|  |  |  |
| --- | --- | --- |
| Bangalore Shivacharan – CEO | David Hong – product manager | Nayanjeet Medhi |
| Jaikumar Madhava | Satya Govindu | Jaychand |

**Approach**

Take FDA provided XML data to work with. OpenFDA API based data mining will take time but it is still being pursued as Plan B (see bottom of the mail). Also, plan to have MySQL replaced with Maria (its  the pure open source equivalent)

**There are three teams involved now – Hadoop team (Akshay/Pradeep DV/Shridhar), REST API team (Pradeep TR), UI team (Jagan/Sajil)**

**Team/Technologies/Responsibilities**

**Hadoop Team:**

  Download XML files from FDA site and convert them into CSV files (denormalize)

  Load them into Maria DB for research purpose

  Set up Hadoop cluster on US Marvin server

  Load CSV data into Hadoop file system

  Create “Dimension” data schema on Maria DB

  Write PIG scripts to generate “dimension” data (counts or summarized data)

  Manually analyze counts which are generated in Maria

  Write PIG scripts to generate “Mean/Average” and “Spike incidents” on a per week/month/Quarter/Annual basis and also per Severity/Seriousness of reactions

**REST API team:**

  Build basic framework for REST API using Spring Boot

  Analyze Maria “Dimension” DB and establish JSON contracts (input) for different REST APIs

  Start exposing REST APIs

**UI Team:**

  Put together AngularJS initial project set up

  Start understanding at Maria DB “Dimension” DB schema

  Start understanding REST API

  Put together HTML5/Bootstrap mockups and get them ready for review

  Look at JS based reporting tools

Tasks for Hadoop team:

1.       Load the FDA XML dump into Hadoop and use it for the back-end processing

2.       The star schema will define how we organize the data

3.       The chart can be plotted based on the star schema

4.       We need to develop an algorithm that helps us identify the peak

5.       Once the peak is defined, we can make use of the OpenFDA APIs to correlate the Drug and Enforcement data to be presented on the UI

6.       Process related –

-          Git hub repository – Everyone can clone the project and check into their own branches and the admin merges the changes onto the master

-          Nayan to look at the Digital Services Playbook and make sure that we follow the rules of the game

7.       Work breakup for today –

-          JC and Jai to review and work with Sajil to define the STAR schema  dimensions to be used

-          Jai and Sajil to come up with an algorithm to identify the spike on the graph

-          Nayan and Sagnik to understand the data come up with the common fields which could be used to correlate the Event and Enforcement data

-          Srikanth to take over from the India team the data ingestion process and load as much of the XML data into the local Hadoop repository