvrtke koja želi izdvojiti podatke o prodaji proizvoda na temelju lokacije kupaca, tvrtka mora znati gdje je kupac bio u trenutku kupnje u odnosu na njegovu stalnu adresu stanovanja kako bi utvrdila radi li se o kupnji lokalnog stanovništva ili kupci dolaze iz nekog drugog dijela zemlje ili okruženja.

2.1.2. Struktura sustava Data Warehouse

Prema [1], sustav DW sastoji se od nekoliko slojeva koji kao rezultat obrade podataka daju izvještaje. Sustav DW najbolje je objasniti na primjeru tvrtke koja se sastoji od različitih odjela pa je na slici 2.1.2 prikazan pregled strukture sustava DW koji se sastoji od:

- izvora podataka,

- integracijskog sloja,
- skladišta podataka,
- izvještaja.



Sl. 2.1. *Pregled strukture sustava DW*

Podaci iz različitih izvora kopiraju se i odlaze u integracijski sloj gdje se događa postupak svođenja podataka na isti oblik. Nakon toga, slijedi postupak učitavanja podataka u sustav DW nakon čega se dobije konačan rezultat u obliku generalnog izvještaja koji je rezultat prikupljenih i obrađenih podataka. Također, između izvora podataka, integracijskog sloja, sustava DW i izvještaja nalazi se specifičan postupak manipuliranja podacima (eng. *Extraction, Transformation and Loading*, ETL) postupak.

ETL se dijeli u tri faze:

- raspakiravanje (eng. *extraction*),
- oblikovanje (eng. *transofrmation*),
- učitavanje (eng. *loading*).

Raspakiravanje podataka iz izvora podataka u integracijski odjel vrši se kako bi se sačuvala originalnost podataka svakog izvora.

Jednom kada se podaci nalaze u integracijskom odjelu potrebno ih je svesti na zajednički oblik. Nakon oblikovanja podataka, slijedi učitavanje istih u sustav DW kako bi se dobio konačan rezultat, a to su izvješća koja mogu služiti u različite svrhe; ako se promatra na primjeru tvrtke onda izvješća pomažu vodstvu tvrtke u donošenju odluka koje su od vitalnog značaja za poslovanje tvrtke.

Kada su podaci učitani u sustav DW, kreće izrada izvještaja koji su sastavljeni od podataka prikupljenih iz različitih izvora. Izvještaji su pravilno sastavljeni ako su jednostavni te ako daju konkretne informacije na sva relevantna pitanja koja krajnji korisnik sustava DW želi dobiti.

Problem je lakše riješiti na manjem skupu podataka nego na većem i upravo je to način na koji funkcioniraju izvještaji dobiveni putem sustava DW.

2.1.3. Primjena i ponuditelji sustava Data Warehouse

Prema [2], na tržištu postoji nekoliko velikih isporučitelja tehnologija zasnovanih na sustavu DW koji su prikazani na slici 2.2.:



Sl. 2.2. *Ponuditelji sustava DW*

2.2. PRINCIP BIG DATA

Princip Big Data predstavlja skup programskih alata koji omogućavaju prikupljanje i spremanje raznovrsnih, često nestrukturiranih podataka koji se generiraju unutar neke organizacije, a u sebi mogu sadržavati nove uvide u međuovisnosti pojava iz kojih nastaju. Prema [3], svrha principa Big Data je opaziti potencijalne probleme, pružiti uvid u postojanje potencijalnih problema te precizno predvidjeti buduća događanja na ekonomski učinkovit način.

2.2.1. Karakteristike principa Big Data

Prema [4], Big Data predstavlja pojam koji opisuje veliku količinu podataka (strukturiranih i nestrukturiranih), koja se prikuplja svakodnevno iz različitih izvora interesantnih korisniku. Međutim, nisu svi podaci koji se prikupljaju korisni te je korisniku potreban sustav ili princip kojim se odvaja bitno od nebitnog.

Pojam Big Data relativno je nov čak i u krugovima inženjera u sektoru informacijskih tehnologija (eng. *Information Technology*, IT). Korijeni principa Big Data postavljeni su prije 15-ak godina i do danas se usavršavaju. Postoje dvije vrste podataka koje se prikupljaju:

- strukturirani podaci,

- nestrukturirani podaci.

Prema [5], strukturirani podaci su oni tipovi podataka koji su korišteni od strane tvrtki desetljećima i predstavljaju brojčane vrijednosti, činjenice kao što su financijski rezultati, karakteristike kupaca i karakteristike proizvoda koji se pohranjenu u relacijskim bazama podataka.
Prema [5], nestrukturirani podaci predstavljaju digitalni sadržaj kao što su slike, videozapisi, tekstualne poruke i sl. Pojam nestrukturiranih podataka sličan je kao i pojam Big Data uz nedostatak podataka kao što su podaci prikupljeni sa senzora, lokacijskih podataka i strojnih podataka koji su strukturirani podaci i koji zajedno s nestrukturiranim podacima čine pojam Big Data.
Princip Big Data koristi se onda kada nije moguće klasičnim putem pomoću jednog računala obraditi podatke i dati odgovarajući rezultat iz razloga što računalo podatke prima i obrađuje serijski.
Upravo zato, potrebna su paralelna računala (poslužitelji), koji će primati manje količine podataka, obrađivati ih te slati na jedno računalo koje će sastavljati konačan rezultat u obliku izvještaja koji daju korisniku uvid u informativnost prikupljenih i obrađenih podataka.

2.2.2. Struktura principa Big Data

Princip Big Data sastoji se od nekoliko programa otvorenog koda (eng. *Open Source*) koji međusobno rade kako bi veliku količinu podataka koja dolazi iz različitih izvora obradili na što bolji i brži način. Ono što je karakteristika principa Big Data je paralelna obrada podataka. Platforma koja omogućava rad principa Big Data zove se Hadoop, a programski modeli koji sačinjavaju navedenu platformu su:

- HDFS (eng. *Hadoop Distributed File System*),
- Mapiranje i Reduciranje (eng. *MapReduce*).

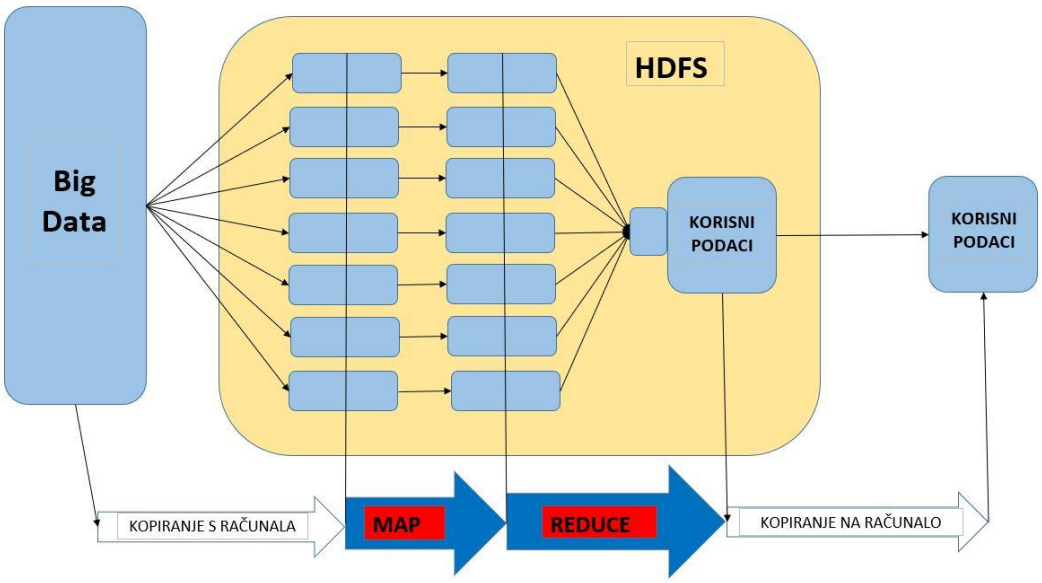
2.2.3. Hadoop i Map Reduce

Prema [6], Hadoop predstavlja platformu otvorenog koda koji služi za pohranjivanje podataka i pokretanje određenih aplikacija na računalima koji su međusobno povezani. Hadoop predstavlja temelj bez kojeg nema paralelne obrade podataka u sustavima koji primjenjuju princip Big Data te odvaja princip Big Data u odnosu na klasične baze podataka i sustav DW. Trend pada cijena uređaja za pohranu podataka omogućava dodavanje memorije bez značajnih povećanja troškova korisnicima koji koriste princip Big Data za obradu podataka. Bitno je napomenuti da se sustav DW i princip Big Data međusobno ne isključuju nego se nadopunjavaju te princip Big Data predstavlja nadogradnju sustava DW.
Vrste podataka koje podržava Hadoop su :

- strukturirani,
- polustrukturirani,
- nestrukturirani.

