Jayanth Balina

Data Engineer

+19402187753 ◇ [jayanthbalina135@gmail.com](mailto:jayanthbalina135@gmail.com) ◇ Denton, Texas, United States ◇ Open to Relocate ◇ [LinkedIn](http://www.linkedin.com/in/jayanthbalina)

# SUMMARY

Experienced Data Engineer and Analyst with 7 years of expertise in designing robust data workflows and infrastructure on Azure and AWS. Proficient in big data technologies, real-time data processing, and data visualization. Skilled in building and optimizing ETL pipelines to improve data efficiency, quality, and accessibility. Adept at leveraging tools like Apache Airflow, Databricks, and SQL-based technologies to create scalable data solutions. Experienced in creating interactive dashboards and reports using Power BI and Tableau to support data-driven decision-making. Proven track record of collaborating with cross-functional teams to deliver data-driven insights for business growth.

**EDUCATION**

**Master’s in computer science, University of North Texas –** Denton, USA (GPA: 3.8) **Jan 2023 – Dec 2024**

Focused on specialized coursework in data engineering and optimization techniques to enhance skills directly applicable to the field of

data engineering.

**Bachelor's in Electronics and Communication Engineering,** JNTU Kakinada, India (GPA: 3.5) **Aug 2015 — Aug 2019**

The capstone project focused on implementing data-driven solutions within an engineering context, demonstrating the practical

application of technical expertise relevant to Data Engineering.

**EXPERIENCE:**

**Senior Data Engineer Jan 2023 – Present**

**Capital One – Plano, Texas**

## 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%.

## Data Engineer/Data Analyst Dec '21 – Dec '22

## IQVIA - Bengaluru, India

## 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%.

## Collaborated with cross-functional teams to deliver data-driven solutions aligned with business goals, improving overall productivity and decision-making.

## Associate Data Engineer/Data Analyst Nov '19 – Dec '21

## UST - Hyderabad, India

## Designed and deployed scalable data lake infrastructure using AWS big data technologies, including Redshift, S3, Glue, and EMR, streamlining data integration and improving reporting 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.

## Data Engineer Intern – Big Data May '18 – Nov '19

## Unitel Software - Hyderabad, India

## 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.

## Documented best practices for big data tools and workflows, aiding in knowledge sharing and team onboarding processes.

**SKILLS**

* + **Programming Languages**: Python, Java, SQL
  + **Databases**: MySQL, Oracle, PostgreSQL, Microsoft SQL Server, MongoDB, Cassandra, Azure Cosmos DB
  + **Big Data Technologies**: Hadoop, Hive, Cloudera, Kafka, Apache Airflow, dbt, Snowflake, Polars, DuckDB, ETL, Apache Spark, Scala, Rust
  + **Cloud Platforms**: GCP (BigQuery, Dataflow, Dataproc, Pub/Sub, Cloud Storage), AWS (S3, Glue, Redshift, EMR, Kinesis), Azure (Databricks, Data Lake Storage)
  + **Data Visualization and BI Tools**: Tableau, Power BI
  + **Additional Skills**: Data Modeling, Advanced SQL Optimization, Orchestration Tools (Apache NiFi, Prefect)

## CERTIFICATIONS

## Microsoft Certified Azure Data Engineer Associate