

Pristupi za dobijanje dimenziono modelovanih podataka iz bežičnih senzorskih mreža

Aleksandar M. Gavrić, dr Dušan M. Vujošević

Sadržaj — U radu su prikazane primene tehnike dimenzionog modelovanja, te konceptata dejta marta i skladišta podataka nad podacima iz bežičnih senzorskih mreža. Dimenzione strukture podataka namenjene su uspostavljanju efikasnog sistema za podršku odlučivanju i analizu podataka iz bežičnih senzorskih mreža. Dimenzioni modeli se formiraju iz ugla senzora i iz ugla misije. Opisuju se koraci transformacije podataka na putu od operacione baze do dejta mart šeme. Do sada u literaturi nije bilo moguće naći primere dimenzionih modela podataka iz bežičnih senzorskih mreža.

Cljučne reči — bežične senzorske mreže, dejta mart, integracija podataka, zvezdasta šema, skladište podataka

I. UVOD

BEŽIČNE SENZORSKE MREŽE generišu velike podatke. Potrebno je uložiti dosta truda da bi se omogućilo izvođenje sveobuhvatnih analiza tih podataka. Analize treba da se izvode u cilju izvlačenja zaključaka od interesa. U praksi sa bežičnim senzorskim mrežama često nailazimo na aplikacije gde je neophodno doneti zaključke o podacima u kratkom vremenu u odnosu na period u kome su podaci generisani, a pritom se podaci sa senzora neprekidno prikupljaju i akumuliraju. Ovaj rad se hvata u koštac s navedenim izazovom i predlaže nekoliko odgovora na njega primenama tehnike dimenzionog modelovanja podataka i tehnologija ekstrakcije, transformacije i punjenja tokova podataka, kroz transformacije od operacione baze do zvezdaste šeme.

Aleksandar M. Gavrić, Računarski fakultet, Beograd, Srbija (e-mail: agavric2020m@raf.rs).
prof. dr Dušan M. Vujošević, Računarski fakultet, Beograd, Srbija
(e-mail: dvujosevic@raf.rs).

II. PREGLED KONCEPATA

Bežične senzorske mreže (eng. Wireless Sensor Networks, WSN) odnose se na mreže prostorno rasutih, namenskih senzora koji nadgledaju i beleže fizičke uslove fenomena iz okoline i prikupljaju podatke u centralnoj stanici. Ovakvi senzori često mere uslove okoline kao što su temperatura, zvuk, nivo zagađenja, vlažnost, karakteristike vetra i slično [1][2].

WSN su slične bežičnim ad hoc mrežama u smislu da se oslanjaju na bežično povezivanje i spontano formiranje mreža tako da se podaci senzora mogu bežično prenositi. Savremene mreže su dvosmerne – prikupljaju podatke i omogućavaju kontrolu aktivnosti senzora [3][4]. Razvoj ovih mreža bio je motivisan vojnim aplikacijama poput nadzora bojnog polja [5].

WSN su izgrađene od „čvorova“ - od par njih do nekoliko stotina ili hiljada, pri čemu je svaki čvor povezan sa drugim sensorima. Svaki takav čvor obično ima nekoliko delova: radio primopredajnik sa unutrašnjom antenom ili vezom sa spoljnom antenom, mikrokontroler, elektronsko kolo za povezivanje sa sensorima i izvorom energije, obično baterijom ili ugrađenim oblikom prikupljanja energije. Čvor senzora može se po veličini kretati od ormana do (teoretski) zrna prašine, iako mikroskopske dimenzije tek treba da se ostvare. Ograničenja veličine i troškova ograničavaju resurse kao što su energija, memorija, računaska brzina i propusni opseg komunikacije. Topologija WSN-a može varirati od jednostavne zvezdaste mreže do napredne multi-hop bežične mreže. Algoritmi za propagaciju podataka kroz mrežu mogu koristiti rutiranje ili poplavu [6][7].

Dimenziono modelovanje (DM) deo je metodologije poslovnog dimenzijskog životnog ciklusa koji je razvio Ralf Kimbal, a koji uključuje skup metoda, tehnika i koncepata za upotrebu u dizajnu skladišta podataka [8][9]. Pristup se fokusira na identifikovanje ključnih poslovnih procesa, koji se modeluju prvo, odnosno pre dodavanja dodatnih poslovnih procesa, tj. pristupom odozdo prema gore [8].

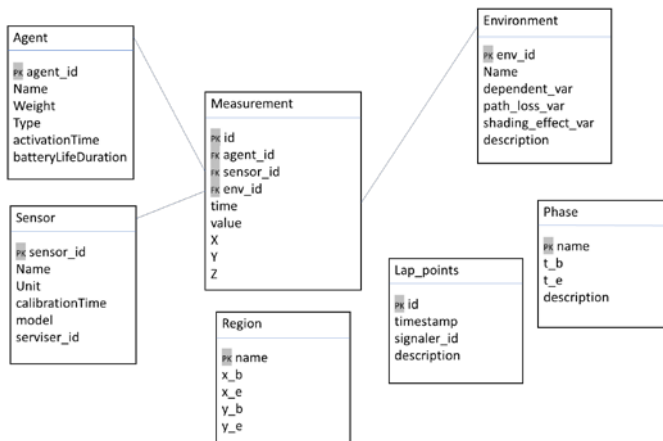
Dimenziono modelovanje uvek koristi koncepte fakata iliti činjenica (mere) i dimenzija (kontekst). Činjenice su često numeričke vrednosti koje se mogu agregirati, a dimenzije su grupe hijerarhija i deskriptora koji definišu činjenice. Dimenzioni modeli se grade prema području poslovnog procesa. Budući da različita područja poslovnog procesa dele neke, ali ne sve dimenzije, efikasnost u dizajnu, radu i doslednosti postiže se korišćenjem usklađenih dimenzija, tj. korišćenjem jedne kopije zajedničke dimenzije u predmetnim oblastima. Dimenziono modelovanje ne uključuje nužno implementaciju relacione baze podataka. Isti pristup modelovanju, na

logičkom nivou, to jest nivou korisničkog interfejsa, može se koristiti za bilo koji fizički oblik podataka, poput višedimenzione strukture podataka ili čak ravnih datoteka. Orijentisano je na razumljivost skladištenih podataka i performanse. Dimenzioni model je izgrađen na šemama nalik zvezdi i pahuljici, sa dimenzijama koje okružuju tabelu činjenica [9][10].

