



## TARA REDDY

Senior Data Engineer – AWS | Azure | Python | Spark | Snowflake | ADF | Databricks | Data Lake

Ph:(443)-420-8421 | saitara529@gmail.com

### Professional Summary

- Sr Data Engineer with 10+ years of experience tackling Data Engineering, Business Intelligence problems, with competence in Analysis, Design, Development, Documentation, Deployment, and Integration using leading industry tools.
- Experience in Azure Cloud, Azure Data Factory, Azure Data Lake storage, Azure Synapse Analytics, Azure Analytical services, Azure Cosmos NO SQL DB and Data bricks.
- Experience in Server infrastructure development on AWS Cloud, extensive usage of EC2, IAM, RDS, S3, EBS, Route53, ECS, VPC, CloudFormation, CloudWatch, Redshift, lambda, DynamoDB, SNS, SQS, EMR, CloudTrail, AWS Configuration.
- Hands on exp Data extraction (extract, Schemas, corrupt record handling and parallelize code), transformations and loads (user -defined functions, join optimizations) and Production (Optimize and automate Extract, Transform and Load).
- Extensively worked with source code Version Control tools like SVN (Subversion), GIT, GitHub, Bitbucket.
- Expertise in using build tools like Maven for the building deployable artifacts such as Jar, War, Ear and Zip extension files.
- Proficiency in multiple databases like MongoDB, Cassandra, MySQL, ORACLE, and MS SQL Server.
- Developed Amazon Lambda functions using python Amazon S3 folder watcher.
- Build servers using AWS, importing volumes, launching EC2, RDS, creating security groups, Auto-scaling, Elastic load balancing (ELB) in the defined virtual private connection.
- Hands on exp on Unified Data Analytics with Databricks, Databricks Workspace User Interface, Managing Databricks Notebooks, Delta Lake with Python, Delta Lake with Spark SQL.
- Worked on projects in waterfall and agile methodology.
- Exp on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL database.
- Exp in Developing Spark applications using Spark – SQL in Databricks for data.
- Good understanding of Bigdata Hadoop and Yarn architecture along with various Hadoop Demons such as Job Tracker, Task Tracker, Name Node, Data Node, Resource/Cluster Manager and Kafka.
- Excellent understanding, solving big data problems using Hadoop ecosystem components such as HDFS, Map Reduce, Hive, Oozie, Autosys.
- Owning expertise in Power BI and Business Intelligence solutions as well as a thorough knowledge of the product and its operational requirements.
- Significant Experience with ELT and ETL loads using Azure Data Factory (ADF) V2
- Experience on Migrating SQL database to Azure Data Lake, Azure Data Lake Analytics, Azure SQL Database, Azure SQL Data warehouse and controlling and granting database access.
- Extensive experience of implementing solutions using AWS services like (EC2, S3, and Redshift), Hadoop HDFS architecture and Map-Reduce framework.
- Hands on working in AWS environment for development and deployment of custom Hadoop applications.
- Extensive knowledge in MPP Systems like Azure Synapse Analytics, Databricks, Hadoop.
- Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Databricks, NoSQL DB), SQL Server.
- Experience with MS SQL Server Integration Services (SSIS), T-SQL skills, stored procedures, triggers.
- Good understanding of Hadoop architecture and Hands on experience with Hadoop components such as Resource Manager, Node Manager, Name Node, Data Node and Map Reduce concepts and HDFS Framework.
- Expertise in writing Python codes for performing custom EL-TL tasks.
- Expertise in Apache Spark high level API programming.
- Experience in Spark transformations and Data frame creation.
- 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.
- Solid knowledge of Spark Streaming, Spark SQL, Data frame API, Data set API, and Spark RDD.
- Experience Spark SQL and Spark programming for huge volume of data workloads.
- Good understanding and knowledge of Microsoft Azure services like HDInsight Clusters, BLOB, ADLS, Data Factory and Logic Apps.
- Worked with various formats of files like delimited Parquet, text files, JSON files, XML Files. Mastered in using different columnar file formats like RC, ORC and Parquet formats.
- Good understanding of Amazon Web Services (AWS) like EC2 for computing and S3 as storage mechanism and EMR, Step functions, Lambda, RedShift, DynamoDB.

