**Priya Kadaluri**

**Sr. Data Engineer**

[**priyakadaluri@gmail.com**](mailto:priyakadaluri@gmail.com) **| +1 (713)-489-8967 |** [**Linkedin**](https://www.linkedin.com/in/priyakadaluri/)

**SUMMARY:**

* Over 10 years of extensive experience as a Data Engineer with expertise in Data Analytics, Data Architecture, Design, Development, Implementation, Testing, and Deployment of Software Applications across Banking, Finance, Insurance, Retail, and Telecom domains.
* Designed and implemented complete end-to-end data infrastructure using AWS services such as S3, Redshift, EMR, Glue, Kafka, Lambda, and EC2, with hands-on experience in Hadoop ecosystem components like HDFS, MapReduce, Hive, HBase, Sqoop, and Spark.
* Developed DataFrames and performed analysis using Spark SQL on Amazon EMR.
* Acute knowledge of Spark Streaming and Spark Machine Learning Libraries.
* Hands-on expertise in writing various RDD (Resilient Distributed Datasets) transformations and actions using Scala, Python, and Java.
* Excellent understanding of Spark Architecture and Framework, including Spark Context, APIs, RDDs, Spark SQL, DataFrames, Streaming, and MLlib.
* Worked in agile projects delivering end-to-end continuous integration/continuous delivery (CI/CD) pipelines by integrating Jenkins and AWS services for VM provisioning and data pipeline automation.
* Created automatic scripts for monitoring file systems and key services on AWS.
* Implemented CI/CD through Jenkins for Hadoop jobs and integrated AWS CodePipeline and CodeBuild for data pipelines.
* Good knowledge of AWS services such as S3, Redshift, Lambda, EC2, EMR, and Glue, with experience in integrating GCP services like Google Cloud Storage and BigQuery where applicable.
* **Excellent understanding of Hadoop Architecture**, including components like MapReduce, HDFS, HBase, Hive, Sqoop, Cassandra, Kafka, and Databricks, with expertise in data testing and monitoring.
* Utilized Sqoop to import data from RDBMS into HDFS and Hive, and used storage formats such as Text, Avro, Parquet, Sequence File, and ORC with compression codes like Snappy and Gzip.
* Performed transformations on imported data and exported it back to RDBMS, Amazon Redshift, and Databricks.
* Experience writing queries in HQL (Hive Query Language) and SQL for Redshift and Databricks to perform data analysis.
* Created Hive External and Managed Tables and implemented Partitioning and Bucketing on Hive tables for optimization.
* Used Apache Flume to ingest data from various sources into sinks like Avro, HDFS, Amazon S3, and Databricks.
* Implemented custom interceptors for Flume to filter data and defined channel selectors to multiplex data into different sinks, including Databricks.
* Excellent knowledge of Kafka Architecture and integrated Flume with Kafka and Databricks, utilizing it for activity tracking and log aggregation.Experienced in writing Oozie workflows and coordinator jobs to schedule sequential Hadoop jobs, as well as orchestrating workflows using AWS Step Functions.
* Worked with various file formats including Text, Sequence files, XML, Parquet, JSON, ORC, AVRO, and Click Stream log files.
* Familiar with data architecture including data ingestion pipeline design, Hadoop architecture, data modeling, and advanced data processing, including optimization of ETL workflows using AWS Glue and Data Pipeline.
* Good exposure to Data Quality, Data Mapping, and Data Filtration using ETL tools such as Talend, Informatica, DataStage, Ab Initio, and AWS Glue.
* Experienced in creating dashboards with reporting tools such as SAS, Tableau, Power BI, BusinessObjects, and QlikView, and using AWS QuickSight for data visualization.
* Worked with various databases including Oracle, Teradata, Informix, DB2, and NoSQL databases like MongoDB, HBase, and PostgreSQL, including Greenplum on AWS.
* Experienced in the complete Software Development Life Cycle (SDLC) including Analysis, Design, Development, Testing, Implementation, and Support using Agile and Waterfall methodologies.
* Demonstrated a comprehensive understanding of Fact/Dimension data warehouse design models, including star and snowflake schema methods, and implemented similar models using Amazon Redshift.Top of Form