Dve zvezdaste šeme koje dele makar jednu tabelu čine skladište podataka. U računarstvu, skladište podataka (eng. Data Warehouse, DW ili DWH), poznato i kao skladište podataka preduzeća (EDW), jeste sistem koji se koristi za izveštavanje i analizu podataka i smatra se ključnom komponentom poslovne inteligencije. Centralno je spremište integrisanih podataka iz jednog ili više različitih izvora [11]. Čuva trenutne i podatke iz prošlosti na jednom mestu, a ti podaci se kasnije koriste za izradu analitičkih izveštaja. Podaci mogu proći kroz operativno skladište podataka i može biti potrebno čišćenje podataka za dodatne operacije kako bi se osigurao kvalitet podataka pre nego što se u DW koriste za izveštavanje. Izdvajanje, transformisanje, učitavanje (ETL) i izdvajanje, učitavanje, transformisanje (ELT) su dva glavna pristupa koja se koriste za izgradnju sistema skladišta podataka [12][13].

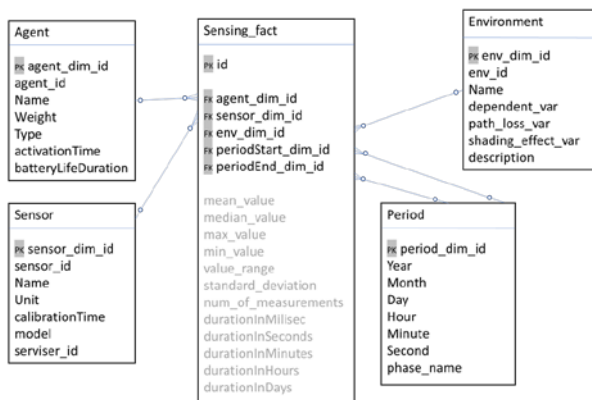
III. METODOLOGIJA

Za projektovanje dimenzionih modela koji skladište podatke iz bežičnih senzorskih mreža koristićemo dva pristupa. Prvi pristup je kroz strategiju iz ugla merenja sa senzora a drugi kroz strategiju iz ugla misije koja je krajnji cilj sistema kao celine.

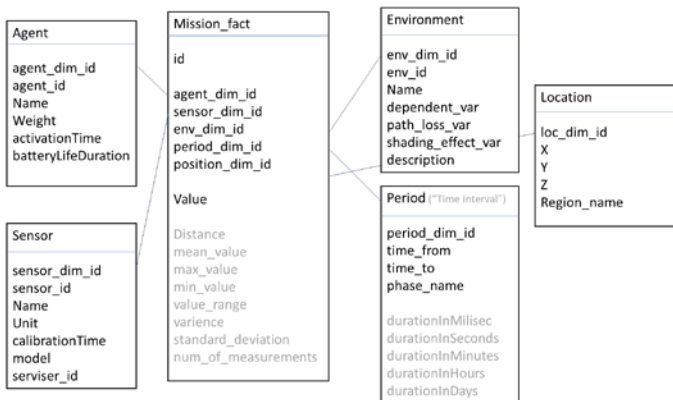


Slika 1 – šema operacione (OLTP) baze podataka

Na Slici 1 je prikazana jedna moguća operaciona ili osnovna (transakciona, OLTP) baza podataka sistema bežičnih senzorskih mreža kojim se evidentira geografski prostor podeljen na regione, vremenski prostor podeljen na faze i slično. Primer je namenjen ovom radu, pa je jednostavniji nego što bi se očekivalo u realnim okolnostima. Slike 2 i 3 prikazuju projektovane šeme tabela činjenica kroz dve različite strategije za istu bežičnu mrežu o čemu će se diskutovati u narednom odeljku.



Slika 2 – šema strategije dimenzionog modelovanja iz ugla merenja



Slika 3 – šema strategije dimenzionog modelovanja iz ugla misije

U ovom odeljku se diskutuje kreiranje dejta marta (dimenzionog modela) za procese sistema oko bežične senzorske mreže koja je sačinjena od više pokretnih čvorova (agenata) koji su nosioci senzora. Modluje se proces skupljanja merenja sa različitih agenata u različitim koordinatama trodimenzionog koordinatnog sistema kroz vreme. Na sensorima se meri snaga određenog radio signala i na osnovu saradnje agenata koji vrše merenja u svojim lokacijama pa je potom moguće da se identifikuje gde se u koordinatnom sistemu nalazi odašiljač tog određenog radio signala. Geografski prostor je podeljen u regione a vreme je podeljeno na faze. Razlikujemo različite osobine prirodnog okruženja koje utiču na naša merenja kroz faktore uticaje prirode (eng. Enviroment factors). Implementirana je skripta koja generiše podatke koji se upisuju u MySQL bazu podataka koja se smatra OLTP bazom podataka. Generatorska skripta pri svom pozivu očekuje GET parametar “samples” koji predstavlja broj simuliranih merenja koja se dodaju bazi podataka.

Primer poziva generatorske skripte: *generate_data.php?samples=1000*.

Inicijalno je potrebno kreirati tabele u bazi podataka što se može postići izvršavanjem SQL naredba koje su date u nastavku. Generatorska skripta očekuje da u bazi podataka postoji korisnik 'root' čija je šifra prazno polje.

