BHARATHWAZ KALAMSETTI

Tampa, FL | +1 716-939-6955 | bharathwazkalamsetti@gmail.com | LinkedIn

---

PROFESSIONAL SUMMARY

- Experienced Data Engineer with 4+ years building and optimizing scalable data pipelines, lakehouses, and cloud-based analytics platforms across public sector, logistics, and insurance domains.
- Skilled in developing end-to-end ETL/ELT workflows using Apache Spark, AWS Glue, Airflow, and Azure Data Factory, supporting both batch and real-time processing at scale.
- Adept at integrating AI and automation into data ecosystems using Python, OpenAI APIs, and Flask, streamlining documentation, reducing manual overhead, and enabling intelligent workflows.
- Proven expertise in designing actionable dashboards with Power BI, Synapse, and DAX to drive KPI monitoring, executive reporting, and operations intelligence for 30+ business units.
- Strong background in data governance, schema enforcement, CI/CD automation, and compliance (HIPAA, CMS, SOX), with hands-on experience using Git, Jenkins, Snowflake, and Delta Lake.

TECHNICAL SKILLS

Programming & Scripting:
 Python, Java, Scala, SQL, Bash, SparkSQL, DAX

Data Engineering & Orchestration: Apache Spark, Databricks, Apache Airflow, AWS Glue, Azure Data Factory, Apache NiFi, Informatica, SSIS

Cloud & Data Platforms: AWS (S3, Redshift, Lambda, Glue, EMR), Azure (Synapse, Blob Storage, App Services), Snowflake, Delta Lake

Big Data & Storage Formats: Hadoop, Hive, HDFS, Kafka, Parquet, Avro, ORC

Machine Learning & AI Integration: Pandas, NumPy, scikit-learn, OpenAI API, Hugging Face Transformers, Flask

Business Intelligence & Reporting: Power BI, Tableau, Amazon QuickSight, Excel (PivotTables, Macros, VLOOKUP)

DevOps & CI/CD: Git, Docker, Jenkins, Azure DevOps, GitHub Actions

Data Governance & Quality: Data Lineage, Metadata Management, UAT, Schema Validation, Data Profiling, Version Control

---

PROFESSIONAL EXPERIENCE

Senior Data Engineer   Jan 2025 - Present
Arc Erie County, NY

- Architected AWS Glue and Redshift pipelines to automate ingestion of structured and semi-structured data from 20+ municipal departments, cutting reporting delays from two days to under 12 hours.
- Consolidated fragmented workforce and budget data into a unified Delta Lake architecture using Spark, enabling analytics teams to generate insights 40% faster through pre-aggregated views.
- Introduced an AI-based summarization utility built with OpenAI's API and Flask to automate weekly briefings and compliance documents, reducing manual reporting time by over half.
- Translated complex operational metrics into intuitive Power BI dashboards with role-based access and DAX-powered KPIs, helping department heads track project status and resource allocation in real time.
- Applied PySpark and SQL for schema validation, anomaly detection, and data profiling across 300+ daily data files, lifting ingestion accuracy to over 99.7% and meeting ISO data quality thresholds.
- Converted outdated SSIS jobs into reusable Airflow DAGs with checkpointing and alerting mechanisms, improving ETL reliability while decreasing support tickets by 30% month over month.
- Developed a metadata framework by integrating SharePoint lists with Python-driven documentation automation, reducing time-to-data for analysts by 50% and improving asset discoverability.
- Facilitated agile delivery by managing Jira boards, GitHub repositories, and sprint reviews, aligning engineering output with evolving federal data compliance and audit requirements.

Data Engineer   Jan 2024 - Dec 2024
HCL Technologies - Remote

- Transitioned 40+ SSIS pipelines to AWS Glue with Spark-based transformations, enhancing throughput and cutting monthly compute costs by 30%.
- Connected IoT telemetry from warehouse robots and delivery trucks to Redshift, enabling predictive maintenance that reduced equipment downtime.
- Built Airflow DAGs with SLA monitors and custom alerts, streamlining delivery data ingestion and improving pipeline reliability by 25%.
- Introduced a Synapse-driven reporting layer that consolidated supply chain metrics, reducing ad-hoc query dependencies for analytics teams.
- Delivered 20+ Power BI dashboards featuring order delays, stock movement, and carrier efficiency, directly supporting weekly ops reviews.

