

BI Solution Concept

Researching of

the World Bank funding

Document Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Updated by | Comments |
| 0.1 | 26.07.2013 | Anzhela Papova | Initial Version |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Contents

[1. Overview 3](#_Toc362612273)

[1.1. Business Background 3](#_Toc362612274)

[1.2. Benefits 3](#_Toc362612275)

[2. Requirements 4](#_Toc362612276)

[2.1. Business Requirements 4](#_Toc362612277)

[2.2. Technical Requirements 4](#_Toc362612278)

[3. Solution Sketch 5](#_Toc362612279)

[3.1. Source Tables structure 5](#_Toc362612280)

[3.2. Summarize Data Plan 7](#_Toc362612281)

[4. DWH Solution Concept 8](#_Toc362612282)

[4.1. Logical Diagram 8](#_Toc362612283)

[4.2. Physical diagram 9](#_Toc362612284)

[4.3. Dimensions 9](#_Toc362612285)

[4.3.1. Dimension Types 9](#_Toc362612286)

[4.3.2. Dimension Hierarchies 11](#_Toc362612287)

[4.4. Facts 11](#_Toc362612288)

[4.4.1. Facts Aggregations 11](#_Toc362612289)

[4.4.2. Periodic Snapshot Fact Tables 11](#_Toc362612290)

[4.5. Data Warehouse Architecture 12](#_Toc362612291)

[4.6. Dataflow Diagram 12](#_Toc362612292)

[4.7. Partitioning rules 14](#_Toc362612293)

[4.8. Strategy of Parallel execution 14](#_Toc362612294)

# Overview

The World Bank is a source of financial and technical assistance for developing countries. The World Bank is made up of two unique development institutions owned by 187 member countries — the International Bank for Reconstruction and Development (IBRD) and the International Development Association (IDA). These institutions play different but collaborative roles to advance the vision of an inclusive and sustainable globalization. The IBRD focuses on middle-income and creditworthy poor countries, while IDA focuses on the poorest countries. Together they provide low-interest loans, interest-free credits, and grants to developing countries for a wide array of purposes, including investments in education, health, public administration, infrastructure, financial and private sector development, agriculture, and environmental and natural resource management.

The current Solution Concept includes Requirements, Solution sketch and DWH Solution concept for The World Bank analyzing allocation of funds within programs and countries.

## Business Background

The aim of this project is to identify unjustified country financing by the World Bank. The World Bank accumulates funds from various sources and invests them different countries (within programs). This is matched to ambitious goals of The World Bank Group – to end extreme poverty within a generation and boost shared prosperity.

For each program are defined the purpose, duration and leader – manager who responsible for the reasonable allocation of funds within program. Manager appointed for a fixed period of time. One program can include financing of several countries. Financing of each country during one year can be carried out within one program.

The World Bank provides financial assistance to countries not more often than once a month in a single tranche. One tranche for one country has one finance source.

The possibility of financing is determined by level of the country's budget deficit. The level of the country's budget deficit is defined as the ratio of the deficit / surplus of the budget to the index of gross domestic product (GDP) for the period, expressed as a percentage. Deficit/ surplus of the budget is determined as the difference between revenues and expenses of the country’s government. Government excludes public corporations and quasi corporations (such as the central bank). Units of government at many levels meet this definition, from local administrative units to the national government.

The level of the budget deficit of the country is ranked in groups. World Bank funds cannot be allocated to countries with a budget surplus or low level of the budget deficit in the 12 months preceding the date of disbursement.

The final decision about financing is made by the program manager (by adopting the appropriate order). Funds are disbursed in the month of approval of the relevant order.

The purpose of this project is to provide customer with analytical information about country financing by the World Bank within programs (and their managers) to identify unjustified financing (to countries with a budget surplus or low level of the budget deficit in the 12 months preceding the date of disbursement).

## Benefits

Researching of the World Bank funding will help:

* to identify those program managers who made the final decision about financing countries with a budget surplus or low level of the budget deficit in the 12 months preceding the date of disbursement;
* to identify program managers who always made the correct decision about financing countries.

