

|  |
| --- |
| Migration business from Oracle to Microsoft Azure  **BOOK SALES** |
|  |

Contents

[1 Referenced documents 3](#_Toc95929175)

[2 Business Description 4](#_Toc95929176)

[3 Dimensions of a Business 4](#_Toc95929181)

[4 ARCHITECTURE DIAGRAM ON MICROSOFT AZURE 10](#_Toc95929182)

[5 Logical Scheme 14](#_Toc95929184)

[6 Data Flow 14](#_Toc95929185)

[7 Partitioning strategy 16](#_Toc95929186)

[8 REPORT 17](#_Toc95929187)

[9 STRUCTURE THE REPOSITORY OF PROJECT 18](#_Toc95929188)

# 

# Referenced documents

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Title** | **Date** | **Path** |
| 1 | Business\_Template\_BookSales.docx | 10/02/22 | <https://github.com/ELENA-VM/DWH_ON_PREMISE/blob/31e468e095627323c175dfe70477f4333a9a47b9/docs/Business_Template_Elena_Mozhenkova.docx> |
| 2 | Repository with the solution | 10/02/22 | https://github.com/ELENA-VM/DWH\_ON\_CLOUD |
|  |  |  |  |

# Business Description

## ThE GOAL

The aim of the program is to move the on-premises solution "Book sales" to the Microsoft Azure cloud.

## Business background

Books are companions for every person during in the life. In current times many people prefer to read electronic books. The Internet allows finding necessary literature doesn't go out from home. But paper books are also demanded in this hard competition. So that the business will succeed, you should very responsible approach for analyzing different metrics influencing sales. To collect, transform and analyze data you can use special tools.

## Problems because of poor data management

Poor data management doesn’t let to correctly analyze the situation in the sales market. You can't do truth choice for growing the business. If you don’t use special instruments for common analysis business data, you won’t be competitive on market book sales.

## Benefits from implementing a Data Warehouse

Using of a data warehouse can help you with the problems described above. Implementing a data warehouse can answer the following questions:

* Which category of books has the highest sales?
* What distribution sales by region?
* Quantity of new customers by period.
* Profit by Customer.
* Category book ranking.

Further processing data would also let you:

* Analyze places for business extension.
* Analyze the time period when increase sales.
* Customer preferences by book category.
* Price dynamics.
* And many other.

# Dimensions of a Business

Design the dimensional model using four steps.

**Step 1. Select business project**

The main target to analyze the business – sales books from different stores and shops.

**Step 2. Declare the Grain**

Analyzing the business process can define the grain entity «sale» as the atomic element for research.

**Step 3. Identify the dimensions**

Define dimension table:

dim\_store – data about shops and warehouses

dim\_book\_scd – data about products of the company (slowly changing dimensions type 2 – storing history change row from the source system)

dim\_customers – retail and wholesale buyers

dim\_employee - store information about the staff of the business

dim\_type\_payments – kind of payments

dim\_address – geography

dim\_dates – date values

All dimension tables exclude “dim\_time” generate their own unique surrogate primary key and

store link with source system use composite natural key(source\_id + table\_name + system\_name).

**Step 4. Identify the Facts**

Table fact «fct\_sales» store information:

- composite natural key – sales\_id + table\_namу + system\_name

- foreign keys to dimension tables (book\_surr\_id, time\_id, customer\_suur\_id, store\_surr\_id, employee\_surr\_id, address\_surr\_id, type\_payment\_surr\_id)

- additive facts: quantity, sale\_amount

- degenerate field: num\_invoice

**The Star schema model «Book sales»**

The new solution migrated to Azure SQL, so some data types have been changed:

* VARCHAR2(XX) -> VARCHAR(XX)
* NUMBER for primary keys -> BIGINT
* NUMBER(10.2) -> NUMERIC(10.2)
* other NUMBER -> INTEGER

