**AKM Arafin**

**Mail: arafinakm@gmail.com| LinkedIn : https://www.linkedin.com/in/akm-arafin-4a1b061ab/**

**Professional Summary:**

* Data Engineer having around 10+ years of experience in Information and Technology which includes Big data and Hadoop Ecosystem components.
* Solid hands-on experience in dealing with Apache Hadoop components like HDFS, Hive, Yarn, Sqoop, Zookeeper, Apache Spark.
* IT experience in all phases of Hadoop Development, ETL Development along with experience in Application Development & Data modelling through various roles over the years.
* Experienced working with Spark Streaming, Spark SQL and Kafka for real-time data processing.
* Extensive knowledge in programming with Resilient Distributed Datasets (RDDs) and Data Frames in Spark.
* Good knowledge on integration of Kafka Infra administration resource with Spark for real time data processing.
* Extensive experience in working with various distributions of Hadoop Enterprise versions of Cloudera(CDH5), Hortonworks and good knowledge on AWS services.
* Knowledge in end-to-end Hadoop Infrastructure including Hive, Sqoop, and Zookeeper.
* Managed data coming from different sources and involved in HDFS loading of structured and unstructured data.
* Experience in developing data pipeline using SPARK, and knowledge on Flume to extract the data from weblogs and store in HDFS.
* Design and develop ETL integration patterns using Python on Spark.
* Imported the data from various sources like Local file, Linux system into Spark RDD.
* Experience in importing and exporting the data using Sqoop from Relational Database to HDFS and reverse on Linux systems. Developed Spark SQL to load tables into HDFS.
* Worked on loading CSV/AVRO/PARQUET files using Scala/Java language in Spark Framework
* Expertise in writing Hadoop Jobs for analysing data using Hive QL (Queries).
* Developed various Shell Scripts to automate Spark jobs and hive scripts.
* Created User Defined Functions (UDF’s), User Defined Aggregated Functions (UDAF’s) in Hive.
* Experience in loading the data from the different Data sources like (Teradata,SQLserver) into HDFS using Sqoop, Spark and load into Hive tables, which are partitioned.
* Knowledge on Snowflake database.
* Extensively worked on MRV1 and MRV2 Hadoop architectures and wrote MapReduce programs, Pig & Hive scripts.
* Designed and created Hive external tables using shared meta-store instead of Derby with dynamic partitioning and bucketing.
* Experienced in importing and exporting data using Sqoop from Relational Database Systems to HDFS and vice-versa.
* Extensively used Kafka Infra administration resource to load the log data from multiple sources directly into HDFS. Knowledge on RabbitMQ. Loaded streaming log data from various webservers into HDFS using Flume.
* Experienced in building Pig scripts to extract, transform and load data onto HDFS for processing. Excellent knowledge of data mapping, extract, transform and load from different data source. Experience in writing HiveQL queries to store processed data into Hive tables for analysis. Extended Hive and Pig core functionality by writing custom UDFs.
* Excellent understanding and knowledge of NOSQL databases like HBase and Cassandra.
* Designed Databases, created and managed schemas, wrote stored procs, functions, DDL, DML, SQL queries and data modeling
* Extensive experience in ETL Architecture, Development, enhancement, maintenance, Production support, Data Modeling, Data profiling, Reporting including Business requirement, system requirement gathering.
* Hands-on experience in shell scripting. Knowledge on cloud services AWS and MS Azure.
* Proficient in using RDMS concepts with Oracle, SQL Server, MariaDB and MySQL.
* Experienced in Object Oriented Analysis, Design (OOAD) and development of software using UML Methodology, good knowledge of J2EE design patterns and Core Java design patterns.
* Experienced in project life cycle (design, development, testing and implementation) of Client Server and Web applications.
* Experience in processing different file formats like XML, JSON and sequence file formats.
* Designed, deployed, maintained and lead the implementation of Cloud solutions using MS Azure and underlying technologies Implemented HA deployment models with Azure Classic and Azure Resource Manager and configured Azure Active Directory and managed users and groups. Worked on Continuous Integration CI/Continuous Delivery (CD) pipeline for Azure Cloud Services using Chef. Migrating Services from On-premise to Azure Cloud Environments.
* Good Knowledge in Amazon AWS concepts like EMR and EC2 web services for fast and efficient processing of Big Data.
* Good Experience in creating Business Intelligence solutions and designing ETL workflows using Tableau. ⮚ Collaborate with development and QA teams to maintain high-quality deployment ⮚ Designed Client/Server telemetry adopting latest monitoring techniques.
* Configured Azure Traffic Manager to build routing for user traffic
* Infrastructure Migrations: Drive Operational efforts to migrate all legacy services to a fully Virtualized Infrastructure.
* Exhibited strong written and oral communication skills. Rapidly learn and adapt quickly to emerging new technologies and paradigms
* Performed systems analysis for several information systems documenting and identifying performance and administrative bottlenecks.
* Good understanding and extensive work experience on SQL and PL/SQL
* Experience in designing and implementing of secure Hadoop cluster using Kerberos.
* Monitor health of the platforms and Generate Performance Reports and Monitor and provide continuous improvements

