

# Mankirat Singh

Plano, TX | 945-217-8662 | [mankirat.tech@gmail.com](mailto:mankirat.tech@gmail.com) | [LinkedIn](#)

## SUMMARY

Data Engineer with strong expertise in SQL, distributed systems, and AWS-based pipelines. I design and optimize scalable data platforms - delivering high-performance ETL, clean modeling layers, and production-grade analytics systems at companies like JPMorgan Chase and DXC.

## SKILLS

**Methodology:** SDLC, Agile, Waterfall

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

**IDEs:** PyCharm, Jupyter Notebook, IntelliJ IDEA, Visual Studio, NetBeans

**Big Data Ecosystem:** Hadoop, MapReduce, Hive, Apache Spark, Pig, HDFS, Yarn, Data Lake, Data Warehouse

**Visualization and Reporting Tools:** Power BI, Tableau, SAS, Argos, Looker, Amazon QuickSight, Google Data Studio

**ETL Tools:** SSIS, Apache Kafka, Airflow, Alteryx

**Cloud Technologies and Infrastructure:** AWS (S3, Glue, IAM, Lambda, EC2, Redshift, CloudFormation, Event Bridge, Athena, RDS, CloudWatch, EMR, SageMaker, DynamoDB), GCP, Azure, Snowflake

**Packages:** NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow

**Development and Operations Tools:** Git, GitLab, Jenkins (CI/CD), Docker, Kubernetes, Grafana, Splunk, Jira, GitHub

**Databases:** MySQL, PostgreSQL, MongoDB, Neo4j, RDBMS, NoSQL, Cassandra, HBase, Elasticsearch

**Operating Systems:** Windows, Linux

## EDUCATION

**Master of Science in Business Analytics**

The University of Texas at Dallas

**May 2024**

## EXPERIENCE

### DXC Technology, TX | Jan 2024 – Current | Data Engineer

- Designed and implemented ETL pipelines using Python, SQL, and Apache Airflow to process high-volume enterprise data into cloud-based data warehouses leveraging GitLab for version control.
- Migrated legacy on-premise data systems to AWS S3 and Redshift, implementing IAM and CloudFormation for secure, automated infrastructure provisioning.
- Optimized data lake architecture on AWS Glue and Athena, improving query performance by 30% and reducing storage cost through partitioning and compression strategies.
- Integrated streaming data ingestion using Apache Kafka, Lambda for near real-time analytics dashboards in Power BI, Tableau.
- Developed data quality frameworks with PySpark and Pandas, automating validation checks and anomaly detection across diverse data sources.
- Partnered with cross-functional teams to enable business intelligence and advanced analytics, ensuring scalable and reliable data models that support predictive insights.

### JPMorgan Chase & Co, TX | June 2023 – August 2023 | Software Engineer intern

- Optimized high-volume SQL Server/PostgreSQL queries supporting Chase lending workflows, reducing execution time through index tuning, plan analysis, and temp-table refactoring.
- Updated data models and table structures in PostgreSQL/Aurora Postgres, improving downstream join performance and simplifying schema changes for reporting teams.
- Enhanced ETL jobs built on Python + SQL + AWS Glue/RDS to load operational data into Aurora clusters, cutting refresh cycles from hourly to ~30 minutes.
- Performed source-to-target validations and resolved ingestion bottlenecks by monitoring CloudWatch logs, RDS performance insights, and Glue job metrics, preventing incorrect records from entering decisioning systems.

### Infinite Infolab, India | Jan 2020 - July 2022 | Data Engineer

- Built scalable big data pipelines using Hadoop, Hive, Pig, and Spark to handle multi-terabyte data processing and analytics.
- Designed and managed relational and NoSQL databases (MySQL, Cassandra, MongoDB) for high-availability and optimized query performance.
- Automated data ingestion and transformation workflows using Apache Airflow and SSIS, enabling faster data delivery to downstream applications.
- Created interactive dashboards and advanced reports in Tableau, Power BI, and Looker, driving actionable insights for clients across multiple industries.
- Deployed machine learning feature pipelines with Pandas, Scikit-learn, and TensorFlow to support predictive models and improve decision-making.
- Implemented CI/CD pipelines using GitLab, Jenkins, Docker - enabling rapid, reliable deployment of data engineering solutions.
- Enhanced data governance and compliance by establishing metadata management, lineage tracking, and security policies across data systems.

## CERTIFICATIONS

**Microsoft Azure Fundamentals - AZ-900, AWS Academy Graduate, Microsoft Power BI Data Analyst Associate, Tableau eLearning certified Data Analyst, Google Analytics 4 certified, Campaign Manager 360 certified, Alteryx Designer Core**