The Great Mind Challenge - Project Scenario Template

Note: Already filled information should not be changed

1.	Name of the Project	Generic Data Analyzer
2.	**Objective/ Vision**	Data analysis and mining of huge multi-terabyte data sets generated in the current world of Internet advertising
3.	**Users of the System**	Any System generating huge amount of data, just to name a few : A. Weather reporting System B. Traffic Monitoring System C. Web Crawlers
4.	**Functional Requirements (Atleast Eight)**	i. Should be pluggable for any type of text data available ii. Should have the facility to specify the format if input structures of the data and adapt accordingly iii. Provide functionalities to access distinct elements/top hitters or frequent occurrences/occurrences in a specified range which can be dates or values for example iv. Should be able to understand new queries (in the predefined format) v. Must be scalable and be able to access data spread across machines. vi. Should use distributed data stores like Hadoop and use advanced techniques like Map Reduce vii. Should have the ability to tag unique data sets viii. Response time should be reasonable and provide detail analysis of different task executions done
5.	**Non-functional requirements (Atleast Four)**	i. Secure access of confidential data (user's details). SSL can be used. ii. 24 X 7 availability iii. Better component design to get better performance at peak time iv. Flexible service based architecture will be highly desirable for future extension
6.	**Optional features**	a. Easy to operate and understand b. Include an alerting System (email or sms alerts) if any abnormal/serious behavior detected in the data analyzed
7.	**User interface priorities**	A. Professional look and feel B. Use of AJAX atleast with all registration forms C. Browser testing and support for IE, NN, Mozila, and Firefox.

		D. Use of Graphical tool like JASPER to show strategic data to admin E. Reports exportable in .XLS, .PDF or any other desirable format
8.	**Reports**	**A. Generate reports for all the functionalities in the system**
9.	**Other important issues**	**A. Use a NoSql data store which is easily scalable rather than traditional RDBMS**
10.	**Team Size**	**2 - 4**
11.	**Technologies to be used**	**UML, J2EE, XML, e-Forms, AJAX, Web 2.0, Web-services, SOA**
12.	**Tools to be Used**	• **ROSE/RSA / WebSphere Modeler** • **Eclipse/ RAD / Lotus Forms Designer / Portlet Factory** • **WebSphere Portal/ WAS/ WAS CE / WPS** • **DB2 Express – 'C' or DB2 UDB** • **Tivoli CDP/TSM / Tivoli Directory Server** • **Linux will be the preferred OS.**
13.	**Final Deliverable must include**	**A. Online or offline help to above said users, Application deployment executive and developer** **B. Application archive (.war/.ear) with source code** **C. Database backup and DDL Script** **D. Complete Source code**