Work Breakdown Structure - FullContact

Tasks

1. Become Familiar with FullContact's Identigraph and data
2. Write Spark jobs to track inputs and their effects on the graph
    a. RawObservations
    b. ObservationEdges
    c. CrudeEdges
    d. RefinedVertexes
    e. RefinedEdges
    f. EdgeWithPropogatedNegatives
3. Develop an application that returns insights like the path of the data through the graph based on input records
    a. Create a microservice to submit a spark job to the EMR cluster
    b. API design
    c. Return query information on each Identigraph job
    d. Add functionality to add and delete query information from database
4. Add Functionality to the app so it returns insights based on output clusters as well
5. Create a web based, interactive UI that make it easy for users to make queries to the app and get response in minutes

Features

- Web-hosted user interface: Available internally on FullContact's AWS-hosted cloud infrastructure and presents a UI to make common queries easy
- Big Data Scale: Operate on multiple-terabyte datasets of parquet and sequence files
- Cost: Run on <$500 of AWS resources per month
- Interactive: Respond to user queries in minutes, rather than hours or days
- Bidirectional: Allow users to generate insights on graph behavior using either input records or output clusters as a starting point