Bottom of Form

**PROFESSIONAL EXPERIENCE**

**Apple, Austin, TX Oct 2024 – Present**

**Sr. Data Engineer.**

**Responsibilities:**

* Designed and developed scalable data pipelines for batch and real-time processing using Apache Kafka and gRPC.
* Architected microservices-based systems to ensure modularity, fault tolerance, and high scalability in distributed environments.
* Built and maintained REST APIs for seamless data integration across applications and external systems.
* Automated testing for REST APIs and microservices using tools like JUnit, TestNG, and Postman to ensure reliability and functionality.
* Designed and implemented Kafka Producers, Consumers, and Stream Processing solutions for real-time data streaming.
* Monitored and optimized data pipeline performance to handle high-throughput and low-latency workloads.
* Migrated monolithic systems to microservices architectures, enhancing system scalability and reducing maintenance complexity.
* Implemented robust authentication and authorization mechanisms within REST APIs and gRPC services for secure data transfer.
* Enhanced data pipeline reliability through testing automation frameworks and continuous integration practices.
* Collaborated with data science and business intelligence teams to deliver actionable insights through data engineering solutions.
* Conducted performance tuning for Java-based applications and microservices to optimize resource utilization.
* Integrated event-driven architectures using Kafka to enable real-time processing and reactive systems.
* Designed and optimized relational and NoSQL database schemas to support data processing workflows.
* Utilized containerization tools like Docker and orchestration platforms like Kubernetes for microservices deployment.
* Implemented CI/CD pipelines for automated deployment and testing of data engineering solutions.
* Conducted root cause analysis and troubleshooting to resolve issues in real-time data pipelines and microservices.
* Designed scalable data solutions with a focus on fault-tolerance and high availability.
* Developed monitoring and logging solutions for Kafka, REST APIs, and gRPC services using tools like Prometheus, Grafana, and ELK stack.
* Built data validation frameworks to improve data quality and accuracy across pipelines.
* Ensured compliance with data governance standards and regulations through secure data handling practices.
* Led efforts to integrate new data sources into existing pipelines, enabling expanded analytics capabilities.
* Designed and implemented distributed caching solutions to enhance data retrieval speeds and system responsiveness.
* Worked closely with stakeholders to translate business requirements into technical specifications and data workflows.
* Conducted training sessions for team members on data engineering best practices, Kafka, and microservices architecture.
* Researched and implemented emerging technologies to optimize data processing and improve system scalability.

**Best Buy, Charlotte, NC Nov 2022 – Aug 2024**

**Sr. Data Engineer.**

**Responsibilities:**

* Implemented installation and configuration of multi-node clusters on AWS using EC2 instances, EMR, and Databricks.
* Utilized AWS management tools such as Amazon CloudWatch for monitoring and AWS CloudTrail for logging.
* Stored log files in Amazon S3. Applied versioning in S3 for managing and protecting sensitive information.
* Integrated Amazon DynamoDB with AWS Lambda to store values and back up the DynamoDB Change Feed.
* Automated regular AWS tasks such as snapshot creation and backups using Python scripts.
* Designed data warehouses on Amazon Redshift, Databricks, and other high-performance platforms within AWS.
* Installed and configured Apache Airflow for AWS services and Databricks, creating Directed Acyclic Graphs (DAGs) to run Airflow workflows.
* Prepared scripts to automate the data ingestion process using PySpark and Scala through various sources such as APIs, Amazon S3, Teradata, Redshift, and Databricks.
* Created multiple scripts to automate the ETL/ELT process using PySpark from diverse sources.
* Developed PySpark scripts utilizing SQL and RDD in Spark for data analysis and storing data back into Amazon S3 and Databricks.
* Developed PySpark code to load data from staging to the data hub, implementing business logic.
* Developed Spark SQL code for implementing business logic, with Python as the programming language.
* Developed and trained a Retrieve-and-Generate (RAG) system to automatically generate text summaries from documents, including fine-tuning language model parameters for context retrieval and text generation.
* Constructed a lifelong machine learning system (LLMS) capable of continuously learning from text data streams while mitigating catastrophic forgetting. Created an evaluation approach to measure model stability.
* Designed, developed, and delivered jobs and transformations over the data to enrich it for consumption in the Pub layer of the data lake.
* Worked on Sequence files, map-side joins, bucketing, and partitioning for performance enhancement and storage improvement in Redshift and Databricks.
* Wrote, compiled, and executed programs using Apache Spark in Scala and Databricks to perform ETL jobs with ingested data.
* Used Spark Streaming to divide streaming data into batches as input to the Spark engine for batch processing.
* Managed Kubernetes patches and upgrades on Amazon EKS and Databricks.
* Administered multiple Kubernetes clusters in a production environment using Amazon EKS and Databricks.
* Designed and incorporated MLOps deployment for machine learning models using AWS services and Databricks. Enabled low-latency serving of model predictions.
* Developed Spark applications for data validation, cleansing, transformation, and custom aggregation using the Spark engine and Spark SQL on Databricks, and provided results to data scientists for further analysis.
* Created various UDFs in MapReduce and Python for Redshift, EMR, and Databricks.
* Conducted data integrity checks using Redshift queries, EMR, Spark, and Databricks.
* Performed transformations and actions on RDDs and Spark Streaming data using Scala and Databricks.
* Implemented machine learning algorithms using Spark with Python on Databricks.
* Profiled structured, unstructured, and semi-structured data across various sources to identify patterns and implemented data quality metrics using appropriate queries or Python scripts based on the source.
* Designed and implemented Scala programs using Spark DataFrames and RDDs for data transformations and actions on Databricks.
* Improved Redshift performance by implementing partitioning and clustering and optimized file formats (ORC) where applicable.

