# Initial Guidelines of Star Aggregation Layer Modeling

The core objectives of building and implementing naming standard in an enterprise are:

* Business as well as technical users should be able to describe any data entity or data element just by looking at its name. Users can be internal as well as external (vendors) to the organization.
* The name decided by more than one professional for an entity or a data element should be same if they are exposed to same business and technical descriptions of the data asset.

Describing and naming data correctly is critical. If it is done right, it can help an enterprise:

* Minimize misunderstandings among business functions, which can reduce the amount of total effort needed in a BI/DW project.
* Facilitate operational efficiency and strategic use of the data.
* Reduce time to deploy new business challenges.
* Build Mapping business Dictionary to technical realizations.

The main goal of a DB objects Name convention chapter is to standardize all naming technics overall development process. Next naming standards will be described below:

* Table Naming Conventions
* Dimensions
* Slowly Changing Dimensions
* Facts Tables
* Aggregates
* Commenting

## Table Naming Conventions

The common Data warehouse objects for Star aggregation and DW layer have to be named as shown below:

|  |  |  |  |
| --- | --- | --- | --- |
| N | **Table category** | **Layer** | **Naming templates** |
| 1 | **Dimensions SCD Type 1** | SAL | **DIM\_<Name in Plural>**  *DIM\_BOOKS* |
| 2 | **Dimensions SCD Type 2** | SAL | **DIM\_<Name in Plural>\_SCD**  *DIM\_BOOKS\_SCD* |
| 4 | **Dimensions: General** | SAL | **DIM\_GEN\_<Name in Plural>**  *DIM\_GEN\_PERIODS*  *DIM\_GEN\_PARAMS*  *DIM\_GEN\_INDICATORS* |
| 5 | **Dimensions: Calendar** | SAL | **DIM\_TIME\_<Level>**  *DIM\_TIME\_DAY*  *DIM\_TIME\_MM* |
| 6 | **Fact tables** | SAL | **FCT\_<Name in Plural>\_<Level>**  *FCT\_BOOKS\_DD* |
| 9 | **Metadata tables** | SAL | **MTA\_<Name in Plural>**  *MTA\_DWH\_INSTANCES* |
| 10 | **Work tables** | DW | **WRK\_<Name in Plural>**  *WRK\_ BOOKS* |
| 11 | **Lookup tables** | DW | **LKP\_<Name in Plural>**  *LKP \_BOOK\_TYPES* |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

The columns have to match rules below:

|  |  |  |  |
| --- | --- | --- | --- |
| N | **Objectcategory** | **Layer** | **Rules** |
| 1 | **Date Columns** | SAL | Postfix:**<Column Name>\_DT**  *EVENT\_DT* |
| 2 | **All Columns**  Except KPI (Metrics) | SAL | Column Attribute: **ARE NOT NULL**  *BOOK\_IDNUMBER NOT NULL* |
| 3 | **Time series** | SAL | Use 2 columns to describe period:  *from* **START\_DT***till* **END\_DT** |
| 4 | **Natural Keys**  ID’s from source systems | DW | Source: <Column Name>\_ID  DW: **<ColumnName>\_SRCID** |
| 5 | **Technical Attributes** | SAL  DW | Prefix: **TA\_<Column Name>**  *TA\_VER\_INSERT\_DT* |
| 6 | **ID’s columns** | SAL | Column Attribute:**NUMBER** |

Whitelist object abbreviations:

|  |  |  |  |
| --- | --- | --- | --- |
| N | **Abbreviation** | **Category** | **Description** |
| 1 | **ACT** | Postfix | **Actual Data** |
| 2 | **AVG** | Postfix | **Average KPI (Metrics)** |
| 3 | **BUS** | Prefix | **Business information** |
| 4 | **CNT** | Postfix | **Count KPI (Metrics)** |
| 5 | **DT** | Postfix | **Data & Time** |
| 6 | **EOP** | Postfix | **End of Period** |
| 7 | **GRP** | Prefix | **Group data** |
| 8 | **NO**  **NUM** | Postfix | **Number or Numeration** |
| 9 | **OPT** | Prefix | **Options** |
| 10 | **SRC** | Postfix | **Source system data** |
| 11 | **SOP** | Postfix | **Start of Period** |
| 12 | **TOT** | Postfix | **Total** |

## Data Warehouse objects: Dimension SCD Type 1

The common script design for SCD Dimensions Type 1 without complex hierarchies has looks like:

CREATE TABLE **DIM\_<Name in Plural>**(  
**<Name>\_ID** NUMBER NOT NULL,  
**<Name>\_CODE** VARCHAR2(<n>) NOT NULL,  
**<Name>\_DESC** VARCHAR2(<m>) NOT NULL,  
 ...  
**INSERT\_DT** DATE NOT NULL,  
**UPDATE\_DT** DATE NOT NULL  
);

Next rules are obligatory:

|  |  |  |
| --- | --- | --- |
|  | **Example** | **Description** |
| 1 | <Name>**\_ID** <Name>**\_CODE** <Name>**\_DESC** | Eachdimensions tables has obligatory columns \_ID, \_DESC. (Optional column \_CODE) |
| 2 | Table**DIM**\_<Name in Plural> <Name>**\_ID**  *Example:*  *Table* ***DIM\_BOOKS***  ***BOOK\_ID*** | Main hierarchy columns names have correspond to table name in the singular. |
| 3 | **INSERT\_DT** DATE  **UPDATE\_DT**DATE | The tables has obligatory columns last time changed information columns:  **INSERT\_DT –** Date of row creation  **UPDATE\_DT –**Date of last time row modification |
| 4 | <Name>**\_ID**  **-99“n.d.” Not Defined** **-98 “n.a.” Not** **Available** | Every dimension has two obligatory default rows:  -98 A default value for the case when the data a generally sensible, but not available at the current case  -99 A default value for the case when values do not make sense |

The script design for SCD Dimensions Type 1 with hierarchies has looks like:

CREATE TABLE **DIM\_SAMPLES** (  
 SAMPLE**\_ID** NUMBER NOT NULL,  
 SAMPLE**\_CODE** VARCHAR2(10) NOT NULL,  
 SAMPLE**\_DESC** VARCHAR2(100) NOT NULL,  
 ...  
 SAMPLE\_**GROUP\_ID** NUMBER NOT NULL,  
 SAMPLE\_**GROUP\_DESC** VARCHAR2(100) NOT NULL,  
 SAMPLE\_**GROUP\_OWNER** VARCHAR2(400) NOT NULL,  
 SAMPLE\_**GROUP\_TOPIC** VARCHAR2(400) NOT NULL,  
 ...  
 SAMPLE\_**FAMILY\_ID** NUMBER NOT NULL,  
 SAMPLE\_**FAMILY\_CODE** VARCHAR2(10) NOT NULL,  
 SAMPLE\_**FAMILY\_DESC** VARCHAR2(100) NOT NULL,  
 ...  
 INSERT\_DT DATE NOT NULL,  
 UPDATE\_DT

Default Hierarchy level names must follow template below:

* CLASS
  + CATEGORY
    - TYPE
      * FAMILY
        + GROUP

SUBGROUP

ITEMS

## Data Warehouse objects:Slowly Changing Dimension

The common script design for SCD Dimensions Type 2:

CREATE TABLE **DIM\_<Name in Plural>\_SCD** (  
<Name>\_**SURR\_ID**NUMBER NOT NULL,  
<Name>\_**ID** NUMBER NOT NULL,  
<Name>\_**CODE**VARCHAR2(<n>) NOT NULL,  
<Name>\_**DESC**VARCHAR2(<m>) NOT NULL,  
...  
<Name>\_**<HIER>\_ID** NUMBER NOT NULL,  
<Name>\_**<HIER>\_DESC** VARCHAR2(<m>) NOT NULL,  
<Name>\_**<HIER>\_OWNER** VARCHAR2(<m>) NOT NULL,  
...  
**START\_DT** DATE NOT NULL,  
**END\_DT** DATE NOT NULL,  
**IS\_ACTIVE**VARCHAR2(4) NOT NULL,  
**INSERT\_DT** DATE NOT NULL  
);

Next rules are obligatory:

|  |  |  |
| --- | --- | --- |
|  | **Example** | **Description** |
| 1 | <Name>**\_SURR\_ID** <Name>**\_ID** <Name>**\_CODE** <Name>**\_DESC** | Eachdimensions tables has obligatory columns \_ID, \_DESC. (Optional column \_CODE) |
| 2 | Table**DIM**\_<Name in Plural> <Name>**\_SURR\_ID** <Name>**\_ID**  *Example:*  *Table* ***DIM\_BOOKS***  ***BOOK\_SURR\_ID***  ***BOOK\_ID*** | Main hierarchy columns names have correspond to table name in the singular. |
| 3 | **START\_DT** DATE **END\_DT**  DATE **IS\_ACTIVE**VARCHAR2(4) **INSERT\_DT** DATE | The tables has obligatory columns for application of the slow changing functionality:   * **START\_DT –** Start Date of period in which row values was in Active status. Default value: 01/01/1990 , “DD/MM/YYYY” * **END\_DT –**End Date of period in which row values was in Active status. Default value: 31/12/9999 , “DD/MM/YYYY” * **IS\_ACTIVE** –Indicator of row values status at the current “storage” day. Default values: Y / N * **INSERT\_DT –**  Date of row creation (Initial date) |
| 4 | <Name>**\_SURR\_ID**  **-99“n.d.” Not Defined -98 “n.a.” Not Available** | Every dimension has two obligatory default rows:   * -98 A default value for the case when the data a generally sensible, but not available at the current case * -99 A default value for the case when values do not make sense |
| 5 | <Name>**\_SURR\_ID  SEQ**\_<Name>\_**SURR\_ID.**NEXTVAL | All surrogatekeys have to correspond with unique sequence. Surrogate keys have to be filled by new values only from sequences. |
| 6 | <Name>**\_ID** <Name>**\_CODE** <Name>**\_DESC** | Each Hierarchy has obligatory columns **\_ID** and **\_DESC.** Column **\_CODE**is optional for all levels. |
| 7 | Table**DIM**\_<Name in Plural> <Name>\_**GROUP**\_ID  <Name>\_**GROUP**\_LINK | Attributes that refer to a specific hierarchy has hierarchy abbreviation or specific prefix. |
| 8 | Default PK: <Name>\_SURR\_ID NUMBER  VALID\_TODATE | Default Primary Key:  <Name>**\_SURR\_ID** NUMBER NOT NULL**,**  **VALID\_TO**DATE NOT NULL |

## Data Warehouse objects: Facts table

The common script design for fact table has look like figure below:

CREATE TABLE FCT\_SAMPLE\_MM

(

**EVENT\_DT** DATE NOT NULL,

<DIM NAME>**\_**ID NUMBER NOT NULL,

<DIM NAME>**\_SURR\_ID** NUMBER NOT NULL,

<DIM NAME>**\_SURR\_ID**\_<LVL> NUMBER NOT NULL,

...

**FCT\_**<NAME>\_<UNIT> NUMBER,

**FCT\_USG\_SECONDS** NUMBER,

**FCT\_USG\_CNT\_SMS** NUMBER,

...

**INSERT\_DT** DATE NOT NULL,

**UPDATE\_DT** DATE NOT NULL

);

Next rules are obligatory:

|  |  |  |
| --- | --- | --- |
|  | **Example** | **Description** |
| 1 | **EVENT\_DT** DATE NOT NULL | Eachfacttables (type: timing fact aggregation) has obligatory column **EVENT\_DT**. This column store main time key – date of event. The hierarchy of time dimension **EVENT\_DT** should solve by parsing postfix of Fact table name.  FCT\_SAMPLE\_MM **-> MM** ->**Monthly granularity.**  **EVENT\_DT**store Months. |
| 2 | <DIM NAME>**\_SURR\_ID** <DIM NAME>**\_ID** | Fact table dimension keys should have “abstract reference” to Dimension table only by **\_ID** or **\_SURR\_ID**columns.  **For Star Aggregation layer (SAL)**: Creation of DB constraint references are strongly **PROHIBITED.**  *ADD CONSTRAINT fk\_fct\_books2dim\_book FOREIGN KEY (book\_id)*  *REFERENCES DIM\_BOOK*  *t\_lng\_scopes (book\_id);*  **NOT ALLOWED!** |
| 3 | **DIM\_<Name in Plural>\_SCD**  <DIM NAME>\_**SURR\_ID**  <DIM NAME>\_**SURR\_ID\_<LVL>** | In case of using hybrid SCD type 2 dimensions with different Hierarchies level add additional postfix to the column name <LVL>. Postfix should correctly solve level of hierarchy in hybrid Dimension = Level Code.  *DIM\_CHANNEL\_SURR\_ID - Low Hier Level  Channel ID*  *DIM\_CHANNEL\_SURR\_ID\_ICH – Hier: Initial Init\_Channel\_ID*  *DIM\_CHANNEL\_SURR\_ID\_OCH – Hier: Operational Oper\_Channel\_ID* |
| 4 | **INSERT\_DT** DATE  **UPDATE\_DT** DATE | The tables has obligatory columns last time changed information columns:  **INSERT\_DT –** Date of fact row creation  **UPDATE\_DT –**Date of fact last time row modification |