**Technical Skills:**

|  |  |
| --- | --- |
| **Programming Languages:** | Java/J2EE, Java 8, SQL, PL/SQL, C, C++. |
| **Technologies & web Services:** | JSP2.3, JSF2.2, JDBC4.0, Servlets, JSP, EJB, JDBC 4.0, JPA, JMS, Hibernate (ORM) ESB Web Services (REST, SOAP) WSDL, UDDI, SOA, JAX-RS, AXIS. |
| **Frameworks:** | Struts2, CSS, Frameworks, GWT, spring 4(Dependency Injection, Spring MVC, Spring Access, core, Spring Context, Spring AOP, Spring DAO, Spring IOC, Spring JDBC, Spring with Hibernate), Hibernate 4.0, Micro services, Angular.JS, Node.js, Bootstrap, JSF2.2. JUnit, Log4J. |
| **Databases:** | Oracle (8i/9i/10g/11g), DB2, PL/SQL, My SQL, SQL Server, NoSQL, Microsoft SQL server, DB2, Mongo DB, Cassandra. |
| **Message Brokers:** | Rabbit MQ, Kafka. |
| **Application Servers:** | Web Logic 8.1/10.3, IBM Web sphere 5.1/6.0/7.0, Apache Tomcat 8, JBoss5.X, Glassfish, Jetty. |
| **Operating Systems:** | Windows XP/VISTA, UNIX/LINUX, MAC. |
| **Big Data:** | HDFS, Hive, HBase, Sqoop, Spark, MapReduce, Kafka |
| **Integrated Development Tools:** | Eclipse4.5, Net Beans, Red Hat Developer, TDD, Intelli J, Rational Application Developer. |
| **Application Tools:** | Eclipse, RAD, Maven, Jenkins, JIRA |
| **Testing Tools:** | JUnit 4.9, Selenium, Firebug, Mockito, Log4j. |
| **Web Technologies:** | HTML5, CSS3, Java Script, AJAX, XML, JSON, React JS, Node JS, Dom, Angular 2. |
| **Version Control:** | CVS, SVN, Rational clear case, GIT, Github. |
| **SDLC Methodologies:** | Agile, Waterfall and Scrum models, RAD.  |

