



Софийски университет „Св. Кл. Охридски”

Факултет по математика и информатика

Курсов Проект

на тема:

Data Warehouse for the Core Banking System (CBS)

Изготвил/и: **Елена Георгиева**, Ф.Н. 81555

Специалност: КН, Курс: 4 , Учебна година: 2019/20

Преподавател: **Тодор Кичуков**

Дата: 18.12.2019

1. Business requirements

Целта на проекта е да се реализира Data Warehouse, който да съдържа данните на банковата система във форма, подходяща за получаване на отчети (reporting) и анализиране на данните (analyzing) с цел взимане на оптимални бизнес решения.

Оперативният модел на данните (3NF-нормализиран), с който разполагаме, се вижда на фигура 1.

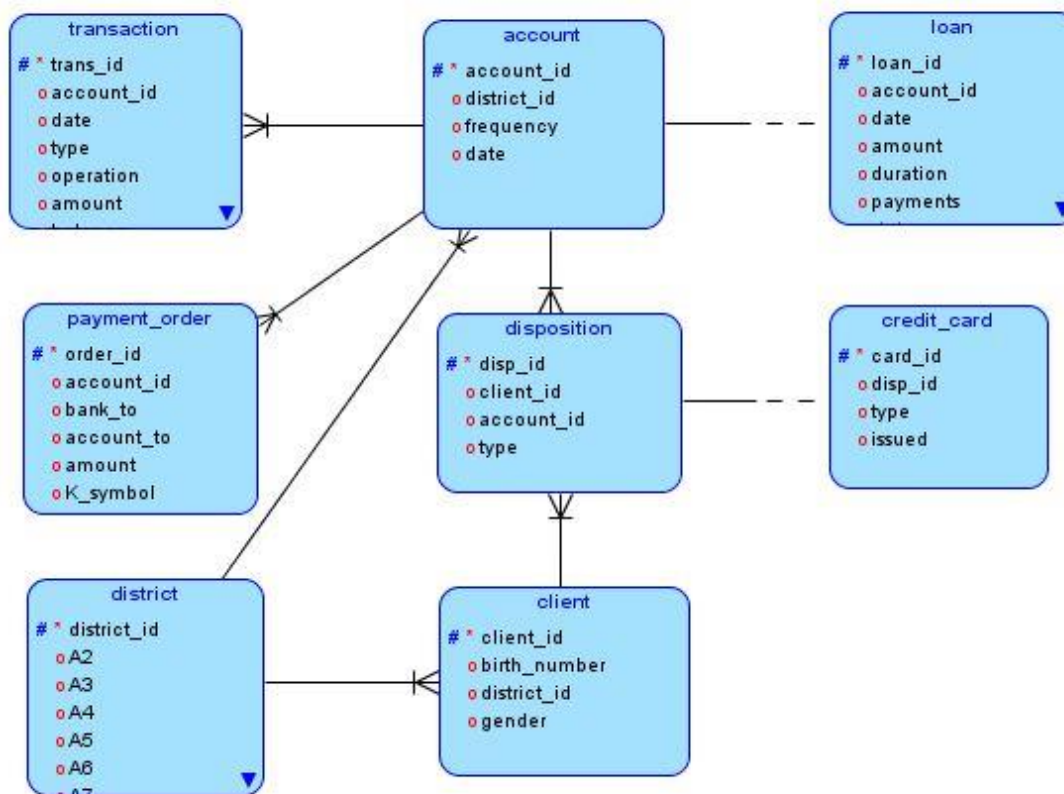


Figure 1 - 3NF model

2. Проектиране и реализация

2.1. Проектиране

Проектът реализира димензионен модел на данните, предоставени от банката. [1] Димензионният модел е организиран около 4 business events на

банката, които се случват на дневен базис и са количествено измерими. Те са представени във факт таблици (fact tables), които съдържат съответните measurements. Тези събития и репрезентиращите ги факт таблици са:

- Транзакции (transaction_fact)
- Платежни нареждания (order_fact)
- Отпускане на заеми (loan_fact)
- Издаване на кредитни карти (credit_card_fact)

Информацията, описваща тези бизнес събития, е организирана в 4 димензионни таблици, общи за всички факт таблици (conformed dimensions):

- account_dimension
- client_dimension
- district_dimension
- date_dimension

Всички случаи на dimension-to-dimension references са моделирани чрез поставяне на foreign key към реферираната dimension table във факт таблицата, вместо да правим outrigger dimension и да имаме dimension-to-dimension връзка в димензионния модел. (Съгласно източник [2])

Всички атрибути от факт таблиците, които не представляват measurement, а описание на факт, но няма друга информация за тях освен primary key, остават във факт таблицата като degenerate dimensions (DD) , например type, operation, k_symbol. (Съгласно източник [3])

credit_card_fact е factless fact table, в която няма количествени измерения, но отчита и протоколира събитието „издаване на кредитна карта” като случило се в даден момент от времето. [4]

Фигури 2 и 3 илюстрират димензионния модел.

