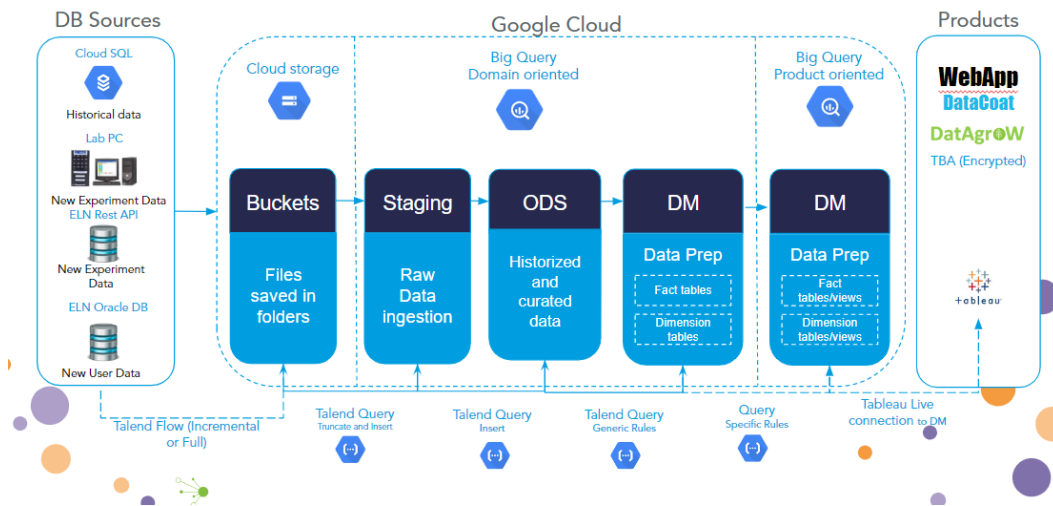


[TARGET 2023] ALB Data Architecture - General

In order to reach more standardized practices, improving the abilities on product development and maintenance, Application Labbooster is target standard data architecture principles.

The Data Architecture principles must provide an efficient data structuration to serve the business needs, not only for the product under development, but also for all R&I function, under the vision to be integrated in a next future into the Solvay Data Ocean. The Data Architecture brings common definitions on Staging, ODS and Data Mart Data structuration stages across the dev team, and through continuous improvement, it capitalizes on the good practices (for instance data compression, table partitions, star schema)

Target Data Architecture for ALB



DEFINITIONS AND CONVENTIONS:

Buckets:

Buckets

Staging:

This area is used for consolidating and load Buckets into the GCP.

The consolidation will guarantee that we have one staging table per bucket set and avoid duplicate records/information.

Also use to making some normalization process.

Data Conversion:

- Copy data AS IS from source (Buckets)

Views and Materialized Views:

- in this area we don't have VIEWS or MT Views

ODS:

Operational data store is a repository that provides a snapshot of the latest data from multiple transactional systems for operational reporting.

The ODS will contain historical data for all the extraction coming from the different sources.

Group all the periods extracted to be stored in the google cloud storage.

Data Mart:

Structure / access pattern specific to [data warehouse](#) environments, used to retrieve client-facing data.

The data mart is a subset of the data warehouse and is usually oriented to a specific business line or team.

Data set name

- DM_projectname

Data Mart Tables name

Table Type	Table Name	Description
Fact Table	FCT_xxxx	consists of the measurements, metrics or facts of a business process .
Dimension Table	DIM_xxxx	Contains relatively static data which can change slowly but unpredictably, rather than according to a regular schedule.

Data Mart Rules

- Full historical data
- Data will be managed in dimensions and facts
- Data will be cleaned & integrated at enterprise level
- All the ELT need to be implemented with Talend Tool
- Only use approved acronyms which are known in the organisation

Views and Materialized Views:

- No business rules implemented on the views, the views at this level are only for consulting data.
- All views should be suffixed with "v_" and materialised views to be suffixed with "mv_"