**Abiram.K**

**raam.sarma23@gmail.com**

**+1(847) 859 9220**

[**LinkedIn**](https://www.linkedin.com/in/abi-ram-219275242/)

**Sr. Data Engineer**

**PROFESSIONAL SUMMARY**

* Data Engineer professional with 10+ years of combined experience in the fields of Data Engineer, Big Data implementations and Spark technologies.
* Experience in Big Data ecosystems using Spark, PySpark, Hadoop, Pig, Hive, HDFS, Sqoop, Storm, Spark, Airflow, DB2
* Teradata, Flume, Kafka, Yarn, Oozie, and Zookeeper.
* High Exposure on Big Data technologies and Hadoop ecosystem including Map Reduce and Hadoop Infrastructure.
* Multiple years of IT experience in a Healthcare, Insurance & Financial Domains working on Big Data technology using technologies such as Cloudera and Hortonworks distributions.
* Strong programming experience with Scala, Java, Python, SQL, T-SQL, R.
* Proficient in utilizing modern cloud-based technologies, particularly AWS, Azure, Google Cloud Platform (GCP) and Databricks to build scalable and efficient data pipelines.
* Hands on experience with Data modeling, Physical Datawarehouse designing & cloud data warehousing technologies including Snowflake, Redshift, BigQuery, Synapse.
* Experience with major cloud providers & cloud data engineering services including AWS, Azure, GCP & Databricks.
* Created and optimized Talend jobs for data extraction, data cleansing, and data transformation.
* Designed & orchestrated data processing layer & ETL pipelines using Airflow, Azure Data Factory, Oozie, Autosys, Cron & Control-M.
* Hands on experience with AWS services including EMR, EC2, Redshift, Glue, Lambda, SNS, SQS, CloudWatch, Kinesis, Step functions, Managed Airflow instances, Storage & Compute.
* Hands on experience with Azure services including Synapse, Azure Data Factory, Azure functions, EventHub, Stream Analytics, Key Vault, Storage & Compute.
* Hands on experience with GCP services including DataProc, VM, Big Query, Dataflow, Cloud functions, Pub/Sub, Composer, Secrets, Storage & Compute.
* Hands on experience with Databricks services including Notebooks, Delta Tables, SQL Endpoints, Unity Catalog, Secrets, Clusters.
* Have Extensive Experience in IT data analytics projects, Hands on experience in migrating on-premises data & data processing pipelines to cloud including AWS, Azure & GCP.
* Experienced in fact dimensional modeling (Star schema, Snowflake schema), transactional modeling and SCD (Slowly changing dimension)
* Hands on experience in MS SQL Server with Business Intelligence in SQL Server Integration Services (SSIS), SQL Server Analysis Services (SSAS), SQL Server Reporting Services (SSRS).
* Experience in Data Governance & Master Data Management through Collibra & Informatica. Standardization to improve Master Data Management (MDM) and other common data management issues.
* Hands-on experience in developing and deploying enterprise-based applications using major Hadoop ecosystem components like MapReduce, YARN, Hive, HBase, Flume, Sqoop, Spark MLlib, Spark GraphX, Spark SQL, Kafka.
* Adept at configuring and installing Hadoop/Spark Ecosystem Components.
* Proficient with Spark Core, Spark SQL, Spark MLlib, Spark GraphX and Spark Streaming for processing and transforming complex data using in-memory computing capabilities implemented in Scala.
* Worked with Spark to improve efficiency of existing algorithms using Spark Context, Spark SQL, Spark MLlib, Data Frame, Pair RDD's and Spark YARN.
* Extensive experience in ETL/ELT processes, data integration, and transformation using tools such as Apache Beam and Cloud Dataflow.
* Proficient in leveraging Azure services such as Azure Data Factory, Azure Databricks, Azure SQL Database, and Azure Synapse Analytics to build robust and scalable data pipelines.
* Implemented data governance practices, including data lineage and metadata management, using Alteryx, promoting data traceability, and facilitating data governance initiatives.
* Hands-on experience working Amazon Web Services (AWS) using Elastic Map Reduce (EMR), Redshift, and EC2 for data processing.
* Designed and implemented data integration solutions using Informatica PowerCenter, enabling seamless extraction, transformation, and loading (ETL) of data from various sources into target systems.
* Experienced in implementing data governance practices, ensuring data security, privacy, and compliance with regulatory requirements within Azure environments.
* Proven track record of optimizing data processes and improving performance by leveraging Azure services, such as parallel processing and auto-scaling capabilities.
* Developed and maintained data models and reports using Oracle Business Intelligence Enterprise Edition (OBIEE).
* Utilized PL/SQL programming language to develop and optimize data transformation routines, ensuring efficient data processing and adherence to business rules.
