MANISH SOMBANSH

412-773-0606 | msombans@andrew.cmu.edu | 210 Wendover, 5562 Hobart St., Pittsburgh, PA 15217

EDUCATION

Carnegie Mellon University, Pittsburgh, PA — Aug 2018 – Dec 2019

Master of Information Systems Management

Relevant coursework: Cloud Computing, Distributed Systems, Object Oriented Programming, Object Oriented Analysis and Design, Data Mining, Statistics for IT Managers, Database Management, NoSQL Database Management, Linux & Open Source, Web Application Development, Agile Methods

BITS Pilani, India — May 2016

Bachelor's in Information Systems

WORK EXPERIENCE

Carnegie Mellon University, School of Computer Science – Graduate Research Assistant — Sep 2019 – Present

- Working as a project member in the TEEL (Technology for effective and efficient learning) Lab of CMU.
- Responsible for developing learning platforms focused on Cloud Computing technologies in collaboration with Microsoft.

Salesforce, San Francisco – Software Engineer Intern — May 2019 – Aug 2019

- Developed a containerised performance load tester service following the microservices pattern using Spring Boot.
- Built and exposed multiple REST endpoints varying from triggering a performance test, deleting test results and fetching test results from a containerized environment.
- Developed an auxiliary service to access the primary performance service due to VPC restrictions and integrated with Jenkins.
- Deployed the service over Core PaaS platform (AWS) of MuleSoft using Docker and Kubernetes.

Oracle - Applications Engineer II — July 2016 – July 2018

- Analysed code independently for performance bottlenecks, tuned SQL queries, identified underperforming java methods which consumed a lot of time under high stress and rewrote them using efficient data structures. As a result, the efficiency of our product improved by nearly 25%.
- Worked on enhancements across multiple releases using various areas of Oracle ADF like task flows, view objects, entity objects, application modules, transaction management, beans, application scopes, lifecycle , JSF pages and developed RESTful Web Services - (GET and POST) for our product.
- Trained fellow team members of varying experience levels on capturing performance reports and analysing them.
- Awarded "AssimilatOR" by Rick Jewell, Senior Vice President of Application Development for my outstanding contribution towards Fusion Supply Chain organization.

Oracle - Software Development Intern — July 2015 - Dec 2015

- Gained experience on building Fusion Cloud Web Applications and worked on developing the entire Financial Trade Agreements User interface from scratch.
- Helped the development team eliminate close to 30% redundant code in the existing application.

TECHNICAL SKILLS

Languages: C++, Java, Python, Scala, SQL, R

Databases: MySQL, Oracle 11g, Mongo DB, AWS RDS, AWS Aurora, Neo4j, HBase, Elasticsearch

Cloud: MapReduce, AWS (EC2, S3, EMR, LightSail, CloudWatch, ELB, Lambda), Azure, GCP, Heroku, Docker, Kubernetes, Redis, Apache Spark, Kafka, Samza, Terraform, Helm, Jenkins

Tools: GIT, NetBeans, Spring Boot, Flask, jMeter, IntelliJ-IDEA, R Studio, Eclipse, JIRA, Android Studio, Confluence, Oracle ADF, ADE

PROJECTS

Scalable High-Performance Twitter Analytics Web Service (AWS, Java, HBase, Spark, Scala) — Mar 2019 – Apr 2019

- Implemented a high performance, fault-tolerant web service on AWS EC2 clusters for 1TB Twitter data analysis with read queries and achieved over 25,000 RPS on a limited budget given to work upon.
- Processed 1TB raw data using Apache Spark and Scala to perform ETL and load data into MySQL and HBase database systems. Used Undertow framework and load balancer for front end.
- Employed backend using MySQL and HBase, improved performance by implementing several optimizations ranging from schema design, database parameter tuning, horizontal scaling, replication and sharding, and monitored performance using AWS CloudWatch to identify performance bottlenecks.
- Migrated the web service to use managed web services provided by Amazon in the final phase of the project. Used AWS LightSail for the front end and RDS, Aurora for the backend.

