**AARTHI**

**Email:** [**vaarthi617@gmail.com**](mailto:vaarthi617@gmail.com) **PH: 980-666-8991**

**Azure Data Engineer**

**Professional Summary**

* Overall 8+ Years of experience in Software industry and around 4+ years as Azure data engineer.
* Experience in building data pipelines using Azure Data factory, Azure databricks and loading data to Azure Data Lake, Azure SQL Database, Azure SQL Data warehouse and controlling and granting database access.
* 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.
* Good understanding of Spark Architecture including Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors and Tasks.
* Good understanding of 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) .
* 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.
* Strong skills in visualization tools Power BI, Confidential Excel - formulas, Pivot Tables, Charts and DAX Commands.
* Experience in analyzing data using HiveQL, and MapReduce Programs.
* Experienced in ingesting data into HDFS from various Relational databases like MYSQL, Oracle, DB2, Teradata, Postgres using sqoop.
* Experienced in importing real time streaming logs and aggregating the data to HDFS using Kafka and Flume.
* 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.
* Experience in Developing Spark applications using Spark - SQL, Pyspark and Delta Lake 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.
* Good understanding of Spark Architecture, MPP Architecture, including Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors and Tasks.
* 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.
* Experienced and Implemented SAN TR migrations like Host based and Array based migrations.
* Hands on Experience in performing Host based online SAN migrations.
* Deep Knowledge in ServiceNow tool, Creating New change requests.
* Working as Cloud Administrator on Microsoft Azure, involved in configuring virtual machines, storage accounts, resource groups.
* Good understanding of Spark Architecture including Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors and Tasks.
* Experience with MS SQL Server Integration Services (SSIS), T-SQL skills, stored procedures, triggers.
* Design and develop 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.
* Good understanding of 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).
* Azure Data Factory (ADF), Integration Run Time (IR), File System Data Ingestion, Relational Data Ingestion.
* Respond rapidly to system maintenance needs, including on evenings and weekends.
* Generate incident reports, change reports, turnovers Summary report weekly basis.

**Professional Experience**

**Azure Data Engineer**

**Molina healthcare, Bothell, WA Jan 2021 to Present**

**Responsibilities:**

* Develop, design data models, data structures and ETL jobs for data acquisition and manipulation purposes.
* Develop deep understanding of the data sources, implement data standards, maintain data quality and master data management.
* Expert in developing JSON Scripts for deploying the Pipeline in Azure Data Factory (ADF) that process the data.
* Expert in using Databricks with Azure Data Factory (ADF) to compute large volumes of data.
* Performed ETL operations in Azure Databricks by connecting to different relational database source systems using jdbc connectors.
* Developed Python scripts to do file validations in Databricks and automated the process using ADF.
* Developed an automated process in Azure cloud which can ingest data daily from web service and load in to Azure SQL DB.
* Developed Streaming pipelines using Azure Event Hubs and Stream Analytics to analyze data for dealer efficiency and open table counts for data coming in from IOT enabled poker and other pit tables.
* Analyzed data where it lives by Mounting Azure Data Lake and Blob to Databricks.
* Used Logic App to take decisional actions based on the workflow.
* Developed custom alerts using Azure Data Factory, SQLDB and Logic App.
* Developed Databricks ETL pipelines using notebooks, Spark Dataframes, SPARK SQL and python scripting.
* Used Python and Shell scripts to Automate Teradata ELT and Admin activities.
* Performed Application level DBA activities creating tables, indexes, and monitored and tuned Teradata BETQ scripts using Teradata Visual Explain utility.
* Performance tuning, monitoring, UNIX shell scripting, and physical and logical database design.
* Developed UNIX scripts to automate different tasks involved as part of loading process.
* Worked on Tableau software for the reporting needs.
* Worked on creating few Tableau dashboard reports, Heat map charts and supported numerous dashboards, pie charts and heat map charts that were built on Teradata database.
* Implement Copy activity, Custom Azure Data Factory Pipeline Activities.
* Design and implement database solutions in Azure SQL Data Warehouse, Azure SQL.
* Collaborate with application architects on infrastructure as a service (IaaS) applications to Platform as a Service (PaaS).

