

FLIGHT DATA ANALYSIS

A BIG DATA ENGINEERING PROJECT

INTRODUCTION

- This project explores the analysis of flight data for the year 2020-2022, focusing on flight delays
- The analysis leverages a comprehensive dataset encompassing various flight details like departure and arrival times, delays, cancellations, origin and destination airports, airlines, and more. By analyzing this data, we aim to gain valuable insights into airline operations and flight performance metrics.

AIM

Design the infrastructure of a Big Data
Application for Flight Data Analysis.

PROBLEM DEFINITION

Handling and analyzing large volumes of flight schedule data
for Flight schedules, flight Delay Analysis and Airline Delays

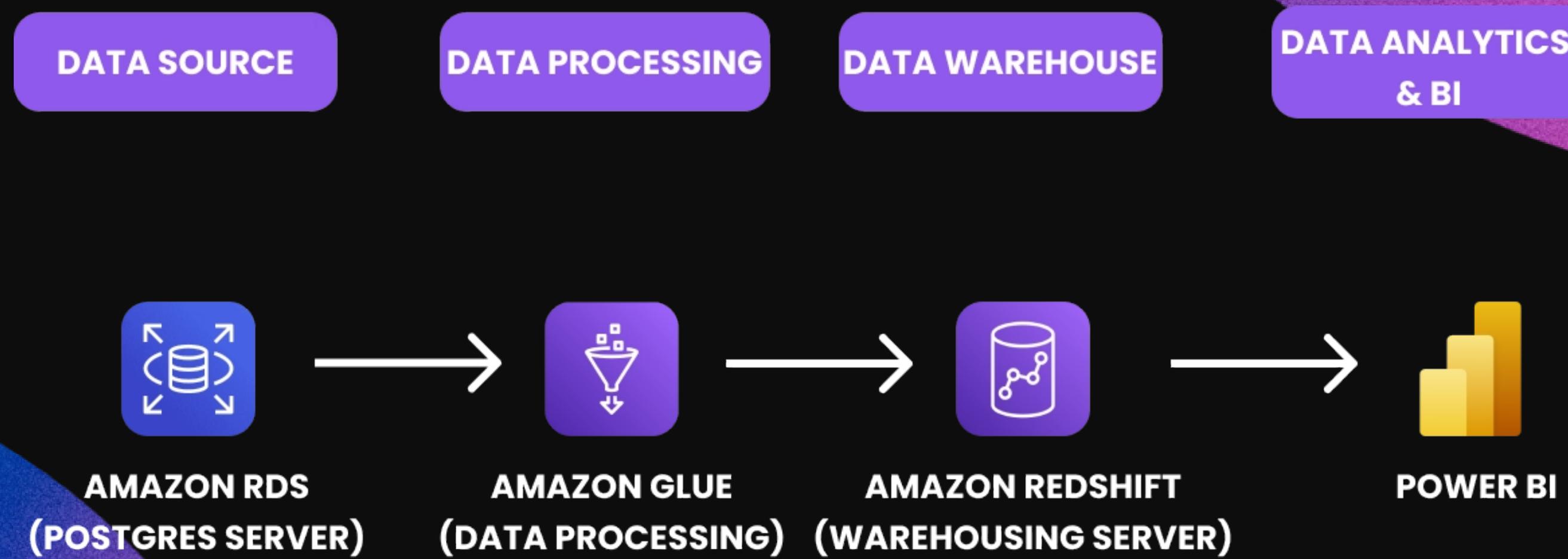
DATA DESCRIPTION

This dataset comprises comprehensive flight information, including cancellations and delays by airline, dating back to January 2020.

It is a dataset with high volume. (5.31 GB of data)

The dataset contains detailed information on thousands of flights, including parameters such as departure delays, arrival delays, airlines, and airports.

DATA PIPELINE:



RDS

The image displays two screenshots illustrating the integration of PostgreSQL and Amazon RDS.

pgAdmin 4 Screenshot:

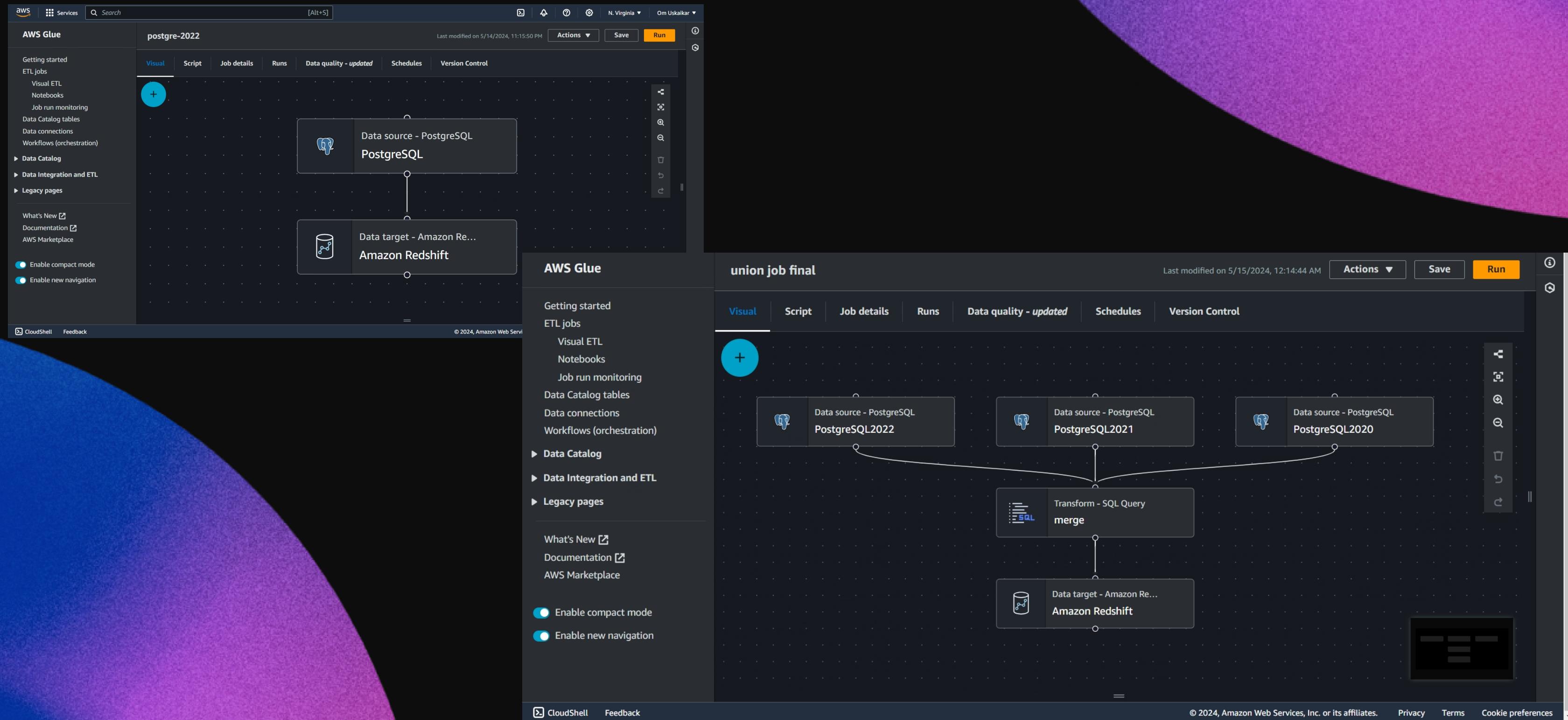
- Object Explorer:** Shows a tree view of database objects including Domains, FTS Configurations, FTS Dictionaries, FTS Parsers, FTS Templates, Foreign Tables, Functions, Materialized Views, Operators, Procedures, Sequences, and Tables (8).
 - Tables (8):** Contains entries for airlines, flight_data, flight_data_1, flight_data_2018, flight_data_2019, flight_data_2020, flight_data_2021, and flight_data_2022.
- Query Editor:** Displays the following SQL query:

```
1 SELECT *
2 FROM flight_data_2020
3 ORDER BY FlightDate DESC
4 LIMIT 10;
```

AWS RDS Console Screenshot:

- Summary:** Shows the DB identifier as database, Status as Available, Role as Instance, Engine as PostgreSQL, and Recommendations as 4 Informational.
- Connectivity & security:** Details the Endpoint (database.cnq4oac8k9bf.us-east-1.rds.amazonaws.com), Port (5432), Networking (Availability Zone: us-east-1c, VPC: vpc-0cf8fea394b99bb42, Subnet group: default-vpc-0cf8fea394b99bb42), and Security (VPC security groups: default (sg-0488514da6fcccacfb), Active).

GLUE JOBS (ETL)



REDSHIFT (DATA WAREHOUSE)

The screenshot displays the AWS Redshift Query Editor v2 interface, showing two separate query sessions.

