**Aamir Khan**

**9452459232**

**Professional Summary:**

* Over all 10+ Years of Information Technology experience, specifically in the areas of Database Development, ETL and Azure Data Engineer.
* In depth knowledge on Microsoft Azure ETL tools Azure data factory, AZ copy, Adlcopy & Azure CLI etc.
* Currently working on Azure Data Lake environment using Azure Resources Azure Blob Storage & Data Lake Store, Azure Data Lake Analytics & Azure Data bricks.
* Proficient knowledge on Python and PowerShell scripting
* Sound knowledge & expertise in Retail, CPG domain which includes Sales, Digital Marketing Manufacturing and POS data and having good exposure at analyzing Sales and Manufacturing related data flow.
* Expertise in using Cloud based managed services for data warehousing/analytics in Microsoft Azure (Azure Data Lake Analytics, Azure Data Lake Storage, Azure Data Factory, Azure Table Storage, Stream Analytics, HDInsight, DATABRICKS etc.)
* Hands on exp on Unified Data Analytics with Databricks, Databricks Workspace User Interface, Managing Databricks Notebooks, AWS S3, Delta Lake with Python, Neo4j, Delta Lake with Spark SQL
* Worked on various stages of soft development life cycle including Requirements, analysis, design, coding, development, and testing.
* Good understanding of Bigdata Hadoop and Yarn architecture along with various Hadoop Demons such as Job Tracker, Task Tracker, Name Node, Data Node, Resource/Cluster Manager and Kafka.
* Hands on experience in application development using Java, RDBMS (SQL), and Linux/Unix shell scripting.
* Expert in Agile/SCRUM model with onsite-offshore model.
* Excellent knowledge on C# programming to develop business objects and develop solutions as per the business requirement.
* Adept in Python, Scala, Java and Shell scripting, with significant experience in UNIX/Linux environments.
* Experience in creating SSIS Packages for integrating data using OLE DB connection from heterogeneous sources (Excel, CSV, Oracle, flat file, Text Format Data) by using multiple transformations provided by SSIS such as Data Conversion, Conditional Split, Bulk Insert, merge and union all.
* Exp in Developing Spark applications using Spark – SQL in Databricks for data.
* Developed SSRS reports and knowledge on rendering reports, managing subscriptions, creating reports with Report builder, report snapshot and report cache.
* Designed and Developed Spark work on Aws using Scala for data pull from AWS S3 bucket and Snowflake applying transformations on it.
* Very Good experience in working with clients & meeting their requirements and willing to adapt to new technologies/work environments, well organized Analytical & Leadership skills. Excellent communication, detail oriented, time bound and responsible team player.

**Technical Skills:**

**Operating System** : WINDOWS XP, WINDOWS SERVER 2000/2003/2008, WINDOWS 7, WINDOWS 10

**Environment** : Client/Server

**Database**  : SQL Server 2005/2008/2012/2014/2016, Oracle 9I/10G/11G, Teradata,

**ETL & Reporting Tools** : SSMS, SSIS, VS 2010/2012/2013/2016,

**Tools**  : SQL DEVELOPER 2.1/1.2, VSS, TFS, Azure Data Factory & Azure Storage Explorer

**Languages** : JAVA, C, SQL, PL/SQL, R programming, Python, Power shell, GIT, JSON

**Other** : Data Modeling, UML, XML

**Functional:**

**Management and Leadership Skills:**

Project Management, Business Analysis and Development, Excellent Communication and Interpretation skills Experienced in environments requiring direct Customer interaction during specifications, development, and project implementation phases.

**Training & Development:** ANSI SQL, PL/SQL, C#, USQL, PYTHON, Windows Azure Cloud Development.

**Education:**

Bachelor’s in computer science in JNTU Kakinada – 2011

Master’s in computer science in University of Central Missouri - 2016

**Certifications:**

* Certified in Databricks
* Azure Certification 70-775 Perform Data Engineering on Microsoft Azure HDInsight.
* Edx Certified Azure Datalake store/Analytics
* Edx Certified Azure Data Factory
* Edx certified Azure Cosmos DB
* Apache Spark and Scala Advance Course Certification

**Professional Experience:**

**Client: Microsoft Corporation June 2023 – Present**

**Location: Charlotte NC**

**Role: Azure Data Engineer**

**Responsibilities:**

* Responsible for calculation of ~ $10 million incentives in each monthly cycle.
* Analyze and brainstorm business requirements, collaborate with cross-functional teams including program managers and business teams to simplify requirements.
* Knowledge of Azure Cloud Services (PaaS & IaaS), Storage, Data-Factory, Data Lake (ADLA & ADLS), Logic

Apps, Azure Monitoring, Active Directory, Synapse, Key Vault, and SQL Azure.

* Expertise in developing end-to-end ETL (Extract, Transform, Load) pipelines using GCP services
* Assemble large, complex data sets that meet business needs.
* As an Azure data engineer, I oversaw implementing the ETL process for loading data from various

sources into Databricks tables and Azure Synapse Tables.