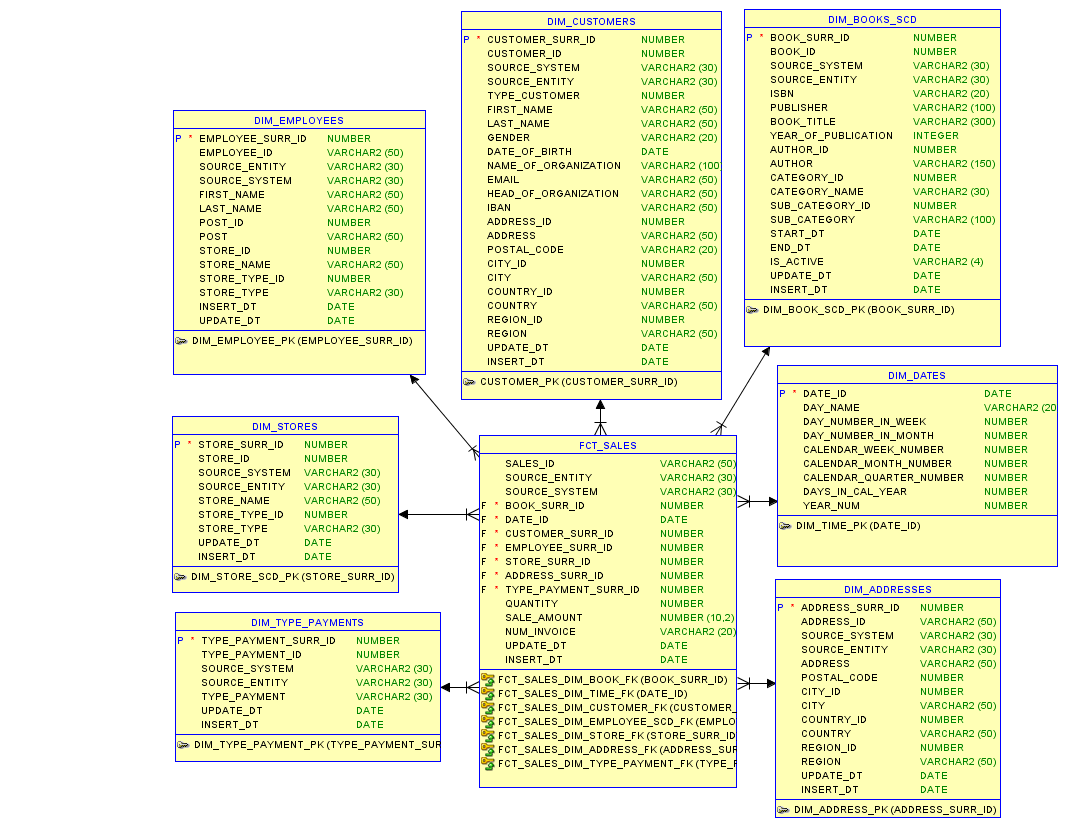
****

Fig. 1 – Oracle Data Modeler star schema



Fig. 2 – MSSM star schema

**The 3NF schema model «Book sales»**

For creating many-level architecture DWH also should create a high-level data model in 3NF. Model 3NF creating with a description business process and yet created the star model. The dimension tables were divided into different tables. For all tables exclude “ce\_book\_author” generate their own unique surrogate primary key and store link with source systems using a composite natural key (source\_id + source\_entity + source\_system).