A. Inicijalizacija lokalne operacione baze podataka (OLTP baze)

Potrebno je izvršiti naredbu za kreiranje tabele Sensor u bazi wsn_oltp sledećom naredbom.

```
CREATE TABLE `wsn_oltp`.`Sensor` ( `id` INT NOT NULL , `name` VARCHAR(30) NOT NULL , `unit` VARCHAR(10) NOT NULL , PRIMARY KEY (`id`)) ENGINE = MyISAM;
```

Postavljamo indeks tabele da bude jedanaestocifreni celobrojni broj i da se prilikom upisa vrši autoinkrementiranje prethodno upisanog indeksa ukoliko nije drugačije naznačeno naredbom upisa.

```
ALTER TABLE `Sensor` CHANGE `id` `id` INT(11) NOT NULL AUTO_INCREMENT;
```

Sledećom naredbom dodajemo redove u tabeli senzora gde proizvoljno odabranim vrednostima simuliramo zapise o sensorima koji mere:

- Snagu radio signala u dBm jedinici

- Količinu energije u bateriji merene u procentima od ukupnog kapaciteta
- Ugao pod kojim se meri radio signal u radijanima
- Temperaturu u stepenima Celzijusa
- Pritisak u paskalima.

```
INSERT INTO `Sensor` (`name`, `unit`) VALUES ('received power', 'dBm'), ('battery life', '%'), ('received angle', 'rad'), ('temperature', 'deg C'), ('pressure', 'Pa');
```

Kreiramo tabelu Agent koja predstavlja zapise o mobilnim senzorskim čvorovima.

```
CREATE TABLE `wsn_oltp`.`Agent` ( `id` INT NOT NULL AUTO_INCREMENT, `name` VARCHAR(30) NOT NULL, `type` VARCHAR(30) NOT NULL, `weight` DECIMAL NULL, `activationDateTime` DATETIME NULL, `batteryLifeDuration` INT NULL, PRIMARY KEY (`id`)) ENGINE = MyISAM;
```

Dodajemo zapise u tabelu sa proizvoljno odabranim vrednostima pri čemu simuliramo čvorove odnosno nosioce senzora kao makro uređaje sa kontrolnom logikom nad više senzora. Proizvoljno smo odabrali da simuliramo dva mobilna telefona (Android i iPhone tipa), laptop i 3 drona.

```
INSERT INTO `Agent` (`id`, `name`, `type`, `weight`, `activationDateTime`, `batteryLifeDuration`) VALUES (NULL, 'DJI1', 'dron', '1120', '2021-04-27 05:53:22', '24'), (NULL, 'DJI2', 'dron', '1125', '2021-04-27 05:53:22', '24'), (NULL, 'DJI3', 'dron', '1120', '2021-04-27 05:53:22', '18'), (NULL, 'Android1', 'phone', '380', '2021-04-27 05:54:22', '510'), (NULL, 'iPhone1', 'phone', '380', '2021-04-27 05:53:22', '480'), (NULL, 'LaptopHP', 'pc', '3200', '2021-04-27 05:53:25', '590');
```

Kreiramo tabelu koja skladišti zapise o karakteristikama prirodnog okruženja u kome nam se nalaze senzori individualno.

```
CREATE TABLE `wsn_oltp`.`Environment` ( `id` INT NOT NULL AUTO_INCREMENT, `name` VARCHAR(30) NOT NULL, `env_dependent_var` DECIMAL NOT NULL, `path_loss` DECIMAL NOT NULL, `shading_effect` DECIMAL NOT NULL, PRIMARY KEY (`id`)) ENGINE = MyISAM;
```

Dodajemo proizvoljne zapise o prirodnim okruženjima koje smo označili kao sobe i bašte. Fizičke veličine koje nam definišu okruženje su efekat senke, koeficijent gubljenja signala i veličina koja sumarizuje sve ostale uticaje na signal kao šum koji se dodaje na vrednost signala i ima Gausovu raspodelu greške u odnosu na realnu vrednost jačine signala.

```
INSERT INTO `Environment` (`id`, `name`, `env_dependent_var`, `path_loss`, `shading_effect`)
VALUES (NULL, 'garden1', '0.684', '0.70', '0.4'), (NULL, 'room8', '0.843', '0.18', '0.11'), (NULL,
'stadion', '0.484', '0.54', '0.34'), (NULL, 'garden2', '0.612', '0.81', '0.48');
```

Sledećom naredbom se kreira tabela za regione na koje smo podelili geografski prostor.

```
CREATE TABLE `wsn_oltp`.`region` ( `name` VARCHAR(30) NOT NULL, `x_b` INT NOT NULL,
`x_e` INT NOT NULL, `y_b` INT NOT NULL, `y_e` INT NOT NULL, PRIMARY KEY (`name`))
ENGINE = MyISAM;
```

Proizvoljno delimo geografski prostor na pravougaone regione koji su definisani sa `x_b`, `x_e`, `y_b`, `y_e` parametrima koji predstavljaju:

- Početnu koordinatu regiona po X osi
- Krajnju koordinatu regiona po X osi
- Početnu koordinatu regiona po Y osi
- Krajnju koordinatu regiona po Y osi

```
INSERT INTO `region` (`name`, `x_b`, `x_e`, `y_b`, `y_e`) VALUES ('Zone 1', '0', '300', '0', '500'),
('Zone 2', '0', '300', '500', '1000'), ('Zone 3', '300', '1000', '0', '200'), ('Zone 4', '300', '1000', '200',
'1000');
```

Kreiramo tabelu za faze na koje delimo vremensku dimenziju.

```
CREATE TABLE `wsn_oltp`.`Phase` ( `name` VARCHAR(30) NOT NULL, `t_b` INT NOT NULL, `t_e`
INT NOT NULL, `description` TEXT NOT NULL, PRIMARY KEY (`name`)) ENGINE = MyISAM;
```

Proizvoljno dodajemo faze koje su definisane trenutkom početka i trenutka završetka faze. Moguće je da se faze preklapaju. Primeri faza su faza kalibracije i slično.

```
INSERT INTO `Phase` (`name`, `t_b`, `t_e`, `description`) VALUES ('search', '0', '13', 'search
phase'), ('learn', '14', '39', 'learn phase'), ('listen', '40', '73', 'listen phase'), ('calibration', '74',
'100', 'calibration phase');
```

Kreiramo tabelu za merenja sa senzora i definišemo da je za sve zapise o merenjima zajedničko da su određeni tačkom u geografskom prostoru gde su izmereni, kroz x, y i z geografsku koordinatu, te da su definisani vremenskim trenutkom u kome su izmereni.

```
CREATE TABLE `wsn_oltp`.`Measurement` ( `id` INT NOT NULL AUTO_INCREMENT, `x` INT NOT
NULL, `y` INT NOT NULL, `z` INT NOT NULL, `time` TIMESTAMP NOT NULL, `time_milliseconds`
```

```
INT NOT NULL, `sensor_id` INT NOT NULL, `agent_id` INT NOT NULL, `env_id` INT NOT NULL,
`value` DECIMAL NOT NULL, PRIMARY KEY (`id`)) ENGINE = MyISAM;
```

