

Varun Bhaskara

📞 9632935145 | 📍 Boulder, CO | @varun.bhaskara@colorado.edu

EDUCATION

Master of Science, Computer Science

University of Colorado Boulder

May 2024 (Expected)

Boulder, CO

Bachelor of Engineering, Information Science and Engineering

Dayananda Sagar College Of Engineering, Viswesvaraya Technological University (CGPA: 9.1/10)

May 2019

Bangalore, India

EXPERIENCE

Hewlett Packard Enterprise

Data engineer

Bangalore, India

Feb '21 - Jul '22

- Developed a customizable standard data ingestion framework using Spark that allowed for easy setup of batch or streaming pipelines to ingest structured data from various sources and improved developer productivity by 40
- Implemented unit tests, data quality checks (Null check, length check, date format check, etc.), audit log capture, and job failure alerts using Oozie, Spark, and Scala that have contributed to a reduction of bugs in production pipelines by up to 60%.
- Setup Vertica DB as a hot repository and created and scheduled Scala scripts to migrate data from the data lake to and decreased OLAP query times by up to 50
- Designed internal API endpoints using NodeJs, Postgres, and TypeORM to serve data from the data lake for consumption by the front-end tools with latency of less than 30ms.
- Created PySpark scripts to migrate 50TB of data from the enterprise data warehouse into the data lake as delta tables.

Hewlett Packard Enterprise

Associate data engineer

Bangalore, India

Aug '19 - Jan '21

- Improved the performance of existing Spark data pipelines by up to 40% by leveraging concepts like column pruning and pre-partitioning and RDD transformations.
- Built a Spark-based batch ETL pipeline to create facts and SCD Type-2 dimensions for supply chain and logistics domains.
- Created custom Hive UDF's to calculate vital metrics such as Order status based on hold codes applied, the turnaround time for various stages in the order life-cycle, % of completed orders, etc., providing end-to-end insights to stakeholders from the order management team.

Hewlett Packard Enterprise

Business Intelligence developer intern

Bangalore, India

Jan '19 - Jul '19

- Developed optimized stored procedures to aggregate fact tables for Configure, Price, Quote team at various intervals, which resulted in a three-fold decrease in query times.
- Built a POC dashboard in QlikSense, to visualize key performance indicator metrics from HPE's Next Generation Quoting tool.

MAJOR PROJECTS

Text recognition in images using Machine Learning

Jan '19 - May '19

- Built a ML model consisting of a series of CNN, RNN, and connectionist temporal classification(CTC) layers to recognize handwritten text in images and achieved an accuracy of 87 percent and presented the paper at IEEE sponsored Ninth International Conference on 'Recent Engineering and Technology' held in Bangalore.

E-Commerce application to sell old text-books

Aug '18 - Dec '18

- Built a full-stack e-commerce website to sell old text-books online using MongoDB, Express, React and NodeJS and scaled it to receive over 100 orders.

SKILLS

- **Programming:** Python, Scala, SQL, JavaScript, Shell scripting, C, Java
- **Big data Technologies:** Apache Spark, Hadoop, Hive, Oozie, Sqoop
- **Cloud Technologies:** Azure Databricks, Azure Datafactory, Azure Event Hub, AWS RedShift, AWS EMR
- **Databases:** Postgres, Vertica, SQL Server, MongoDB, Hbase
- **Other Technologies:** MS Excel, Kafka, ReactJs, NodeJs, Redux, Express, Docker

LEADERSHIP AND HONORS

- Spotlight award for seamless delivery and going above and beyond in July 2021 at HPE for my work in migrating enterprise data warehouse data into the data lake.
- Performance excellence award in April 2020 at HPE for setting up data pipelines to create data models for shipment and logistics domains.