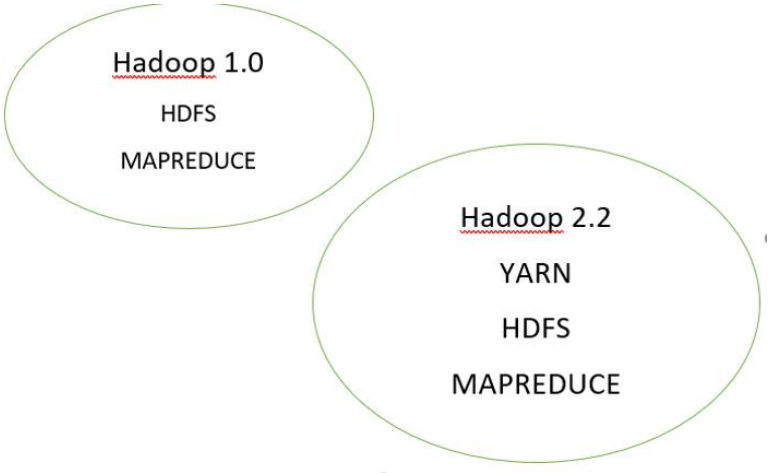
Hadoop zbog svojih karakteristika ima sposobnost obrade gotovo neograničene količine zadataka i poslova koji pristižu istovremeno na obradu.

Na slici 2.3. prikaz je način rada u Hadoop okruženju.



Sl. 2.3. *Rad u Hadoop okruženju*

Na slici 2.4. prikazane su inačice Hadoopa 1.0 (lijevo) i 2.2 (desno) te programski modeli koje svaka inačica obuhvaća.



Sl. 2.4. *Inačice Hadoopa*

Prema [6], Hadoop 1.0 obuhvaća HDFS i MapReduce. HDFS je ključan element Hadoop sustava koji podržava brzi prijenos podataka između računala koja su međusobno povezana. Usko je vezan za MapReduce koji je zadužen za obradu podataka.

HDFS je sustav koji prima podatke, raspodijeli ih na manje blokove podataka te šalje na različita računala kako bi se podaci obrađivali paralelno. Podaci koji se šalju na računala kopiraju se u slučaju da dođe do kvara istih pa se zato HDFS naziva još i sustav koji ima veliki prag tolerancije na kvarove. HDFS koristi *master/slave* arhitekturu sličnu kao i PLC (eng. *Programmable logic controller*) uređaji u industriji.

Uz HDFS, u inačici 1.0 Hadoop sadrži i MapReduce. MapReduce predstavlja tehniku obrade podataka koji su raspodijeljeni na većem broju međusobno povezanih računala.
Prema [7], sastoji se od dva dijela:

- Map,
- Reduce.

Map predstavlja radnju koja uzima podatke koji su u jednom obliku i pretvara ih u podatke koji su u drugom obliku na način da svaki podatak dobije ključ ovisno o tome što predstavlja podatak. Nakon toga, slijedi Reduce radnja koja raspoređuje podatke ovisno o njihovim ključevima i šalje obrađene podatke na izlaz sustava korisniku.
MapReduce može obavljati dvije funkcije:

- upravljanje resursima računala,
- obrada podataka.

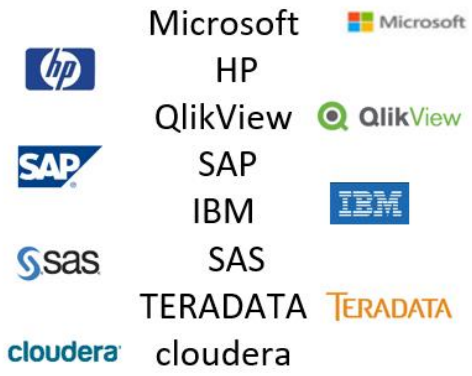
U inačici 2.2 Hadoop, uz HDFS i MapReduce sadrži i YARN (eng. *Yet Another Resource Negotiator*).
Prema [8], YARN se pojavio kao projekt Apache tvrtke koja je zadužena za programski alat Hadoop. Pojavila se potreba za upravljanjem resursima računala nevezano za MapReduce koju je preuzeo projekt YARN.

2.3. PRIMJENA I PONUDITELJI PRINCIPA BIG DATA

Princip Big Data svoju primjenu ima u velikim korporacijama koje su sastavljene od više manjih tvrtki koje su međusobno povezane. Služi za obradu svih vrsta podataka i izvlačenja korisnosti iz podataka kako bi se postigao određeni cilj korporacije. Big Data ne isključuje eksplicitno sustav DW nego se nadopunjuje i stvara ozbiljnu platformu za prikupljanje i obradu podataka. Prema [9], princip Big Data je korišten u rješavanju raznih poslovnih problema, a neke od tvrtki koje su uspješno primijenile princip Big Data su:

- Walmart (primjena kod upravljanja supermarketom),
- CERN (primjena kod istraživanja svemira),
- Netflix (primjena kod sustava prijedloga programa za korisnika),
- Rolls-Royce (primjena kod proizvodnje luksuznih automobila),
- Facebook (primjena kod promatranja korisnika),
- BBC (primjena u televizijskim medijima).

Prema [10], na tržištu postoji nekoliko velikih ponuditelja tehnologija Big Data koji su prikazani na slici 2.5.:



Sl. 2.5 *Ponuditelji tehnologija Big Data*

3. USPOREDBA SUSTAVA DATA WAREHOUSE I PRINCIPA BIG DATA

Sustav DW i princip Big Data se vrednuju kako bi krajnji korisnik dobio uvid u to što sustav ili princip predstavlja te donio odluku što je prikladnije za njegove potrebe. Sustavi i principi koji rade u stvarnom vremenu i pružaju rezultate prikupljanja i obrade podataka u razumnom roku mogu se promatrati kroz tri kategorije:

- održivost implementacije sustava,
- dostupnost sustava,
- korisnost u poslovanju.

3.1. ODRŽIVOST IMPLEMENTACIJE SUSTAVA DATA WAREHOUSE I PRINCIPA BIG DATA

Prema [11], svaki korisnik koja upravlja velikom količinom podataka trebao bi imati sustav DW koji omogućava prikupljanje i obradu podataka iz različitih izvora koji su mu zanimljivi. Alternative sustavu DW često su improvizirane baze podataka koje ne zadovoljavaju standarde današnjice u smislu obrade podataka u razumnom roku. Ukoliko korisnik smatra da će količina podataka koju će prikupljati u budućnosti narasti toliko da će imati potrebu nadograditi sustav DW na način da će se obrada podataka vršiti paralelno, onda se poseže za principom Big Data koji predstavlja nadogradnju sustava DW.

Sustav DW se koristi kod strukturiranih podataka dok su nestrukturirani podaci nepoznanica. S druge strane, princip Big Data se sastoji od skupa alata koji su uglavnom besplatni i služe za analizu svih vrsta podataka neovisno o tome jesu li strukturirani ili nestrukturirani.

Hadoop predstavlja samo jedan od alata kojim se manipuliraju svi prikupljeni podaci. Ulaganje u sustave za obradu i pohranu velike količine podataka je uzaludno ukoliko tvrtka ne ulaže u vlastite servere i obrazovane ljude koji su obučeni za korištenje sustava DW i principe Big Data. Dugoročno gledano, ulaganjem u ljude koji su sposobni koristiti resurse tvrtke na efikasan način štedi se vrijeme i novac koji se može ulagati u daljnje projekte radi poboljšanja cjelokupnog poslovanja tvrtke.