Za svako merenje pamtimo na kom tipu senzora je izvršeno, na kom agentu je bio taj konkretan senzor i u kom prirodnom okruženju se nalazio agent u trenutku merenja.

```
ALTER TABLE measurement ADD FOREIGN KEY (sensor_id) REFERENCES sensor(id);
ALTER TABLE measurement ADD FOREIGN KEY (agent_id) REFERENCES agent(id);
ALTER TABLE measurement ADD FOREIGN KEY (env_id) REFERENCES environment(id);
```

Kreiramo tabelu u kojoj ćemo skladištiti informacije o realnim vrednostima koje definišu fenomen koji posmatramo što je u našem primeru odašiljač radio signala. Na osnovu ovih vrednosti možemo naknadno evaluirati uspešnost našeg sistema u prediktovanju ovih karakteristika.

```
CREATE TABLE `wsn_oltp`.`target` ( `x` INT NOT NULL, `y` INT NOT NULL, `z` INT NOT NULL,
`time` TIMESTAMP NOT NULL, `time_milliseconds` INT NOT NULL, `id` INT NOT NULL
AUTO_INCREMENT, PRIMARY KEY (`id`)) ENGINE = MyISAM;
```

U svrhu ilustracije se prilaže naredba kojom se dodaje jedan zapis o merenju u bazu.

```
INSERT INTO `measurement` (`id`, `x`, `y`, `z`, `time`, `sensor_id`, `agent_id`, `env_id`, `value`)
VALUES (NULL, '3', '3', '3', '2021-04-27 19:33:22', '1', '1', '1', '3189');
```

Generisanje narebi za popunjavanje tabele o merenjima se vrši simulatorom koji je implementiran kroz skriptu.

B. Skripta za generisanje simuliranih podataka sa senzora

Skripta je implementirana kao PHP skripta koja ostvari konekciju sa bazom podataka. Kreira nasumične podatke koji po formi odgovaraju podacima koje bi kreirala mreža koju opisujemo.

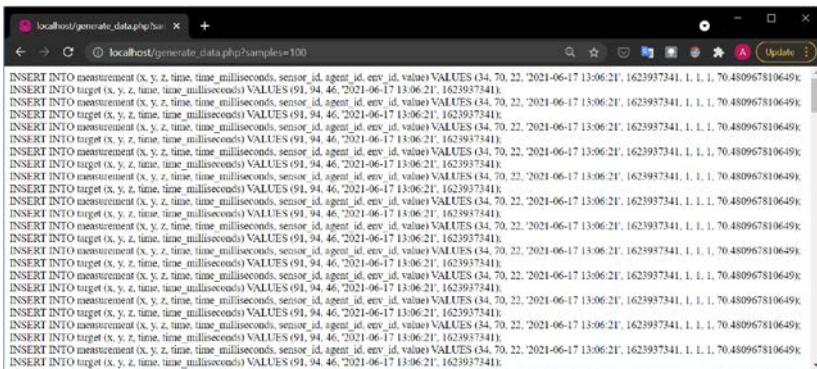
Implementirana je funkcija koja ima za cilj da samo povremeno za nijansu promeni vrednost koju meri senzor kako bi simulirala retkost i postepenost u promeni karakteristika fenomena koji se opservira.

<pre>function promenaVrednosti (staraVrednost, prag) if (rand(0, 100) > prag return staraVrednost + 5 * ((rand(0,1) == 0)?1:(-1)) else</pre>

return staraVrednost

Fukcija simulira nove vrednosti koje generišu senzori. Ukoliko parametar za kontrolu retkosti promene (parametar *prag*) postavimo na vrednost 100 u tom slučaju se nikad neće promeniti vrednost koju meri senzor i sva nova merenja će imati iste vrednosti kao stara merenja. Ukoliko parametar za kontrolu retkosti promene postavimo na vrednost 0 u tom slučaju će se prilikom svakog merenja menjati vrednost u odnosu na onu prethodno izmerenu. Podešavanjem parametra za kontrolu retkosti promene na vrednost od 0 do 100 možemo podesiti stabilnost odnosno varijabilnost vrednosti koje simuliramo da su izmerene na senzorima.

Na Slici 3 su prikazane neke od SQL naredbi koje su izvršene nad bazom podataka kada izvršimo skriptu sa parametrom `samples = 100`.

The image shows a browser window at the URL `localhost/generate_data.php?samples=100`. The page content consists of a long list of SQL `INSERT INTO` statements. Each statement follows a similar pattern: `INSERT INTO measurement (x, y, z, time, time_milliseconds, sensor_id, agent_id, env_id, value) VALUES (34, 70, 22, '2021-06-17 13:06:21', 1623937341, 1, 1, 1, 70, 480967810649);`. The values for `x, y, z` and `time` vary slightly, while `time_milliseconds` is constant at 91, 94, or 46. The `sensor_id` is 91, 94, or 46. The `agent_id` is 91, 94, or 46. The `env_id` is 91, 94, or 46. The `value` is 34, 70, or 22. The `VALUES` list is followed by a closing semicolon. The statements are repeated multiple times, representing the data generated for 100 samples.

Slika 3 – neke od SQL naredbi koje su izvršene nad bazom podataka kada izvršimo generatorski skriptu sa parametrom `samples = 100`.

U okviru dalje diskusije kreiranja dimenzionog modela, biće prikazani svi koraci pri kreiranju transformacija za definisanje dimenzija zvezdaste šeme na osnovu informacija koje se pribavljaju iz baze podataka. Takođe, biće prikazano definisanje dimenzije na osnovu novokreiranih podataka, kao i proces definisanja table činjenica zvezdaste šeme.

