Solution Concept of KFC

Mastykina Elizaweta

<u>Brief overview of the brand KFC:</u>

KFC (**Kentucky Fried Chicken**) is an American fast food restaurant chain headquartered in Louisville, Kentucky, that specializes in fried chicken. It is the world's second-largest restaurant chain (as measured by sales) after McDonald's, with 22,621 locations globally in 150 countries as of December 2019.

Let this clothing brand face certain difficulties related to incorrect data handling. Some of the possible problems:

      1. lack of business intelligence from multiple sources;

      2. reduced query and system performance;

      3. lack of timely access to data;

      4. the lack of historical intelligence.

      We offer the customer to use DWH(we will consider 2 types of storage schemes: star scheme and snowflake scheme) for the following reasons:

      1. It is better to make decisions. Corporate decision makers will no longer have to make important business decisions based on limited data and guesswork. The data warehouse will store reliable facts and statistics, and decision makers will be able to extract this information from the data warehouse based on their personal needs.

      2. Quick and easy access to data. Speed is an important factor that puts the company above its competitors. Business users can quickly access data from multiple sources from the data warehouse, which means that precious time will not be wasted extracting data from multiple sources. This allows the company to make fast and accurate decisions, with little or no support from the IT Department.

      3. The quality and consistency of data. As data warehouses collect information from various sources and convert it into a single and widely used format, departments will produce results that are consistent and consistent

with each other. When data is standardized, a company can be confident in its accuracy, and accurate data is what makes it possible to make strong business decisions.

Business Requirements:

ID	Description
B01	Calculation of information about volume of sales on monthly basis
B02	Calculation of information about total sales for each product for each country
B03	Calculation of information about discounts and coupons to optimize margin
B04	Each office has its own geographical location. The warehouse must take this fact into account
B05	Storing information about products in the warehouse, supporting easy and quick search of the selected product

Technical Requirements:

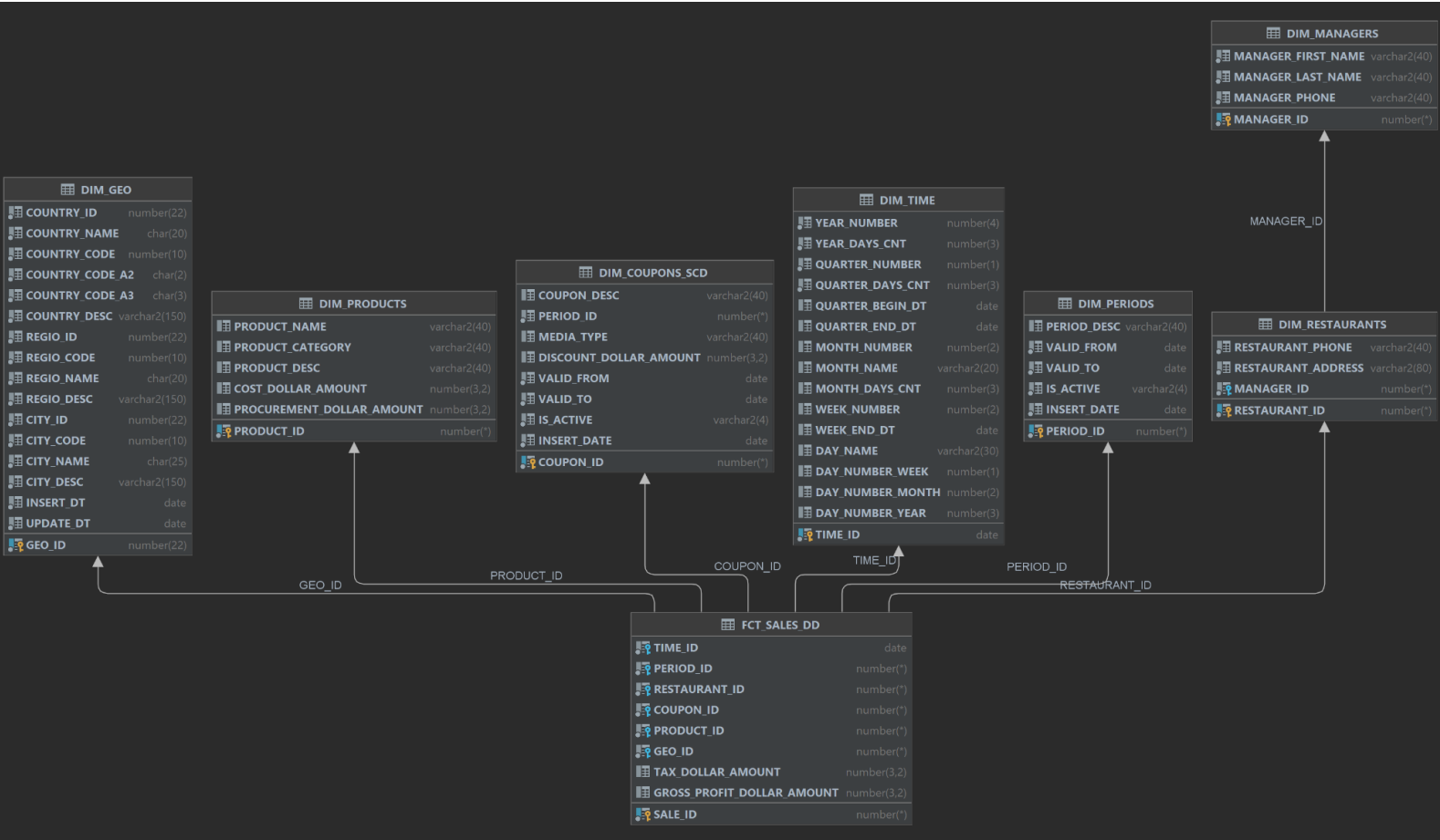| ID	Description
T01	Ability to process large amounts of information per day (1 million rows)
T02	The System should be available at least during the restaurant operating hours and must be recovered within an hour or less if it fails. The system should respond to the requests within two seconds or less.
T03	The system should provide consistent performance with easy tracking of records and updating of records.
T04	The software should be easily maintainable and adding new features and making changes to the software must be as simple as possible.