* Highly proficient in programming languages such as Python, Java, or Scala for data manipulation and automation.
* Experience in Extraction, Transformation and Loading (ETL) data from various sources into Data Warehouses, as well as data processing like collecting, aggregating, and moving data from various sources using Apache Flume, Kafka, Power BI and Microsoft SSIS.
* Hands on experience with ETL, Hadoop and Data Governance tools such as Tableau, Informatica Enterprise Data Catalog
* Implemented data masking and data anonymization techniques within IDQ to protect sensitive information in accordance with data privacy regulations.
* Designed and executed data profiling and data quality assessment processes within IDQ to identify data anomalies, inconsistencies, and data quality issues.
* Solid Experience and understanding of Implementing large scale Data warehousing Programs and E2E Data Integration Solutions on Snowflake Cloud, GCP, Redshift, Informatics Intelligent Cloud Services (IICS - CDI) & Informatica Power Center integrated with multiple Relational databases (MySQL, Teradata, Oracle, Sybase, SQL server, DB2)
* Hands-on experience with Hadoop architecture and various components such as Hadoop File System HDFS, Job Tracker, Task Tracker, Name Node, Data Node and Hadoop MapReduce programming.
* Comprehensive experience in developing simple to complex Map reduce and Streaming jobs using Scala and Java for data cleansing, filtering and data aggregation. Also possess detailed knowledge of MapReduce framework.
* Experience working with NoSQL databases like Cassandra and HBase and developed real-time read/write access to very large datasets via HBase.
* Expertise in writing end to end Data processing Jobs to analyze data using MapReduce, Spark and Hive.
* Experience with Apache Spark ecosystem using Spark-Core, SQL, Data Frames, RDD's and knowledge on Spark MLLib.
* Experienced in data manipulation using python for loading and extraction as well as with python libraries such as NumPy, SciPy and Pandas for data analysis and numerical computations.
* Hands-on experience in SQL and NOSQL databases such as Snowflake, HBase, Cassandra and MongoDB.
* Hands on experience in setting up workflow using Apache Airflow and Oozie workflow engine for managing and scheduling Hadoop jobs.
* Created tableau dashboards designing with large data volumes from data source SQL server.
* Proficient in working with MS SQL Server, including database design, optimization, and administration tasks such as performance tuning, indexing, and query optimization.
* Extensive experience in building ETL (Extract, Transform, Load) processes to integrate data from various sources into a centralized data warehouse using Data Bricks and MS SQL Server Integration Services (SSIS).
* Strong knowledge of SQL programming, including complex queries, stored procedures, and functions, to extract and manipulate data effectively from relational databases.
* Extract, Transform and Load (ETL) source data into respective target tables to build the required data marts.
* Practical understanding of the Data modeling (Dimensional & Relational) concepts like Star-Schema Modeling, Snowflake Schema Modeling, Fact and Dimension tables.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Big Data Technologies** | Hadoop, MapReduce, Spark, HDFS, Sqoop, YARN, Oozie, Hive, Impala, Zookeeper, Apache Flume, Apache Airflow, Cloudera, HBase |
| **Programming Languages** | Python, PL/SQL, SQL, Scala, C, C#, C++, T-SQL, Power Shell Scripting, JavaScript, Perl script. |
| **Cloud Technologies** | AWS, Microsoft Azure, GCP, Databricks, Snowflake |
| **Cloud Services** | Azure Data Lake Storage Gen 2, Azure Data Factory, Blob storage, Azure SQL DB, Databricks, Azure Event Hubs, AWS RDS, Amazon SQS, Amazon S3, AWS EMR, Lambda, AWS SNS, Data Flow, Big Query, VM, Delta Tables, Cloud functions, Clusters. |
| **Databases** | MySQL, SQL Server, IBM DB2, Postgres, Oracle, MS Access, Teradata, and Snowflake |
| **NoSQL Data Bases** | MongoDB, Cassandra DB, HBase |
| **Development Strategies** | Agile, Lean Agile, Pair Programming, Waterfall, and Test-Driven Development. |
| **ETL, Visualization & Reporting** | Tableau, Data Stage, Informatica, Talend, SSIS, and SSRS |
| **Frameworks** | Django, Pandas, NumPy, Matplot Lib, TensorFlow, PyTorch |
| **Version Control & Containerization tools** | Jenkins, Git, CircleCI and SVN |
| **Operating Systems** | Unix, Linux, Windows, Mac OS |
| **Monitoring tool** | Apache Airflow, Control M |
| **Tools** | PyCharm, Eclipse, Visual Studio, SQL\*Plus, SQL Developer, TOAD, SQL Navigator, Query Analyzer, SQL Server Management Studio, SQL Assistance, Eclipse, Postman |
| **Machine Learning Techniques:** | Linear & Logistic Regression, Classification and Regression Trees, Random Forest, Associative rules, NLP and Clustering. |

