**OnLine\_Shop**

**Gulliver**

****

solution

Version number 1.0

|  |  |
| --- | --- |
| **Submission Date** | *07/26/2013* |
| **Requested By** | *Kyril Bucha* |
| **Business Owner** | *Ihar\_Yelin* |
| **Contact Info.** | *Ihar\_Yelin@epam.com* |

# Overview

First of all, we must determine what we have for solving the problem. The client provide some external files with different information:

|  |  |
| --- | --- |
| Data of interest | Description |
| STUCTURE | Information about departments (or single agents), employees |
| TIMES | Time period for analysis: day, week, month, quarter, year |
| PRODUCTS | Category of products, description of each product, their cost |
| LOCATIONS | Countries, region, part, continent |
| CLIENTS | Detail information about clients (first name, last name, email and etc.) |
| DELIVERY | Full information about delivery products, localization of warehouse |
| SUPPLIERS | Proven suppliers of the shop |

## Business Background

Online shop market manufactured goods all over the world. So it has warehouses in different countries, large amount of departments and single agent in different cities. Shop needs statistics by each country in different periods in time.

## Benefits

1. Depending on where the best sold, we can plan where better to locate warehouses for cheaper and sooner delivery products.
2. Opportunity to analyze employee skills in marketing.
3. Depending on sold, analyze to solve problem with the size of warehouse, amount of employees .
4. We can determine the best suppliers for the future cooperation.
5. To determine clients with the largest amount of purchase for provide different benefits.
6. Help to decide what product to buy and how much.

# Requirements

## Business Requirements

1. Analyze data for several years for required period of time for planning.
2. Possibility to change duty value.
3. Possibility to set percent of allowance.
4. Calculate amount sold for every week.
5. Calculate per countries for main departments every day.

## Technical Requirements

1. Update data every day.
2. Store data in 4 servers (4 countries, main departments).
3. Data stored for 2 years.

# Solution Sketch

## Source Tables structure

STRUCTURE

|  |  |  |
| --- | --- | --- |
| Column name | Data Type | Description |
| emp\_id | number(8,2) | id of employee |
| emp\_first\_name | varchar(50) | first name |
| emp\_last\_name | varchar(50) | last name |
| emp\_experience | number(8,2) | amount years of experience |
| emp\_tel\_number | number(8,2) | telephone number |
| emp\_email | varchar(50) | employee’s email |
| dept\_id | number(8,2) | id of department |
| dept\_desc | varchar(50) | description of the department |
| dept\_mng\_id | number(8,2) | id of manager in this department |
| dept\_city | varchar(50) | city where department located |
| dept\_tel\_number | number(8,2) | telephone number |
| localization\_id | number(8,2) | all information of department location |

## PRODUCTS

|  |  |  |
| --- | --- | --- |
| Column name | Data Type | Description |
| prod\_id | number(8,2) | id of product |
| prod\_desc | varchar(50) | description of product |
| prod\_category | varchar(50) | product category |
| prod\_cost | number(8,2) | cost of each product |

## LOCATIONS

|  |  |  |
| --- | --- | --- |
| Column name | Data Type | Description |
| country\_id | number(8,2) | country unique code |
| country\_ desc | varchar(50) | country name |
| country\_code | varchar(50) | country abbreviation |
| region\_id | number(8,2) | region unique code |
| region\_ desc | varchar(50) | region name |
| region\_code | varchar(50) | region abbreviation |
| part\_id | number(8,2) | part unique code |
| part\_ desc | varchar(50) | part name |
| part\_code | varchar(50) | part abbreviation |
| continent\_id | number(8,2) | continent unique code |
| continent\_ desc | varchar(50) | continent name |