- Enforced row-level security across Power BI models to protect vendor-level insights while enabling role-specific access for 200+ stakeholders.
- Prepared feature-ready data using Spark and SQL for machine learning teams, contributing to a 12% gain in delivery time prediction accuracy.
- Defined QA protocols and executed test plans during the warehouse data lake migration, ensuring complete accuracy during go-live transitions.

Data Engineer   **Aug 2020 - Dec 2023**
Softtek, Bangalore, India
- Designed Spark-based data pipelines to process over 6TB of raw claims and policy data daily, which helped reduce ETL runtime by 40% and supported near real-time claims adjudication.
- Introduced schema validation and Delta Lake versioning to protect ingestion workflows from upstream file structure changes, decreasing manual rework and improving data reliability across reporting layers.
- Created PySpark-driven validation scripts that automatically flagged mismatched CPT codes, duplicate claim entries, and field-level null surges, cutting recurring quality issues nearly in half.
- Consolidated policyholder and provider data into Snowflake using S3 and Redshift staging zones, improving actuarial model readiness and enabling faster CMS regulatory submissions.
- Wrote SQL-based logic to detect schema drift, volume spikes, and missing partition data during daily loads, which reduced SLA violations for high-priority claims dashboards by 28%.
- Developed audit trail logic in Python for each transformation stage, embedding record-level metadata for traceability and meeting SOX and CMS compliance audit standards.
- Built and maintained Power BI dashboards for claims lifecycle tracking, claim denial patterns, and financial variance reporting, supporting faster decision-making by billing and reimbursement teams.
- Led deployment and rollback workflows via Git and Jenkins for all production pipeline releases, minimizing post-release incidents and maintaining full visibility into version history.

PROJECTS

AI-Powered Document Summarization & Workflow Automation
Tools: Python, OpenAI API, Flask, SharePoint, Azure Blob, Power BI
- Developed a Flask-based microservice deployed on Azure App Services to automate text extraction and summarization from operational reports, using the OpenAI API to condense lengthy documents into 3–5 sentence briefings.
- Integrated SharePoint for document ingestion and Azure Blob for file storage, with Python scripts handling version control, metadata tagging, and access logging to ensure security and regulatory traceability across departments.
- Enabled real-time monitoring of summarization usage by implementing log parsers and Power BI dashboards, allowing leadership to track adoption trends across 30+ teams and reduce manual reporting time by 70%.

Enterprise Data Intelligence Portal with Power BI
Tools: Power BI, SQL Server, Azure Synapse, DAX, Excel
- Consolidated operational data from logistics, HR, finance, and project delivery into Azure Synapse, then modeled and exposed curated datasets in Power BI for 30+ department-specific dashboards.
- Applied advanced DAX calculations to generate YOY growth metrics, variance indicators, conditional RAG statuses, and predictive forecasting visuals, helping leadership quickly interpret financial and operational health.
- Automated daily refresh cycles using Power BI Gateway, configured role-based access with row-level security, and designed interactive drill-down views and slicers that improved decision-making speed by 60%.

EDUCATION
Master's in data science   **Aug 2023 - Dec 2024**
University at Buffalo, SUNY

B.Tech in Computer Science   **Sep 2017 - May 2021**
Chennai Institute of Technology, India

CERTIFICATIONS
- Google Cloud Data Engineering Professional Certificate - **Coursera**
- IBM Data Engineering Professional Certificate - **Coursera**
- Databricks Lakehouse Fundamentals - **Databricks Academy**
- Data Engineering with Python - **DataCamp**
- Power BI Data Analyst Professional Certificate - **Coursera**