**Environment**: GCP, JMeter, Kafka, Ansible, Jenkins, Docker, Maven, Linux, Red Hat, GIT, Cloud Watch, Python, Shell Scripting, Golang, Web Sphere, Splunk, Tomcat, Soap UI, Kubernetes, Terraform, PowerShell.

## **Encompass Health, San Antonio, TX June 2020 – Oct 2022**

**Sr. Data Engineer.**

**Responsibilities:**

* Configured data loads from Amazon S3 to Amazon Redshift using AWS Glue and Databricks.
* Utilized Amazon Redshift and Databricks for extracting, transforming, and loading data from various heterogeneous sources and destinations.
* Created tables, stored procedures, and extracted data using SQL for business users as needed.
* Performed data analysis and design, maintaining large, complex logical and physical data models, and metadata repositories using AWS Glue Data Catalog and Databricks.
* Wrote a shell script to trigger data staging jobs in AWS.
* Assisted service developers in finding relevant content in existing reference models.
* Worked with data sources like Google Sheets, Amazon RDS, and flat files using connectors and transformations provided by AWS Glue and Databricks.
* Utilized Spark SQL API in PySpark on Amazon EMR and Databricks to extract and load data and perform SQL queries.
* Developed a PySpark script to encrypt raw data using hashing algorithms on client-specified columns on Databricks.
* Led the migration of on-premises enterprise data warehouse (EDW) to Snowflake’s cloud data platform and Databricks, reducing query times and infrastructure costs.
* Responsible for designing, developing, and testing databases and creating stored procedures, views, and triggers.
* Created visualizations and reports using Amazon QuickSight and Tableau, and Databricks.
* Designed data models that correlate all metrics to provide valuable insights.
* Optimized SQL queries by working on indexes and execution plans to reduce runtime.
* Explored and optimized algorithms using Spark on EMR, Spark SQL, Databricks, PostgreSQL, DataFrames, Kubernetes, and Talend.
* Integrated Hadoop clusters with the Spark engine to perform batch processing and GraphX operations on Databricks.
* Conducted data preprocessing and feature engineering for predictive analytics using Python Pandas on Databricks.
* Generated reports on predictive analytics using Python and Amazon QuickSight, including visualizing model performance and predictions on Databricks.
* Implemented data ingestion activities using AWS Glue, custom Glue pipeline activities, and Databricks.
* Managed data migration using SQL, Amazon RDS, Amazon S3, AWS Glue, AWS Lambda, and Databricks.
* Implemented medium to large-scale BI solutions on AWS using Redshift, AWS Glue, EMR, NoSQL databases, and Databricks.
* Migrated on-premises data (Oracle, SQL Server, DB2, MongoDB) to Amazon S3, Redshift, and Databricks using AWS Glue.
* Developed detailed project plans and managed the data conversion migration from legacy systems to the target Snowflake database and Databricks.
* Designed, developed, and tested dimensional data models using Star and Snowflake schema methodologies following Kimball methods.
* Implemented ad-hoc analysis solutions using Redshift, AWS Glue, and Databricks.
* Developed data pipelines using Spark, Hive, Pig, Python, Impala, HBase, and Databricks for customer data ingestion.
* Converted Hive/SQL queries into Spark transformations using Spark RDDs, Python, Scala, and Databricks.
* Utilized Amazon QuickSight for direct queries to compare legacy data with current data, generating reports and dashboards, and Databricks for advanced analytics.
* Designed ETL processes to extract, transfer, and load (ETL) data into Redshift and Databricks from various sources. Created and formatted reports and dashboards using Amazon QuickSight and Tableau.
* Developed filters, parameters, and calculated fields for dashboards and reports using Tableau, Amazon QuickSight, and Databricks.
* Created visualizations and dashboards using Tableau, Amazon QuickSight, and Databricks.
* Adhered to ANSI SQL language specifications where possible and provided context about similar functionality in other industry-standard engines (e.g., referencing PostgreSQL documentation).
* Implemented Slowly Changing Dimensions (SCD) transformation in ETL processes to maintain historical data in Redshift and Databricks.
* Performed ETL testing activities including job execution, data extraction, transformation, and loading into Redshift and Databricks.
* Created dashboards for analyzing POS data using Tableau and Databricks.

