

# TEJA MUMMANA

Email: tejamummana12@gmail.com

Mobile: +19452136457

Data Engineer

---

## PROFESSIONAL SUMMARY:

- Proven Data Engineer with 5+ years of experience configuring and ingesting structured, unstructured, and semi-structured data sources into cloud-based data pipelines, supporting analytics and DevOps delivery cycles.
- Expertise in maintaining end-to-end operational integrity of data transfers, ensuring compliance with security postures, including implementing and automating DoD STIGs, and infrastructure reliability across cloud platforms.
- Proficient in multiple programming/scripting languages (e.g., Python, Java, Scala, Bash, etc.) for automating configuration management processes using modern DevOps tools and techniques, ensuring scalable solutions.
- Strong experience with Big Data systems including Apache Spark and Databricks for distributed data processing and real-time analytics, identifying opportunities for automation and system right-sizing.
- Adept at implementing CI/CD pipelines with tools like Jenkins, GitLab CI, and Azure DevOps, ensuring smooth integration and deployment workflows, while monitoring the health of the cloud environment.
- Hands-on experience in developing and deploying serverless solutions using AWS Lambda, Azure Functions, and GCP Cloud Functions to streamline data processing pipelines and prototype innovative solutions.
- In-depth knowledge of cloud architecture and secure data operations, including data lake and data warehouse solutions such as Amazon Redshift, Azure Synapse Analytics, and Google BigQuery.
- Experienced in building real-time data pipelines using Apache Flink and Spark Streaming for high-throughput data ingestion and processing, while collaborating with multiple functional groups.
- Solid experience with containerization and orchestration tools like Docker and Kubernetes to manage scalable and portable data applications, applying expertise in cloud-native technologies.
- Expertise in using Terraform and CloudFormation for Infrastructure as Code (IaC) to automate cloud resource provisioning and build/maintain cloud infrastructure using vendor consoles.
- Extensive experience in ETL technologies such as Informatica, Talend, Apache Nifi, and SSIS to design, develop, and optimize data extraction, transformation, and loading processes.
- Proficient in using Sqoop for efficient data transfer between Hadoop and relational databases, Impala for interactive SQL queries, and Flume for reliable ingestion of streaming data.
- Developed a data quality assurance policy leading to a 20% decrease in data errors, while identifying and implementing scalable, maintainable, and high-performance coding approaches.
- Adept in managing large-scale data platforms AZURE, AWS, GCP using comprehensive suite of languages including SQL, Java and Python, and enhancing existing web scraping tools.
- Driven by designing powerful systems, enhancing user experiences, and looking forward to contributing value in a dynamic, cutting-edge technology environment to meet dynamic client demands.

## TECHNICAL SKILLS:

- **Programming Languages** - Python, Java, Scala, Bash
- **Data Processing** - Apache Spark, Hadoop, Hive, Pig, Kafka, Apache Flink, Spark Streaming, Apache Beam, Dataflow, Databricks
- **Cloud Platforms** - AWS, Azure, GCP, MilCloud 2.0
- **CI/CD & DevOps Tools** - Jenkins, GitLab CI, Azure DevOps, GitHub Actions, Cloud Build, Chef, Puppet
- **Databases** - PostgreSQL, MySQL, MongoDB, Cassandra, DynamoDB, Azure SQL, Cosmos DB
- **ETL Tools** - Informatica, Talend, Apache Nifi, SSIS, Azure Data Factory, AWS Glue
- **Data Warehousing** - Amazon Redshift, Azure Synapse Analytics, Google Big Query, Snowflake
- **Orchestration & Scheduling** - Apache Airflow
- **Infrastructure as Code** - Terraform, CloudFormation
- **Serverless Technologies** - AWS Lambda, Azure Functions, GCP Cloud Functions
- **Containerization & Orchestration** - Docker, Kubernetes, GKE, AKS

- **Real-time Data Tools** - Kafka, Event Hubs, Pub/Sub, Flume, Zookeeper
- **Reporting & BI Tools** - Power BI, Looker, Tableau, Quick Sight
- **Machine Learning Platforms** - Amazon SageMaker, Azure Machine Learning, Data Robot, TensorFlow, AI Platform Pipelines
- **Monitoring & Alerting** - Grafana, Prometheus, Splunk
- **API Development** - Flask, Django
- **Query Engines** - Impala, Hive
- **Scripting & Automation** - PyTest, SQL Unit Testing, Shell Scripting
- **Data Transfer & Integration** - Sqoop, SAP Integration, GCP SDK, Google Container Builders
- **Others** - Web Scraping

