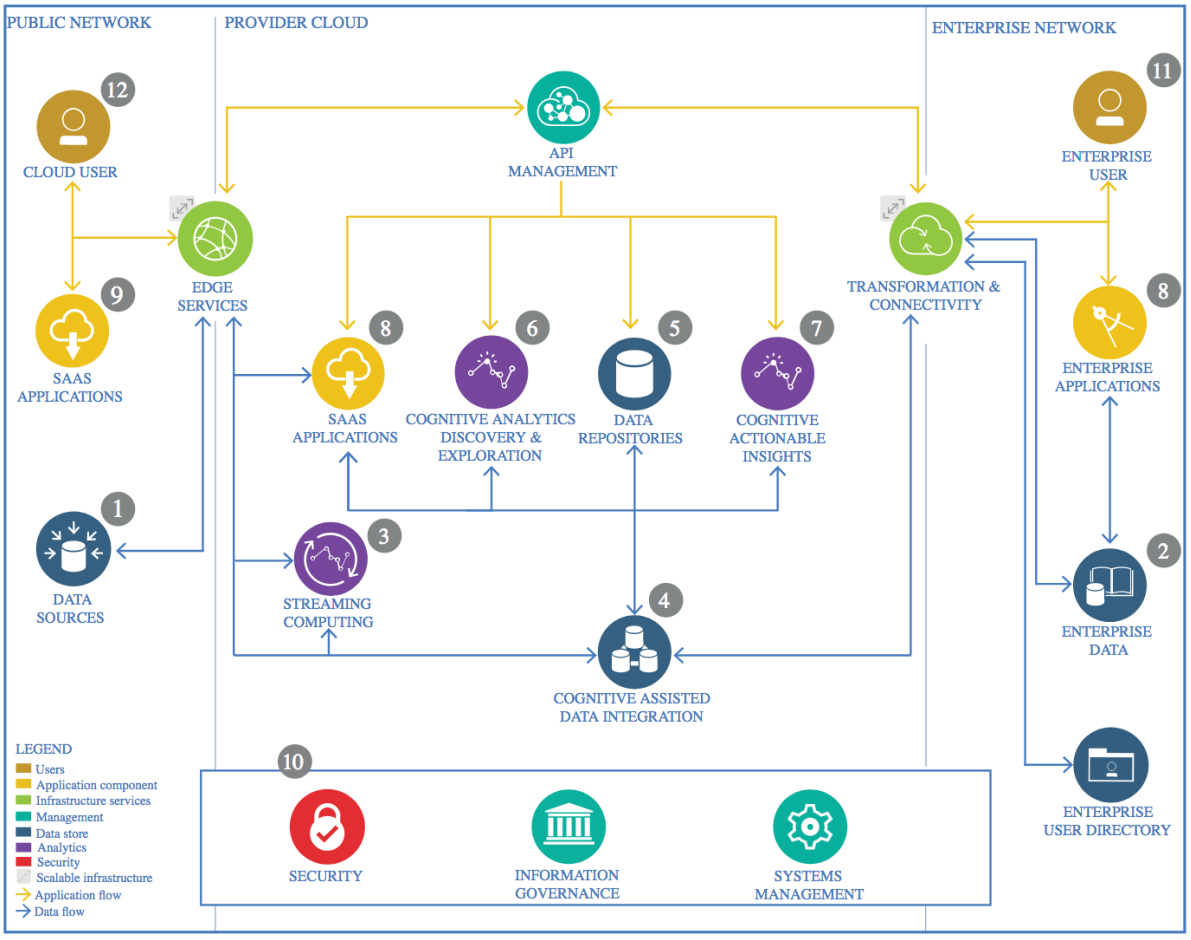
Mammogram Abnormality Detection

Architectural Decisions Document

(The Lightweight IBM Cloud Garage Method for Data Science)

1 Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

1.1 Data Source

1.1.1 Technology Choice

The images are from the Cancer Imaging Archive, under The Chinese Mammography Database (CMMD), consisting of DICOM mammogram images from 1775 patients from China with benign or malignant breast disease who underwent mammography examination between July 2012 and January 2016. Due to image imbalances, more images were also obtained from the same source, under the Curated Breast Imaging Subset of DDSM (CBIS-DDSM).

1.1.2 Justification

The Cancer Imaging Archive consists of a global trusted source of cancer-related images available for public use. Such a comprehensive database is valuable and hard to gather individually.

1.2 Enterprise Data

1.2.1 Technology Choice

IBM Cloud Pak for Data is used for the project. For deployment, Google Colab was also used for quick web application prototyping.

1.2.2 Justification

IBM Watson Studio is already a familiar tool used within the course, and provides ease of scale and deployment. Google Colab is also a valuable free resource.

1.3 Streaming analytics

1.3.1 Technology Choice

Not applicable.

1.3.2 Justification

This is not a real time model.

1.4 Data Integration

1.4.1 Technology Choice

The DICOM images and features are identified before being used for model development.

1.4.2 Justification

The image and features are required as input into a deep learning model for training in image classification.

1.5 Data Repository

1.5.1 Technology Choice

The IBM Object Storage is being used. All images are stored in file folders, with supporting meta-data.

1.5.2 Justification

It is already a familiar tool used within the course, and provides ease of scale and deployment.

1.6 Discovery and Exploration

1.6.1 Technology Choice

The mammogram images are in DICOM format, with a CC and MLO view on the side of the breast with calcification or mass object. All images are classified as having calcification, mass or both. The Convolution Neural Network DeepLearning approach will be used as an effective way to do image classification. Both calcification and mass can appear inside a mammogram together. Hence, the image will be multi-labeled as 'benign/malignant-calcification' and 'benign/malignant mass', For assessment of the model, the Accuracy score and Confusion Matrix will be used.

1.6.2 Justification

The image abnormality detection model requires knowing whether the object is benign or malignant, for classification into various classes.

1.7 Actionable Insights

1.7.1 Technology Choice

The radiologist will be able to have aided abnormality detection on patients' mammogram images via a web application.

1.7.2 Justification

This will help to further improve the current human diagnosis accuracy of around 87%.

1.8 Applications / Data Products

1.8.1 Technology Choice

It will be a cloud based web accessible model for ease of anomaly detection. The programming language selected will be python which is popular and easy to maintain. TensorFlow/Keras libraries will be used for DeepLearning.

1.8.2 Justification

The model can be used anywhere and anytime. A cloud based web application is easily scalable/deployable and easy to use. Python is also easy to upskill compared to other traditional programming languages.

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice

The access to images are restricted for radiologists only, with certain locations permissible only.

1.9.2 Justification

The patients' data are highly confidential and sensitive, and protected under PDPA and MOH guidelines.