Contents

I. Project Information 2
II. Overview 2
III. Datasets 3
IV. Environment Details 4
V. Data definition and code 5
VI. Design Details 5
A. **Architecture**: (describe high level architecture of the solution) 5
B. Data Flow: 6
C. **Data processing**: (please provide details of how the data will be ingested, processed, transformed, and stored) ..... **Error! Bookmark not defined.**

Technical Design Document

I. Project Information

Project Name	Aldenone – Connecticut Data Load
Sprint	
Date	12/22/2024
Prepared by	[redacted]
Reviewed by	

II. Overview

Objective:

Business requires the CT AldenOne application data to be made available in the EDP environment. It helps the business to utilize EDP as the source for their various reporting needs. As part of this requirement EDP must create a new space\catalog for keeping the AldenOne data.

Alden is the vendor and AldenOne is the application utilized to track the third-party attachments to Eversource poles.

Existing System

CT AldenOne data is getting pulled by Eversource from Alden's storage account utilizing SFTP on a daily basis and getting stored in SSMS database tables. AldenOne dataset consists of asset & config files.

- Alden's (source) storage account: edpaldenonetst
- Asset files: There are 12 asset files which contains details about poles, attachments, ducts etc.
- Config files: There are 45 config files which contains conversation (workflow) details and reports

Proposed System

AldenOne data files have to be pulled into the EDP platform for the business to utilize for reporting purposes.

- File Categories:XML is the source for Alden data files

- Asset files: Currently there are 12 asset files available from Alden, which contains data about poles, attachments, duct etc.
- Config files: Currently there are 45 config files available from Alden, which contains conversation details and Reports
- Mapping document for the Alden files to EDP is provided in the link below
  AldenFiles_EDP_mapping.xlsx

EDP Work Scope

- New pipeline to pull source data file to the target table in the EDP space
- Create new catalog\tables for Alden data in Databricks environment
- Data fields & data types of target table should be same as the input file
- Pull the dataset from the .psv files provided by Alden and include in the corresponding tables in EDP space.
- No change or transformations to the source data needed while adding to the target tables.
- Scheduled daily refresh for the Alden data set in EDP space

III. Datasets

Source system details:

System name:	edpaldenonetst
Data format:	psv
Data frequency:	Daily : hourly schedule

Target system details:

System name:	ADLS
Data format:	psv
Data destination (DB/Table/Storage):	ADLS Storage Storage Account: saedpaldenoneproduse2 saedpaldenoneproduse2 - Microsoft Azure

System name:	ADLS
Data format:	psv
Data destination (DB/Table/Storage):	ADLS Storage Storage Account: salzidfproduse2 salzidfproduse2 - Microsoft Azure

System name:	ADLS
Data format:	psv
Data destination (DB/Table/Storage):	ADLS Storage Storage Account: saizidfproduse2 saizidfproduse2 - Microsoft Azure

System name:	ADLS
Data format:	PSV
Data destination (DB/Table/Storage):	ADLS Storage Storage Account: saprivatepzidfproduse2 https://portal.azure.com/#home

System name:	EDP
Data format:	
Data destination (DB/Table/Storage):	edp_intgerated_prod Schema:Alden_ct

System name:	EDP
Data format:	
Data destination (DB/Table/Storage):	edp_finance_prod Schema:singlepoleadmin

IV. Environment Details

Development Environment:

Tools/Technologies:	Databricks, ADF, ADLS, Azure SQL Server
URLs/ Access Info:	Databricks ADF ADLS

Testing Environment:

Tools/Technologies:	
URLs/ Access Info:	