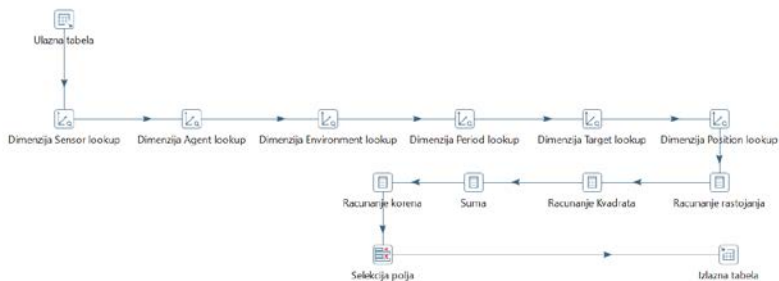
U okviru ovog demonstrativnog uputstva i diskusije će se koristiti alati:

- Pentaho Data Integration alat za kreiranje ETL procesa,

- Baza podataka postavljena na lokalnom serveru kao izvor podataka
- Pentaho Business Analytics Server za smeštanje dejta marta.

C. Tok podataka za transformaciju tabele činjenica o merenjima

Transformacija je mreža logičkih zadataka koji se nazivaju koraci ili komponente. Transformacije su u osnovi tokovi podataka. Transformacije čine, u suštini, usmereni graf logičkog skupa konfiguracija transformacije podataka. Na Slici 4 je prikazana transformacija podataka koja kreira tabelu činjenica za merenja.



Slika 4 – Transformacija podataka koja kreira tabelu fakata za merenja

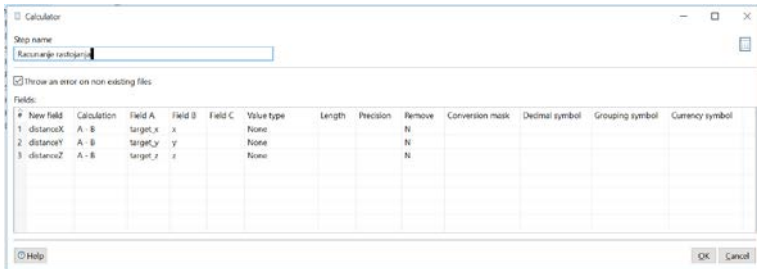
U nastavku odeljka će se opisati svaki od koraka kroz koji prolaze podaci na putu kroz transformacije. Prvi korak odnosno komponenta transformacije je Ulazna tabela koja izvršava naredbu `SELECT * FROM MEASUREMENTS`.

Sensor lookup korak koristi komponentu za pretraživanja / ažuriranja dimenzija. Korak pretraživanja / ažuriranja dimenzija nam omogućava da primenimo sporo promenljivu dimenziju koju definiše Ralf Kimbal za oba tipa: Tip I (ažuriranje) i Tip II (umetak) zajedno sa nekim dodatnim funkcijama. Ne samo da ovaj korak može da ažurira tabelu dimenzija, već takođe se može koristiti za traženje vrednosti u dimenziji.

Sledećih četiri koraka u toku podataka su primene komponente Kalkulator. Korak kalkulator pruža nam unapred definisane funkcije koje se mogu izvršiti na vrednostima polja za unos. Pored argumenata (Polje A, Polje B i Polje C) mora se navesti i tip povratne vrednosti funkcije. Takođe se može odabrati da

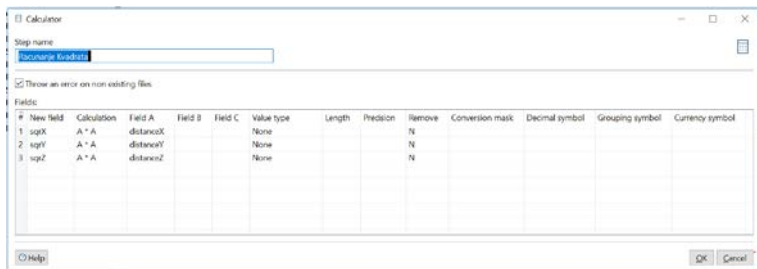
se ukloni polje iz rezultata (izlaza) nakon što se izračunaju sve vrednosti; ovo je korisno za uklanjanje privremenih vrednosti.

Koristimo četiri koraka računanja čija su podešavanja prikazana na slikama 5 – 8 kako bi u postupku transformacije podataka izračunali rastojanje između agenata koji je nosilac senzora i odašiljača radio signala čija se snaga meri na sensorima.

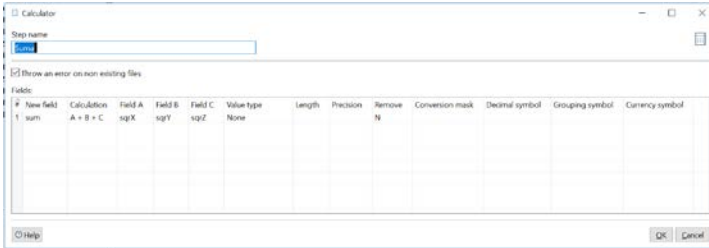


Slika 5 - Podešavanje koraka računanja razlika koordinata senzora i obzerviranog fenomena.

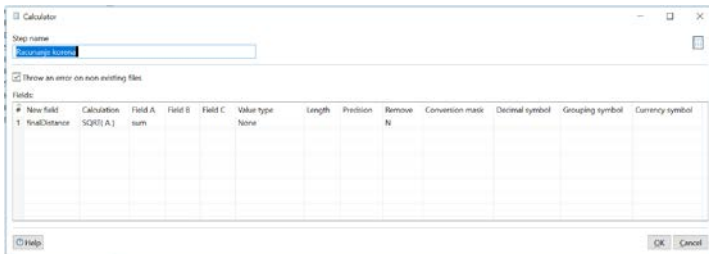
Prvenstveno računamo razliku koordinata između koordinata po svakoj od osa koordinatnog geografskog sistema. Potom računamo zbir kvadrata tih razlika rastojanja jer koren tog rezultata predstavlja rastojanje od agenata koji nosi senzor do odašiljača signala koji senzor obzervira.



Slika 6 - Podešavanje koraka računanja kvadrata razlika koordinata senzora i opserviranog fenomena.



Slika 7 - Podešavanje koraka računanja razlika koordinata senzora i opserviranog fenomena.