- Knowledge of Data mining and Data warehousing using ETL Tools and Proficient in Building reports and dashboards in Tableau (BI Tool).
- Developed visualization dashboards using Power BI with analysis views of drill down options using slicers and with various charts like Bar, Line, Scatter, Donut, Map and so on.
- Developed analysis reports and visualization using DAX functions like table function, aggregation function and time functions.
- Experience in Database Design and development with Business Intelligence using SQL Server 2014/2016, Integration Services (SSIS), DTS Packages, SQL Server Analysis Services (SSAS), DAX, OLAP Cubes, Star Schema and Snowflake Schema.
- Writing the Stored Procedures, Query to fetch the data for SSRS reports.
- Having good experience over T-SQL statements, Joins, Constraints, Views, Tables, Stored procedures.
- Used Jira Kanban board & Scrum – followed Agile methodology.
- Excellent written and verbal communication skills including presentation creation.

#### Skill Set

AWS environment	EMR, EC2, EBS, RDS, S3, Athena, Glue, Elasticsearch, Lambda, SQS, DynamoDB, Redshift, ECS, Quick Sight, Kinesis.
Azure environment	Azure Databricks, Azure Data Lake, Blob Storage, Azure Data Factory, SQL Database, SQL Data Warehouse, Cosmos DB, AAD, Azure batch.
Scripting Languages	Python, PySpark, Advanced SQL, Scala, Shell, PowerShell, HiveQL.
Databases	Snowflake, MySQL, Oracle, Teradata, MS SQL SERVER, PostgreSQL, DB2
NoSQL Databases	HBase, DynamoDB
Big Data Ecosystem	HDFS, Yarn, MapReduce, Spark, Kafka, Hive, Airflow, StreamSets, Sqoop, HBase, Flume, Ambari, Oozie, Zookeeper, Nifi, Apache Hadoop, Cloudera CDP, Hortonworks HDP
Others	Jenkins, Tableau, Power BI, Grafana
SDLC- Methodologies	Agile, Waterfall, Hybrid

#### Professional Experience:

Senior Data Engineer @ Elevance Health, Alexandria, VA

Aug 2022 – Till Date

##### Responsibilities:

- Analyze, design, and build Modern data solutions using Azure PaaS, SaaS to support visualization of data. Understand current Production state of application and determine the impact of new implementation on existing business processes.
- Creating Multiple Data Sets in Azure and working in migration of SQL Server 2008/2012/2014 to SQL Server 2017.
- Working in Azure Data Factory, Azure Databricks, Azure SQL, SQL Server Azure Data Lake, Cosmos DB.
- Building the pipelines to copy the data from source to destination in Azure Data Factory (ADF V2)
- Creating triggers and activities in Azure Data factory.
- Creating Stored Procedure and Scheduled them in Azure Environment.
- Monitoring Produced and Consumed Data Sets of ADF.
- Creating Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform, and load data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards.
- Developing JSON Scripts (ARM templates) for deploying the Pipeline in Azure Data Factory (ADF) that processes the data using the SQL Activity.
- Design and Develop ETL Processes in AWS Glue to migrate Campaign data from external sources like S3, ORC/Parquet/Text Files into AWS Redshift.
- 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.
- Implementing Spark applications for huge data loads.
- Estimating the cluster size, monitoring, and troubleshooting of the Spark data bricks cluster.
- Creating different visualizations like Line, Bar, Scatter Plot, Waterfall Chart, Heat map, Tree map, Card, Multi card, Table, Area chart, Pie chart, Donut chart and paginated reports using Power BI
- Creating Custom Visuals and Groups Creation.
- Implemented Installation and configuration of multi-node cluster on Cloud using Amazon Web Services (AWS) on EC2.
- Handled AWS Management Tools as Cloud watch and Cloud Trail.
- Stored the log files in AWS S3. Used versioning in S3 buckets where the highly sensitive information is stored.
- Integrated AWS Dynamo DB using AWS lambda to store the values of items and backup the DynamoDB streams.

