Sreenath bathula

Email: sreenadhc7@gmail.com   **PH:470-724-3135**

Sr. Data Engineer

Professional Summary

- With over 8 years of IT experience, I've specialized in Business Requirements Analysis, Application Design, Development, Coding, testing, and implementation, primarily using **Azure Synapse Analytics, Azure Data Bricks, Azure Data Factory, MapReduce, and Hadoop Ecosystem**.
- Professional in developing pipelines, activities, data sets, linked services for Azure Data Factory.
- Expertise on creating pipelines, implementing spark pool notebooks, views for serverless and dedicated SQL pools on **Azure Synapse Analytics**.
- I have significant experience in deploying **Hadoop** Ecosystems, designing and supporting Data Warehouses using SQL Server and Azure, and handling large-scale data extraction via **Azure Data Factory**.
- Exposure on usage of **Apache Kafka** to develop data pipeline of logs as a stream of messages using producers and consumers.
- Expertise in writing and tuning complex SQL queries, developing data pipelines with **Apache Kafka,** creating and managing Azure Data Factory elements, and implementing event-based data movement on Azure platforms.
- I've worked with **AWS Cloud Formatio**n for IAM Roles and architecture deployment, provisioned EC2 Instances using Terraform, and have experience with **Snowflake** multi-cluster warehouse. I've also designed and deployed AWS Solutions using various tools and services.

Technical skills:

Big Data Stack	Spark, MapReduce, Kafka, Hadoop
Cloud Services	Azure data bricks, Azure blob storage, Azure data lake , Azure Data factory, Azure synapse Analytics,
Programming Languages	Python, Scala, PL/SQL, C, C++and JAVA
Relational Databases	MySQL, SQL Server, PostgreSQL,
NoSQL databases	Cassandra, HBase, Cosmos DB
Libraries	Pandas, NumPy, Scikit-learn, Matplotlib, Tweepy, Seaborn, TensorFlow, Keras, MLlib, Boto3.
Data Analysis Skills	Data Cleaning, Data Visualization, Feature Selection, Pandas.
Web technologies	Django, Flask, HTML/CSS, JavaScript
BI Tools	Tableau, Tableau server & Reader, Power BI
Operating Systems	Unix, Linux, Windows
Version Control Systems	Bitbucket, GIT, SVN, GitHub

Professional Experience

Sr. Data Engineer

Charles schwab, TX   **Jan 2022 to Present**

Responsibilities:

- Used **Azure Data Factory** as an orchestration tool for integrating data from upstream to downstream systems.
- Used Azure Data Factory extensively for ingesting data from disparate source systems.
- Created data model for **Azure Synapse Analytics** to analyze Bugsplat data.
- Created, provisioned different **Data bricks clusters** needed for batch.
- Implemented **Azure Data Bricks clusters, notebooks**, jobs and auto scaling.
- Implemented **Airflow** for the scheduling jobs.
- Automated jobs using different triggers (Event, Scheduled and Tumbling) in **ADF.**
- Implemented several DAX functions for various fact calculations for efficient data visualization in **Power BI.**

- Developed batch processing solutions by using **Data Factory and Azure Data bricks.**
- Designed and built modern data solutions on Azure Cloud using Data platform capabilities like **Azure Data Lake,**

  Azure Data Factory, Azure data bricks and Azure SQL Data warehouse.

- Actively involved in data requirements gathering and translating them into technical documentation.
- Developed JSON Scripts for deploying the Pipeline in **Azure Data Factory (ADF)** that process the data using the SQL

  Activity.

- Integrated **Azure Active Directory** authentication to every **Cosmos DB** request sent and demoed feature to

  Stakeholders.

- Designed end to end scalable architecture to solve business problems using various Azure Cloud Services like Data factory, Data Lake, function Apps and SQL Server.
- Created Application Interface Document for the downstream to create new interface to transfer and receive the files through **Azure Data Share.**
- Creating pipelines, data flows and complex data transformations and manipulations **using ADF** and **PySpark with Databricks.**

Environment: Python, SQL , Spark , ETL , Azure Cloud, Azure Data Factory (ADF), Azure functions Apps, Azure Data Lake, BLOB Storage, SQL Server, Azure synapse, Shell scripting, Azure PowerShell, Data Bricks, Python, Data Modelling Tool and Azure Cosmos DB.

Sr. Data Engineer   **Dec 2019 to Jan2022**
United Healthcare, MN
Responsibilities:

- Developed **Spark applications using PySpark** and Spark-SQL for data extraction, transformation, and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
- Designed and built modern data solutions on Azure Cloud using Data platform capabilities like Azure Data Lake, **Azure Data Factory, Azure data bricks** and **Azure SQL Data warehouse.**
- Extract, Transform and Load data from Source 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 Ingression to one or more Azure Services **(Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in Azure Databricks.**
- Experience in **creating PowerShell scripts** for automating clusters and storage accounts in production environments.
- Moved on-premises applications to Azure cloud with **Azure Data factory**. Created Self-hosted integration run time to provide data-integration capabilities across different network environments.
- Deployed **snowflake** workloads and data warehouse on azure cloud platform.
- Used various Transformations **in SSIS Dataflow**, control Flow using for loop Containers and Fuzzy lookups.
- Developed **Informatica** mappings to be implemented based on client requirements and for the analytics team.
- Developed JSON Scripts for deploying the Pipeline in **Azure Data Factory (ADF)** that process the data using the SQL Activity.
- Have Extensive experience in creating pipeline jobs, schedule triggers using **Azure Data Factory**.
- Created Linked services to land the data from **SFTP location and Azure SQl Database to Azure Data Lake.**
- Involved in migrating the client data warehouse architecture from on-premises into **Azure cloud**.
- Hands-on experience in Azure Cloud Services (PaaS and IaaS), Azure Synapse **Analytics, SQL Azure**, Data Factory, Azure Analysis services, Application Insights**, Azure Monitoring, Key Vault, Azure Data Lake.**
- Developed batch processing solutions by using Data Factory and **Azure Data bricks**. Implemented **Azure Data Bricks clusters**, notebooks, jobs and auto scaling.