3.2. DOSTUPNOST SUSTAVA DATA WAREHOUSE I PRINCIPA BIG DATA

Prema [12], sustav je dostupan kada korisnici mogu raditi s njim bez osjećaja prekida rada. Stvara se slika da je sustav dostupan u trenucima kada korisnik ne osjeti prekid.

Mjerilo za dostupnost sustava je korisnik i njegov subjektivni dojam o tome je li nešto dostupno ili nije. Ako sustav prema ugovoru treba biti dostupan 360 dana u godini, a dostupan je 359, tada se smatra da sustav nije visoko dostupan. U dostupnost sustava uvijek se uračunavaju i prekidi zbog održavanja sustava koji su planirani i najavljeni te se ne uzimaju u obzir kad se računa broj dana kada je sustav bio u zastoju.

Najčešći uzroci prekida su:

- sklopovlje (eng. *hardware*),
- programska podrška (eng. *software*),
- ljudski faktor (pogreške korisnika i/ili operatera),
- logički prekidi (stanke zbog nedosljednosti u podacima pohranjenim u bazama podataka).

Nastoje se smanjiti prekidi sustava poboljšanjem kvalitete komponenata koje čine sustav, kako sklopovlja tako i programske podrške.

Ako sustavi s kojima korisnik radi nisu potpuno automatizirani nego zahtijevaju povremenu intervenciju ljudskog faktora, mogući su prekidi za koje nisu krivi, ni sklopovlje, ni programska podrška, nego osoba koja se koristi sustavom. Zbog toga, često u statistiku o prekidima rada sustava ulaze i prekidi uzrokovani korisnikom pa se može stvoriti loša slika o tome koliko je određeni sustav pouzdan i dostupan u određenom trenutku.

Dostupnost sustava DW se često dovodi u pitanje zbog problema dupliciranja podataka u bazama podataka. Kako bi se otklonio navedeni problem, potrebno je stvoriti on-line kopiju podataka koju koristi i off-line kopija podataka koja se koristi prilikom ETL postupka i drugih aktivnosti vezanih za održavanje baze podataka kako bi se povećao dojam dostupnosti sustava DW zajedno s brzinom obrade podataka. S druge strane, Hadoop je alat koji je napravljen da bi se mogao primijeniti princip Big Data u obradi podataka, a sastoji od dva dijela: HDFS (zadužen za spremanje i posluživanje podataka) te MapReduce (zadužen za obradu podataka). HDFS sustav se bazira na distribuciji računala koja paralelno obrađuju podatke što omogućava nastavak rada cijelog sustava ukoliko se dogodi kvar na jednom od računala koje obrađuje podatke. U tom slučaju zadatke računala u kvaru preuzimaju ostala računala koja paralelno obrađuju podatke.

3.3. KORISNOST SUSTAVA DATA WAREHOUSE I PRINCIPA BIG DATA U POSLOVANJU

Sustav DW predstavlja centralizirani sustav za pohranu podataka koji omogućava korisniku prikaz informacija prikupljenih iz više izvora na jednom mjestu. Takve informacije omogućavaju pravovremeno donošenje odluka koje se baziraju na izvještajima dobivenim putem navedenog sustava. Izolacija podataka i njihova optimizacija predstavljaju imperativ u današnjem poslovanju.

Poboljšana poslovna inteligencija predstavlja donošenje odluka na temelju potpunih informacija dobivenih putem sustava DW. Segmentacija tržišta, upravljanje zalihama proizvoda, financijsko upravljanje i prodaja samo su neki od dijelova poslovanja koji se mogu poboljšati pravovremenom informacijom.

S druge strane, važnost principa Big Data ne ogleda se u tome koliko podataka korisnik posjeduje nego kako koristi prikupljene i obrađene podatke. Princip Big Data prati različite izvore iz kojih pristižu podaci te paralelno obrađuje podatke koji su različitog oblika.

Prema [13] i [14], u tablici 3.1. i tablici 3.2. prikazani su prednosti i nedostaci sustava DW i principa Big Data iz koje se može zaključiti da je princip Big Data nadogradnja sustava DW koji se koristi kada korisnik posjeduje veliku količinu podataka različitog oblika koju nije moguće obraditi u razumnom roku.

Tab. 3.1. *Prednosti sustava DW i principa Big Data principa*

Prednosti sustava DW	**Prednosti principa Big Data**
• Analiziranje podataka iz više izvora na jednom mjestu • Analiziranje velike količine podataka na učinkovit način • Pravilno donošenje odluka i potencijalno veća dobit tvrtke • Poboljšana poslovna inteligencija • Poboljšane performanse u odnosu na klasične baze podataka • Pravovremeni pristup podacima • Dosljednost i standardizacija podataka • Smanjenje troškova tvrtke	Big Data uključuje sve navedene prednosti sustava DW, ali i slijedeće: • Smanjenje vremena potrebnog za obradu podataka • Razvijanje novih proizvoda • Razumijevanje tržišta i njegovih uvjeta • Poboljšanje poslovanja tvrtke pravilnom interpretacijom dobivenih informacija • Povjerenje suradnika korištenjem novih metoda analize podataka

Tab. 3.2. *Nedostaci sustava DW i principa Big Data principa*

Nedostaci sustava DW	**Nedostaci principa Big Data**
• Visoki troškovi održavanja • Serijska obrada podataka • Kriva procjena vremena potrebnog za odvijanje ETL postupka • Dug proces uvođenja sustava DW u svakodnevni rad tvrtke	• Vizualizacija prilikom velike količine podataka • Strah od promjena (nepoznatog) • Potencijalni sigurnosni napadi zbog povjerljivih podataka tvrtke

4. PRIMJENA SUSTAVA I PRINCIPA ZA POHRANU I OBRADU VELIKE KOLIČINE PODATAKA U DALJINSKOM OČITAVANJU POTROŠNJE PRIRODNOG PLINA

Za potrebe ovog rada dobio se uvid u opremu koja se koristi za analizu, pohranu i obradu podataka prikupljenih daljinskim putem te je oprema detaljno opisana u poglavlju 4.2. Tvrtka distributer za glavni cilj ima dobiti dnevno očitanje potrošnje prirodnog plina kako bi smanjila potencijalne troškove prilikom naručivanja plina od dobavljača.

4.1. OPIS PILOT PROJEKTA

Prema [15], pilot projekt daljinskog očitanja potrošnje prirodnog plina u tvrtki distributeru prirodnog plina pokrenut je 2016. godine te je finaliziran 2017.godine. Putem daljinskog očitanja prikuplja se određena količina podataka koja pomaže u analiziranju i procjeni potreba kupaca za plinom kako bi se što bolje procijenila količina plina koju tvrtka naručuje na dnevnoj bazi te da se izbjegnu penali koji se plaćaju u slučaju manjka ili viška plina u sustavu.
Nakon pilot projekta daljinskog očitanja potrošnje prirodnog plina planira se pokretanje pilot projekta koji će se baviti tematikom pohrane i obrade podataka prikupljenih putem daljinskog očitanja. Sustav DW i princip Big Data su načini pohrane i obrade podataka te su detaljno opisani u prethodnim poglavljima. Velika količina podataka zahtjeva obradu podataka različitog oblika u razumnom roku. Ako se radi o primjeru tvrtke koja uvodi sustave i principe kako bi dobila obrađene podatke u razumnom roku, na tvrtki je kako će iskoristiti dobivene podatke zbog toga što obrađeni podaci sami po sebi ne predstavljaju vrijednost ukoliko osoba koja interpretira podatke ne analizira iste na primjeren način. Nakon što su podaci prikupljeni i obrađeni, potrebno je sastaviti izvještaje koji će pokazati što obrađeni podaci znače kako bi tvrtka mogla pravovremeno djelovati po pitanju naručivanja plina te kako bi izbjegla dodatne troškove u slučaju prevelike ili premale količine plina u sustavu.

