

PAMETNA OMREŽJA IN IZRABA VELIKE KOLIČINE PODATKOV

ARSO SAVANOVIĆ
SMART COM d.o.o. Ljubljana
arso.savanovic@smart-com.si

TOMAŽ BRAČIČ
SMART COM d.o.o. Ljubljana
tomaz.bracic@smart-com.si

UROŠ HROVAT
SMART COM d.o.o. Ljubljana
uros.hrovat@smart-com.si

PETER CEFERIN
SMART COM d.o.o. Ljubljana
peter.ceferin@smart-com.si

Povzetek – V pametnih omrežjih, kjer se število inteligentnih naprav, ki merijo, zbirajo, pošiljajo ali izmenjujejo podatke med seboj ali z aplikacijskimi rešitvami naglo povečuje. Podatki, ki se zgolj kopičijo v podatkovnih bazah niso polno izkoriščeni, lahko pa nosijo informacije, iz katerih s primernim pristopom lahko pridobimo dodatne funkcionalnosti v pametnih omrežjih, ki se odražajo v analitskih obdelavah, korelacijah in potencialno tudi v predikcijah. Fenomen velike količine podatkov v pametnih omrežjih (ang. Big Data) je potrebno obravnavati sistematsko, saj so možnosti ki jih nudijo tehnološke rešitve široke, za uporabne rešitve oz. izdelke pa potrebno tesno sodelovanje strokovnjakov s področja elektroenergetike ter IT.

V članku bo prikazano področje velike količine podatkov, kje se pojavlja ter kakšne so možnosti uporabe v elektroenergetskem sistemu pametnih omrežjih, ter kakšne so tehnološke možnosti razvoja tega področja..

SMART GRIDS AND USE OF BIG DATA

Abstract – Number of different active devices in Smart Grid infrastructure is rapidly growing, thus causing rapid growth of the data volume exchanged in the network and towards control centres of e.g. distribution companies. Traditional data base approach is not providing solutions, required by Big data phenomena. Collecting, storage, analytics and correlations are becoming important, since raw data cannot provide added value. The article is dealing with the technological principles, which are providing possibilities to ensure further exploitation of high data volume in Big Data systems.

I. UVOD

Smernice razvoja v pametnih omrežjih elektrogospodarskih podjetij kažejo na vse večjo informatizacijo in avtomatizacijo omrežja. Cilji npr. elektrodistribucijskega podjetja ali družbe, ki z električno energijo trguje in jo kupuje, je priti do čim boljšega poznavanja elektroenergetskega omrežja in preko tega do možnosti vplivanja nanj tudi preko temeljitega poznavanja končnega uporabnika, njegovih navad in njegovega vplivanja na omrežje. Merilni sistemi in sistemi avtomatskega upravljanja elementov omrežja se zato širijo tudi na nižje napetostne nivoje, bliže uporabnikom. Spreminja se tudi paradigma delovanja uporabnika v sistemu, ki ni več samo pasivni odjemalec ali proizvajalec energije, temveč vse bolj aktivno sodeluje pri obratovanju omrežja. Vse te spremembe bodo zahtevale porast aktivnih naprav v elektroenergetskem sistemu, čemur smo priča že danes, tudi v slovenskem elektroenergetskem sistemu (primer: uvajanje pametnih števecv in AMI infrastrukture): Vse to prinaša velike količine podatkov, ki pa jih bo potrebno zajemati, obdelovati in tudi ustrezno prikazati oz. iz zajetne množice podatkov pridobiti uporabne informacije, bodisi za spremljanje infrastrukture, analitiko, načrtovanje ali za potrebe novih storitev z dodano vrednostjo za akterje na trgu električne energije.

II. PROBLEM VELIKE KOLIČINE PODATKOV

Velike količine podatkov predstavljajo vedno večji izziv na vseh področjih, kjer je prisotna visoka stopnja informatizacije. Uporablja se angleški izraz "big data", ki opisuje stanje, ko količina podatkov s katero upravljamo presega trenutno obvladljive zmožnosti za zajemanje, hranjenje, upravljanje in njihovo analizo. Tu gre za praktično definicijo pojmovanja velike količine podatkov, saj lahko vsakemu posamezniku, ki se trudi pridobiti prednost na podlagi pridobljenih podatkov, velika količina podatkov prinaša drugačno vrednost.

