**Pedestrian Counting System**

Functional Specifications Document

**DOCUMENT HISTORY** ------------------------------------------------------------------------------------------------------------------------1

VERSION HISTORY----------------------------------------------------------------------------------------------------------------- 1

APPROVAL LOG---------------------------------------------------------------------------------------------------------------------1

**INTRODUCTION**--------------------------------------------------------------------------------------------------------------------------------2

OBJECTIVE----------------------------------------------------------------------------------------------------------------------------2

BUSINESS NEEDS------------------------------------------------------------------------------------------------------------------- 2

PROJECT STAKEHOLDERS---------------------------------------------------------------------------------------------------------2

**SCOPE**---------------------------------------------------------------------------------------------------------------------------------------------2

SOURCE FILES------------------------------------------------------------------------------------------------------------------------3

EXTRACT FILE LOCATION/ENVIRONMENT------------------------------------------------------------------------------------3

GRANULARITY-----------------------------------------------------------------------------------------------------------------------3

HISTORY------------------------------------------------------------------------------------------------------------------------------3

DATA ELEMENTS-------------------------------------------------------------------------------------------------------------------3

DELTA FILES--------------------------------------------------------------------------------------------------------------------------4

INITIAL LOAD------------------------------------------------------------------------------------------------------------------------4

SCHEDULING------------------------------------------------------------------------------------------------------------------------ 4

OBFUSCATION---------------------------------------------------------------------------------------------------------------------- 4

**OUT OF SCOPE----------------------------------------------------------------------------------------------------------------------------------**4

**ASSUMPTIONS AND RISKS**------------------------------------------------------------------------------------------------------------------4

**DEPENDENCIES AND PREREQUISITES**----------------------------------------------------------------------------------------------------4

**RELATED DOCUMENTS**-----------------------------------------------------------------------------------------------------------------------5

**FUNCTIONAL OVERVIEW--------------------------------------------------------------------------------------------------------------------5**

Data Flow/Architecture-----------------------------------------------------------------------------------------------------------5

Data Model--------------------------------------------------------------------------------------------------------------------------5

**NON -FUNCTIONAL REQUIREMENTS-----------------------------------------------------------------------------------------------------6**

# Document History

Version History

| Date Modified | Version | Modified By | Description |
| --- | --- | --- | --- |
| 03/24/2023 | Draft | Ankit Chhabra | Created draft document |
|  |  |  |  |

Approval Log

| Date Modified | Approved By | Description |
| --- | --- | --- |
|  |  |  |

# Introduction

## Objective

This Functional Specification document is intended to describe the functional design of the solution. It provides the design and build team all the key information, assumptions, rules, and logic that are required to accurately design and build the interfaces to efficiently consume in-scope Data – Pedestrian Monthly Count Files in Data Lake to Operational Data Store (ODS) database in the Pedestrian system being setup for <Organization>.

The Technical Specification will be created to capture technical details.

## Business Needs

Deliver short and long-term distribution data and analytics capabilities, enabling data-driven experiences for the Analytics team of <org> as per its existing vision and strategy. Specifically, this data will provide insights into Pedestrian information for different Sensors across Melbourne which in turn will be consumed by the Government of Australia. As part of this release, the in-scope Pipeline Data – Source Count Monthly files will be ingested to Data Lake (S3 Buckets) . This solution will enable cleansed and conformed pipeline data for various data and analytics processes within <Org>.

Benefits of ingesting Monthly Counts Data - into S3 include:

* Supporting business reporting and analytical requirements
* Streamlining analytical capabilities, including drill-down, by creating a single source for Sensor information

## Project Stakeholders

Any person or department with a vested interest in the outcome of this project is a stakeholder. The table below categorizes the stakeholders associated with Sensor Data – Monthly Count Files.

|  |  |  |
| --- | --- | --- |
| **Name** | **Company** | **Role** |
| Renata Romila | Belong | Business Sponsor |
| Shenaz Syed | Belong | Program Lead |

# Scope

This document describes the ingestion of the Sensor Monthly Count data file from Source Data Pedestrian Counting System to the <ORG> ODS and the essential functional and non-functional requirements for its implementation.

This functional specification outlines the data scope and conceptual data flow. The detailed data scope will be defined by the depth (filter criteria, history, and granularity) and breadth (specific elements) of data. The depth is defined in the following sections, and the breadth is defined in the source-to-target mapping (STTM) document. The STTM outlines required translation rules, reference data, and quality checks to move all data from the landing to staging layers and from the staging to ODS layers of <Org>.