4.2. OPREMA I PODACI KOJI SE SKUPLJAJU PRILIKOM DALJINSKOG OČITANJA

Daljinsko očitanje potrošnje prirodnog plina bitno je kako za kućanstva tako i za pravne subjekte. Udio potrošača tvrtke distributera prirodnog plina, kada se radi o pravnim subjektima, iznosi od 5-8% dok je ostatak rezerviran za kućanstva. Prema [16], udio u potrošnji je nešto drugačiji. Kućanstva zauzimaju 52,97% ukupne potrošnje prirodnog plina, dok 5-8% pravnih subjekata potroši ostatak plina, točnije 47,03%. Prilikom mjerenja potrošnje prirodnog plina koriste se različite vrste plinomjera ovisno o veličini i potrebama poslovnog korisnika za razliku od membranskih plinomjera koji su najčešći kod kućanstava.

Za daljinsko očitanje podataka potrebna je nadogradnja klasičnih plinomjera putem određenog uređaja. Tvrtka distributer prirodnog plina koristi OKO l505 GPRS/SMS uređaj za prikupljanje i pohranu podataka (eng. *data logger*, DL) za plinomjere s regulacijom opskrbe.

Kartica mobilnog operatera koja se nalazi u DL-u koristi *GPRS/SMS* protokol kako bi izvršila slanja podataka koja sadrže dnevno očitanje potrošnje prirodnog plina. Programska podrška koja radi zajedno s karticama mobilnog operatera naziva se *Cisco Jasper* koji predstavlja upravljački centar za kartice mobilnog operatera i radi na *cloud* principu te omogućava upravljanje komunikacijom svih kartica mobilnih operatera i uređaja koji primaju informaciju koje kartice šalju. *Jasper* nakon toga komunicira s programskom podrškom OKO I505 DL-a.

Prema [17], DL omogućava daljinsko upravljanje ventilom bez pristupa prostorijama korisnika; na zahtjev zatvara ventil i daje dozvolu za sigurno ručno očitanje potrošnje prirodnog plina.. Na slici 4.1. prikazan je OKO 1505 GPRS/SMS uređaj za prikupljanje i pohranu podataka.



Sl. 4.1. *OKO l505 GPRS/SMS data logger*

Izvor: [17]

Podaci koji se prikupljaju prilikom daljinskog očitanja potrošnje prirodnog plina su:

- trenutno stanje uređaja,
- temperatura,
- dnevna potrošnja,
- GSM sažetak,
- detekcija neovlaštenog rukovanja: demontaža uređaja, vanjski magnet, prekoračenje temperature i sl.

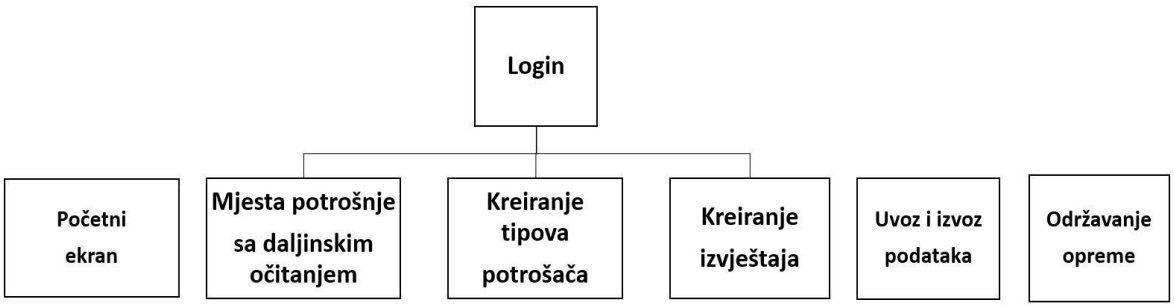
Na satnoj bazi se očitavaju svi navedeni podaci te se vrši komunikacija DL-a i programske podrške putem *GPRS/SMS* protokola svakog dana u isto vrijeme gdje je komunikacija otvorena određeni vremenski period (nekoliko minuta) te se nakon toga komunikacija prekida radi uštede baterije DL-a. Nakon toga kreće novi ciklus satnog prikupljanja podataka kako bi se idući dan u isto vrijeme otvorila komunikacija i poslali podaci o očitanim vrijednostima. Na slici 4.2. simbolično je prikazana komunikacija daljinsko očitanja potrošnje prirodnog plina.



Sl. 4.2. *Prikaz daljinsko očitanja prirodnog plina*

Izvor: [17]

Svi prikupljeni podaci moraju biti prikazani pomoću programske podrške. Razvojni tim tvrtke je 2017. godine dobio smjernice za izradu programske podrške. Na slici 4.3. prikazana je struktura programske podrške koju je razvojni tim tvrtke dobio za zadatak izraditi.



Sl. 4.3. *Struktura programske podrške*

Uz sve navedeno, potrebno je dodati komponentu koja nedostaje nakon što se podaci prikupe daljinskim očitanjem. Trenutno broj spojenih potrošača na daljinsko očitanje potrošnje prirodnog plina nije velik te se obrada podataka vrši klasičnim putem. Međutim, ukoliko se većina korisnika distribucijske mreže odluči spojiti na sustav daljinskog očitanja potrošnje prirodnog plina doći će do pristizanja ogromne količine podataka koju će biti potrebno obraditi na određeni način koji nije klasičan. Distribucijska mreža iznosi 595 km s ukupno 10.594 potrošača. Ako svaki potrošač na dnevnoj bazi pošalje 6 različitih informacija , to je ukupno 63.564 podatka koje treba obraditi na dnevnoj bazi. Kada se brojke tako prikažu, vidljiva je potreba za sustavom obrada podataka. Upravo tom problematikom bavit će se pilot projekt za prikupljanje i obradu velike količine podataka u daljinskom očitanju potrošnje prirodnog plina.

4.3. ZAHTJEVI PILOT PROJEKTA

Svaki projekt ima određene ciljeve i zahtjeve koji moraju biti ispunjeni kako bi se projekt smatrao uspješno obavljenim. Zahtjevi pilot projekta koji su navedeni u tablici 4.1. bit će poslani na adrese potencijalnih isporučitelja opreme i programske podrške koje čine cjelinu sustava za prikupljanje i obradu podataka.

Tab. 4.1. *Karakteristike DW/BD pilot projekta*

ZAHTJEVI PILOT PROJEKTA	OPIS ZAHTJEVA
CIJENA PROIZVODA	Cijena je jedan od bitnijih faktora. Optimalna cijena se formira po principu uloženo – dobiveno.
VRIJEME ISPORUKE	Poštivanje rokova isporuke važan je iz dva razloga: • zadovoljstvo kupaca i isporučitelja, • penali koji se plaćaju u slučaju nepoštivanja vremena isporuke.
NAČIN SPREMANJA PODATAKA	Dvije su vrste spremanja podataka: • lokalno (on premises), • *cloud.* Isporučitelj koji ponudi lokalno spremanje imat će prednost prilikom odabira izvođača projekta, zbog specifičnih zahtjeva naručitelja na sigurnost podataka.
MODULARNOST	Mogućnost nadogradnje bez izmjene cjelokupnog sustava. (izmjena određenih parametara, dinamičko osvježavanje te generiranje izvještaja, dodavanje novih funkcionalnosti).
TRAJNOST INVESTICIJE	Dostupnost ugrađene opreme u svrhu izvršavanja projekta mora biti dostatna za minimalno 5 godina kako programski tako i sklopovski.

SLA UGOVOR (*eng. Service Level Agrement*)	Koliko dugo je dostupna usluga održavanja sustava, korisnička podrška, mogućnost educiranja vlastitog tima za uporabu i certificiranje te koliko će sustav godišnje biti nedostupan, koji je rok za rješavanje prigovora i prijave nedostataka.
BRZINA SPREMANJA I OBRADE PODATAKA	Razina optimizacije, koliko su brzo podaci obrađeni.
SNAPSHOT BAZE PODATAKA	Statički pregled baze podataka, omogućava prikaz podataka koji su pohranjeni ranije u svrhu provjere i izrade izvještaja. Može se koristiti i kao osiguranje podataka u slučaju neuspješne nadogradnje ili pada sustava.

4.4. POTENCIJALNII PONUDITELJI RJEŠENJA PILOT PROJEKTA

Za potencijalne ponuditelje rješenja sustava pohrane i obrade podataka odabrani su idući:

- Microsoft,
- Qlik,
- SAS.

