

# Saksham Harish Arora

Baltimore, MD | [saksham.arora501@gmail.com](mailto:saksham.arora501@gmail.com) | +1 4436835213 | [LinkedIn](#) | [GitHub](#)

## EDUCATION:

### University of Maryland Baltimore County, Baltimore, MD

Aug 2021-May 2023

Master of Science - Computer Science, GPA ~ 4.0

Graduate Assistant(GA) for Design & Algorithms class.

Courses - Machine Learning, Data Visualization, Principles of AI

### University of Pune, India

July 2013-May 2017

Bachelor of Engineering – Computer Engineering, GPA ~ 3.6

Courses- Data Structures, Operating Systems, Object Oriented Programming, Distributed Systems, Software Engineering, Data Mining

## SKILLS:

- |                                 |  |
|---------------------------------|--|
| • <b>Languages and Database</b> | Python, Java, Scala, Bash, MS SQL Server, PostgreSQL   |
| • <b>Big Data Ecosystem</b>     | Apache Spark, Hadoop, Hive, YARN, HDFS, Sqoop, Kafka, Cloudera Manager, Hue  |
| • <b>Cloud</b>                  | AWS - EMR, EC2, S3, IAM, Spectrum, Athena, RedShift, Lambda, Secrets   |
| • <b>Tools and Technologies</b> | VSCode, Tableau, Anaconda, MS Excel, Docker, REST, Maven, Spring Boot, git, Jira, Airflow, CI/CD, Jenkins, Postman, bitbucket, PuTTY, MobaXterm, HTML, OAuth |

## PROFESSIONAL EXPERIENCE:

### Data Engineer | Deloitte USI

May 2021-Aug 2021

- Collaborated with a cross-functional team following **Scrum Agile** methodology for development and enhancement of Financial and Payments data lake using **PySpark, AWS** and **Shell Scripting**.
- Designed and developed **ETL** framework to validate and transform data using PySpark and loaded the final data on **AWS S3**. Additionally, executed these data pipelines to automate data transformation phase on **AWS EMR** cluster.
- Handled data skewness on a daily basis by fine tuning the spark jobs based on the cluster size, executor cores, etc.
- Constructed **SQL queries** to migrate data from ingestion to transformation phase which helped in automation of datalakes.
- Analyze large amounts of data sets using pandas & spark sql dataframes to determine optimal ways to aggregate & report on it.

### Senior Systems Engineer | Infosys Ltd

Dec 2017-Apr 2021

- Orchestrated framework in Spark to do ETL and look-ups on user input data by defining data dictionaries and data lists.
- Maintained data pipeline up-time of 98% while ingesting Semi-Structured(**JSON and XML**) and Structured data from five different sources. Additionally, automated these workflows using **Airflow DAGs** to reduce physical workload by 26%.
- Implemented a workflow for historical and incremental data(~5 GBs) utilizing **hive** and **kudu** tables(using **impala engine**) for faster data processing in multiple layers. Enforced this workflow for different file formats like avro, parquet and flat files.
- Created **Unix** shell scripts to automate the data load processes to the target hive tables which reduced 64% of manual work.
- Performed fine tuning of **hive** tables using buckets and partitioning which resulted in efficient storage of data.
- Experienced in defining data quality, **data modeling**, **data warehouse**, Star/Snowflake Schema Designs with requirements analysis - definition, database design, testing and implementation, and Quality process.
- Worked on broadcast variables in Spark to resolve small file issues which improved efficiency of spark jobs by 12%.
- Assisted in development of new APIs for integration with e-commerce websites using **Java and Spring framework**.
- Good Knowledge on working with **OAuth2.0** to provide authentication and authorization to **REST APIs** services by exchanging Access tokens.

## PROJECTS:

### Employee Attrition: Predictive Analysis System | Python, Jupyter Notebook

[Link](#)

- Designed a predictive model using Python that assesses reasons for Attrition.
- The model predicted the attrition using classifiers such as Logistic Regression, Decision Tree, Random Forest and achieved a F1 score of 0.92.

### Amazon Clone | ReactJS, NodeJS, Express, Firebase

[Link](#)

- Developed a full functional clone of the Amazon website which included features like authentication, registration and payments. Integrated the payments system with Stripe API. Additionally, deployed the frontend and backend on firebase.

## AWARDS & CERTIFICATIONS:

- Awarded with 'Insta' Award twice for quick resolution of customer issues and deployment of new data lake.
- Microsoft Technology Associate by Microsoft - .Net Individual