**Environment**: MS SQL Server 2016, T-SQL, GCP, SQL Server Integration Services (SSIS), SQL Server Reporting Services (SSRS), SQL Server Analysis Services (SSAS), Management Studio (SSMS), Advance Excel (creating formulas, pivot tables, Hlookup, Vlookup, Macros), Spark, Python, ETL, Power BI, Tableau, Presto, Hive/Hadoop, Snowflakes, Power BI, AWS Data Pipeline, IBM Cognos 10.1, Data Stage, Cognos Report Studio 10.1, Cognos 8 & 10 BI, Cognos Connection, Cognos office Connection, Cognos 8.2/3/4, Data stage and Quality Stage 7.5

**Ross Stores, Bay Area, CA. Sep 2018 – June 2020**

**Data Engineer.**

**Responsibilities:**

* Processed web server logs by developing multi-hop data pipelines using AWS Glue and Databricks, and loaded data into Amazon DynamoDB for further analysis. Extracted and processed files from DynamoDB using Glue.
* Expertise in Amazon DynamoDB, NoSQL data modeling, tuning, and disaster recovery, utilizing CRUD operations for distributed storage and processing on AWS and Databricks.
* Extracted and restructured data into Amazon DynamoDB using the AWS CLI and Databricks.
* Configured fan-out workflows in AWS Glue and Databricks to design a V-shaped architecture for ingesting data from multiple sources into a single sink.
* Experience in creating, dropping, and altering tables at runtime without blocking updates and queries using DynamoDB, Amazon Redshift, and Databricks on AWS.
* Worked with various join patterns and implemented both Map and Reduce Side Joins in Redshift and Databricks.
* Configured AWS Glue pipelines and Databricks for importing streaming log data into DynamoDB.
* Imported transactional logs from web servers into Amazon S3 using Glue and ingested the data into Redshift and Databricks.
* Utilized Amazon S3 and Glue for loading data from local systems to Redshift and Databricks.
* Installed and configured Apache Beam on AWS Glue and Databricks, and wrote Beam scripts to convert data from text files to Avro format.
* Created partitioned tables in Redshift and Databricks, and managed them using SQL queries.
* Loaded data into DynamoDB using batch and streaming methods, and Databricks for enhanced processing.
* Used Jenkins for continuous integration and automated deployment processes for cloud-based applications on AWS and Databricks.
* Integrated Redshift with Amazon QuickSight and Databricks to create and publish interactive reports and dashboards.
* Developed custom data processing scripts using Apache Beam in Java on AWS Glue and Databricks to parse raw data and populate staging tables.
* Set up and debugged log forwarding to Amazon CloudWatch Logs for Apache logs.
* Developed Spark code using Scala and Spark SQL/Streaming on Amazon EMR and Databricks for efficient data processing and testing.
* Analyzed SQL scripts and designed solutions for implementation using Scala in Spark and Databricks.
* Utilized Spark SQL to load JSON data, create schemas with RDDs, and load data into Redshift and Databricks, handling structured data with Spark SQL.
* Implemented Spark scripts using Scala and Spark SQL to access and process Redshift tables on AWS and Databricks.
* Performed ETL operations from source systems to Amazon S3 and Redshift using a combination of AWS Glue, T-SQL, Spark SQL, and U-SQL with Azure Data Lake Analytics. Ingested data into Azure services and processed it in Azure Databricks.
* Tested Apache Beam for building high-performance batch and interactive data processing applications on Glue, Databricks, and Redshift jobs.
* Explored and optimized algorithms in Hadoop using Spark context, Spark SQL, PostgreSQL, Scala, DataFrames, and Amazon EMR and Databricks.
* Set up data pipelines using Glue, Talend, Redshift, and Databricks based on data load sizes.
* Implemented real-time analytics on data using DynamoDB with Glue and Databricks.
* Designed column families in DynamoDB, ingested data from RDBMS, performed transformations, and exported data to DynamoDB and Databricks.
* Led testing efforts across a broad technology landscape, including Unix, AngularJS, AWS, AWS CloudWatch Logs, Cucumber JVM, DynamoDB, GitHub, Bitbucket, SQL, NoSQL databases, APIs, Java, Jenkins, and Databricks.

