**Ajith Kumar**

**Data Engineer**

**Email: ajkum941@gmail.com**

**Mobile: 4084951714**

**Professional Summary:**

* Overall 7+ years of experience as a Data Analyst and Data Engineer with analytical programming using Python and high proficiency in Big Data Analytics, Predictive Modelling, Text mining, and Machine Learning.
* Experience in data analysis, designing experiments, interpreting data behavior, business decision support.
* Experience in diverse set of procedures but not limited to methods such as machine learning, deep learning, Bayesian algorithm, Regressions, cluster analysis, decision trees, time series, resampling and regularization, NLP and other techniques.
* Strong working experience with various python libraries such as NumPy, SciPy for mathematical calculations
* Experience in SQL Queries and optimizing the queries in Oracle, SQL Server, MONGODB, PostgreSQL and Teradata.
* Experience in data transformations using Azure HDInsight, HIVE for different file formats.
* Experience in using Tableau, creating dash boards and quality story telling.
* Excellent visual representation of data and communicating analysis with Python, Tableau, to all levels of business users within the organization, automate analyses and build analytics data pipelines via SQL and python based ETL framework.
* Experience with common Machine Learning Algorithms like Regression, Classification and Ensemble models.
* Good knowledge in establishing classification and forecast models, automate processes, text mining, sentiment analysis, statistical models, risk analysis, platform integrations, optimization models, models to increase user experience, A/B testing using R, Python, tableau, etc.
* Strong familiarity in Build the data models by extracting and stitching data from various sources, integrated systems with R to cater efficient data analysis.
* Extensive experience working with AWS Cloud services and AWS SDKs to work with services like AWS API Gateway, Lambda, S3, IAM, and EC2.
* Proficient with Azure Data Lake Services (ADLS), Databricks & iPython Notebooks formats, Databricks Deltalakes & Amazon Web Services (AWS).
* Experience in installation, configuration, management, supporting, and monitoring Hadoop clusters using various distributions such as Apache Hadoop, Cloudera Hortonworks, and various cloud services like AWS, Azure
* 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 Developed complex database objects like Stored Procedures, Functions, Packages and Triggers using SQL and PL/SQL.
* Experience in installing, configuring, and maintaining the databases like PostgreSQL, Oracle, Big DataHDFS systems.
* Expertise in applying data mining techniques and optimization techniques and proficient in Machine Learning, Data/Text Mining, Statistical Analysis and Predictive Modeling.
* Proficient in Data Warehousing using Snowflake Schema, Data Mining concepts and ETL transformations from Source to target systems.
* Excellent Working knowledge of Hadoop, Hive, Sqoop, pig, HBase &amp; Oozie in real time environment and worked on many modules for performance improvements and architecture designing.
* Good working experience on Spark (park streaming, spark SQL), Scala and Python.
* Extensively worked on Spark using Python and Scala on cluster for computational (analytics), installed it on top of Hadoop performed advanced analytical application by making use of Spark with Hive and SQL/Oracle/Snowflake.
* Strong experience and knowledge of real time data analytics using Spark Streaming, Kafka and Flume.
* Good knowledge on Amazon’s Elastic Cloud Compute (EC2) infrastructure for computational tasks and Simple Storage Service (S3) as Storage mechanism.
* Good working experience on using Sqoop to import data into HDFS from RDBMS and vice-versa.

**Technical Skills:**

|  |  |
| --- | --- |
| **Programming languages** | SQL (Oracle, MySQL), SAS- Base/SAS-SQL/SAS-Macro/SAS-ODS, Java, UNIX shell, Python |
| **Database querying tools** | SQL Developer, MS SQL Server, DB2, Toad for Oracle, Snowflake |
| **Visualization** | Tableau, Business Objects, Microsoft Excel |
| **Web development** | HTML5, CSS3, JavaScript, XML, JSON, HTTP, SOAP, REST |
| **Application Development** | Eclipse, Visual Studio, SAS EG |
| **ETL tools** | IBM Data stage, MQ, Dimensions CM, CAWA, Tivoli, Oracle Scheduler |
| **Retail products** | ORMS, ORIB, RMS |
| **RPA** | UiPath, IDEA |
| **Development Tools** | Microsoft SQL Studio, IntelliJ, Eclipse, NetBeans. |
| **Operating Systems** | All versions of Windows, UNIX, LINUX |
| **NO SQL Databases** | HBase, MongoDB, MariaDB |

