



ETL Data warehouse

VECNACARES

Data warehouse concepts



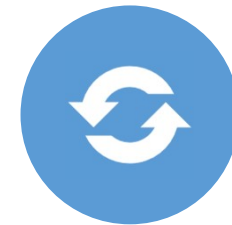
STAR SCHEMA
DESIGN



DIMENSIONS



FACTS



ETL

VECNACARES

Star Schema



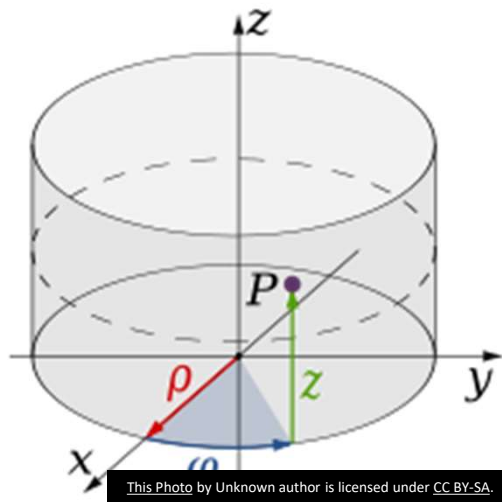
A star schema is composed of one or more central fact tables, a set of dimension tables, and the joins that relate the dimension tables to the fact tables



Advantages:

- Query Performance
- Load Performance and administration
- Built-in referential integrity
- Simplicity

Dimensions



- Dimension tables store descriptions of the characteristics of a business. A dimension is usually descriptive information that qualifies a fact. Dimensions do not change, or change slowly over time. The shape of dimension tables is typically wide and short because they contain few records and many columns

Slowly Changing Dimensions



SCD1: Overwrite old data with new data. Does not track history



SCD2: Tracks history by adding a new row for updated data



SCD3: Modifies tables structure by adding new columns. Tracks limited history



SCD 4: Tracks history using history table

Facts

A fact table contains data columns for the numeric measurements of a business. It also includes a set of columns that form a concatenated, or composite key. Each column of the concatenated key is a foreign key drawn from a dimensional table primary key. Fact tables usually have few columns and many rows, which result in relatively long and narrowly shaped tables.

