Jayanth Balina

**Dallas, TX |** [**jayanthbalina135@gmail.com**](mailto:jayanthbalina135@gmail.com) **| (940) 218-7753 |** [**LinkedIn**](https://www.linkedin.com/in/jayanthbalina/)

# PROFESSIONAL SUMMARY

**Results-driven Azure and Databricks Cloud Data Engineer with 7+ years of experience** in designing and implementing end-to-end data solutions using Azure, Databricks, Python, PySpark, and SQL. Skilled in large-scale data migrations, ETL pipeline development, Unix shell scripting, and performance optimization, with a strong foundation in software engineering principles, data structures, and multi-threading. Adept at working in Agile environments and collaborating with cross-functional teams to drive digital transformation in a global delivery model. Industry expertise includes Banking and Financial Services, with hands-on experience in Snowflake and big data frameworks. Strong analytical and problem-solving skills, with a passion for delivering scalable, high-quality solutions while staying updated with emerging technologies.

# TECHNICAL SKILLS

**Programming Languages**: Python, Java, Scala, SQL, PL/SQL, T-SQL, NoSQL, Unix Shell Scripting

**Big Data Technologies:** Hadoop, Spark, Hive, Pig, Sqoop, MapReduce, HBase, EMR

**Methodologies:** SDLC, Agile, Waterfall

**DevOps Tools**: Git, Jenkins, Docker, Kubernetes

**ETL Tools:** SSIS, Informatica PowerCenter, Erwin, Talend, Data Stage

**Databases:** SQL Server, PostgreSQL, MySQL, Oracle, Snowflake, DynamoDB, MongoDB, CosmosDB

**Pipelines:** Apache Airflow, AWS Step Function, Luigi, Prefect, Oozie

**Streaming Technologies:** Amazon Kinesis, Apache Spark, Apache Kafka, Apache Hive

**Libraries:** Pandas, NumPy, Matplotlib, SciPy, ScarPy, TensorFlow, PyTorch, Scikit-learn, NLTK, Plotly, Keras, PyMc3

**Data Visualization:** Microsoft Excel, Power BI, Tableau, IBM Cognos, QlikView, QuickSight, Seaborn, SSRS

**Cloud Platforms:** AWS (EC2, S3, Redshift, Glue), Azure (Azure Data Factory, Azure Databricks), GCP

**Data Warehousing:** Amazon (Redshift, DynamoDB, RDS, Athena), Azure (Synapse, Blob, Data Lake), BigQuery, Teradata, Snowflake

**Additional Skills:** Design patterns, Collections, Multi-threading, Memory management, Concurrency.

# PROFESSIONAL EXPERIENCE

**Charter Communications** **| TX| *Senior Data Engineer* January 2024 – Present**

* Designed and deployed end-to-end scalable data pipelines leveraging Azure Data Factory, Databricks, and Data Lake Storage, reducing data processing time by 45%.
* Engineered a real-time data ingestion framework using Kafka and Spark Structured Streaming, ensuring seamless integration and availability of live data streams across business domains.
* Optimized data models and ETL pipelines for structured and unstructured data, improving query performance and reducing latency by 35%.
* Collaborated with cross-functional teams to develop and implement data governance strategies, ensuring compliance with industry standards and data quality metrics, which improved data accuracy by 50%.
* Integrated Azure Synapse Analytics and Power BI to deliver actionable insights through dashboards, enabling executive decision-making and reducing reporting time by 40%.
* Conducted advanced performance tuning for big data systems, enhancing cluster utilization and reducing costs by 25% across enterprise data solutions.
* Migrated legacy on-premises data systems to the cloud, leveraging Azure Databricks and Snowflake, which reduced maintenance overhead and increased scalability by 60%.
* Supported and mentored junior engineers in implementing best practices for data pipeline design, performance optimization, and cloud-native solutions, fostering a high-performing team culture.
* Implemented CI/CD pipelines for data workflows using Apache Airflow and Azure DevOps, reducing deployment cycles by 30%.

**IQVIA | India | *Data Engineer* December 2021 – December 2022**

* Designed and deployed end-to-end scalable data workflows using Azure Data Factory, reducing data processing and transformation time by 40%.
* Engineered robust data lake architectures on Azure Data Lake Storage Gen2, enabling faster data retrieval and improved organization, leading to a 50% increase in operational efficiency.
* Utilized Azure Databricks for advanced data exploration, manipulation, and machine learning model building, improving data accuracy and business insights by 30%.
* Developed real-time data ingestion pipelines with Spark Structured Streaming and Kafka, enhancing the availability and processing of live data streams by 35%.
* Created interactive dashboards and reports using Power BI and Azure Synapse Analytics, empowering stakeholders with
* actionable insights and reducing reporting time by 45%.
* Established security best practices and compliance protocols for data workflows, ensuring confidentiality and integrity, and reducing data breaches and incidents by 20%.

**UST | India | *Data Engineer* November 2019 – December 2021**

* Designed and deployed scalable data lake infrastructure using AWS big data technologies, including Redshift, S3, Glue, and EMR, streamlining data integration and improving report efficiency by 50%.
* Engineered batch and real-time streaming architectures to enable seamless data ingestion and processing, reducing data latency by 40% and ensuring real-time availability for critical business operations.
* Developed and optimized ETL pipelines for structured and unstructured data, enhancing data transformation and processing efficiency by 60%, meeting both ad-hoc and pre-scheduled reporting requirements.
* Conducted advanced performance tuning for data queries and big data applications, improving system reliability and achieving an 80% boost in efficiency, aligning with SLAs and business objectives.
* Collaborated with business stakeholders, data analysts, and engineering teams to gather requirements and deliver robust data solutions, supporting strategic decision-making across multiple domains.
* Implemented best practices for data governance and security, ensuring compliance with industry standards and safeguarding sensitive business data.
* Automated data workflows using AWS Lambda and Step Functions, reducing manual intervention and accelerating deployment cycles for data pipelines.

**Unitel Software | India | *Junior Data Engineer* May 2017 – November 2019**

* Assisted in the design and development of data pipelines using Hadoop and Hive, enabling efficient batch processing for large datasets.
* Contributed to setting up and maintaining scalable data lakes, improving data accessibility and reducing retrieval time by 30%.
* Implemented data ingestion workflows using Apache Kafka, facilitating real-time data streaming for critical business applications.
* Supported ETL pipeline development to transform raw data into structured formats, enhancing reporting capabilities across teams.
* Conducted data validation and quality checks to ensure accuracy and consistency, reducing errors in downstream analytics by 25%.
* Gained hands-on experience with AWS services such as S3 and Redshift, optimizing data storage and query performance.
* Collaborated with senior engineers to troubleshoot and resolve performance issues in distributed data systems, boosting overall system reliability.

# EDUCATION

**Master’s in Computer Science | Jan 2023 – Dec 2024**

University of North Texas, Denton, TX, USA

**Bachelor’s in Computer Science | Aug 2015 – Apr 2019**

Jawaharlal Nehru Technological University, India

**CERTIFICATIONS**

Microsoft Certified: Azure Data Engineer Associate.

AWS Certified: Developer – Associate.

Snowflake: SnowPro Core Certification.

Google Cloud Certified: Professional Data Engineer.