Svaki od navedenih potencijalnih ponuditelja u dosadašnjem radu ima iza sebe projekte vezane uz pohranu i obradu podataka te svaki projekt ima svoje specifičnosti.

4.4.1. Microsoft

Kada se gleda sa stajališta pohrane i obrade podataka, Microsoft ima rješenja koja su dominantna na tržištu. Proizvod koji pruža poznatu okolinu koja je prilagodljiva, platformu koja je pogodna za analizu u *cloud* okruženju naziva se Azure SQL Data Warehouse (eng. *Data Warehouse*, DW).

Prema [18], Azure SQL DW ima sljedeće karakteristike:

- brza implementacija,
- visoke performanse,
- dostupnost u cijelom svijetu,
- sigurnost na visokoj razini,
- pohrana i paralelna obrada u *cloudu*,
- kompaktibilnost s klasičnim alatima poslovne inteligencije (eng. *Business Inteligence,* BI).

Microsoft ima infrastrukturu koje je moćna i kao takav pogodan je za pohranu i obradu podataka što je dobro za tvrtke koji nemaju vlastitu infrastrukturu. Međutim, dovodi se u pitanje privatnost podataka u *cloud* načinu rada. Većina tvrtki želi svoje prikupljene podatke imati samo za sebe kako bi na temelju informativnosti istih poboljšala svoje poslovanje u odnosu na konkurente. Ono što je prednost kod *cloud* načina pohrane podataka je to što korisnik prema svojim potrebama može proširivati kapacitete pohrane kao i računalnu snagu koja je potrebna za obradu sve veće količine podataka. Proces proširenja kapaciteta spremanja podataka dostupan je u manje od 60 sekundi na zahtjev korisnika. Postupak raspakiravanja, oblikovanja i učitavanja podataka koji se koristi kod skladištenja podataka inače traje od tri do četiri sata

ako se radi o velikoj količini podataka. To je podatak ukoliko se obrada podataka vrši putem jednog računala. Međutim, Azure SQL DW nudi paralelnu obradu podataka koja je inače dostupna samo u principu Big Data pohrane i obrade podataka što predstavlja naprednu inačicu sustava DW. Također, bitno je spomenuti da Azure SQL DW ima mogućnost korištenja *polybase* baza podataka koje imaju mogućnosti pokretanja strukturiranih podataka kroz Azure SQL DW kao i polustrukturiranih i nestrukturiranih podataka kroz Azure *cloud* ili putem vlastite infrastrukture korisnika. Takav princip rada je hibridni gdje se koristi najbolje od sustava DW i principa Big Data kako bi se dobili obrađeni podaci u razumnom roku. Ono što je važno da tvrtka plaća onoliko prostora koliko joj je potrebno. Isto vrijedi i za računala koja služe za obradu podataka; ako tvrtka koristi malu procesnu moć i cijena je manja u odnosu na veliku obradu podataka. Fleksibilnost je ono što Microsoft naglašava kod svojih proizvoda.

Power BI je program koji je dostupan u sklopu Microsofta te omogućuje vizualizaciju dobivenih rezultata koji su prikupljeni i obrađeni kako bi se otkrila informativnost podataka.

Na slici 4.4. prikazano je sučelje Power BI programskog alata.

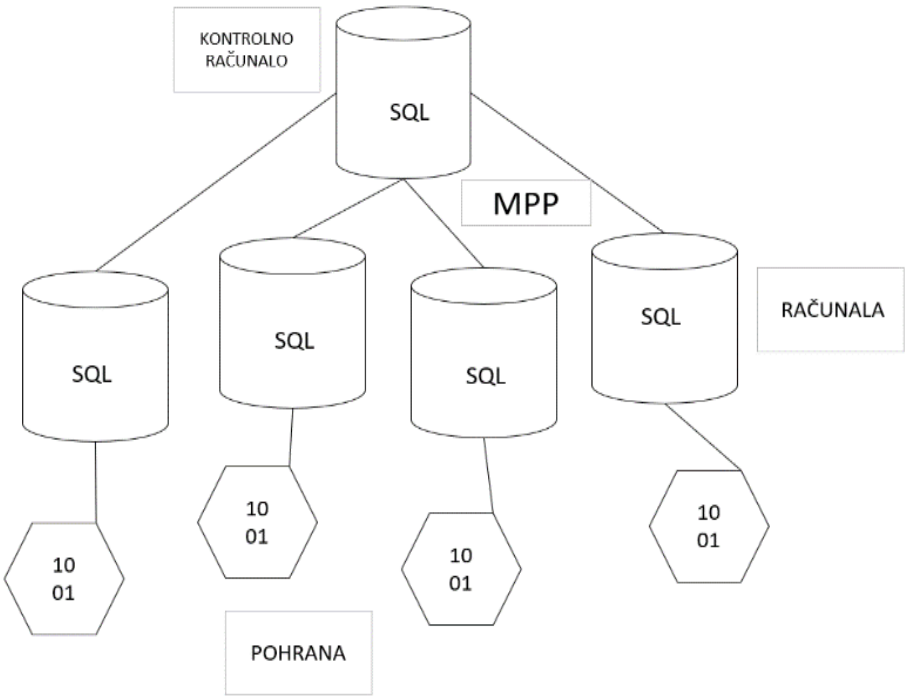


Sl. 4.4. *Sučelje Power BI programskog alata*

Izvor: [19]

Prema [20], razlika između klasične SQL baze podataka (eng. *Database*, DB) i SQL DW je u tome što je SQL DB samo dio SQL DW-a. SQL DB bavi se klasičnom obradom podataka s jednim računalom dok je SQL DW dizajniran na način da se koristi paralelna obrada podataka. Microsoft koristi tehnologiju za paralelnu obradu podataka (eng. *Massively parralel processing*, MPP) koja omogućava obradu podataka do maksimalno 60 računala u istom trenutku.

Na slici 4.5. prikazana je arhitektura SQL DW-a:



Sl. 4.5. *Paralelna obrada podataka*

Druga bitna stavka odnosi se na prostor koji je dostupan za pohranu podataka; SQL DB omogućava maksimalno 4 TB, ali će se u budućnosti omogućiti do 10 TB zbog sve veće potrebe za pohranom podataka. S druge strane, SQL DW nudi onoliko prostora koliko korisnik može platiti: ako je korisniku potrebno 1000 TB i ako korisnik to može priuštiti, Microsoft će pružiti tu uslugu. Naravno da takva usluga ima i svoju cijenu. SQL DW je znatno skuplji u odnosu na tradicionalnu SQL DB. Cijena za SQL DB iznosi početnih 4.20 EUR/mjesečno, dok Azure SQL DW iznosi oko 900 EUR/mjesečno što ne uključuje troškove pohrane podataka koji su uključeni u cijenu SQL DB-a. Cijena jednog TB kod SQL DW-a je oko 125 eura mjesečno po TB. Maksimalne cijena SQL DB iznosi 13.500 EUR dok cijena SQL DW-a može narasti do 57.000 EUR u što nije uključena cijena pohrane podataka. Međutim, kada se pogleda razlika u arhitekturi, vidljivo je zašto SQL DW košta neusporedivo više u odnosu na SQL DB.

4.4.2. Qlik

Drugi potencijalni isporučitelj je tvrtka Qlik. Proizvod koji je Qlik-u donio najveću popularnost je QlikView koji je fokusiran na *front-end* aspekt obrade i vizualizacije podataka. Drugi proizvod koji se ističe je QlikSense koji nudi kompletan spektar prikupljanja, transformacije,

spremanja, analize i vizualizacije podataka. QlikSense je potpuno novi proizvod kojega dodatno karakterizira modularna arhitektura gdje je svaki sloj rada s podacima servis za sebe. Pruža besplatnu podršku za analizu podataka i pruža korisnicima mogućnost izgradnje podataka i web aplikacija pomoću API-a (eng. *Application programming interface*).

U konkurenciji s Microsoftom, široj masi nije toliko poznata tvrtka Qlik i na pitanje koji je bolji većina ljudi odgovara Microsoft. Međutim, potrebno je definirati što znači bolje i koju logiku koristi tvrtka Qlik u svojim programskim rješenjima.