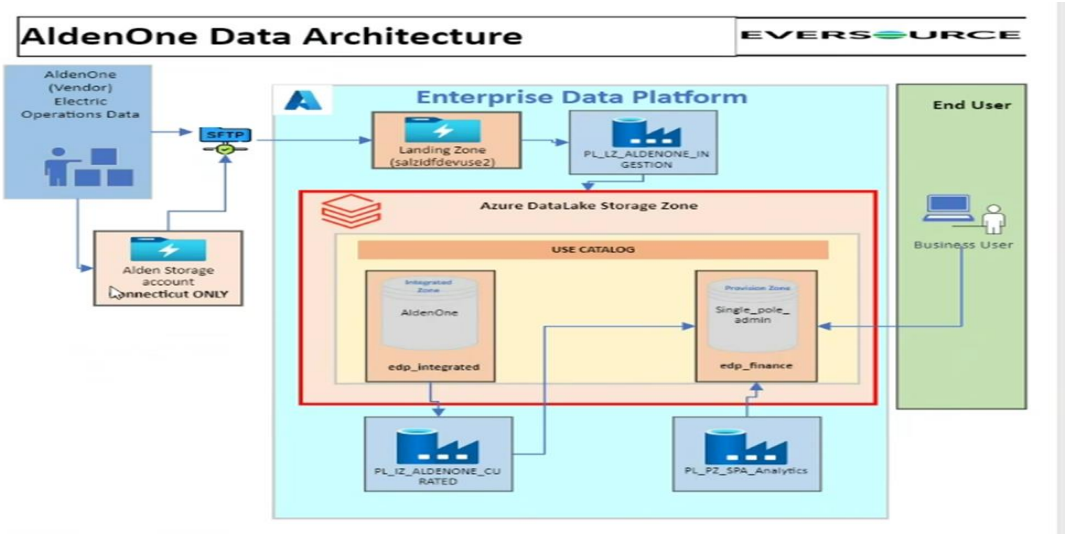
V. Data definition and code

Notebook code details: (link to notebooks used)

Repository Link:	Repo_Workspace_EDP : EDP_Alden Aldenone CT Asset confiles Table comments update Aldenone CT Assetfiles Integrated Data Ingestion Aldenone CT Assetfiles Provision Data Ingestion Aldenone CT catalog schema Create Aldenone CT confiles Integrated Data Ingestion Aldenone CT confiles Provision Data Ingestion DDL's for CT Aldenone Asset confiles tables create Aldenone CT schema Create - Databricks
Branch:	
Commit ID:	

VI. Design Details

A. Architecture



Data Sources:

- edpaldenonetst

Data Storage Layers:

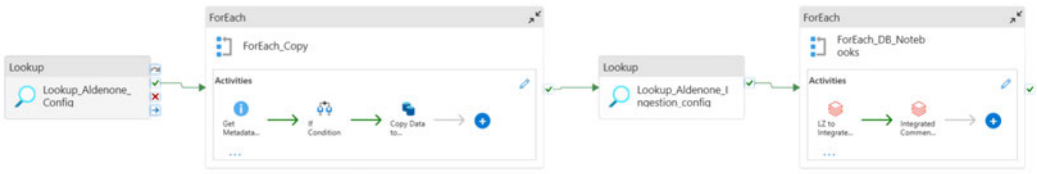
- saedpaldenoneproduse2 (Alden Zone)
- saizidfproduse2 (Integrated zone)
- salzidfproduse2 (landing zone)
- saprivatepzidfproduse2 (Private zone)

Components involved:

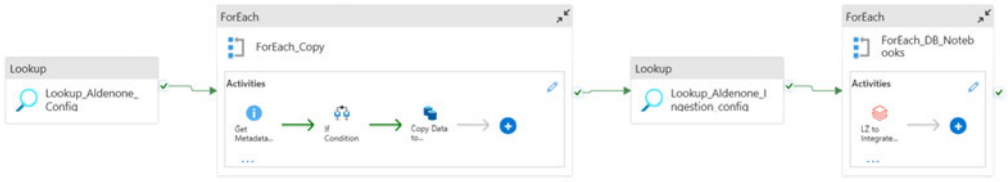
- Azure Data Factory
- Azure Blob Storage
- Data bricks

B. Data Flow:

pl_Aldenone_CT_Assetfiles_EDP_Ingestion



pl_Aldenone_CT_Confiles_EDP_Ingestion



Data Processing:

1. On Daily Basis Asset and Conversation Data (psv files) are copied from Alden Source storage Account (edpaldenonetst) to Landing Zone/Integrated and provision Zone thru copy Activity of ADF Ingestion Pipeline

   Below are the storage accounts files are copied to
   Saedpaldenoneproduse2/salzidfproduse2/ saizidfproduse2/ saprivatepzidfproduse2

2. Configuration are set on SQL server Database and passed to ADF and Notebook thru lookup activity

   Server: srv-data-idf-prod-use2.database.windows.net
   Database: sql-data-idf-prod-use2
   Schema:dbo
   Tables:
   Aldenone_Asset_CT
   Aldenone_Asset_Ingestion_CT
   Aldenone_Conversation_CT
   Aldenone_Conversation_Ingestion_CT

3. Databrick Notebook are executed thru ADF pipelines to load 12 Asset and 45 conversation files from source (psv files) to corresponding Tables on Integrated and Provision Catalog on EDP

   Aldenone CT Assetfiles Integrated Data Ingestion

   Aldenone CT Assetfiles Provision Data Ingestion

   Aldenone CT confiles Integrated Data Ingestion

   Aldenone CT confiles Provision Data Ingestion

Approval

Prepared by:	[redacted]
Reviewed by:	[redacted]
Approved by:	[redacted]