**Professional Experience:**

**Client: Delta Airlines, Atlanta, GA July 2021 - Present**

**Role: Azure Data Engineer**

**Responsibilities:**

* Analyzed, designed, and built modern data solutions using Azure PaaS services to facilitate data visualization. Assessed the current production state of applications and evaluated the impact of new implementations on existing business processes.
* Utilized Azure Databricks notebooks for transforming complex files that couldn't be handled efficiently in Azure Data Factory.
* Imported data from various sources such as HDFS/HBase into Spark RDD and developed data pipelines using Kafka and Storm for storing data into HDFS.
* Modeled key risk indicators in Splunk and constructed comprehensive Hive & SPL queries to analyze behavior across the customer lifecycle.
* Extracted, transformed, and loaded data from source systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL, and U-SQL in Azure Data Lake Analytics. Performed data ingestion into one or more Azure Services (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processed data in Azure Databricks.
* Utilized Azure services like ADLS and Synapse Analytics for small data sets.
* Employed Python scripts for updating content in databases and file manipulation.
* Utilized Sqoop for channeling data from various sources to HDFS and RDBMS.
* Created pipelines in ADF using Linked Services, Datasets, Pipeline activities to Extract, Transform, and Load data from different sources such as Azure SQL, Blob storage, Azure SQL Data Warehouse, and write-back tools.
* Developed Spark applications using PySpark and Spark SQL for data extraction, transformation, and aggregation from multiple file formats to analyze and transform data to reveal insights into customer usage patterns.
* Converted Hive/SQL queries into Spark transformations using Spark RDDs.
* Integrated Kafka with Spark Streaming for real-time data processing.
* Integrated Oozie with Hue and scheduled workflows for multiple Hive, Pig, and Spark jobs.
* Designed and deployed Hadoop clusters and various Big Data analytic tools including Pig, Hive, HBase, Oozie, Zookeeper, Sqoop, Flume, Spark, Impala, and Cassandra with Hortonworks Distribution.
* Estimated cluster sizes, monitored, and troubleshooted Spark Databricks clusters.
* Tuned Spark applications by setting appropriate batch interval times, levels of parallelism, and memory configurations.
* Documented requirements, including available code, to be implemented using Spark, Hive, HDFS, HBase, and Elasticsearch.
* Developed UDFs in Scala and PySpark to meet specific business requirements.
* Created JSON scripts for deploying pipelines in Azure Data Factory (ADF) that process data using SQL activities.
* Implemented build and release processes for multiple projects (modules) in the production environment using Visual Studio Team Services (VSTS).

**Environment:** Azure Data Factory(V2), Azure Databricks, Python, SSIS, Azure SQL,

**Client: Mindtree, Hyderabad, India Nov 2018 – Sept 2020**

**Role: Azure Data Engineer**

**Responsibilities:**

* Participated in requirement grooming meetings which involves understanding functional requirements from a business perspective and providing estimates to convert those requirements into software solutions.
* Design and Develop & Deliver the Code to IT/UAT/PROD and validate and manage data Pipelines from multiple applications with fast-paced Agile Development methodology using Sprints with JIRA Management Tool
* Responsible for checking data in DynamoDB tables and checking EC2 instances upon running for (DEV, QA, CERT, and PROD) in AWS.
* Analysis of existing data flows and Creating high-level/low-level technical design documents for business stakeholders that confirm technical design aligns with business requirements.
* Created and deployment of Spark jobs in different environments and loading data to no SQL database Cassandra/Hive/HDFS. Secure the data by implementing encryption-based
* Implemented AWS solutions using E2C, S3, RDS, EBS, Elastic Load Balancer, Auto scaling groups, Optimized volumes, and EC2 instances and created monitors, alarms, and notifications for EC2 hosts using Cloud Watch.
* Responsible for deployments to DEV, QA, PRE-PROD (CERT), and PROD using AWS.
* Created quick Filters Customized Calculations on SOQL for SFDC queries, Used Data loader for ad hoc data loads for Salesforce
* Extensively worked on Informatica power center Mappings, Mapping Parameters, Workflows, Variables, and Session Parameters.
* Responsible for facilitating load data pipelines and benchmarking the developed product with the set performance standards.
* Used a Debugger within the Mapping Designer to test the data flow between the source and target and to troubleshoot the invalid mappings.
* Worked on SQL tools like TOAD and SQL Developer to run SQL Queries, SSIS and validate the data.
* Study the existing system and conduct reviews to provide a unified review on jobs.
* Involved in Onsite & Offshore coordination to ensure the deliverables.
* Involved in testing the database using complex SQL scripts and handling the performance issues effectively.

**Environment:** Apache spark, Scala, Cassandra, HDFS, Hive, GitHub, Jenkins, Kafka, Informatica PowerCenter, SQL Server, SSIS, Salesforce Cloud, Visio, TOAD.

**Client: Narvee Technologies, India Jul 2016 – Oct 2018**

**Role: Data engineer**

**Responsibilities:**

* Created High & Low level design documents for the various modules. Review the design to ensure adherence to standards, templates and corporate guidelines.
* Coordinated with the Application support team and help them assist understand the business and necessary components for the Integration, Extraction, Transformation and load data.
* Analyzed and develop Data Integration templates to extract, cleanse, transform, integrate and load to data marts for user consumption. Review the code against standards and checklists.
* Created a Deployment document for the developed code and provide support during the code migration phase.
* Performed Analysis on the existing source systems, understand the Informatica/Teradata based applications and provide the services which are required for development & maintenance of the applications.
* Worked with Google Cloud(GCP) Services like Compute Engine, Cloud Functions, Cloud DNS, Cloud Storage and Cloud Deployment Manager and SaaS, PaaS and IaaS concepts of Cloud Computing and Implementation using GCP.
* Used enqueue/Init Lambda to trigger the EMR Process
* Create Initial Unit Test Plan to demonstrate that the software, scripts and databases developed conforms to the Design Document.
* Used python Boto 3 to configure the services AWS glue, EC2, S3.
* Develop framework for converting existing PowerCenter mappings and to PySpark(Python and Spark) Jobs.
* Provides support during the integration testing and User Acceptance phase of the project. Also provide hyper care support post deployment.
* Designed the tables and loaded the data to the Redshift databases using Amazon best practices.
* Fully in charge of site Cycle from design to Service verification to continuous Optimization.

**Environment:** Informatica power center, EMR-AWS, Optimization Expertise Power exchange, Oracle, GCP, Python, Business Objects, machine learning Tableau, Pyspark.