**Mani Madhav Goud Chetlapally**
Senior Data Engineer


**E-mail:** [redacted]
**Contact**: [redacted]
**LinkedIn:** https://www.linkedin.com/in/manimadhav/

---

## PROFESSIONAL SUMMARY

- Over 10+ years of experience in the IT industry, specializing in leveraging Azure tools and services, including **Azure ADLS GEN2**, **Azure Blob Storage**, **Azure Synapse Analytics**, **Azure Data Factory**, **Azure Functions**, **Azure Stream Analytics**, **Azure Logic Apps** and **Azure Cosmos DB**.
- Utilized **Azure ADLS GEN2** for creating data lake, and triggered **Azure Functions** upon file changes.
- Effectively managed the **Azure Data Lake Storage**, overseeing metadata management for efficient organization of tables, partitions, and databases.
- Led the migration of on-premises data to **Azure Synapse Analytics**, optimizing data storage and streamlining analytics processes, resulting in a 20% reduction in query response times.
- Implemented real-time analytics solutions on **Azure Databricks**, providing stakeholders with timely insights and improving decision-making processes by 20%.
- Engineered comprehensive **data models**, enabling seamless integration of diverse data sources, optimizing performance, and facilitating advanced analytics for informed decision-making.
- Spearheaded complex **data migration** initiatives, employing meticulous mapping and conversion strategies, resulting in streamlined processes, improved data accuracy, and enhanced operational efficiency.
- Orchestrated successful **data conversion** projects, ensuring smooth transition from legacy systems to modern platforms, mitigating risks, and preserving data integrity throughout the migration process.
- Transformed data modeling with **DBT**, optimizing data pipelines and increasing analytics accuracy, resulting in streamlined insights and informed decision-making.
- Led **Snowpark** integration in **Snowflake** for complex data transformations, enhancing **ETL** efficiency and enabling advanced analytics within **Snowflake**'s virtual data warehouse.
- Achieved a secure **Azure Virtual Machine** environment by configuring Network Security Groups, implementing Azure Active Directory roles, and regularly updating and patching instances.
- Implemented advanced access controls in **Azure Active Directory**, defining policies with conditions based on IP address, time, and Virtual Network, enhancing security and compliance measures.
- Employed **Azure Monitor** to collect, store, and analyze Azure services and application logs efficiently.
- Enhanced flexibility, expertly selecting and managing diverse database engines on **Azure SQL Database**; **MySQL**, **PostgreSQL**, **Oracle, SQL** Server, for optimal performance.
- Proficient in developing and optimizing large-scale Spark applications with **PySpark**, utilizing RDDs, DataFrames, and **Spark SQL** for efficient analytics and processing.
- Experienced in architecting data pipelines with **Azure Event Hubs**, managing partitions and topics, seamlessly integrating with **Azure Stream Analytics** for continuous analysis of streaming data.
- Optimized **T-SQL** query efficiency by implementing data partitioning and indexing, significantly reducing data scanning and enhancing overall query performance.
- Experienced in No-SQL databases like **Azure Cosmos DB** and **Cassandra**, proficient in schema design, data modeling, and query optimization, enabling scalable and high-performance storage and retrieval of unstructured data.
- Successfully automated workflows with **Azure Logic Apps**, ensuring security, scalability, and task orchestration.
- Successfully implemented **Power BI** solutions, collaborating with Data Analysts to create visualizations.
- Skilled in **Azure DevOps** version control, adept at branching, merging, and conflict resolution, ensuring code integrity.
- Proficient in Azure DevOps **CI/CD** pipelines, automating software build, test, and deployment for continuous integration and delivery.
- Streamlined project workflows using **Git, BitBucket, JIRA**, Confluence, and Notion, fostering transparent collaboration and accelerating project delivery.
- Expert in **Agile**, leading cross-functional teams through sprint planning, stand-ups, and retrospectives for software delivery.

