

# Akhilesh Santosh Jadhav

Santa Clara, CA | akhileshjadhav@nextitjobs.com | +1 (408)780-1597 | [Portfolio](#) | [LinkedIn](#) | [GitHub](#)

## PROFESSIONAL SUMMARY

Certified Data Analytics Engineer with 3.5+ years of experience in ETL pipelines, real-time data streaming, and cloud-based architectures using Python, SQL, Spark, Kafka, and AWS. Skilled in big data processing, database optimization, and machine learning integration to enhance decision-making and operational efficiency. Proficient in data analysis, statistical modeling, and data visualization to extract actionable insights for strategic business growth. Experienced in migrating data warehouses, ensuring data governance, and reducing data latency across finance, IoT, and e-Commerce domains.

## WORK EXPERIENCE

### Citi

San Francisco, CA

Data Engineer

Jun 2024 – Present

- Architected scalable ETL pipelines using Apache Spark, Python, and Airflow, optimizing financial transaction processing from SQL, NoSQL, and APIs into AWS S3 and Snowflake data lake, accelerated data processing by 40%.
- Pioneered real-time fraud detection system using Apache Kafka, Flink, and Python, synchronized with Citi's risk engine and machine learning models TensorFlow and Scikit-learn, transformed response time to identify anomalies from hours to seconds.
- Spearheaded migration of legacy data warehouses from Oracle to Google BigQuery, achieved 60% faster query performance through SQL optimization, partitioning, and indexing strategies for risk analytics.
- Established data governance framework using Apache Atlas and AWS Glue, strengthened data lineage tracking and adhering to GDPR compliance across enterprise platform.

### PTC

Boston, MA

Data Engineer

Jul 2023 – May 2024

- Engineered ETL pipelines using AWS Glue, Apache Spark, and Python to process industrial IoT data from ThingWorx, enabling predictive maintenance analytics.
- Devised data lake on AWS S3 and Snowflake, integrated structured/unstructured data from PLM (Windchill), CAD systems, and IoT sensors, improved accessibility by 40%.
- Optimized real-time data streaming using Apache Kafka and AWS Kinesis, decreased ingestion latency by 30%, enabling faster decision-making for industrial automation clients.
- Restructured SQL/NoSQL databases (PostgreSQL, MongoDB, DynamoDB) performance for storing and retrieving IoT sensor data, achieved 35% faster queries through indexing and partitioning.
- Collaborated with cross-functional teams to deploy machine learning models using MLflow, AWS SageMaker, and dbt, minimized equipment downtime by 25% through anomaly detection in IoT data.

### Sigma Infosolutions Ltd.

Bangalore, IN

Data Analyst

Jun 2020 – Jul 2022

- Developed advanced statistical analysis using Python Scikit-Learn, R, and Tableau, identifying key customer trends and optimizing product recommendations, increased e-Commerce conversion rates by 15%.
- Designed interactive dashboards in Power BI and Tableau, incorporating key performance indicators (KPIs) to transform complex financial data into actionable insights for digital lending and improve stakeholder decision-making.
- Automated data validation using Python and SQL while ensuring data integrity, streamlined manual data cleaning efforts by 50% across multiple reporting platforms.
- Conducted A/B testing for e-Commerce analytics project using Google Analytics, Tableau, and Excel VBA, to measure the impact of UI/UX changes on customer retention increased user engagement by 18%.
- Formulated optimized data pipelines using SQL, Python Pandas, NumPy, and Apache Spark, streamlining data collection and processing for e-Commerce and FinTech solutions, accelerated data retrieval by 30%.
- Created ETL workflows using SSIS, Airflow, and Snowflake, unified data from AWS S3, Google BigQuery, and MySQL to create centralized data warehouse, improved efficiency by 40%.

## CERTIFICATIONS

- AWS Certified Data Engineer [\[Credentials\]](#)
- Databricks Fundamentals - Academy Accreditation [\[Credentials\]](#)
- Microsoft Technology Associate - Python [\[Credentials\]](#)

## TECHNICAL SKILLS

**Programming Languages:** Python, Java, Scala, SQL, R, Bash

**Big Data & Distributed Systems:** Hadoop(HDFS, Hive, MapReduce, Yarn, HBase), Apache Spark, Kafka, Nifi, Airflow, Flink, Teradata

**Cloud Platforms:** AWS (EC2, S3, EMR, Redshift, Glue, Lambda, Cloudwatch), Azure (Databricks, Synapse), GCP (BigQuery, Dataflow, Pub/Sub)

**Databases:** MySQL, SQL Server, Snowflake, PostgreSQL, MongoDB, Cassandra, DynamoDB

**Reporting Platforms:** SSRS, Tableau, Microsoft Power BI, Sigma Computing, Metabase, Looker

**Devops & Version Control:** Git, GitHub, Bitbucket, Jenkins, Docker, Kubernetes, Terraform, Alteryx

**Data Analytics & Architecture:** Data Visualization, Cleaning, Masking, Manipulation, Data Mesh, Schema Design, Data Lakes

**ETL/ELT Tools & File Formats:** SSIS, SSAS, Informatica, Talend, Matillion, DBT, Parquet, Avro, ORC, JSON

**Methodologies:** Agile/Scrum, Waterfall, SDLC

## EDUCATION

### Syracuse University

Syracuse, NY

Master of Science in Computer Science

Aug 2022 – May 2024

### Vishwakarma Institute of Technology

Pune, IN

Bachelor of Technology in Computer Engineering

Jul 2018 – Jun 2022