T05	Only authorized users must be able to access the system and view and modify the data.
T06	All the information must be protected according to the company's security Policy

Solution Sketch
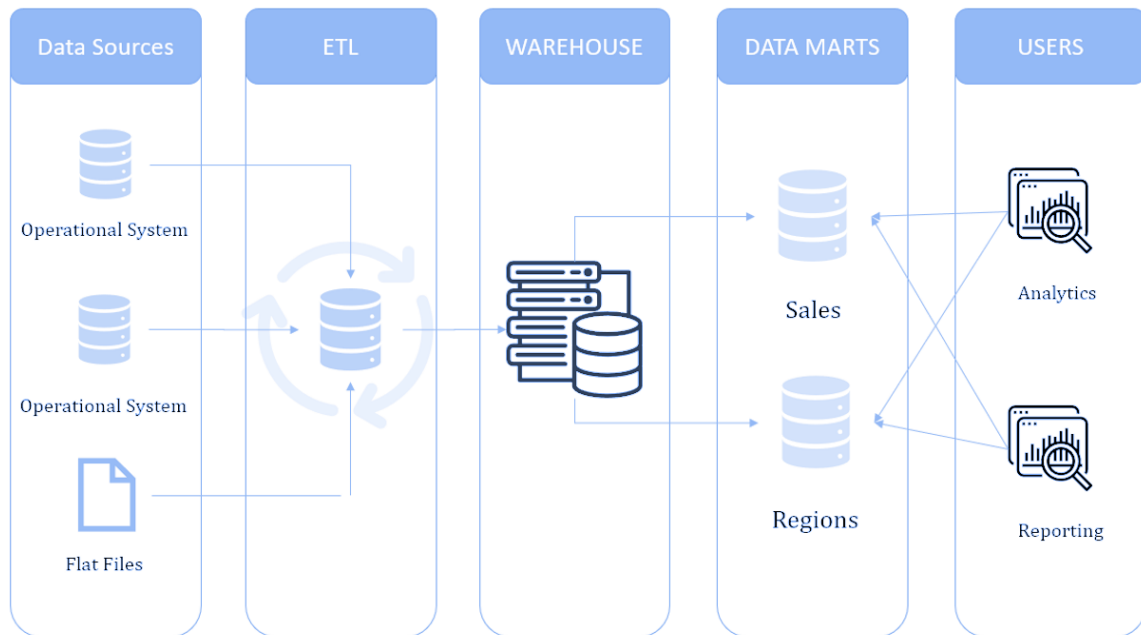
Conception Design

Physical Diagram

Dimension Types Description

Name	Type	Size	DW – Merged Dimensions	Descriptions
DIM _TIME	SCD1	BIG	DW.T_DAYS, DW.T_WEEKS, DW.T_MONTHS, DW.T_QUARTERS, DW.T_YEARS	A Calendar, contain wide information about days, weeks, months, quarters and years.
DIM_PERIODS	SCD1	BIG	PERIOD_DESC, VALID_FROM, VALID_TO, IS_ACTIVE, INSERT_DATE	Processing Dimension – Useful in Context of Data Aggregation.
DIM_MANAGER	SCD1	BIG	FIRST_NAME, LAST_NAME, PHONE	This dimension contain all information about managers.