**Professional Summary:**

**Comcast, Philadelphia, PA Oct 2022 – Present**

**Big Data Engineer**

**Roles & Responsibilities:**

* Used Spark Data Frames Operations to perform required Validations in the data and to perform analytics on the Hive data.
* Developed Apache Spark applications by using spark for data processing from various streaming sources.
* Implemented Spark using Java spark and SparkSQL for faster processing of data.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs.
* Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
* Utilized Global Relay, a leading provider of cloud-based message archiving, compliance, and supervision solutions.
* Used Spark SQL on data frames to access hive tables into spark for faster processing of data.
* Imported the data from different sources like Unix, Local file system into Hadoop.
* Create data pipelines that transform and enrich data to usable forms.
* Loading Data from the edge node to Hadoop using Spark framework.
* Extracted data from various data sources using Cosmos DB and/or Azure Data Explorer (Kusto) for data analytics and reporting purposes.
* Proficient in data modeling, data analytics, business intelligence, master data management, and data quality.
* Developed interactive and responsive user interfaces using React.js.
* Designed and implemented data warehousing solutions using Snowflake, leveraging its cloud-native architecture for scalability and performance
* Utilized CDP tools to build unified customer profiles for multi-dimensional segmentation
* Led the design, development, and maintenance of mission-critical web and server applications in a dynamic Agile environment, ensuring robustness, scalability, and high performance.
* Developed and maintained data processing pipelines using PySpark, enabling efficient ETL operations on large datasets
* Created and managed Snowflake databases and warehouses to support analytical workloads efficiently.
* Managed and optimized Microsoft SQL Server databases for various clients in diverse industries.
* Integrated Salesforce with popular applications such as Microsoft Office 365, Google Workspace, and marketing automation platforms.
* Migrated on-premises data infrastructure to Snowflake, reducing maintenance overhead and improving data accessibility.
* Set up, enabled analytics, and maintained diverse data sources.
* Worked closely with development teams in an embedded DBA/DBM role to provide database-related support and guidance.
* Collaborated with cross-functional teams to deploy and maintain web applications, databases, and services on Unix-based servers.
* Worked effectively in an Agile environment with Scrum Master/Product Owner, consistently delivering results.
* Led and contributed significantly to database development projects, utilizing strong PL/SQL skills to design, implement, and maintain efficient and robust database solutions.
* Utilized Python to create custom data analysis scripts, resulting in streamlined data exploration and insights extraction.
* Clean, transform and analyse raw data from various mediation sources using Spark to provide ready-to-use data.
* Develop data processing pipelines using JAVA Spark.
* Implemented Monte Carlo simulations to assess and mitigate project risks, enabling better project planning and resource allocation.
* Managed and deployed Microsoft SQL Server databases on the Amazon Web Services (AWS) cloud platform.
* Implemented data integration solutions using StreamSets Data Collector
* Registering, ingesting, and transforming the data and loading it to the Hadoop environment.
* Querying Hive table using SPARK SQL after data ingestion to check for data quality.
* Converting parquet Hadoop files to data frames for clients using spark-shell.
* Loading data into Teradata and its integration with the Hadoop cluster.
* Developed and maintained efficient data pipelines to ingest data into Cosmos DB or Kusto clusters, ensuring high data availability and reliability.
* Utilized Snowflake's virtual warehouses for resource allocation and scaling, optimizing query performance and cost-efficiency.
* Collaborated closely with cross-functional teams, including product owners, designers, and QA engineers, to gather requirements, prioritize tasks, and deliver high-quality technical solutions.
* Leveraged partitions, DBMS jobs, packages, procedures, functions, and Oracle user-defined types to optimize database performance and manage data effectively.
* Proficiently handled data lake, data marts, and data warehouse systems, showcasing strong expertise
* Conducted training sessions to empower non-technical team members in using Power BI for self-service analytics.
* Designed and optimized SQL queries to extract, transform, and load data from various sources into a unified database for reporting purposes.
* Comprehensive understanding of existing data source connectors, provided support, and extended connectors for enhanced data engineering capabilities.
* Played a crucial role in ensuring data privacy compliance, conducting regular audits and assessments to identify potential vulnerabilities.