2.2.Реализация

Скриптът **tables_script.sql** създава и зарежда с данни всички релации, дадени в условието на задачата. Данните в таблицата client.csv са обработени предварително, като информацията за дата на раждане и пол е разделена на два отделни атрибута.

Допълнително е създадена таблица `date.csv`, съдържаща информация за всички дати от 01.01.1993 до 01.01.1999г.

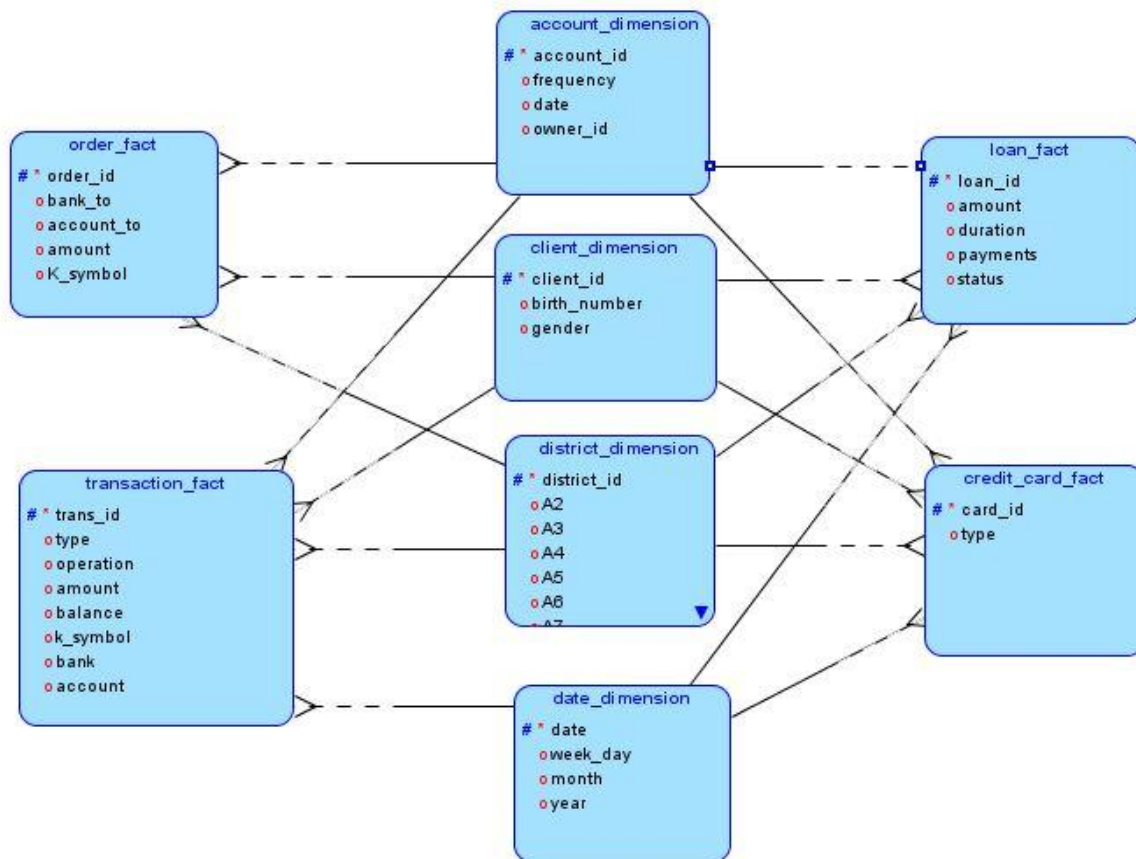


Figure 2 - Dimensional model (logical)

Следва генерирането на димензионния модел, като първо се зареждат с данни димензионните таблици, а след това факт таблиците, рефериращи димензионните таблици.

