**Professional Summary**

11+ years of IT experience which includes cross-functional and technical experience in handling large-scale Data warehouse delivery assignments in the role of Azure data engineer and ETL developer.

**Core Qualifications**

* Experience in developing data integration solutions in Microsoft Azure Cloud Platform using services Azure Data Factory [ADF], Azure Synapse Analytics, Azure SQL Database [ADB], Azure Blob Storage, Azure Data Lake Storage [ADLS] and Azure DevOps.
* Hands on experience troubleshooting, configuring Azure services like Azure Data factory, Azure Data Lake
* Experience in creating various datasets in ADF using linked services to connect to different source and target systems like SQL Server, Oracle, Azure Blob Storage, Azure Data Lake Storage and Azure Synapse Analytics and Azure SQL DB.
* Developed DataBricks ETL pipelines using notebooks with data frames and pyspark.
* Implemented various parameterized Azure Data Factory pipelines using activities like Copy activity, Custom Azure Data Factory Pipeline Activities.
* Self-hosted integration runtime has been created to copy the files from On-premises VM using various activities like metadata, foreach and copy and loaded into ADLS GEN2 and azure synapse analytics.
* Developed the pipelines in Azure Data factory for various scenarios to meet business requirement using blob storages and ingesting the data into azure synapse analytics.
* Expertise in writing complex SQL queries, Joins, Stored procedures using Azure Synapse analytics, SQL Server, and Oracle.
* Created Azure key vault for storing connection string details, certificates and used the key vaults in Azure Data factory while creating linked services.
* Experience in Extraction, Transformation, loading (ETL) of data from various sources into Data Warehouses and Data Marts using Informatica PowerCenter Tools (Repository Manager, Designer, Workflow manager and Workflow Monitor)
* Experienced in Power BI, Power Pivot, and complex DAX queries.
* Practical understanding of the Data modeling (Dimensional & Relational) concepts like Star-Schema Modeling, Fact and Dimension tables.
* Expertise in Database design, entity Normalization & database creation, maintenance, and monitoring.
* Extensive database programming experience in writing T-SQL, User Defined Functions, Triggers, Views, Temporary Tables Constraints, and Indexes using various DDL and DML commands.
* Experienced in development of a framework for common and consistent use of the intended design for batch and real time streaming of data ingestion, data processing, Predictive analysis, and delivery of massive datasets.
* Strong experience in data Injection, data storage and data processing using Hadoop Ecosystem tools like Sqoop, Hive, Pig, Spark, MapReduce, Spark Streaming, MapReduce, Flume, Kafka, HBase, Oozie, Zookeeper, and HDFS.
* Good experience in Scala, Python, Unix scripts and writing UDF’s including SQL, JSON, XML.
* Excellent experience in importing and exporting data using Sqoop from RDBMS to HDFS and vice-versa.
* Well-versed in and implemented Partitioning, Dynamic-Partitioning and bucketing concepts in Hive to compute data metrics.
* Experience in data modeling, predictive analytics and develop best-practices within data integration, data analytics and operational solutions.
* I worked extensively with SQL Server, HBase, and MySql.
* Excellent understanding and hands on experience using NOSQL databases like Mongo DB and HBase.
* Strong working in global delivery model (Onsite-Offshore model) involving multiple vendors and cross functional engineering teams.
* Have good knowledge of Gaming and Telecom domains.
* Excellent oral and written communication skills and great team player.

**Technical Skills**

|  |  |
| --- | --- |
| **Data Integration Tools** | Azure Data Factory, Azure Data bricks, Talend |
| **Databases** | Azure Synapse Analytics, Azure SQL Database, Oracle, MS SQL Server |
| **Azure Storage Accounts** | Azure Blob Storage, Azure Data Lake Storage |
| **ETL Tools** | Power BI |
| **Code Migration** | Azure DevOps, GitHub |
| **Languages** | SQL, PL/SQL, Shell Scripting, Java/J2EE, Python, Pyspark, Scala, JSON, XML |
| **Big Data** | Hive, Pig, Sqoop, Oozie, HBase, Zookeeper, YARN, Kafka, Spark, Scala,Flume |
| **Version Control** | GIT, Perforce |
| **Platform** | Linux/Unix, Windows |
| **Agile Tools** | JIRA |

**Education**

* Master of science (M.sc) in Electronics and Physics from S.K University in 2007.