---

## TECHNICAL SKILLS

| | |
|---|---|
| **Cloud technologies**: | Azure HDInsight, Azure Data Factory, ADLS GEN2, Azure Blob Storage, Azure Synapse Analytics, Azure DataBricks, Azure Cosmos DB, Azure DevOps, Purview, Azure Functional Apps, Azure Logic Apps, Azure Synapse Analytics, Entra ID, Azure Resource Manager, Azure Virtual Machines, Azure Load Balancer. |
| **Big Data**: | Spark, Hadoop, HDFS, MapReduce, YARN, Hive, Oozie, Pig, Sqoop, Presto, Zeppelin, Flink, ZooKeeper. |
| **Programming languages**: | Python, Scala, Java, SAS, PySpark, SQL, PL/SQL, T-SQL. |
| **Database**: | HBase, MongoDB, MYSQL, SQL SERVER, Oracle, PostgreSQL, Snowflake, Teradata. |
| **Data Visualization tools**: | Tableau, Power BI |
| **Machine Learning Libraries:** | Sci-kit learn, Pandas, NumPy, PyTorch, TensorFlow, Azure ML. |
| **Version control**: | Git, GitHub, BitBucket. |
| **Scripting languages**: | Shell scripting, Power Shell, Bash, UNIX/Linux |
| **Streaming platforms**: | Kafka, Confluent Kafka, Azure Event Hubs. |

---

## PROFESSIONAL EXPERIENCE

**State Of Oregon, Salem, Oregon**
**Senior Data Engineer** **Nov 2023 - Current**

**Use case:**

Designed and implemented a data pipeline solution for the State of Oregon's Person Management Redesign (PMR) project to improve internal IT operational efficiency and employee experience. The primary beneficiaries were IT support teams and agency employees, who gained real-time insights into helpdesk call trends by analyzing structured call data sourced from 8x8 software logs. This enabled the IT team to identify and prioritize recurring issues such as software bugs, hardware failures, and access bottlenecks leading to reduced call volumes and faster resolution times. The initiative enhanced system reliability, minimized employee downtime, and supported the PMR program's strategic goal of improving data-driven decision-making across modules.

**Responsibilities:**

- Led the development of a customer insights dashboard, utilizing diverse data sources including **Azure SQL Database**, **Azure Blob Storage**, **Azure Data Lake Storage GEN2**, and **Azure HDInsight**, employing **JSON**, **CSV**, **Parquet**, and **ORC** formats.
- Orchestrated efficient data streaming into **Azure Event Hubs** for initial processing, leveraging producers and consumers. Optimized topic structures, saving costs by improving efficiency by 40%.
- Implemented **Azure Synapse Analytics** for Spark transformations, leveraging Synapse catalogs for metadata storage, and crafting 100+ Spark jobs to process data from various sources including Azure Blob Storage and transactional servers.
- Optimized **Azure Data Factory** pipelines, reducing data processing time by 40%, enhancing **ETL** efficiency, and ensuring timely and accurate data delivery for analytics and reporting.
- Orchestrated **data migration** project, ensuring seamless transition between legacy and modern systems through meticulous **data modeling**, **mapping**, and **conversion strategies**, minimizing disruption.
- Led **data mapping** initiative, reconciling disparate data sources into a unified model, streamlining processes, and enhancing data integrity across the organization.
- Executed complex **data migration**, transforming legacy datasets into new formats while preserving integrity and accuracy, resulting in improved efficiency and accessibility.
- Integrated **Unity Catalog** with **Azure Databricks**, fortifying security, meeting compliance standards, and elevating **data governance** for enhanced regulatory adherence.
- Implemented complex workflows with **Apache Airflow**, ensuring seamless orchestration, efficient task scheduling, and reliable data pipelines for analytics and reporting.
- Engineered transformations on individual files using **Azure Functions**, scheduled through **Azure Logic Apps**, ensuring timely execution and maintenance of data processing tasks with minimal manual intervention.
- Designed and managed workflows using **Azure Logic Apps**, creating DAGs with **Python**, **SQL**, and Bash operators, ensuring seamless execution and monitoring of tasks with defined **CRON** expressions.
- Architected data pipelines for optimal transformation and loading into **Azure Synapse Analytics**, enabling large-scale data processing with reduced storage costs savings through efficient data management strategies.
- Designed and enforced granular Azure **RBAC** policies with custom roles and conditional access controls, reducing security risks by 85% while enabling least-privilege access across 200+ data resources.
- Spearheaded the migration of on-premises data to **Snowflake** on Azure, reducing query times by 40% and enhancing overall data accessibility for the organization.
- Implemented a robust data governance framework on **Snowflake**, ensuring compliance with industry regulations and improving data quality, leading to more informed decision-making.