**PROFESSIONAL EXPERIENCE**

**CapitalOne, Mclean, VA Mar 2022 – Present**

**Senior Data Engineer**

**Responsibilities:**

* Designed and implemented end-to-end data solutions on Azure, leveraging services such as Azure Data Factory, Azure Databricks, Azure Data Lake Storage, and Azure SQL Database.
* Developed data pipelines using Azure Data Factory to orchestrate data movement and transformations across diverse data sources and destinations.
* Built and optimized scalable data processing workflows using Azure Databricks, leveraging Spark for data ingestion, transformation, and analysis.
* Proficient in writing Spark jobs using Scala and Python in Azure Databricks notebooks for data cleansing, feature engineering, and advanced analytics.
* Implemented real-time data processing solutions using Azure Event Hubs, Azure Stream Analytics, and Databricks Structured Streaming.
* Led the implementation of data quality initiatives using Informatica Data Quality (IDQ) to ensure the accuracy, consistency, and reliability of critical data assets.
* Developed and maintained data profiling and data quality assessment processes within IDQ to identify data anomalies, inconsistencies, and data quality issues.
* Designed end-to-end data pipelines in Azure Data Factory for data extraction, transformation, and loading.
* Integrated Teradata as a data source within Azure Data Factory pipelines for seamless data movement.
* Extensive experience in implementing data lake architectures using Azure Data Lake Storage for storing and processing large volumes of structured and unstructured data.
* Designed and developed data warehousing solutions using Azure Synapse Analytics, including data modeling, SQL script development, and query optimization.
* Establish and enforce data governance practices on Databricks Feature Store, including metadata management, data lineage tracking, and data quality monitoring, to ensure data integrity and reliability throughout the feature lifecycle.
* Implemented machine learning workflows on Azure using Azure Databricks and Azure Machine Learning, including data preparation, model training, and deployment.
* Developed and maintained data pipelines to ensure seamless and efficient data ingestion into Oracle Database.
* Demonstrated experience in handling large-scale data processing using Teradata.
* Integrated and leveraged Azure Cosmos DB, a globally distributed NoSQL database, to handle large-scale and highly responsive applications with low latency requirements.
* Developed data ingestion and extraction processes, ensuring data quality and integrity using Azure Data Factory and related data integration technologies.
* Implemented data security and access controls in Azure data products, ensuring compliance with regulatory requirements and protecting sensitive data.
* Proficient in Teradata database design, development, optimization, and maintenance.
* Utilized Azure Monitor to proactively monitor and troubleshoot data pipelines and services, identifying and resolving performance issues and bottlenecks.
* Integrated Azure Application Insights to monitor the performance and usage of data engineering solutions, identifying areas for optimization and improvement.
* Proficient in using control structures, loops, and conditional logic within Teradata stored procedures.
* Utilized Azure Log Analytics to analyze and visualize log data, facilitating effective troubleshooting and monitoring of data pipelines and systems.
* Proficient in implementing and managing big data clusters on Azure Databricks, including cluster provisioning, monitoring, and optimization.
* Implemented data governance practices in Azure, ensuring data security, privacy, and compliance with industry standards and regulations.
* Monitored and optimized the performance of Oracle databases, including SQL query optimization and database configuration tuning specific to Oracle Database.
