**Naveen**

**Email: jack@trinity-world.com**

**Phone: 248 876 0464**

**SUMMARY:**

* Overall 9+ Years of experience can be headhunted for a Lead level position across any functional sectors within an IT organization of repute
* Experience on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse and Controlling and granting database access and Migrating On premise databases to Azure Data lake store using Azure Data factory.
* 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.
* Wrote AWS Lambda functions in python for AWS's Lambda which invokes python scripts to perform various transformations and analytics on large data sets in EMR clusters.
* Involved in all the stages of Software Development Life Cycle Primarily in Database Architecture, Logical and Physical modeling, Data Warehouse/ETL development using MS SQL Server 2012/2008R2/2008, Oracle 11g/10g, and ETL Solutions/Analytics Applications development.
* Good experience in Spark Architecture including Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors and Tasks.
* Instantiated, created, and maintained CI/CD (continuous integration & deployment) pipelines and apply automation to environments and applications. Worked on various automation tools like GIT, Terraform, Ansible.
* Good experience in Big Data Hadoop and Yarn architecture along with various Hadoop Demons such as Job Tracker, Task Tracker, Name Node, Data Node, Resource/Cluster Manager, and Kafka (distributed stream-processing).
* Hadoop Ecosystem, AWS Cloud Data Engineering, Data Visualization, Reporting and Data Quality Solutions.
* Good experience in Amazon Web Services like S3, IAM, EC2, EMR, Kinesis, VPC, Dynamo DB, RedShift, Amazon RDS, Lambda, Athena, Glue, DMS, Quick Sight, Amazon Elastic Load Balancing, Auto Scaling, CloudWatch, SNS, SQS and other services of the AWS family.
* 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.
* Experienced in importing real time streaming logs and aggregating the data to HDFS using Kafka and Flume.
* Unit Testing, Integration Testing and Performance Testing of Informatica, Talend, IRIS tool jobs and stored procedures.
* Well versed with various Hadoop distributions which include Cloudera (CDH), Hortonworks (HDP), Azure HD Insight.
* Extending HIVE and PIG core functionality by using custom User Defined Function's (UDF), User Defined Table-Generating Functions (UDTF) and User Defined Aggregating Functions (UDAF) for Hive and Pig.
* Experience working on NoSQL Databases like HBase, Cassandra and MongoDB.
* Experience in Python, Scala, shell scripting, and Spark.
* Experience with Testing Map Reduce programs using MRUnit, Junit and EasyMock.
* Experience on ETL methodology for supporting Data Extraction, transformations and loading processing using Hadoop.
* Worked on data visualization tools like Tableau and also integrated the data using ETL tool Talend.
* Hands on development experience with JAVA, Shell Scripting, RDBMS, including writing complex SQL queries, PL/SQL, views, stored procedure, triggers, etc.
* Preparing ETL test strategy, designs and test plans to execute test cases for ETL and BI systems
* Proficient in SQL and other relational databases.
* Experience working in a cross-functional AGILE Scrum team.
* Good working knowledge of Amazon Web Services(AWS) Cloud Platform which includes services likeEC2,S3,VPC,ELB, IAM, DynamoDB, Cloud Front, Cloud Watch, Route 53, Elastic Beanstalk (EBS), Auto Scaling, Security Groups, EC2 Container Service (ECS), Code Commit, Code Pipeline, Code Build, Code Deploy, DynamoDB, Auto Scaling, Security Groups, Red shift, CloudWatch, Cloud Formation, CloudTrail, Ops Works, Kinesis, IAM, SQS, SNS, SES
* Good experience in working with cloud environment like Amazon Web Services (AWS) EC2 and S3.
* Expertise working with AWS cloud services like EMR, S3,Redshift, EMR cloud watch, for big data development.
* Proficiency in SQL across several dialects (we commonly write MySQL, PostgreSQL, Redshift, SQL Server, and Oracle)

**TECHNICAL SKILLS:**

**Hadoop/Big Data:**HDFS, MapReduce, Spark, Yarn, Kafka, Apache NiFi, PIG, HIVE, Sqoop, Storm, Flume, Oozie, Impala, HBase, Hue, Zookeeper.

