Bharat Kumar Reddy-Data Engineer
PH : (602)-562-7676 |
Gmail : bharatreddyvt@gmail.com
Loc: San Jose, CA
LinkedIn: https://www.linkedin.com/in/bharat-kumar-redd-m-03a2b5162/

Professional Summary :

- **Master's-level** Data Engineer with 5+ years of experience designing and implementing scalable data pipelines, data lakes, and analytics platforms across cloud environments (**AWS, Azure**). Proven expertise in **Databricks**, **PySpark**, **Snowflake,** and streaming technologies like **Kafka** for batch and real-time data processing.
- Hands-on experience building **Retrieval-Augmented Generation** (RAG) pipelines using **OpenAI** embeddings and **Azure** Cognitive Search for semantic search and contextual Q&A over unstructured data. Strong background in healthcare and manufacturing data environments with **HIPAA, FDA**, and CFR Part 11 compliance.
- Skilled in **CI/CD** automation (Jenkins, **Terraform,** Docker), data modeling (**DBT, Snowflake**), data validation, and BI dashboards (**Power BI, Looker**). Adept at cross-functional collaboration with product, engineering, and analytics teams to deliver reliable, secure, and high-impact data solutions.

Technical Skills:

Programming Languages	Python, SQL, C#, .Net PySpark, Linux, Java
Data Processing	Apache Spark, Kafka, Hadoop (HDFS, MapReduce, Hive), Spark SQL, EDA, Databricks.
AI & ML Skills	OpenAI API, Retrieval-Augmented Generation (RAG), Azure Cognitive Search, Embeddings, Vector Search, TensorFlow, PyTorch, Scikit-learn, NLP, LLMs, GenAI.
ETL Tools	Apache Airflow, Talend, Informatica, Azure Data Factory, DBT, Fivetran.
Cloud Platforms	AWS (S3, Glue, EMR, Kinesis, Athena), Azure (Data Factory, SQL Database, Data Lake), GCP (Big Query, Dataflow)
Data Warehousing	Amazon Redshift, SQL Server, Hive, Snowflake, OLTP, OLAP.
Database Management	MongoDB, PostgreSQL, Oracle, MySQL
Data Visualization	Power BI, Tableau, Looker, Excel.
CI/CD & DevOps	Jenkins, Docker, Kubernetes, Terraform, Azure DevOps
Data Governance	IAM Roles, Collibra, GDPR, HIPAA, CCPA compliance.
Certifications	AWS solution Architect, Python, SQL, Snowflake, Databricks certified Data engineer.
Other	TensorFlow, PyTorch, Apache Nifi, JIRA, GitHub, Git, Machine learning.

Professional Experience
KLA – Data Engineer
May 2024 – Present- Currently working
Responsibilities:

- Designed and developed data ingestion and processing pipelines using **Azure Data Factory**, **Databricks**, and **PySpark** to enable real-time analysis of wafer test data, driving insights into production yield and defect trends.
- Built scalable **Azure Data Lake** storage solutions to manage and analyze terabytes of equipment telemetry data, supporting predictive maintenance use cases for manufacturing tools.
- Integrated **Azure Synapse Analytics** to centralize data warehousing and enable unified reporting across engineering and quality assurance teams.
- Developed a Retrieval-Augmented Generation (RAG) pipeline using Azure Cognitive Search and OpenAI API with embeddings and vector search principles, enabling contextual Q&A over semiconductor defect documents.
- Integrated **LLM-powered assistants** within Databricks notebooks to support dynamic code generation, pipeline debugging, and metadata explanations, enhancing productivity across data engineering and analytics teams.
- **Applied advanced data structures** to optimize telemetry data processing pipelines in Databricks and PySpark, enabling faster defect classification and real-time yield anomaly detection across semiconductor manufacturing systems.
- Collaborated with cross-functional stakeholders (product, manufacturing, data science) to define data modeling standards that align with **semiconductor defect classification systems**.
- Provided **L2/L3** support for high-throughput data pipelines handling wafer inspection and metrology data, troubleshooting failures related to defect classification logic, fab tool telemetry ingestion, and **yield analytics** within semiconductor manufacturing workflows.
- Refactored legacy code into optimized **Databricks notebooks**, reducing pipeline runtime by 40% and improving maintainability across yield analytics use cases.
- Created and deployed **APIs** to serve processed manufacturing KPIs, enabling seamless integration with engineering dashboards and internal tools.
- Engineered high-throughput data pipelines to process wafer inspection and metrology datasets, enabling yield analytics across 300mm semiconductor fabs.