## TIMES

|  |  |  |
| --- | --- | --- |
| Column name | Data Type | Description |
| time\_id | date | time unique code |
| day\_name | varchar(50) | Name of the day |
| day\_number\_in\_week | number(8,2) | Day number in week |
| day\_number\_in\_month | number(8,2) | Day number in month |
| day\_number\_in\_year | number(8,2) | Day number in year |
| calendar\_week\_number | number(8,2) | Calendar week number |
| week\_ending\_date | date | Week ending date |
| calendar\_month\_number | number(8,2) | Calendar month number |
| Days\_in\_cal\_month | number(8,2) | Days in calendar month |
| End\_of\_cal\_month | date | End of calendar month |
| Calendar\_month\_name | varchar(50) | Calendar month name |
| Days\_in\_cal\_quarter | number(8,2) | Days in calendar quarter |
| Beg\_of\_cal\_quarter | date | Beg of calendar quarter |
| end\_of\_cal\_quarter | date | End of calendar quarter |
| Calendar\_quarter\_number | number(8,2) | Calendar quarter number |
| Calendar\_year | number(8,2) | Calendar year |
| Days\_in\_calendar\_year | number(8,2) | Days in calendar year |
| End\_of\_cal\_year | date | End of calendar year |

CLIENTS

|  |  |  |
| --- | --- | --- |
| Column name | Data Type | Description |
| client\_id | number(8,2) | client unique code |
| first\_name | varchar(50) | first name |
| last\_name | varchar(50) | last name |
| gender | varchar(2) | gender |
| street\_address | varchar(100) | address |
| postal\_code | varchar(50) | postal code |
| city | varchar(50) | city in which live |
| location\_id | number(8,2) | location |
| account | number(8,2) | amount of money |
| email | varchar(50) | email of client |

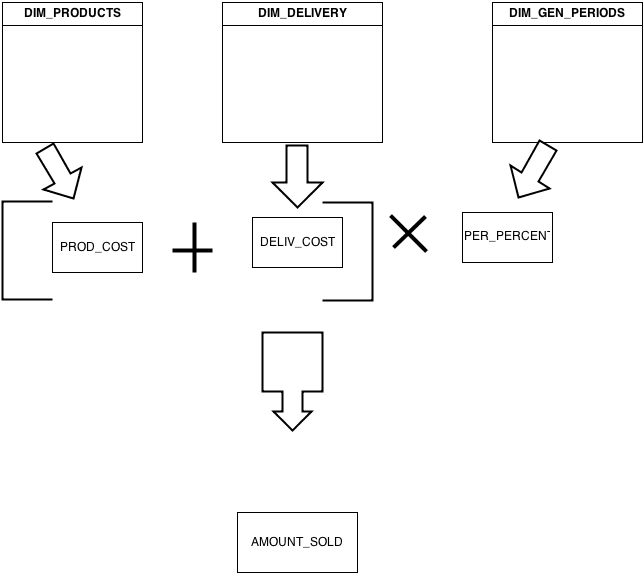
## DELIVERY

|  |  |  |
| --- | --- | --- |
| Column name | Data Type | Description |
| delivery\_id | number(8,2) | delivery unique code |
| category | varchar(50) | category of delivery |
| cost | number(8,2) | cost of each delivery |
| localization\_id | number(8,2) | location of warehouse |

## SUPPLIERS

|  |  |  |
| --- | --- | --- |
| Column name | Data Type | Description |
| suppl\_id | number(8,2) | supplier unique code |
| name | varchar(50) | name |
| category | varchar(50) | category |
| tel\_number | number(8,2) | telephone number |
| email | varchar(50) | email |
| localization\_id | number(8,2) | location of supplier |

