I

Implementation of ETL workflow

Version 1.0

12-January-2019

Table of Contents:

1.1 Project Overview

Development of Application which implements ETL Workflow in SAP Hana Database to extract and load data from External system.

1.2 Purpose and scope of this Document

This document describes the requirement to implement Extract, Transform and Load data from a csv file stored in External system into SAP Hana database via an ETL Workflow.

1.3 Functional Requirements

1. **Extract data from external sources**

Application should be able to read data and extract data from two types of data:

a. Non-Cooperative sources: Flat file of type csv
b. Cooperative sources: DB Triggers

The goal is fast extraction of data from external source system.

2. **Transform data**

Application should be able to transform data:

a. Syntactic Transformation: Data loaded from csv file should be structured as per the tables designed in SAP Hana database for this application.
b. Sematic Transformation: Definition of structure of data, format and interpretation of data stored in normalized tables.

It should perform normalization/denormalization to desired format depending on source. Data must be precise, complete, consistent, unique and timely.

Data Cleansing should be performed to remove unnecessary attributes, handle inconsistent data formats, combine data with common key, handle removal of duplicates, uniform treatment of NULL, avoid problems in joins.

3. **Load Data**

Application should be able to perform fast loading in SAP HANA. The extraction should be done from only changes in the last load (Delta Loading).

Parallelization: Dimensions, Fact Tables, Partitions can be loaded concurrently.

1.4 Non-Functional Requirements

Application should ensure:

1. Availability: Application should remain accessible at all times for e.g. during delta loading.
2. Integrity: Data Integrity should be maintained during and after all stages of ETL workflow have been executed.
3. Performance: Performance of application should not be hampered w.r.t to time, space and cost of operations
4. Accuracy: Application should always maintain accurate data
5. Consistency: Application should remain in consistent state before data is extracted, transformed and loaded in the system