Smart Grids oziroma pametna elektroenergetska omrežja niso prvo področje, kjer se danes soočamo z velikimi količinami podatkov, zato so se na področju IT-ja že pred časom uveljavile nove tehnologije imenovane NoSQL (not only sql). Najbolj znana so velika spletna podjetja, ki nudijo visoko skalabilne spletne aplikacije, kot so Google, Facebook, Amazon. Sedaj tudi podjetja iz drugih tehnoloških področij te rešitve sprejemajo in jih prilagajajo svojim potrebam in zahtevam.

Tako morajo tehnologije "Big Data" v elektroenergetskih podjetjih zadostiti določenim zahtevam, vključujoč obdelavo podatkovnega toka,

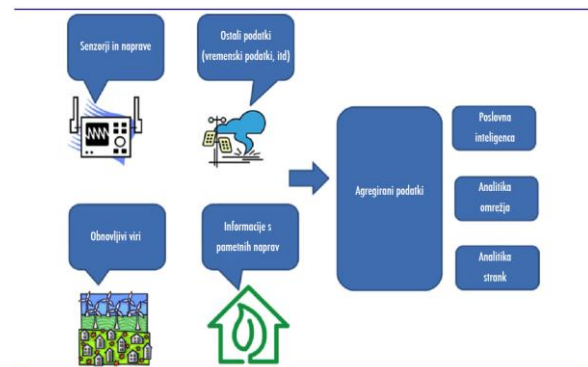
indeksacijo podatkov, rudarjenje nad podatki, vizualizacijo, itd.

Količina podatkov bo v prihodnjih letih strmo naraščala, zato je pomembno, da so izbrane tehnologije sposobne obvladovanja pojava in porasta velike količine podatkov.

Zaradi kompleksnih podatkovnih struktur se uvajajo novi pristopi k integraciji ter izrabi analitičnih možnosti, s tem pa se spreminja in dvigajo meje, ki smo jih poznali v okviru tradicionalnega upravljanja s podatki.

Podatkovna skladišča za velike podatkovne zbirke, kot so podatki zbrani iz različnih virov v pametnih omrežjih, morajo biti sposobna obvladovanja, shranjevanja in analiziranja hitro rastoče količine izmenjanih podatkov v ustreznih časovnih obdobjih.

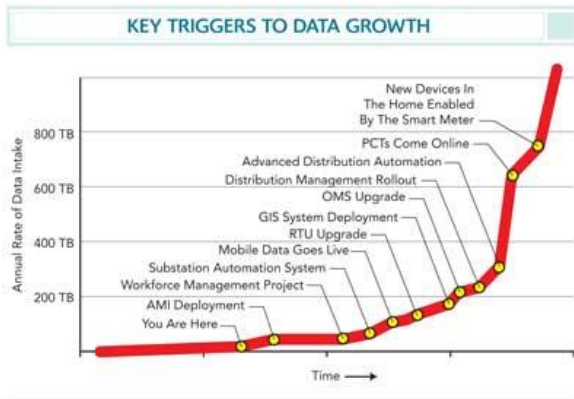
Te tehnologije se v tujini že pojavljajo na področju upravljanja elektroenergetskih omrežij oz. Smart Grids. Primer je podjetje Tennessee Valley Authority, ki predstavlja enega večjih dobaviteljev električne energije končnim uporabnikom v ZDA, za shranjevanje in manipulacijo podatkov, ki jih prejema iz množice aktivnih naprav v omrežju pa že uporablja zgoraj omenjeno tehnologijo.



Slika 1: Informacijski pretok analitičnih podatkov v pametnih omrežjih od vira do uporabnika

Na področju pametnih elektroenergetskih omrežij je, kot že omenjeno, vedno večji izziv obvladovanje velike količine podatkov. Velikost podatkovnih sistemov raste v velikostne razrede Tb (Tera bajt) in Pb (peta bajt).

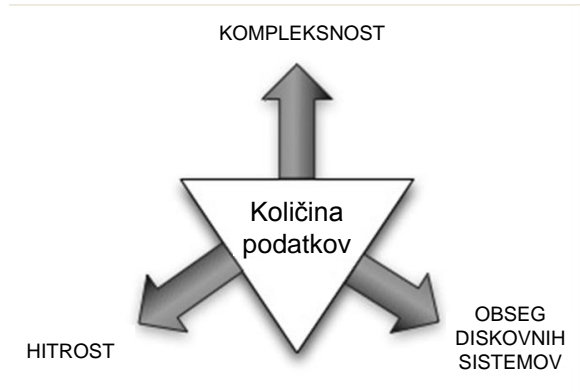
Trendi v tej smeri se nadaljujejo na račun dela z vedno novimi in novimi množicami podatkov. Slika 2 zelo nazorno prikazuje rast mase podatkov na račun modernizacije elektroenergetskih sistemov oz. vključevanja konceptov pametnih omrežij..