Machine Learning on the Cloud (Python, GCP) Apr 2019

• Performed data exploration, cleaning and feature engineering for building a machine learning model using XGBoost algorithm to predict cab fare prices in NYC.
• Deployed the model onto Google cloud machine learning engine and used Hyperparameter tuning to optimize the model and increase accuracy.
• Built a fare prediction web service using Google App Engine to predict fare prices on basis of attributes like pickup_datetime, pickup_longitude, dropoff_longitude etc by using the model hosted on ML engine.
• Developed a machine learning application which accepts speech queries and responds with a speech result in the context of fare prediction by using GCP API's (text to speech, speech to text, NLP, Google Maps).
• Developed a machine learning application which accepts speech queries and responds with a speech result in the context of fare prediction by using GCP APIs (text to speech, speech to text, NLP, Google Maps). Enhanced the fare prediction API by accepting images of landmarks as the trip pickup and drop off locations using Google Cloud Vision and AutoML API's.

AWS Cloud Resources Autoscaling (AWS ELB, Java) Feb 2019

• Developed solutions that invoke cloud APIs to programmatically provision and de-provision resources based on the current load to optimize cloud resource usage and minimize costs.
• Configured an Elastic Load Balancer and Auto Scaling Rules in AWS to distribute load across cloud servers and provision against resource failure.

Uber Like Application for Rider-Driving matching (Java, Apache Kafka, Samza, AWS EMR) Apr 2019

• Generated streams of data using Kafka Producer and made it available for Samza consumer.
• Designed and implemented a realistic solution to join and process multiple streams of GPS data using Samza API to setup a driver matching service like Uber.
• Developed an application to process streams of IoT device data and static data using Samza to setup an advertising system of restaurants for the rider once he gets into the cab based on his interests.

Functions as a service – Video Processing Pipeline (AWS S3, SNS, AWS Lambda, FFmpeg) Feb 2019

• Configured and provisioned AWS S3, Lambda and SNS with a thumbnail generator function using Java to generate thumbnails for uploaded videos.
• Used AWS Lambda and Rekognition API to detect labels from the thumbnails and use them to classify an image.
• Implemented a video searching feature for a YouTube like application using AWS CloudSearch for users to search videos.
• Generated gifs for uploaded videos when user hovers on a particular video search result.

WeChat As A Service (Java, AWS, Azure, GCP, Docker, Kubernetes) Feb 2019

• Created a WeChat like application with Microservices pattern; Designed and developed 3 services, login, group chat, and profile; Modeled the interaction between services as directed acyclic graph.
• Implemented the REST APIs of login, group chat and the profile service using Java, Spring web-socket, Helm and used Redis Pub/Sub as a messaging service to improve the scalability of the group chat service.
• Deployed the services using Docker, Kubernetes, and Helm across multiple cloud platforms (GCP and Azure) for fault-tolerance; Monitored CPU utilization metrics, automatically augmented the pods and added the auto-scaling feature with Kubernetes HPA.

Social Networking Timeline with Heterogenous Backends (GCP, MySQL, Neo4j, MongoDB) Mar 2019

• Configured, populated and deployed heterogeneous SQL and NoSQL, databases to build a mini social networking website.
• Implemented login feature using MySQL. Stored social graph relationship using graph database Neo4j to fetch list of followers for a particular user. Fetched posts and comments made by a social user on a user's timeline using MongoDB.
• Built a timeline page by integrating all previous features to demonstrate a mini social networking site like Facebook.

Distributed Key-Value Stores with Strong (Causal) and Eventual Consistencies (Java, AWS) Mar 2019

• Implemented a replicated cross-regional cloud storage system on AWS to handle requests using strong and eventual consistency on basis of timestamp.
• Used multithreading and priority queue for controlled access to the key value store and synchronizing all data centers.

Most Influential Users from Twitter Social Graph (Spark, Scala, Azure) Apr 2019

• Implemented an iterative computation on a Twitter social graph (10GB) to rank each user by their influence using the graph-parallel framework GraphLab to implement the PageRank algorithm using Apache Spark and Scala.
• Deployed the Spark program on Azure HDInsight and optimized it to meet performance objectives by tuning parameters in Spark configuration and writing effective code in Spark-Scala.
• Deployed the same application on Azure Databricks to understand performance boosts and platform improvements when compared to Azure HDInsight while running Spark jobs.