**Environment**: Azure Data Factory, Tableau, Shell Scripting, Teradata, python scripting, Azure data bricks, Azure data lake storage, Blob storage, Azure SQL Database, Azure synapse analytics, Azure synapse workspace, Synapse sql pool,

**Azure Data Engineer**

**Chewy Dania Beach, FL Mar 2018 to Dec 2020**

**Responsibilities:**

* Understand requirements, build codes, and guide other developers in the course of development activities in order to develop high standard stable codes within the limits of Confidential and client processes, standards and guidelines.
* Develop Informatica mappings to be implemented based on client requirements and for the analytics team.
* Perform end to end system integration testing
* Involve in functional testing and regression testing
* Review and write sql scripts to verify data from source systems to target
* Worked on transformations to transform the data required by analytics team for visualization and business decisions.
* Review plan and provide feedback on gaps, timeline and execution feasibility etc. as required in the project
* Participate in KT sessions conducted by customer/ other business teams and provide feedback on requirements
* Involved in migrating the client data warehouse architecture from on-premises into Azure cloud.
* Create pipelines in ADF using linked services to extract, transform and load data from multiple sources like Azure SQL, Blob storage and Azure SQL Data warehouse.
* Creating storage accounts which involved with end to end environment for running jobs.
* Implement Azure Data Factory operations and deployment into Azure for moving data from on-premise into cloud.
* Design data auditing and data masking for security purpose.
* Monitoring end to end integration using Azure monitor.
* Implementation of data movements from on-premises to cloud in Azure.
* Develop batch processing solutions by using Data Factory and Azure Data bricks
* Implement Azure Data bricks clusters, notebooks, jobs and auto scaling.
* Design for data auditing and data masking
* Design for data encryption for data at rest and in transit
* Design relational and non-relational data stores on Azure
* Preparing ETL test strategy, designs and test plans to execute test cases for ETL and BI systems.
* Creating ETL test scenarios and test cases and plans to execute test cases.
* Interacting with business users and understanding their requirements.
* Good understanding of data warehouse concepts.
* Good exposure and understanding of Hadoop Ecosystem
* Proficient in SQL and other relational databases.
* Good exposure to Microsoft Power BI.
* Good understanding and working knowledge of Python language

**Environment**: SQL Database, Azure data factory, Azure data lake storage, Azure synapse analytics, Azure synapse workspace, Synapse sql pool, Power BI, Python

**BigData Engineer**

**Ascena Retail Group, Patskala, Ohio Aug 2015 to Feb 2018**

**Responsibilities:**

* Worked on analyzing Hadoop cluster and different big data analytical and processing tools including Sqoop, Hive, Spark, Kafka and Pyspark.
* Worked on MapR platform team for performance tuning of hive and spark jobs of all users.
* Using Hive TEZ engine to increase the performance of the applications.
* Working on incidents created by users for platform team on hive and spark issues by monitoring Hive and Spark logs and fixing it or else by raising MapR cases.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.
* Tested the cluster Performance using Cassandra-stress tool to measure and improve the Read/Writes.
* Worked on Hadoop Data Lake for ingesting data from different sources such as Oracle and Teradata through INFOWORKS ingestion tool.
* Worked on ARCADIA for creating analytical views on top of tables as if the batch is loading also no issue in reporting or table locks as it will point to arcadia view.
* Worked on Python API for converting assigned group level permissions to table level permission using MapR ace by creating a unique role and assigning through EDNA UI.
* Queried and analyzed data from Cassandra for quick searching, sorting and grouping through CQL.
* Migrating various Hive UDF's and queries into Spark SQL for faster requests.
* Configured to receive real time data from the ApacheKafka and store the stream data to HDFS using Kafka connect.
* Hands on experience in Spark using Scala and Python creating RDD's, applying operations -Transformation and Actions.
* Extensively perform complex data transformations in Spark using Scala language.
* Involved in converting Hive/SQL queries into Spark transformations using Scala.
* Used Pyspark and Scala languages to process the data.
* Used Bitbuket and Git repositories.
* Used text, AVRO, ORC and Parquet file formats for Hive tables.
* Experienced Scheduling jobs using Crontab.
* Used Sqoop to import data from Oracle, Teradata to Hadoop.
* Created Master Job Sequences for integration, (ETL\_Control) logic to capture job success, failure, error and audit, information for reporting.
* Used TES Scheduler engine to manage interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Java map-reduce Hive, Spark, Kafka and Sqoop.
* Experienced in creating recursive and replicated joins in hive.
* Experienced in developing scripts for doing transformations using Scala.
* Written Map Reduce code to process and parsing the data from various sources and storing parsed data into HBase and Hive using HBase-Hive Integration.
* Experienced in creating the shell scripts and made jobs automated.

