The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document Template

# Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

## Data Source

COVID-19 Case Surveillance Public Use Data

### Common Core

|  |  |
| --- | --- |
| Publisher | CDC |
| Contact Name | Surveillance Review and Response Group |
| Contact Email | [eocevent394@cdc.gov](mailto:eocevent394@cdc.gov) |
| Suggested Citation | Centers for Disease Control and Prevention, COVID-19 Response. COVID-19 Case Surveillance Public Data Access, Summary, and Limitations (version date: December 31, 2020). |
| Bureau Code | 009:20 |
| Program Code | 009:020 |

### Data Quality

|  |  |
| --- | --- |
| Geographic Coverage | US |
| Temporal Applicability | 2020-01-01/2020-12-16 |
| Update Frequency | Monthly |

### Technology Choice

Saved as CSV file

### Justification

This was the format given and it’s fine.

## Enterprise Data

[COVID-19 Case Surveillance Public Use Data | Data | Centers for Disease Control and Prevention (cdc.gov)](https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf)

### Technology Choice

Website

### Justification

Has reliable and updated data

## Streaming analytics

Cognos Dashboard: <https://dataplatform.cloud.ibm.com/dashboards/b22c220a-58e7-4772-9c80-dbbcdd34d5c4/view/4f60bc0838ec6cc77ff2d4e407ca7905283f255de7bbd55282827b490b342797a93d4098c82b1d5fdc450035a0ed1250cf>

### Technology Choice

Cognos Analytics

### Justification

Dashboard is interactive and the Data can be updated monthly

## Data Integration

### Technology Choice

None

### Justification

Only one source is needed

## Data Repository

### Technology Choice

Data is saved on local computer—if needed by others, it would be available on GitHub

### Justification

GitHub makes the data accessible to all team members

## Discovery and Exploration

### Technology Choice

Bar graphs in Python notebook and the above Cognos Dashboard

### Justification

Both have good graphics for exploration

Features and Label:

|  |  |
| --- | --- |
| **current\_status** | Case Status: Laboratory-confirmed case; Probable case |
| **sex** | Sex: Male; Female; Unknown; Other |
| **age\_group** | Age Group: 0 - 9 Years; 10 - 19 Years; 20 - 39 Years; 40 - 49 Years; 50 - 59 Years; 60 - 69 Years; 70 - 79 Years; 80 + Years |
| **race\_ethnicity\_combined** | Race and ethnicity (combined): Hispanic/Latino; American Indian / Alaska Native, Non-Hispanic; Asian, Non-Hispanic; Black, Non-Hispanic; Native Hawaiian / Other Pacific Islander, Non-Hispanic; White, Non-Hispanic; Multiple/Other, Non-Hispanic |
| **hosp\_yn** | Hospitalization status |
| **icu\_yn** | ICU admission status |
| **medcond\_yn** | Presence of underlying comorbidity or disease |
|  |  |
| **death\_yn** | Death status |

**Deidentified Patient**

## Actionable Insights

### Technology Choice

From Keras in Python (gridsearch):

* **Loss = binary\_crossentropy, with=adam**
* **Loss=categorical\_crossentropy with optimizer=adam**
* **Loss=mean\_squared\_error with optimizer=adam**
* **Loss = binary\_crossentropy, with optimizer=SGD**
* **Loss=categorical\_crossentropy with optimizer=SGD**
* **Loss=mean\_squared\_error with optimizer=SGD**
* **Loss = binary\_crossentropy, with optimizer=Adadelta**
* **Loss=categorical\_crossentropy with optimizer=Adadelta**
* **Loss=mean\_squared\_error with optimizer=Adadelta**

For Logistic Regression in Python:

* **Accuracy**
* **Confusion matrix**
* **Feature Importance**

### Justification

Python is arguably the best for these!

## Applications / Data Products

Please specify and justify the technologies used for model definition and training in the ADD.

### Technology Choice

Python notebook on GitHub

### Justification

The notebook (with Markdown) explains the code in detail

## Security, Information Governance and Systems Management

### Technology Choice

Data is open sourced

### Justification

Very reliable data

Notebook Link:

<https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/592428c9-02d1-4b56-b08f-a9d157b9522e/view?access_token=c46f0f04af11c0b36d0e001c92aac715c89d09b20ea8a04054e00c758bc784da>

### or

[Covid\_Neural\_Networks-and-Logistic\_Regression/Covid\_Neural\_Networks and Logistic\_Regression.ipynb at main · kwpeters58/Covid\_Neural\_Networks-and-Logistic\_Regression (github.com)](https://github.com/kwpeters58/Covid_Neural_Networks-and-Logistic_Regression/blob/main/Covid_Neural_Networks%20and%20Logistic_Regression.ipynb)