Thus the researching will help to optimize managers’ structure and to improve the distribution of financial resources between the countries with the objective of achieving the purposes of the World Bank.

# Requirements

## Business Requirements

Business requirements of this project are listed below.

BR1. To get information divided by periods it’s necessary to create dimension DIM\_TIME\_MM.

BR2. To get information by countries it’s necessary to create hierarchy dimension DIM\_COUNTRIES\_SCD with saving historical changes (SCD type 2).

BR3. To get information about financing by sources it’s necessary to create dimension DIM\_FIN\_SOURCES.

BR4. To get information about financing within programs and their managers it’s necessary to create dimension DIM\_PROGRAMS\_SCD with saving historical changes (SCD type 2).

BR5. To range information by groups of deficit level it’s necessary to create universal dimension DIM\_GEN\_PERIODS.

BR6. Structure of dimensions listed above should comply with Physical Diagram presented in chapter 4.2.

BR7. To storage and extract analytical information it’s necessary to create fact table FCT\_WB\_FIN\_COUNTRIES\_MM containing fact columns FIN\_AMOUNT, GDP, BUD\_DEFICIT which depend on dimensions: DIM\_TIME\_MM, DIM\_COUNTRIES\_SCD, DIM\_FIN\_SOURCES, DIM\_PROGRAMS\_SCD, DIM\_GEN\_PERIODS.

BR8. Government finance statistics (values of columns FIN\_AMOUNT, GDP, BUD\_DEFICIT) are stored and reported in USD.

BR9. Level of granularity information in the fact table is monthly.

BR10. In the fact table data should be loaded and stored since 2000.

BR11. Data should be updated monthly by overwriting data for previous 12 months.

BR12. All translatable data should be presented in English.

## Technical Requirements

Technical requirements of this project are listed below.

TR1. All objects should be named according to the Initial Guidelines of Star Aggregation Layer Modeling.

TR2. Dimensional entities SCD type 1 (except DIM\_TIME\_MM) and fact entity should contain last time changed information columns: INSERT\_DT (date of row creation) and UPDATE\_DT (date of last time row modification).

TR3. Dimensional entity DIM\_TIME\_MM should be loaded once that’s why last time changed information columns are not required.

TR4. Dimensional entities SCD type 2 should contain date of row creation in column INSERT\_DT.

TR5. Data about financing of particular country should be updated at night according to the distribution of time zones.

TR6. Data warehouse should be monthly backed up at night according to the distribution of time zones on the day before updating.

# Solution Sketch

## Source Tables structure

Data on government revenues and expenses are collected through questionnaires to member countries and by the Organization for Economic Cooperation and Development.

Schema <FINANCE> is used as source system where data is physically stored and used to analytical needs.

Source table structures are presented in tables below.

**Table 3.1 Finance\_sources**

|  |  |  |
| --- | --- | --- |
| Column\_name | Data\_type | Comment |
| Fin\_source\_ID | NUMBER | Finance source code |
| Fin\_source\_name | VARCHAR2(100) | Finance source name |

This table contains information which should be loaded in dimensional entity DIM\_FIN\_SOURCES.

**Table 3.2 Programs**

|  |  |  |
| --- | --- | --- |
| Column\_name | Data\_type | Comment |
| Program\_code | VARCHAR2(10) | Program code |
| Program\_name | VARCHAR2(100) | Program name |
| Program\_purpose | VARCHAR2(200) | Program purpose |
| Manager\_FN | VARCHAR2(50) | Program manager first name |
| Manager\_LN | VARCHAR2(50) | Program manager last name |
| Start\_date | DATE | Date of appointment |
| End\_date | DATE | Date of validity |

This table contains information which should be loaded in dimensional entity DIM\_PROGRAMS\_SCD.

Ranging information by groups of deficit level is made by analytics using special table with constant values. Structure and values which should be loaded in dimensional entity DIM\_GEN\_PERIODS are presented below.