* Experienced in implementing data replication and synchronization using Azure Data Factory and Azure Databricks to enable hybrid data integration scenarios.
* Developed automated data quality checks and monitoring solutions using Azure services such as Azure Monitor and Azure Log Analytics.
* Designed, developed, and maintained the data architecture for SSRS reports, ensuring efficient data storage, retrieval, and processing.
* Collaborated with cross-functional teams to gather requirements, design data solutions, and deliver projects on time and within budget.
* Designed and executed data cleansing and data standardization routines in IDQ to transform and enrich data for downstream applications.
* Collaborated with data analysts and business stakeholders to define data quality rules and data quality scorecards within IDQ, aligning with business requirements.
* Implemented data partitioning and optimization techniques to improve data processing performance and reduce costs in Azure Databricks.
* Optimized the performance of SSRS reports and underlying data queries to ensure fast response times and efficient resource utilization.
* Implemented data encryption and implemented security controls in Azure to ensure data protection and compliance with organizational policies.
* Experience in implementing Azure Data Bricks Auto-Scaling and Auto-Termination policies to optimize resource utilization and cost management.
* Implemented CI/CD pipelines using Azure DevOps to automate deployment of data pipelines and workflows.
* Implemented data lineage and metadata management solutions to track and document data transformations and lineage using Azure services.
* Experience in optimizing and fine-tuning Spark jobs and SQL queries in Azure Databricks to improve performance and resource utilization.
* Implemented data archival and data retention policies using Azure services such as Azure Blob Storage and Azure Data Lake Storage.
* Integrated SSRS reports with other systems and applications as required, including configuring data connections, APIs, and data synchronization processes.
* Developed and deployed machine learning models using Azure Machine Learning, integrating them into production data workflows and Databricks pipelines.
* Implemented data security measures, including role-based access control (RBAC), encryption, and data masking techniques in Azure environments.
* Tested Oracle Database backup and recovery procedures regularly to ensure reliability and effectiveness.
* Proficient in troubleshooting and resolving issues related to Azure services, Databricks clusters, and data pipelines.
* Implemented data cataloging and metadata management solutions using Azure Data Catalog and Databricks Delta Lake.
* Implemented data streaming and real-time analytics solutions using Azure Event Hubs, Azure Stream Analytics, and Azure Databricks.
* Monitor Oracle database growth and forecast future capacity requirements specifically considering Oracle Database's performance characteristics.
* Experience in migrating on-premises data and workloads to Azure, including re-architecting and optimizing data processes for the cloud.
* Implemented data-driven insights and visualizations using Azure services such as Azure Data Explorer, Azure Synapse Studio, and Power BI.
* Implemented data access controls and auditing mechanisms to ensure data governance and compliance with regulatory requirements.
* Experience in working with Azure Cognitive Services and integrating them with data pipelines for text analytics, image recognition, and natural language processing.
* Stayed up to date with the latest developments in Azure and Databricks, exploring new features.

**Environment:** Azure SQL, Azure Storage Explorer, Azure Storage, Azure Blob Storage, Azure Backup, Azure Files, Azure Data Lake Storage, SQL Server Management Studio 2016, Teradata, Visual Studio 2015, VSTS, Azure Blob, Power BI, PowerShell, C# .Net, SSIS, DataGrid, ETL Extract Transformation and Load, Business Intelligence (BI).