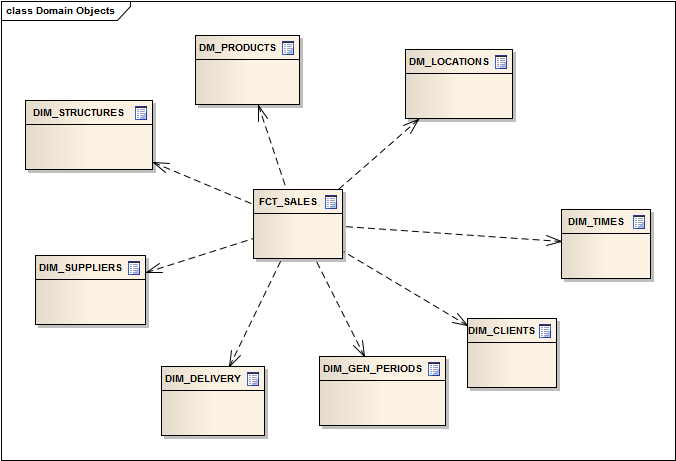
## Summarize Data Plan



# DWH Solution Concept

## Logical Diagram

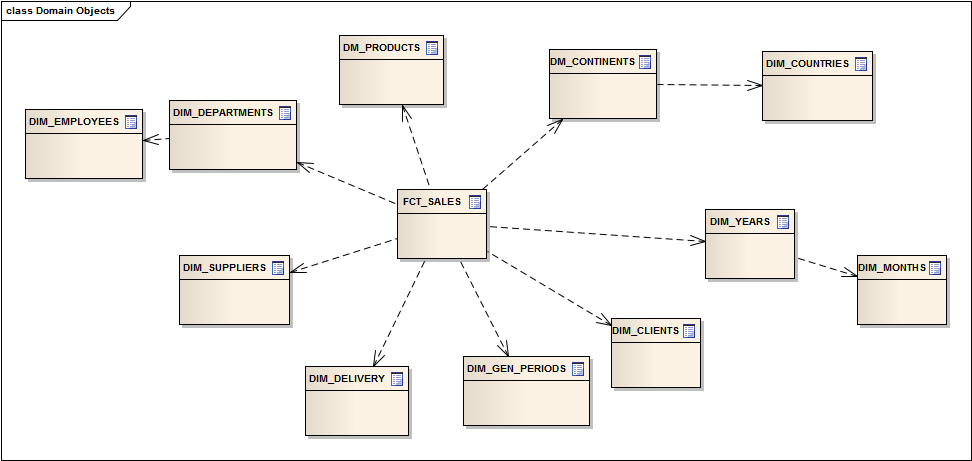
STAR



|  |  |
| --- | --- |
| DIM | DESC |
| DM\_STUCTURES | Information about departments (or single agents), employees |
| DM\_TIMES | Time period for analysis: day, week, month, quarter, year |
| DM\_PRODUCTS | Category of products, description of each product, their cost |
| DM\_LOCATIONS | Countries, region, part, continent |
| DM\_CLIENTS | Detail information about clients (first name, last name, email and etc.) |
| DM\_DELIVERY | Full information about delivery products, localization of warehouse |
| DM\_SUPPLIERS | Proven suppliers of the shop |
| DM\_GEN\_PERIODS | Contain information of different values for different periods of time |

