Akhilesh. T

Phone: +1 834-410-3706

Email Id: kalyang@lgltechnologies.com

**Professional Summary:**

Around 9 years of working experience in Data Engineering as well as in Data Analysis/Modelling including planning, analysis, design, development, testing and implementation.

* Hands On experience on developing UDF, DATA Frames and SQL Queries in Spark SQL.
* Experience in integrating Kafka with Spark streaming for high-speed data processing.
* Good understanding and exposure to Python programming.
* Good experience in conducting Joint Application Development (JAD) sessions with SMEs, Stakeholders.
* Hands on experience in writing Ad - hoc Queries for moving data from HDFS to HIVE and analysing the data using HIVE QL.
* Strong Experience in migrating data using Sqoop from HDFS to Relational Database Systems and vice-versa.
* Have good experience working with Azure BLOB and Data Lake storage and loading data into Azure SQL Synapse analytics (DW).
* Good experience in technical consulting and end-to-end delivery with data modelling, data governance.
* Experienced with cloud platforms like Amazon Web Services, Azure, Databricks (both on Azure as well as AWS integration of Databricks).
* Spark for ETL follower, Databricks Enthusiast, Cloud Adoption & Data Engineering enthusiast in Open-source community.
* Compute engine, cloud load balancing, cloud storage, cloud SQL, stack driver monitoring, and cloud deployment manager are some of the Google Cloud Platform (GCP).
* Development, Implementation, Deployment, and Maintenance of analytics technologies such as HDFS, HDF5 MapReduce, Spark, Scala, Yarn, Kafka, PIG, HIVE, Sqoop, Flume, Oozie, Impala, and HBase. Airflow, Zookeeper, Ambari, NiFi, AWS, Azure, Google Cloud Platform.
* Used cloud shell SDK in GCP to configure the services Data Proc, Storage, Big Query, coordinated with team and Developed framework to generate Daily ad hoc reports.
* Responsible for managing the GCP services such as Compute Engine, App Engine, Cloud Storage, VPC, Load Balancing, Big Query, Firewalls, and Stack Driver.
* Proficient with Azure Data Lake Services (ADLS), Databricks & I Python Notebooks formats, Databricks Delta lakes & Amazon Web Services (AWS).
* Solid Excellent experience in creating cloud-based solutions and architecture using Amazon Web services and Microsoft Azure.
* Experience in configuring the Cloud Firewall rules to allow the inbound traffic to GCP GKE cluster.
* Deployed the Micro services with Spring Boot on GCP / GKE Kubernetes Cluster configured with master nodes and worker nodes.
* Experience in conducting Joint Application Development (JAD) sessions with SMEs, Stakeholders and other project team members for requirements gathering and analysis.
* Excellent proficiency in Software Development Life cycle (SDLC), Agile/Scrum and waterfall methodologies.
* Extensive experience in using ER modelling tools such as Erwin and ER/Studio, Power Designer.
* Experience in writing Build Scripts using Shell Scripts, MAVEN and using CI (Continuous Integration) tools like Jenkins.
* Experienced in creating shell scripts to push data loads from various sources from the edge nodes onto the HDFS.
* Good Understanding of Hadoop architecture and underlying framework including storage management.
* Experience in assisting the Deployment team in setting up Hadoop cluster and services.
* Experience in Storage, Querying, Processing and analysis of big data's by utilizing the quality components of Hadoop.
* Experienced working with JIRA for project management, GIT for source code management, Jenkins for continuous integration and Crucible for code reviews.

**Technical Skills:**

|  |  |
| --- | --- |
| Hadoop/Spark Ecosystem | Hadoop, CDH, MapReduce, Hive/impala, YARN, Kafka, Sqoop, Spark, Spark SQL, PY Spark, Kusto, Scala |
| Databases | SQL Server, PostgreSQL, HBase, Snowflake, Cassandra, MongoDB, Azure Cloud DB, Dynamo DB. |
| BI Tools | Business Objects XI, Tableau 9.1, Power BI |
| Query Languages | SQL, PL/SQL, T-SQL |
| Scripting Languages | Unix, Python, PY Spark |
| Reporting tools | Eclipse, Visual Studio, Net Beans, Azure Data bricks, UNIX Eclipse, Visual Studio, Net Beans, Junit, CI/CD, Linux, Google Shell, Unix, Power BI, SAS and Tableau |
| Google Cloud Platforms | GCP, Big Query, google Cloud, Data Flow, Cloud SQ |
| Cloud computing | AWS (Amazon Redshift, Athena, Glue, AWS lambda, EMR, S3), MS Azure (Azure blob storage, Azure Data Factory, Azure Synapse) |
| Data Modelling Tools | Erwin 9.7/9.6, ER Studio v17, and Power Designer |
| Methodologies | RAD, JAD, RUP, UML, System Development Life Cycle (SDLC), Agile, Waterfall Model |
| Data Warehousing | Informatica PowerCenter, Talend Data Integration, AWS Glue, Azure Data Factory |
| Operating System | Windows and Linux |