**Vanguard, Malvern, PA Feb 2020 to Feb 2022**

**Sr. Data Engineer**

**Responsibilities:**

* Responsible for the execution of big data analytics, predictive analytics, and machine learning initiatives.
* Implemented a proof of concept deploying this product in AWS S3 bucket and Snowflake.
* Utilize AWS services with focus on big data architect /analytics / enterprise Data warehouse and business intelligence solutions to ensure optimal architecture, scalability, flexibility, availability, performance, and to provide meaningful and valuable information for better decision-making.
* Developed ETL (Extract, Transform, Load) scripts using AWS Glue's serverless Spark-based ETL engine to cleanse, transform, and enrich raw data.
* Designed, developed, and implemented data ingestion pipelines using AWS Glue to extract data from various sources such as databases, data lakes, and streaming platforms.
* Created a snowflake warehouse strategy and set it up to use PUT scripts to migrate a terabyte of data from S3 into Snowflake.
* Tuned Spark configurations, data partitioning strategies, and caching mechanisms in Java to improve job performance.
* Designed and implemented appropriate indexing strategies on database tables to improve query performance, reduce disk I/O, and enhance overall system efficiency.
* Developed robust data ingestion pipelines on Databricks to efficiently extract, transform, and load data from various sources into the storage layer, ensuring data quality, consistency, and reliability.
* By creating a customized read/write Snowflake utility function in python, data was transferred from an AWS S3 bucket to Snowflake.
* Strong expertise in developing and scheduling ETL workflows using AWS Glue jobs to automate data extraction, transformation, and loading processes.
* Extensive experience in designing, implementing, and managing data warehousing solutions using Amazon Redshift.
* Created S3 buckets and managed S3 bucket policies, as well as using S3 buckets for storage and backup on AWS.
* Developed and implemented data governance policies and procedures to ensure the accuracy, integrity, and security of data within the organization.
* Rewrote and restructured SQL queries to leverage database indexes, minimize table scans, reduce join complexity, and optimize execution plans for faster query processing.
* Designed and implemented a Snowflake data warehouse architecture to support efficient data storage, retrieval, and analysis.
* Ensured data quality, integrity, and consistency within Delta Live Tables by implementing data quality checks, validation rules, and governance policies.
* Designed and managed schemas for structured and semi-structured data stored in Databricks, ensuring compatibility with downstream analytics and machine learning processes while accommodating evolving business requirements.
* Implemented Data Vault methodology for data modeling, ensuring flexibility, scalability, and easy adaptability to changing business requirements.
* Conducted performance testing and benchmarking of SQL queries under various load conditions, data volumes, and concurrency levels to validate optimizations and identify potential performance regressions.
* Collaborated with cross-functional teams to identify and define key data elements, data lineage, and data quality rules for master data management (MDM) initiatives.
* Designed and implemented a robust data quality framework using Snowflake's built-in features and DBT (Data Build Tool) to perform data profiling, validation, and cleansing.
* Designing and implementing automated data quality checks and validation rules using Java-based frameworks or libraries to monitor data quality in real-time or as part of batch processing workflows.