DIM_RESTAURANS	SCD1	MEDIUM	RESATURANT_DESC, RESATURANT_ADDRESS, MANAGER_ID	This dimension contain all information about restaurnts.
DIM_GEO	SCD1	MEDIUM	DW.T_COUNTRIES, DW.T_REGIONS, DW.T_CITIES	This one stored data about all available locations all over the world.
DIM_PRODUCTS	SCD1	MEDIUM	PRODUCT_NAME, PRODUCT_CATEGORY, PRODUCT_DESC, COST, PROCUREMENT_AMOUNT,	Dimensions stores information about products and their characteristics.

DIM_COUPONS_SCD	SC D2	BIG	DESCRIPTION, MEDIA_TYPE, DISCOUNT_AMOU NT, VALID_FROM, VALID_TO, IS_ACTIVE, INSERT_DATE	Dimension is stored detailed information about coupons. This Table was chosen as SCD2 dimension type for rejecting data merging.

Dimension Hierarchies

DIM_TIME:

Hierarchy DAY-WEEK-MONTH- QUARTER -YEAR

Name	LEVEL_CODE	LEVEL_DESC	LEVEL_NATURAL_KEY
DAYs	DAY	Store all day at the calendar	DAY_NUMBER_YEAR
WEEKs	WEEK	Store all weeks at the calendar year	WEEK_NUMBER

MONTHs	MONTH	Store all months at the calendar year	MONTH_NUMBER
QUARTERs	QUARTER	Store all quarter at the calendar year	QUARTER_NUMBER
YEARs	YEAR	Store all years	YEAR_NUMBER

DIM_GEO:

Hierarchy CITY-REGIO-COUNTRY

Name	LEVEL_CODE	LEVEL_DESC	LEVEL_NATURAL_KEY
CITYs	CITY	Store all cities in the country	CITY_ID
REGIOs	REGIO	Store all regios in the country	REGIO_ID
COUNTRYs	COUNTRY	Store all countries	COUNTRY_ID

DIM_PERIODS:

Hierarchy START_DATE – END_DATE – PERIOD_NAME

Name	LEVEL_CODE	LEVEL_DESC	LEVEL_NATURAL_KEY
START DATEs	START_DT	Store start date of looking period	START_DT
END DATEs	END_DT	Store end date of looking period	END_DT
PERIOD NAMEs	PERIOD_ID	Store all periods	PERIOD_ID

DIM_MANAGER:

Hierarchy FIRST_NAME – LAST_NAME –PHONE-MANAGER

Name	LEVEL_CODE	LEVEL_DESC	LEVEL_NATURAL_KEY
FIRST NAMEs	FIRST_NAME	Contain information about all managers first names	FIRST_NAME
LAST NAMEs	LAST_NAME	Contain information about all managers last names	LAST_NAME
PHONEs	PHONE	Contain phone numbers. Is for subjectification while searching	PHONE
MANAGERs	MANAGER	Contain Information about all existing managers	MANAGER_ID

Facts aggregation:

Name	Code	Table Name	Additive	Descriptions

TAX_DOLAR_AMOUNT	TAX_DOLAR_AMOUNT	FCT_SALES_DD	+	The estimated total taxes in the fact table for the corresponding period.
GROSS_PROFIT_DOLAR_AMOUNT	GROSS_PROFIT_DOLAR_AMOUNT	FCT_SALES_DD	+	The estimated total gross profit in the fact table for the corresponding period.