## **PROFESSIONAL EXPERIENCE:**

### **Bread Financials**

**April 2024 – Present**

#### **Azure Data Engineer**

##### **Responsibilities:**

- Configured and ingested structured, unstructured, and semi-structured data sources into cloud-based data pipelines, supporting analytics and DevOps delivery cycles, ensuring data quality and integrity. This involved working with diverse data formats and sources to build robust pipelines.
- Maintained end-to-end operational integrity of data transfers, ensuring compliance with security postures and infrastructure reliability, implementing monitoring solutions and automated alerts for proactive issue resolution. This ensured data security and availability.
- Automated configuration management processes using modern DevOps tools and techniques, including Azure DevOps and GitHub Actions, streamlining deployments and ensuring consistent environments across development, testing, and production stages. This improved efficiency.
- Monitored the health of the cloud environment using dashboards, logs, alerts, and performance metrics, proactively identifying and resolving issues to maintain optimal performance and availability of data pipelines and infrastructure components. This ensured system stability.
- Analyzed existing data architectures to identify opportunities for automation, optimization, cost savings, and system right-sizing, implementing solutions to improve efficiency and reduce operational expenses while maintaining performance and scalability. This improved resource utilization.
- Applied expertise in cloud-native technologies to build and maintain cloud infrastructure using vendor consoles and infrastructure-as-code practices, leveraging Terraform for provisioning and managing Azure resources, ensuring consistency and repeatability. This improved infrastructure management.
- Collaborated with cross-functional groups to understand mission needs and propose scalable, secure, and efficient technical solutions, working closely with stakeholders to gather requirements and design solutions that meet their specific needs. This improved collaboration.
- Implemented scalable, maintainable, and high-performance coding approaches using Python, Scala, and SQL, ensuring code quality and adherence to best practices, while optimizing performance for large-scale data processing and analytics workloads. This improved code quality.

### **Cardinal Health**

**Dec 2023 – March 2024**

#### **AWS Data Engineer**

##### **Responsibilities:**

- Configured and ingested structured, unstructured, and semi-structured healthcare data sources into cloud-based data pipelines supporting analytics and DevOps delivery cycles, utilizing AWS Glue and Lambda for efficient data processing and transformation. This improved data accessibility.
- Maintained end-to-end operational integrity of data transfers, ensuring compliance with security postures and infrastructure reliability, implementing AWS CloudWatch for monitoring and alerting, and adhering to security best practices for data protection. This ensured data security.
- Automated configuration management processes using modern DevOps tools and techniques, including Terraform and CloudFormation, streamlining infrastructure deployments and ensuring consistent environments across development, testing, and production stages. This improved efficiency.

- Monitored the health of the cloud environment using dashboards, logs, alerts, and performance metrics, proactively identifying and resolving issues to maintain optimal performance and availability of data pipelines and infrastructure components. This ensured system stability.
- Analyzed existing data architectures to identify opportunities for automation, optimization, cost savings, and system right-sizing, implementing solutions to improve efficiency and reduce operational expenses while maintaining performance and scalability. This improved resource utilization.
- Applied expertise in cloud-native technologies to build and maintain cloud infrastructure using vendor consoles and infrastructure-as-code practices, leveraging Terraform and CloudFormation for provisioning and managing AWS resources, ensuring consistency and repeatability. This improved infrastructure management.
- Collaborated with multiple functional groups to understand mission needs and propose scalable, secure, and efficient technical solutions, working closely with stakeholders to gather requirements and design solutions that meet their specific needs. This improved collaboration.
- Implemented scalable, maintainable, and high-performance coding approaches using Python, Scala, and SQL, ensuring code quality and adherence to best practices, while optimizing performance for large-scale data processing and analytics workloads. This improved code quality.
- Proven background with ETL pipelines and automation of ingest processes using AWS Glue, ensuring data quality and efficient data flow from various sources into the data lake for analytics. This improved data pipeline efficiency.