## Source files

As a part of this release, <OrG> will source data through a single in-scope extract consumed from Source Data Pedestrian Counting System. <Org> extracts will be created as per the defined frequency and the records that will be written to the Database.

1. <ORG> IT team to support creation of any required views on the current tables for sourcing data
2. The extracts would be placed in S3 Bucket and would be consumed by <Org>

The Pedestrian Data –source filename and its expected frequency is

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Source** | **File Name** | **Description** | **File Load** | **Frequency** |
| 1 | Pedestrian | pedestrian-counting-system-sensor-locations \_YYMMDD\_HHMMSS.csv | Sensor Information | Full File | Monthly/Yearly |
| 2 | Pedestrian | Pedestrian\_Counting\_System\_Monthly\_counts\_per\_hour\_YYMMDD\_HHMMSS.csv | Contains Hourly Counts for sensors | Delta File | Monthly |

## Extract file location/environment

The Pedestrian files will be sourced from AWS S3 buckets. Location of the S3 buckets are:

|  |  |
| --- | --- |
| **Filename** | **S3 Location** |
| pedestrian-counting-system-sensor-locations \_YYMMDD\_HHMMSS.csv | s3://<bucket>/datalake/raw/sensor/location/ |
| Pedestrian\_Counting\_System\_Monthly\_counts\_per\_hour\_YYMMDD\_HHMMSS.csv | s3://<bucket>/datalake/raw/ sensor/counts/ |

## Granularity

The granularity will be same from what is received from Source data. The ODS layer of <org> will include all in-scope elements at the lowest level of granularity available in the data files.

## History

Historical load is required to be loaded for Year 2009-2022

## Data elements

The list of all in-scope elements for loading into the ODS layer of <Org> is defined by the “Target Column” column in the “Stage to ODS” tab of the STTM for Pedestrian. The STTM document can be used for data lineage to identify the associated Data – Pedestrian source file element names for each of these in-scope elements.

To determine the list of in-scope elements for inclusion in <org>, multiple working sessions were conducted to understand the data and finalize the attributes. The attributes were then consolidated and de-duplicated.

All the attributes identified as part of the R&A sessions will be extracted from Source data and those extracts will be loaded to tables.

All needed transformations will be captured as part of the STTM.

## Delta Files

Extracts will be created on a Monthly basis

## Initial load

For the initial load will consist of a full file from Pedestrian tables via the extract generation mechanism . The data from 2009 which is approximately a 14 years’ worth of historical data would be loaded

## Scheduling

The timing of the extract file generation / processing / loading should be completed to meet the operational needs of the in-scope functional areas. If additional requirements are identified in the future, the job execution timing and sequence need to be discussed between IT and business stakeholders. The extracts will be generated from Source every first day of Month.

## Obfuscation

Based on the security guidelines defined by <org> security team, as of now there are no attributes which contain any identified attribute(s) for obfuscation.

# Out of Scope

In general, the following are out of scope for this release. Exceptions will be documented separately.

* Any Pedestrian attributes outside of what is explicitly listed as an attribute in the STTM will not be included

# Assumptions and Risks

The high-level assumptions and risks associated with the ingestion of Pedestrian files are outlined below:

* No additional SOX compliance process will be developed.
* Based on the security guidelines defined by <org> security team, as of now there are attributes which contain PII, and the identified attribute(s) require obfuscation
* All the Pedestrian source files will be made available in s3 data lake and will require minimal manipulations to be consumed by the Ingestion process
* There is no requirement of data movement from Cloud to On-prem or any outbound extract
* To begin the initial development process, the team would leverage manual extraction and ingestion of Pedestrian source files and later in the build phase the team would look to determine the process to automate the extraction and ingestion process
* A copy of entire Source Data files would be required to begin the development
* Development and QA testing will be conducted on the same dataset
* 2 additional sets of entire source files will be required to test the delta process
* UAT testers would need to identify the timeline for testing as well as the month of data that would need to be validated as part of UAT; An additional copy of dataset would be needed to facilitate the UAT testing
* Resolving any source data, file or integration issue will not be part of the existing release

# Dependencies and Prerequisites

The high-level dependencies and prerequisites associated with the ingestion of Pedestrian Data files to <Org> are outlined below:

* Access to upstream systems (Source Data) for the extract generation process
* On time arrival and successful validation of files (identification of high-level errors, failures)
* Availability of all the in-scope Pedestrian Data
* Reference information update and successful ingestion of source extracts into ODS