- Automated Regular AWS tasks like snapshots creation using Python scripts.
- Designed data warehouses on platforms such as AWS Redshift, Azure SQL Data Warehouse, and other high-performance platforms.
- Built data workflows by using GCP, HBase, Big Query, AWS EMR, Spark, Spark SQL, Scala, and Python
- Prepared scripts to automate the ingestion process using Pyspark and Scala as needed through various sources such as API, AWS S3, Teradata and Redshift.
- Extract Transform and load data from sources systems to Azure data storage services using a combination of Azure Data Factory, Spark SQL and U-SQL, T- 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 Data bricks.
- Worked on Amazon Web service (AWS) to integrate EMR with Spark 2 and S3 storage and Snowflake.
- Creating pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform, and load data from different sources like Azure SQL, Blob Storage, Azure SQL data warehouse, write-back tool and backwards.
- Practice marketplace timeline slicer, Hierarchy slicer, Drill down and Drill up, Text filter and Word cloud on Power BI.
- Writing required DAX Queries to generate computed columns in Power BI.
- Role based Access creation in Power BI and providing access to Dashboards in Power BI Service based on the request from BA.

**Environment:** Azure Data Factory, Azure Databricks, Azure Logic Apps, Azure SQL, Delta Lake, Azure DW, Amazon Web Services (AWS Redshift, AWS Glue, AWS Athena, AWS S3, AWS RDS), AWS S3 PySpark, SQL, T-SQL, U-SQL, Apache Spark, Power BI, CosmosDB, MongoDB, Snowflake, Blob Storage, Scala, Spark v2.0.2, Hadoop 2.x (HDFS, MapReduce, Yarn), Airflow, Hive, Sqoop, HBase.

Azure Data Bricks Engineer @ Ally Bank, Detroit, MI

Apr 2021 – Jun 2022

#### Responsibilities:

- Involved in understanding client requirements, prepare design documents.
- Writing hive queries based on transformation logic and filters.
- Developing spark applications using Spark-SQL in Databricks for data extraction, transformation, and aggregation from multiple file formats for analyzing and transforming the data to uncover insights into the customer usage patterns.
- Extract Transform and load data from sources systems to Azure data storage services using a combination of Azure data factory, T-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.
- Develop spark applications using py-spark and spark SQL for data extraction, transformation, and aggregation from multiple file formats for analyzing and transforming the data uncover insight into the customer usage patterns.
- Developing SQL scripts for automation
- Demonstrated a full understanding of the Fact/Dimension data warehouse design model, including star and snowflake design methods.
- Developed a detailed project plan and helped manage the data conversion migration from the legacy system to the target snowflake database.
- Design, develop, and test dimensional data models using Star and Snowflake schema methodologies under the Kimball method.
- Responsible for estimating the cluster size, monitoring, and troubleshooting of the spark data bricks cluster.
- Ability to apply the spark Data Frame API to complete data manipulation within spark session.
- Involved in spark architecture including spark core, spark SQL, Data Frame, Spark streaming, Driver Node, Worker Node, Stages, Executors and Tasks, Deployment Modes, the execution hierarchy, fault tolerance and collection.
- Contributed on Data Modelling, data lineage and Data mapping.

**Environment:** Redshift, SQL, MPP, Databricks, Azure Data Factory, Logic App, Azure Service, Delta Lake, Azure Data Lake, HDInsight-HDFS, Hive Tables, Spark architecture, Spark Core, Spark SQL, Data Frames, Spark Streaming, Hadoop, Yarn, and Kafka.

Data Engineer @ Sentara Health, Hampton Roads, VA

Jan 2020 - Feb 2021

#### Responsibilities:

- Involved in understanding client requirements, prepare design documents.
- Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs & Python
- Analyzed the SQL scripts and designed the solution to implement using Spark.
- Extract Transform and load data from sources systems to Azure data storage services using a combination of Azure Data Factory, Spark SQL and U-SQL, T- 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 Data bricks.
- Created pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform, and load data from different sources like Azure SQL, Blob Storage, Azure SQL data warehouse, write-back tool and backwards.

- Experienced in performance tuning of Spark Applications for setting right batch interval time, correct level of parallelism and memory tuning.
- Architect & implement medium to large scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/Databricks, NoSQL DB).
- Built Spark applications with Pyspark and Spark-SQL for data extraction, transformation, and aggregation from various file formats for analysis and transformation of the data to reveal insights into customer usage patterns.
- To meet specific business requirements wrote UDF's in Pyspark
- Contributed on Data Modelling, data lineage and Data mapping.
- Developing SQL scripts for automation purpose

**Environment:** Azure, Snowflake, Spark Core, Spark SQL, Node, Hadoop, SQL ADF, HDInsight, u-SQL, t-SQL, spark SQL, VMS, PAAS, OLAP, DB, PROD.

