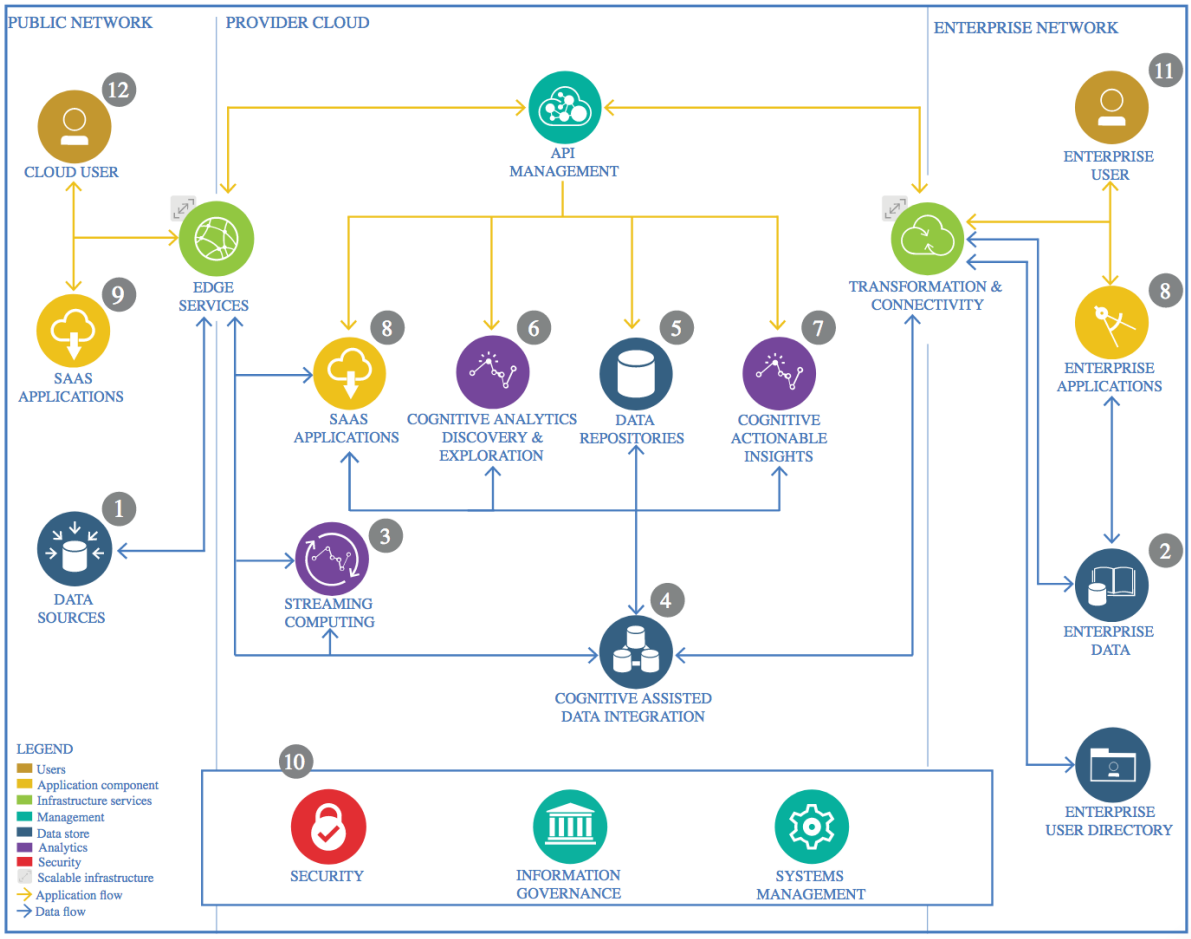
The Lightweight IBM Cloud Garage Method for Data Science

Architectural Decisions Document Template

1 Architectural Components Overview



IBM Data and Analytics Reference Architecture. Source: IBM Corporation

1.1 Data Source

1.1.1 Technology Choice

The technology choice here is the friendly API provided by Lichess.org, a possible source of data related to chess games.

1.1.2 Justification

My goal was to retrieve data about chess games online, so I chose the website Lichess.org because it provides users with a powerful API, which enables collection of any given user's game history. I actually used a publicly available dataset obtained in this way.

1.2 Enterprise Data

1.2.1 Technology Choice

Not used

1.2.2 Justification

In this case I did not use any enterprise data, or, to be more precise, the data of the enterprise Lichess was publicly available and coincided with the data source I used.

1.3 Streaming analytics

1.3.1 Technology Choice

Not implemented

1.3.2 Justification

My analysis is static, although it could have also been conducted in a real-time way, so I did not consider any streaming analytics tool or technology.

1.4 Data Integration

1.4.1 Technology Choice

Not in this case

1.4.2 Justification

The source of data I collected is unique, so there is no need for integration in this case.

1.5 Data Repository

1.5.1 Technology Choice

Either my own disk or IBM Cloud *Object Storage*.

1.5.2 Justification

I worked on my project partly locally on my laptop and partly on Watson Studio, so I chose accordingly where to store the dataset.

1.6 Discovery and Exploration

1.6.1 Technology Choice

Data Exploration with summary statistics and some forms of inferential statistics, visualization techniques such as Histograms, Bar plots, Pie Charts or similar.
Also, some outlier and NaN basic detection techniques.

1.6.2 Justification

I chose these commonly used approaches because they allow a complete and significant understanding of the most relevant features, how they correlate, if there are inconsistent values, empty or duplicates entries etc.

1.7 Actionable Insights

1.7.1 Technology Choice

A PDF Report using Jupyter Notebook

1.7.2 Justification

I chose the PDF report because I want to show stakeholders that it is possible to predict with a certain accuracy how a chess game will end, highlighting why and how, to provide them with enough information to enable them to make appropriate decisions regarding the whole gaming experience based on the report results and considerations.

1.8 Applications / Data Products

1.8.1 Technology Choice

Not created, optional

1.8.2 Justification

One interesting application I had in mind was to implement in real-time the algorithm created in the project to stop a game or suggest a player stop a game when he/she has a very low chance to win, in order to start a new game and avoid wasting time, to keep the game entertaining.

1.9 Security, Information Governance and Systems Management

1.9.1 Technology Choice

Not implemented

1.9.2 Justification

I did not concentrate on the security aspects of this project, but if it were necessary as it is I would think of CrateDB. CrateDB is a cybersecurity database that combines a real-time SQL engine built on top of a NoSQL foundation. Thanks to this, it provides scalability, performance, and flexibility to process logs and network traffic in real-time or in great volumes to support a wide variety of cybersecurity analytics use cases.