- Orchestrated the integration of external data sources with **Snowflake** on Azure, enabling real-time analytics and providing valuable insights that contributed to a 25% increase in operational efficiency.
- Developed **Azure Functions** and **API** Management for seamless financial data submission.
- Identified and implemented cost-saving measures on **Azure Databricks** by optimizing resource utilization, resulting in a 15% reduction in cloud infrastructure costs.
- Employed **Spark** and **PySpark** for scalable data processing, accelerating analytics workflows and handling large datasets effectively.
- Leveraged **PySpark's** RDD and DataFrame APIs within Spark for distributed data processing, enhancing performance and scalability.
- Applied Spark's machine learning libraries, utilizing **PySpark** for model development, training, and evaluation on diverse datasets.
- Established a robust **CI/CD** pipeline using **Azure DevOps** and **Azure Functions** for efficient financial data processing.
- Deployed **Snowflake** on Azure, utilizing **Snow SQL** for scalable data querying and management, enhancing analytics with features like automatic scaling and native support for semi-structured data.
- Worked with **Python** and **Scala** to transform **Hive/SQL** queries into **Spark** (RDDs, DataFrames, and Datasets), customizing them for financial data processing.
- Led the design of real-time data pipelines to support Sanctions Screening, **AML** transaction monitoring, and **KYC analytics** using Azure Event Hubs and Databricks, enabling proactive fraud detection and regulatory adherence.
- Built secure, auditable data frameworks aligned with **FATF, FinCEN, and GDPR** standards, incorporating access controls, encryption, and audit trails using Azure RBAC and Snowflake governance features.
- Deployed PySpark and Azure ML for ML-driven anomaly detection pipelines tailored for fraud and risk scoring, increasing early threat detection accuracy by 30%.
- Architected high-availability and fault-tolerant pipeline systems across Azure Synapse and Databricks, ensuring continuous processing of high-volume financial transactions with minimal downtime.
- Collaborated directly with **AML, Fraud, and Compliance teams** to translate regulatory needs into technical solutions, streamlining workflows and boosting regulatory reporting efficiency by 25%.
- Expertise in using the **Scala** programming language to build **microservices** for financial data applications.
- Demonstrated experience with **Azure Event Hubs** for publish-subscribe messaging as a distributed commit log, with a particular focus on managing financial data streams.
- Leveraged **Azure Logic Apps** for debugging and monitoring scheduled jobs, streamlining troubleshooting processes within the workflow management system.
- Facilitated seamless integration of transformed data with **Power BI** for visualization, empowering stakeholders to derive actionable insights that led to informed decision-making with significant cost savings.
- Collaborated effectively with cross-functional teams, including business analysts, data scientists, and data engineers, ensuring alignment with business objectives and delivery of high-impact data solutions.

**Environment:** Azure Blob Storage, Azure SQL Database, Azure Data Lake Storage, Azure HDInsight, Azure Databricks, Unity Catalog, Azure Logic Apps, Azure Synapse Analytics, Azure Event Hubs, Azure Functions, Azure DevOps, Snowflake on Azure, Python, Scala, Spark (PySpark, SparkSQL), Kafka, Power BI, Linux, Java, Airflow, PostgreSQL, Oracle PL/SQL, Flink.