Prema [21], Qlik je orijentiran na krajnjeg korisnika i ono što krajnji korisnik vidi i može učiniti putem programske podrške na što lakši način predstavlja prioritet Qlik-u.

Treba napomenuti da je Qlik tehnologija, osim toga što je pristupačna za korištenje u smislu *drag and drop* načina korištenja sučelja, vrlo moćna i pruža korisniku brojne mogućnosti. Ukoliko korisnik odluči koristiti Qlik kao jedinu tehnologiju sustava DW, to pruža mnoge mogućnosti u pogledu brzine razvoja programske podrške. Korisniku je, nakon kupnje proizvoda, dozvoljena modifikacija programske podrške u određenoj mjeri kako bi se prilagodila potrebama korisnika. Ako su podaci smješteni u tablice, Qlik-u je dovoljan samo jedan pristup po tablici kako bi dobio sve podatke koji se nalaze u njoj te više nema potrebu pristupati cijelom sadržaju tablice nego se bazira na promjene koje se događaju u tablicama koje se uključuju u izvještaje koje programska podrška kreira kao rezultat obrade podataka. Navedeni postupak bitno skraćuje vrijeme učitavanja podataka.

Qlik platforma ne koristi klasičnu bazu podataka (relacijsku ili ne-relacijsku) za spremanje podataka. Umjesto toga, Qlik koristi vlasnički (*eng. propriatory)* sustav zapisivanja u *flat-file* infrastrukturu koja je optimizirana za brzinu i jednostavnost upisa i dohvaćanja podataka. Nadalje, naglasak na daljnju manipulaciju podacima u višim slojevima se prebacuje na radnu memoriju servera čime se dodatno ubrzava procesiranja podataka.

Ono što je nedostatak Qlik-a je to što se izvori podataka mogu koristiti isključivo pomoću Qlik programske podrške dok se druge baze podataka otvorenije i tradicionalnije te se mogu koristiti i putem ostalih BI alata.

Qlik napreduje iz godine u godinu i zauzima sve veći udio kod korisnika koje ne zanima što se događa u pozadini sustava nego im je bitno da su podaci obrađeni u razumnom roku kako bi mogli izvlačiti korisne informacije iz njih.

4.4.3. Statistical Analysis System (SAS)

Treći potencijalni ponuditelj je SAS koji predstavlja tvrtku koja ima više od 40 godina iskustva u pružanju analitičkih programskih rješenja svojim klijentima. SAS pripada kategoriji zrelijih (*eng. legacy*) platformi od ostalih potencijalnih ponuditelja kao što su Microsoft i Qlik.

SAS pruža programsku podršku koja može mijenjati, upravljati i preuzimati podatke iz raznih izvora i obavljati statističku analizu na njima. Također, grafičko korisničko sučelje pruža dodatne pogodnosti krajnjem korisniku ukoliko je korisnik struke koja nije tehnička. Za napredne korisnike SAS nudi mogućnost modifikacije sučelja i programa. Tvrtka ima različite proizvode koji se nalaze na tržištu:

- SAS Data Loader for Hadoop,
- SAS Enterprise Guide,
- SAS Enetrprise Miner,
- SAS Model Manager,
- SAS Visual Analytics,
- SAS Visual Statistics.

Za razliku od Microsofta i Qlika, SAS nudi nešto kompleksnija rješenja koja imaju cijenu koja nije konkurentna u odnosu na druge potencijalne isporučitelje. SAS koristi skupinu programa koji su pomno razvijani u svrhu dobivanja što veće informativnosti podataka. Tako postoji poseban program za obradu podataka koji se bavi samo time dok se vizualizacijom bavi poseban program pod nazivom SAS *Visual Analytics*. Takvim pristupom dobiveni se specijalizirani programski alati koji su napredniji od konkurencije, no često nude opcije koje nisu potrebne većini korisnika pa se za navedeni način odlučuju samo napredniji korisnici koji posjeduju veliku količinu novaca.
Izrazito moćni programski alati u samom svjetskom vrhu koji pruža mnoštvo mogućnosti koji imaju jednu manu, a to je da svaki program sam za sebe nema svrhu nego mora surađivati s alatima tvrtke SAS kako bi se dobila informativnost podataka zbog koje se radi projekt pohrane i obrade podataka. Svaki korisnik treba odabrati programsku podršku koja najbolje odgovara njegovim potrebama. Sukladno tome, SAS proizvodi imaju svoj krug korisnika koji su zadovoljni omjerom uloženo – dobiveno.

4.5. USPOREDBA POTENCIJALNIH PONUDITELJA

Nakon što se željene karakteristike pilot projekta pošalju potencijalnim isporučitelja rješenja, na red dolazi odabir najbolje ponude za tvrtku. Konkretna problematika kojom se tvrtka bavi je pohrana i obrada podataka nakon daljinskog očitanja potrošnje prirodnog plina. Potencijalni ponuditelji moraju zadovoljiti zahtjeve koji se postavljeni pred njih i odlučiti se na to hoće li sudjelovati u izradi projekta ili neće. Uzet ćemo za pretpostavku da su sva tri potencijalna ponuditelja poslala svoju ponudu te na temelju njihovih dosadašnjih javno dostupnih informacija dati mišljenje koji od njih bi bio najprimjereniji za izvršenje pilot projekta spremanja i obrade daljinski očitanih podataka vezanih za potrošnju prirodnog plina. U tablici 4.5 prikazane su neke od karakteristika koje su javno dostupne vezano za potencijalne isporučitelje.

Tab. 4.2. *Usporedba karakteristika potencijalnih isporučitelja*

ZAHTJEVI	MICROSOFT	QLIK	SAS
Cijena proizvoda	Azure SQL DW iznosi oko 900 EUR/mj. što ne uključuje troškove pohrane podataka Cijena se može popeti do 57.000 EUR/mj. u što nije uključena cijena pohrane podataka.	Qlik View Najskuplja licenca: 15.000 USD/korisnik Osobna inačica programa besplatna. Iznajmljivanje servera: 35.000 USD za velike tvrtke 8.400 USD za male tvrtke (uključuje 25 korisnika).	SAS nudi popis cijena za svoje proizvode koji su poprilično dugački. Daleko najskuplji od konkurencije, no nudi najefikasnije alate za obradu podataka (interesantno velikim tvrtkama).
Način spremanja podataka	*Cloud*, na zahtjev može i vlastita infrastruktura.	*Cloud*, na zahtjev može i vlastita infrastruktura.	*Clud*, na zahtjev može i vlastita infrastruktura.

Modularnost	Velika modularnost ako se radi o *cloud* načinu pohrane podataka, ogromna infrastruktura na raspolaganju.	U pogledu programske podrške vrlo velika, najfleksibilnije rješenje.	Omogućeno je naprednim korisnicima mijenjati programsku podršku putem posebnih alata i jezika koje proizvodi SAS.
Trajnost investicije	Zadovoljava potrebe koliko god je korisniku potrebno.	Zadovoljava potrebe koliko god je korisniku potrebno.	Zadovoljava potrebe koliko god je korisniku potrebno.
Brzina spremanja i obrade podataka	Paralelna obrada podataka do 60 računala u isto vrijeme.	Učitavanje podataka iz tablica samo jednom, svaki idući put učitavaju se samo promjene koje se događaju u tablicama što znatno smanjuje vrijeme obrade podataka.	Podaci o brzini spremanja i obradi podataka ovisi o vrsti sklopljenog ugovora te može odstupati u ovisnosti o potrebama naručitelja.
Snapshot baze podataka	DA	DA	DA

4.6. ISPLATIVOST PROJEKTA

Kada se govori o isplativosti projekta uvijek treba gledati nekoliko aspekata:

- kako se došlo na ideju projekta,
- koliki je omjer uloženog i dobivenog,
- što projekt donosi kao rezultat u budućnosti.