Extract Transform Load



REST DATA
BETWEEN TASKS



LOAD DATA
INCREMENTALLY



ENFORCE
IDEMPOTENCY



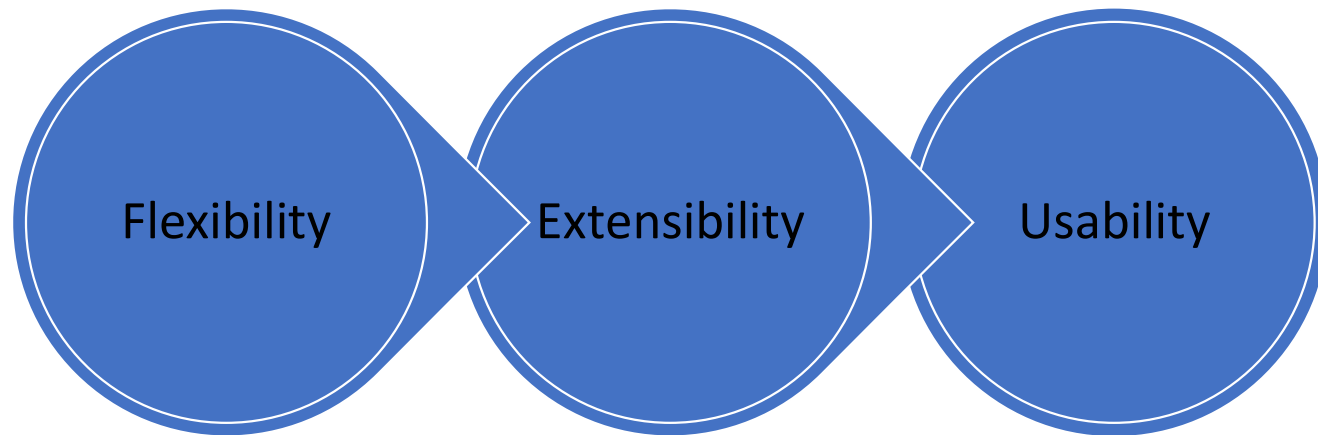
PARALLELIZE
TASKS

VECNACARES

ETL Tools

- Airflow
- Luigi
- Pentaho
- CloverETL

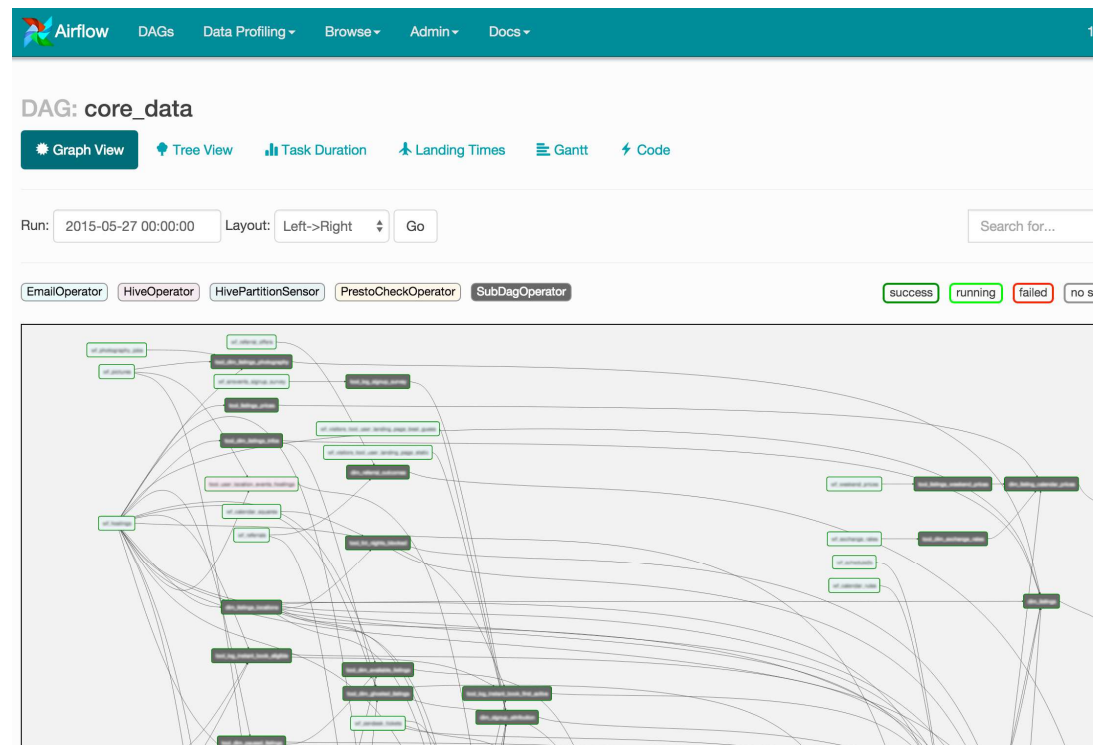
Design goals



VECNACARES

Introduction to Airflow

Airflow is a platform to programmatically author, schedule and monitor workflows defined as directed acyclic graphs (DAGs) of tasks. The airflow scheduler executes your tasks on an array of workers while following the specified dependencies



Future Work



CREATING
VIEWS



CREATING
REPORTS



CREATING UIS



ACCESS
CONTROL



AUDITING

VECNACARES

Fin

Q&A

Please contact:

Sameer.Motwani@Vecna.com

Faiith.Nassiwa@Vecna.com

Paul.amendola@Vecna.com

VECNACARES