**TCS, San Fransisco, CA**
**Sr. Data Engineer** **Sep 2021 – Oct 2023**

**Responsibilities:**

- Collaborated with data scientists and utilized **Azure Data Factory**, including Azure Data Factory data flows and **data catalogs**, to develop a Spending Classification model on corporate card data, enhancing data organization and analysis capabilities.
- Designed and deployed Snowflake-based data solutions for a financial services client, implementing secure, scalable data models and compliance-aligned reporting for banking operations.
- Conducted extensive data exploration, gathering data from **Azure Synapse Analytics**, **Azure SQL Database**, and **Azure Data Lake storage GEN2** to facilitate comprehensive analysis and derive meaningful insights for informed decision-making processes.
- Leveraged **Azure API Management** for seamless API calls, employed **Azure Functions** to integrate data from Concur Expense management platform, and dynamically created **Azure Data Lake** Storage to streamline data management processes.
- Managed global financial data, leveraging **Azure Cosmos DB** and **Azure SQL Database** for efficient and scalable data storage.
- Developed **data migration** strategies, overseeing end-to-end execution to seamlessly transfer data between systems, ensuring minimal downtime and preserving data integrity throughout the process.
- Implemented robust **data modeling** techniques to design scalable and adaptable data structures, enabling efficient storage, retrieval, and analysis of critical business information.
- Conducted thorough **data mapping** exercises to identify relationships and dependencies between disparate datasets, optimizing data integration processes and enhancing data quality.
- Executed **data conversion** projects, transforming data from legacy systems to modern formats, enhancing accessibility, and facilitating the adoption of advanced analytics and reporting capabilities.

- Collaborated cross-functionally to define **data migration** requirements, establish conversion methodologies, and validate results, ensuring successful migration of business-critical data with minimal disruption.
- Leveraged **Azure Data Factory** data flows, and **Azure RBAC** for efficient metadata storage, and data management.
- Implemented feature engineering on large datasets using **Azure HDInsight** Spark clusters, optimizing data for model development and achieved a 20% improvement in processing efficiency.
- Successfully integrated machine learning models into **Azure Databricks** pipelines, enabling predictive analytics and improving business forecasting accuracy by 25%.
- Automated **SQL** queries and built **Azure Logic Apps** workflows for streamlined data processing and delivery, reducing manual intervention.
- Delivered data to data scientists in two modes - **Azure Data Lake Storage** and **Azure Synapse Analytics**, using **Azure Functions** for flexible and automated data access.
- Conducted data modeling in **Snowflake** on Azure, implementing **STAR schema & Snowflake schema**, and utilized **Azure SQL Database** for structured data representation with **Azure RBAC** ensuring secure access.
- Implemented **Azure Data Factory** ETL processes, enhancing data integration from diverse sources to **Azure Synapse Analytics**, optimizing performance, and ensuring seamless processing for analytics.
- Executed data cleansing and transformation tasks with **PySpark** on **Azure Databricks**, harnessing **Spark's** parallel processing capabilities for enhanced efficiency and performance.
- Established collaborative data science workflows on **Azure Databricks**, fostering cross-functional collaboration between data scientists, analysts, and engineers, leading to a 20% increase in productivity.
- Employed **Spark SQL** queries for seamless integration with various data sources, significantly improving data extraction, transformation, and loading (**ETL**) processes by 30%.
- Designed and executed **Azure Logic Apps** for orchestrating data workflows, significantly streamlining and automating complex data processing tasks across Azure services.
- Successfully integrated **Azure Synapse Analytics** for ad-hoc querying of data stored in **Azure Data Lake Storage**, providing users with quick insights, and facilitating dynamic, on-demand analysis.
- Deployed **Azure Event Hubs** for real-time data streaming, ingesting and processing high-velocity data, and enabling timely analytics for dynamic, event-driven applications.
- Orchestrated data migration and synchronization between on-premises databases and Azure using Azure **Database Migration Service (DMS)**.
- Configured **Azure Monitor** for comprehensive monitoring of data pipelines and infrastructure, proactively identifying and addressing performance issues, ensuring optimal system reliability and performance.
- Automated the ingestion of web server log data using **Azure Stream Analytics**, streamlining the process of storing data in **Azure Data Lake Storage**.
- Successfully implemented advanced techniques such as Partitioning, Dynamic Partitions, and Buckets in Hive, contributing to improved performance and logical data organization.
- Utilized **Apache Airflow** to automate and streamline data workflows, automating processes and reducing data engineers' overheads, allowing them to focus on more productive tasks.
- Enhanced data warehousing on **Snowflake** on Azure, ensuring scalability, multi-cloud flexibility, secure collaboration, time travel, versioning, and integration, including star schema & snowflake schema design for efficient analytics.
- Developed robust solutions for real-time data streaming using **Azure Kafka**, **Azure Stream Analytics**, and **Azure Databricks**, enabling immediate access and analysis of continuously generated data.
- Streamlined and optimized **ETL** processes using **Azure Databricks**, resulting in a 25% reduction in data processing time and increased data availability for business users.

