

SAI SUDEEP TADIKAMALLA

Data Engineer

Email: vss.tadikamalla@gmail.com | Phone: +1(732) 723-8619 | LinkedIn: <https://www.linkedin.com/in/tadikamalla-sudeep/>

PROFESSIONAL SUMMARY:

- Results-oriented Data engineer with over 4+ years of experience in designing, developing, and maintaining scalable data infrastructure. Proficient in data modelling, ETL processes, and data warehousing techniques.
- Enterprise-level system creation and implementation may be practically mastered through the use of components such as MapReduce, YARN, Hive, HBase, Flume, Sqoop, Spark GraphX, Spark SQL, and Kafka. Expert in complex data processing and transformation utilizing Spark Core, Spark SQL, Spark GraphX, and Spark Streaming—a Scala-written in-memory computing framework.
- Collaborated together with Spark to improve the efficiency of the present techniques using Spark Context, Spark SQL, Spark MLlib, Data Frames, Pair RDDs, and Spark YARN.
- Practical knowledge of Databricks Workspace User Interface, Managing Databricks Notebooks, and Unified Data Analytics with Databricks, as well as Delta Lake with Spark SQL and Python.
- Configuring Data Bricks with AWS and Microsoft Azure, Using Data Bricks as a Workspace for Business Analytics, Managing Data Brick Clusters, and Managing the Machine Learning Lifecycle

TECHNICAL SKILLS:

Database	SQL Server, SQL, MySQL, MongoDB, PostgreSQL, Teradata SQL, Assistant, PyCharm, Autosys, ZOS and windows, Oracle
Programming and Markup Languages:	Python, Java, R, C++, Scala, Unix shell script, Cobol, SQL, and, PL/SQL, JavaScript, TypeScript, JSON, MATLAB, Shell, Bash, Python
Script Languages:	JavaScript, jQuery, Python.
Frameworks:	Pandas, NumPy, Scikit-learn, TensorFlow, KERAS, Flask, React, Flask, Apache Spark, Apache Flink, Kafka, Jenkins, GitHub, Docker, Kubernetes
Hadoop	HDFS, Hive, Pig, Sqoop, Yarn, Spark, SQL, Kafka, Horton work and ClouderaHadoop
Technologies and Tools:	Git, Docker, Amazon Web Service (AWS), Tableau, Google Cloud Platform, Kubernetes, Power BI, Apache Airflow, Talend, Amazon Redshift, Google Big Query
Version Control:	Git, GitHub, Git Flow, AWS, EC2, S3
Automation:	Terraform, AWS Cloud Formation, AWS Cloud Development Kit(CDK), AWS, MYSQL, RDS, Aurora Postgres, Ansible, Windows, Linux.
Monitoring and Logging:	Prometheus, Cloud Watch, ELK Stack
Fundamentals:	Machine Learning Algorithms, Exploratory Data Analysis, A/B Testing, Time Series Analysis
Machine Learning Models:	Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbor (KNN), Principal Component Analysis, Linear Regression, Naïve Bayes.
OLAP/Reporting:	SQL Server Analysis Services and Reporting Services.

PROFESSIONAL EXPERIENCE:

Dell Technologies Data Engineer

Feb 2023 to Present

- Involved in Requirement gathering, Business Analysis, Design and Development, testing, and implementation of business rules.

- Experience in developing Spark applications using Spark-SQL in Databricks for data extraction, transformation, and aggregation from multiple file formats for Analyzing & transforming the data to uncover insights into the customer usage patterns.
Experienced in building and optimizing data processing pipelines with Apache Spark, leveraging Spark SQL, Data Frames, and Spark Streaming for scalable big data analytics.
- Extract Transform and Load data from sources Systems to Azure Data Storage services using a combination of Azure Data factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics.
- Data ingestion to one or more Azure services (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in Azure Databricks.
Proficient in creating interactive dashboards and reports with Power BI, using DAX and Power Query to transform data and generate business insights.
- Understand business use cases, integration business, write business & technical requirements documents, logic diagrams, process flow charts, and other application-related documents.
- Design and develop ETL integration patterns using Python on Spark.
- Develop framework for converting existing PowerCenter mappings and to Pyspark (Python and Spark) Jobs. Create Pyspark frame to bring data from DB2 to Amazon S3.
- Translate business requirements into maintainable software components and understand impact (Technical and Business).
- Managed and maintained scalable cloud architectures on AWS and Azure, ensuring high availability and fault tolerance.
- Automated provisioning and deployment of cloud resources using Terraform, AWS CloudFormation, and Azure Resource Manager (ARM).
- Orchestrated all Data pipelines using Azure Data Factory and built a custom alerts platform for monitoring.
- Created custom alert queries in Log Analytics and used Webhook actions to automate custom alerts.
- Used Azure Key vault as a central repository for maintaining secrets and referenced the secrets in Azure Data Factory and also in Databricks notebooks.
- Built Teradata ELT frameworks which ingest data from different sources using Teradata Legacy load utilities.
- Built a common sftp download or upload framework using Azure Data Factory and Databricks. Maintain and support Teradata architectural environment for EDW Applications.
- Involved in the full lifecycle of projects, including requirement gathering, system designing, application development, enhancement, deployment, maintenance, and support
- Developing Data Extraction, Transformation, and Loading jobs from flat files, Oracle, SAP, and Teradata Sources into Teradata using BTEQ, Fast Load, Fast Export, MultiLoad, and stored procedures.

Trigent Software

May 2018 to Jul 2021

Data Engineer

- Worked on Informatica Advanced concepts & also Implementation of Informatica Push down Optimization technology and pipeline partitioning.
- Performed bulk data load from multiple data sources (ORACLE 8i, legacy systems) to TERADATA RDBMS using BTEQ, MultiLoad, and Fast Load.
- Skilled in setting up and managing Apache Kafka clusters, developing Kafka producers and consumers, and implementing real-time data streaming solutions.
- Used various transformations like Source qualifiers, Aggregators, lookups, Filters, Sequence generators, Routers, Update Strategy, Expression, sorters, Normalizer, Stored Procedures, Union, etc.
- Used Informatica Power Exchange to handle the change data capture (CDC) data from the source and load into Data Mart by following the slowly changing dimensions (SCD) Type II process.
- Used Power Center Workflow Manager to create workflows, and sessions, and also used various tasks like a command, event wait, event raise, and email.
-
-
- Developed ETL (Extract, Transform, Load) workflows using Apache Spark, ensuring efficient data transformation and integration with various data sources.
- Designed, created, and tuned physical database objects (tables, views, indexes, PPI, UPI, NUPI, and USI) to support normalized and dimensional models. Used Python and Shell scripts to Automate Teradata ELT and Admin activities. Performed Application-level DBA activities creating tables, and indexes, and monitored and tuned Teradata BTEQ scripts using Teradata Visual Explain utility.
Utilized Apache Kafka for real-time data ingestion and Spark Streaming for real-time analytics, enabling quick

decision-making in dynamic business environments.

- Performance tuning, monitoring, UNIX shell scripting, and physical and logical database design.
- Developed UNIX scripts to automate different tasks involved as part of the loading process.
- Worked on creating a few Tableau dashboard reports, Heat map charts and supported numerous dashboards, pie charts, and heat map charts that were built on the Teradata database.

EDUCATION:

M.S – Computer Science

Stevens Institute Of Technology – Hoboken, New Jersey Aug 2021 – Dec 2022

CGPA: 3.768/4.0

Bachelors in Computer Science and Engineering

RMK Engineering College – Anna University, Chennai, India Sep 2017 – Jun 2021

CGPA: 7.81/10