**Environment**: HDFS, Hadoop, Python, Hive, Sqoop, Flume, Spark, Map Reduce, Scala, Oozie, YARN, Tableau, Spark-SQL, Spark-MLlib, Impala, Nagios, UNIX Shell Scripting, Zookeeper, Kafka, Agile Methodology, SBT.

**Hadoop Engineer**

**Maisa Solutions Private Limited Hyderabad, India Nov 2014 to April 2015**

**Responsibilities:**

* Involved in importing the real time data to Hadoop using Kafka and implemented the Oozie job for daily data.
* Loaded the data from Teradata to HDFS using Teradata Hadoop connectors.
* Import the data from different sources like HDFS/HBase into Spark RDD
* Developed Spark scripts by using Python shell commands as per the requirement
* Issued SQL queries via Impala to process the data stored in HDFS and HBase.
* Used the Spark - Cassandra Connector to load data to and from Cassandra.
* Used Restful Web Services API to connect with the MapRtable. The connection to Database was developed through restful web services API.
* Involved in developing Hive DDLs to create, alter and drop Hive tables and storm, & Kafka.
* Worked with application teams to install operating system, Hadoop updates, patches, version upgrades as required.
* Experience in data migration from RDBMS to Cassandra. Created data-models for customer data using the Cassandra Query Language.
* Responsible for building scalable distributed data solutions using Hadoop cluster environment with Horton works distribution
* Involved in developing Spark scripts for data analysis in both Python and Scala. Designed and developed various modules of the application with J2EE design architecture.
* Implemented modules using Core Java APIs, Java collection and integrating the modules.
* Experienced in transferring data from different data sources into HDFS systems using Kafka producers, consumers and Kafka brokers
* Installed Kibana using salt scripts and build custom dashboards that can visualize aspects of important data stored by Elastic search.
* Used File System Check (FSCK) to check the health of files in HDFS and used Sqoop to import data from SQL server to Cassandra
* Streaming the transactional data to Cassandra using Spark Streaming/Kafka
* Implemented a distributed messaging queue to integrate with Cassandra using Apache Kafka and Zookeeper.
* Written ConfigMap and Daemon set files to install File beats on Kubernetes PODS to send the log files to Log stash or Elastic search to monitor the different type of logs in Kibana.
* Created Database on Influx DB also worked on Interface, created for Kafka also checked the measurements on Databases.
* Installed Kafka manager for consumer lags and for monitoring Kafka Metrics also this has been used for adding topics, Partitions etc.
* Successfully Generated consumer group lags from Kafka using their API.
* Ran Log aggregations, website Activity tracking and commit log for distributed system using Apache Kafka.
* Involved in creating Hive tables, and loading and analyzing data using hive queries.
* Developed multiple MapReduce jobs in java for data cleaning and pre-processing. Loading data from different source (database & files) into Hive using Talend tool.
* Used Oozie and Zookeeper operational services for coordinating cluster and scheduling workflows.
* Implemented Flume, Spark, and Spark Streaming framework for real time data processing.

**Environment**: Hadoop, Python, HDFS, Hive, Scala, MapReduce, Agile, Cassandra, Kafka, Storm, AWS, YARN, Spark, ETL, Teradata, NoSQL, Oozie, Java, Cassandra, Talend, LINUX, Kibana, Hbase

**Data Modeler**

**IBing Software Solutions Private Limited Hyd India July 2013 to Oct 2014**

**Responsibilities:**

* Actively involved in the design phase of the development, conducted detailed data modeling for the database to be developed.
* Ensured that business requirements can be translated into data requirements
* Designed and build logical and physical data models for OLTP OLAP systems.
* Developed, tested, managed and documented data warehouse operations and tasks.
* Assisted in the design the overall ETL processes and architecture.
* Performed complex data analysis in support of ad-hoc requests.
* Analyzes data to identify root-cause of issues related to data warehouse and ETL processes and provided troubleshooting expertise in resolving technical issues.
* Developed processes that define, support and enforce information and data quality standards.

**Environment**: Erwin, Oracle 10g, SQL Server 2005, Business Object Data Integrator 6.X, VB Script, PL/SQL, Microsoft Project/Office.