Проектът е моделиран чрез *Oracle SQL Developer Data Modeler*, а базата данни и всички таблици са създадени в *Microsoft SQL Server Management Studio*.

Моделът служи за удовлетворяване на reporting заявки от вида:

- Кой райони се отличават с най-много клиенти с непогасени заеми?
Каква е средната заплата и нивото на безработица в тези райони?

- Клиентите на кой филиал са направили най-много транзакции през май 1995?
- Колко кредитни карти са издадени през 1996?
- Кои са клиентите, които са погасили всичките си заеми, и от кои райони са?

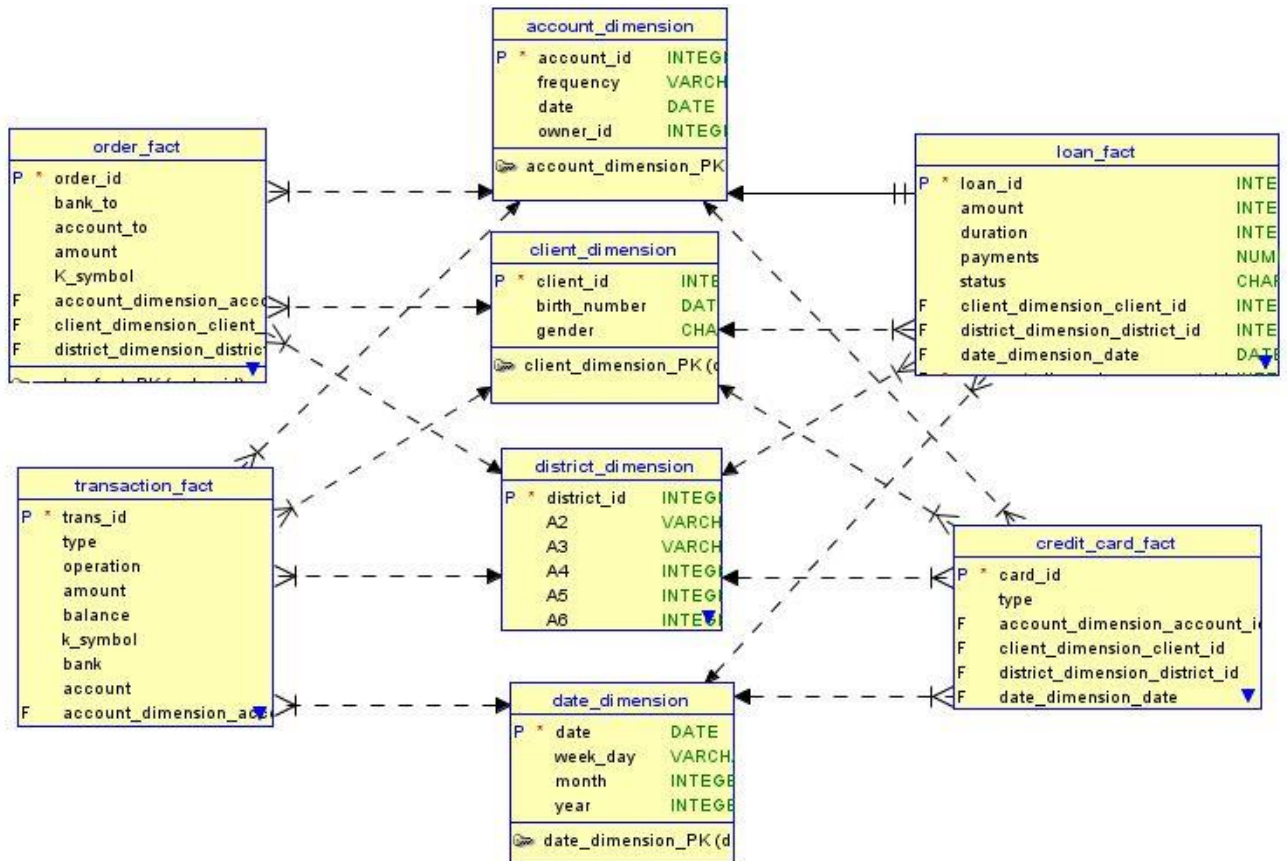


Figure 3 - Dimensional model (Relational)

Използвани източници

- [1] <https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/>
- [2] <https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/dimension-to-dimension-join/>

[3] <https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/degenerate-dimension/>

[4] <https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/factless-fact-table/>