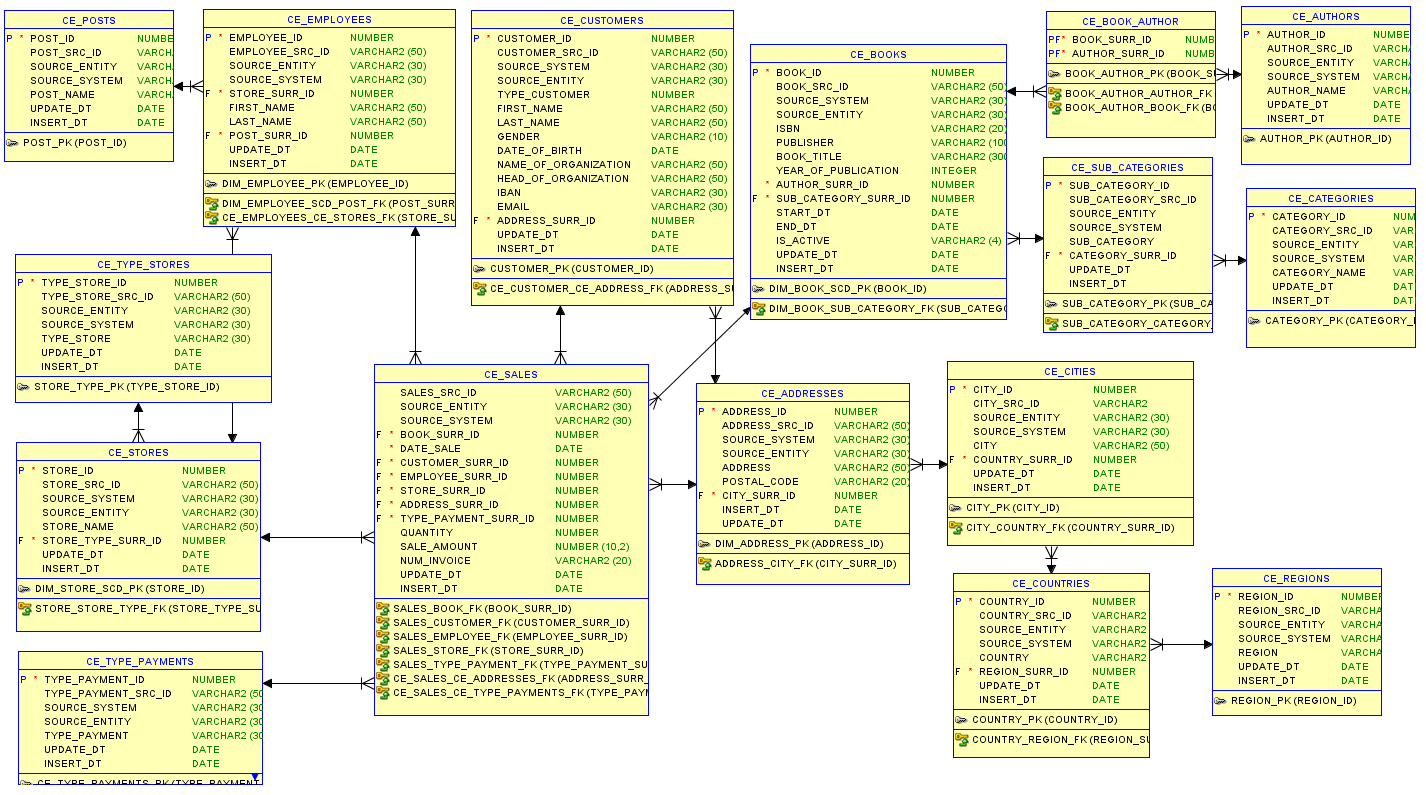


Fig. 3 – Oracle Data Modeler 3NF schema



Fig. 4 – MSSM 3NF schema

# ARCHITECTURE DIAGRAM ON MICROSOFT AZURE

Diagram showing how on-premises solution can migrate and modernize with Azure Blob Storage, SQL Database, Azure Data Factory, and other services. This scenario demonstrates a data pipeline that integrates large amounts of data from two sources into a unified analytics platform in Azure.

# Diagram Description automatically generated with medium confidence

Fig. 5 – Architecture diagram

The data flows through the solution as follows:

* The semi-structured data such as CSV files enter the Data Blob Storage for further analysis and storage
* For each data source, any updates are exported periodically into a staging area in Azure Blob storage
* Azure Logic App “RunETL” (fig. 6) runs the main pipeline (fig. 7)
* Data Factory pipelines sa\_retail\_pipeline\_part\_1, sa\_retail\_pipeline\_part\_2, sa\_stock\_pipeline, bl\_3nf\_pipeline, bl\_dm\_pipeline (figs. 8-12) run for executing stored procedures in Azure SQL
  + Firstly, stored procedures load the data from Blob storage into staging tables (sa\_retail, sa\_stock)
  + Secondly, stored procedures incrementally load the data into 3NF layer (BL\_3NF)
  + Lastly, stored procedures incrementally load the data into star layer (BL\_DM)
* Actions “Web Email notifications” runs Azure logic app “SendEmailProcessEtl” for sending an email about the result run pipelines (fig. 13)
* Pipeline “move\_blob\_pipeline” (fig 14) move blob files from “input” to “output” containers
* Business analysts use Microsoft Power BI to analyze warehoused data.

Diagram

Description automatically generated

Fig. 6 – Azure Logic App “RunETL”

Graphical user interface

Description automatically generated

Fig. 7 – The pipeline “Load ETL”

Graphical user interface, application

Description automatically generated

Fig. 8 – The pipeline “sa\_retail\_pipeline\_part\_1”

Graphical user interface

Description automatically generated

Fig. 9 – The pipeline “sa\_retail\_pipeline\_part\_2”

Graphical user interface, application, chat or text message

Description automatically generated

Fig. 10 – The pipeline “sa\_stock\_pipeline”

Graphical user interface, application

Description automatically generated

Fig. 11 – The pipeline “bl\_3nf\_pipeline”

Graphical user interface, text, application, chat or text message

Description automatically generated

Fig. 12 – The pipeline “bl\_dm\_pipeline”

Diagram

Description automatically generated

Fig. 13 – Azure Logic App “SendEmailProcessEtl”

Graphical user interface, application, Teams

Description automatically generated

Fig. 14 – The pipeline “move\_blob\_pipeline”

# Logical Scheme

Create logical model of DWH load

Diagram

Description automatically generated

# Data Flow

Design data flow diagrams for DWH load

**DIMENSION TABLES**

Graphical user interface, application

Description automatically generated

Graphical user interface, diagram

Description automatically generated

Graphical user interface, application, PowerPoint

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

**FACT TABLE**

Graphical user interface, application

Description automatically generated

# Partitioning strategy

For optimization to work with the Datawarehouse will create partitions on the layer BL\_3NF for table CE\_SALES and on the dimension layer for table FCT\_SALE. It improves load data to DWH and query performance. These tables were chosen for partition because contain historical data and the size table will greater than 2 GB.

Since the analysis of sales data is planned to execute by period, the partitioning strategy by range is selected. Selected the period - month as the unit of the partition.

# REPORT

The dashboard has been prepared for visualization and analyzing the data of DWH.

Map

Description automatically generated

# STRUCTURE THE REPOSITORY OF PROJECT

The infrastructure of the Azure solution has been implemented as code using Azure Resource Manager templates (dwso/arm\_templates/\*.json). Steps of the deployments flow are described in 01\_configure.ps1 using PowerShell scripts.

The directories for each schema of the Azure SQL databases are located in the “dbs” folder. The next level is directories for each type of database object. The script for creating all structure databases is saved in the script\_NAME\_SCHEMA.sql files.

The dashboard is stored in the “reports” folder.

Text

Description automatically generated with medium confidence