Ivan

Software Engineer, Sofia, Bulgaria

Responsible for developing innovative solutions and ensuring their successful technical delivery

Tasks

- Create new collector ID in order to identify the data source in Super Collider
- Create data pipelines for ingestion data coming from different sources
- Validate that created tables follow the correct data schema, before promoting to production
- Ensure the new data schema changes are promoted to production
- Monitor ingestion and fix errors if there are any that occur during the ingestion process.

Goals

- Design correct, convenient and effective data schema to ensure easier analytics later on.
- Ensure that data is ingested successfully and with correct schema in Super Collider
- Monitor the ingestion process and ensure there are no ingestion errors.

Pain Points

- No clear guidance to understand data regulation requirements and how to make sure we comply with them.
- Hard to understand the significance of the Super Collider specific columns in the data schema.

Emma

Data Engineer, Sofia, Bulgaria

Responsible for developing, testing and maintaining data pipelines and data architectures

Tasks

- Bring data from multiple sources to ensure service 360 view.
- Develops sophisticated models that later enables data analysts to put data into action
- Create data jobs to automate the data consolidation and data model creation
- Monitor, debug and maintain data jobs and produced data sets

Goals

- Provision clean, structured, quality data for data analysts
- Maintain high quality datasets and ensure data is up-to-date
- Ensure consumers of my datasets understand its meaning and are aware of any upcoming changes to the datasets and the data model.

Pain Points

- Difficulty on some occasions to find and understand the data in data lakes and data marts.
- No easy way to ensure the data you are using is up-to-date.
- Inability to effectively communicate datasets schema and lifecycle changes between producers and consumers.

Ben

Data Analyst, Palo Alto, California

Responsible for collecting, organizing and interpreting statistical information to help business and stakeholders use it and make decisions.

Tasks

- Explore the data catalog to find if relevant datasets that are needed for a particular report, are already available and how they are structured.
- Extract and aggregate relevant data from the datasets to build accurate and actionable reports.
- Visualize data to help its easy consumption and facilitate the decision-making process.
- Monitor and troubleshoot already created reports.

Goals

- Enable the business to make data-driven decisions by building reports that provides full visibility of the service performance and customer consumption behavior.
- Educate reports' consumers to understand the data, so that they can make data driven decisions.
- Support the assessment of already implemented business decisions, by analyzing the data.

Pain Points

- No clear guidance who the owner of the data is and how to contact them.
- Difficulty on some occasions to find and understand the data in data lakes and data marts.
- Timeout queries on big data, which prevent us to create big data reports to production and force us to find workarounds or redefine requirements.