# Data Engineering Challenge

Contents

[Data Engineering Challenge 1](#_Toc107768404)

[Section 1: Data Pipelines (Deliverable: Github link, sec1Target.csv) 2](#_Toc107768405)

[Section 2: Databases (Deliverable: Documentation) 6](#_Toc107768406)

[Section 3: System Design (Deliverable: Power Point presentation) 9](#_Toc107768407)

[Section 4: Charts and APIs (Deliverable: Documentation with screenshots) 9](#_Toc107768408)

[Section 5: Machine Learning (Documentation with screenshots) 12](#_Toc107768409)

Azure suite of services have been used for the following exercises expect for Sec 2 and 3. The Github configuration has been added for Azure Data Factory alone.

## Section 1: Data Pipelines (Deliverable: Github link, sec1Target.csv)

1. Created a data factory detcadf1 in Azure portal
2. Created a blob storage account – detcbs

Graphical user interface, text, application, email

Description automatically generated

The two datasets are uploaded into the sec1dp container under Source folder

Graphical user interface, application

Description automatically generated

A linked Service detcbs\_LinkedService is created from ADF to connect to the above Blob Storage Account:

Graphical user interface, text, application, email

Description automatically generated

A dataset DSSource is created within ADF to read both files from the Blob Storage Account:

Graphical user interface, text

Description automatically generated

**Filter:**

Graphical user interface, text, application, email

Description automatically generated

**Derived Columns:**

Text

Description automatically generated with medium confidence

**first\_name** - left(name,instr(name,' ')-1)

**last\_name** - right(name,length(name)-instr(name,' '))

**price** - ltrim(price,'0')

**above\_100** - case(toDecimal(price, 20, 4)>100,'true','false')

Chart, radar chart

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Target File: sec1Target.csv is created

The ADF job above has been scheduled via an ADF trigger as below:

Graphical user interface, text, application, email

Description automatically generated

The trigger is then attached to the adf pipeline:  
  
A screenshot of a computer

Description automatically generated with medium confidence

Graphical user interface, text, application, email

Description automatically generated

## Section 2: Databases (Deliverable: Documentation)

Docker Desktop was configured as the first step. The docker file extraction wasn’t possible.

With a new ID - ndsudhakar

Graphical user interface, text, application, email

Description automatically generated

Postgresql DB Created:

Text

Description automatically generated

A picture containing diagram

Description automatically generated

Table Creation Scripts:

Text

Description automatically generated

Text

Description automatically generated

**Queries required:**

Graphical user interface, application

Description automatically generated

## Section 3: System Design (Deliverable: Power Point presentation)

DE Challenge Sec3 has been attached

## Section 4: Charts and APIs (Deliverable: Documentation with screenshots)

PowerBI has been used to create a dashboard that displays data with different granularity:

Graphical user interface, text, application, email

Description automatically generated

URL Used in PowerBI:

https://api.covid19api.com/country/singapore/confirmed?from=2021-06-01T00:00:00Z&to=2022-06-30T00:00:00Z

Graphical user interface, application

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart

Description automatically generated

## Section 5: Machine Learning (Documentation with screenshots)

Azure ML was used to build an ML model. A regression model was built to predict the price of a car using the given test dataset:

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface

Description automatically generated

