CS486C – Senior Capstone Design in Computer Science
Project Description

Project Title:	Data Visualization Analytics on Enterprise Storage	
Sponsor Information:	**Forrest Townsend** fetownse@us.ibm.com Daniel Boros dboros@us.ibm.com	Christopher J. Ruskay Software Development Manager IBM Spectrum Protect 520.799.4086 ruskay@us.ibm.com

Project Overview:

Eighty percent of the world's data is unstructured – pictures, media files, movies, documents, et cetera – data that doesn't fit well into a traditional row and column model. As the world's data volume increases, data is forecasted to become more unstructured at a rate of sixty to seventy percent per year and enterprises are not able to digest all this information in a cohesive way. The amount of data in the world is increasing exponentially – these enterprises need ways to understand the data they manage to make better business decisions and reduce costs. Data visualization can provide these representations of the underlying relationship of data and display it using interactive graphics which gives the customer more metrics and information with which to make better decisions.

IBM has established itself as a leader in cognitive solutions and cloud as-a-service offerings which are driven by innovation in big data analytics. Analytics paired with visualization allows customers to respond to change much quicker by visually understanding their situation and reactively make appropriate business decisions that could save and make them money.

For example, consider this scenario drawn from our team's specific area of focus: a customer is backing up data and they are missing their backup windows... that means they are at risk of not having a copy of their data backed up and they have zero insight as to why this is happening. It would be beneficial if they had some way to view the health of their enterprise storage network and configuration at a glance. This would allow the customer to actively visualize resource usage, discover pain points within their current storage configuration, and forecast how their data and storage needs will trend in the future.

The customer may find – after much diagnosing and support from IBM – that they do not have enough fast, expensive disk to fulfill their current backup requirements. This could have been prevented by visualizing the relationship between backups schedules and storage media, in this case revealing the fact that they do not have enough high-speed resources in order to service the current backup requirements. This could have enabled the customer to realize upfront that their current backup expectations are unrealistic in their current setup and allow them to actively make the changes necessary to fulfill business needs.

On the other hand, the customer may also be interested in how their backups are distributed across their storage network. They would like to see a visualization of how much data is being stored on high- cost media in order for them to allocate that data somewhere else and use the high-speed storage for the primary backups, increasing the efficiency when restore time comes.

In the realm of storage and backup, your proof of concept will provide customers the ability to visualize data in an enterprise network. This will provide insight into the current state of their data and how that data might impact their infrastructure. Ideally, the solution might also provide some sort of "what-if" analysis to assist in planning system modifications (additions or deletion of capacity).

The nature of the project is open-ended and somewhat exploratory, but definite deliverables will be expected by the end:

- As an overall project model, you will create a secure Web Application that will provide the overall interface to the product.
- The product will be rooted in a "dashboard" where users are taken upon logging in. This dashboard will contain some key overview "status" graphics to show the overall state of storage system. It might possibly also be configurable to display specific summary data desired by the user.
- The dashboard will also serve as access to a variety of visual tools for exploring and more deeply analyzing the state and performance of the user's storage system. In particular, you will be expected to create a set of visualizations for classifying data in order to provide meaning to a customer's data set in a large scale enterprise backup solution. Creativity on your part will be expected…but, of course, we will ultimately work collaboratively to discuss and settle on targeted visualization during the design phase.

Knowledge, skills, and expertise that will need to be learned for this project:

As usual, the proficiencies that will need to be developed to tackle this project are not entirely known in advance and will be clarified as we progress through requirements and design. They are likely to include:

- Knowledge of Web2.0 architectures: frameworks, cloud services, etc.
- JavaScript and general web application knowledge.
- Suggested knowledge of D3.js.
- Added bonus: doing visual testing with Selenium.
- A willingness to collaborate with your team and us.

Equipment Requirements:

- There should be no equipment or software required other than a development platform and software/tools freely available online.

Software and other Deliverables:

- The specified webapp, installed and running on a platform determined during the development process.
- A personal demo for sponsors of product and all of its features by the team. In-person or via teleconf.
- A GUI that adheres to good design principles, along with a viable end-user manual.

- A strong as-built report detailing the design and implementation of the product in a complete, clear and professional manner. This document should provide a strong basis for future development of the product.
- GitHub, BitBucket, or some other version control repository which we can view. Code should be fully documented to a professional level.