**Environment:** Azure Blob Storage, Azure Data Factory, Azure Synapse Analytics, Azure SQL Database, Azure Data Lake Storage, Azure Databricks, Azure Logic Apps, Azure HDInsight, Azure Event Hubs, Azure Functions, Azure RBAC, Python, Scala, Spark (PySpark, SparkSQL), Kafka, Azure Synapse Analytics, Azure Cosmos DB, Linux, Java, Apache Airflow, PostgreSQL, Snowflake on Azure.

**CVS, Richardson, TX.**
**Data Engineer** **July 2020 to Oct 2021**

**Responsibilities:**

- Strengthened data security protocols by implementing **Azure Identity and Access Management (IAM)** policies and Virtual Network (VNet) configurations, ensuring restricted access to sensitive healthcare information.
- Implemented and optimized **Azure HDInsight** clusters for parallelized processing of large-scale healthcare datasets, reducing processing time and enhancing analytics capabilities.
- Streamlined healthcare data workflows by developing automated processes with **Azure Functions** and **Logic Apps**, improving efficiency and reducing manual intervention.
- Orchestrated scalable and cost-effective healthcare data storage solutions on **Azure Blob Storage**, facilitating seamless access and retrieval for various analytical purposes.
- Configured and optimized **Azure Virtual Machines (VMs)** to host healthcare applications, ensuring optimal performance and

responsiveness for healthcare professionals and end-users.

- Implemented Azure **Power BI** to create interactive and insightful dashboards, providing healthcare stakeholders with real-time visualizations for data-driven decision-making.
- Facilitated collaborative healthcare data analysis by creating shared data environments on Azure, fostering cross-functional teamwork and knowledge sharing.
- Developed automated reporting solutions using **Azure Functions** and **Azure Blob Storage**, ensuring timely and accurate generation of **healthcare** reports for internal and external stakeholders.
- Designed and implemented IAM policies and VNet configurations to ensure secure healthcare data management, achieving compliance with HIPAA/GDPR and strengthening data integrity.
- Implemented cost-saving measures by optimizing resource allocation and usage across **Azure VMs**, **Azure Blob Storage**, and other Azure services, ensuring efficient healthcare data infrastructure management.
- Implemented **Azure Databricks** for efficient analysis and visualization of relationships in large-scale datasets.
- Leveraged **Azure Stream Analytics** for real-time data streaming, enabling rapid processing and analysis of streaming data sources.
- Applied **Azure Machine Learning** pipelines for end-to-end model development, from data preparation to model deployment.
- Participated in all stages of **SDLC**, including requirement analysis, design, coding, testing, and production, for big data projects on Azure.
- Extensively utilized **Azure Data Factory** to import/export data between RDBMS and **Azure Data Lake**, creating data pipelines for last saved value, and performing incremental imports.
- Implemented efficient data storage solutions on **Azure Data Lake Storage**, optimizing healthcare data accessibility and retrieval for diverse analytical needs.
- Leveraged **Azure Databricks** for scalable and resource-efficient healthcare data processing, ensuring seamless scalability to handle growing volumes of data.
- Implemented **Azure Cosmos DB** for real-time processing of healthcare data, enabling immediate access and analysis of continuously generated information for timely insights.
- Utilized **Azure Synapse Analytics** to query and analyze structured healthcare data, optimizing performance and resource utilization for analytical purposes.
- Developed and optimized **PySpark** scripts for processing healthcare datasets, incorporating Azure Synapse SQL for structured data analysis, and creating Directed Acyclic Graphs (DAGs) for efficient workflow orchestration.

**Environment:** Azure Virtual Machines (VM), Azure Blob Storage, Azure Functions, Azure Logic Apps, Azure HDInsight, Azure RBAC (Role-Based Access Control), Power BI, Hive, Hadoop, Spark, SparkSQL, Scala, PySpark, Python, Sqoop, Kafka, Oracle.