**Table 3.3 Dimensional entity DIM\_GEN\_PERIODS**

| PERIOD\_ID | PERIOD\_ code | PERIOD\_ desc | VALUE\_ FROM\_ NUM | value\_ TO\_NUM | VALUE\_ FROM\_ DT | value\_ TO\_ DT | VALUE\_ FROM\_ CHAR | value\_ TO\_ CHAR | Level\_ code |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Lower 60% | Extremely high | -999,99 | -60,00 | 01/01/1800 | 01/01/1800 | n.d. | n.d. | Deficit level |
| 2 | 20-59,99% | High | -59,99 | -20,00 | 01/01/1800 | 01/01/1800 | n.d. | n.d. | Deficit level |
| 3 | 3-19,99% | Middle | -19,99 | -3,00 | 01/01/1800 | 01/01/1800 | n.d. | n.d. | Deficit level |
| 4 | 0-2,99% | Low | -2,99 | 0,00 | 01/01/1800 | 01/01/1800 | n.d. | n.d. | Deficit level |
| 5 | Surplus | Surplus | 0,01 | 999,99 | 01/01/1800 | 01/01/1800 | n.d. | n.d. | Deficit level |

*\*01/01/1800 – means that value of date column is not defined*

**Table 3.4 Fact\_financing**

|  |  |  |
| --- | --- | --- |
| Column\_name | Data\_type | Comment |
| Date | DATE | Date of fact financing |
| Country | VARCHAR2(100) | Country name |
| Program\_code | VARCHAR2(10) | Program code |
| Fin\_source\_ID | NUMBER | Finance source code |
| Amount | NUMBER | Disbursement amount |
| Loan\_charge | NUMBER | Loan charge, % |
| End\_date | DATE | Plan return date |

This table contains information (column AMOUNT) which should be loaded in fact entity FCT\_WB\_FIN\_COUNTRIES\_MM (column FIN\_AMOUNT).

**Table 3.5 GDP\_countries**

|  |  |  |
| --- | --- | --- |
| Column\_name | Data\_type | Comment |
| Year | VARCHAR2(4) | Year |
| Month | VARCHAR2(30) | Month name |
| Country | VARCHAR2(100) | Country name |
| GDP | NUMBER | GDP |

This table contains information which should be loaded in fact entity FCT\_WB\_FIN\_COUNTRIES\_MM (column GDP).

The next two tables are used to derive column BUD\_DEFICIT in fact entity FCT\_WB\_FIN\_COUNTRIES\_MM. Calculation procedure is presented in chapter Summarize Data Plan.

**Table 3.6 Finance\_items**

|  |  |  |
| --- | --- | --- |
| Column\_name | Data\_type | Comment |
| Fin\_item\_ID | NUMBER | Finance item code |
| Fin\_item\_name | VARCHAR2(100) | Finance item name |

**Table 3.7 Finance\_countries**

|  |  |  |
| --- | --- | --- |
| Column\_name | Data\_type | Comment |
| Year | VARCHAR2(4) | Year |
| Month | VARCHAR2(30) | Month name |
| Country | VARCHAR2(100) | Country name |
| Group | VARCHAR2(2) | Finance item group (R – revenue; E – expense) |
| Fin\_item\_ID | NUMBER | Finance item code |
| Amount | NUMBER | Revenue / expense fact amount |

## Summarize Data Plan

Summarize Data Plan presented below illustrates calculation procedure of columns BUD\_DEFICIT, GDP in fact entity FCT\_WB\_FIN\_COUNTRIES\_MM and value of budget deficit (to define its level using dimensional entity DIM\_GEN\_PERIODS) for particularly country for each month.

**GDP\_countries**.

GDP

**Finance\_countries**. Amount

**TOTAL EXPENSES**

Is defined as sum Finance\_countries. Amount where Finance\_countries. Group = ‘E’

**TOTAL REVENUES**

Is defined as sum Finance\_countries. Amount where Finance\_countries. Group = ‘R’

**FCT\_WB\_FIN\_COUNTRIES\_MM**. GDP

**FCT\_WB\_FIN\_COUNTRIES\_MM**. BUD\_DEFICIT

Is defined as TOTAL REVENUES minus TOTAL EXPENSES

VALUE to define **DIM\_GEN\_PERIODS.**PERIOD\_SURR\_ID

Value of (FCT\_WB\_FIN\_COUNTRIES\_MM. BUD\_DEFICIT \* 100 / GDP) should be between DIM\_GEN\_PERIODS.VALUE\_FROM and VALUE\_TO for LEVEL\_CODE = ‘Deficit level’

**FCT\_WB\_FIN\_COUNTRIES\_MM**. DIM\_GEN\_DEF\_LEV\_SURR\_ID

Is equal DIM\_GEN\_PERIODS.PERIOD\_SURR\_ID

Red blocks contain information from source tables; blue blocks contain information from fact table; grew blocks contain calculation rules.