* Developed and maintained data dictionaries, data catalogs, and data lineage documentation for improved data understanding and traceability.
* Worked on end-to-end deployment of the project that involved Data Analysis, Data Pipelining, Data Modelling, Data Reporting and Data documentations as per the business needs.
* Construct the AWS data pipelines using VPC, EC2, S3, Auto Scaling Groups (ASG), EBS, Snowflake, IAM, CloudFormation, Route 53, CloudWatch, CloudFront, CloudTrail.
* Launching and configuring of Amazon EC2 (AWS) Cloud Servers using AMI s (Linux/Ubuntu) configuring the servers for specified applications.
* Authoring JavaScripts for custom UDF’s for Row/ Column manipulations, merges, aggregations, stacking, data labeling and for all Cleaning and conforming tasks.
* Gathering the data stored in AWS S3 from various third-party vendors, optimizing it and joining with internal datasets to gather meaningful information.
* Design and Develop ETL Processes in AWS Glue to migrate Campaign data from external sources like S3, ORC/Parquet/Text Files into AWS Redshift
* AWS CI/CD Data pipeline and AWS Data Lake using EC2, AWS Glue, AWS Lambda.
* Migrate data from on-premises to AWS storage buckets.
* Developed a python script to transfer data from on-premises to AWS S3. Developed a python script to hit REST API’s and extract data to AWS S3.
* Worked on Ingesting data by going through cleansing and transformations and leveraging AWS Lambda, AWS Glue and Step Functions.
* Developed Spark applications using PySpark and Spark-SQL for data extraction, transformation, and aggregation from multiple file formats.
* Developed PySpark and SparkSQL code to process the data in Apache Spark on Amazon EMR to perform the necessary transformations based on the source to target mappings developed.
* Developed Data mapping, Transformation and Cleansing rules for the Data Management involving OLTP and OLAP.
* Skilled in visualizing and presenting data using Tableau, creating interactive dashboards, and generating meaningful insights for stakeholders.
* Used AWS EMR to transform and move large amounts of data into and out of other AWS data stores and databases, such as Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB.
* Collected, cleaned, and preprocessed data from various sources to ensure high-quality input for machine learning models. This involves understanding data requirements, performing data wrangling, and handling missing or inconsistent data. Including algorithms data imputation techniques such as k-Nearest Neighbors (k-NN) or algorithms for outlier detection like Isolation Forest.
* Collaborated with data scientists and machine learning engineers to assist in model development tasks such as model selection, hyperparameter tuning, and algorithm optimization. Provided input on data-related considerations that can impact model performance. Algorithms like Decision Trees, Random Forests, are used for classification and regression tasks.
* Performed benchmark tests to read data from database, object store using pandas and PySpark API s to compare results, identify potential improvement areas and provide recommendations.
* Read and write Parquet, JSON files from S3 buckets using Spark, Pandas data frame with various configuration.
* Develop code to assign default life cycle policy to buckets, objects and auto purge objects based on default policy in mercury which is internal implementation of AWS S3.
* Designed and implemented data visualizations and charts in Tableau to effectively communicate complex data insights and trends to non-technical users.
* Work closely with the application customers to resolve JIRA tickets related to API issues, data issues, consumption latencies, onboarding, and publishing data.