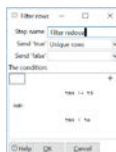


Slika 8 - Podešavanje koraka računanja korena prethodno prosleđenog rezultata kako bi se izračunalo finalno rastojanje

Korak Izlazna tabela predstavlja poslednji korak u transformaciji podatka o merenjima. Kao konekciju koristi „dm konekciju“, što označava da svoj rezultat upisuje u bazu za dejta mart, a ne u operacionu OLTP bazu podataka od koje smo počeli. Ovim korakom je obrada podataka gotova i popunjava se dejta mart baza odgovarajućim transformisanim podacima.

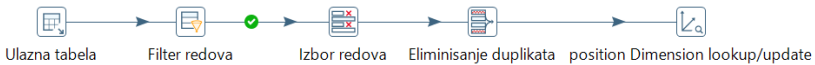
D. Tok podataka za transformaciju tabele dimenzija faza i regiona

U transformacija podatka za dimenziju perioda, podaci se iz ulazne tabele pribavljaju naredbom `SELECT * FROM measurement CROSS JOIN phase`. Potom se selektuju redovi od interesa i nad njima se primenjuje filter čija je konfiguracija prikazana na Slici 9.

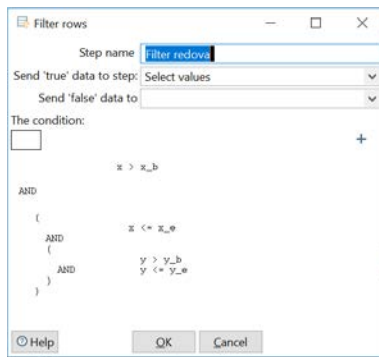


Slika 9 – podešavanje koraka za filtriranje redova koji ispunjavaju zadat uslov

Transformacija podatka za dimenziju pozicije je priložena na Slici 10. Podaci se iz ulazne tabele pribavljaju naredbom `SELECT * FROM measurement CROSS JOIN regions`. Potom se selektuju redovi od interesa i nad njima se primenjuje filter čija je konfiguracija prikazana na Slici 11.



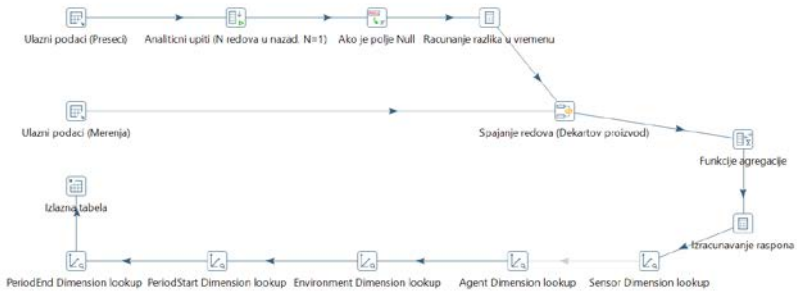
Slika 10 – Transformacija podataka koja kreira tabelu dimenzije za poziciju



Slika 11 – podešavanje koraka za filtriranje redova koji ispunjavaju zadat uslov

E. Tok podataka za transformaciju podataka za agregacije i sumarizacije

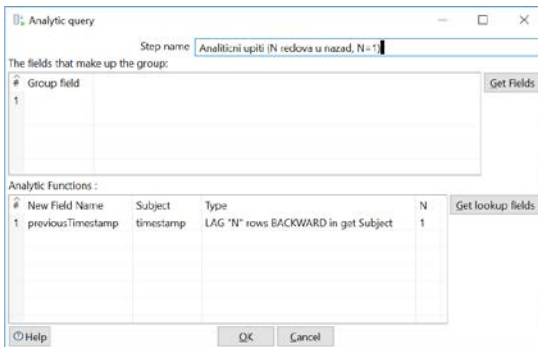
Usled potencijalnih analiza koje će se sprovoditi nad skladištenim podacima, od značaja nam je da pripremimo sumarizovane verzije podataka kako ne bismo morali iznova da ih računamo u postupku traženja smisla u tim podacima. Sumarizacija podataka primenom funkcija agregacija nad podacima olakšava vizualizaciju podataka i kasnije izveštavanje odnosno pravljenje izveštaja na osnovu upita nad podacima. U sledećem primeru ćemo sumarizovati podatke od merenja na sensorima tokom jednog ograničenog vremenskog perioda. Transformacija podataka će podrazumevati grupisanje merenja tokom jednog ograničenog vremenskog perioda i primenu funkcija agregacija nad tim grupama. Tok podataka za ovu transformaciju je prikazan na Slici 12.



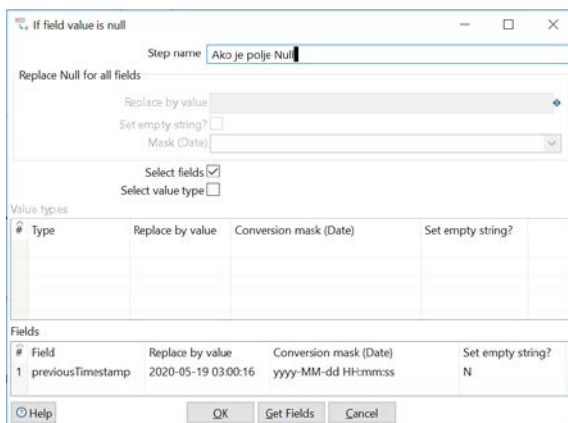
Slika 12 – tok podataka za transformaciju podataka za agregacije i sumarizacije

Korak Ulazni podaci (preseci, pritanje, kapanje od eng. *Lap points*) obezbeđuje trenutke u kojima smo simulirali presečne tačke u vremenu, koje treba da nam služe kao momenti koji razdvajaju vremenske periode koje na taj način definišemo. Svaki presek osim prvog i poslednjeg označava kraj jednog perioda i početak narednog. Prvi presek podrazumeva početak prvog perioda a poslednji presek kraj poslednjeg perioda.

