NTSB Accidents Database Project Requirements Specification Document

Faraazuddin Syed

2025-02-22

Version 1.0A

Project Overview

The **NTSB Accidents Database Project** (NTSB-ADP) is designed to compile and organize aviation accident data from 2000 to the present into a structured, relational database. The dataset will be sourced from the National Transportation Safety Board (NTSB), Aviation Safety Network (ASN), Bureau of Transportation Statistics (BTS), and other verified open sources.

The primary goal of this project is to standardize and store accident data in a PostgreSQL database, making it easier to query, analyse, and generate reports. The database will help identify trends in aviation safety, providing structured insights into accident frequency, causes, and aircraft performance over time.

A key aspect of this project is examining the relationship between aircraft type and safety records. This database will serve as a reference tool for aviation analysts, safety regulators, and industry stakeholders who require reliable data for research, reporting, and risk assessments.

Project Objectives

1. Enable historical analysis of aviation incidents.
2. Provide structured and validated accident data to client.
3. Support aviation safety research and reporting.
4. Allow efficient querying and data extraction.

Scope

The project scope is defined in the below figure:

Includes	Excludes
Collection and standardization of aviation accident data	Real-time flight tracking measurements
Relational database design for PostgreSQL	Predictive Analytics
Data dictionary and schema documentation	
Backup and recovery strategy	
Acquisition, extraction, transformation, and loading (ETL) process.	
Development of PostgreSQL views for efficient data retrieval	

The acquisition of data will involve CSV imports from the Case Analysis and Reporting Online (CAROL) tool provided by the National Transportation Safety Board (NTSB) in the United

States. This dataset comprises 44,567 items across 892 pages and has an approximate file size of 21.0 MB. Given its size, minimal optimization is anticipated during the ETL process.

A key aim of this project is to develop PostgreSQL views that enable quick lookups for common search cases. These common search cases will be identified based on high-level fields from the CSV file, which will be detailed in the subsequent sections of this document. The views will enhance the efficiency of data retrieval for users conducting analyses or generating reports based on the dataset.

Functional Requirements

The underlying functional requirements for this project include data requirements, performance and scalability requirements, and search and query requirements.

Data Requirements

2.0… The system shall store the following data fields for each accident record:

NTSB report reference number
Event type
Event Date
City
State (if applicable)
Country
Tail number
Safety recommendation binary
Report type
Original publish date
Highest injury level
Fatal injury count
Serious injury count
Minor injury count
Probable cause
Latitude
Longitude
Aircraft make / manufacturer
Aircraft model
Aircraft category
Airport ID
Airport name
Amateur built Boolean
Engine count
Scheduled / unscheduled binary
Purpose of flight
Aircraft damage
Weather conditions
Operator
Report status

2.1… The system shall ensure that all stored data is structured and indexed for efficient retrieval.

2.2… The system shall prevent duplicate accident entries by enforcing unique constraints on accident ID and report reference number.

2.3… The system shall support data retention policies, allowing administrators to archive or purge old records.

In consideration of the high number of fields, the data will be split into their relevant tables, more on this in subsequent sections of this document.

Performance and Scalability Requirements

Expected query response times will be optimised to ensure efficient retrieval of accident records. Performance benchmarks will be established after database implementation, with target response times under 500ms for filtered queries and under 2s for complex aggregations. Indexing, partitioning, and caching strategies will be employed to meet these targets.

Search and Query Requirements

3.0… The system shall allow users to search accident records based on multiple criteria (date, location, severity, aircraft type).

3.1… The system shall support full-text search on report descriptions and probably cause.

3.2… Query results shall be sortable, filterable, and paginated for better usability.

3.3… The system shall allow users to import accident records in CSV, JSON, and XML formats.

Non-Functional Requirements

The nonfunctional requirements of this project availability requirements, and compliance requirements.

Availability Requirements

5.0…. The system shall provide 99.9% uptime for database access and query operations.

5.1… A failover mechanism shall be implemented to ensure continuous availability.

5.2… Regular automatic backups shall be scheduled to prevent data loss.

5.3… The system shall have a disaster recovery plan in place.

Compliance Requirements

6.0… The system shall comply with NTSB data policies and FAA regulations.

6.1… Data storage and processing shall adhere to relevant data protection laws.

6.2… Audit logs shall be maintained for a minimum of five years to ensure regulatory compliance.

System Architecture

PostgreSQL will be the primary database system, handling data, storage, indexing, and query execution. A pipeline will be responsible for acquiring, transforming, and loading aviation accident data into the database. Queries will be optimised through PostgreSQL views to enhance retrieval performance.

Data Flow