Environment: HDFS, Oozie, Spark, Kafka, Zookeeper, Java, Python, HBase, Flume, AWS (EMR, S3, EC2, VPC, LAMBDA, EBS, CloudFront, IAM, CloudWatch, Redshift, DynamoDB), Teradata, DBT, Kafka, PySpark, Snowflake, Tableau.

**BCBS, Chicago, IL August 2018-Jan 2020**

**Data Engineer**

**Responsibilities:**

* Extensive experience in designing, implementing, and managing data solutions on AWS, leveraging services such as Amazon S3, Amazon Redshift, Amazon Glue, and Amazon EMR.
* Proficient in building and optimizing scalable data pipelines using AWS Data Pipeline, AWS Glue, and Databricks to extract, transform, and load data from various sources.
* Hands-on expertise in implementing big data processing and analytics workflows using Databricks, leveraging Apache Spark for data ingestion, transformation, and analysis.
* Strong understanding of Spark architecture and proficient in writing Spark jobs using Scala or Python, leveraging Databricks notebooks for data exploration, feature engineering, and model development.
* Experienced in building and managing Spark clusters on AWS Databricks, including cluster provisioning, configuration, and performance optimization.
* Define DBT models to represent your business logic and transformations on raw data.
* Design and create database schemas and tables based on data modeling principles. Leverage Snowflake's support for schema evolution to make changes to existing structures.
* Implemented real-time data processing and streaming analytics solutions using AWS services such as Amazon Kinesis and Databricks Structured Streaming.
* Wrote and executed tests in DBT to validate the accuracy and integrity of your transformed data.
* Developed Spark applications using Scala for easy Hadoop transitions.
* Used Spark and Spark-SQL to read the parquet data and create the tables in hive using the Scala API.
* Performed advanced procedures like text analytics and processing, using the in-memory. computing capabilities of Spark using Scala.
* Developed Spark code using Scala and Spark-SQL for faster processing and testing.
* 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.
* Responsible for loading Data pipelines from web servers and Teradata using Sqoop with Kafka and Spark Streaming API.
* Developed Kafka producer and consumers and Spark along with components on HDFS, Hive.
* Populated HDFS and HBase with huge amounts of data using Apache Kafka.
* Used Kafka to ingest data into Spark engine.
* Wrote stored procedures in Snowflake for automating and orchestrating data processing tasks.
* Proficient in utilizing AWS Glue for data cataloging, metadata management, and automatic schema inference to streamline data discovery and exploration.
* Implemented data warehousing solutions on AWS using Amazon Redshift, including schema design, SQL script development, and query optimization.
* Experience in deploying and managing machine learning models on AWS using Databricks and Amazon SageMaker, integrating them with data pipelines for real-time predictions.
* Implemented data governance practices on AWS, including data security, privacy, and compliance with regulations such as GDPR and HIPAA.
* Proficient in leveraging AWS Lambda for serverless data processing, event-driven architectures, and data transformation tasks.
* Experience in implementing data lakes on AWS using Amazon S3 for scalable storage and processing of structured and unstructured data.
* Implemented data replication and synchronization solutions using AWS services such as AWS Data Sync and AWS Glue to ensure data consistency across multiple environments.
* Implemented data security measures on AWS, including encryption at rest and in transit, identity, and access management (IAM), and network security controls.
* Proficient in deploying and managing ETL processes on AWS using AWS Glue, leveraging its data discovery, schema evolution, and data transformation capabilities.
* Implemented data archival and backup strategies using AWS services such as Amazon S3 Glacier and AWS Backup for long-term data retention.
* Implemented data streaming solutions on AWS using Amazon Kinesis Data Streams and Databricks Structured Streaming for real-time data ingestion and analysis.
* Experience in monitoring and optimizing AWS resources and services for performance, cost efficiency, and scalability using AWS CloudWatch, AWS Cost Explorer, and Databricks monitoring.
* Collaborated with cross-functional teams to gather requirements, design data solutions, and deliver projects on time and within budget.
* Exploring with the Spark for improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, Pair RDD's, Spark YARN.
* Design & implement ETL process using Talend to load data from Worked extensively with Sqoop for importing and exporting the data from HDFS to Relational Database systems/mainframe and vice-versa. Loading data into HDFS.
* Created and altered HBase tables on top of data residing in Data Lake and Created external Hive tables on the Blobs to showcase the data to the Hive Meta Store.
* Utilized Agile Scrum Methodology to help manage and organize a team of 4 developers with regular code review sessions.
* Stayed up to date with the latest developments in AWS and Databricks, exploring new features and functionalities, and actively participating in relevant communities and forums.

**Environment:** Agile Scrum, Spark, Scala, Hive, Kafka, Python, AWS (EC2, S3, EBS, ELB, RDS, SNS, SQS, VPC, Cloud formation, CloudWatch, ELK Stack), Bitbucket, Ansible.