- Enforced data governance practices using **Azure Purview** and RBAC to ensure secure and compliant access to high-sensitivity production data.
- Leveraged OpenAI APIs with Azure-hosted embeddings to develop contextual document search tools for semiconductor defect classification and knowledge retrieval.
- Conducted root cause analysis of data anomalies using advanced SQL and Spark logic, contributing to early detection of process issues in fab environments.

Fagron Sterile Services (Hired by Wichita state university , I was directly hired by WSU, for Fagron to work as a BI Developer and Post- Production Specalist and later prompoted to data engineer role)

Data Engineer   **Wichita, KS(Dec 2022 – May 2024)**

Fagron (FSS) is a global group of innovative **pharmaceutical companies** focused on creating the **future of personalized medicine.** Worked closely with both IT teams and business units.

Responsibilities:
- Engineered ETL workflows using **AWS Glue**, **S3**, and **Redshift** to manage and analyze controlled drug formulation data, supporting FDA and USP <797>/<800> compliance audits.
- Developed automated **Snowpipe** ingestion into Snowflake for real-time integration of ERP and LIMS system data, ensuring traceability of sterile product batches.
- Partnered with QA and regulatory teams to establish **data validation rules and lineage tracking** using **Apache Airflow** and **Snowflake tasks**, improving inspection readiness.
- Used **AWS Lambda** to automate notifications and data quality checks for critical production events, ensuring proactive monitoring of batch inconsistencies.
- Modelled production data across **bronze-silver-gold layers** using **DBT and Databricks**, enhancing operational visibility into order fulfillment, sterility test cycles, and environmental monitoring.
- Delivered interactive dashboards via Power BI and QuickSight to monitor sterile drug production, batch fulfillment, and environmental metrics, supporting FDA inspections and operational efficiency.
- Ensured compliance through **Unity Catalog** and fine-grained access controls in cloud environments, aligning with HIPAA and CFR Part 11 standards.
- Partnered with pharmacists and regulatory stakeholders to define metadata standards for cleanroom operations, aiding future ML-driven contamination risk models.
- Collaborated with pharmaceutical scientists and operations teams to expose clean datasets for machine learning initiatives related to contamination risk prediction.

Dentsu, Chennai (May 2020 – May 2022)

Data Engineer *(Databricks, Azure cloud, Spark, Kafka, PowerBI)*
- Designed and implemented advanced **ETL** pipelines using Azure Data Factory (ADF) to orchestrate workflows across diverse data sources, integrating on-premises databases such as Postgres and SQL Server with cloud platforms like Azure Data Lake Storage (**ADLS**) and Databricks.
- Automated data ingestion from FTP servers, **REST AP**Is, and SaaS applications via ADF, optimizing pipelines with PolyBase and data partitioning to enhance loading efficiency and data availability.
- Streamlined Databricks workflows using Spark to process millions of customer transactions, achieving an 18% reduction in processing time, and implemented SCD Type 2 frameworks for robust data quality management using PySpark and Spark SQL.
- Developed advanced SQL queries, stored procedures, views, and Python scripts to support efficient data extraction and analysis, while integrating Kafka with Spark Streaming for real-time data processing and automated quality checks.
- Defined and tracked key business metrics in Power BI, utilizing DAX time intelligence functions and advanced visualizations to deliver actionable insights into sales KPIs, resulting in a 5% improvement in customer retention.

Education Details:

Masters in Computer Science

University: <mark>Wichita State University</mark>, Wichita (3.5 Gpa out of 4.0)

PERSONAL QUALITIES:
- Strategic Outreach & Client Engagement
- Time Management & Prioritization
- Adaptability & Team Collaboration
- Customer Satisfaction Focus