# Related Documents

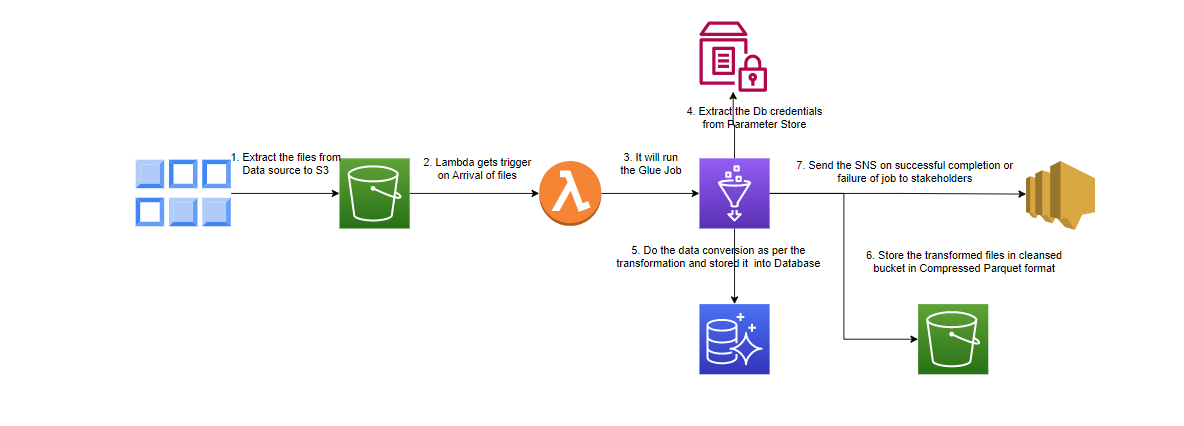
The following documents are related to this functional specification or are referenced in this document.

| Document |
| --- |
| Logical Data Model (LDM) |
| Source-to-Target-Mapping (STTM) |

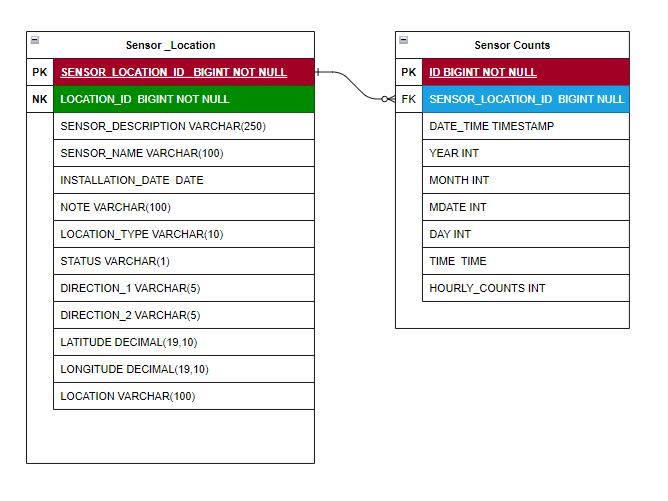
# Functional Overview

## Data Flow

This document covers the data flow from the ingestion of the Pedestrian Data to the ODS layer by using Glue and Lambda. Details of the data flow are as follows



Data Model



1. Sensor\_Location is a Dimensional table- It loads the information related to Sensor like Location/Id/Installation date etc. This table is loaded with the source file: pedestrian-counting-system-sensor-locations.csv
   1. Primary Key: Sequence generator id i.e., sensor\_location\_id
   2. Natural Key: location\_id from the source file- which uniquely identifies the record in the source file
2. Sensor Counts is a Fact table – it loads the information related to the sensor counts. This table is loaded from Pedestrian\_Counting\_System\_Monthly\_counts\_per\_hour\_may\_2009\_to\_14\_dec\_2022. Some of the attributes are dropped from the source file as not required like sensor id/sensor description as this all information can be obtained from Sensor\_location table
   1. Primary Key: Sequence generator id i.e., ID
   2. Foreign Key – Sensor\_Location\_ID – this value is fetched using the Sensor Id from the source file by joining this with the Sensor\_Location table (Location\_Id) attribute
   3. Natural key: Sensor\_ID – But this won’t be present in the table as using this we would fetch the location ID

# NON FUNCTIONAL REQUIREMENTS

This section outlines the major high-level non-functional requirements that are applicable for source Data. They include:

* **Data integrity**
  + The source system shall inform the IT team about any structural changes ahead of time, to ensure the change management process can be initiated and subsequently IT team can begin the plan to accommodate the changes as needed. For the initial set of production files, any source file change needs to be communicated no later than a month before the targeted source system go-live date
* **Data archiving / retention**
  + **S3 archival**. There is no archival requirement for S3 files as of now