Projekt daljinskog očitanja potrošnje prirodnog plina je počeo 2016. godine. Ideja projekta je bila dobiti satno očitanje potrošnje prirodnog plina koji će se skupljati svaki sat i putem *GPRS/SMS* protokola poslati svaki dan u isto vrijeme. Takvo očitanje omogućava planiranje kupnje plina koji se naručuje na dnevnoj bazi kako bi se izbjegli penali koji se plaćaju u slučaju da se sav naručeni plin ne potroši ili ako plina ima nedovoljno u sustavu. Jednoznačno odgovoriti isplati li se ili ne projekt daljinskog očitanja nije moguće. Na slici 4.6 prikazan je odnos troškova opremanja i udjela u ukupnoj količini plina kumulativno po tarifnim modelima za poduzetništvo.



Sl. 4.6. *Odnos troškova opremanja i udjela u ukupnoj količini plina*

Izvor: [16]

Nakon što se podaci prikupe, potrebno ih je pohraniti i obraditi radi dobivanja informativnosti. Kako se trenutno radi o malom broj potrošača spojenih na sustav daljinskog očitanja potrošnje prirodnog plina, za obradu podataka može se koristiti klasičan način koji obavlja čovjek. Međutim, uslijed velike količine podataka koristi se paralelna računalna obrada podataka koja daje rezultate u razumnom roku. Uvođenjem sustava DW i principa Big Data ulaže se određena količina financijskih sredstava što na prvi pogled izgleda nelogično. Postavlja se pitanje zašto se ne očitava potrošnja prirodnog plina jednom mjesečno kao do sada uz pomoć studenta koje je jeftinije plaćati nego uvoditi posebne sustave; odgovor se krije u tome što tvrtka posjeduje 595 km distribucijske mreže koji nije jednostavno pokriti malom količinom ljudi te prikupljanje podataka iziskuje određeni vremenski period. Također, očitavanje jednom mjesečno u današnje vrijeme nije pogodno zbog toga što tvrtka nema uvid cijeli mjesec u potrošnju plina, a u mjesec dana se mogu dogoditi ogromni gubici pogotovo kod poslovnih subjekata čiji će udio u ukupnom dijelu potrošača u budućnosti rasti. Ako se problem daljinskog očitanja, spremanja i obrade podataka sagleda s navedenih aspekata, vidljivo je da je u konačnici veoma isplativo uvesti daljinsko očitanje potrošnje prirodnog plina te nakon toga sustav za pohranu i obradu podataka potrošnje prirodnog plina.

4.7. ODABIR NAJBOLJEG POTENCIJALNOG PONUDITELJA PILOT PROJEKTA POHRANE I OBRADE VELIKE KOLIČINE PODATAKA

Na temelju istraživačkog rada i prikupljenih podataka o potencijalnim ponuditeljima potrebno je odabrati optimalnu ponudu u odnosu na potrebe tvrtke. Sva tri potencijalna ponuditelja su renomirane tvrtke i imaju vrhunske proizvode u području analitike i obrade podataka. Međutim, svaki korisnik gleda i odabire proizvode i rješenja prema svojim potrebama. Najkompleksnije rješenje nudi ponuditelj SAS. Tvrtka se isključivo bavi analitikom, obradom i manipulacijom podacima što joj daje odlična znanja u tim područjima. Sam pogled na popis proizvoda i cijena koje SAS nudi pruža dovoljnu informaciju o tome o kakvoj se tvrtki radi i da svi ti proizvodi koštaju popriličnu sumu novca. SAS omogućava vrhunsku modularnost, brzina spremanja i obrade podataka na visokom nivou, nudi se *snapshot* baze podataka, način spremanja podataka prilagođava se kupcu proizvoda, međutim puno skuplje je ukoliko se tvrtke odluče spremati i obrađivati podatke putem vlastite infrastrukture ukoliko je ne posjeduju. Također, SAS ne posjeduje jedan alat koji je dovoljan za pohranu, obradu te vizualizaciju podataka nego postoje zasebni alati koji se bave pojedinačno svakim od navedenih zadataka. Samim time tvrtka SAS

predstavlja preveliko ulaganje za pilot projekt pohrane i obrade podataka prikupljenih daljinskim očitanjem potrošnje prirodnog plina.

Druga tvrtka koja se nameće kao potencijalni isporučitelj u izvođenju pilot projekta je Microsoft koji svoju DW/Big Data platformu gradi na skupu postojećih rješenja poput baza podataka (*eng. SQL Server*), Analytics Services (OLAP), izvještavanja (*eng. Reporting Services*) te Power BI programu koji služi za analitiku i vizualizaciju rezultata. Azure SQL DW je odličan proizvod po svojim karakteristikama. Nudi veliku procesnu moć u paralelnoj obradi podataka; zahvaljujući ogromnoj infrastrukturi Microsoft tvrtke koja nudi mogućnost uporabe do 60 računala istovremeno u svrhu paralelne obrade podataka. Instant integrabilnost jedna je od prednosti Microsofta s ostalim proizvodima poput Office paketa kojeg koristi veliki broj tvrtki. Način spremanja podataka vrši se putem *clouda*, no moguće je na zahtjev kupca i spremanje na vlastitu infrastrukturu što drastično povećava ulaganje u projekt ako tvrtka ne posjeduje vlastitu infrastrukturu. Konkurentnost Microsofta očituje se kroz licenciranje u *cloud* sustavu gdje su moguće različite inačice ugovora o količinskom licenciranju. Međutim, ukoliko se tvrtke odluče za *on-premise* rješenje, potrebno je sve navedene servise implementirati on-premise i to u najskupljim inačicama (Professional/Enterprise), što znatno poskupljuje implementaciju i buduće troškove. Osim toga, svaki od navedenih servisa je platforma za sebe i ima vlastiti način rada, kao i vlastiti jezik za rad s platformom. Određeni podaci se šalju Microsoftu bez obzira koji način spremanja podataka tvrtka odabere. Treba naglasiti da je danas privatnost podataka veoma bitna te da se sve veći broj tvrtki odlučuje investirati u vlastitu infrastrukturu kako bi zaštitili podatke svojih klijenata, a samim time i sebe. Cijena proizvoda Microsoft tvrtke druga je po veličini od sva tri potencijalna isporučitelja što je smješta u zlatnu sredinu prilikom odabira.

Treća tvrtka koja je u eksponencijalnom rastu i ima veliki potencijal je tvrtka Qlik sa svojim proizvodima QlikView i QlikSense koji zauzima veliki dio tržišta kada se radi o obradi i vizualizaciji podataka. Platforma nudi kompletan spektar gdje su svi slojevi, od učitavanja do analitike i vizualizacije, integrirani u jedno *on-premise* serversko rješenje. *On-premise* server se može i paralelizirati na način da više računala vrše paralelno procesiranje radi povećanja brzine. Iako jednostavno rješenje, nije monolitno već se sastoji od niza slojeva što uvelike olakšava integrabilnost. Cijela platforma omogućava jednak način rada kroz sve slojeve.

Dodatnu prednost Qlik-u daje i model po kojem posluju: koliko korisnika tvrtka priključi toliko licenci plati. Ako tvrtka u dvije godine naraste toliko da ima potrebu priključiti još 5 dodatnih korisnika, platit će 5 licenci koje koštaju 15.000 USD po licenci. Uz licencu potrebno je iznajmiti servere ukoliko tvrtka nema vlastite servere, no mogućnost odabira spremanja putem

clouda ili vlastite infrastrukture dodatan je plus. Također, Qlik je najfleksibilnije rješenje od sva tri ponuđena u pogledu modularnosti te u pogledu brzine spremanja i obrade podataka zauzima prvo mjesto.

Iz svega navedenog smatram da tvrtka Qlik najviše odgovara tvrtki distributeru prirodnog plina za potrebe daljinskog očitanja potrošnje prirodnog plina te pohrane i obrade podataka kako bi se naglasila informativnost podataka.