* Provided Unix production support for critical applications, ensuring high availability and minimal downtime.
* Worked with Snowflake, a cloud-based data warehousing platform, to store and analyze large volumes of data.
* Conducted database migrations and upgrades to newer versions of MSSQL, ensuring minimal downtime and data loss.
* Worked on loading CSV, EBCDIC, ASCII, PARQUET files using JAVA in the SPARK framework.
* Developed various shell scripts to automate Spark jobs.
* Utilized Visual Studio Code (VS Code) for efficient coding, debugging, and version control, increasing team productivity by 90%.
* Developed ETL pipelines using Snowflake's native features and integrations with tools like Apache Spark and Talend.
* Validated control-m XML and scheduled the jobs using control-m.
* Collaborated with stakeholders to gather requirements and design creative UI/UX solutions that align with the organization's branding and user expectations.
* Developed test scripts to compare data in edge node and HDFS. Also took reports from Hadoop tenant path to provide to clients.
* Well-versed in using Containers like ECS and Docker, enhancing application deployment efficiency.
* Utilized AWS RDS (Relational Database Service) for SQL Server to automate database setup, backups, and scaling.
* Extensive experience in developing custom Apex code to enhance Salesforce functionality and automate business processes.
* Integrated StreamSets with different data platforms, such as Hadoop, Spark, and cloud services (e.g., AWS S3, Azure Blob Storage).
* Proficient in HTML, CSS, and JavaScript to translate design concepts into responsive and visually appealing web interfaces.
* Collaborated with legal and compliance teams to interpret and implement data privacy regulations, ensuring alignment with industry standards.
* Applied Agile development methodologies and Test-Driven Development (TDD) principles to streamline development processes, resulting in increased productivity and reduced time-to-market for features.
* Managed the integration of Electronic Communications data from various sources into a centralized system.
* Managed and configured Global Relay platform for efficient message archiving and retrieval.
* Creating JSON structures to match with source schemas and to load data into TERADATA and Hadoop.
* Preparing Design Documents (Request-Response Mapping Documents, Hive Mapping Documents).
* Experienced with batch processing of data sources using Apache Spark and Elastic search.
* Experienced in implementing Spark RDD transformations, actions to implement business analysis
* Migrated Hive QL queries on structured into Spark QL to improve performance
* Collaborated closely with UI teams and applications, effectively bridging the gap between backend and frontend development. Used cursors, XML, JSON, and other technologies to ensure seamless integration.
* Configured, deployed and maintained a single node storm cluster in DEV environment
* Developed solutions to pre-process large sets of structured, semi-structured data, with different file formats (Text file, Avro data files, Sequence files, Xml and JSon files, ORC and Parquet).
* Implemented AWS Security features like Virtual Private Cloud (VPC), Security Groups, and Network ACLs to secure database instances.
* Leveraged Snowflake's robust support for semi-structured data, enabling efficient processing of JSON, XML, and other formats.
* Utilized tools like Jira, Confluence, SAFe, and DevOps methodologies, contributing to efficient development.
* Leveraged expertise in backend technologies such as Java and J2EE to architect and implement scalable solutions, utilizing Tomcat for efficient application deployment.
* Handled importing of data from RDBMS into HDFS using Sqoop.
* Proficient in implementing and customizing Salesforce Experience Cloud (formerly known as Community Cloud) to build self-service portals for customers, partners, and employees.
* Collaborated with cross-functional teams to understand data requirements and design appropriate data models in Cosmos DB or Kusto to optimize data retrieval.
* Managed data across Raw, Analytical, and Process stores, leveraging Azure's capabilities.
* Developed Spark scripts to import large files from Amazon S3 buckets.
* Developed Spark core and Spark SQL scripts using Scala for faster data processing.
* Experienced in writing Hive Scripts for analyzing data in Hive warehouse using Hive Query Language
* Involved in creating Hive tables, loading with data, and writing hive queries to process the data.
* Developed Spark code and Spark-SQL/Streaming for faster testing and processing of data.
* Created scripts to automate the process of Data Ingestion.
* Designed and implemented data integration workflows to extract, transform, and load Electronic Communications data.
* Worked on configuring Azure services such as Azure App Service, Azure SQL Database, and Azure Functions for seamless application hosting and management.
