HARSHDEEP HARIKESH MISHRA

New York, USA | harshdeepmishra82@gmail.com | +1-716-426-8682 | LinkedIn | Portfolio| GitHub

EDUCATION

University at Buffalo, The State University of New York **Expected: December 2024**

Master of Science in Computer Science and Engineering. ***(GPA: 3.7)*** *Buffalo, USA*

- Relevant coursework: Algorithms, Data Intensive Computing, Network Concepts, Database Systems

University of Mumbai - Thadomal Shahani Engineering College **June 2021**

Bachelor of Engineering in Computer Engineering. **(GPA: 9.15/10)** *Mumbai, India*

- Relevant coursework: Software Engineering, Big Data & Analytics, Distributed Computing, Data warehousing & Mining

TECHNICAL SKILLS

- **Programming languages**: C/C++, Java, Python, JavaScript, SQL.
- **Web**: React.js, Node.js, HTML, CSS, REST APIs
- **Data**: Spark, Beam, MySQL, Hadoop, Hive, Snowflake, Apache oozie, Kafka, NoSQL, Data Lake, Data Modeling.
- **Cloud platforms**: Google Cloud Platform (GCP) - Cloud functions, Composer, BigQuery, Dataproc, Pub/Sub, Storage, Compute Engine, Cloud SQL; Amazon Web Services (AWS) - EC2, S3, Athena, Redshift, AWS Glue, lambda, API Gateway.
- **Additional Skills**: Docker, Git, Linux/Unix, Excel, terraform, Shell Scripting (Bash), CI/CD.

PROFESSIONAL EXPERIENCE

Haver AI Inc, Boston, USA (Remote) **June 2024 – August 2024**

DATA ENGINEER INTERN

- Developed monthly data ingestion jobs using **AWS Lambda (Python)** and **AWS Eventbridge** to fetch claims data from BCDA API, leading to 100% reduction in manual effort. Enhanced system reliability with real-time notifications via **AWS SNS**.
- Utilized **AWS Glue** and **PySpark** for preprocessing and integrating **JSON** data from various sources into **AWS Redshift,** achieving a 50% increase in data processing speed.
- Designed and executed backend functionality for serving **REST API** requests using **AWS API Gateway** and **Lambda**.
- Developed a scalable **MLOps pipeline** that automates the execution of a 6-step machine learning workflow in **AWS SageMaker**, triggered by user interactions on the frontend. Utilized AWS SageMaker **AutoML** to generate predictions as part of the pipeline.

Skuad, Mumbai, India **December 2022 – May 2023**

DATA ENGINEER

- Implemented business KPIs for a telecommunications client leveraging **PySpark** on Google Cloud Platform (**GCP**) and Amazon Web Services (**AWS**) with Customer 360 data.
- Collaborated with product manager to understand client requirements and develop highly efficient data processing pipelines with **AWS Glue**, **S3**, **Athena**, and GCP's **Dataproc** ephemeral clusters and **BigQuery**, resulting in a 40% increase in data processing speed and enabling timely insights for decision-making.
- Provided comprehensive training to **5+** developers on a range of AWS services.

Quantiphi, Mumbai, India **November 2021 – November 2022**

DATA ENGINEER

- Spearheaded creation of a cloud solution on **GCP** with **Apache Airflow** and **Cloud Vision API** for parallel processing of 300,000 PDF documents. Achieved a **65% optimization** in processing time, reducing it from 28 to 10 days.
- Implemented a comprehensive logging strategy with **Cloud Logging** leveraging structured log formats, routing logs to **BigQuery** via **log sinks** for seamless integration. Exported 100 logs per minute, enabling detailed **analytics**.
- Led end-to-end migration of critical **workflow orchestration** projects from **Hadoop** ecosystem to **GCP Composer** (Airflow) and **Dataproc**, improving overall resource utilization by 40%.

DATA ENGINEER INTERN **July 2021 – November 2021**

- Utilized a Google Cloud **Virtual Machine (VM)** to execute **Python** scripts for testing real-time and batch endpoints on **Vertex AI**, Google Cloud's **AutoML** platform.
- Created a custom **Cloud Dataflow** template with **Apache Beam** for data transformations and employed **Airflow** to automate daily **ETL jobs**.
- Gained extensive training and hands-on experience with AWS, GCP, big data, **advanced SQL**, **Informatica**, **Tableau**, and **Snowflake**, enhancing overall technical proficiency and versatility.

PROJECTS

Tracking Bitcoin Transactions on Dark Web by a Crawler

- Engineered a system to assist government authorities in tracking and analyzing illicit cryptocurrency transactions on dark web by deploying a dark web **crawler** using **selenium** on **Tor** written in **Python**.
- Crawled onion websites using seed URLs such as "The Hidden Wiki" and extracted and stored 78 bitcoin addresses in a **MongoDB** database. Conducted data gathering for **14 days, visiting 72,000** websites on a dual-core CPU.
- Explored the addresses further on the **blockchain** using **WalletExplorer** and displayed them graphically with **PetriNets**.

PROFESSIONAL CERTIFICATIONS / ACHIEVEMENTS

- Google Cloud Platform Associate Cloud Engineer (link) (January 2022), offered by Google.
- SnowPro Core Certification (link) (March 2022), offered by Snowflake.
- Secured 1st place at the Smart India Hackathon (SIH) 2019 amongst **50+** participants held at NIT Trichy.