[](https://www.credly.com/badges/7b8f57c7-fdf8-4747-8ed4-c9041464dce2/)A black diamond with white text and red star

Description automatically generated

**SALMA SULTHANA PATAN**

**AWS Data Engineer**

Sumanth@codcores.com

214-227-8816

**EDUCATION & CERTIFICATIONS/ACHIEVEMENTS:**

* Bachelor’s Degree in Computer Science Engineering from JNTU, India (2016)
* Master’s Degree in Computer Science Engineering from Southern Arkansas Univeristy
* Linked In : <https://www.linkedin.com/in/salma-sulthana-p/>
* DBT Fundamentals : <https://credentials.getdbt.com/8f198103-59ad-4fd4-813d-f820142a46c8>
* Snowflake - SnowPro Core Certified

CredlyE-verify - <https://www.credly.com/badges/7b8f57c7-fdf8-4747-8ed4-c9041464dce2/>

**PROFESSIONAL SUMMARY:**

* 8+ years of experience in Analysis, Design, Development and Implementation as a Data Engineer, with IT industries
* 5+ years of working Experience on Big Data Technology using technologies such as cloudera and Hortonworks Distributions. Hadoop working Environment includes Hadoop, Spark, MapReduce, Kafka, Hive, Ambari Sqoop, HBase and Impala, PySpark& Spark SQL, Python, Snowflake, S3 storage, Airflow
* Experience in ETL tools like **DBT** DataStage, Informatica, Oracle
* Expert in providing ETL Solutions for any type of business models, optimized and tuned ETL processes and SQL Queries for better performance
* Experience in SnowflakeCloud data warehousing shared technology environment for providing stable infrastructure, architecture, best practices, secured environment, reusable generic frameworks, robust design architecture, technology expertise, best practices and automated SCBD (Secured Database Connections, Code Review, Build Process, Deployment Process) utilities.
* Experience with AWS (Amazon EC2, Amazon S3, Amazon RDS, Amazon Elastic Load Balancing, Amazon SQS, AWS Identity and access management, Amazon SNS, AWS Cloud Watch, Amazon EBS, Amazon CloudFront, VPC, DynamoDB, Lambda and Redshift)
* Experience in Snowflake streams, Snowpipe, tasks, Data sharing, Views, Snow SQL
* Create Hive Tables loaded data & write Hive Ad-hoc queries to run internally in MapReduce & Spark
* Expertise in integration of various data sources like RDBMS, Spreadsheets, Text files, JSON and XML
* Experience with Relational & dimensional data modelling, Snowflake schema, fact and dimension tables and process mapping using the top-down and bottom-up approach.
* Experience in doing performance tuning for map reduce jobs & hive complex queries.
* Experience in DW, Data Architecture & ETL from sources into DW and Data Marts using DataStage
* Experience in Data Modelling, MDM, Data Models data marts, OLAP (MOLAP, ROLAP).
* Experience in Data Modelling, Data requirement analysis & data mapping for ETL processes, Predictive modelling, Data Visualization, Data mining with large data sets
* Experience in execution, optimization and troubleshooting SQL queries - T-SQL, HIVE SQL.
* Involved in building Data Models & Dimensional Modelling with 3NF, Star and Snowflake schemas for OLAP and Operational data store (ODS) applications.
* Involved in data testing and validation processes, Experience in scheduling batch jobs in Control-M.
* Experience in creating architectural process for Automation with Python.
* Experience in working with Databases like Teradata and proficiency in writing complex SQL, Snow SQL, PL/SQL for creating tables, views, indexes, stored procedures, and functions.
* Experienced in design, management and visualization of databases using Oracle and MySQL.
* Experience in validating batch jobs, input files and output files in UNIX environment.
* Experience in full software development life cycle (SDLC) of ETL process including requirement analysis, design, development, support of testing and migration to production.
* Experience in Building Management tools- DBT, Version Control Git &PM tools - Agile, Jira, Jenkins
* Demonstrating developed Stories, Features to Product Owners and Business teams.
* Experience in Waterfall and Agile methodologies, attendmeetings & hand off calls.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Programming Languages | Python, PySpark, Shell Scripting, SQL, HiveQL, Java |
| ETL Tools | Informatica Power Center, DataStage, Tableau |
| Databases | HBase, AWS RDS, Teradata, SQL Server, PostgreSQL, Oracle,DB2, Aurora |
| Bigdata Ecosystem | Spark, Spark SQL, Spark Streaming, Python Spark, Hive |
| Integration Tools | Git, GitHub, Jenkins, Docker, Kubernetes |
| Cloud Technologies | Snowflake, Azure, AWS Redshift, S3, Redshift, Spectrum,RDS,AWS Glue,Athena,LambdaCloudwatch,HIVE,Presto, EMR |
| Scripts | Advanced SQL, Python, PL/SQL, Unix shell scripting, XML,JSON |
| Management Tools | Skype, Teams, Slack, Jira, QTest |
| IDE | IntelliJ, PyCharm, Eclipse, Visual Studio Code |
| Build Tools | Date Build Tool, Gradle, Maven |
| Other Tools | Apache Storm, Putty, GitLab, Control-M, Normalization/ De- normalization |
| Operating System | Windows, iOS, Unix, Linux |