* Designed and implemented data ingestion pipelines to load data from various sources into Snowflake.
* Integrated Apex with external systems and APIs to facilitate data exchange and process automation.
* Experience in using Testing Frameworks of Bigdata world, MRUnit, PIG Unit for testing raw data and executed performance scripts. Extract, transform, and load (ETL) data from multiple federated data sources (JSON, relational database, etc.) with Data Frames in Spark.
* Created interactive and visually appealing dashboards using Power BI, providing actionable insights to stakeholders and decision-makers.
* Supported Data Science capabilities within the Advanced Analytics Group.
* Utilized Snowflake's automatic clustering and partitioning to enhance data organization and retrieval efficiency.
* Developed RESTful APIs that facilitated seamless communication between different components of the application ecosystem, enhancing overall system flexibility and integration capabilities.
* Extend the capabilities of Data Frames using User Defined Functions in Python and Scala.
* Responsible for the design, development, testing, and support of the Big Data solutions on cloud (AWS) in collaboration with cross functional teams.
* Created custom connectors and processors using StreamSets SDK for handling specific data sources or formats.
* Defining, developing, and documenting standard requirements and methodologies
* Collaborate with key stake holders and translate business requirements to technical requirements and implement solutions under the guidance of technical leads.
* Managed and orchestrated big data processing tasks using Amazon EMR, utilizing frameworks like Hadoop and Spark for parallel processing.
* Collaborated with cloud teams to facilitate data loading in and out of the Cloudera Data Platform (CDP) from the Data Lake, ensuring smooth data flow and synchronization.
* Dive into large, noisy, and complex real-world customer TV and Digital Ad viewing data to do pre-campaign planning and post campaigns performance measurement using Big data Platform (DataBricks, Pyspark, Glue, EMR).
* Responsible for building data pipelines to ingest data, integrate data from multiple data sources (On-Premises & Cloud) and create aggregated data sets for reporting needs.
* Designs new software and web applications, supports applications under development and customizes current applications. Develops software update process for existing applications. Assists in the roll-out of software releases.
* Oversees the researching, writing, and editing of documentation and technical requirements, including evaluation plans, test results, technical manuals and formal recommendations and reports.
* Implemented compliance policies and procedures using Global Relay tools to ensure regulatory requirements were met.
* Configured and administered user accounts, access controls, and permissions using Unix file system security.
* Implemented fine-grained access controls and data masking within Snowflake, ensuring data security and compliance with regulations.
* Familiarity with Oracle Exadata, including its architecture and performance optimization strategies.
* Developed complex DAX (Data Analysis Expressions) calculations to derive meaningful metrics and KPIs from raw data.
* Created and customized Experience Cloud themes and templates to align with brand guidelines and provide a consistent user experience.
* Leveraged indexing and partitioning strategies in Cosmos DB or Kusto to enhance query performance and reduce data retrieval times.
* Implemented data partitioning and optimization techniques in Glue and EMR workflows, resulting in faster data processing and reduced costs.
* Integrated DataBricks with cloud services like Azure, AWS, or GCP for scalable data storage and computation
* Keeps current with technological developments within the industry. Monitors and evaluates competitive applications and products. Reviews literature, patents and current practices relevant to the solution of assigned projects.
* Works with Quality Assurance team to determine if applications fit specification and technical requirements.
* Displays expertise in process design and redesign skills. Presents and defends architectural, design and technical choices to internal audiences.
* Managed user accounts, permissions, and access control using Unix file system permissions and access control lists (ACLs).
* Developed and delivered data privacy training sessions to educate employees on best practices and policies.
* Created executive reports and dashboards using Power BI Reporting.
* Deployed and managed web applications (Apache, Nginx) on Unix servers, handling configuration, virtual hosts, and SSL certificates.
* Configured Snowflake's data sharing capabilities to securely share data with external partners and stakeholders.
* Researches, writes, and edits documentation and technical requirements, including software designs, evaluation plans, test results, technical manuals and formal recommendations and reports.
* Knowledge of and ability with process design and redesign skills

