ITSM Analytics Pipeline

This project implements a complete data pipeline for analyzing IT Service Management (ITSM) ticket data from ServiceNow. The pipeline includes data ingestion, transformation, and visualization using a modern data stack: Apache Airflow, DBT, PostgreSQL, and Apache Superset.

Architecture

The pipeline follows a typical modern data architecture pattern:

1. **Data Ingestion**: Raw CSV data is loaded into PostgreSQL using an Airflow DAG
2. **Data Transformation**: DBT models clean, transform, and aggregate the data
3. **Data Visualization**: Apache Superset connects to the transformed data to provide analytics dashboards

Project Structure

Copy

```
itsm-analytics/
├── airflow/
│   └── dags/
│       └── itsm_data_pipeline.py
├── dbt/
│   ├── models/
│   │   ├── staging/
│   │   │   └── stg_tickets.sql
│   │   └── marts/
│   │       ├── resolution_time_by_category.sql
│   │       ├── closure_rate_by_group.sql
│   │       ├── monthly_ticket_summary.sql
│   │       └── open_tickets_by_priority.sql
│   ├── dbt_project.yml
│   └── schema.yml
├── superset/
│   └── dashboards/
│       └── itsm_performance_dashboard.json
└── README.md
```

Setup and Installation

Prerequisites

- Docker and Docker Compose
- Git

- Python 3.8+

Step 1: Clone the Repository

bash   Copy

```bash
git clone https://github.com/your-username/itsm-analytics.git
cd itsm-analytics
```

Step 2: Configure Environment

Create a `.env` file with the necessary configurations:

Copy

```
POSTGRES_USER=postgres
POSTGRES_PASSWORD=postgres
POSTGRES_DB=itsm
AIRFLOW_UID=50000
```

Step 3: Start the Services

bash   Copy

```bash
docker-compose up -d
```

This will start PostgreSQL, Airflow, and Superset.

Step 4: Initialize DBT

bash   Copy

```bash
cd dbt
dbt deps
dbt seed --profiles-dir .
```

Step 5: Configure Airflow Connections

1. Navigate to http://localhost:8080 (default credentials: airflow/airflow)

2. Go to Admin -> Connections

3. Add a new PostgreSQL connection:
   - Conn Id: postgres_default
   - Conn Type: Postgres
   - Host: postgres

- Schema: itsm

- Login: postgres

- Password: postgres

- Port: 5432

Step 6: Configure Airflow Variables

In the Airflow UI, go to Admin -> Variables and add:

- Key: itsm_data_path Value: /opt/airflow/data

- Key: dbt_project_path Value: /opt/airflow/dbt

Step 7: Import Superset Dashboard

1. Navigate to http://localhost:8088 (default credentials: admin/admin)

2. Go to Dashboards -> Import Dashboards

3. Upload the `superset/dashboards/itsm_performance_dashboard.json` file

Running the Pipeline

Manual Execution

To manually trigger the pipeline:

1. In the Airflow UI, navigate to DAGs

2. Find the `itsm_data_pipeline` DAG

3. Click the "Play" button to trigger it

Scheduled Execution

The DAG is configured to run daily at 1:00 AM. You can modify the schedule in the DAG file.

DBT Models

Staging Layer

- `stg_tickets`: Cleans and standardizes the raw ticket data

Marts Layer

- `resolution_time_by_category`: Calculates average resolution time by category and priority

- `closure_rate_by_group`: Calculates the ticket closure rate by assigned group

- `monthly_ticket_summary`: Creates monthly aggregations of ticket metrics

- `open_tickets_by_priority`: Provides a view of currently open tickets by priority

Dashboard

The Superset dashboard includes:

1. **Ticket Volume Trends**: Line chart showing daily ticket creation
2. **Resolution Time**: Bar chart comparing average resolution time across Categories
3. **Closure Rate**: Pie chart of closure rate by Assigned Group
4. **Ticket Backlog**: Table displaying open tickets grouped by Priority
5. **Filters**: Week, Category, and Priority filters

Assumptions

1. The CSV data follows a consistent format with the specified columns
2. PostgreSQL is used as the data warehouse
3. Tickets have a linear lifecycle with defined status values
4. The "Resolved Date" and "Resolution Time" fields may contain null values for open tickets