Pear Group Solutions Company

Sales history BI Solution Proposal.

# DWH Solution Concept

## Logical Diagram



## Physical diagram



## Dimensions

### Dimension Types

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Type | Size | DW – Merged Dimensions | Descriptions |
| DIM\_GEN\_TIMES | SCD1 | BIG | DW.T\_DAYS, DW.T\_WEEKS, DW.T\_MONTHS, DW.T\_QUARTERS,  DW.T\_YEARS | It is a specific type of dimension. Appears at the DWR and contains all the time since the introduction of the project until its completion. |
| DIM\_GEO\_LOCATIONS | SCD1 | 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 | This kind of dimension contains information about all countries, subregions, regions of the world. And also enters information on the types of economic development and unions according to the international classification. |
| DIM\_CUSTOMERS | SCD1 | BIG | DW.T\_CUSTOMERS  DW.T\_GENDER  DW.T\_MARITAL\_STATUS  DW.T\_ADDRESS  DW.T\_EMAIL | This kind of dimension contains detailed information about clients (including postal code and address of the residence for informative presentation). |
| DIM\_PRODUCTS | SCD2 | BIG | DW.T\_PRODUCT\_DESC  DW.T\_PROD\_SUBCAT  DW.T\_PROD\_CAT | This kind of dimension contains detailed information about the company's products, including the name of an individual product, category and subcategory. To do so, provided the opportunity for dimension Type SCD 2 perfectly partitions history because each detailed version of a dimensional entity is correctly connected to the span of fact table records for which that version is exactly correct |
| DIM\_CHANNELS | SCD1 | SMALL | DW.T\_CHANNEL\_DESC  DW.T\_CHANNEL\_CLASS | Provides information about channels of sales (description and class) |
| DIM\_GEN\_PERIOD | SCD1 | SMALL | DW.T\_PERIOD\_DESC  DW.T\_PER\_START  DW.T\_PER\_END  DW.T\_LEVEL\_CODE | Dimension specific type, which allows grouping of facts on the basis of logic (the age of customers). |
|  |  |  |  |  |

### Dimension Hierarchies

**DIM\_GEN\_TIME:**

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

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

**DIM\_PRODUCTS:**

**Hierarchy PRODUCTS –-SUBCATEGORY-- CATEGORY**

|  |  |  |  |
| --- | --- | --- | --- |
| Name | LEVEL\_CODE | LEVEL\_DESC | LEVEL\_NATURAL\_KEY |
| PRODUCTS | PROD\_NAME | Store all possible products for each category. | PROD\_ID |
| SUBCATEGORIES | PROD\_SUBCATEGORY | Store all product subcategories for each category. | PROD\_SUBCATEGORY\_ID |
| CATEGORIES | PROD\_CATEGORY | Store all product categories of our company. | PROD\_CATEGORY\_ID |
|  |  |  |  |

**DIM\_GEO LOCATIONS:**

**Hierarchy COUNTRY –-SUBREGION--REGION**

|  |  |  |  |
| --- | --- | --- | --- |
| Name | LEVEL\_CODE | LEVEL\_DESC | LEVEL\_NATURAL\_KEY |
| COUNTRIES | COUNTRY\_NAME | Store all countries for each region. | COUNTRY\_ID |
| SUBREGIONS | COUNTRY\_SUBREGION | Store all subregions for each region . | COUNTRY\_SUBREGION\_ID |
| REGIONS | COUNTRY\_REGION | Store all regions of the world. | COUNTRY\_REGION\_ID |
|  |  |  |  |

## Facts

### Facts Aggregations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Code | Table Name | Additive | Descriptions |
| Counts amount sold products | AMOUNT\_SOLD | FCT\_SALES | + | Calculate distinct values of Sales at the event\_dt period. |
| Counts quantity sold products | QUANTITY\_SOLD | FCT\_SALES | + | Calculate distinct values of Saled Products at the event\_dt period. |

## Dataflow Diagram



# Data Warehouse Architecture

## Tablespaces

|  |  |  |  |
| --- | --- | --- | --- |
| Level Type | Object Name | Tablespace | Desctiption |
| Storage level  SA\_\* | U\_SA\_CUSTOMERS | ts\_sa\_customers\_data\_01  (AUTOALLOCATE,  SEGMENT SPACE MANAGEMENT AUTO,  LOGGING,  Size 150M,  Autoextend clause ON next 50M) | Loading from db storage system. Contains Customer, products, channels information. |
| U\_SA\_SALES | ts\_sa\_SALES\_data\_01  (AUTOALLOCATE,  SEGMENT SPACE MANAGEMENT AUTO,  LOGGING,  Size 150M,  Autoextend clause ON next 50M) | Loading from flatfile storage system. Contains information about sales. |
| DW - Cleansing Level | U\_DW\_CL | ts\_DW\_CL  (AUTOALLOCATE,  SEGMENT SPACE MANAGEMENT AUTO,  NOLOGGING,  Size 100M,  Autoextend clause ON next 50M) | LOADING from stage level system. Contains all information and prepare it for further usage (cleaning it). |
| DW – Level | U\_DW\_DATA | ts\_DW\_DATA\_01  (AUTOALLOCATE,  SEGMENT SPACE MANAGEMENT AUTO,  LOGGING,  Size 150M,  Autoextend clause ON next 50M) | LOADING data from cleansing system. Contains clean information tending to the 3rd normal form ready for preparing star schema. |
| DW– Prepare Star Cleansing Level | U\_DW \_STR\_CLS | ts\_ DW \_STR\_CLS  (AUTOALLOCATE,  SEGMENT SPACE MANAGEMENT AUTO,  NOLOGGING,  Size 150M,  Autoextend clause ON next 50M) | LOADING data from DW system. Contains views merging objects from DW level. |
| STAR - Cleansing | U\_SAL\_CL | ts\_SAL\_CL  (AUTOALLOCATE,  SEGMENT SPACE MANAGEMENT AUTO,  NOLOGGING,  Size 150M,  Autoextend clause ON next 50M) | LOADING data from DW\_CL system. Contains views from previous level but clean any redundancy. |
| STAR – Level | U\_STR\_DATA | ts\_ STR\_DATA  (AUTOALLOCATE,  SEGMENT SPACE MANAGEMENT AUTO,  LOGGING,  Size 200M,  Autoextend clause ON next 50M) | LOADING data from star cleansing system. Contains information about facts. |
| U\_STR\_REFERENCES | ts\_ STR\_REFERENCES  (AUTOALLOCATE,  SEGMENT SPACE MANAGEMENT AUTO,  LOGGING,  Size 150M,  Autoextend clause ON next 50M) | LOADING data from star cleansing system. Contains information about dimensions. |