Kako bismo nastavili transformaciju podataka do oblika pogodnog za agregaciju, koristimo komponentu Analitični upiti u koraku gde želimo podacima da dodamo kolonu koja će takođe prikazivati redom preseke koje smo simulirali ali sa bitnom razlikom da je svaki od tih redova uparen sa odgovarajućim redom iz kolone preseka koji je na indeksu za N manji od tekućeg reda novogenerisane kolone. Za uzetu vrednost $N = 1$ postizemo da se u jednom redu tabele nalaze uzastopni preseci u dve različite kolone, pri čemu je izuzetak prvi red koji nema inicijalnu vrednost za svog prethodnika. Vrednost preseka prethodnika u prvom redu upisujemo u koraku čije je podešavanje prikazano na Slici 14 dok je podešavanje analitičnog upita prikazano na Slici 13.

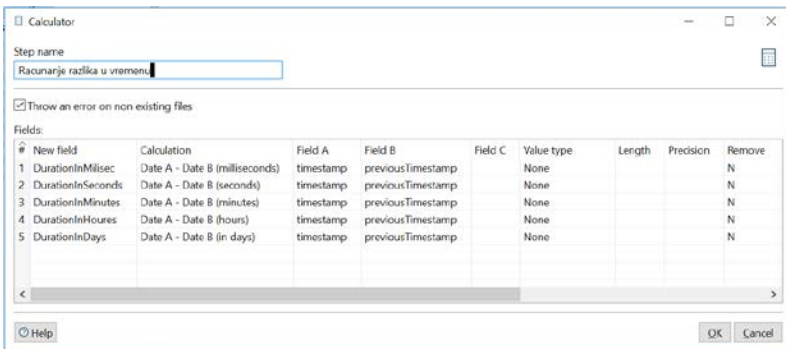


Slika 13 – kreiranje kolone vremeskih presečnih tačaka u svojstvu prethodnika u odnosu na presečnu tačku u tekućem redu



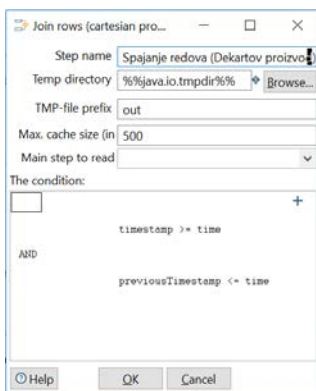
Slika 14 – podešavanje koraka za ispunu vrednosti prvog reda kolone prethodnika vremeskih preseka

Sada kada u jednom redu imamo informaciju o vremenskom trenutku početka i kraja vremenskog perioda koji smo ograničili tim vrednostima, može se izračunati dužina trajanja tog vremenskog perioda u različitim jedinicama kao što su minuti, sekunde, milisekunde i slično. Korak računanja dužine trajanja vremeskih perioda je prikazan na Slici 15.



Slika 15 – podešavanje koraka računanja dužine trajanja vremenskog perioda ograničenog sa dva uzastopna vremenska preseka

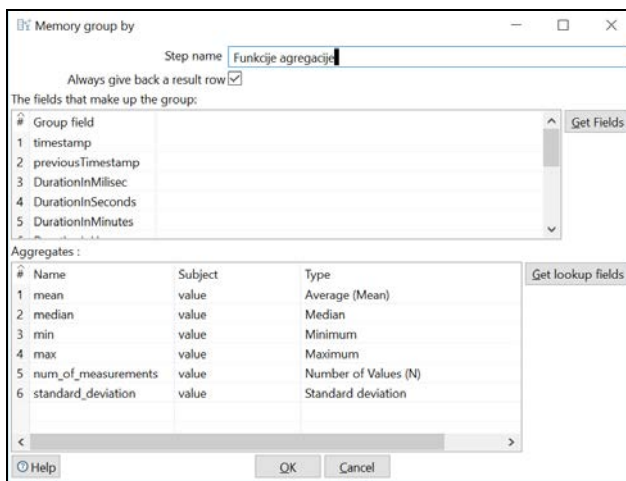
Iz druge komponente za Ulaznu tabelu pribavljamo podatke naredbom `SELECT id, time, sensor_id, agent_id, env_id, value FROM measurement`. Sledeći korak je Spajanje redova korišćenjem Dekartovog proizvoda nad dva izvora podataka koja su pripremljeni. Podešavanje koraka spajanja redova je prikazano na Slici 16, pri čemu je rezultat ovog koraka tabela sa merenjima sa senzora koji su pridruženi odgovarajućem periodu koji je u ranijim koracima izračunat.



Slika 16 – podešavanje koraka spajanja redova i filtriranja merenja na izračunate periode

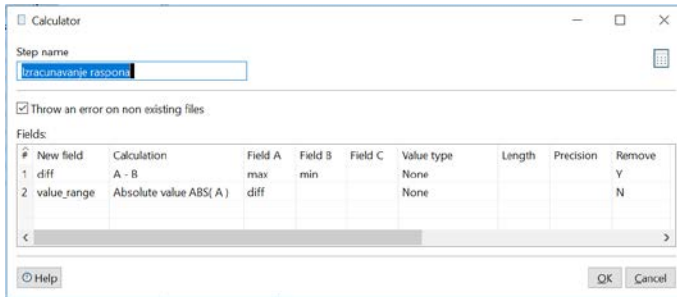
U sledećem koraku možemo konačno primeniti funkcije agregacije nad kreiranim tabelama pri čemu ćemo formirati predstavnike svakog perioda na osnovu neke od funkcija kao što su srednja vrednost merenja tokom perioda,

medijana, minimum, maksimum, standardna devijacija i prebrojavanje vrednosti. Podešavanje ovog koraka je prikazano na Slici 17.



Slika 17 – korak formiranja predstavnika svakog perioda na osnovu kriterijuma srednja vrednost merenja tokom perioda, medijana, minimum, maksimum, standardna devijacija, prebrojavanje vrednosti

Nad izračunatim vrednostima možemo dalje primenjivati računice i tako formirati nove atribute. U koraku Izračunavanje raspona koristimo komponentu Kalkulator da odredimo raspon vrednosti koji su imala merenja sa senzora tokom jednog perioda. Raspon računamo tako što izračunamo razliku između najveće i najmanje vrednosti merenja u periodu. Iako će očigledno uvek umanjjenik biti veći od umanjitelja pa će razlika biti uvek pozitivna, radi ilustracije mogućnosti transformacije nad podacima računa se prvo razlika maksimuma i minimuma kao medjurezultat koji se neće dalje pamtili budući da je odabrana opcija uklanjanja, a potom se nad tim rezultatom primenjuje funkcija za apsolutnu vrednost. Ovaj korak je prikazan na Slici 18.



