

DATA ARCHITECTURE

This part of the project comprehends the models and standards we need to define which data needs to be collected and its format, how it is stored, arranged, integrated, and used for machine learning predictions and business analysis.

- **Architecture and schema**

- Creation of a data model of the structure of our Data Warehouse, where the extracted and transformed data will be stored.
- Definition of attributes and their data types and their relationship and connection through foreign keys (referential integrity and entity integrity).
- Deliverables are a DWH logical data model and the SQL script for the creation of the tables

- **1st-stage-op-dbs – Connect to databases with raw data**

Make connections to MONGODB and INFLUXDB databases to get raw data. Read and write to them to validate a successful connection.

- **2nd-stage-dwh – Connect to Snowflake**

Make connections to DWH (Snowflake) from Python environment to read, write and guarantee data access. Snowflake offers a scalable data warehouse as a service from cloud and it is the data platform where the company stores all the data.

- **3rd-stage-etl – Extract data, Transform data, Load data**

Based on the new database schema and structures designed and developed in the “Architecture and schemas” stage, the data is transformed to meet the requirements and characteristics needed to be imported. Once the values are correct, each dataset (ID connector, Playout connector, and the Sessions connector) is sent and stored in the cloud Snowflake database.