* Developed U-SQL scripts to process large datasets on Azure Data Lake Analytics.
* Strong understanding of Logical and Physical database models and entity-relationship modelling.
* Debug and resolve the issues related to pipelines failures, pipelines getting stuck, pipelines not producing output, performance issues or any other deployment issues in Azure Data factory.
* Designed and Developed Spark work on using Scala for data pull from AWS S3 bucket and Snowflake applying transformations on it.
* Designed and implemented multiple dashboards for internal metrics using Azure Synapse – PowerPivot & Power Query tools.
* Developed Hive/Pig scripts to process the data using custom Java functions to convert data types. Created documentation for production team.
* Involved in loading data from the system generated data sources to HDFS and experienced in writing multiple java-based Map-Reduce jobs for cleaning, processing the data.
* Writing SQL stored procedures and queries and scheduling using ADF for Daily/Monthly run.
* Push data into Azure Data Lake Store using Azure Data factory by creating pipelines as per requirements on daily, monthly basis so that the data can be analyzed in the downstream systems.
* Making sure the pipelines that are built using Azure Data factory to process massive scale of data have all the best practices in the context of Data Warehousing world so that they get a stable system in production.
* Develop Azure Resource Management templates to move data factory from one resource group to another and thus decrease the manual work.
* Transferring the log incoming log files to the Parser (specially written Java Code to load in to HDFS, HBASE) by using Kafka message broker.
* Created PySpark notebooks to execute on Databricks.
* Created Power BI dashboard to view summarized earnings per month.
* Subject Matter Expert for Online Incentives and Azure Market Place programs in Partner Incentives project
* Follow scrum agile methodology and code development & continues integration using Azure Dev ops.

**Environment:** Windows and Mac OS X, Visual Studio tools for Azure Data Lake Analytics & Data Factory tools, Azure Power Shell ISE for Azure ARM template deployment, Spark, SSMS 2017 for Azure SQL Managed Instance, GIT, JSON, U-SQL, Python, Azure power shell, Azure .Net SDK & T-SQL.

**Client: Mindtree Feb 2019 – March 2022**

**Location: India**

**Role: Azure Data Engineer (Azure Developer)**

**Project Description:**

Altria & ALCS current direction is to adapt industry trending business model as insights driven organization. Part of that is to migrate data sources which include Sales & Digital Marketing data from Altria on-premises to Microsoft Azure cloud storage as flat files and transform the data sets to build enterprise analytics.

**Responsibilities:**

* Analyze the Sales & DMS data sets required to be moved to Azure cloud storage and document the data flow strategy for each data set.
* Ability to apply the spark Data Frame API to complete data manipulation with in spark session
* Developed Spark applications in Databricks using PySpark and Spark SQL to perform transformations

and aggregations on source data before loading it into Azure Synapse Analytics for reporting.

* Develop and maintain the data flow using Microsoft Azure SQL Server Utilities and BI Tools like SSIS & Bulk copy utility to extract data and generate flat files.
* Reading the Mainframe copy book and the mainframe files by Java language and migrating the data to
* Hadoop ecosystems with the metadata
* Develop Azure Logic Apps & Function Apps to move/copy files between Azure Data Lake Store & Azure Blob Storage
* Develop and Azure power shell scripts to export files from On-premises to Cloud storage using Azure power shell SDK.
* Closely work with offshore team and have daily communications as a subject expert and help them to achieve the business targets in terms of coding & unit testing.
* Performed Java MapReduce programs on log data to transform into structured way to find user location, age group, spending time
* Integrated Oozie with the rest of the Hadoop stack supporting several types of Hadoop jobs out of the box (such as Java map-reduce, Streaming map-reduce, Pig, Hive, Sqoop and Distcp) as well as system specific jobs (such as Java programs and shell scripts).
* We make heavy use of Spark to process and transform data.
* Execute integration testing & User acceptance testing in coordination with client based on the planned releases & deploy the deliverables into Production environment.
* Follow scrum agile methodology and code development & repository using Microsoft VSTS tools.

**Environment:** Windows and Mac OS X, Visual Studio tools for Azure Data Lake Analytics & Azure Data Factory, Azure Logic Apps & Function Apps, Azure Power Shell ISE, SQL Server Bulk copy utility, Spark, SQL Server Integration services & SQL Agent for job scheduling, GIT, JSON C#, T-SQL, Azure power shell & U-SQL.

**Client: IBM Jan 2014 – Jan 2019**

**Location: India**

**Role: Hadoop Developer**

**Responsibilities:**

* Responsible for calculation of ~ $10 million incentives in each monthly cycle.
* Planned, installed, and configured the distributed Hadoop Clusters.
* Ingested data using Sqoop to load data from MySQL to HDFS on regular basis from various sources.
* Configured Hadoop tools like Hive, Pig, HBase, Zookeeper, Flume, HBase, Impala and Sqoop.
* Built relational view of data using Catalog.
* Ingested data into HBase tables from MySQL, Pig and Hive using Sqoop.
* Used Tez for faster execution of MapReduce jobs.
* Performed a needs analysis and implemented ESRI GIS software at enterprise level.
* Setup workflow for mapping of geologic data from Acquire database using ESRI software.