- Implemented several DAX functions for various fact calculations for efficient **data visualization in Power BI.**

Environment: SQL Database, Python, Spark , Azure data factory, Azure data lake storage, Azure synapse analytics, Azure synapse workspace, Synapse sql pool, Power BI, Python, Data masking, Azure data bricks, Azure SQL Data warehouse, Azure Cosmos DB, Azure stream analytics.

Sr. Data Engineer
Early Warnings, Scottsdale, AZ   **Jan 2019 to Dec2020**
Responsibilities:

- Working on implementing data pipeline solutions on **AWS EMR** to perform aggregations and ETL transformations on different data sources based on business requirements by developing Spark applications using **Pyspark.**
- Designed and Developed ETL Processes using AWS Glue to migrate Campaign data from external data sources like **S3** into **AWS Redshift.**
- Migrated the on-prem database to **AWS RDS (mySQL)** by creating Data Retention on multi-AZ and configured read replicas to increase performance.
- Implemented stored procedures to handle data transformations and aggregates based on complex business logic on AWS RDS and created tables, views to enhance the performance of database.
- Migrated **RDBMS (mySQL)** structured data to **AWS S3 by utilizing Sqoop.**
- Integrated and Consolidated data from various data sources which include **S3 , RDS and KNESIS to perform real-time data analytics by feeding data to AWS Redshift Datawarehouse.**
- Build best practice **ETLs with Apache Spark** to load and transform raw data into easy-to-use dimensional data for self-service reporting.
- Used **AWS step functions** to synchronize the data pipelines which ensures the sequential execution of jobs.
- Created **AWS lambdas** as webhooks for triggering event notifications.
- Using **python scripts** to automate Data pipeline.
- Setting up data in AWS using **S3 bucket** and configuring automated instance backups to S3 bucket.
- Expertise in building **CI/CD on AWS environment** using AWS Code Commit, Code Build, Code Deploy and

    Code Pipeline.

- Implemented workflow tools including Stack storm, **Airflow and AWS Glue.**
- Installed and configured Apache Airflow on server to automate data ingestion from multiple data sources to **AWS S3 bucket** by creating **DAGS** to run the scheduled jobs**.**
- Running **Pyspark** jobs to perform data transformation using data frames and **Pyspark sql** functions.
- ETL pipelines in and out of data warehouse using combination of python and snowflakes **snow SQL** queries against **snowflake.**

        Environment: Python, SQL , Spark, AWS, Redshift , AWS Glue, EMR and S3.

Data Engineer
Creator Technologies, Hyd   **Dec 2017 - Dec 2018**
Responsibilities:

- Coded model level validation and provide guidance in making long term architectural design decisions and also used Agile Methodology and **SCRUM process**.
- Utilized the existing **Python and Django modules** and rewritten to deliver data in required formats.
- Built database Model, **APIs** and Views utilizing Python, to build an interactive web-based solution.
- Worked on object-oriented programming (OOP) concepts using **Python and Linux**.
- Embedded **AJAX** in UI to update small portions of the web page avoiding the need to reload the entire page.
- Used Python in-built libraries urllib2 and beautiful soup modules for web scraping.
- Installed and configured **Hadoop ecosystem** like **HBase, Flume, Pig, and Sqoop.**
- Designed and developed **Big Data analytic** solutions on a **Hadoop**-based platform and engage clients in technical discussions.

- Installed, Configured, and Maintained the **Hadoop cluster** for application development and Hadoop ecosystem components like **Hive, Pig, HBase, Zookeeper, and Sqoop.**
- Designed Django REST web services using Python and Django to get and post data.
- Used **Python and Django** creating graphics, XML processing of documents, data exchange and business logic implementation between servers.
- Logged user stories and acceptance criteria in **JIRA** for features by evaluating output requirements and formats.

Environment: Python, Django, HTML5, Hadoop, **Hive**, Git, Jira, Agile, Windows.

Data Analyst
Yana Software Pvt. Ltd, Hyderabad   **Jan 2016 to Nov 2017**
Responsibilities:

- Documented the system requirements to meet end-state requirements and complied Software Requirement Specification Document and **Use Case document.**
- Prepared ETL (Extract, Transform and Load) standards, naming conventions and wrote **ETL flow** documentation.
- Used **Microsoft** SharePoint to upload, manage all project related documents and have version control.
- Prepared Dashboards using calculations, parameters in **Tableau** and generated KPI reports to be analysed by management.
- Worked on **SQL queries** for data manipulation.
- Arranged weekly team meetings to **assign testing tasks** and acquisition of status reports from individual team members.
- Effectively managed change by deploying change management techniques such as **Change Assessment, Impact Analysis and Root Cause Analysis.**
- Used advanced **Excel functions** to generate spreadsheets and pivot tables.
- Presented solutions in written reports while analyzing, designing, **testing,** and monitoring systems in a **waterfall methodology.**