# DWH Solution Concept

Final solution will be implemented using a Star Schema approach because of the next benefits from using Star Schema:

* Star Schema can easily be augmented by adding new dimensions, as long as they fit in with the fact entity;
* Star Schema rolls subset dimensions into single entities from a multiple dimensional hierarchy of a Snowflake schema. Therefore, the number of joins in queries will be reduced and queries should execute faster.

Logical and physical diagrams of solution are presented below.

## Logical Diagram

**PERIODS**

PERIOD\_ID

PERIOD

VALUE\_FROM

VALUE\_TO

**FINANCIAL SOURCES**

FIN\_SOURCE\_ID

FIN\_SOURCE

**FINANCING OF COUNTRIES**

FIN\_AMOUNT

GDP

BUD\_DEFICIT

**COUNTRIES**

COUNTRY\_ID

COUNTRY

REGION

PART

GEO\_SYSTEM

SUB\_GROUP

GROUP

GRP\_SYSTEM

**TIME**

TIME\_ID

MONTH

QUARTER

YEAR

**PROGRAMS**

PROGRAM\_ID

PROGRAM\_CODE

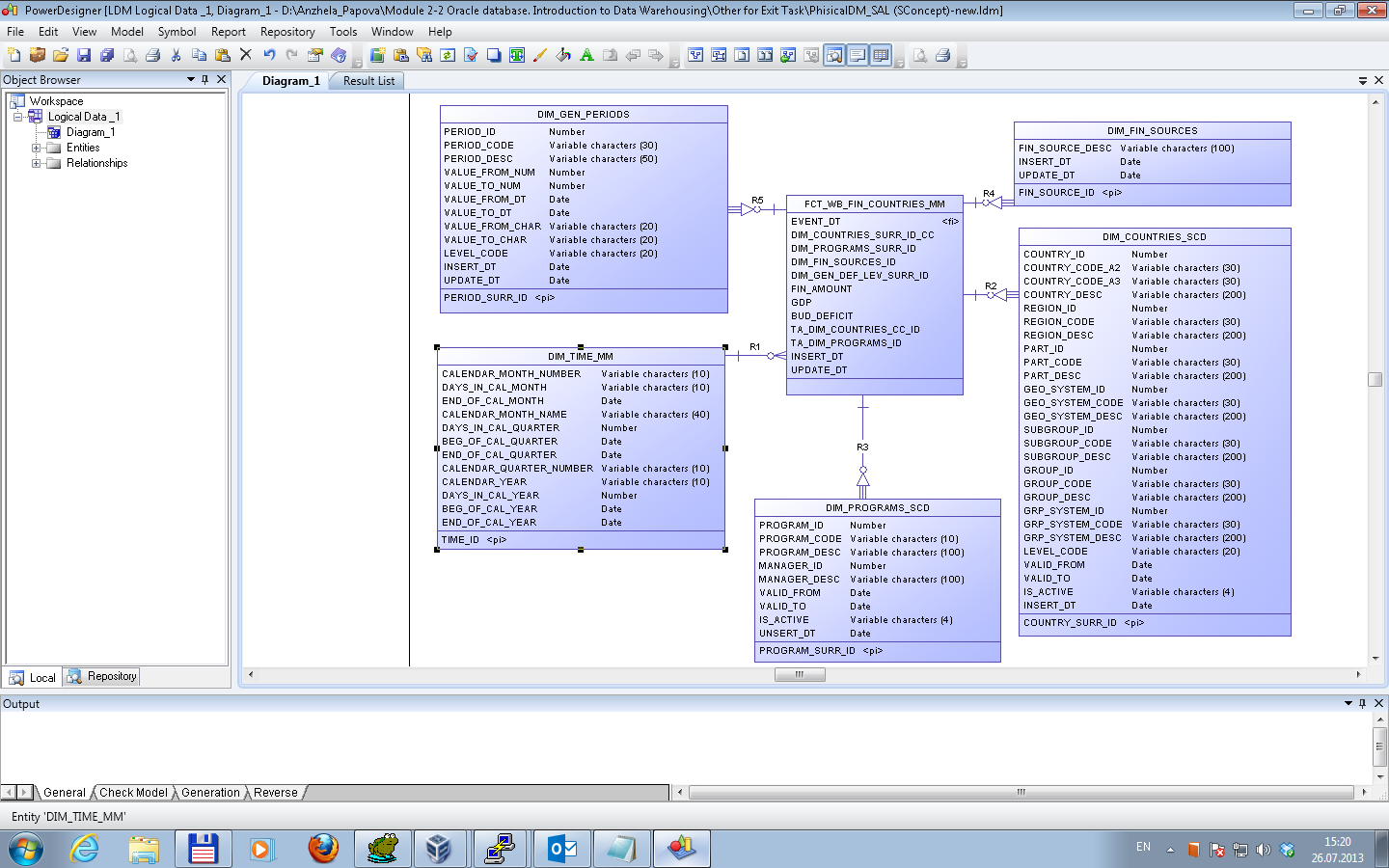
PROGRAM

MANAGER

MANAG\_VALID\_FROM

MANAG\_VALID\_TO

## Physical diagram



## Dimensions

### Dimension Types

Project Star schema contains five dimensions; each of them is described in table below.

**Table 4.1 Dimension Types**

| Name | Type | Size | DW – Merged Dimensions | Descriptions |
| --- | --- | --- | --- | --- |