**Professional Experience**

**Cincinnati, OH Nov 2022- Till date**

**Azure Data Engineer**

**Responsibilities**

* Designed and developed data ingestion pipelines using ADF and processing layer using Databricks and notebooks with PySpark.
* Created ADF pipelines to extract data from on premises source systems to azure cloud data lake storage. Extensively worked on copy activities and implemented copy behaviors such as flatten hierarchy, preserve hierarchy and Merge hierarchy. Implemented Error Handling concept through copy activity.
* Extensively worked on Azure Data Lake Analytics with the help of Azure Data bricks to implement SCD-1, SCD-2 approaches.
* Designed dimensional data modeling for various inventories in datalake architecture via ELT process in databricks.
* Developed Spark notebooks to transform and partition the data and organize files in ADLS.
* Implemented end-to-end cloud architecture using Azure databricks, Azure synapse and power BI.
* Worked with client to understand the business requirements and designed the analytical trend and reports in Power BI.
* Utilizing Azure Databricks to monitor inventory levels and make data-driven inventory decisions.
* Developed ELT process using Azure databricks to pull data from various sources like GCP, csv, json, Microsoft SQL Server, and stored in delta files.
* Implementing data quality checks and validation processes to ensure data accuracy and consistency. Identifying and resolving data anomalies and discrepancies in pyspark.
* Integrating data from different retail systems, including point-of-sale (POS) systems, inventory management, and customer databases, to provide a unified view of the business.
* Designed incremental load using SCD type 2 and orchestrating pipelines using terraform.
* Writing and optimizing ETL (Extract, Transform, Load) processes in Databricks to cleanse, preprocess, and enrich raw retail data in raw, silver, and curated layer.
* Utilizing Azure Databricks to monitor inventory levels and make data-driven inventory decisions.
* Captured the whole process status using metric tables (Azure log monitoring).
* Keeping up to date with the latest advancements and best practices in Databricks and Big Data technologies to drive innovation and efficiency in retail data solutions.
* Generated month-wise and year-wise PowerBI reports for different inventories in retail.
* Demonstrated good understanding of Spark architecture including Spark Core, Spark SQL, DataFrame, Spark streaming, Driver Node, Stages, Executors, and Tasks.

**Environment:** Azure Data Factory, Azure Data bricks, Azure Synapse Analytics, Azure SQL Database, Data Lake, Power BI, Azure DevOps

**Las Vegas, USA Jan 2021 - oct 2022**

**Azure Data Engineer**

**Responsibilities**

* Engineered a re-useable Azure Data Factory based data pipeline infrastructure that transforms provisioned data to be available for consumption by Azure SQL Data warehouse and Azure SQL DB.
* Created ADF pipelines to extract data from on premises source systems to azure cloud data lake storage. Extensively worked on copy activities and implemented copy behaviors such as flatten hierarchy, preserve hierarchy and Merge hierarchy. Implemented Error Handling concept through copy activity.
* Extensively worked on Azure Data Lake Analytics with the help of Azure Data bricks to implement SCD-1, SCD-2 approaches.
* Developed Spark notebooks to transform and partition the data and organize files in ADLS.
* Worked on Azure Data bricks to run Spark-Python Notebooks through ADF pipelines.
* Worked on migration of data from On-prem SQL server to Cloud databases (Azure Synapse Analytics (DW) & Azure SQL DB).
* Exposure to Azure Data Factory activities such as Lookups, Stored procedures, if conditions, for each, Set Variable, Append Variable, Get Metadata, Filter etc.
* Created Linked Services for multiple source systems (i.e. Azure SQL Server, ADLS, BLOB, and Rest API).
* Implemented delta logic extractions for various sources with the help of control tables; implemented the Data Frameworks to handle the deadlocks, recovery, and logging the data of pipelines.
* Created a Power BI data model based on analysis of the end-user workflow data provided by the client.
* Imported data from SQL Server DB, Azure SQL DB to Power BI to generate reports. Deploy Power BI reports to Web service and develop Dashboards.
* Developed analysis reports and visualization using DAX functions like table function, aggregation function.
* Explore data in a variety of ways and across multiple visualizations using Power BI Desktop App.
* Configured the logic apps to handle email notification to the end users and key shareholders with the help of web services activity; created a dynamic pipeline to handle multiple source extraction 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.
* Created Azure Stream Analytics Jobs to replicate the real time data to load to Azure SQL Data warehouse.
* Deployed the codes to multiple environments with the help of CI/CD process and worked on code defects during the SIT and UAT testing and provide support to data loads for testing; Implemented reusable components to reduce manual interventions.
* Processed the Structured and semi structured files like JSON, XML using Spark and Databricks environments.
* Prepared the data models for Data Science and Machine Learning teams. Worked with the teams in setting up the environment to analyze the data using Pandas.
* Worked with VSTS for the CI/CD Implementation.
* Reviewing individual work on ingesting data into azure data lake and provide feedback 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, Azure Synapse Analytics, Azure SQL Database, Data Lake, Power BI, Azure DevOps

