https://github.com/synthetichealth/synthea/wiki/Architecture

=================================Project-X1===============================

Build a healthcare product recommendation engine using PySpark (ALS and/or other model like FPM)

**DataSet (CSV):**

<https://drive.google.com/drive/u/0/folders/1iqmypphulNV66QwfeHD14WloFJo8vApE>

Dataset Relationships:

<https://docs.google.com/document/d/1VbA5APdc4xQSvCAv0Y2WXNQlnZp9c8YMEb9Orjt3RLA/edit>

Optional:

FHIR data

**Must-have Technologies:**

<https://spark.apache.org/docs/latest/api/python/pyspark.ml.html>

<https://spark.apache.org/docs/latest/api/python/pyspark.sql.html>

Recommendation Module

Tuning Module

Evaluation Module

Nice-to-have Technologies:

Spark Graphframes

MLFlow

Koalas

Redash

Spark-NLP package from spark-packages

=================================Project-X2===============================

Build an encounter cost estimator/predictor engine using PySpark (logistic regression and/or other model)

**DataSet (CSV):**

<https://drive.google.com/drive/u/0/folders/1iqmypphulNV66QwfeHD14WloFJo8vApE>

Dataset Relationships:

<https://docs.google.com/document/d/1VbA5APdc4xQSvCAv0Y2WXNQlnZp9c8YMEb9Orjt3RLA/edit>

Optional:

FHIR data

**Must-have Technologies:**

<https://spark.apache.org/docs/latest/api/python/pyspark.ml.html>

<https://spark.apache.org/docs/latest/api/python/pyspark.sql.html>

Regression Module

Tuning Module

Evaluation Module

Nice-to-have Technologies:

Spark Graphframes

MLFlow

Koalas

Redash

Spark-NLP package from spark-packages

=================================Project-X3===============================

Build a patient sickness/wellness prediction module using PySpark (clustering module and/or another model)

**DataSet (CSV):**

<https://drive.google.com/drive/u/0/folders/1iqmypphulNV66QwfeHD14WloFJo8vApE>

Dataset Relationships:

<https://docs.google.com/document/d/1VbA5APdc4xQSvCAv0Y2WXNQlnZp9c8YMEb9Orjt3RLA/edit>

Optional:

FHIR data

**Must-have Technologies:**

<https://spark.apache.org/docs/latest/api/python/pyspark.ml.html>

<https://spark.apache.org/docs/latest/api/python/pyspark.sql.html>

Clustering Module

Tuning Module

Evaluation Module

Nice-to-have Technologies:

Spark Graphframes

MLFlow

Koalas

Redash

Spark-NLP package from spark-packages

=================================Project-X4===============================

Build a “promoting pediatric primary prevention model” using PySpark (sql module and/or other module like ml.stat module)

* Number of well-child visits (number of visits/number of children in the target population)
* Well-child visit disparities (comparing measures across two or more subpopulations)
* Doses administered of specific vaccines; coverage of the primary pediatric vaccination series (for example, number of MMR vaccines given/number of children 12-18 months in the target population)
* Disparities within the community on a specific immunization metric (comparing measure across two or more subpopulations)
* Optional - personalized guidance on healthy nutrition, exercise and safety.
* Optional - schedule recommendation for kids to be seen in primary care as soon as possible for catch-up vaccination.

**DataSet (CSV):**

<https://drive.google.com/drive/u/0/folders/1iqmypphulNV66QwfeHD14WloFJo8vApE>

Dataset Relationships:

<https://docs.google.com/document/d/1VbA5APdc4xQSvCAv0Y2WXNQlnZp9c8YMEb9Orjt3RLA/edit>

Optional:

FHIR data

**Must-have Technologies:**

<https://spark.apache.org/docs/latest/api/python/pyspark.ml.html>

<https://spark.apache.org/docs/latest/api/python/pyspark.sql.html>

SQL Module

Tuning Module

Evaluation Module

Nice-to-have Technologies:

Spark Graphframes

MLFlow

Koalas

Redash (Trend Analysis and slice/dice analysis using sub-populations - <http://www.h4-technology.com/nhis-system/>)

Spark-NLP package from spark-packages

De Identification

PPML

Baseline

https://towardsdatascience.com/pyspark-forecasting-with-pandas-udf-and-fb-prophet-e9d70f86d802

ACIP Recommendations (https://www.cdc.gov/vaccines/acip/recommendations.html)

Vaccination Schedules (https://www.cdc.gov/vaccines/schedules/hcp/index.html)

=================================Project-X5===============================

Build an encounter cost estimator/predictor engine using Tensorflow/MMLSpark

**DataSet (CSV):**

<https://drive.google.com/drive/u/0/folders/1iqmypphulNV66QwfeHD14WloFJo8vApE>

Dataset Relationships:

<https://docs.google.com/document/d/1VbA5APdc4xQSvCAv0Y2WXNQlnZp9c8YMEb9Orjt3RLA/edit>

Optional:

FHIR data

**Must-have Technologies:**

Keras

Tensorflow

Tensorboard

Nice-to-have Technologies:

TFX

TFE

TFF

TF Serving

TF Cloud

Horovod

MMLSpark (https://github.com/Azure/mmlspark)

AutoKeras