**PROFESSIONAL EXPERIENCE:**

**Verizon, Irving, TX Jun2021 - Current**

**Snowflake / Data Engineer**

**Responsibilities:**

* Use DBT cloud to connect to Snowflake and set the development, deployment environments
* Extensively used jinja syntax to bring programmability on SQL before writing to Snowflake
* Working on Snowflake pipeline to configure data loads from AWS S3 to into Snowflake Virtual DataWarehouse
* Use JSON &XML for serialization and deserialization to load JSON,XML data into Hive tables
* Imported data from AWS S3 into Spark RDD’s and performed Transformations & Actions on RDD’s
* Working on integration engagements in services such as EC2, S3, RDS Worked with AWS ecosystem (EC2, S3, Lambda, RDS, Redshift)
* Worked on AWS Data pipeline to configure data loads from S3 to into Redshift
* Using **AWS** Redshift, Extracted, transformed & load data from various data sources &destinations
* Setting up HBase column-based storage repository for archiving data on daily basis
* Worked with SQL across several dialects (we commonly write HBase, MySQL, PostgreSQL, Redshift, SQL Server, and Oracle), Like Access, Excel, CSV, Oracle, flat files using connectors, tasks and transformations provided by **AWS Data Pipeline**
* Scala programming proficiency for developing data processing applications and ETL (Extract, Transform, Load) processes.
* Query performance optimization in Snowflake and involved other teams for user experience Source Data from various feeds-Semi/Structured data including flat files, Parquet, JSON &XML
* Created Snow pipe for continuous data load from staged data residing on cloud gateway server Implemented data transformations on Snowflake using various built-ins like Snow SQL, Snowpipe
* Use Flatten object table to produce output in several formats – variant, object & array columns
* Worked on UNIX Shell scripting and batch execution in Putty.
* Designed Power BI Reports and published on Power BI server.
* Experience with Scala libraries and frameworks commonly used in AWS data engineering, such as Apache Spark for distributed data processing.
* Use SparkSQL to load JSON data & create schema RDD & loaded it into Hive tables
* Responsible for Design, Development, testing of the database and developed Stored Procedures, Views, and Triggers
* Participate in development of DQ sync framework using PySpark. Application compares parties between source systems and Master Data Management (MDM) system to determine missing parties and mismatched attributes
* Expertise in writing complex DAX functions in Power BI. Worked on calculated Columns, Measures queries in Power BI desktop to show good data analysis techniques.
* Worked on all types of transformations that are available in Power BI query Editors.
* Worked on PySpark programming daily to perform transformations for applying business logic
* Converting data load pipeline algorithms written in python and SQL to Scala spark and PySpark
* Utilized Spark SQL API in PySpark to extract and load data and perform SQL queries
* Involved in developing PySpark script to encrypting the raw data by using hashing algorithms concepts on client specified columns
* Worked in the CI/CD pipeline management for managing the weekly releases
* Provide capacity planning, performance tuning and monitoring
* Attending Stand-ups daily and providing updates on assigned stories
* Defining user stories out of features in Program meetings, iteration planning & track them in JiraSprint planning, Story Demonstration and PI planning,

**Environment:**PySpark, SQL, AWS (AWS Datalake, EC2, S3, Lambda, RDS, Redshift, Glue Components), Snowflake, Oracle, Unix, PL/SQL, PostgreSQL, MDM, REST APIs, dbt, UNIX, Snowflake Web UI, Snow SQL Snowflake – Streams, Snowpipes, ETL, AWSEC2, S3, RDS, Tasks, Snow SQL (Snowflake CLI), SQL, Python, JSON, XML, PowerShell, UNIX

