**HARISH**

**Azure Data Engineer**

**G-Mail:** **sujal@nextcoderit.com**

**LinkedIn url:** **www.linkedin.com/in/harish-r-433035249**

**Ph: 6783919250**

**PROFESSIONAL SUMMARY**

Having around 8 years of IT experience, and 4 years in complete software development life Cycle Azure Data Engineering.

* Working experience Pipelines in ADF using Linked Services/Datasets/Pipeline to Extract and load data from different sources like On-premises MSSQL, Azure SQL, ADLS Gen2, Blob storage, Azure SQL Data warehouse.
* Working with Azure Data Factory Control flow transformations such as For Each, Lookup Activity, Until Activity, Web Activity, Wait Activity, If Condition Activity.
* Implemented multi tables full load from on-premises to cloud.
* Implemented multi tables incremental load from on-premises to cloud.
* Implemented Audit in Azure Data Factory using stored procedure.
* Hands-on experience on Integration Runtime
* Implemented the scheduling in Azure Data Factory using Triggers.
* Implemented Dynamic Data loading using a Config Table using components Lookup, Foreach and Copy Data activities for the tables which are configured in the Config. Table.
* Hands on experience on Key Vaults
* Experience in creating the Azure blob, Data Lake Storage.
* Implemented alert notification in ADF using Logical App.
* Good understanding of Spark Architecture including Spark Core, Spark SQL, Data Frames, Driver Node, Worker Node, Stages, Executors and Tasks.
* Working experience in different connectors in Azure databricks like creating mount point to connect to data lake storage, Blob Storage.
* Experience in Developing Spark applications using Spark - SQL/PYSPARK in Databricks for ETL from multiple file (Parquet, CSV file) formats for analyzing & transforming the data to customer usage patterns.
* Hands on experience on creating parameters in the databricks notebooks.
* Hands on experience on reading and writing the data from data lake into databricks
* Hands on experience on implementing different types of joins in pyspark and sal
* Implemented delta lake concepts in databricks
* Hands on experience in implementing SCD1, SCD2 using delta lake.
* Hands on experience on implementing data validations using pyspark
* Hands on experience on scheduling databricks notebooks using databricks jobs.
* Hands on experience in implementing audit logs in every notebook.
* Hands on experience in implementing the exception handling and logging in every databricks notebook.
* Excellent troubleshooting skills with strong technical background and good interpersonal
* Quick learner and excellent team player, ability to meet deadlines and work under pressure.

**TECHNICAL SKILLS:**

|  |  |  |
| --- | --- | --- |
| **Hadoop Ecosystem** |  | HDFS, MapReduce, Hive, Sqoop, Spark, Yarn, Strom, Zookeeper, Oozie, Flume |
| **Programming Languages** |  | Python, PySpark, T-SQL |
| **Big Data Platforms** |  | Cloudera |
| **Operating Systems** |  | Linux, Windows, UNIX |
| **Databases** |  | MySQL, HBase, MongoDB, Cassandra |
| **Development Methods** |  | Agile/Scrum, Waterfall |
| **IDE’s** |  | PyCharm, IntelliJ, Ambari |
| **Data Visualization** |  | Tableau, BO Reports, Splunk |
| **Azure Cloud Technologies** |  | Azure Data Factory, Azure Data Lake, Azure Data Bricks, Azure SQL Database, Azure Synapse Analytics, Active Directory, Application Insights, Azure Monitoring, LogicApps, Key Vault, Azure functions. |

**EDUCATION**

* Bachelor’s in computer science engineering, from JNTU, India, 2016.
* Master’s in computer science from Southern New Hampshire University, Aug 2022.

**PROFESSIONAL EXPERIENCE**

**Rockefeller Capital Management, Hamilton, NJ Sep 2022 to present**

**Azure Data Engineer**

**Responsibilities:**

* Attending client meetings to understand the requirements and business.
* Attending the DSM (Daily Standup Meetings) to update the task status.
* Installed Self hosted IR in on premise machine and created the connection from ADF.
* Created Azure Data Factory V2 Linked Services to connect to different sources like on premise SQL server, blob storage and SQL Datawarehouse.
* Created dynamic pipelines to copy multiple folders using dataset parameters.
* Created the azure data factory V2 pipelines for ingesting data from on premise file system to ADLS Gen2 as full load.
* Created the azure data factory V2 pipelines for ingesting data from on premise file system to ADLS Gen2 as Incremental load.
* Created triggers in azure data factory V2 to run pipelines on scheduled time.
* Created the stored procedure to update the audit logs of the ADF V2 copy activity.
* Created Azure Data Factory V2 Linked Services to connect to Azure Databricks.
* Created the azure data factory V2 pipelines to invoke the Azure databricks notebooks.
* Created the stored procedure to update the audit logs of the ADF V2 copy activity.
* Implemented alert notification in ADF using Logical App
* Created the mount point to connect the data lake Storage and Blob Storage to databricks.
* Implementing data validations using pyspark.
* Implemented audit framework in databricks notebooks to identify missing records.
* Developing Spark applications using Spark - SQL/PYSPARK in Databricks for ETL from multiple file (Parquet, CSV, Json file) formats for analyzing & transforming the data to customer usage patterns.
* Implementing different types of joins in pyspark and sql
* Implementing delta lake concepts in databricks.
* Implementing SCD1, SCD2 using delta lake.
* To scheduling the databricks notebooks. using databricks jobs.

