



Abhinav

Mobile: **832 476 1948**

DATA ENGINEER | | AZURE | | BIG DATA

Email: basha@americanvassal.com

Summary

- Around 8 years of work experience in Development and Implementations of Data Warehousing solutions.
- Experienced in Azure Data Factory and preparing CI/CD scripts for the deployment.
- 4+ years of development experience on cloud platforms (Azure).
- Solid experience on building ETL ingestion flows using Azure Data Factory.
- Experience in building Azure stream Analytics ingestion spec for data ingestion which helps users to get sub second results in Realtime.
- Experience in building ETL (Azure Data Bricks) data pipelines leveraging PySpark, Spark SQL.
- Extensively worked on Azure Databricks.
- Experience in building the Orchestration on Azure Data Factory for scheduling purposes.
- Experience working with Azure Logic APP Integration tool.
- Experience working with Data warehouse like Teradata, Oracle, SAP.
- Experience on Implementation of Azure log analytics providing Platform as a service for SD-WAN firewall logs.
- Experience in building the data pipeline by leveraging the Azure Data Factory.
- Selecting appropriate low cost driven AWS/Azure services to design and deploy an application based on given requirements.
- Expertise on working with databases like Azure SQL DB, Azure SQL DW.
- Solid programming experience on working with Python, Scala.
- Experience working in a cross-functional AGILE Scrum team.
- Happy to work with the team who are in middle of the road with some Big Data challenges for both on prem and cloud.
- Hands-on experience in Azure Analytics Services – Azure Data Lake Store (ADLS), Azure Data Lake Analytics (ADLA), Azure SQL DW, Azure Data Factory (ADF), Azure Data Bricks (ADB) etc.
- Orchestrated data integration pipelines in ADF using various Activities like Get Metadata, Lookup, For Each, Wait, Execute Pipeline, Set Variable, Filter, until, etc.
- Have knowledge on Basic Admin activities related to ADF like providing access to ADLS using service principle, install IR, created services like ADLS, logic apps and Cortana analytics.
- Good knowledge on Polybase external tables in SQL DW.
- Involved in production support activities.
- Experience on major components in Hadoop Ecosystem like Hadoop Map Reduce, HDFS, HIVE, PIG, Hbase, Zookeeper, Oozie and Flume.
- Expertise in setting up processes for Hadoop based application design and implementation.
- Experience in importing and exporting data using Sqoop from HDFS to Relational Database Systems and vice-versa.
- Experience in managing and reviewing Hadoop log files.
- Experienced in processing Big data on the Apache Hadoop framework using MapReduce programs.
- Excellent understanding and knowledge of NOSQL databases like HBase and Mongo DB.
- Profound understanding of Partitions and Bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance.
- Good understanding on Amazon Web Services (AWS).
- Proficiency in SQL across several dialects (we commonly write MySQL, PostgreSQL, Redshift, SQL Server, and Oracle).
- Extensively worked with Teradata utilities Fast export, and Multi Load to export and load data to/from different source systems including flat files.
- Experienced in building Automation Regressing Scripts for validation of ETL process between multiple databases like Oracle, SQL Server, Hive, and Mongo DB using Python.



Technical Skills:

Operating system	Linux, Windows, Ubuntu, Unix
Big Data Tools	Hadoop Ecosystem: Map Reduce, Spark 2.3, Airflow 1.10.8, Nifi 2, HBase 1.2, Hive 2.3, Pig 0.17 Sqoop 1.4, Kafka 1.0.1, Oozie 4.3, Hadoop 3.0
Database Tools	Nosql, MangoDB, Teradata, Oracle, SQL AZURE DW, Oracle, MySQL, SQL Server, PostgreSQL, HBase, Snowflake, Cassandra, MongoDB
Methodologies	System Development Life Cycle (SDLC), Agile
Scripting Languages	Python, Scala, shell
Azure Cloud Stack	Azure Data Factory, Azure data bricks, Gen2 storage, Blob Storage, Event Hub, Log analytics, Sentinel analysis, Cosmos DB, ADLA, ADLS.
Languages	Scala, R, Python, C, C++, Java, Spark

Professional Experience:

PNC BANK, Birmingham, AL
Role: AZURE DATA ENGINEER

May 2017 – Current

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.
- 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.
- 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.
- Understand the latest features like (Azure DevOps, OMS, NSG Rules, etc...) introduced by Microsoft Azure and utilized it for existing business applications
- Worked on migration of data from On-prem SQL server to Cloud databases (Azure Synapse Analytics (DW) & Azure SQL DB).
- 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
- 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 tables ,Datawarehouse and Databases in snowflake environments.
- Worked with the snowflake team on creating snow pipeline for data flow.
- 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.