**Las Vegas, NV April 2018 – Dec 2020**

**Azure Data Engineer**

**Responsibilities:**

* 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.
* Ingested huge volume and variety of data from disparate source systems into Azure Data Lake Gen2 using Azure Data Factory V2 by using Azure Cluster services.
* Developed ADF Pipelines to load data from on prem to AZURE cloud Storage and databases.
* Extensively worked on Spark Context, Spark-SQL, RDD's Transformation, Actions and Data Frames.
* Created pipelines, data flows and complex data transformations and manipulations using ADF and PySpark with Databricks.
* Created, provisioned multiple Databricks clusters needed for batch and continuous streaming data processing and installed the required libraries for the clusters.
* Handle the requests for SQL objects, schedule, business logic changes and Ad-hoc queries from customer and analyzing and resolving data sync issues.
* 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.
* Developed custom ETL solutions, batch processing and real-time data ingestion pipeline to move data in and out of Hadoop using PySpark and shell scripting.
* Developed PySpark notebook to perform data cleaning and transformation on various tables.
* Created several Databricks Spark jobs with Pyspark to perform several tables to table operations.
* Created Linked service to land the data from SFTP location to Azure Data Lake.
* Perform ongoing monitoring, automation, and refinement of data engineering solutions.
* Experience in working on both agile and waterfall methods in a fast-paced manner.

**Environment:** Azure Data Factory, Azure Data bricks, Azure Synapse Analytics, Azure SQL Database, Data Lake, Power BI, Azure DevOps

**Las Vegas, NV Sep 2017 - Mar 2018**

**Power BI Developer**

**Responsibilities**:

* Worked on data warehouse based on business requirements to provide quality information for publishing Power BI and SSRS Reports for the business executives.
* Worked in creating Stored Procedures, Triggers, Functions, Indexes, Tables, and Views for applications.
* Designed and developed databases for OLTP and OLAP Applications.
* I also have experience in maintaining the reports in power bi services and sharing with another team members in the organization.
* Creating row level security for the visuals and had experience in scheduling to refresh the data.
* I do have experience in Power bi admin, like creating content packs, how many users are accessing the dashboard, how many users are using power bi Pro license.
* Used Power BI Power Pivot to develop data analysis prototype and used Power View and Power Map to visualize reports.
* Designed and developed Power BI dashboards, Reports and published it on Power BI sites.
* Developed the pipelines to bring data from disparate data sources, which included SQL server databases, store into Azure SQL DW.
* Implemented process to import Google Analytics data and Google AdWords data into Data warehouse system.
* Implemented ETL process to download files by email attachments and import data into Data warehouse system through SSIS packages.
* Optimized performance of various SQL scripts, stored procedures, and triggers by using embedded UDFs, CTEs and System stored procedures.
* Developed and published reports and dashboards using Power BI and written effective DAX formulas and expressions.
* Configuring Power BI Gateway for SSAS Live connection and Direct Query.
* Worked on several types of SSIS control tasks like WMI data task, MSMQ task, Web service task, Script task, Ftp task to upload and download files, worked with several types of connection managers like file connection manager, flat file connection manager, cache connection manager, ftp connection manager and data quality connection services manager. Also, used XML task to validate XML files and XSLT for generating html pages from XML.
* Extensive use of DAX (Data Analysis Expressions) functions for the Reports and for the Tabular Models.
* Created reports in Power BI preview portal utilizing the SSAS Tabular via Analysis connector.
* Developed and published reports and dashboards using Power BI and written effective DAX formulas and expressions.
* Generated matrix reports, drill down, drill through, sub reports, chart reports, multi parameterized reports and Dashboards.
* Designed SSRS reports with sub reports, expressions, and queries, and prepared Ad-Hoc reports through report builders.
* Implementing SSRS solutions, used dashboards and gauges to enhance for the business users.
* Worked on MDX Queries to pull the data from the SSAS Cubes for the reporting purposes.

