

****

**Mani   
 Sr Data Architect/ AWS Architect**

**910-593-2336**

**imran@zelecloud.com**

**SUMMARY**

* AWS certified Professional with **17+ Years** of IT Experience including 8+ Years of work experience in **Big Data**, **Hadoop Architecture**, **Spark Development**, **design**, **ETL development** and **Ecosystem Analytics** in Banking and Insurance domains.
* Over Seven plus years of high-performance experience with **Oracle E-Business suite** in technical, functional and management roles.
* Extensive experience in **Apache Spark**, **MapReduce**, **YARN**, **Scala**, **PySpark**, **Hive**, **Impala**, **Pig**, **Sqoop**, **Kafka**, **Hue**, **Hbase**, **Python**, **Oozie**, **Azure**, **Teradata**, **MongoDB**, **Unix shell scripting**, **Core Java**, **Scala, DMX-h tool**, **Mainframe** and **ETL** development.
* Done the **Teradata 12 Basics Certification**.
* Worked extensively on Python,Pyspark,Databricks,Kafka
* Worked on Snowflake, SnowSQL and Snowpipe
* Worked on Matillion tools
* Worked on Snowflake integration and Oracle to build a DataSet on Snowflake
* Good working experience with AWS and Azure
* 3+ years of extensive Snowflake Cloud data warehouse implementation on AWS
* 3+ years of good working experience with **Amazon Redshift, S3, EMR, GLUE, Aurora, Event Bridge and Lambda**
* Excellent gap analysis, project management, resource management, solution architecting and troubleshooting skills repeatedly tested in high-pressure environments. Worked extensively with EBS Setups, Architecture, and Development tools, on Financials, SCM, OTC, Manufacturing, PTP, and CRM modules.
* Strong working experience with Apache Airflow
* Worked with Unity Catalog Data Bricks for centralized access control, auditing, lineage, and data discovery capabilities across Databricks workspaces.
* Experience in GCP Dataproc, GCS, Cloud functions, BigQuery.
* Used cloud shell SDK in GCP to configure the services Data Proc, Storage, BigQuery
* Work related to downloading BigQuery data into pandas or Spark data frames for advanced ETL capabilities.
* Created Data Ingestion Framework in Snowflake for the Batch Data from different file formats (XML, JSON, Avro) using Snowflake Stage and Snowflake Data pipe.
* 2+ years of extensive knowledge of Data Science, Data Analytics, and data engineering using Azure Data Bricks.
* Worked closely with the Data Governance team and Data Quality teams in identifying and correcting data flaws.
* Good Working experience with Azure Data Bricks
* Worked on Devops Operations
* Ingested data into Azure Blob storage and processed the data using Azure Databricks.
* Strong knowledge of implementation of **SPARK core - SPARK SQL**, **GraphX, Flink and Spark streaming**
* Experienced in writing **PySpark** scripts and doing the **joins/ETL** operations on data frames.
* Good Knowledge on **Scala-based** **Spark programs**.
* Implemented cloud services **IAAS, PAAS and SaaS** which includes **Openstack, Docker and Openshift**
* Good Working Experience on **Google Big Query and GCP**
* Used **Kubernetes** to orchestrate the deployment, Scaling, and management of Docker Containers.
* ng data using **Sqoop** from HDFS to **Relational Database Systems/Non-Relational Database** Systems and vice-versa. And handled huge incremental data using Sqoop.
* Experience in using Hive partitioning, bucketing, and executing different types of **joins** on **Hive** tables and implementing **Hive SerDes like JSON and Avro**.
* Good understanding of NoSQL databases and hands-on work experience in writing applications on **NoSQL databases like HBase and MongoDB**.
* Good working knowledge of RESTful API’s used for HTTP requests to GET, PUT, POST and DELETE Data.
* Strong Experience in working with Databases like **Teradata**, **Oracle**,and **MySQL** and proficiency in writing **complex SQL** queries.
* Good knowledge of processing **semi-structured** and **unstructured** data.
* Working on different file formats like **JSON**, **Parquet**, **Avro**, **XML**, **CSV**, **XLS** etc.
* Good knowledge of **Normalization**, Fact Tables and Dimension Tables, also dealing with **OLAP** and **OLTP** systems.
* Experience in identifying and resolving **ETL production root cause** issues.
* Excellent understanding of **Hadoop architecture**, Hadoop Distributed File System, and various components such as HDFS, Name Node, Data Node, Job Tracker, Task Tracker, YARN, **Spark Architecture** and MapReduce concepts.
* Experienced working with **JIRA** for project management, **GIT** for source code management, **JENKINS** for continuous integration and **ServiceNow** for change request/Incident management.