Query 1:

```
1 SELECT *  
2 FROM ... "dev"."public"."flight_data_2022";
```

Result 1 (100):

flightdate	airline
2022-07-12	American Airlines Inc
2022-03-07	Delta Air Lines Inc.
2022-01-25	United Air Lines Inc.
2022-05-26	Endeavor Air Inc.
2022-02-27	Comair Inc.
2022-06-20	SkyWest Airlines Inc.
2022-04-30	Envoy Air
2022-06-05	American Airlines Inc

Query 2:

```
3 *,  
4 ROW_NUMBER() OVER (PARTITION BY year ORDER BY FlightDate DESC) AS rn  
5 FROM  
6 "dev"."public"."flight_data_merge"  
7 )  
8 SELECT *  
9 FROM RankedFlights  
10 WHERE  
11 rn <= 3
```

Result 1 (9):

flightdate	airline	origin	dest	cancelled	diverted
2020-12-31	Delta Air Lines Inc.	ATL	PHX	false	false
2020-12-31	Allegiant Air	LAS	CVG	false	false
2020-12-31	American Airlines Inc.	ORD	CLT	false	false
2021-12-31	SkyWest Airlines Inc.	BOI	DEN	false	false
2021-12-31	Air Wisconsin Airlines Corp	GSO	IAD	false	false
2021-12-31	Southwest Airlines Co.	AUS	SJC	false	false
2022-07-31	Commutair Aka Champlai...	PWM	IAD	false	false
2022-07-31	Delta Air Lines Inc.	PDX	MSP	false	false
2022-07-31	Horizon Air	BOI	SEA	false	false

Query ID 206543 Elapsed time: 298 ms Total rows: 9

POWER BI DASHBOARD:

FLIGHT DATA ANALYSIS

Month:

Select all

1.00

2.00

3.00

4.00

5.00

6.00

7.00

8.00

9.00

10.00

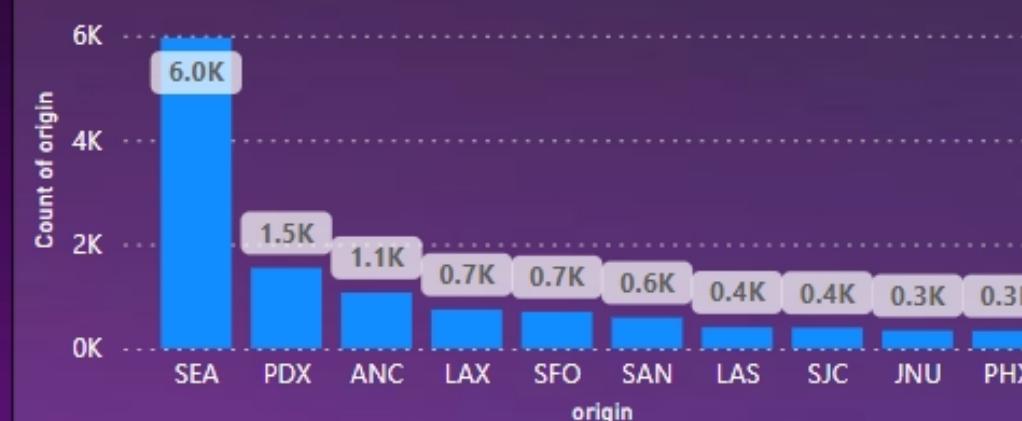
11.00

12.00

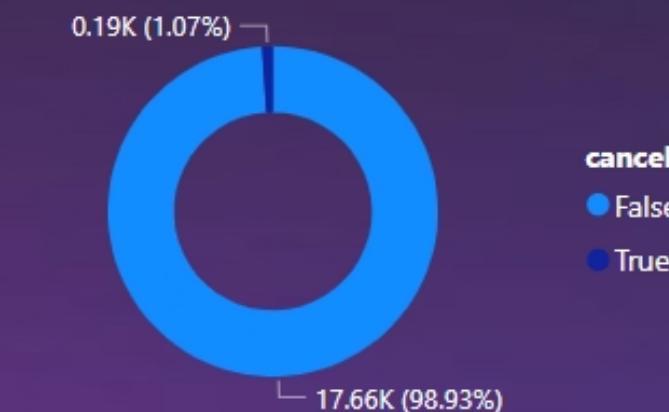
Year:

2021

Top 10 Busiest Airports



Cancelled Flights Percentage



Airline Names

Air Wisconsin Airlines Corp
Alaska Airlines Inc.
Allegiant Air
American Airlines Inc.
Capital Cargo International
Comair Inc.
Commutair Aka Champlain Enterprises, Inc.

Total Departure Delay by Airline



Day wise Average Arrival Delay



Average Departure Delay in Minutes

4.13





Want to make a presentation like this one?

Start with a fully customizable template, create a beautiful deck in minutes, then easily share it with anyone.

[Create a presentation \(It's free\)](#)