Slika 18 – podešavanje koraka za određivanje raspona vrednosti koji su imala merenja sa senzora tokom jednog perioda

Sledeći i naš poslednji korak predstavlja trivijalno upisivanje sračunatih tabela u odgovarajuću bazu dejta marta.

IV. ZAKLJUČAK

Rad je pružio kompletan prikaz projektovanja rešenja koje omogućava efikasno prikupljanje, transformisanje i skladištenje podataka koje generišu umreženi senzori kroz tok podataka, od operacione baze, do zvezdaste šeme. Rad je prikazao primene teorijskih metoda tehnike dimenzionog projektovanja u rešavanju konkretnih inženjerskih problema u domenu bežičnih senzorskih mreža, te prikaze praktičnih novih rešenja i ideja u aspektu transformacije podataka. Originalna vrednost rada je činjenica da je u njemu po prvi put prikazana, koliko je poznato autoru, mogućnost primene tehnike dimenzionog modelovanja nad podacima iz bežičnih senzorskih mreža.

Prednost rada je da su transformacije koje su realizovane nad podacima u sklopu ETL procesa univerzalne za veći broj aplikacija bežičnih senzorskih mreža budući da su primenjive nad podacima koji su definisani vremenskim poretkom i prostornom određenošću. Nedostatak rada se ogleda u nedostatku automatizacije procesa projektovanja i sprovođenja postupka koji je opisan i u specifičnosti šema koje su predložene.

Dalja istraživanja u domenu teme ovog rada bi po mišljenju autora bila usmerena ka automatizaciji primene postupaka prikazanih u ovom radu. Automatizacija postupaka se može postići primenom mašinskog učenja kroz obučavanje prediktivnog modela koji će tokom treniranja nad projektovanim zvezdastim šemama za konkretne sisteme zasnovane na bežičnim senzorskim mrežama uzimati u obzir karakteristike tih sistema kao što su tipovi podataka

koji se skladište, njihova kardinalnost, standardna devijacija vrednosti podataka i slično, te automatski generisati odgovarajuće zvezdaste šeme za te sisteme...

LITERATURA

- [1] William Inmon. *Building the Data Warehouse* (2005) John Wiley and Sons, ISBN 978-81-265-0645-3
- [2] Ullo, Silvia Liberata; Sinha, G. R. (2020-05-31). "Advances in Smart Environment Monitoring Systems Using IoT and Sensors". *Sensors* (Basel, Switzerland). 20 (11): 311 Bibcode:2020Senso...20.3113U. doi:10.3390/s2011311 ISSN 1424-8220. PMC 7309034. PMID 32486411.
- [3] FrancescoMario, Di; K, DasSajal; AnastasiGiuseppe (2011-08-01). "Data Collection in Wireless Sensor Networks with Mobile Elements". *ACM Transactions on Sensor Networks* (TOSN). 8: 1–31. doi:10.1145/1993042.1993049. S2CID 15576441.
- [4] Xia, Feng; Tian, Yu-Chu; Li, Yanjun; Sun, Youxian (2007-10-09). "Wireless Sensor/Actuator Network Design for Mobile Control Applications". *Sensors* (Basel, Switzerland). 7 (10): 2157–217 Bibcode:2007Senso...7.2157X. doi:10.3390/s7102157. ISSN 1424-8220. PMC 3864515. PMID 28903220.
- [5] "Wireless sensor networks for battlefield surveillance" (PDF). 2006.
- [6] Dargie, W. and Poellabauer, C. (2010). *Fundamentals of wireless sensor networks: theory and practice*. John Wiley and Sons. pp. 168–183, 191–192. ISBN 978-0-470-99765-9.
- [7] Sohrawy, K., Minoli, D., Znati, T. (2007). *Wireless sensor networks: technology, protocols, and applications*. John Wiley and Sons. pp. 203–209. ISBN 978-0-471-74300-2.
- [8] Connolly, Thomas; Begg, Carolyn (26 September 2014). *Database Systems - A Practical Approach to Design, Implementation and Management* (6th ed.). Pearson. Part 9 Business Intelligence. ISBN 978-1-292-06118-4.
- [9] Moody, Daniel L.; Kortink, Mark A.R. "From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design" (PDF). *Dimensional Modelling*.
- [10] Matteo Golfarelli; Stefano Rizzi (26 May 2009). *Data Warehouse Design: Modern Principles and Methodologies*. McGraw-Hill Osborne Media. ISBN 978-0-07-161039-1.
- [11] Ralph Kimball; Margy Ross (26 April 2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling* (Second ed.). Wiley. ISBN 0-471-20024-7.
- [12] Ralph Kimball; Margy Ross; Warren Thornthwaite; Joy Mundy; Bob Becker (January 2008). *The Data Warehouse Lifecycle Toolkit* (Second ed.). Wiley. ISBN 978-0-470-14977-5.
- [13] Ralph Kimball; Margy Ross; Warren Thornthwaite; Joy Mundy (10 January 2008). *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses* (Second ed.). Wiley. ISBN 978-0-470-14977-5.

ABSTRACT

The paper presents the applications of dimensional modeling techniques, as well as of the concepts of data mart and data warehousing, in a context of the data from wireless sensor networks. Dimensional data structures are intended to establish an efficient system for decision support and analysis of data from wireless sensor networks. Two dimensional models are formed – one from a viewpoint of the sensor and the other from a viewpoint of the mission. The steps of data transformation in a data flow from the operational database to the data mart scheme are described. It was not possible to find applications of dimensional modeling of data from wireless sensor networks in the literature prior to this paper.

**APPROACHES FOR DIMENSIONAL MODELING OF
PROCESSES IN WIRELESS SENSOR NETWORKS**

Aleksandar M. Gavrić, dr Dušan M. Vujošević