Slika 2: Prikaz narave rasti količine podatkov, ki jih prinašajo pametna omrežja

Podatkovni sistemi postajajo tako kompleksni, da njihovo upravljanje in obvladovanje z dosedaj množično uporabljenimi klasičnimi podatkovnimi orodji postaja vedno bolj neučinkovito, posledično pa tudi drago. Ključne ovire in neučinkovitost izvira iz težave, ki nastanejo pri zajemanju, hranjenju, iskanju, deljenju, analitiki, vizualizaciji, itd. Zbiranje podatkov seveda ni samo sebi namen, zato takoj nastopijo nadaljnje težave, ki onemogočijo učinkovito upravljanje s temi podatki, analitske operacije na njih, shranjevanja in nadaljnja uporaba pridobljenih informacij, itd.

Zaradi velike količine strukturiranih, delno strukturiranih in nestrukturiranih podatkov, ki jih bomo konstanto pridobivali iz elektroenergetskega omrežja, kmalu pridemo do spoznanja, da za učinkovito upravljanje s takimi podatki izjemno hitro narašča kompleksnost in zahtevnost obvladovanja le teh, v kolikor uporabljamo klasične metode z uporabo standardnih RDBMS sistemov. Vse skupaj hitro vpliva tudi na stroškovni vidik, saj klasični pristop pomeni vedno večje in močnejše strežnike ter največkrat še centralne diskovne sisteme. Danes se pri uporabi klasičnih RDBMS sistemov tem problemom zoperstavimo s širjenjem tehnologije navzgor, kar pomeni večje, boljše in dražje diskovne sisteme. Poleg porasta zahtev glede diskovnih enot smo soočeni tudi z zahtevo po boljših, zmogljivejših strežnikih, porastom njihove količine, itd. Slika 3 prikazuje tri področja, na katera pristop s klasično infrastrukturo najbolj vpliva, jim povečuje kompleksnost in jih tudi najbolj draži.



Slika 3: Področja na katera vpliva velika količina podatkov

Kompleksnost (complexity) - Podatki že dolgo niso več samo tekst in števila. Vedno več podatkov je vezanih na dogodke v realnem času (real-time) oz. trenutnih dogodkov in vrednosti iz različne opreme v omrežju, itd. Smernice kažejo na vedno večje število različnih vrst podatkov in njihove pojavnosti. Vedno več je tudi nestrukturiranih podatkov. Apliciranje klasičnih poznanih algoritmov za iskanje, hranjenje in kategorizacijo postaja vedno bolj kompleksno.

Hitrost (speed) – parameter je naveden v smislu hitrosti zajemanja podatkov, kot je primer video visoke kvalitete, ali t.i. streaming (kontinuirano pošiljanje podatkov), itd. Obsežnost informacij, ki se širi na tak način je ogromna. Hitrosti so običajno zelo visoke. Če je potreba, da se vse to računalniško obdeluje, prenaša, hrani, mora tudi infrastruktura na ustrezen način to omogočati.

Obseg diskovnih sistemov (volume) - Vse podatke, ki jih hranimo, je potrebno hraniti varno in zagotavljati njivo uporabo za daljše časovno obdobje. Lahko se sicer uvede politika periodičnih brisanj, vendar določenega tipa podatkov ne smemo brisati zaradi politike oz. narave podatkov samih.

III. REŠITVE

Elektroenergetska podjetja bodo na področju obdelave in analize pametnih omrežij imele največ koristi od tehnologij, ki lahko obdelujejo tako strukturirane, kot tudi nestrukturirane podatke in omogočajo ostalim aplikacijam v sistemu zagotoviti zgodovinske podatke in informacije v realnem času.