| DIM\_TIME\_MM | SCD1 | BIG | DW.T\_MONTHS,  DW.T\_QUARTERS,  DW.T\_YEARS | Entity contains continuous list of months, quarters and years from 01/01/2000 to 31/12/2020 |
| DIM\_FIN\_SOURCES | SCD1 | SMALL | DW.T\_FIN\_SOURCES | Entity contains list of finance sources (Government funds, Corporations’ funds, etc.) |
| DIM\_GEN\_PERIODS | SCD1 | SMALL | DW.T\_GEN\_PERIODS | Entity contains ranged in groups levels of budget deficit (3-19,99%, 0-2,99%, surplus, etc.) |
| DIM\_COUNTRIES\_SCD | SCD2 | SMALL | DW.T\_GEO\_OBJECTS,  DW.T\_GEO\_OBJECT\_LINKS,  DW.T\_GEO\_TYPES,  DW.T\_COUNTRIES,  DW.LC\_COUNTRIES,  DW.T\_GEO\_REGIONS,  DW.LC\_GEO\_REGIONS,  DW.T\_GEO\_PARTS,  DW.LC\_GEO\_PARTS,  DW.T\_GEO\_SYSTEMS,  DW.LC\_GEO\_SYSTEMS,  DW.T\_CNTR\_SUB\_GROUPS,  DW.LC\_CNTR\_SUB\_GROUPS,  DW.T\_CNTR\_GROUPS,  DW.LC\_CNTR\_GROUPS,  DW.T\_CNTR\_GROUP\_SYSTEMS  DW.LC\_CNTR\_GROUP\_SYSTEMS | Entity contains list of countries and their regions, parts, geo\_systems, sub\_groups, groups, grp\_systems |
| DIM\_PROGRAMS\_SCD | SCD2 | SMALL | DW.T\_PROGRAMS,  DW.T\_MANAGERS | Entity contains lists of the World Bank programs (Child illnesses, Poverty, etc.) and managers who responsible for finance allocation within the program (John Smith, etc.) |

### Dimension Hierarchies

Project schema contains two hierarchy dimensions which described in the tables below.