**TECHNICAL SKILLS**

* **Big Data Technologies –**  Hadoop, MapReduce, YARN, Hive, Pig, HBase, Impala, Hue, Sqoop, Kafka
* **Spark components –**  RDD, Spark SQL (Data Frames and Dataset), Spark Streaming.
* **Programming Languages:** SQL, PL/SQL, Core Java, Python, Scala, Shell Scripting, Cobol
* **ERP Tools** - Oracle applications R12 and 11i (Financials, SCM like Account Payables, Account Receivables, General Ledger, Order Management, Purchasing, Inventory, Work in Progress, Bill of Materials, Enterprise Asset Management, AOL, Alerts, System Administration)
* **Databases –**  Oracle 12c/11g, Teradata 15/14, MySQL, SQL Server2016/2014, DB2 and Mongo DB.
* **Scripting and Query Languages –**  Shell scripting, PL/SQL, Java Script, HTML, DHTML, XML
* **Version Control / Schedule –**  Git, GitHub, Tortoise SVN, Autosys, CA7, Oozie
* **Cloud Infrastructure –**  Amazon Web Services (S3, Athena, Glue, EMR, Lambda), Microsoft Azure, Snowflake Integrations with Informatica
* **ETL tools –**  Syncsort DMX-h and Informatica
* **Operating Systems –**  Windows, UNIX/Linux, and Mainframe z/OS
* **Tools –**  Eclipse, IntelliJ, SQL Assistant, SQL Plus, SSH Tecta, Super Putty, JIRA, Incident Management (IM)

**EDUCATION**

* Bachelors in computer science at JNTU,
* Year – 2007, India

**CERTIFICATIONS**

* AWS Architect
* SnowPro Core

**WORK EXPERIENCE**

**Zelecloud (Dallas,TX)**

**Sr Snowflake Architect/ AWS Engineer**

**Dec’2022 – Till Date**

**Project: FannieMae**

**Position: Sr Snowflake Architect/ AWS Engineer**

**Responsibilities:**

* Created HBase tables to load large sets of structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios.
* Worked on REST API’s that uses HTTP request for communication with web services. These were following certain constraints like, Client-server architecture, Stateless and Cacheable.
* Worked on moving some of the Data Objects on to Amazon AWS
* Supporting the current Production System’s with daily scheduled jobs and other on demand tickets
* Developed pig scripts to transform the data into structured format and it are automated through Oozie coordinators.
* Worked on Unity Catalog for data bricks setting up the below roles,

Create a storage container where the megastore’s managed table data will be stored. ...

Create an identity that Azure Databricks uses to give access to that storage container. ...

Provide Azure Databricks with the storage container path and identity.

* Developed end-to-end data analytics framework
* Worked on Snowflake, SnowSQL and Snowpipe

Worked on Matillion tools

Worked on Snowflake integration and Oracle to build a DataSet on SnowflakeAutomated resulting scripts and workflows using Apache Airflow and Shell Scripting to ensure daily executing in Production.