**Professional Experience:**

|  |
| --- |
| **Client: Advent Health, FL March 2023 - Present**  **Role: Sr. Data Engineer** |

**Responsibilities:**

* As a Senior Data Engineer understand current Production state of application and determine the impact of new implementation on existing business processes.
* Assisted in leading the plan, building, and running states within the Enterprise Analytics Team.
* Installed and configured various components of Hadoop ecosystem and maintained their integrity on Cloudera.
* Extensively involved in Cluster Capacity planning, Hardware planning, Installation, Performance Tuning of the Hadoop Cluster.
* Followed agile methodology and Scrum process.
* Designed and Configured Azure Cloud relational servers and databases analysing current and future business requirements.
* Perform analyses on data quality and apply business rules in all layers of data extraction transformation and loading process.
* Responsible for data governance rules and standards to maintain the consistency of the business element names in the different data layers.
* To meet specific business requirements wrote UDF’s in Scala and PY Spark.
* Design and implement end-to-end data solutions (storage, integration, processing, and visualization) in Azure.
* Exported the analysed data to the relational databases using Sqoop.
* Extracted and loaded data into Data Lake environment (MS Azure) by using Sqoop which was accessed by business users.
* Involved in creating Azure Data Factory pipelines.
* Built the data pipelines that will enable faster, better, data-informed decision-making within the business.
* Analyse, design and build Modern data solutions using Azure PaaS service to support visualization of data.
* Experience in Migrating the Legacy application into GCP Platform.
* Developed JSON Scripts for deploying the Pipeline in Azure Data Factory (ADF) that process the data using the SQL Activity.
* Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform and load data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards.
* Performed end-to-end delivery of PY spark ETL pipelines on Azure - data bricks to perform the transformation of data orchestrated via Azure Data Factory (ADF) scheduled through Azure automation accounts and trigger them using Tidal Schedular.
* Developed business intelligence solutions using SQL server data tools 2019 version and load data to SQL & Azure Cloud database.
* Developed Spark applications using PY Spark and Spark-SQL for data extraction, transformation and aggregation from multiple file formats for analysing & transforming the data to uncover insights into the customer usage patterns.
* For Log analytics and for better query response used Kusto Explorer and created alerts using Kusto query language.
* Experience in moving data between GCP and Azure using Azure Data Factory also building power bi reports on Azure Analysis services for better performance.
* Worked on creating tabular models on Azure analysis services for meeting business reporting requirements.
* Migrate data into RV Data Pipeline using Data Bricks, Spark SQL and Scala.
* Used Databricks for encrypting data using server-side encryption.
* Responsible for estimating the cluster size, monitoring and troubleshooting of the Spark Databricks cluster.
* Used Apache airflow in GCP composer environment to build data pipelines and used various airflow operators like bash operator, Hadoop operators and python callable and branching operators.
* Developed dashboards and visualizations to help business users analyse data as well as providing data insight to upper management with a focus on Power BI.
* Used JIRA to track issues and Change Management
* Involved in creating Jenkins jobs for CI/CD using GIT and Maven.
* Involved in daily Scrum meetings to discuss the development/progress and was active in making scrum meetings more productive.
* Extensively working on Spark using Python for testing and development environments.
* Executed tasks for upgrading cluster on the staging platform before doing it on production cluster.
* Working with production support team to provide necessary support for issues with CDH cluster and the data ingestion platform.

