Indiana University, Bloomington
Data Science
INFO I591 – Graduate Internship
Internship Report

Samvat Rastogi
samrasto@indiana.edu

About Company:

Palo Alto Research Center (PARC), a Xerox Company, is a research and development center based in Palo Alto, California. It was founded in 1970 as a part of Xerox Cooperation and has been responsible for some major breakthrough inventions of the time like laser printing, object oriented programming, Ethernet, Graphical User Interface[1].

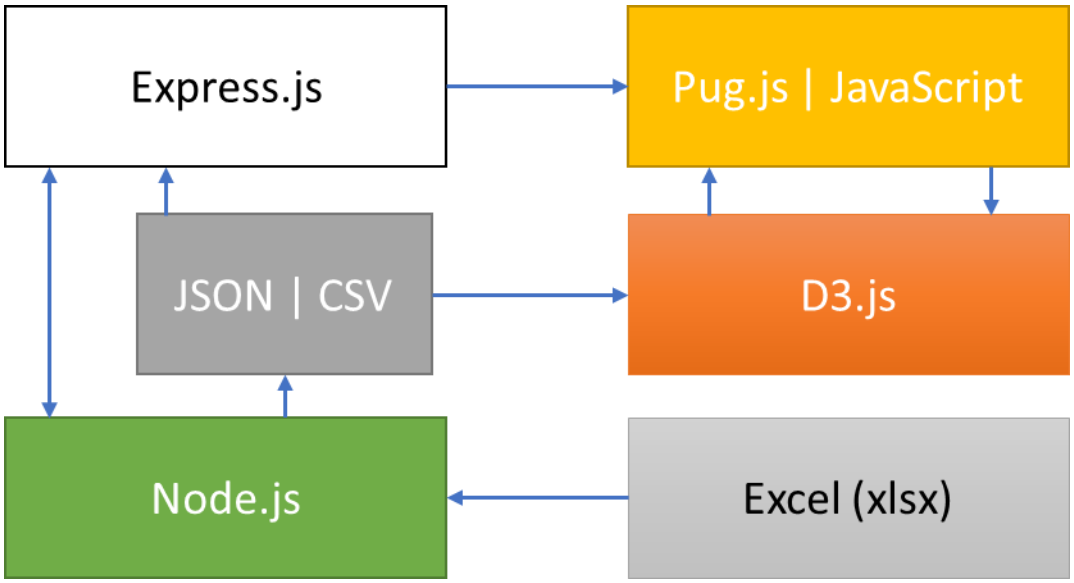
About Internship:

PARC offered me a position of Data Visualization and Simulation Intern at Palo Alto for 11 weeks. The main objective of this internship was to visualize the output of a simulation of energy management in a closed system. This visualization was a web dashboard based on D3.js[2].

Technologies used:

As this dashboard was web-based, a completed javascript stack was used. I used Node.js, Express.js, D3.js, Pug.js with vanilla javascript to develop the complete application stack. The data was available in Excel (.xlsx) format.

Dashboard Stack:



This application uses Node.js as the base of the stack with two data layers. The first data layer serves as an input to the application and the second layer is the processed data for the D3.js at the client. It uses Express.js for communication and routing. The front end extensively uses Pug.js, Bootstrap and JavasScript. The input data is the output of the simulation and is available as flat files (.xlsx) to the dashboard. The processed data is formatted as JSON and CSV to keep the data flow inside the application intact. The JSON data acts a backup of the actual

data and serves any data request made within the application where CSV acts as input to D3.js. The front end is rendered in HTML5 using Pug.js, which is an HTML rendering engine.

So the basic function of the application was to read data, process it using a custom data pipeline and display it on a webpage. The application will be used by a client of PARC to determine the best scenario out of a pool of scenario which can happen with the closed system. Here the closed system is defined as a system with limited resources which may exhaust if the system remains to be in the closed state.

The visualizations using D3.js is interactive and have features like filtering, mouse over data display. Each visualization like stacked area graphs, multi-line charts, pie charts uses a dedicated data pipe to avoid any performance issue with the dashboard.

Inspiration:

Data visualization has been a part of my interest since 2013. Prior to my masters, I have used D3.js as part of my industrial experience. This was one of the major reason, I opted for Data Science at IU. My interest in Data Visualization was magnified by many folds when I took Information Visualization course under Professor Katy Borner. This course gave me a complete freedom to experiment with visualizations and helped to learn various visualization technologies like Tableau, Sci2, Shiny, Gephi and Processing.js. This prompted me to apply for data visualization specific roles for an internship as I maintained a GitHub repo (https://github.com/samvat/Z637-Info-Viz-Spring2017) for this course.

This internship proved to be a very integral part of my data science curriculum as I was able to implement my learning in the internship and was able to learn more in developing visualization. I discovered a whole new aspect of data visualization at PARC. This opportunity in a research center provided me a blend of the complete freedom to experiment, like an academic project and provide the quality oriented result, like an industrial project.

I strongly believe that this internship opportunity will hugely contribute to my future aspirations. As the data visualization is the language spoken by data analytics, a strong hold in data visualization will help me a lot choosing my career path post my masters.

Challenges:

1. As JavaScript is a loosely typed language, I faced a major challenge in dealing with various data structures. But with time, I was able to get a hold on it and overcome it.
2. My prior experience did not give a hands-on experience with Node.js and its stack. However, this application gave an exposure to same to a very great extent.

Future Work and Current Status:

1. Adding drill down graphs to view lower level data or unaggregated data.
2. The audience of this dashboard is supposed to be technical. So adding a parallel dashboard for business analysts or non-technical audience.
3. Current Status: Completed and gave the knowledge transfer to the current project team.

References:

[1]: Wikipedia, Xerox PARC (https://en.wikipedia.org/wiki/PARC_(company))

[2]: D3.js (https://d3js.org/)