* Install and configure Apache Airflow **for S3 bucket** and Snowflake data warehouse and created dags to run all Airflows.
* Strong working experience with transforming data onto **Snowflake** using **DBT**.
* Worked majorly on **Snow pipe** for connecting between AWS S3 and **Snowflake** for ETL process.
* Performed data quality issue analysis using Snow SQL by building analytical warehouses on **Snowflake**
* Create and maintain highly scabble and fault tolerance multi-tier AWS environments spanning across multiple available zones using Terraform!
* Write Terraform scripts. From scratch for building Dev, Staging and Prod Environments.
* Performed migration of large sets of data to Databricks (Spark). Create and Administer Cluster, load data, configure data pipelines.
* Involved in transforming the relational database to legacy labels to HDFS, and HBASE tables using Sqoop and vice versa.
* Worked on building AWS DataLakes
* Consumed the data from Kafka queue using Spark. Configured different topologies for Spark cluster and deployed them on regular basis.
* Written PySpark code to import data from MongoDB collection to HDFS and ingested the data back to MongoDB collection.
* Worked on different file formats (AVRO, PARQUET, TEXTFILE, CSV) and different compression codecs (GZIP, SNAPPY, LZO).
* Worked on AWS Services like Athena, Glue, EC2 and S3.
* Worked on ETL using Spark, Hive, HBase, Sqoop, Oozie on Hadoop.
* Experience in Azure Data Services like Azure Data Factory, Azure Data Lake, Azure Databricks.
* Worked within Azure Cloud Services (PaaS & IaaS), Azure Databricks, Azure Synapse Analytics, SQL , Azure Data Factory, Azure HDInsight, Key Vault, Azure Data Lake for data ingestion, ETL process, data integration, data migration.
* Ingested data into Azure Blob storage and processed the data using Azure Databricks.
* Created Data Ingestion Framework in Snowflake for the Batch Data from different file formats (XML, JSON, Avro) using Snowflake Stage and Snowflake Data pipe.
* Worked with Infrastructure and Network Security teams to implement Security and Authentication between Snowflake and AWS.
* Creating the database structures to replicate existing structures and applications in Teradata, which are to be migrated to the Big Data stack such as Hive, Spark, etc.
* Worked on Spark SQL and Data frames for faster execution of Hive queries using Spark.
* Worked on ETL tools like TPT loads, Sqoop run book and stream ingestion.
* Created HBase tables to load large sets of structured and unstructured data coming from UNIX, NoSQL, and a variety of portfolios.
* Involved in transforming the relational database to legacy labels to HDFS, and HBASE tables using Sqoop and vice versa.
* Consumed the data from Kafka queue using Spark. Configured different topologies for Spark cluster and deployed them on regular basis.