Data Engineer @ State Farm, Bloomington, IL

Oct - 2018 – Nov -2019

Responsibilities:

- Writing hive queries based on transformation logic and filters.
- Provide guidance and direction to the offshore team and follow standard methodologies to drive towards business goals.
- Performed end to end automation of ETL / Data Ingestion process, Kubernetes Salesforce Developer for various datasets that are being ingested into the Azure.
- Worked with Informatica, PowerCenter, BDQ, IDQ, BDM, PIM, MDM, PIM, Looker, Mode Admin, Big Data, IBM DataStage ETL, ESB (Enterprise Service Bus), Logic Apps, API, BI, Tableau, DW to help form a cloud-based engineering team to deliver platform automation and security.
- Built data workflows with ETL, SSIS, BI, DW, Salesforce, AWS EMR, Data Lake, Redshift, Hadoop, Spark, Spark SQL, Scala, and Python
- We use Presto and Redshift for ad-hoc queries and analysis.
- Involved in generating reports and data creation.
- Extract Transform and load data from sources systems to Azure data storage services using a combination of Azure data factory, T-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.
- Contributed on Data Modelling, data lineage and Data mapping.
- Prepare design document and suggest feasibility and complexity of task.
- Solution road map for future growth, process architecture and mapping.
- Provide inputs on solution architecture based on evaluation/understanding of solution alternatives, frameworks and products.

**Environment:** Azure Data Factory, HDInsight, Databricks, ETL, Power BI, Data Lake, Azure MS SQL Server 2017 Services

Data Engineer @ Equifax, Atlanta, GA

Sep 2017 - Aug 2018

Responsibilities:

- Extract Transform and load data from sources systems to Azure data storage services.
- Performance tuning of Spark Applications for setting right batch interval time, correct level of parallelism and memory tuning.
- To meet specific business requirements wrote UDF's in Pyspark
- Contributed on Data Modelling, data lineage and Data mapping.
- Hands on experience on developing SQL scripts for automation purpose.
- Involved in generating reports and data creation.
- Interact with clients to elicit architectural and non-functional requirements like performance, scalability, reliability, availability, maintainability.
- Develop Power BI dashboards and reports to provide data insights and visualizations for stakeholders.
- Automated repeated tasks (Pipeline Monitoring, Data Quality check between Azure Data warehouse and OLAP Cube) using Azure Automation Account and Azure Function App
- Developed complex SQL queries using stored procedures, common table expressions (CTEs), temporary table to support Power BI reports.
- Developed solution for Azure Storage (Tables, Cosmos DB, SQL, Blob)
- Working with complex SQL views, Stored Procedures, Triggers, and packages in large databases from various servers.
- Create complex distributed systems that handle a lot of data, gather metrics, develop data pipelines, and use analytics.
- Created Complex ETL Packages using SSIS for extracting, cleaning, transforming, and loading data from staging tables to partitioned tables with incremental load.
- Perform performance optimizations on Hadoop, Oracle and Shell Scripts
- Design applications based on identified architecture and support implementation of design by resolving complex technical issues.

**Environment:** Azure SSAS, SSIS, SQL Azure, PowerBI, ETL, Cosmos DB, MS SQL Server 2012 Services

Responsibilities:

- Transformed raw data into actionable insights by incorporating various statistical techniques and using data mining tools such as PySpark and python.
- Responsible for tuning and optimizing Spark Jobs for Performance and scaling to larger datasets.
- Wrote batch process for validation and transformation of data using hive tables JSON and parquet files.
- Wrote Spark streaming consumer for Kafka to consume data from multiple topics and push in to HDFS.
- Increased marketing efficiency by 10% and reduced churn by leveraging advanced machine learning algorithms and statistics to provide data leading to better targeted messages and customer intervention.
- Extract Transform and load data from sources systems to Azure data storage services.
- Performance tuning of Spark Applications for setting right batch interval time, correct level of parallelism and memory tuning.
- To meet specific business requirements wrote UDF's in Pyspark
- Contributed on Data Modelling, data lineage and Data mapping.
- Hands on experience on developing SQL scripts for automation purpose.
- Involved in generating reports and data creation.
- Interact with clients to elicit architectural and non-functional requirements like performance, scalability, reliability, availability, maintainability.
- Perform performance optimizations on Hadoop, Oracle and Shell Scripts
- Design applications based on identified architecture and support implementation of design by resolving complex technical issues.