

Manav Bagai

mbagai@asu.edu | (480)-570-8392 | <https://www.linkedin.com/in/manav-bagai/> | <https://manavbagai.github.io/>

Education

Masters in Computer Science

Arizona State University, Tempe, Arizona, USA

May 2020 (Expected)

GPA: *In progress*

Bachelor of Technology in Computer Engineering

Aligarh Muslim University, Aligarh, India

June 2016

GPA: 8.1/10

Experience

Big Data Engineer, Exadatum Software Services, Pune, India

November 2016-February 2018

• Ingestion, Analysis, and Visualization of Cardiovascular Patient Data

June 2017- February 2018

Technologies: Python, Scala, Play Framework, Hadoop, Spark, Docker, Airflow, Druid, Superset, SaltStack, Jenkins, AWS

- Related to Healthcare domain in a 3-member team.
- Worked on creating Application Docker with Java, Python, Scala, Hadoop, Spark, Airflow, and Zeppelin.
- Responsible for creating and deploying Database Docker with Druid and Superset installed.
- Contributed towards designing and developing ETLs in Spark and Scala.
- Contributed towards optimizing the data algorithm to work for large data and in distributed mode in PySpark.
- Responsible for creating and maintaining client environment which includes AWS, Docker, Jenkins and SaltStack.
- Responsible for orchestrating the whole flow using Apache Airflow.
- Developed a micro service in Play Framework which acts as middleware between UI and Airflow.

• Recommendation System and Chat-bot

November 2016- May 2017

Technologies: Java, Python, Neo4J, Nutch Crawler, Play Framework, Elastic Search, API.AI, AWS

- Related to Retail domain in a 4-member team.
- Worked on creating the knowledge base by crawling web data using Apache Nutch. This involved writing an HtmlParseFilter plugin in which the useful data was extracted with the help of regular expressions.
- Written Python and Pandas scripts to clean the data and generate required csv and Jsons out of it.
- Written REST service in Play Framework to ingest data in Neo4J and query elastic search.

• XStream

January 2018- February 2018

- Containerized and deployed the Exadatum in-house product XStream in a Docker with Hadoop, Java, Scala, Python, Spark, Zookeeper, Kafka, SQL, Druid, Superset, Hive, Maven and Redis installed.

Internships

Deep Learning Trainee- Radix Info Solutions, New Delhi, India

March 2018- July 2018

Data Science Internship in Healthcare- CSIR- Central Drug Research Institute under Dr. Sukant Khurana

March 2018- July 2018

Web Developer – Django- Learning While Travelling

March 2018- April 2018

Academic Projects

Query-Focused Multi-Document Summarization

Technologies: Java, Stanford CoreNLP, WordNet, Rida-WordNet, Lucene

- This project involves taking documents and query as input and generating summary related to the query as an output.
- NLP Tasks: Pre-processing both document and query by performing stemming and stop word removal using Stanford CoreNLP library. The documents were broken into sentences and query was expanded by generating synset using WordNet.
- Represented sentences and query as vectors using term frequency and calculated cosine similarity to select similar sentences.

SpeedAhead.com

Technologies: PHP, HTML, CSS, Bootstrap, JavaScript, MySQL, Wampp Server

- Designed SpeedAhead.com - a dynamic e-commerce website for selling cars. It is designed for 3 specific users: customers, admin and seller. Different functionalities according to different users are added.

8085 Microprocessor Simulator

- The project was developed in C language. This project involved simulating registers, main memory and instructions related to data transfer, arithmetic and logical operations, and branching.

Skills

Programming Languages: Python, Java, C, Scala

Dev-Ops: Linux, Docker, SaltStack, AWS- EC2 and VPC, Jenkins

Backend Web Development: Play Framework, Django

Big Data and Data Science: Hadoop, Spark, Pandas, Numpy, Nutch, Airflow

Databases: SQL, Neo4J, Hive

ML and AI: Tensorflow, Stanford CoreNLP, WordNet

Achievements

- Published a research paper on Plagiarism Detection in an International Journal.
- Neo4J Certified Graph Database Developer.
- GRE Score- 317