**Environment:** Java, HTML5, Angular 8+ higher, jQuery, JavaScript, Spring Batch, Node.js, Bootstrap, Spring 4.0, Apache 7.0, log4j, Hibernate, Maven, Tortoise SVN, Spring Boot, Rest web services, Docker, AWS, Microservices, maven, jQuery, JIRA, Eclipse Kepler, XML, MySQL.

**Bank of America AUG 2021 – SEP 2022**

**Hadoop Developer**

**Roles & Responsibilities:**

* Worked on client AML KYC applications, migrating data from Teradata to Hive for OMNI project applications. In this application which provides derailed insights about the usage of coll abortion tools and dashboards for the meetings.
* Application analyzed data SOR in mainframe and Teradata warehouse.
* Converted Teradata to Hive DDL through BB analysis.
* Designed and implemented a comprehensive ETL process for a new data source, resulting in accurate and timely data integration into the existing system.
* Demonstrated a strong proficiency in setting up API integrations and connectors.
* Created hive compatible DDL and Database project.
* Stayed up-to-date with industry trends and best practices to continuously enhance the creativity and innovation of UI/UX designs.
* Played a key role in developing end-to-end Extract, Transform, Load (ETL) processes, ensuring data accuracy and reliability throughout the data pipeline.
* Utilized AWS Glue for automated ETL processes, enabling efficient data extraction, transformation, and loading from various sources.
* Designed and implemented Snowflake data warehousing solutions, optimizing data storage and query performance.
* Monitored database performance using AWS CloudWatch and implemented auto-scaling to handle varying workloads effectively.
* Familiarity with Camunda workflow automation platform, leveraging it to streamline business processes and reduce manual intervention.
* Create design document to migrate data into Hadoop(Hive).
* Utilized Apex test classes and unit testing best practices to ensure code reliability and minimize deployment issues.
* Worked on BTEQ, MLOAD, TPT, T2H code conversion to load the data from Teradata to Hive Staging/temp table using pySpark and then load the data into Hive target tables using HiveQL and Shell script.
* Experiencing performing YTD, QTD and MTD analysis.
* Worked on DataBricks' collaborative features, facilitating seamless teamwork and version control.
* Creating Autosys Box and command job.
* Worked with AWS Database Migration Service (DMS) to migrate on-premises SQL Server databases to AWS with minimal downtime.
* Acquired a basic understanding of MDM solutions, CDP platforms, and digital marketing concepts.
* Contributed to the adoption of DevOps practices by integrating tools like Artifactory, SonarQube, Selenium, and Jenkins into the development pipeline, enabling automated testing and continuous integration.
* Migrated on-premises data to Snowflake, leading to improved data accessibility and reduced maintenance overhead.
* Performed complex queries and analytics using Snowflake's SQL-based querying language and integrated external tools like Tableau or Power BI.
* Attained AEP Developer, showcasing expertise in Adobe Experience Platform (AEP) development and proficiency in building omni-channel personalization implementations.
* Good in implementing advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in pyspark.
* Developed customized reports and dashboards using Global Relay analytics for management and compliance purposes.
* Used Spark for interactive queries, processing of streaming data and integration with popular NoSQL databases for huge volumes of data.
* Provided documentation and training to end-users on accessing and utilizing integrated Electronic Communications data.
* Collaborated with cross-functional teams to design and implement Camunda-based solutions, contributing to improved operational workflows.
* Proficient in Apex best practices, governor limits, and bulkification to optimize code performance and maintain platform efficiency.
* Demonstrated expertise in SQL and PL/SQL tuning, identifying and addressing performance bottlenecks to enhance application speed and responsiveness.
* Implemented and supported Adobe Experience Cloud products, including Adobe Analytics, Adobe Target, Adobe Launch, and Adobe MarketoEngage, driving data-driven decision-making and personalized customer interactions.
* Integrated Cosmos DB or Kusto with other tools and services, such as Power BI, Azure Data Factory, or custom applications, to enable seamless data visualization and insights.
* Conducted performance testing and tuning of StreamSets pipelines to ensure scalability.
* Developed Spark scripts by using Scala shell commands as per the requirement.
* Enhanced and optimized product Spark code to aggregate, group and run data mining tasks using the Spark framework.
* Used Spark RDD for faster Data sharing.
* Experienced in querying data using SparkSQL on top of Spark engine for faster data sets processing.
* Worked on Ad hoc queries, Indexing, Replication, Load balancing, Aggregation in MongoDB
* Extracted and restructured the data into MongoDB using import and export command line utility tool.
* Worked on the large-scale Hadoop YARN cluster for distributed data processing and analysis using Hive and MongoDB.
* Experience in workflow Scheduler Oozie to manage and schedule job on Hadoop cluster for generating reports on Day and weekly basis.
* Developed automated scripts and workflows to schedule data refreshes and updates, ensuring that the data in Cosmos DB or Kusto remains up-to-date and accurate.
* Designed and implemented ETL processes using DataBricks to transform raw data into valuable insights.
* Monitored and tracked message communications through Global Relay to identify and resolve any issues or violations.
* Responsible for analysing and cleansing raw data by performing Hive/Impala queries and running Pig scripts on data.
* Involved in moving log files generated from various sources to HDFS for further processing through Flume.
* Worked on partitioning the HIVE table and running the scripts in parallel to reduce the run time of the scripts.
* Programmed pig scripts with complex joins like replicated and skewed to achieve better performance.
* Developed data pipeline using SPARK for loading data into HDFS.
* Worked closely with legal and compliance teams to respond to regulatory inquiries and eDiscovery requests using Global Relay.
* Utilized monitoring and logging tools such as CloudWatch, Nagios, Splunk, and Dynatrace to ensure real-time system health and performance monitoring, proactively addressing potential issues.
* Utilized Python for various tasks, such as automating routine processes and interacting with databases. Developed scripts using libraries like requests, json, and pandas to streamline data manipulation and reporting.
* Stayed updated with the latest features and enhancements in StreamSets to leverage new functionalities in the data integration process.
* Migrated data from MySQL server to Hadoop using Sqoop for processing data.
* Exported the analysed data to the relational databases using Sqoop for visualization and to generate reports for the BI team.
* Configured and managed DataBricks clusters to ensure high availability and performance.
* Experienced in developing Shell scripts and Python scripts for system management.
* Worked with application teams to install operating system, Hadoop updates, patches, version upgrades as required.
* Worked with the SCRUM team in delivering agreed user stories on time for every Sprint.

**Environment:** Java/J2EE, spring, JPA, Java Script, Angular Js, jQuery, Hibernate, SQL DB, Oracle or MySQL, Ant or Maven, JUNIT, Firebug, Mongo DB, Cloud Services.