* 1. **Object Privileges**

|  |  |  |  |
| --- | --- | --- | --- |
| Level Type | Object Name | Tablespace | Granted privileges |
| Storage level  SA\_\* | U\_SA\_CUSTOMERS | ts\_sa\_customers\_data\_01 | Loading data from db storage system customer’s, product’s, channel’s information (Read only). Hold it in its tablespace (create tables, create views on this dirty tables and push data upstairs (update upper user(ts\_DW\_CL)). |
| U\_SA\_SALES | ts\_sa\_SALES\_data\_01 | Loading data from flatfile (Read only) storage system the most secure information about sales. Hold it in its tablespace (create tables and views push data to the upper level (update objects ts\_DW\_CL) user). |
| DW - Cleansing Level | U\_DW\_CL | ts\_DW\_CL | LOADING data from stage level system (Read objects ts\_sa\_SALES\_data\_01 and ts\_sa\_customers\_data\_01). Hold it in its tablespace (create tables and views and push it upstairs (update objects ts\_DW\_DATA\_01) ). |
| DW – Level | U\_DW\_DATA | ts\_DW\_DATA\_01 | LOADING data from DW - Cleansing Level (Read objects ts\_DW\_CL). Hold it in its tablespace (create tables and views and push it upstairs (update objects ts\_ DW \_STR\_CLS) ). |
| DW– Prepare Star Cleansing Level | U\_DW \_STR\_CLS | ts\_ DW \_STR\_CLS | LOADING data from DW – Level (Read objects ts\_DW\_DATA\_01). Hold it in its tablespace (create tables and views and push it upstairs (update objects ts\_SAL\_CL) ). |
| STAR - Cleansing | U\_SAL\_CL | ts\_SAL\_CL | LOADING data from DW– Prepare Star Cleansing Level (Read objects ts\_ DW \_STR\_CLS). Hold it in its tablespace (create tables and views and push it upstairs (update objects ts\_ STR\_DATA) ). |
| STAR – Level | U\_STR\_DATA | ts\_ STR\_DATA | LOADING data from STAR - Cleansing Level (Read objects ts\_SAL\_CL). Hold it in its tablespace (create tables and views ). |
| U\_STR\_REFERENCES | ts\_ STR\_REFERENCES | LOADING data from STAR - Cleansing Level (Read objects ts\_SAL\_CL). Hold it in its tablespace (create tables and views ). |

# Partitioning rules

For our project it is advisable to divide (introduce partitions) fact table by *event\_dt* by month (as the lowest division unit for analyses). For each partition it is useful to introduce 5 sub partitions by hash by channel\_id column.

PARTITION BY RANGE (time\_dt)

subpartition by hash(channel\_id) subpartitions 2

( PARTITION part\_1 VALUES LESS THAN(to\_date('01/07/20012','dd/mm/yyyy'))

( subpartition part\_1\_sub\_1,

subpartition part\_1\_sub\_2 ),

PARTITION part\_2 VALUES LESS THAN(to\_date('01/01/20013','dd/mm/yyyy'))

( subpartition part\_2\_sub\_1,

subpartition part\_2\_sub\_2 )

); (SUBPARTITION sp\_04 TABLESPACE ts\_sal\_data\_014));

# Strategy of Parallel execution

1. Using parallel query to perform SELECT operation to see some resulting data, this query could use some number of parallel sessions, break the BIG\_TABLE into small, non-overlapping slices, and then ask each parallel session to read the table and count its section of rows. It may result big savings in time, and as our system is able to process large amounts of data for fact and some dimension tables.
2. Parallel DML. On a multi-CPU machine with plenty of I/O bandwidth, the potential increase in speed may be large for mass DML operations. It is useless for OLTP systems as in an OLTP system (with a lot of users all doing short, fast transactions), you do not want to give a user the ability to fully take over the machine resources. But for OLAP systems oracle may use many parallel execution servers to perform your INSERT, UPDATE, DELETE, or MERGE instead of a single serial process. For our DWH system it may save plenty of time to update dimensions and facts.
3. Parallel DDL. With parallel execution we will have big savings in time using all hardware resources to execute following operations : create tables as select; create index; alter index rebuild; during reorganizing/updating structure of our DW and star levels.

This activities may process for better performance by automatic jobs in time whem our system undergoes a minimum load (at night for example).