**SiriusXM Feb 2019–Mar2021**

**Data Engineer**

**Responsibilities:**

* Migrated applications from Django to Flask &NoSQLto SQL, used Jenkins, AWS, Bitbucket
* Worked on Automating, Configuring and deploying instancesAWS EC2, Cloud watch, Cloud Formation and managing security groups on AWS
* Troubleshoot Production issues pertaining to AWS Cloud Resources and Application Infrastructure
* Demonstrated ability to write efficient and scalable Scala code for data manipulation, transformation, and analysis.
* Develop Shell scripts for cleansing, validating files using AWK, SED, grep & UNIX commands
* Developed Shell scripts implementing PL/SQL queries for data migration & batch processing
* Designed multiple Python packages that were used within a large ETL process used to load 2TB of data from an existing Oracle database into a new PostgreSQL cluster
* Wrote python scripts to parse JSON documents and load the data in PostgreSQL database
* Worked on generating various dashboards in Tableau/Power BI using various datasources like HANA, Snowflake, Salesforce, Oracle, SQL server, Excel, MS Access.
* Used Apache Airflow to schedule &monitor workflow & to orchestrate complex data pipelines
* Knowledge of integrating Scala-based applications with AWS services such as Amazon S3, Amazon Redshift, Amazon EMR, and AWS Glue.
* Worked on Using **ETL** to implement the Slowly Changing Dimension Transformation, to maintain Historically Data in Data warehouse, worked with JSON based REST Web services
* Designed, Implement & solutions for using Docker, Jenkins, & Puppet for Microservices &CI/CD
* Worked on improvement of existing machine learning algorithms to extract data points accurately.
* Responsible for setting up Python REST API framework using Flask
* Engaged in Design, Development, Deployment, Testing and Implementation
* Create user stories of features in Program meetings, iteration planning &tracking them using Jira
* Attending Standups meetings and Updating Tasks in Jira on time
* Integrated custom visual based on business requirements using Power BI Desktop.
* Worked in converting Hive/SQL queries into Spark transformations using Spark RDDs and experience in using Spark-Shell and Spark Streaming
* Wrote python scripts to parse JSON documents and load the data in PostgreSQL database
* Implemented Data Governance, Data quality frame work including data cataloging, data lineage, MDM, business glossary, data steward, metadata -operational.
* Worked on SQL for Querying the database in the UNIX environment
* Established and maintained comprehensive data model documentation including detailed descriptions of business entities, attributes, and data relationships
* Proposed users to use Power BI in replacement of Analysis for Office, to run reports connecting to Snowflake.
* Worked on Kubernetes for the runtime environment of the CI/CD system to build, test deploy
* Worked on Flask and Snowflake queries
* Responsible for loading the customer’s data and event logs from Kafka into HBase using REST API
* Used Jenkins for continuous integration and delivery platform over GIT
* Worked on Spark using Python, Scala & utilizing Data frames &Spark SQL API for processing of data
* Performing ETL testing activities like running the Jobs, Extracting the data using necessary queries from database transform, and upload into the Data warehouse servers
* Automated most of the daily task using python scripting
* Defining user stories out of features in Program Increment meetings, planning & tracking on Jira
* Attending Standups meetings and Updating Tasks in Jira on time.

**Environment**:PySpark, AWS, Docker, NoSQL, Jenkins, MDM, UNIX, SCDs, Hive/SQL, Spark SQL API, PostgreSQL, GIT, CI/CD, REST, Flask, Snowflake, Jira.

**United Health Group, India Mar2016–Feb 2019**

**Data Analyst**

**Responsibilities:**

* Created action filters, parameters, calculated sets for preparing dashboards, worksheets in Tableau
* Created Tableau scorecards, dashboards using stack bars, bar graphs, scattered plots, geographical maps, Gantt charts functionality, created SQL Queries for testing the tableau Dashboards
* Analysed long running queries and create the indexes and recommend for query tuning
* Created and managed Tableau sites, projects, and workbooks, groups, data views, data sources, and data connections, Worked on database Design, Development and Support
* Used principles of Normalization to improve the performance. Involved in ETL code using PL/SQL to meet requirements for Extract, transformation, cleansing & loading of data from source to target data structures, Publishing dashboards and extracts to Tableau server
* Create claims dashboards & reports in Tableau. Created report schedules on Tableau server
* Combined visualizations into Interactive Dashboards and publish them to the web.
* Involved in installation and configuration of Tableau Server.
* Defined best practices for Tableau report development.
* Worked on Shell scripts implementing PL/SQL queries for data migration & batch processing
* Worked on Creating Shell Scripts for automated execution of batch process Wrote SQL Queries, Dynamic-queries, sub-queries, Stored Procedures, Triggers, User-defined Functions
* Performed Data mapping between claims source systems to Target systems, created class diagrams and Diagrams and used SQL queries to filter data
* Attending Standups meetings and Updating Tasks in Jira on time
* Performed Business data gathering and report generation for Insurance and claims using Tableau.
* Gathered user and business requirements through surveys, Clients and subject matter experts to define the scope and requirements of Health Insurance.
* Resolving Service request & Critical request raised for any technical issues in processing the claims.
* Performed Document Analysis, Brainstorming and JAD Sessions to gather the requirements.
* Performed Data cleaning and normalization along with profiling to ensure Data Quality
* Gave permissions to specified Business users to access the Report using Tableau server
* Developed plans for verification and validation of requirements at all levels so that developed software successfully fulfilled user expectations and business needs.
* Generating reports for Forecasting and Trend Analysis and provide data driven recommendations.
* Experience in Portfolio analysis and risk management by creating the performance matrix
* Developing test plans for ensuring delivery of quality dashboard or reports
* Generating reports and Power point presentation for Insurance
* Providing solutions to ensure the controls are well understood & incorporated for Data Analytics
* Using Decision support system for Translating data into meaningful information that can be used by operational management to make decisions
* Facilitated requirements discussions with business & technology partners, assisted the quality assurance department in their test efforts by clarifying requirements & addressing issues

**Environment:**UNIX, Tableau, PL/SQL, SQL server, MS Excel, Erwin.