**Morgan Stanley, New York, NY SEP 2019 – JULY 2021**

**Hadoop developer**

**Roles & Responsibilities:**

* Working on developing architecture document and proper guidelines
* Involved in all phases of Software Development Life Cycle (SDLC) and Worked on all activities related to the development, implementation and support for Hadoop.
* Installed and Configured Apache Hadoop clusters for application development and Hadoop tools like Hive, Pig, HBase, Zookeeper and Sqoop.
* Adding/installation of new components and removal of them through Cloudera Manager
* Played a key role in installation and configuration of the various Hadoop ecosystem tools such as Solr, Kafka Infra administration resource, Pig, HBase and Cassandra.
* Configured roles, profiles, and sharing settings to control access and permissions within the Experience Cloud community.
* Implemented multiple Map Reduce Jobs in java for data cleansing and pre-processing.
* Wrote complex Hive queries and UDFs in Scala and Python.
* Involved in implementing an HDInsight version 3.3 clusters, which is based on spark version 1.5.1.
* Responsible for data extraction and data ingestion from different data sources into Hadoop Data Lake by creating ETL pipelines using Pig, and Hive
* Integrated Experience Cloud with Salesforce data and external systems to provide a seamless user experience and access to relevant information.
* Worked with systems engineering team to plan and deploy new Hadoop environments and expand existing Hadoop clusters and Experience in converting MapReduce applications to Spark.
* Job duties involved the design, development of various modules in Hadoop Big Data Platform and processing data using Map Reduce, Hive, Pig, Sqoop and Oozie.
* Design developed and tested Map Reduce programs on Mobile Offers Redemptions and Send it to the downstream applications like HAVI.
* Installed Oozie workflow engine to run multiple Map Reduce, Hive HQL and Pig jobs.
* Worked with Control-M workflow.
* Developed HDFS with huge amounts of data using Apache Kafka.
* Collected the log data from web servers and integrated into HDFS using Flume.
* Created and maintained Technical documentation for launching HADOOP Clusters and for executing Hive queries and Pig Scripts and Experience in managing and reviewing Hadoop log files
* Constructed System components and developed server-side part using Java, EJB, and Spring Frame work. Involved in designing the data model for the system.
* Used J2EE design patterns like DAO, MODEL, Service Locator, MVC and Business Delegate.
* Worked with cloud services like Amazon Web Services (AWS) and involved in ETL, Data Integration and Migration
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDD, Scala and Python.
* Implemented best income logic using Pig scripts and UDFs.
* Monitored and optimized DataBricks jobs and clusters for cost efficiency and resource utilization.
* Component unit testing using Azure Emulator Analyze escalated incidences within the Azure SQL database
* Moderate and contribute to the support forums (specific to Azure Networking, Azure Virtual Machines, Azure Active Directory, Azure Storage) for Microsoft Developers Network including Partners and MVPs.
* Built a prototype Azure Data Lake application that accesses 3rd party data services via Web Services. The solution dynamically scales, automatically adding/removing cloud-based compute, storage and network resources based upon changing workloads.
* Worked with Azure ExpressRoute to create private connections between Azure datacenters and infrastructure for on premises and in colocation environment.
* Worked with XML files, to configure the .NET core application as well as PowerShell code.
* Proactively worked as the complete Software Development Life Cycle including Requirement Analysis, Design, Implementation, Testing and Maintenance
* Experience in deployment of Hadoop Ecosystems like Map Reduce, Yarn, Sqoop, Flume, Pyspak,Pig, Hive, Hbase, SPARK, SCALA, Cassandra, Zoo Keeper, Storm, Impala, Kafka Infra administration resource
* Worked on data processing and transformations and actions in spark by using Python (Pyspark) language.
* Expertise in developing data driven applications using Python 2.7, Python 3.0 on Pycharm and Anaconda Spyder IDE's.
* Worked on Hadoop, Hive, JAVA, python, Scala Struts web framework.

**Environment:** Oracle/SQL Server and PL/SQL, spring, Hibernate, Ant, Apache, Tomcat, JBOSS, Web logic, UNIX, RDBMS, HTML, CSS, Java Script, JDBC, Eclipse, Multi-threading.