Bottom of Form

**Environment**: Hadoop (HDFS, MapReduce), Databricks, Spark, Talend, Impala, Hive, GCP, PostgreSQL, Jenkins, Nifi, Scala, Mongo DB, Cassandra, Python, Pig, Sqoop, Hibernate, spring, Oozie, AWS Services EC2, S3, Autoscaling, Azure, Elastic Search, DynamoDB, UNIX Shell Scripting, TEZ.

**BMO USA,** [**Chicago, IL**](https://en.wikipedia.org/wiki/Chicago,_Illinois) **May 2017 – Sep 2018**

**Data Engineer**

**Responsibilities:**

* Gathered data and business requirements from end users and management. Designed and built data solutions to migrate existing source data from the Data Warehouse to AWS S3 (Big Data). Utilized Spark and Databricks for large-scale data processing.
* Analyzed huge volumes of data and devised simple and complex Hive and SQL scripts to validate data flow in various applications. Performed report validation using Cognos. Made use of AWS Glue and Databricks for validating Data Profiling & Data Lineage.
* Devised PL/SQL statements - Stored Procedures, Functions, Triggers, Views, and packages. Utilized AWS RDS (Relational Database Service) and indexing, aggregation, and materialized views to optimize query performance. Incorporated Spark and Databricks for enhancing ETL processes.
* Created reports using AWS QuickSight and other tools for data validation. Employed Spark and Databricks for generating and processing data for reporting purposes.
* Involved in creating Tableau dashboards using stack bars, bar graphs, scatter plots, geographical maps, Gantt charts, etc., using the "Show Me" functionality. Developed dashboards and stories as needed using Tableau Desktop and Tableau Server. Utilized Spark and Databricks for pre-processing data to ensure accuracy in Tableau visualizations.
* Performed statistical analysis using SQL, Python, R Programming, and Excel. Integrated Spark and Databricks for large-scale statistical computations and Scala for implementing complex data analysis workflows.
* Worked extensively with Excel VBA Macros and Microsoft Access Forms. Employed Spark and Databricks to handle large data sets, ensuring seamless integration with VBA and Access solutions.
* Imported, cleaned, filtered, and analyzed data using tools such as SQL, Hive, and Pig. Enhanced data processing capabilities by integrating Spark and Databricks for large-scale data manipulation.
* Used Python & SAS to extract, transform, and load source data from transaction systems, generated reports, insights, and key conclusions. Incorporated Spark and Databricks for efficient ETL processes and data handling.
* Developed storytelling dashboards in Tableau Desktop and published them on Tableau Server, allowing end users to understand the data on the fly with the usage of quick filters for on-demand information. Utilized Spark and Databricks for data pre-processing to optimize dashboard performance.
* Analyzed and recommended improvements for better data consistency and efficiency. Leveraged Spark and Databricks to identify and address data quality issues and streamline data processing workflows.
* Designed and developed data mapping procedures for ETL - Data Extraction, Data Analysis, and Loading processes for integrating data using R programming. Incorporated Spark and Databricks to enhance ETL pipelines and data integration processes.
* Effectively communicated plans, project status, project risks, and project metrics to the project team and planned test strategies by project scope. Used Spark and Databricks to support data-driven decision-making and project insights.
* Data ingestion from AWS Glue, AWS Data Pipeline, and Databricks from Oracle database. Employed Spark and Databricks for efficient data ingestion and transformation tasks.
* Responsible for wide-ranging data ingestion using AWS Glue, AWS S3 commands, and Databricks. Accumulated partitioned data in various storage formats like text, JSON, Parquet, etc. Involved in loading data from the Linux file system to AWS S3. Used Spark and Databricks for optimizing data ingestion processes.
* Stored data files in AWS S3 buckets daily. Used AWS Data Pipeline, AWS Glue, and Databricks to develop and maintain AWS cloud-based solutions. Integrated Spark and Databricks for data processing and management.
* Experience working with AWS services for storage and handling terabytes of data for customer BI reporting tools. Utilized Spark and Databricks for scalable data processing within AWS environments.
* Experience with dimensional modeling (Star schema, Snowflake schema), transactional modeling, and SCD (Slowly Changing Dimension). Applied Spark and Databricks for dimensional data processing and Scala for implementing complex data models.
* Devised PL/SQL Stored Procedures, Functions, Triggers, Views, and packages. Utilized AWS RDS and indexing, aggregation, and materialized views to optimize query performance. Integrated Spark and Databricks for enhancing data transformation processes.
* Hands-on experience with AWS services including AWS Redshift, AWS Glue, AWS Lambda, Amazon EMR, and Databricks. Utilized Spark and Databricks for big data processing and Scala for managing data workflows.
* Implemented Apache Airflow for authoring, scheduling, and monitoring data pipelines on AWS and Databricks. Leveraged Spark and Databricks for orchestrating complex data pipelines and ensuring efficient data processing.
* Worked with Confluence and Jira, skilled in data visualization using Matplotlib and Seaborn libraries. Integrated Spark and Databricks for supporting data-driven visualizations and insights.
* Experience implementing machine learning back-end pipelines with Pandas, NumPy on AWS SageMaker and other AWS ML services. Utilized Spark and Databricks for scalable machine learning workflows and Scala for advanced data processing.