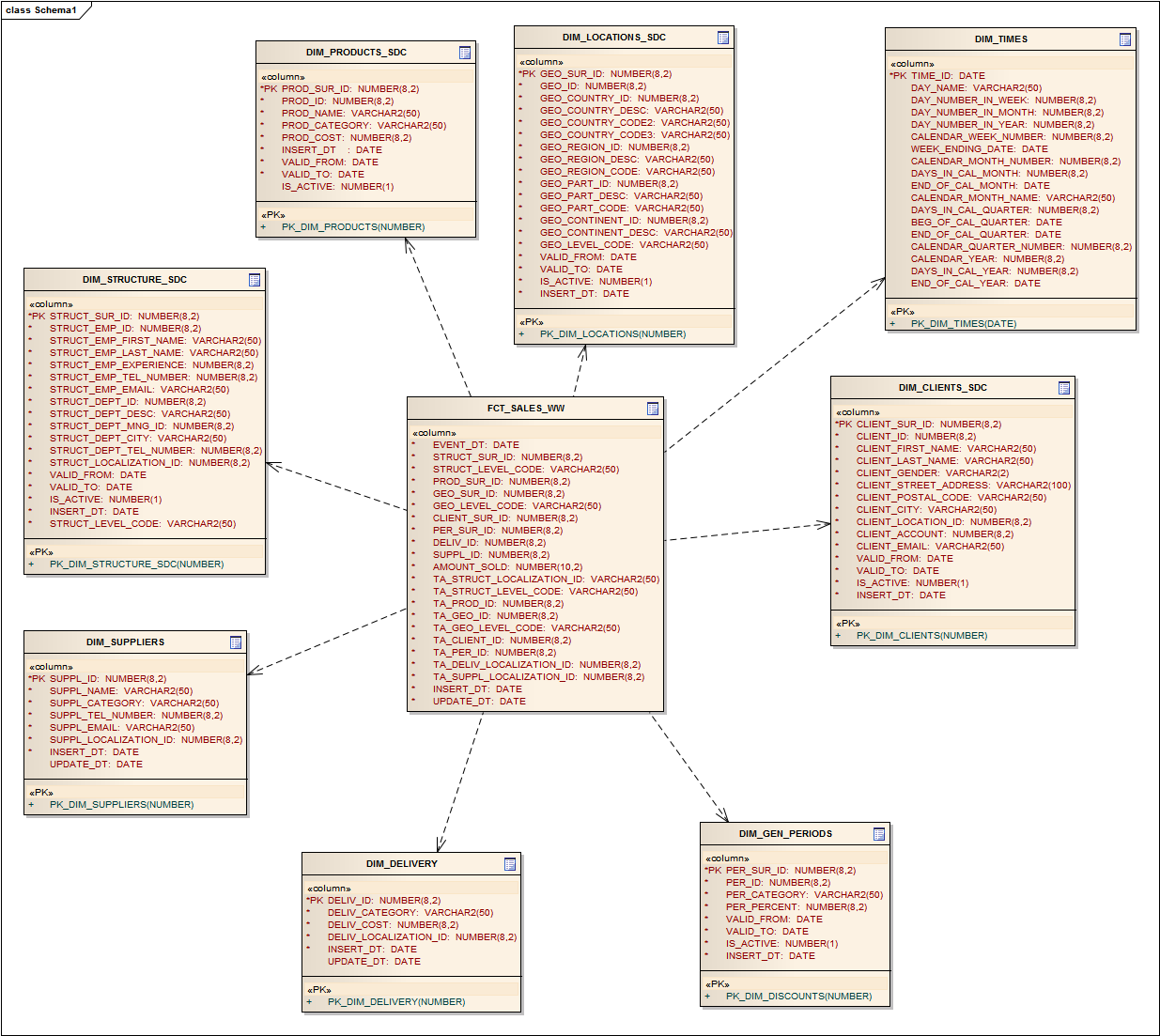
## 

## SNOWFLAKE



Need additional joins, so it would be bad in performance.

## Physical diagram



## Dimensions

### Dimension Types

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Type | Size | DW – Merged Dimensions | Descriptions |
| DIM \_TIMES | SCD1 | BIG | DW.T\_DAYS, DW.T\_WEEKS, DW.T\_MONTHS, DW.T\_QUARTERS,  DW.T\_YEARS | Dimension table with list of time parameters for analytics |
| DIM\_LOCATIONS | SCD2 | SMALL | DW.T\_COUNTRIES  DW.T\_CNTR\_GROUPS  DW.T\_CNTR\_SUB\_GROUPS  DW.LC\_CNTR\_GROUPS  DW.T\_GEO\_TYPES  DW.T\_GEO\_SYSTEMS  DW.LC\_GEO\_SYSTEMS  DW.T\_GEO\_PARTS  DW.T\_GEO\_REGIONS  DW.T\_GEO\_OBJECTS  DW.T\_CNTR\_GROUP\_SYSTEMS  DW.LC\_CNTR\_GROUP\_SYSTEMS  DW.LC\_CNTR\_SUB\_GROUPS  DW.LC\_ GEO\_PARTS  DW.LC\_COUNTRIES  DW.LC\_ GEO\_REGIONS | Dimension table with full information about all countries, regions, subregions, economic types and some other classifications etc. |
| DIM\_GEN\_PERIODS | SCD2 | SMALL | DW.T\_DISCOUNTS | Dimension table with information for different periods |
| DIM\_DELIVERY | SCD1 | SMALL | DW.T\_DELIVERY | SCD2 Type dimension with list of all currencies and types of them (convertible or not) |
| DIM\_SUPPLIERS | SCD1 | SMALL | DW.T\_SUPPLIERS | Dimension table with details information about supplier |
| DIM\_STRUCTURE\_SDC | SCD2 | SMALL | DW.T\_EMPLOYEES  DW.T\_DEPARTMENTS | Contain all information about employees and their departments. |
| DIM\_PRODUCTS\_SDC | SCD2 | BIG | DW.T\_PRODUCTS | Full information about product |
| DIM\_CLIENTS\_SDC | SCD2 | BIG | DW.T\_CLIENTS | Need information about client |

