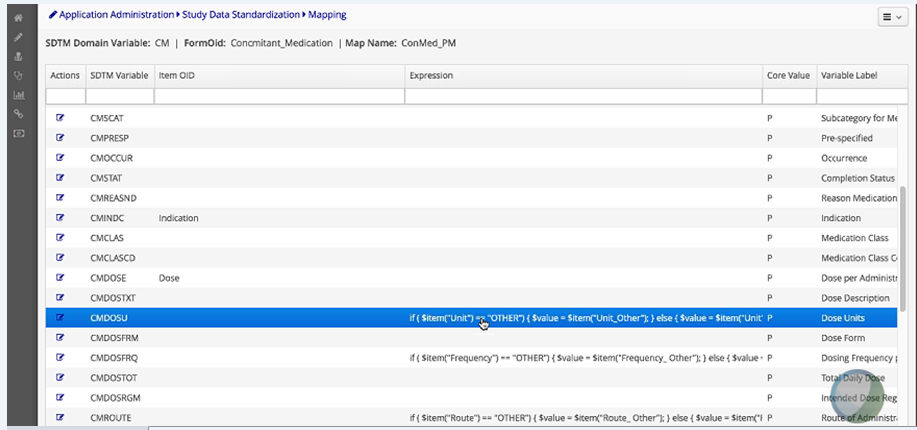
The Project to to Develop the Risk Based Analytics:

1. This is a data for Pharmaceutical Company
2. The company develops drugs and tests the safety and effectiveness (efficacy) of the drugs on human population and submits the results to FDA for approval
3. The work flow is described as below:
4. The Testing Drug is called “Study”
5. The Study has a defined layout as defined in a Protocol
6. For example a Cancer Drug. Study is called “test study” and the Pharmaceutical study will send the Drug for this study to different Sites (Enrolled Sites, Sites are Clinics across country or globally).
7. The Drug is dispensed at these Sites by the doctors to Subjects (patients) who are enrolled into the study
8. The Drug is dispensed and study is conducted at these sites based on the protocol that the sponsor (pharmaceutical company) provides/ instructs the sites
9. The Protocol written by Pharma company for the study will have the Visits defined (that is the time intervals the patients should come to the clinic, the assessments they should go through at each visit (for example, Height and Weight is collected only at Visit 1 and Drug is dispensed at Visit 2 and Any symptoms happening to the patients while on this drug is called as Adverse Events data)
10. The data is collected at the Sites in an Electronic Data Capturing System
11. So this is the basic structure
12. Study is a Protocol written at Pharma company example, Brain Tumor study
13. Sites are Enrolled or Recruited by Sponsor (Pharma Company) to conduct the study
14. Sites Enroll the subjects (patients) based on the inclusion / exclusion reasons, age of patient etc etc as defined in the protocol
15. Subjects in the study come to the clinics (sites) as per visits (week1, Week 2 etc) defined in the protocol to get the assessments (labs, vital signs, study drug given by sites)
16. So the Sites conduct the study for the Phama company which is a sponsor and the data is entered in Electronic data capturing system. The Sites need to be monitored by Pharma company that they are conducting the study properly
17. Now the Sites need to be monitored regularly that they are conducting the study for the Pharma company and the Patient data is getting recorded properly by the Sites. Sites get paid by Pharma company
18. Now the Sites need to be Monitored and this is the Project
19. Site Monitoring
20. Site Monitoring is done by people employed by pharma company or it is outsourced to a CRO (contract organization)
21. However, we need to do analytics on Site Monitoring and site Performance so that the study is completed at minimum cost
22. This is called Risk Based Monitoring. Not ALL data collected at Sites can be monitored. So we need to make assessments as to which data is most critical which need to be 100% accurate etc. Hence we do Risk based monitoring instead of all monitoring because Sites when monitored there are millions of records and cannot be monitored on all data
23. We need to develop a system that will help the Pharma company do the analytics and decide which site is performing very bad and requires more monitoring. For this we need to create the Risk Plan. The industry follows a guideline that has been published (it is called Trancelerator guide. I am forwarding that guide to you. Briefly go through the guide. It will help you understand the Risk categories. The guide is a reference for all companies to follow a Risk Plan )
24. For this we need to create a Front end Application:
25. The application will have screens to input Study Information.
26. Site Information Tab. The Sites assigned to the Study (example, Study is Brain Tumor. Sites are California Clinic, Arizona, Philadelphia Memorial Hospital etc etc). Sites are stored as Site ID and Site Name, Address. Sites have Principal Investigators (Doctors) who recruit Patients (subjects) for the study
27. The Sites associated to the study are uploaded from other system (The other system is Electronic Data Capturing system). We can get this information into this system through APIs
28. The Patient information is recorded in the Electronic Data Capturing system and this information is pulled into our new system through APIs
29. The Sites are monitored by Site Monitors or by CROs. The data collected by Site Monitors as monitoring reports are uploaded from their systems through APIs into this system we build
30. When we define the Risk plan in the system, we assign Risk Categories. For example The categories are Data quality, Safety etc. Data Quality is the quality of data that are collected at the Site. For example, if there is a page of data that needs to be filled into Electronic Data Capturing System for a Subject , like temperature value, Weight , height etc. This data is going into Vitals Page. If there are some fields that are missing a data then it is **missing data** or if the page is not entered for a subject then it is **missing page.** If there are queries on data (for example weight entered value is out of range for the subject, data validation checks are fired and results in **Query issued.** So there are the statistics on the data quality we can pull into our system.
31. For Data Quality category we need to assign a Weight (%). And define the threshold value. For example, Data Quality, Number of Open Queries. The threshold is min 5 and max 25. Weight is 25%. We need to write Triggers for these categories. Once we set up this trigger, we pull the data into our system from Electronic Data Capturing System then it will ve I number of open queries that are not yet answered and this statistics we can assign to this category and the report will turn red traffic light for the Site, for this category and will need more Site Monitoring (we need to build the dashboards on all this data and one of the dashboard is screen shot I included. Here the Sites have traffic signals based on the Triggers we have set up) .
32. The vision for our system will be to read ALL the data from different systems into one standardized database. We will create the mapping into one common structure , irrespective of where the data I coming from, different sources, different formats. This data Is then read into our system that we will build and based on the risk plan we define, the reports for Site Metric Dashboards are generated.
33. Capture.PNG and Capture 2. PNG describe the Study Data Standardization. When the raw data from Electronic Data Capturing System is read into our system, we need to standardize it. The data can be coming from different studies through different systems or though different data transfer file types. Our system should be able to read any source data. However, to read source data into our system into our dashboards etc, we need to maintain 1 standard. This is Study data Standardization tab. We need to create the study data standardization tab. For each study, the Admin or user, should be able to create mapping (fields from source data and mapped to defined tables and variables in our system) for example in screen shot below, the “SDTM” column is standard column in our database. The dataset t(table ) name is CM. The Expression column will have the logic to map. Sometimes it is one to one map. For example, source data file is CM01 and field name is CMSTATUS. We need to map into our standard structure as CM (table ) and Field name is CMSTAT

So for CM, Column SDTM shows our standard variable. Mapping into CMSTAT is CM1. CMSTATUS.

Or if it complex mapping (data coming from different forms and different variables in source and needs manipulation, then we write the complete logic in Expression” column, and with the click of the button , the logic in Expression column for that variable needs to be derived.



**My Questions to you:**

1. For the purposes of the Pilot, will you be able to build quick front end screens to capture the input for Study, Sites, Risk Plan , Risk Categories, Assigning the Weight and Triggers
2. Can you also build a couple of dashboards from the data I will send. That will be in Dashboards menu item
3. The look and feel of the application is like the screen shots I am sending. Please make a little modifications to look and feel so it does not look like exact copy of some other company application
4. Processing the data through Big data analytics is not needed in first 2 weeks. I need the front end application to be developed first. So we can show how it functions as a prototype.
5. Please explain to me how does MongoDB and Hadoop work?
6. I included the model for our project ….screen shot