# Data Architecture Documentation

AMR DATAVISUALIZATION

WEBSCAPE

Introduction

This documentation dedicates explaining the data pipeline for the Capstone Project "AMR Data Visualization" by Webscape team. The explanation includes how the data is processed; system components that are already existed, will be added and components are not going to be applied but worth mentioning.

Data Architecture Diagram
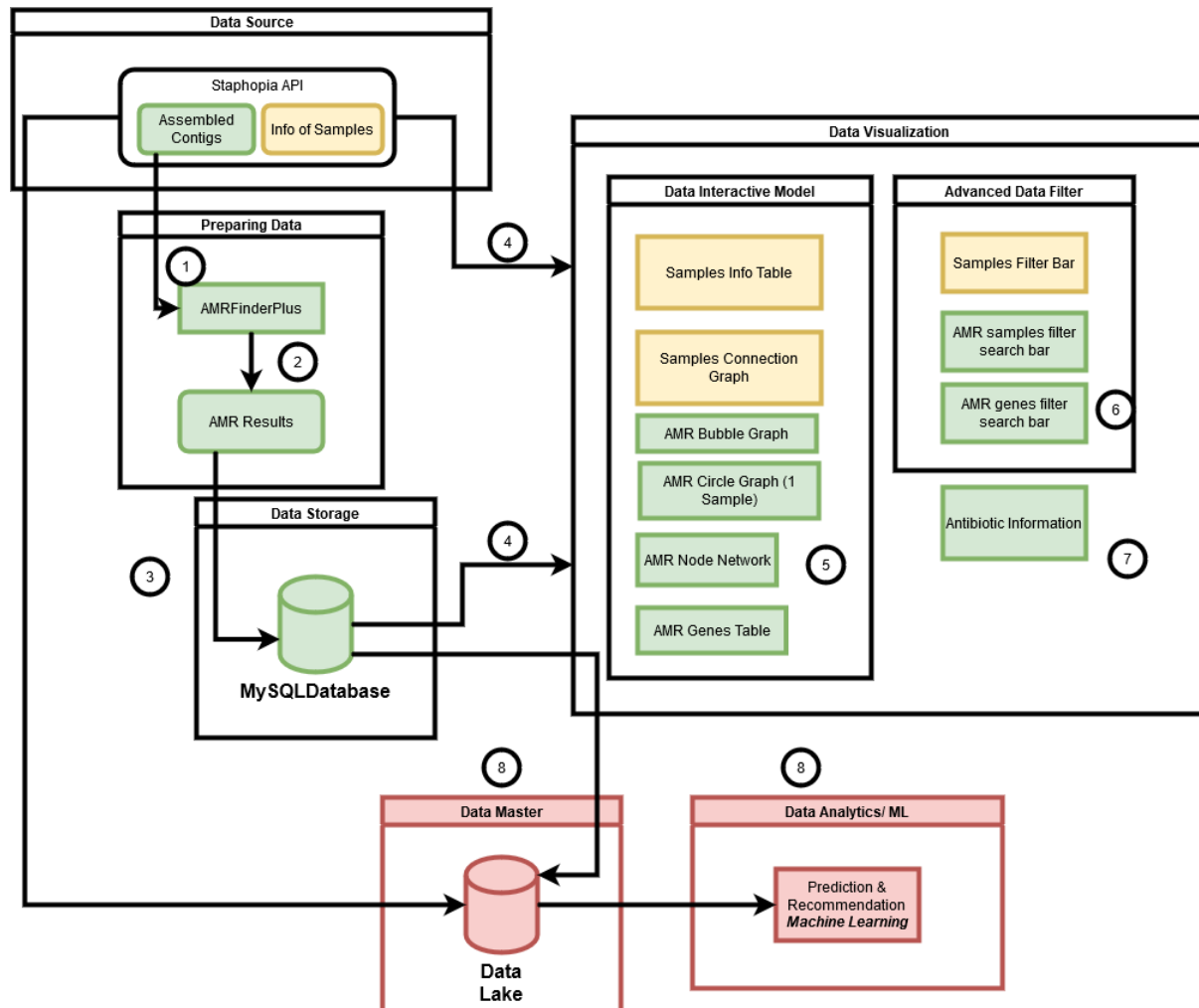
Additional Notes

The note is numbered, which also indicates that the process order and tells that the note is specifically for that system components shown in the diagram

1. Since Staphopia API does not provide any information regarding AMR genes. The team has to prepare AMR data by getting assembled contigs sequence of each samples and use it as an input for AMRFInderPlus to predict AMR genes of those contigs

2. AMRFInderPlus generates the result. The results includes information about the AMR genes and its content (nucleotides), the name of AMR genes, the family of the genes, the antibiotic type that the genes against on.
3. The result is then stored in teams' SQL database
4. The data regarding the information of each sample (origins, metadata, SCCmec, ...) and newly generated AMR genes result from the API are used for the teams' data visualization feature for the website
5. The new models will be added are bubble graph, circle graph (for one sample), node network and a new table. All of these models dedicate for visualizing AMR genes
6. Since new models are added, new feature for filtering data feature is also added. The feature will be presented as a search bar and they are for filtering AMR genes and samples.
7. A brief information of antibiotics will be added as it will help users to understand the antibiotics and AMR genes related to it.
8. The raw data can be prepared and stored in data lake storage and can be used by machine learning model. This component is not being implemented in the project. However, it is worth mentioning as this solution supports machine-learning based in the future

Colour Notation

Already existed component in the previous project

Component will be added in this project

Component not going to be implemented but worth mentioning