**Programming Languages:**Java, PL/SQL, Python, HiveQL, Scala, SQL, Azure Power Shell.

**Development Tools:**Eclipse, SVN, Git, Ant, Maven, SOAP UI

**Databases:**Oracle 11g/10g/9i, Teradata, MS SQL

**No SQL Databases:**Apache HBase, Mongo DB, Cassandra

**Distributed platforms:**Hortonworks, Cloudera, Azure HD Insight

**Operating Systems: UNIX, Ubuntu Linux and Windows 00/XP/Vista/7/8**

**ETL Tools: Automic, Informatica, Talend, Ab-Initio, IRIS**

**Other Technologies: Azure Data lake, Data factory, Azure Databricks, Azure SQL database, Azure SQL Datawarehouse**

**PROFESSIONAL EXPERIENCE:**

**T-Mobile, NJ**

**Jan 2022 – till date**

**AWS/Azure Data Engineer**

**Responsibilities:**

* Implemented Azure Data Factory (ADF) extensively for ingesting data from different source systems like relational and unstructured data to meet business functional requirements
* Design and developed Batch processing and real-time processing solutions using ADF, Databricks clusters and stream Analytics
* Worked on Azure Data Factory to integrate data of both on-prem (MY SQL, Cassandra) and cloud (Blob storage, Azure SQL DB) and applied transformations to load back to Azure Synapse.
* Created numerous pipelines in Azure using Azure Data Factory v2 to get the data from disparate sourcesystems by using different Azure Activities like Move &Transform, Copy, filter, for each, Databricks etc. Maintain and provide support for optimal pipelines, data flows and complex data transformations and manipulations using ADF and PySpark with Databricks.
* Created, provisioned different Databricks clusters, notebooks, jobs and autoscaling.
* Performed data flow transformation using the data flow activity.
* Used Polybase to load tables in Azure synapse.
* Used AWS Glue for transformations and AWS Lambda to automate the process.
* Implemented Azure, self-hosted integration runtime in ADF.
* Improved performance by optimizing computing time to process the streaming data by optimizing thecluster run time.
* Scheduled, automated business processes and workflows using Azure Logic Apps.
* Designed and developed a new solution to process the NRT data by using Azure stream analytics, Azure Event Hub and Service Bus Queue.
* Implementing and Managing ETL solutions and automating operational processes.
* Designed data warehouses on platforms such as AWS Redshift, Azure SQL Data Warehouse, and other high-performance platforms.
* Created Linked services to connect the external resources to ADF.
* Worked with complex SQL views, Stored Procedures, Triggers, and packages in large databases from various servers.
* Used Azure Devops& Jenkins pipelines to build and deploy different resources (Code and Infrastructure) in Azure.
* Written PySpark job in AWS Glue to merge data from multiple tables and in Utilizing Crawler to populate AWS Glue data Catalog with metadata table definitions.
* Experience in resolving any technical issue, Troubleshooting, Project Risk & Issue identification, and management.
* Experience managing Azure Data Lakes (ADLS) and Data Lake Analytics and an understanding of how to integrate with other Azure Services.
* Migration of on-premise data (Oracle/ Teradata) to Azure Data Lake Store(ADLS) using Azure DataFactory(ADF V1/V2).
* Work closely across teams (Support, Solution Architecture) and peers to establish and follow best practices while solving customer problems
* Created Lambda functions to run the AWS Glue job based on the AWS S3 events.
* Created infrastructure for optimal extraction, transformation, and loading of data from a wide variety of data sources.
* Designed and created optimal pipeline architecture on Azure platform.
* Created pipelines in Azure using ADF to get the data from different source systems and transform the data by using many activities.
* Wrote various data normalization jobs for new data ingested into Redshift
* Importing & exporting database using SQL Server Integrations Services (SSIS) and Data Transformation Services (DTS Packages).
* Created Linked service to land the data from different sources to Azure Data Factory.
* Created different types of triggers to automate the pipeline in ADF.
* Created, provisioned different Databricks clusters needed for batch and continuous streaming data processing and installed the required libraries for the clusters.

**Environment:** Azure Data Factory (ADF v2), Azure SQL Database, Azure functions Apps, Azure Data Lake, BLOB Storage, AWS, SQL server, Windows remote desktop, UNIX Shell Scripting, AZURE PowerShell, Data bricks, Python, ADLS Gen 2, Azure Cosmos DB, Azure Event Hub, Azure Machine Learning.