**Table 4.2 Dimension DIM\_TIME\_MM:**

Hierarchy: MONTH-QUARTER-YEAR

|  |  |  |  |
| --- | --- | --- | --- |
| Name | LEVEL\_CODE | LEVEL\_DESC | LEVEL\_NATURAL\_KEY |
| MONTHS | MONTH | Store all months in the calendar year | MONTH\_ID |
| QUARTERS | QUARTER | Store all quarters in the calendar year | QUARTER\_ID |
| YEARS | YEAR | Store all years from 01/01/2000 to 31/12/2020 | YEAR\_ID |

**Table 4.3 Dimension DIM\_COUNTRIES\_SCD:**

Hierarchy1: COUNTRY – REGION – PART – GEO\_SYSTEM

Hierarchy2: COUNTRY – SUB\_GROUP – GROUP – GRP\_SYSTEM

|  |  |  |  |
| --- | --- | --- | --- |
| Name | LEVEL\_CODE | LEVEL\_DESC | LEVEL\_NATURAL\_KEY |
| COUNTRIES | COUNTRY | Store all countries | COUNTRY\_ID |
| REGIONS | REGION | Store all regions | REGION\_ID |
| PARTS | PART | Store all parts (continents) | PART\_ID |
| GEO\_SYSTEMS | GEO\_SYSTEM | Store all geo systems | GEO\_SYSTEM\_ID |
| SUB\_GROUPS | SUB\_GROUP | Store all subgroups | SUB\_GROUP\_ID |
| GROUPS | GROUP | Store all groups | GROUP\_ID |
| GRP\_SYSTEMS | GRP\_SYSTEM | Store all group systems | GRP\_SYSTEM\_ID |

## Facts

### Facts Aggregations

Star Fact table aggregations are presented in table below.

**Table 4.4 Facts aggregations**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Code | Table name | Additive | Descriptions |
| Total sum fin\_amount | Sum\_fin\_amount | FCT\_WB\_FIN\_ COUNTRIES\_MM | + | Summarize values of fin\_amount at the event\_dt period |

### Periodic Snapshot Fact Tables

Using periodic snapshot fact tables in the project is not required.

## Data Warehouse Architecture

Project Data Warehouse Architecture diagram is presented in table below.

**Table 4.5 Data Warehouse Architecture**

| Level Type | Object Name | Tablespace | Description |
| --- | --- | --- | --- |
| Storage level SA\_\* | SA\_Finance | ts\_sa\_finance\_data\_01  (SIZE 100M  AUTOEXTEND ON NEXT 50M  SEGMENT SPACE MANAGEMENT AUTO  LOGGING) | Loading from Flatfile storage system. Contain Customer information  User SA\_FINANCE should be able to load data from source files, create tables and view in this area |
| DW - Cleansing Level | DW\_CL | ts\_dw\_cl  (SIZE 200M  AUTOEXTEND ON NEXT 50M  SEGMENT SPACE MANAGEMENT AUTO  NOLOGGING) | User DW\_CL should be able to create tables and views, upload and select cleaned data |
| DW – Level | DW | ts\_dw\_data\_01  (SIZE 300M  AUTOEXTEND ON NEXT 100M  SEGMENT SPACE MANAGEMENT AUTO  LOGGING) | Contains cleaned and normalized data after cleansing them on DW –Cleansing level.  User DW should be able to create tables and views, select data from DW - Cleansing level |
| DW– Prepare Star Cleansing Level | SAL\_DW\_CL | ts\_sal\_dw\_cl  (SIZE 100M  AUTOEXTEND ON NEXT 50M  SEGMENT SPACE MANAGEMENT AUTO  NOLOGGING) | User SAL\_DW\_CL should be able to select data from DW - Level, create tables and views with cleaned data |
| STAR - Cleansing | SAL\_CL | ts\_sal\_sl  (SIZE 100M  AUTOEXTEND ON NEXT 50M  SEGMENT SPACE MANAGEMENT AUTO  NOLOGGING) | User SAL\_CL should be able to select data from views of DW– Prepare Star Cleansing Level, create tables and views |
| STAR – Level | SAL | ts\_sal\_data\_01  (SIZE 100M  AUTOEXTEND ON NEXT 50M  SEGMENT SPACE MANAGEMENT AUTO  LOGGING) | User SAL should be able to select data from tables and views after STAR – Cleansing Level, create tables and view for data marts |

## Dataflow Diagram

Dataflow Diagram describing refresh process with the flow of data from Stage-level to Star –level of the project is presented below.