### Dimension Hierarchies

**DIM\_ TIMES:**

**Hierarchy DAY-WEEK-MONTH-QUARTER-YEAR**

|  |  |  |  |
| --- | --- | --- | --- |
| Name | LEVEL\_CODE | LEVEL\_DESC | LEVEL\_NATURAL\_KEY |
| DAYs | DAY | Store day at the calendar | DAY\_ID |
| WEEKs | WEEK | Store weeks at the calendar year | WEEK\_ID |
| MONTHs | MONTH | Store months at the calendar year | MONTH\_ID |
| QUARTERs | QUARTER | Store quarters at the calendar year | QUARTER\_ID |
| YEARs | YEAR | Store years at the calendar year | YEAR\_ID |

**DIM\_LOCATIONS:**

**Hierarchy COUNTRY –REGION -- CONTINETS**

|  |  |  |  |
| --- | --- | --- | --- |
| Name | LEVEL\_CODE | LEVEL\_DESC | LEVEL\_NATURAL\_KEY |
| COUNTRIES | COUNTRY\_NAME | All countries in regions | GEO\_COUNTRY\_ID |
| REGIONS | COUNTRY\_REGION | All regions on continents | GEO\_COUNTRY\_REGION\_ID |
| PARTS | COUNTRY\_PART | All part in the world | GE0\_COUNTRY\_PART\_ID |
| CONTINENTS | COUNTRY\_CONTINENT | All continents in the world. | GEO\_COUNTRY\_ CONTINENT \_ID |

**DIM\_STRUCTURES:**

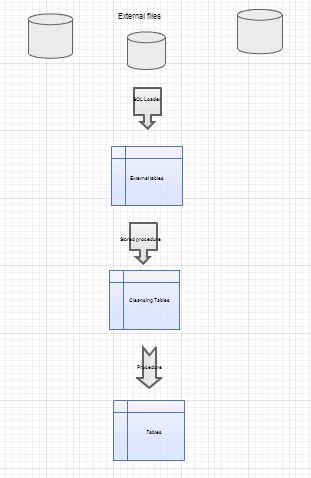
|  |  |  |  |
| --- | --- | --- | --- |
| Name | LEVEL\_CODE | LEVEL\_DESC | LEVEL\_NATURAL\_KEY |
| EMP\_ID | EMPLOYEE | Describe employee | EMP\_ID |
| DEPT\_ID | DEPARTMENT | Describe DEPARTMENT | DEPT\_ID |

## Facts

### Facts Aggregations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Code | Table Name | Additive | Descriptions |
| Calculate amount sold | AMOUNT\_SOLD | FCT\_SALES\_WW | + | Calculate amount sold for need period of time |

## Dataflow Diagram



## Partitioning rules

For the reason that information stored in four different servers, it partition.

Alter table DW.DIM\_SALES\_WW

PARTITION BY HASH (EVENT\_DT)

(

partition part\_1 tablespace TS\_DW1,

partition part\_2 tablespace TS\_DW2,

partition part\_3 tablespace TS\_DW3,

partition part\_4 tablespace TS\_DW4

);

## Strategy of Parallel execution

It would be good performance as for the query (best read from several servers), so for the DML (faster update data), so for DDL (creating or rebuild indexes).