5. ZAKLJUČAK

Cilj ovog diplomskog rada bio je usporediti Data Warehouse i princip Big Data te definirati njihovu primjenjivost na sustave daljinskog očitanje potrošnje prirodnog plina. Iz usporedbe je vidljivo da se radi o sustavu i principu koji se međusobno nadopunjuju i nadograđuju te kao spoj Data Warehouse i princip Big Data čine jedinstvenu cjelinu koja služi za prikupljanje i obradu velike količine podataka različitog oblika. U ovom radu Data Warehouse i princip Big Data uspoređeni su prema održivosti rješenja, dostupnosti sustava te korisnosti u poslovanju. Također su u sklopu ovog rada sastavljeni zahtjevi na pilot projekt u suradnji s lokalnim distributerom prirodnog plina. Glavni cilj pilot projekta je primjena Data Warehouse i principa Big Data na sustave daljinskog očitanja potrošnje prirodnog plina. Na osnovu definiranih zahtjeva opisana su postojeća rješenja iz javno dostupnih informacija vezano za potencijalne ponuditelje programskih rješenja za pilot projekt. Opisani ponuditelji su: Microsoft, Qlik i SAS. Pružen je osvrt na isplativost projekta daljinskog očitanja potrošnje prirodnog plina kao i sustava za pohranu i obradu velike količine podataka. Uspoređena su sva tri isporučitelja te je dana evaluacija o tome čije rješenje daje optimalne rezultate za pilot projekt pohrane i obrade velike količine podataka koja se skuplja putem sustava daljinskog očitanja potrošnje prirodnog plina. Sastavljeni zahtjevi i dano mišljenje pružaju dobru podlogu za buduću provedbu pilot projekta tvrtke distributera potrošnje prirodnog plina.

SAŽETAK

Usporedba Data Warehouse i Big Data principa te njihova primjenjivost na sustave daljinskog očitanja potrošnje prirodnog plina i električne energije

U daljinskom očitanju prirodnog plina prikuplja se velika količina podataka, zbog toga se u tvrtkama koje se bave time javlja problem kako te podatke skladištiti i obraditi. U novije vrijeme dvije osnovne tehnologije koje se pojavljuju u tu svrhu su Data Warehouse i Big Data te svaka od njih omogućuje različitu tehniku obrade podataka; Data Warehouse serijski obrađuje podatke dok Big Data koristi paralelnu obradu podataka. Za potrebe pilot projekta sastavljeni su zahtjevi koji će biti poslani potencijalnim ponuditeljima programskog rješenja pilot projekta. Prikazana je usporedba potencijalnih ponuditelja rješenja te je utvrđeno da za potrebe distributera prirodnog plina najviše odgovara ponuditelj Qlik. Navedeni rezultati također pokazuju da preostali potencijalni ponuditelji nude kvalitetna programska rješenja.

Ključne riječi: Data Warehosue, Big Data, Daljinsko očitanje

ABSTRACT

Comparison of Data Warehouse and Big Data principle and their applicability to the remote readout systems of natural gas and electrical energy

In the remote reading of natural gas, a large amount of data is collected causing the problem of storing and processing such a type of data for the companies involved. More recently, two main technologies that appear for this purpose are Data Warehouse and Big Data, and each of them provides different data processing techniques; Data Warehouse uses serial data processing while Big Data uses parallel data processing. For the purpose of pilot project, in colaboration with the local natural gas distributer, the requirements for processing and storing the data collected are compiled and sent to the potential vendros. A comparison of the software packages provided by the potential vendors had shown that Qlik offers the best software package for the needs of the local natural gas distributor. The results also show that the other potential vendors also offer quality software packages.

Key words: Data Warehouse, Big Data, pilot project, remote reading.

ŽIVOTOPIS

Vedran Kluk rođen je 23.6.1993. u Našicama. Pohađao je osnovnu školu Josipa Jurja Strossmayera u Đurđenovcu. Srednju školu Isidora Kršnjavoga završio je u Našicama i stekao zvanje ekonomista 2012. godine. Iste godine upisuje preddiplomski studij računarstva na Fakultetu elektrotehnike, računarstva i informacijskih tehnologija te isti završava 2016. godine. Trenutno studira na Fakultetu elektrotehnike, računarstva i informacijskih tehnologija u Osijeku na diplomskom studiju, smjer DRAB - procesno računarstvo. Engleski vrlo dobro govori i piše. Posjeduje znanja vezano za programske jezike C i Matlab.

Potpis:

LITERATURA

[1] A.Wicks, Data Warehousing, A Quick Summary [online], YouTube, 2013., dostupno na: https://www.youtube.com/watch?v=zTs5zjSXnvs&t=10s, [15. travnja 2018.]

[2] J. Walker, Top 5 data warehouses on the market today, [online], Monities, 2018, dostupno na: http://www.monitis.com/blog/top-5-data-warehouses-on-the-market-today/, [17. travnja 2018.]

[3] R.Buyya, R.N.Calheiros, A.V.Dastjerdi, Big Data - Principles and Paradigms, Elsevier, New York, 2016.

[4] M. Rouse , Using big data platforms for data management, access and analytics, [online], TechTarget, 2013, dostupno na: http://searchdatamanagement.techtarget.com/definition/Hadoop-Distributed-File-System-HDFS , [5. svibnja 2018.]

[5] S.Williams, Business Intelligence Strategy and Big Data Analycis, Elsevier, New York, 2016.

[6] B. Marr, Big Data:What is Hadoop-An Easy Explanation For Absolutley Anyone, [online], Bernard Marr & Co., 2018, dostupno na: https://www.bernardmarr.com/default.asp?contentID=1080 , [30. travnja 2018.]

[7] Tutorialpoint-simplyeasylearning, [online], Tutorialpoint, 2018, dostupno na: https://www.tutorialspoint.com/hadoop/hadoop_mapreduce.htm, [9. svibnja 2018.]

[8] Apache Hadoop YARN, [online], TechTarget, 2013, dostupno na: https://searchdatamanagement.techtarget.com/definition/Apache-Hadoop-YARN-Yet-Another-Resource-Negotiator, [15 svibnja 2018.]

[9] B.Marr:, Big Data in Practice – How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results, Wiley, Hoboken, 2016.

[10] A. Patrizio, Big Data Companies, [online], Datamation, 2017, dostupno na: https://www.datamation.com/big-data/big-data-companies.html , [19. svibnja 2018.]

[11] Big Data for Sustainable Development, [online] United Nations, 2017, dostupno na: http://www.un.org/en/sections/issues-depth/big-data-sustainable-development/index.html , [19. svibnja 2018.]

[12] I.Bjorhovde, Redefining High Availability for a Data Warehouse, [online], IDUG, 2014, dostupno na: https://www.idug.org/p/bl/et/blogid=278&blogaid=278 , [24. svibnja 2018.]

[13] The Benefits of Data Warehousing and Extract, Transform and Load (ETL), [online], Glow Touch Technologies, 2016, dostupno na: https://www.glowtouch.com/blog/analytics/the-benefits-of-data-warehousing-and-etl/, [1. lipnja 2018.]

[14] Blog, 5 Benefits: Competitive Adventages of Big Data in Business, [online], Newgen apps, 2017, dostupno na: https://www.newgenapps.com/blog/importance-benefits-competitive-advantage-big-data, [8. lipnja 2018.]

[15] D. Guberac, Smjernice za izradu programa za obradu podataka daljinskog očitanja, Prvo plinarsko društvo – distribucija plina d.o.o., Vukovar, 2017.

[16] M. Horvacki Zivalov, Daljinsko očitanje kod pravnih osoba, Prvo plinarsko društvo – distribucija plina d.o.o., Vukovar, 2017.

[17] OKO i505 GPRS/SMS data logger, [online], IMR, 2018, dostupno na: www.smartgasmetering.com, [13. lipnja 2018.]

[18] Microsoft, [online], Microsoft Azure, 2018, dostupno na: https://azure.microsoft.com/en-us/services/sql-data-warehouse/, [17. lipnja 2018.]

[19] Izgled korisničkog sučelja Microsoft Power BI alata, [online], TechNet, 2018, dostupno na: https://blogs.technet.microsoft.com/franmer/2015/07/30/sql-azure-data-warehouse-and-power-bi/, [17. lipnja 2018.]

[20] Microsoft, [online], Bitools, 2017, dostupno na: http://microsoft-bitools.blogspot.com/2017/07/azure-sql-database-vs-azure-sql-data.html,[17. lipnja 2018.]

[21] C. Lupo, Can Qlik technology be cosidered as a full data warehouse?, [online], Element 61, 2017, dostupno na: https://www.element61.be/en/resource/can-qlik-technology-be-considered-full-data-warehouse , [18. lipnja 2018.]