**Environment:** Python, PySpark, Spark SQL MS SQL Server, Azure Data Lake Service (ADLS), Databricks, Azure Data Factory, Logic Apps, Azure Functions

**Virtusa, Chennai, Tamil Nadu Mar 2019 – Jul 2021**

**Role: Data Engineer**

**Responsibilities:**

* Worked as data engineer using Azure Databricks, azure data warehouse, Data Lake Gen1, Azure Data

Factory, Gen2, Azure Analytics services. Performed Process improvement (PI) that helped clients and businesses to have hard and soft saving and improvisations.

* Extract Transform and Load data from Sources Systems to Azure Data Storage services using a

combination of Azure Data Factory, 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 Azure

Data bricks.

* Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform and load data.

from different sources

* Participated in requirements meetings and data mapping sessions to understand business needs.
* Identified and documented detailed business requirement and handled team to get work upon same
* Connected various data sources both on premise and in cloud configuration. i.e spark, azure, SQL,

excel, CSV, Json etc

* Connected various data sources both on premise and in cloud configuration. i.e spark, azure, SQL,

excel, CSV, Json etc

* Used azure data services, ADLS,azure SQL, SAAS, data modeling for report creation.
* Integrated Custom Visuals based on business requirements using Power BI desktop.
* Development/creation of pipelines to get data from different source systems into azure data lake.

environment and to process and perform ETL transformation and get into our common data model.

using Azure Databricks, Spark, ADLS with different file system and format

* Hands on Experience on Azure Databricks script writing to perform ETL transformation on data using spark, python, and load into SQL, Cosmos, or application usage with various life systems.
* Worked on MS SQL studio to write queries, stored procedure, views for data development.
* Worked on development of workflow-based application using power App and power automate.

**Environment:** Azure Devops, Azure Data Factory, Powerapps, AzureDataBricks, ADLS, Pyspark, SparkSQL SQL Server

**Virtusa, Chennai, Tamil Nadu Aug 2017 – Feb 2019**

**Role: Hadoop Developer**

**Responsibilities:**

* Involved in loading and transforming large sets of structured data from router location to EDW using a NIFI data pipeline flow.
* Developed PySpark code and Spark-SQL for faster testing and processing of data.
* Worked on Data serialization formats for converting complex objects into sequence bits by using Parquet, ORC, AVRO, JSON, and CSV formats.
* Created Hive tables to load large data sets of structured data coming from data sources after the transformation of raw data.
* Created reports for the BI team to export data into HDFS and Hive.
* Developed custom NIFI processors for parsing the data from XML to JSON format and filtering broken files.
* Created Hive queries to spot trends by comparing fresh data with EDW reference tables and historical metrics.
* Used Pyspark to convert Panda’s data frame to Spark Data frame.
* Worked on partitioning Hive tables and running scripts parallel to reduce the run time of the scripts.
* Extensively worked on creating an End-to-End data pipeline orchestration using NIFI.
* Implemented business logic by writing UDFs in Spark Scala and configuring CRON Jobs.
* Provided design recommendations and resolved technical problems.
* Assisted with data capacity planning and node forecasting.
* Involved in performance tuning and troubleshooting the Hadoop cluster.
* Developed H-catalog Streaming code to stream the JSON data into Hive (EDW) continuously.
* Administrated Hive, and Kafka installing updates, patches, and upgrades.
* Supported code/design analysis, strategy development, and project planning.
* Managed and reviewed Hadoop log files.
* Evaluated the suitability of Hadoop and its ecosystem to project and implemented various proof of concept applications to eventually adopt them to benefit from the Hadoop initiative.

**Environment:** Spark, Python, Hive, Hadoop, GitHub, Splunk, Sqoop, Flume, Pyspark.

**Influx Info Solutions, Hyderabad, India Jun 2015 – Jul 2017**

**SQL Server Developer**

**Responsibilities:**

* Involved in the Creation of tables, indexes, sequences, and constraints and created stored procedures and triggers that were used to implement business rules.
* Installation of SQL Server on Development and Production Servers, setting up databases, users, roles, and permissions.
* Extensively involved in SQL joins, sub-queries, tracing, and performance tuning for better running of queries
* Provided documentation about database/data warehouse structures and updated functional specifications and technical design documents.
* Designed and created different ETL packages using SSIS and transferred data from heterogeneous databases in different file formats Oracle, SQL Server, and Flat File to SQL server destination.
* Worked on several transformations in Data Flow including Derived column, Slowly Changing dimensions using SSIS Controls, Lookup, Fuzzy Lookup, Data Conversion, Conditional split, and many more.
* Created various reports with drilldowns, drill-through, calculated members, and drill-downs reports by using SQL Server Reporting Services
* Used various report items like tables, sub-reports, and charts to develop the reports in SSRS and upload them to the Report Manager
* Created complex Stored Procedures, Triggers, Functions, Indexes, Tables, Views, SQL joins, and other T- SQL code to implement business rules.
* Used Performance Monitor and SQL Profiler to optimize queries and enhance the performance of database servers.

**Environment:** MS SQL Server 2012/2008R2/2008, T- SQL, SQL Server Reporting Services (SSRS), SSIS, SSAS, Business Intelligence Development Studio (BIDS), MS Excel, Visual Source Team Foundation Server, VB Script