DataFlow Diagram
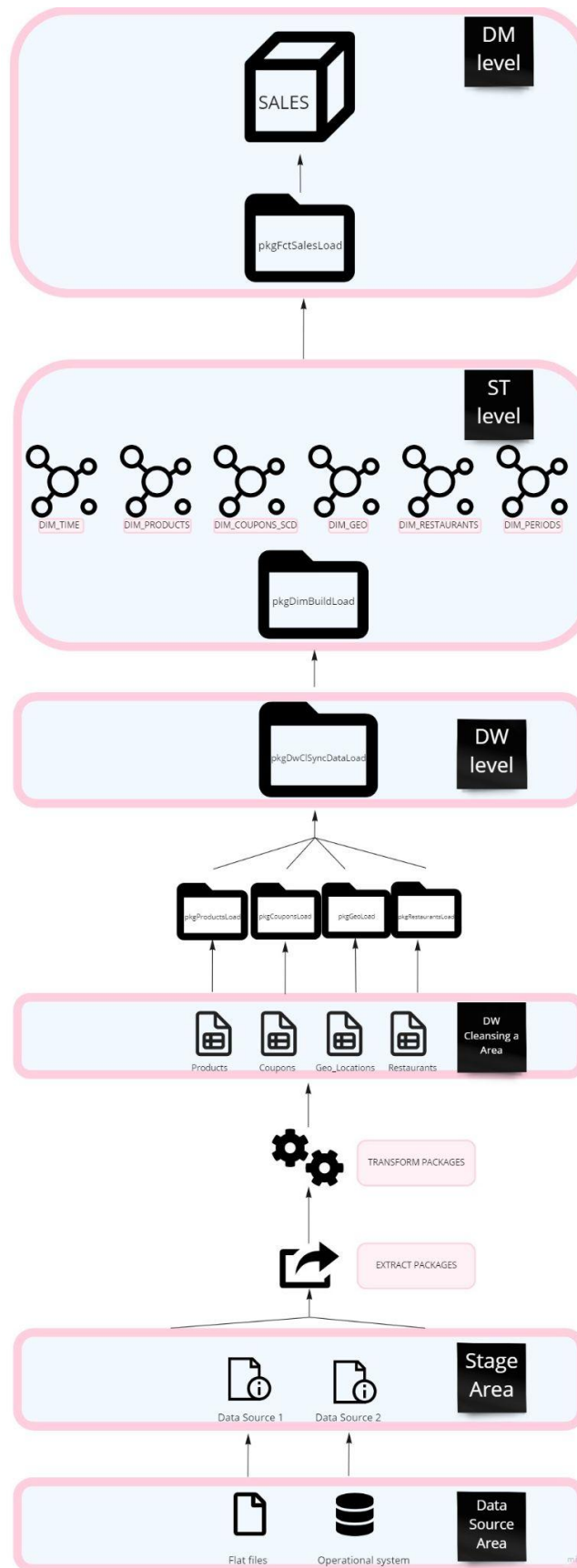
Partitioning Facts

The concept of Fact table is to create opportunity to suggest successful reporting and analyse trends as well perspectives. Right Strategy is to create a composite range-hash partitioned table.

Partitioning by Range on Time_id column will be adoptet by month, cause the reporting will be over the days. Subpartition is represented by hash over the Restaurant_ID by value 4.

```sql
CREATE TABLE  FCT_SALES_DD (
    SALE_ID NUMBER NOT NULL,
    TIME_ID DATE NOT NULL,
    PERIOD_ID NUMBER NOT NULL,
    RESTAURANT_ID NUMBER NOT NULL,
    COUPON_ID NUMBER NOT NULL,
    PRODUCT_ID NUMBER NOT NULL,
    GEO_ID NUMBER NOT NULL,
    TAX_DOLLAR_AMOUNT NUMBER(3,2),
    GROSS_PROFIT_DOLLAR_AMOUNT NUMBER(3,2),
    CONSTRAINT PK_FCT_SALES PRIMARY KEY(SALE_ID),
    CONSTRAINT FK_SALES_PERIODS FOREIGN KEY (PERIOD_ID) REFERENCES DIM_PERIODS (PERIOD_ID),
    CONSTRAINT FK_SALES_TIME FOREIGN KEY (TIME_ID) REFERENCES DIM_TIME (TIME_ID),
    CONSTRAINT FK_SALES_RESTAURANTS FOREIGN KEY (RESTAURANT_ID) REFERENCES DIM_RESTAURANTS (RESTAURANT_ID),
    CONSTRAINT FK_SALES_COUPONS FOREIGN KEY (COUPON_ID) REFERENCES DIM_COUPONS_SCD (COUPON_ID),
    CONSTRAINT FK_SALES_PRODUCTS FOREIGN KEY (PRODUCT_ID) REFERENCES DIM_PRODUCTS (PRODUCT_ID),
    CONSTRAINT FK_SALES_GEO FOREIGN KEY (GEO_ID) REFERENCES DIM_GEO (GEO_ID)
)
PARTITION BY RANGE (TIME_ID) INTERVAL (NUMTOYMINTERVAL(1,'MONTH'))
SUBPARTITION BY HASH (RESTAURANT_ID) SUBPARTITIONS 4
(
    PARTITION P0 VALUES LESS THAN (TO_DATE('1-1-2020','DD-MM-YYYY')),
    PARTITION P1 VALUES LESS THAN (TO_DATE('1-1-2021','DD-MM-YYYY')),
    PARTITION P2 VALUES LESS THAN (TO_DATE('1-1-2022','DD-MM-YYYY'))
)
```

Strategy of Parallel Execution

Parallel Execution is a strong tool to increase Data Warehouse performance. The solution concept includes multiple tables with rapidly-changing data. Therefore, it is necessary to constantly use insert and update commands. Consequently, the best strategy is the using of Parallel DML (PARALLEL INSERT, UPDATE, DELETE, and MERGE), which uses parallel execution mechanisms to speed up or scale up large DML operations against large database tables and indexes.