Report 3: Final Assessment

Please complete all fields.

Name:	**Rahul Ethiraj**
Company:	A9.com, Amazon
Department/Team:	Advertising Technology – Newton team
Project Name:	Amazon Experimentation Framework & metrics- AXF

1. How did you like working for this company?

- The tasks assigned to me for an intern role is really appreciated. I felt that the work that I have been doing is important and would create an impact on the production codebase.
- I had worked on big-data pipeline previously and would love playing with data and performing analytics on it. I felt home being here for the fall internship and would love to come back if given the opportunity.
- I felt happy working for this company and satisfied when I was able to finish up the assigned tasks.

2. Did you complete your project goals?

- Yes, I have completed my project goals.
- The final goal of the project is to populate Amazon Ad Exchange (AAX) metrics report based on hourly logs generated by AAX server within one hour.
- This involves creating a pipeline that generates AAX metrics and also calculates statistical metrics such as p-value and confidence intervals.
- The central idea is to prototype processing AdRunway logs in AWS Elastic Map-Reduce (EMR) cluster using Spark.
- I first spent some time onboarding by reviewing SDE bootcamp sessions and learning basic advertising jargons. I set up the development environment, learnt Spark, explored other existing alternatives for log processing, and came up with an initial design doc. I iteratively improved upon the design based on feedback gathered from several conversations with mentors, teammates, and manager to arrive at the best possible design to successfully complete the project.
- Second milestone involved development of actual code and the deliverables were further broken down into first parsing only a subset of logs and next parsing logs in entirety – bid, impression, click, punt, nobid with source groups adx, mobile_rtb, other, rmx, rtb_other, video_rtb, and ttypes mobile and web.
- The third milestone was to productionize Spark jobs. I used Remo Aegis, created a production Apollo environment, pipeline, package group, version set, DJS job, DJS Agent, and set up Glue crawler and Athena.
- As part of the fourth deliverable, I wrote a package called statsEngine that computes basic statistical metrics such as p-value and confidence intervals. Computing variance was not straightforward as data had to be processed in batches every hour.

3. What experience did you gain?

- The assigned project had several clear milestones – design, development, productionization and statistical metric calculation.
- I have gained so much experience as part of this fall internship ranging from technical – code revision, code testing, automation, debugging, and non-technical - communication with peers, teamwork, team building, formal communication – oral and writing.
- I was exposed to many big-data technologies such as divide and conquer, parallel processing, Hadoop Distributed File System, and Spark programming knowledge.
- I also gained many platform and usage of tools knowledge such as using Elastic-map reduce clusters, s3, Quicksight, Athena, AWS Glue, installing and managing containers and distributed job scheduler.
- Furthermore, I gained knowledge by setting up the entire production spark job in the environment and created a production pipeline having Alpha, Beta, and PROD stages.

4. In what ways do you feel you were well prepared for this job?

- I have worked with Cisco for almost around three years in Server Virtualization team. This work experience has helped me a lot in the first initial days of work.
- I have also completed certification courses like AWS, cloud, big-data which has helped me in understanding the existing AWS environment with ease.
- I have taken courses relating to Advertising, such as CSCI572 Information Retrieval and Web Search Engines by Professor Horowitz, which has guided me on how to approach programming for any Information Retrieval or advertising technologies relating to search engines; web crawling, indexing, querying and quality of results.

5. In what ways do you feel you could have been better prepared?

- My learnings from CSCI 572 at the University of Southern California was well put to practical use in my fall internship.
- My knowledge about big-data was only at basics before joining the team. Since I was put in the Advertising technology domain, I think any in-depth Advertising-Technology course from USC would have helped me to ramp up with the team quickly, since CSCI 572 touched upon basics of Advertising but not in-depth.

6. Other comments?

- My manager Nagesh Balivada kept regular 1:1 meeting every week, to ensure that I am going in the right direction and there are no active blockers that would hinder my project. My team members and mentor were very helpful and I cannot thank them enough for the help that they have done.
- I don't have anything to list about things that didn’t go well, as everything was well-taken care from onboarding to exit formalities.
- Since the project was broken down into surmountable components, I could easily acquire each milestone. Experiment users can now investigate raw data, analyze results and gather statistical measures using statsEngine. These metrics can be used to assess the statistical significance of the difference in behavior` observed between control and treatment groups. The project is winnable, real and worth it.