

BHARGAVI KOMMI

Sr Data Engineer

📞 203 822-7693 ✉ kommibhargavi7@gmail.com [in](#) [linkedin](#)

Professional Summary

- Seasoned **Data Engineer** with over **5 years of experience** delivering robust, scalable, and cloud-native data solutions across **Azure** and **AWS** ecosystems.
- Specialized in building real-time and batch data pipelines using tools like **Azure Data Factory**, **AWS Glue**, **Apache Spark**, **Kafka**, and **Databricks**.
- Proven success in implementing secure, cost-efficient, and high-performance architectures for clients in **e-commerce**, **finance**, and **retail banking**.
- Hands-on experience enabling AI/ML workflows by delivering clean, versioned, and feature-rich datasets for data science teams using **PySpark**, **dbt**, and **Delta Lake**.
- Skilled in collaborating with **cross-functional teams** including data scientists, product managers, compliance officers, and DevOps engineers to ensure business-aligned and compliant data delivery.
- Adept at implementing **data governance, security, and compliance standards** such as RBAC, SOX, and data lineage through tools like **Lake Formation**, **Azure Purview**, and **CloudWatch**.
- Experienced in building **metadata-driven frameworks** for reusable ingestion, transformation, and monitoring to support scalable analytics and reporting platforms.
- Strong analytical and communication skills with a focus on driving data engineering best practices, mentoring junior engineers, and supporting agile delivery cycles.

Education

University of New Haven

Masters in Data Science

Jan 2023 – De 2024

West Haven, CT, USA

Certifications

- Microsoft Certified: Azure Data Engineer Associate
- AWS Certified: Data Engineer Associate

Experience

JP Morgan & Chase Co.

Sr Data Engineer

Feb 2024 – Present

New York, USA

- Leading the build of a data platform on Azure for real-time risk and compliance analytics using Event Hubs, ADLS Gen2, and Azure Functions.
- Collaborating with quant teams and ML engineers to streamline model input data pipelines for market risk and stress testing simulations.
- Engineered end-to-end workflows in Azure Data Factory and orchestrated machine learning batch inference using Synapse and Databricks.
- Developed Delta Lake-based datasets with change data capture (CDC) to track transaction-level audit trails.
- Enabled business teams with self-service analytics through curated Power BI datasets, row-level security, and governance tagging.
- Optimized Spark clusters for heavy workloads, reducing job latency by 45% through caching, partition tuning, and adaptive execution.
- Automated metadata cataloging and schema validation workflows with Azure Purview and dbt.
- Integrated pipeline health dashboards with Azure Monitor and Application Insights to support 24x7 data availability SLAs.
- Worked across data governance, risk compliance, and legal teams to enforce lineage, consent, and data residency controls.

- Provided ongoing mentoring and code reviews, promoting modular and reusable code for ingestion and transformation layers.

(TCS) Citi Bank

May 2021 – Jan 2023

AWS Data Engineer

Hyderabad, India

- Designed and developed secure, scalable ETL pipelines in AWS Glue and PySpark to support regulatory reporting across multiple financial product lines.
- Partnered with data science teams to deliver engineered features for credit risk scoring and customer churn modeling.
- Processed batch and streaming data using S3, Kinesis, Lambda, and Step Functions to enable near real-time transaction monitoring.
- Optimized data pipeline performance and cost using partitioning, bucketing, and lifecycle policies on S3.
- Implemented a reusable metadata-driven ingestion framework to handle 100+ data sources with schema drift handling.
- Integrated Lake Formation and IAM policies to ensure granular access control and encryption compliance (SOX/PCI-DSS).
- Built visualizations using QuickSight to help compliance teams monitor high-risk transactions and anomalies.
- Worked closely with DevOps and security teams to integrate pipelines with centralized monitoring and secrets management via CloudWatch and AWS Secrets Manager.
- Coordinated with QA and business analysts to build test data strategies and validate pipeline outputs in sandbox environments.
- Enabled experimentation with ML fraud detection models by provisioning training data snapshots through version-controlled S3 buckets.

flipkart

May 2019 – April 2021

Data Engineer

Mumbai, India

- Built a unified data ingestion framework using Apache Spark and Hive on Hadoop to process 20TB+ of user activity data daily for Flipkart's product recommendation engine.
- Collaborated with data scientists to preprocess and transform training datasets for personalization models using PySpark.
- Integrated streaming data from Kafka and Flume to enable near real-time user behavior tracking and A/B testing analytics.
- Designed and implemented star schema models for the Product and Customer domains to support executive dashboards.
- Contributed to Flipkart's transition to a data lake architecture by organizing raw, processed, and curated zones on HDFS.
- Enabled fraud detection and inventory anomaly use cases through anomaly detection features built using Spark MLlib.
- Automated pipeline validation, data profiling, and anomaly detection workflows using Python and Airflow.
- Worked in cross-functional teams with product managers and analysts to ensure feature data requirements met business KPIs.
- Conducted proof-of-concept for migrating analytics workloads from Hive to Presto for faster reporting.
- Mentored junior data engineers on Spark best practices and code modularization for reusable ingestion logic.

Technical Skills

Languages: Python, SQL, Java, Scala, Shell Scripting, Bash, YAML, HTML/CSS, JavaScript

Cloud Platforms: AWS (Glue, S3, Lambda, Redshift, EMR, Step Functions, Kinesis, Athena, Lake Formation), Azure (Data Factory, Databricks, Synapse Analytics, ADLS Gen2, Event Hubs, Key Vault, Azure Functions), GCP (BigQuery, Dataflow, Pub/Sub)

Data Engineering Tools: Apache Spark, PySpark, Apache Kafka, Apache Hive, Apache Hudi, Apache Flink, Sqoop, Oozie, Airflow, NiFi, StreamSets, dbt, Talend, Informatica, SSIS

Data Warehousing & Databases: Snowflake, Redshift, BigQuery, Azure Synapse, Teradata, Oracle, PostgreSQL, MySQL, SQL Server, HBase, MongoDB, DynamoDB

DevOps & Automation: Jenkins, GitHub Actions, Azure DevOps, Docker, Kubernetes, Terraform, Ansible, Maven, Gradle, CI/CD Pipelines

Monitoring & Logging: CloudWatch, Prometheus, Grafana, ELK Stack (Elasticsearch, Logstash, Kibana), Splunk, Azure Monitor, Datadog

Data Science & Analytics: Pandas, NumPy, scikit-learn, MLlib, Power BI, Tableau, QuickSight, Jupyter, SageMaker, MLflow

Version Control & IDEs: Git, GitHub, Bitbucket, VS Code, JupyterLab, IntelliJ, Eclipse, PyCharm

Data Governance & Security: Azure Purview, AWS Lake Formation, Collibra, Alation, Ranger, RBAC, IAM, Data Encryption, GDPR/SOX Compliance