Top of Form

Bottom of Form

**Environment**: Hive, AWS, Hadoop, HDFS, Python, PL/SQL, SQL, Python, R Programming, Apache Airflow, NumPy, Pandas, Jira, PIG, Tableau, Spark, Linux, NumPy.

**Zerodha, Bengaluru, India** **May 2013 - July 2016**

**Java Developer**

**Responsibilities:**

* Managed Full Implementation Lifecycle: Specialized in designing and executing custom MapReduce and Hive applications within cloud environments, leveraging services like Azure Synapse Analytics, Azure Data Lake Storage, and Databricks.
* Data Querying and Analysis: Extensively utilized Hive queries, Azure Synapse SQL, and Databricks SQL for querying and analyzing data, including string searches and complex transformations.
* Spark Integration: Implemented Apache Spark solutions using Python and Spark SQL on Azure Synapse Analytics and Databricks for accelerated data processing and performance enhancement.
* Real-time and Batch Processing: Employed Spark on Databricks for interactive queries, streaming data processing, and integration with NoSQL databases such as Azure Cosmos DB.
* Data Connectivity: Used the Cosmos DB Spark Connector on Databricks to efficiently load and extract data from Cosmos DB, facilitating seamless data operations.
* Data Transfer and Migration: Managed data movement between HDFS and SQL databases using Azure Data Factory and Databricks, orchestrating data workflows for efficient processing.
* Advanced Data Processing: Implemented advanced data processing techniques, including in-memory computing with Spark on Databricks for text analytics and large-scale transformations.
* Data Warehouse Migration: Executed a migration strategy to transition from an on-premises SAP Data Warehouse to Azure Synapse Analytics and Databricks, optimizing data storage and analytics capabilities.
* Dynamic Data Loading: Used Data Factory and Databricks to dynamically load data into clusters from various sources, including relational databases and flat files.
* Stream Processing: Leveraged Spark Streaming on Databricks to process streaming data in real-time, enabling timely insights and data-driven decision-making.
* Data Integration: Loaded data from multiple sources (SQL, DB2, Oracle) into HDFS, Hive tables, and Databricks for comprehensive analysis and reporting.
* Event Processing: Conducted real-time event processing by integrating Azure Stream Analytics with Event Hubs and Databricks to process data from multiple servers.
* Reusable Components: Created and managed reusable data integration components like Data Flows and other transformations in Databricks for efficient development.
* Informatica Integration: Worked with Informatica tools (Source Analyzer, Mapping Designer, Mapplet) for data integration and transformation in cloud contexts, integrating with Databricks for enhanced capabilities.

**Environment**: Hadoop, HDFS, Hive, MapReduce, Impala, Sqoop, SQL, Informatica, Python, Flume, PySpark, Yarn, Pig, Oozie, Linux, AWS, Tableau, Maven, Jenkins, Cloudera, SAS (BI & DI), PL/SQL, Autosys, Oracle, SQL Server, No SQL, Teradata.

**EDUCATION**

* BTech | Computer Science | MVGR College of engineering, May-2013.
* Master of Science | Computer Science | Saint Louis University, Dec-2017