**Costco, Miami, FL**
**Big Data Developer** **May 2018 to June 2020**

**Responsibilities:**

- Enhanced **ETL** processes using **Python**, **SQL**, and **Java**, improving efficiency by 30% for a large-scale data pipeline, resulting in faster data retrieval and analysis.
- Implemented **Hadoop** and **Hive** for handling vast datasets, reducing processing time by 40% and enabling seamless analysis of healthcare data.
- Optimized **MapReduce** jobs, leveraging Pig and Java, to process and transform raw data efficiently, enhancing the overall performance of big data processing pipelines.
- Implemented **HBase** for real-time healthcare data processing, enabling immediate access to continuously generated information and supporting timely insights for stakeholders.
- Utilized **YARN** to ensure scalable and resource-efficient data processing, enabling seamless scalability to handle growing volumes of healthcare data.
- Integrated Spark and **PySpark** for real-time data streaming, facilitating rapid processing and analysis of streaming data sources, enhancing responsiveness and insights.
- Applied **Spark SQL** for structured data analysis within healthcare datasets, optimizing performance and resource utilization for analytical purposes.
- Managed and optimized **HDFS** storage solutions, ensuring efficient accessibility and retrieval of healthcare data for diverse analytical needs.
- Implemented data processing solutions using **Teradata** and **Oracle** databases, improving data integrity and facilitating seamless integration with existing systems.
- Developed and executed efficient **PySpark** scripts for processing healthcare datasets, incorporating **SparkSQL** for structured data analysis, and creating DAGs for workflow orchestration.
- Deployed Postgres for structured data analysis, improving analytical capabilities and providing healthcare professionals with enhanced reporting and visualization tools.
- Collaborated in SDLC stages, from requirement analysis to production, ensuring successful implementation and maintenance of big data projects, contributing to improved healthcare data infrastructure.

**Environment:** Hadoop, Map Reduce, HDFS, Hive, Java, SQL, Cloudera Manager, Pig, Sqoop, Oozie, Hadoop, HDFS,Map

Reduce, Hive, HBase, Linux, Cluster Management

**Abbvie, Chicago, IL.**
**ETL Informatica Developer** **Feb 2014 to Mar 2018**

**Responsibilities:**

- Enhanced **ETL** workflows using Informatica, SQL, and Python, optimizing data extraction from Oracle databases, resulting in a 20% improvement in data processing efficiency.
- Developed complex Informatica mappings to transform and load data, leveraging **SQL** and **Python** for efficient data integration, contributing to streamlined **ETL** processes.
- Implemented Java transformations in Informatica, enhancing data processing capabilities and improving overall performance in a large-scale **ETL** environment for critical business systems.
- Utilized **SQL** queries to optimize Oracle database interactions, improving data retrieval efficiency and ensuring seamless integration with Informatica ETL processes for timely insights.
- Collaborated on **ETL** design and implementation, integrating Informatica and Python scripts for data validation, ensuring data accuracy and reliability across multiple systems.
- Conducted performance tuning of Informatica workflows, optimizing **SQL** queries, and enhancing overall **ETL** processing speed, resulting in significant time and resource savings.
- Designed and implemented Informatica workflows to extract, transform, and load data from various sources, utilizing **SQL** for data profiling and quality assurance.
- Applied Python scripting for data cleansing and transformation within Informatica workflows, improving data quality and facilitating accurate reporting for business stakeholders.
- Executed ETL tasks using Informatica PowerCenter, incorporating SQL optimization techniques, and enhancing overall system efficiency for large-scale data integration projects.
- Integrated **Oracle PL/SQL** within Informatica workflows, ensuring seamless communication between ETL processes and **Oracle databases**, enhancing data consistency and reliability.
- Developed Java-based custom transformations in Informatica, enabling complex data manipulations and contributing to the successful implementation of intricate ETL solutions.
- Automated Informatica **ETL** processes using Python scripts, reducing manual intervention, improving workflow reliability, and ensuring data consistency across diverse business systems.

**Environment**: Informatica Power Center 9.6, Oracle 11g, Oracle, Putty, Shell Scripting, Notepad++, Informatica, Oracle,ETL, Manual Testing, UNIX/Linux.

---

**CERTIFICATIONS:**

- **Microsoft Certified: Azure Data Engineer Associate (DP-203)**
- **Snowflake SnowPro Advanced: Data Engineer (DEA-C01)**

**EDUCATION DETAILS:**

- Masters from University of South Florida in Dec 2013
- Bachelor's from Amrita Vishwa Vidyapeetham in May 2011