Environment: Azure Data Factory, Azure Data Bricks,PolyBase,AzureDW,ADLS,AzureDevOps,BLOB, Azure SQL Server, Azure DW, Azure devops , Azure synapse

CENTENE CORPORATION
StLouis MO
Role: DATA ENGINEER

Jun 16 – April 17

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 clients 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
- Using HP quality center to store and maintain test repositories.
- 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.
- Implementing AAD to specific user roles.
- Deploying ADLS accounts and SQL Databases.
- 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

Environment: Azure Data Bricks,PolyBase,AzureDW,ADLS,AzureDevOps,BLOB, Azure DW, Azure devops.

ACI World wide
Omaha, Nebraska
Role: BIG DATA DEVELOPER

Jan 15 – May 16

RESPONSIBILITIES:

- Developed Spark applications using Scala 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.
- Expertise in creating HDInsight cluster and Storage Account with End-to-End environment for running the jobs.
- Processed data into HDFS by developing solutions, analyzed the data using MapReduce, Pig, Hive and produce summary results from Hadoop to downstream systems.
- Used Kettle widely in order to import data from various systems/sources like MySQL into HDFS.
- Did various performance optimizations like using distributed cache for small datasets, Partition, Bucketing in hive and Map Side joins.
- Involved in creating Hive tables, and then applied HiveQL on those tables for data validation.
- Used Zookeeper for various types of centralized configurations.
- Deep understanding of schedulers, workload management, availability, scalability and distributed data platforms.
- Installed and configured Hadoop MapReduce, HDFS, developed multiple MapReduce jobs in Java for data cleaning and pre-processing.
- Involved in loading data from UNIX file system to HDFS.
- Wrote MapReduce jobs to discover trends in data usage by users.
- Involved in managing and reviewing Hadoop log files.
- Involved in running Hadoop streaming jobs to process terabytes of text data.
- Developed HIVE queries for the analysts.
- Implemented Partitioning, Dynamic Partitions, Buckets in HIVE.
- Exported the result set from HIVE to MySQL using Shell scripts.
- Used Git for version control.
- Maintain System integrity of all sub-components primarily HDFS, MR, HBase, and Flume .
- Monitor System health and logs and respond accordingly to any warning or failure conditions.
- Developing Spark (Scala) notebooks to transform and partition the data and organize files in ADLS
- 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.

Environment: Hadoop, HDFS, Map Reduce, Hive, Pig, Sqoop, Java 1.6, UNIX Shell Scripting.

VALUE LABS
Hyderabad India.
Role: SQL Developer

Apr 12 – Oct 13

RESPONSIBILITIES:

- Database development experience with Microsoft SQL Server in OLTP/OLAP environments using integration services (SSIS) for ETL (Extraction, Transformation and Loading). Developed SSIS packages that fetches files from FTP, did transformation on those data base on business need before I loaded to destination.
- Creating Metadata tables to log the activity of the packages, errors and change of the variable. Various techniques such as CDC, SCD and Hash bytes were used to capture the change of the data and execute incremental loading of the dimension tables



- Responsible for Deploying, Scheduling Jobs, Alerting and Maintaining SSIS packages.
- Implementing and managing Event Handlers, Package Configurations, Logging, System and User-defined Variables, CheckPoints and Expressions for SSIS Packages.
- Automating process by creating jobs and error reporting using Alerts, SQL Mail Agent, FTP and SMTP.
- Developed, tested, and deployed all the SSIS packages using project deployment model in the 2016 environment by configuring the DEV, TEST and PROD Environments.
- Created SQL Server Agent Jobs for all the migrated packages in SQL Server 2016 to run as they were running in the 2014 version.
- Created shared dimension tables, measures, hierarchies, levels, cubes and aggregations on MS OLAP/ Analysis Server (SSAS).
- Involved in creating a virtual machine on Azure, installed SQL 2014, created and administered databases then loaded data for mobile application purposes using SSIS from another virtual machine.
- Designed SSAS cube to hold summary data for Target dashboards. Developed Cube with Star Schema.
- Explore data in a Variety of ways and across multiple visualizations using SSRS.
- Responsible for creating SQL datasets for SSRS and Ad-hoc Reports.
- Expert on creating multiple kinds of SSRS Reports and Dashboards.
- Created, Maintained & scheduled various reports.
- Experienced in creating multiple kinds of reports to present story Points.
- Experience in writing reports based on the statistical analysis of the data from various time frame and divisions.

Environment: T-SQL, MS Office, Visual Studio , SQL Server Management Studio, SSIS, SSRS, SSAS, MS Azure, Microsoft Excel Services.

Education :

- Bachelors of technology in Jawaharlal Nehru Technological (2009 – 2013)
- Masters in computer science at Rivier University (2014).