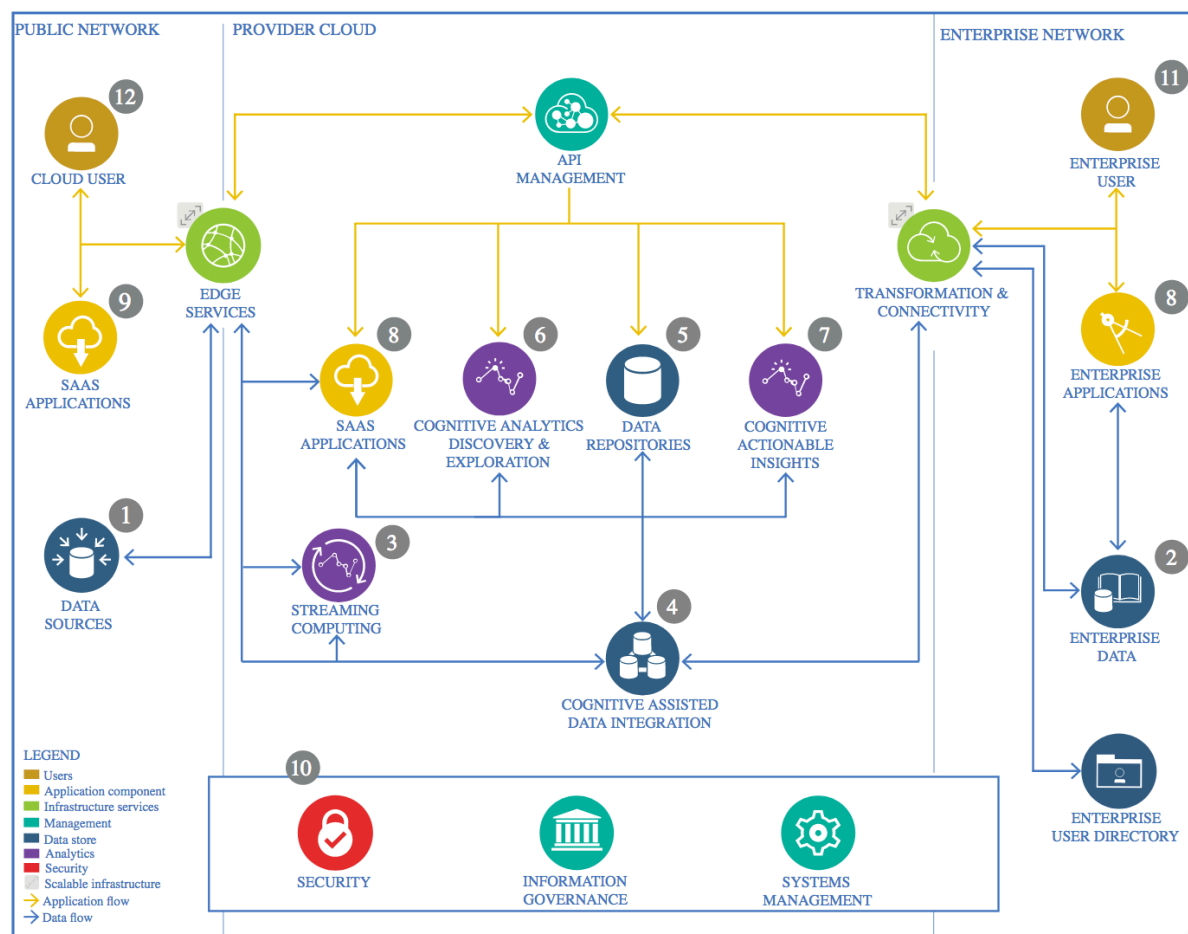# The Lightweight IBM Cloud Garage Method for Data Science

## Architectural Decisions Document

## 1. Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

### 1.1. Data Source

#### 1.1.1. Technology Choice
The data was downloaded from Kaggle (https://www.kaggle.com/ankurzing/sentiment-analysis-for-financial-news).

#### 1.1.2. Justification
Primary reason to download from Kaggle was availability and ease of use.

## 1.2. Enterprise Data

### 1.2.1. Technology Choice
GitHub repository

### 1.2.2. Justification
Ease of use to keep the repository up-to-date and publicly accessible.

## 1.3. Streaming analytics
No devices used for streaming analytics, as this was not needed for the analysis of the underlying data.

## 1.4. Data Integration
No devices used for data integration, as this was not needed for the analysis of the underlying data.

## 1.5. Data Repository
No devices used for data repository, as this was not needed for the analysis of the underlying data.

## 1.6. Discovery and Exploration

### 1.6.1. Technology Choice
The Python 3.6 libraries Pandas and Matplotlib were primarily used for Data Exploration and Visualization.

### 1.6.2. Justification
The underlying data is small enough to be processed on a single computer ruling out the need for distributed processing.

## 1.7. Actionable Insights

### 1.7.1. Technology Choice
The following Python 3.6 libraries have been used for actionable insights:
Pandas
Matplotlib
Wordcloud

### 1.7.2. Justification
Pandas and Matplotlib have been sufficient for analyzing the distributions, shape and size. Wordcloud has been a proper choice for plotting the occurrences of each word.

## 1.8. Applications / Data Products

### 1.8.1. Technology Choice

A Jupyter notebook based report was generated.

### 1.8.2. Justification

A Jupyter notebook based report was considered to be sufficient for the simplicity of the data project.

## 1.9. Security, Information Governance and Systems Management

No devices used for data security and systems management, as this was not needed for the analysis of the underlying data.