

# Jayanth Balina

Data Engineer

+19402187753 ◊ [jayanthbalina135@gmail.com](mailto:jayanthbalina135@gmail.com) ◊ Denton, Texas, United States ◊ Open to Relocate ◊ [LinkedIn](#)

## SUMMARY

Experienced Data Engineer and Analyst with 5 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 Computer Science**, JNTU Kakinada, India (GPA: 3.8)

**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:

**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

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, Go, Julia, Bash, TypeScript

**Databases:** MySQL, Oracle, PostgreSQL, Microsoft SQL Server, MongoDB, Cassandra, Azure Cosmos DB, Redis, Neo4j, Amazon DynamoDB, ClickHouse, ElasticSearch, TimescaleDB, FoundationDB

**Big Data Technologies:** Hadoop, Hive, Cloudera, Kafka, Apache Airflow, dbt, Snowflake, Polars, DuckDB, ETL, Apache Spark, Scala, Rust, Delta Lake, Apache Iceberg, Apache Flink, Apache Beam, Presto, Trino, OpenMetadata, Dremio, Datapalego

**Cloud Platforms:** GCP, BigQuery, Dataflow, Dataproc, Pub/Sub, Cloud Storage, AWS, S3, Glue, Redshift, EMR, Kinesis, Azure, Databricks, Data Lake Storage, Databricks Lakehouse, Athena, Synapse Analytics, Terraform, Kubernetes

**Data Visualization and BI Tools:** Tableau, Power BI, Looker, Apache Superset, Metabase, Plotly, Dash, Mode Analytics

**Data Engineering Automation:** Dagster, Great Expectations, DataOps, Monte Carlo, Datafold, Soda

**Data Integration and Transformation:** Apache NiFi, Prefect, Talend, Matillion, Fivetran, Hevo Data, Stitch, Alteryx

**Streaming and Real-Time Processing:** Kafka, Apache Flink, Pulsar, Materialize

**Machine Learning and Data Science:** MLflow, Feature Store, TensorFlow Extended, H2O.ai, KubeFlow, PyCaret

**Additional Trending Skills:** SSIS, SSMS, SSAS, SAAS, Data modeling, SQL optimization, reverse ETL, GraphQL, CI, CD, dbt metrics layer, event-driven architectures, lakehouse

## CERTIFICATIONS

Microsoft Certified Azure Data Engineer Associate

AWS Certified Developer – Associate Snowflake

SnowPro Core Certification