Data Intensive Systems - Miniproject

Erik Sidelmann Jensen
ejens11@student.aau.dk

Lasse Vang Gravesen
lgrave11@student.aau.dk

Dennis Jakobsen
djakob11@student.aau.dk

1 DIS Miniproject

1.1 Task B

- Sales

- Payments

- Members

- Location

We pick the 'Sales' business process, because it is the most interesting. For that we need to model sales, products, members along with the time and date.

As for granularity, when it comes to sales the time goes down to seconds. We decided to cut it off at minutes, because seconds are not that important. When it comes to members, the granularity goes down to years and there is no more information so that is the only thing that is retained.

For product it is somewhat possible to add extra levels to the hierarchy, such as category of product like dairy or soda.

The data is reasonable for paying customers and detailed enough to ask valuable questions. Examples of questions:

- How much is bought at some point during some day?

- How does the amount sold change over time?

- Which days are the most busy?

- When is it best to restock, given low activity?

- How much revenue is gained each day, week, month, year?

- Which products have changed the most in price from year to year?

- Which department or member have spent the most?

1.2 Task C

Slowly changing dimensions are not that important.

The business process proposed to be modelled in the data warehouse is the 'Sales' business process. With regards to dimensions we pick out product, time, date, and member because they allow for the most interesting queries. Our granularity with regards to time goes down to minutes across two different dimensions to reduce the amount of rows, with regard to product it only has a name attribute, with regard to member it has balance and year. It might be a good idea to split products into categories, but the data does that directly allow for that and it would have to be done manually based on the name. It might also be a good idea to show if its a special day, like if its a holiday or if a day falls in a vacation.

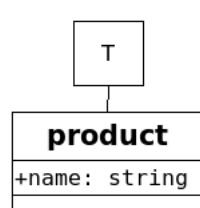The schema for the dimensions can be seen below.
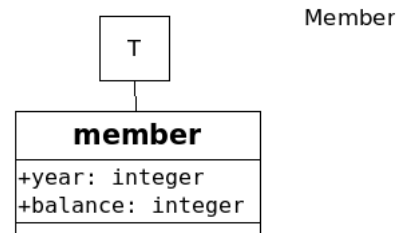
Product

Figure 1.1: The product dimension.

Member

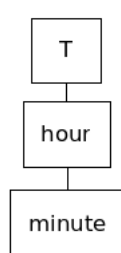Figure 1.2: The member dimension.
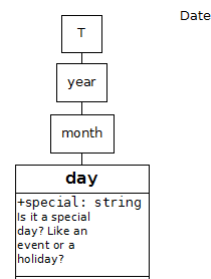
Time

Figure 1.3: The time dimension.

Date

Figure 1.4: The date dimension.

The star schema for it can be seen below. With the schema it is important to note that the dimensions have a surrogate key with a serial integer(it auto increments).
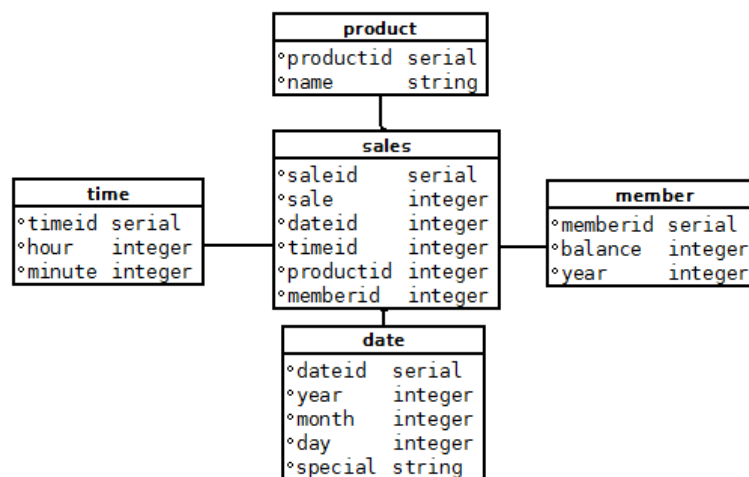
Figure 1.5: The star schema for the data warehouse.

Here is the SQL Table creation text, with some things like serial sequences stripped:

CREATE TABLE date (

```
    dateid integer NOT NULL,
    year integer,
    month integer,
    day integer,
    special text
);

CREATE TABLE member (
    balance integer,
    year integer,
    memberid integer NOT NULL
);

CREATE TABLE product (
    productid integer NOT NULL,
    name text
);

CREATE TABLE sales (
    memberid integer,
    dateid integer,
    timeid integer,
    productid integer,
    sale integer,
    saleid integer NOT NULL
);

CREATE TABLE "time" (
    timeid integer NOT NULL,
    hour integer,
    minute integer
);

ALTER TABLE ONLY date
    ADD CONSTRAINT date_pkey PRIMARY KEY (dateid);

ALTER TABLE ONLY member
    ADD CONSTRAINT member_pkey PRIMARY KEY (memberid);

ALTER TABLE ONLY product
    ADD CONSTRAINT productid PRIMARY KEY (productid);

ALTER TABLE ONLY sales
    ADD CONSTRAINT sales_pkey PRIMARY KEY (saleid);

ALTER TABLE ONLY "time"
    ADD CONSTRAINT timeid PRIMARY KEY (timeid);

ALTER TABLE ONLY sales
```

ADD CONSTRAINT dateid **FOREIGN KEY** (dateid) REFERENCES **date** (dateid);

ALTER TABLE ONLY sales
ADD CONSTRAINT memberid **FOREIGN KEY** (memberid) REFERENCES member (memberid);

ALTER TABLE ONLY sales
ADD CONSTRAINT producid **FOREIGN KEY** (productid) REFERENCES product (productid);

ALTER TABLE ONLY sales
ADD CONSTRAINT timeid **FOREIGN KEY** (timeid) REFERENCES "time" (timeid);