**DELL, NY**

**June 2019 – Dec 2021**

**AWS/AZURE DATA ENGINEER**

**RESPONSIBILITIES:**

* Created Linked Services for multiple source system (i.e.: Azure SQL Server, ADLS, BLOB, Rest API).
* Created Pipeline’s to extract data from on premises source systems to azure cloud data lake storage; Extensively worked on copy activities and implemented the copy behavior’s such as flatten hierarchy, preserve hierarchy and Merge hierarchy; Implemented Error Handling concept through copy activity.
* Exposure on Azure Data Factory activities such as Lookups, Stored procedures, if condition, for each, Set Variable, Append Variable, Get Metadata, Filter and wait.
* Configured the logic apps to handle email notification to the end users and key shareholders with the help of web services activity; create dynamic pipeline to handle multiple source extracting to multiple targets; extensively used azure key vaults to configure the connections in linked services.
* Developed server-side software modules and client-side user interface components and deployed entirely in Compute Cloud of Amazon Web Services (AWS).
* Create data ingestion modules using AWS Glue for loading data in various layers in S3 and reporting using Athena and QuickSight.
* Configured and implemented the Azure Data Factory Triggers and scheduled the Pipelines; monitored the scheduled Azure Data Factory pipelines and configured the alerts to get notification of failure pipelines.
* Migrated on premise database structure to Confidential Redshift data warehouse
* Creating AWS Lambda functions using python for deployment management in AWS and designed and implemented public facing websites on Amazon Web Services and integrated it with other applications infrastructure.
* Created ETL jobs using Informatica, Talend, IRIS tool to load the data into stage and target tables.
* Extensively worked on Azure Data Lake Analytics with the help of Azure Data bricks to implement SCD-1, SCD-2 approaches.
* Created Azure Stream Analytics Jobs to replication the real time data to load to Azure SQL Data warehouse;
* Implemented delta logic extractions for various sources with the help of control table; implemented the Data Frameworks to handle the deadlocks, recovery, logging the data of pipelines.
* Creating different AWS Lambda functions and API Gateways, to submit data via API Gateway that is accessible via Lambda function.
* Created ETL Pipeline using Spark and Hive for ingest data from multiple sources.
* Deployed the codes to multiple environments with the help of CI/CD process and worked on code defect during the SIT and UAT testing and provide supports to data loads for testing; Implemented reusable components to reduce manual interventions
* Created Snowpipe for continuous data load. Load the data from Azure blob storage
* Developing Spark (Scala) notebooks to transform and partition the data and organize files in ADLS
* Working on Azure Data bricks to run Spark-Python Notebooks through ADF pipelines.
* Using Data bricks utilities called widgets to pass parameters on run time from ADF to Data bricks.
* Created Triggers, PowerShell scripts and the parameter JSON files for the deployments
* Worked with VSTS for the CI/CD Implementation
* Reviewing individual work on ingesting data into azure data lake and provide feedbacks based on reference architecture, naming conventions, guidelines and best practices
* Implemented End-End logging frameworks for Data factory pipelines.

**Hess Corporation, Houston, TX**

**Jan 2017 – May 2019**

**Data Engineer**

**Responsibilities:**

* Analyze, design and build Modern data solutions using Azure PaaS service to support visualization of data. Understand current Production state of application and determine the impact of new implementation on existing business processes.
* 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 In Azure Databricks.
* 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.
* 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.
* Responsible for estimating the cluster size, monitoring and troubleshooting of the Spark databricks cluster.
* Experienced in 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 Scala and Pyspark.
* Developed JSON Scripts for deploying the Pipeline in Azure Data Factory (ADF) that process the data using the SQL Activity.
* Implemented OLAP multi-dimensional cube functionality using AzureSQL Data Warehouse.
* Hands-on experience on developing SQL Scripts for automation purpose.
* Created Build and Release for multiple projects (modules) in production environment using Visual Studio Team Services (VSTS).
* Wrote AzurePowershellscripts to copy or move data from local file system to HDFS Blob storage.
* Worked extensively with Dimensional modeling, Data migration, Data cleansing, ETL Processes for data warehouses.
* Worked in Agile Methodology and used JIRA for maintain the stories about project.
* Involved in gathering the requirements, designing, development and testing.

**Environment:** Hadoop, Azure Data Factory, Azure Data Lake, Azure Storage, Azure SQL, Azure DataWarehouse, Azure Databricks, Azure Power Shell, Map Reduce, Hive, Spark, Python, Yarn, Tableau, Kafka, Sqoop, Scala, HBase.

**Ramco Systems, India**

**Nov 2013 – Aug 2016**

**Data Engineer**

**Responsibilities:**

* Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.
* 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).
* Design & implement migration strategies for traditional systems on Azure (Lift and shift/Azure Migrate, other third-party tools.
* Engage with business users to gather requirements, design visualizations and provide training to use self-service BI tools.
* Used various sources to pull data into Power BI such as SQL Server, Excel, Oracle, SQL Azure etc.
* Propose architectures considering cost/spend in Azure and develop recommendations to right-size data infrastructure.
* Develop conceptual solutions & create proof-of-concepts to demonstrate viability of solutions.
* Technically guide projects through to completion within target timeframes.
* Collaborate with application architects and DevOps.
* Identify and implement best practices, tools and standards.
* Design Setup maintain Administrator the Azure SQL Database, Azure Analysis Service, Azure SQL Data warehouse, Azure Data Factory, Azure SQL Data warehouse.
* Build Complex distributed systems involving huge amount data handling, collecting metrics building data pipeline, and Analytics.