ADB: AZURE Data Bricks project - Hospital Admissions.

**Hands-on Project on Azure Data Engineering:**

**Project**: Extract Load and Transform Hospital Admissions Data

**Data**: hospital\_admissions.csv

Project Description:

Create an Azure Data Factory Pipeline which would pull the csv file from on premise folder to ADLS container.

Checks:

1)     Check if the file is available in the path. If it’s not available, there should be timeout after 1 minute.

2)     Check if the file is empty or not. If the file is empty throw an error.

3)     Verify the records in source file with the one in ADLS. Both should be equal.

Write a code to extract, transform and load the data from source to target. For ETL, use Python, Panda etc.

•              If date is NULL or blank, give default date as ‘2020-11-28’. Format of date column should be YYYY-MM-DD.

•              Remove the entries which has URL field value as ‘ERROR’.

•              Transform the values of column country with their acronyms. For eg: Austria would be replaced

by ‘AUST’, Belgium by ‘BELG’ etc. You can find data in given url: [Country Acronyms (state.gov)](https://travel.state.gov/content/travel/en/us-visas/visa-information-resources/fees/country-acronyms.html)

Load the data to target table. Use Azure SQL. Structure of the target table:

Table Name: PROJ.HOSPITAL\_ADMISSIONS Column Names:

|  |  |
| --- | --- |
| **Name** | **Data Type & Constraint** |
| **Country** | varchar(10),NULL |
| **Indicator** | varchar(50),NULL |
| **Date** | date, NOT NULL |
| **Year\_Week** | varchar(10),NULL |
| **Value** | numeric(22,4),NULL |
| **Source** | varchar(50),NULL |
| **Url** | varchar(100),NULL |

Deliverables:

1.     Architecture design as per project requirements.

2.     Best Practice in Pipeline

3.     Best Practice for Database

4.     Best Practice for ETL

5.     Target Tables

Demo:

•              Problem Definition - 5

•              Literature Review - 10

•              Proposed System & Design -10

•              Implementation -10

•              Result & Analysis -10

•              Presentation -55