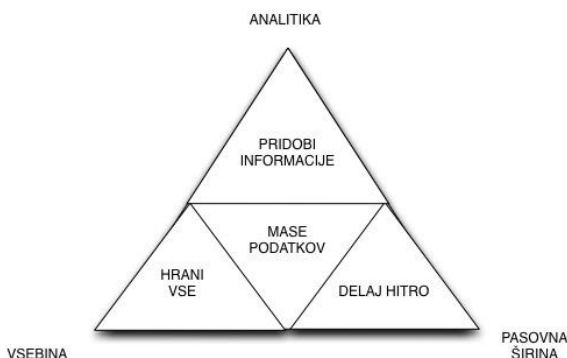
Zato smernice razvoja na področju obvladovanja velike količine podatkov prešli na drugačne koncepte, kot smo jih vajeni dosedaj, saj pri njih naletimo na neobvladljive težave, opisane v prejšnjem poglavju. Konceptualna zamisel, ki nakazuje rešitve je v bistvu precej preprosta in uvaja širitev na istem nivoju, vendar v širino po principu več poceni strežnikov,

povezanih v enovito strukturo, vendar brez centralnih diskovnih sistemov.

Gre za svojevrsten pristop, ki se v mnogih stvareh razlikuje od dosedanje prakse obvladovanja podatkov. Ena glavnih sprememb je zgoraj omenjena širitev infrastrukture v širino. Pri tem konceptu gre predvsem za to, da se problemov ne lotevamo z dražjo opremo, temveč z večjim številom strežnikov, pri katerih so stroški nižji. Koncept zagovarja rešitev brez centralnega diskovnega sistema in le-to nadomešča z diski znotraj strežnikov. Nadzorni strežnik skrbi, da se računske operacije izvajajo na podatkovnih strežnikih, ne da bi bilo potrebno podatke seliti k sistemom, ki zagotavljajo procesorsko moč. Sistem sam skrbi za varovanje in potrebne replikacije blokov s podatki. V primeru okvar na strojni opremi se celoten sklop okvarjene opreme enostavno samo zamenja, kar na nemoteno delovanje strežnika ne vpliva.

Tri pomembna področja uporabe, ki gredo preko meja klasične uporabe infrastrukture in predstavljajo temelje pri obvladovanju velikih količin podatkov poimenujemo s principom ABC (ang. Analytics Bandwidth Content), kar prikazuje slika 4:

- **Analitika (Analytics)** - Analitika na velikih množicah podatkov daje možnost oz. priložnost pridobivanja dodatnih informacij iz nabora strukturiranih in nestrukturiranih podatkov in le to prevesti v koristne informacije ter na ta način pridobiti dodano vrednost za boljše poslovne odločitve.
- **Pasovna širina (Bandwidth)** - Kvalitetno visoko propustno omrežje za podatkovno intenzivne operacije pri velikih hitrostih.
- **Vsebina (Content)** - Skoraj brezmejno podatkovno polje za pisanje in trajno hranjenje podatkov.

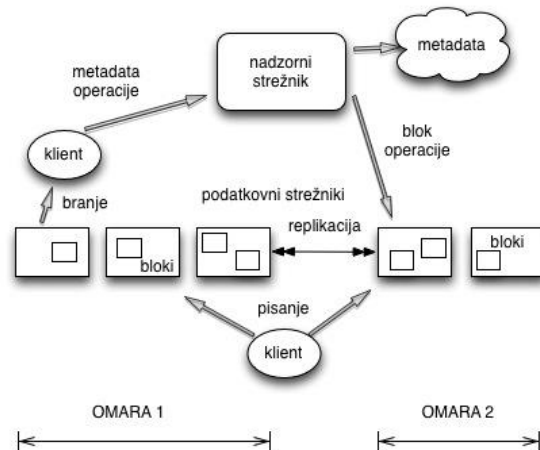


Slika 4: Ključni dejavniki, ki jih moramo upoštevati pri obvladovanju večjih količina podatkov

Tehnoloških rešitev, ki so se pojavile je več, vendar tehnologija na osnovi platforme Hadoop predstavlja eno glavnih področij, ki se uveljavljajo v svetu, med

drugim tudi na področju obvladovanja podatkov, ki jih ustvarijo naprave v pametnih omrežjih – SmartGrids. Predstavlja platformo, ki sodi v nabor NoSQL tehnologij, ki so trenutno v fazi množičnih implementacij. Gre za t.i. Apache Software library framework, ki omogoča distribuirano procesiranje ogromnih količin podatkov preko distribuirane gruče strežnikov nižjega cenovnega razreda. V nasprotju s klasičnimi RDBMS sistemi v tem konceptu ni centraliziranega podatkovnega skladišča, saj Hadoop koristi diskovne enote, ki jih vsebuje vsak strežnik. Diskovni sistemi so preko internih mehanizmov združeni v logične enote.