**RBC, Jersey City, New Jersey May 2017 – AUG 2019**

**Hadoop Admin**

**Roles & Responsibilities:**

* Extensively involved in the Design phase and delivered Design documents. Experience in Hadoop eco system with HDFS, Hive, Pig and Sqoop
* Successfully optimized a critical PL/SQL process, reducing execution time by 40% and significantly improving overall system performance.
* Migrated existing **SQL queries to HiveQL** queries to move to a big **data analytical** platform.
* Involved in moving all log files generated from various sources to HDFS for further processing through **Flume**.
* Created **Hive** Tables, loaded values and generated adhoc-reports using the table data.
* Extending HIVE and PIG core functionality by using custom UDF’s.
* Managing and scheduling Jobs on Hadoop cluster using Oozie.
* Involved in writing MapReduce programs for analytics and also for structuring the data coming from flume sinks.
* Written Java program to retrieve data from HDFS and provide REST services.
* Integrated multiple sources of data (SQL Server, DB2, MySQL) into Hadoop cluster and analysed data by Hive-HBase integration.
* Conducted training sessions and workshops to upskill team members on DataBricks best practices.
* Worked on Partitions, bucketing concepts in Hive and designed both Managed and External tables in Hive for optimized performance.
* Involved in writing optimized Pig Script along with involved in developing and testing Pig Latin Scripts
* Managing and monitoring the Hadoop cluster through **Cloudera Manager**.

**Environment:** JDK, JSP, Spring, Hibernate, HQL, Web Services (REST), JSON, JDBC, JavaScript, HTML, servlet, CSS, JSP, Spring MVC, Servlet XML, AJAX, Eclipse, WebLogic PL/SQL, Oracle 10g, Windows, CVS, Maven, RDMBS.

**HYPE Dhaka, Dhaka Mar 2012 – Dec 2016**

**Java Developer**

**Roles & Responsibilities:**

* Worked on various UI technologies like JSF, HTML, CSS, JavaScript and JSP tag libraries.
* Built Web pages that are more user-interactive using jQuery plugins for Drag and Drop, Auto Complete, JSON, Angular JS, JavaScript, Bootstrap.
* Used advanced JavaScript concepts to validate input fields using regular expression.
* Extensively used CSS (Cascading Style Sheets) along with Responsive Web Design (RWD) to support the web page in Mobile browser as well.
* Built Web pages that are more user-interactive using jQuery Drag and Drop, Auto Complete, JSON, Angular JS, JavaScript, Bootstrap.
* Application is deployed on AWS Cloud Platform using Elastic Beanstalk.
* Configured AWS Multi Factor Authentication in IAM to implement 2 step authentications of user's access using OKTA Verify and AWS Virtual MFA.
* Experience in integrating code quality tools such as SonarQube, in CI/CD pipelines.
* Translate Java based Elasticsearch and MongoDB ingest modules to Spring Boot.
* LDAP user authentication for Spring Boot applications.
* Implemented Teradata RDBMS analysis with Business Objects to develop reports, interactive drill charts, balanced scorecards and dynamic Dashboards.
* Implement Azyuserver-side backend components using Java/JavaEE and Spring Boot.
* Developed user interface using Node JS and developed client-side interface using React JS
* Involved in Enhancement of existing application utilizing React JS, created HTML navigation menu that is role-based menu items changes dynamically, derived from the database in the form of XML
* Used special syntax called JSX to mix HTML with Java Script with the help of React JS.
* Worked with Spring Batch Used Spring ORM module to integrate with Hibernate.
* Used Bootstrap along with CSS to design web-pages by HTML.
* Used Web services (SOAP and RESTful) for transmission of large blocks of XML/JSON.
* Used Jenkins as continuous integration tool for building archive files and snapshots for all applications.
* Worked on DOM manipulation using JavaScript, jQuery and AJAX.
* Added analytics and tagging functionality for all the existing pages to track all the user actions using JavaScript.
* Worked with CSS background, CSS Layouts, CSS positioning, CSS text, CSS border, CSS margin, CSS padding, Pseudo elements and CSS behaviors.
* Created SOAP APIs.
* Implantation of CI/CD end to end pipeline
* Scaled microservices, distributed systems and serverless applications using Simple Queue Service (SQS).
* Modified the APIs and the database to improve the performance of the APIs.
* Tested the application under Scrum (Agile) Methodology.
* Created Test strategy, Test plan, execution plan and Identified test data requirements.
* Worked on major bug fixes, which included UI issues and functionality issues as well

**Environment:** HTML, CSS, JavaScript, jQuery, DHTML, DOM, JSON, Bootstrap, AJAX, Adobe Flash, Photoshop, AEM, PLSQL, Oracle, Windows

**Educational Details:**

Master’s in business administration Southeast University (Dhaka, Bangladesh) Year 2011