**Omni Channel (LTI MINDTREE)**

**Dec 2021 – April 2023**

**Sr. GCP Data Engineer**

**Responsibilities:**

- Configured and ingested structured, unstructured, and semi-structured data sources into cloud-based data pipelines supporting analytics and DevOps delivery cycles, utilizing Google Cloud Pub/Sub and Dataflow for real-time data ingestion and processing. This improved data accessibility.
- Maintained end-to-end operational integrity of data transfers, ensuring compliance with security postures and infrastructure reliability, implementing Google Cloud Monitoring and alerting, and adhering to security best practices for data protection. This ensured data security.
- Automated configuration management processes using modern DevOps tools and techniques, including Cloud Build and GitHub Actions, streamlining infrastructure deployments and ensuring consistent environments across development, testing, and production stages. This improved efficiency.
- Monitored the health of the cloud environment using dashboards, logs, alerts, and performance metrics, proactively identifying and resolving issues to maintain optimal performance and availability of data pipelines and infrastructure components. This ensured system stability.
- Analyzed existing data architectures to identify opportunities for automation, optimization, cost savings, and system right-sizing, implementing solutions to improve efficiency and reduce operational expenses while maintaining performance and scalability. This improved resource utilization.
- Applied expertise in cloud-native technologies to build and maintain cloud infrastructure using vendor consoles and infrastructure-as-code practices, leveraging Terraform for provisioning and managing GCP resources, ensuring consistency and repeatability. This improved infrastructure management.
- Collaborated with multiple functional groups to understand mission needs and propose scalable, secure, and efficient technical solutions, working closely with stakeholders to gather requirements and design solutions that meet their specific needs. This improved collaboration.
- Implemented scalable, maintainable, and high-performance coding approaches using Python, Scala, and SQL, ensuring code quality and adherence to best practices, while optimizing performance for large-scale data processing and analytics workloads. This improved code quality.
- Hands-on experience with Big Data systems including Apache Spark and Databricks on GCP Dataproc, processing large datasets for omni-channel data analysis and reporting, ensuring efficient data processing and scalability. This improved data processing capabilities.

**Sun Pharmaceutical Industries Ltd (ABSYZ)**

**June 2019 – Sep 2021**

**GCP Data Engineer**

**Responsibilities:**

- Configured and ingested structured, unstructured, and semi-structured data sources into cloud-based data pipelines supporting analytics and DevOps delivery cycles, utilizing Google Cloud Storage and BigQuery for efficient data storage and processing. This improved data accessibility.
- Maintained end-to-end operational integrity of data transfers, ensuring compliance with security postures and infrastructure reliability, implementing Google Cloud Monitoring and alerting, and adhering to security best practices for data protection. This ensured data security.
- Automated configuration management processes using modern DevOps tools and techniques, including Google Cloud Build and Cloud SDKs, streamlining infrastructure deployments and ensuring consistent environments across development, testing, and production stages. This improved efficiency.
- Monitored the health of the cloud environment using dashboards, logs, alerts, and performance metrics, proactively identifying and resolving issues to maintain optimal performance and availability of data pipelines and infrastructure components. This ensured system stability.
- Analyzed existing data architectures to identify opportunities for automation, optimization, cost savings, and system right-sizing, implementing solutions to improve efficiency and reduce operational expenses while maintaining performance and scalability. This improved resource utilization.
- Applied expertise in cloud-native technologies to build and maintain cloud infrastructure using vendor consoles and infrastructure-as-code practices, leveraging Google Cloud Deployment Manager for provisioning and managing GCP resources, ensuring consistency and repeatability. This improved infrastructure management.
- Collaborated with multiple functional groups to understand mission needs and propose scalable, secure, and efficient technical solutions, working closely with stakeholders to gather requirements and design solutions that meet their specific needs. This improved collaboration.
- Implemented scalable, maintainable, and high-performance coding approaches using Python, Scala, and SQL, ensuring code quality and adherence to best practices, while optimizing performance for large-scale data processing and analytics workloads. This improved code quality.
- Proven background with ETL pipelines and automation of ingest processes using Apache Airflow on GCP, ensuring data quality and efficient data flow from various sources into BigQuery for reporting. This improved data pipeline efficiency.

#### **Educational Details:**

**Master's in Computer Science** - Wright State University, USA