**Environment:** Azure Synapse, Power BI

**Las Vegas, NV Jan 2017 - July 2017**

**Hadoop Spark Developer**

**Responsibilities:**

* Worked extensively with Sqoop for importing and exporting data from SQL Server.
* Implemented Preprocessing steps using Data Frames for Batch Processing
* Analyzing Data issues for the customers and fixing the issues
* Built the summary tables, implemented call prediction models: player gaming summary models with K-Means Cluster in production using Spark MiLib and Scala.
* Worked with Data scientist partner for Predictive analysis, implemented bonus recommendation Engine using Spark MiLib, and persisted the recommendation results in HBase.
* Bug fixing and QA support.

**Environment:** Hive, Pig, Sqoop, Oozie, HBase, Zookeeper, YARN, Kafka, Spark, Scala, flume

**India Sep 2014 - Oct 2016**

**Hadoop Developer**

**Responsibilities:**

* Worked extensively with Sqoop for importing and exporting data from SQL Server.
* Implemented Preprocessing steps using Data Frames for Batch Processing
* Built the summary tables, implemented call prediction models: player gaming summary models with K-Means Cluster in production using Spark MiLib and Scala.
* Worked with Data scientist partner for Predictive analysis, implemented bonus recommendation Engine using Spark MiLib, and persisted the recommendation results in HBase.
* Used Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time and persists into HBase.
* Developed data ingestion framework to acquire data from SQL Server, and error handling mechanism.
* Partnered with data scientists to perform data analysis, summary datasets and identify the call input predictors and machine learning algorithms using RStudio.
* Developed real time ingestion of System and free form remarks/messages using Kafka and Spark Streaming to make sure the events are available in customer’s activity timeline view in real-time.
* Coordinated with Hadoop admin on cluster job performance and security issues, and Hortonworks team to resolve the compatibility and version related issues of HDP, Hive, Spark, Oozie.
* Automated ingestion and prediction process using Oozie workflows, coordinators jobs and supported in running jobs on the cluster.

**Environment:** Hive, Pig, Sqoop, Oozie, HBase, Zookeeper, YARN, Kafka, Spark, Scala, flume

**India Sep 2012 - Oct 2014**

**Data Analyst**

**Responsibilities**:

* Developed stored procedures in MS SQL to fetch the data from different servers using FTP and processed these files to update the tables.
* Responsible for Designing Logical and Physical data modeling for various data sources on Confidential Redshift.
* Performed logical data modeling, physical Data modeling (including reverse engineering) using the Erwin Data modeling tool.
* Created dimensional model for the reporting system by identifying required dimensions and facts using Erwin.
* Designed and Developed ETL jobs to extract data from Salesforce replica and load it in data mart in Redshift.
* Involved in performance tuning, stored procedures, views, triggers, cursors, pivot, unpivot functions, CTE’s.
* Developed and delivered dynamic reporting solutions using SSRS.
* Extensively used Erwin for Data modeling. Created Staging and Target Models for the Enterprise Data Warehouse.
* Involved in Normalization / De normalization techniques for optimum performance in relational and dimensional database environments.
* Resolved the data type inconsistencies between the source systems and the target system using the Mapping Documents and analyzing the database using SQL queries.
* Worked on ETL testing and used SSIS tester automated tool for unit and integration testing.
* Designed and created SSIS/ETL framework from the ground up.
* Created new Tables, Sequences, Views, Procedure, Cursors and Triggers for database development.
* Created ETL Pipeline using Spark and Hive for ingest data from multiple sources.
* Involved in using SAP and transactions done in SAP - SD Module for handling customers of the client and generating the sales reports.
* Creating reports using SQL Reporting Services (SSRS) for customized and ad-hoc Queries
* Coordinated with clients directly to get data from different databases.

**Environment:** Google Analytics · Process Improvement · Tableau · Python (Programming Language) · SQL