**Empower Retirement, Greenwood Village, CO Mar 2016 - July 2018**

**ETL Developer**

**Responsibilities**:

* Extensive experience in designing and implementing data integration solutions using Informatica PowerCenter for extracting, transforming, and loading data from various sources to target systems.
* Proficient in developing ETL mappings, transformations, and workflows using Informatica PowerCenter to ensure accurate and timely data delivery.
* Strong understanding of data integration best practices and data quality principles to ensure the integrity and consistency of data throughout the ETL process.
* Hands-on experience in implementing Master Data Management (MDM) solutions using Informatica MDM to consolidate and manage enterprise-wide master data.
* Proficient in configuring and customizing Informatica MDM Hub and Data Director for data governance, data stewardship, and data quality management.
* Implemented data profiling and data cleansing techniques using Informatica Data Quality (IDQ) to identify and resolve data quality issues within source systems.
* Designed and implemented data validation rules, match and merge strategies, and data standardization processes using Informatica MDM for accurate and reliable master data management.
* Proficient in integrating Informatica MDM with other systems such as CRM, ERP, and data warehouses to ensure consistent master data across the enterprise.
* Implemented data synchronization processes between source systems and Informatica MDM Hub to ensure data consistency and accuracy in real-time or batch mode.
* Developed custom data validation and enrichment rules using Informatica MDM to improve data quality and completeness.
* Collaborated with business stakeholders and data stewards to define data governance policies, data ownership, and data stewardship processes within Informatica MDM.
* Implemented data profiling and data quality scorecards using Informatica Data Quality to measure and monitor the quality of master data.
* Experienced in designing and implementing data models and metadata structures within Informatica MDM for efficient data management and retrieval.
* Implemented workflows and business rules within Informatica MDM for data validation, approval, and exception handling processes.
* Proficient in designing and implementing data hierarchies, relationships, and data domains within Informatica MDM to ensure accurate and consistent master data representation.
* Implemented data deduplication and data consolidation strategies using Informatica MDM for effective management of duplicate or redundant data.
* Developed custom integrations and APIs using Informatica MDM to enable seamless data exchange with external systems and applications.
* Implemented data governance frameworks and data stewardship processes within Informatica MDM to ensure data quality, compliance, and data lifecycle management.
* Proficient in performance tuning and optimization techniques within Informatica MDM and PowerCenter to improve data processing speed and efficiency.
* Stayed updated with the latest developments in Informatica and MDM, exploring new features and functionalities, and actively participating in Informatica community forums or user groups.

**Environment:** SQL, Informatica, MDM, ETL, CRM, Data Quality, Data Governance, Informatica Power Center.

**VNV Logix pace Private. Ltd, Hyderabad, India June 2014-Dec 2015**

**Data Analyst**

**Responsibilities:**

* Extensive experience in designing, developing, and implementing ETL processes using SSIS for efficient data extraction, transformation, and loading.
* Proficient in creating SSIS packages to integrate data from various sources into SQL. Server databases, data warehouses, and data marts.
* Strong understanding of data modeling concepts and relational databases, ensuring data consistency, integrity, and performance in SSIS solutions.
* Experienced in designing and implementing complex data transformations and business logic using SSIS data flow components and expressions.
* Proficient in deploying and scheduling SSIS packages using SQL Server Agent, ensuring reliable and timely data processing.
* Skilled in troubleshooting and resolving issues related to data quality, performance, and error handling in SSIS packages.
* Implemented incremental loading strategies using SSIS to efficiently process large volumes of data and improve ETL performance.
* Proficient in integrating external systems and APIs with SSIS packages, facilitating data exchange and automation of data workflows.
* Strong understanding of SSIS control flow tasks and event handlers, enabling conditional execution and error handling.
* Experienced in deploying and managing SSIS packages in different environments, including development, testing, and production.
* Proficient in creating and configuring SSIS connection managers for various data sources, including flat files, databases, and web services.
* Developed custom SSIS components using Script tasks and Script transformations, extending the capabilities of SSIS for specialized data processing requirements.
* Implemented data validation and cleansing techniques within SSIS packages using SQL Server data quality tools and custom transformations.
* Experienced in performance tuning and optimizing SSIS packages, including data flow transformations, buffer size tuning, and parallel execution.
* Proficient in creating reports and visualizations using SSRS, including designing report layouts, defining data sources, and implementing interactive features.
* Strong understanding of SQL Server Reporting Services architecture, including report server configuration, security, and report delivery options.
* Implemented complex report calculations, expressions, and parameters within SSRS reports to meet specific business requirements.
* Proficient in developing drill-down, drill-through, and ad-hoc reports using SSRS, providing users with interactive data exploration capabilities.
* Experienced in deploying and managing SSRS reports to different report servers, ensuring secure and controlled access to reports.
* Collaborated with business stakeholders to gather report requirements, design report layouts, and deliver high-quality and visually appealing reports.

**Environment:** SQL, SSIS, SSRS, ETL, SQL Server, Data Warehouse, Data Mart.

**EDUCATIONAL DETAILS**