**City of Memphis Light Gas and Water (Memphis, TN & Dallas,TX)**

**Senior Data Architect/AWS Engineer**

**Sep’2014 – Nov’2022**

**Project: Memphis Light, Gas and Water**

**Responsibilities:**

* Worked on AWS Services like Athena, **Glue, EC2 and S3**.
* Working on legacy data systems SQL Server and had been helping to massage the data for reporting purpose to different teams.
* Hands on experience in setting up Workflow using Apache Airflow
* Worked on ETL using Spark, Kafka, Hive, HBase, Oozie on Hadoop.
* Wrote various data normalization jobs for new data ingested into Redshift
* Bulk loading from the external stage (AWS S3), internal stage to snowflake cloud using the COPY command.
* Loading data into **snowflake** tables from the internal stage using snowsql.
* Worked on DBT for loading, transforming the data into **Snowflake**.
* Responsible for Designing Logical and Physical data modelling for various data sources on Redshift.
* Created Databrick Notebooks to Streamline and Curate the data for various business use cases
* Experience in Azure Data Services like Azure Data Factory, Azure Data Lake, Azure Databricks.
* Created Data Ingestion Framework in Snowflake for the Batch Data from different file formats (XML, JSON, Avro) using Snowflake Stage and Snowflake Data pipe.
* Worked with Infrastructure and Network Security teams to implement Security and Authentication between Snowflake and AWS.
* Creating the database structures to replicate existing structures and applications in Teradata, which are to be migrated to the Big Data stack such as Hive, Spark, etc.
* Worked on Spark SQL and Data frames for faster execution of Hive queries using Spark.
* Worked on ETL tools like TPT loads, Sqoop run book and stream ingestion.
* Created HBase tables to load large sets of structured and unstructured data coming from UNIX, NoSQL, and a variety of portfolios.
* Involved in transforming the relational database to legacy labels to HDFS, and HBASE tables using Sqoop and vice versa.
* Consumed the data from Kafka queue using Spark. Configured different topologies for Spark cluster and deployed them on regular basis.
* Written PySpark code to import data from MongoDB collection to HDFS and ingested the data back to MongoDB collection.
* Worked on different file formats (AVRO, PARQUET, TEXTFILE, CSV) and different compression codecs (GZIP, SNAPPY, LZO).
* Written complex Hive queries involving external dynamic partitioned on Hive tables which stores rolling window time-period user viewing history.
* Building/Maintaining **Docker** container clusters managed by Kubernetes Linux, Bash, GIT
* Utilized **Kubernetes** for the runtime environment of the **CI/CD** system to build, test deploy.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDD, Scala and Python.

**TCS(Tata Consultancy Services), Hyderabad, India**

**Senior Data Engineer**

**Jul 2011 – Aug 2014**

**Project: Great Batch Medical , Bank of America**

The received data from various sources should be massaged and make the data usable for analytical platform, ETL should be done on the received data to EAP cluster and make it available in hive tables.

**Responsibilities:**

* For New applications in Hadoop development of Data Ingestion and ETL workflows for various EAP data sources. Working with the Business analysts and the DBAs for requirements gathering analysis, coding, testing, deployment, and project coordination.
* Worked on ETL using Spark, Kafka, Hive, HBase, Oozie on Hadoop.
* Creating the database structures to replicate existing structures and applications in Teradata, which are to be migrated to the Big Data stack such as Hive, Spark, etc.
* Worked on Spark SQL and Data frames for faster execution of Hive queries using Spark.
* Worked on ETL tools like TPT loads, Sqoop run book and stream ingestion.
* Created HBase tables to load large sets of structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios.
* Good working knowledge on T-SQL
* Worked on REST API’s that uses HTTP request for communication with web services. These were following certain constraints like, Client-server architecture, Stateless and Cacheable.
* Worked on moving some of the Data Objects on to Amazon AWS
* Supporting the current Production System’s with daily scheduled jobs and other on demand tickets
* Developed pig scripts to transform the data into structured format and it are automated through Oozie coordinators.
* Involved in transforming the relational database to legacy labels to HDFS, and HBASE tables using Sqoop and vice versa.
* Consumed the data from Kafka queue using Spark. Configured different topologies for Spark cluster and deployed them on regular basis.
* Written PySpark code to import data from MongoDB collection to HDFS and ingested the data back to MongoDB collection.

**SQLSTAR International Ltd**

**Oracle ERP Techno-Functional**

**Sep 2007 – Jun 2011**

**Projects:**

**Responsibilities:**

* Single handedly involved & implemented Project Management Related Activities like, Estimations, Allocation of Work, Analysis on project requirement scope, Design, Test Cycle Support, coordinated onsite project management activities with the client, functional team, QA Teams, cross flow track teams, Transition Teams, BSOS at various stages of the project life cycle, ensured smooth go live.
* Involved in MD50 Analysis and prepared Technical Specification, Unit Test Case documents. Coding and Unit testing of Forms/Packages/Reports and other PL/SQL objects. Involved in review of other migrated Components.
* Involved in Production Support (Maintenance) and documenting Functional, Technical and migration requirements.