Porazdeljen datotečni sistem (HDFS) je torej primarni datotečni prostor za hranjenje podatkov, ki ga uporabljajo vse nadrejene aplikacije. Slika 5 shematsko prikazuje HDFS, kjer je vidna arhitektura vse do bloka na diskovnem pomnilniku.



Slika 5: Arhitektura HDFS

Porazdeljen datotečni sistem sam skrbi za varovanje podatkov v primeru odpovedi strežniških elementov v sistemu, saj ima vgrajene mehanizme replikacij. Sistemsko se ugotovljajo težave in jih vkolikor je to možno tudi ustrezno odpravlja. Predvsem pa je zagotovljeno, da je vedno na voljo predpisano število varnostno kopiranih blokov.

Velikosti podatkov v okviru arhitekture HDFS se gibljejo od zelo majhnih pa vse do terabajtov ali petabajtov. Velik del strukturiranih, delno strukturiranih in nestrukturiranih podatkov se na podatkovni ravni HDFS, poleg skladiščenja, tudi obdeluje, agregira, itd.

Za operacije nad podatki v arhitekturi HDFS se uporablja razvojno okolje MapReduce. MapReduce je v bistvu programski model, ki je namenjen procesiranju večjih količin podatkov. Model oziroma koncept je začel razvijati Google. Največkrat se MapReduce uporablja za distribuirane računske

operacije na gruči strežnikov. Model oziroma koncept je črpal ideje na podlagi map in funkcionalnosti reduciranja odvečnih podatkov, ki so običajno uporabljene pri funkcionalnem programiranju (ang. functional programming).

Celoten koncept temelji na tezi, da je ceneje in lažje pošiljati računske operacije k podatkom, kot da bi se podatke pošiljalo proti računskim centrom v smislu CPU (central processing unit).

V primeru porazdeljenega datotečnega sistema HDFS se centralni nadzorni strežnik sam odloči, kam bo poslal računske operacije. Pri tem mu pomagajo procesi kot so t.i. "task" in "job tracker", posamezne opravilne vrste, itd. Računanje se nato opravi na oddaljenih podatkovnih strežnikih lokalno, ob uporabi lokalne procesorske moči, kar močno poenostavi sistem in optimizira stroške.

IV. ZAKLJUČKI

V praksi smo že pričali porastu zbrane količine podatkov iz pametnih omrežij. Pri tem gre za svojevrsten paradoks, saj je primarna osredotočenost na širitev IKT infrastrukture čedalje bolj v smeri končnega odjemalca v elektroenergetskem omrežju. Takoj, ko se to zgodi pridobimo možnost zbiranja različnih podatkov, kar pa ne bomo več mogli reševati na dosedaj uveljavljene principe, ki veljajo v IT oddelkih elektroenergetskih podjetij. Dejstvo je, da informatizacija omrežja prinaša veliko novih možnosti, predvsem v smislu pridobivanja novih uporabnih informacij iz zbrane količine podatkov, ki lahko pomembno vplivajo na poslovne modele in lahko podjetjem omogočijo razvoj bodisi novih storitev za potrebe trženja, bodisi za interne procese, kjer je potrebno na podlagi analiz uvajati izboljšave v delovanje elektroenergetskega sistema. Konkretno vsebine, ki jih je možno preliti v nove storitve bodo premet razvoja v okviru posameznih podjetij, vendar se dotikajo tako naprednih sistemov merjenja AMI, upravljanja s porabniki DSM, kot analitike nad sredstvi (ang. asset analytics), analitike procesov v pametnem omrežju (ang. grid analytics), kot so spremljanje dogodkov v realnem času, njihova korelacija, itd. Vpliv razpršenih virov na delovanje elektrodistribucijskega omrežja zaradi dinamične narave proizvodnje energije npr. v fotovoltaičnih elektrarnah je že sedaj prisoten, vendar ni na voljo orodij, ki bi v prvi vrsti nudile monitoring teh vplivov v realnem času. Vkolikor to postane potrebno takoj nastopi zajem in hranjenje velike količine podatkov, s potrebo po dodatnem upravljanju s temi podatki.

Na voljo so tehnološke rešitve, ki se množično uveljavljajo, na drugi strani pa bo vedno več potreb po obvladovanju velike količine podatkov, pri čemer je možno uporabiti v članku opisane rešitve.

REFERENCE

- [1] <http://www.smartgrids.eu/>, SmartGrids.eu. Smart Grids, junij 2012
- [2] Pike Research: Smart Grid Data Analytics, 2012