**Storage level SA\_FINANCE**

GDP\_ countries

Programs

Finance\_ sources

Finance\_ countries

Finance\_ items

Fact\_ financing

**DW - Cleansing Level**

Packages

Procedures

Work tables, views

**DW – Level**

T\_MONTHS, T\_QUARTERS, T\_YEARS

(tables for DIM\_TIME\_MM)

T\_GEO\_OBJECTS, T\_GEO\_OBJECT\_LINKS, T\_GEO\_TYPES, T\_COUNTRIES, LC\_COUNTRIES, … (tables for DIM\_COUNTRIES\_SCD)

T\_Fact\_financing

T\_GDP\_countries

T\_MANAGERS

T\_FIN\_ SOURCES

T\_GEN\_ PERIODS

T\_Finance\_ groups

T\_Finance\_ items

T\_PROGRAMS

T\_Finance\_ countries

Work tables, views

Procedures

Packages

**DW– Prepare Star Cleansing Level**

Work tables, views

Procedures

Packages

**STAR - Cleansing**

**STAR – Level**

DIM\_FIN\_SOURCES

DIM\_ COUNTRIES\_SCD

DIM\_GEN\_PERIODS

**FCT\_WB\_FIN\_ COUNTRIES\_MM**

DIM\_TIME\_MM

DIM\_ PROGRAMS\_SCD

## Partitioning rules

Fact table of the project should be partitioned using partitioning scheme Range by EVENT\_DT (by years). Fact table partitions are presented in the table below.

**Table 4.6 Fact table partitions**

| Partition name | Condition |
| --- | --- |
| Y2000 | VALUES LESS THAN (to\_date('01/01/2001','dd/mm/yyyy')) |
| Y2001 | VALUES LESS THAN (to\_date('01/01/2002','dd/mm/yyyy')) |
| Y2002 | VALUES LESS THAN (to\_date('01/01/2003','dd/mm/yyyy')) |
| Y2003 | VALUES LESS THAN (to\_date('01/01/2004','dd/mm/yyyy')) |
| Y2004 | VALUES LESS THAN (to\_date('01/01/2005','dd/mm/yyyy')) |
| Y2005 | VALUES LESS THAN (to\_date('01/01/2006','dd/mm/yyyy')) |
| Y2006 | VALUES LESS THAN (to\_date('01/01/2007','dd/mm/yyyy')) |
| Y2007 | VALUES LESS THAN (to\_date('01/01/2008','dd/mm/yyyy')) |
| Y2008 | VALUES LESS THAN (to\_date('01/01/2009','dd/mm/yyyy')) |
| Y2009 | VALUES LESS THAN (to\_date('01/01/2010','dd/mm/yyyy')) |
| Y2010 | VALUES LESS THAN (to\_date('01/01/2011','dd/mm/yyyy')) |
| Y2011 | VALUES LESS THAN (to\_date('01/01/2012','dd/mm/yyyy')) |
| Y2012 | VALUES LESS THAN (to\_date('01/01/2013','dd/mm/yyyy')) |
| Y2013 | VALUES LESS THAN (to\_date('01/01/2014','dd/mm/yyyy')) |
| Y2014 | VALUES LESS THAN (to\_date('01/01/2015','dd/mm/yyyy')) |
| Y2015 | VALUES LESS THAN (to\_date('01/01/2016','dd/mm/yyyy')) |
| after\_Y2015 | VALUES LESS THAN (to\_date('31/12/9999','dd/mm/yyyy')) |

Fact table partitioning will allow:

* increase availability of data;
* improve administration of segments with archive data by removing them from the database;
* improve query performance.

## Strategy of Parallel execution

Fact table of the project is quite small (number of rows will increase every year by nearly 3000 – 250 countries with maximum 12 rows by year). Queries will take seconds or less to complete serially, so the introduction of parallel execution and its associated managerial overhead will likely make the entire thing take longer.

In this case the best way to allow Oracle to automatically parallelize processes with optimal degree of parallelism.