**Environment:** Hadoop 3.3, Agile/Scrum, Spark 3.1, Azure, ADF, Scala, data bricks, CI/CD, PY Spark, SQL, Python 3.9, PY Spark 3.0, SQL, NiFi, GCP, Apache, XML, Kafka 3.0, SQL Server DB 2019, CDH, Sqoop 1.4.7, Power BI.

|  |
| --- |
| **Client: Fidelity Investments, Alabama April 2022 – Feb 2023**  **Role: Sr. Data Engineer** |

**Responsibilities:**

* As a Data Engineer helped developers automatically build and deploy software into production multiple times a day safely while maintaining compliance in a highly regulated financial industry.
* Interacted with business partners, Business Analysts and product owner to understand requirements and build scalable distributed data solutions using Hadoop ecosystem.
* Worked with data governance team on data protection regulation projects.
* Installed and configured Hadoop Ecosystem components.
* Installed and configured Hive and written Hive UDFs.
* Provided a streamlined developer experience for delivering small serverless applications to solve business problems The Platform is a Lambda-based platform.
* Implemented a 'serverless' architecture using API Gateway, Lambda, and Dynamo DB.
* Integrated AWS Dynamo DB using AWS lambda to store values of items and backup to Dynamo DB.
* Deployed AWS Lambda code from Amazon S3 buckets.
* Responsible for managing the GCP services such as Compute Engine, App Engine, Cloud Storage, VPC, Load Balancing, Big Query, Firewalls, and Stack Driver.
* Created a Lambda Deployment function, and configured it to receive events from your S3 bucket.
* Wrote scripts in Hive SQL for creating complex tables with high performance metrics like partitioning, clustering, and skewing, Migrated an Oracle SQL ETL to run on google cloud platform (GCP) using cloud data proc & big query, cloud pub/sub for triggering the airflow jobs.
* Performed information purging and applied changes utilizing Data bricks and Spark information analysis.
* Enterprise data from Big Query, can work parallelly in both GCP and Azure Clouds coherently.
* Work related to downloading big Query data into pandas or Spark data frames for advanced EL capabilities, built reports for monitoring data loads into GCP and drive reliability at the site level.
* Extensively utilized Data bricks notebooks for interactive analysis utilizing Spark APIs.
* Designed the data models to be used in data intensive AWS Lambda applications which are aimed to do complex analysis creating analytical reports for end-to-end traceability.
* Created AWS Lambda functions using python for deployment management in AWS.
* Designed, investigated and implemented public facing websites on Amazon Web Services and integrated it with other applications infrastructure.
* Created different AWS Lambda functions and API Gateways, to submit data via API Gateway that is accessible via Lambda function.
* Designed and Developed ETL Processes in AWS Glue to migrate data from external sources like S3, ORC/Parquet/Text Files into AWS Redshift.
* Automated Datadog Dashboards with the stack through Terraform Scripts.
* Created external tables with partitions using Hive, AWS Athena and Redshift.
* Developed the PY Spark code for AWS Glue jobs and for EMR.
* Wrote AWS Glue scripts to extract, transform load the data.
* Integrated lambda with SQS and Dynamo DB with step functions to iterate through list of messages and updated the status into Dynamo DB table.
* Designed and Developed ETL jobs to extract data from Salesforce replica and load it in data mart in Redshift.
* Responsible for Designing Logical and Physical data modelling for various data sources on Redshift.
* Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in in Azure Data bricks.
* Managed multiple Kubernetes clusters in a production environment.
* Participated in daily stand-ups, bi-weekly scrums and PI panning.
* Managed individual project priorities, deadlines and deliverables.
* Connected to Amazon Redshift through Tableau to extract live data for real time analysis.
* Wrote reports using Tableau Desktop to extract data for analysis using filters based on the business use case.
* Created Tableau dashboards, datasets, data sources and worksheets.
* Wrote and deployed data applications within a streaming architecture to ingest, transform, and store data to be used in Machine Learning Models, developing prototypes quickly.
* Worked with application teams to install operating system, Hadoop updates, patches, version upgrades as required.

**Environment:** Hadoop 3.0, Hive, Dynamo DB, AWS Lambda, data bricks, AWS, Azure, Amazon S3, AWS Glue, Redshift, Terraform, Spark, GCP, Hive, SQL, Athena, Big Query, Redshift, AWS Glue, PY Spark, SQS, Salesforce, Kubernetes, ETL, Tableau.

|  |
| --- |
| **Client: Robert Half, CA Nov 2019 – March 2022**  **Role: ETL Developer** |

**Responsibilities:**

* Designed and implemented reporting Data Warehouse with online transaction system data using Informatica PowerCenter and Azure Synapse Analytics.
* Developed and maintained data warehouse for PSN project.
* Created reports and publications for Third Parties using Power BI for Royalty payments.
* Managed user accounts, groups, and workspace creation for different users in Azure Synapse Studio.
* Wrote complex UNIX/Windows scripts for file transfers and email tasks from FTP/SFTP using Azure Logic Apps.
* Worked with PL/SQL procedures and used them in Stored Procedure Transformations.
* Worked extensively on Oracle and SQL Server. Wrote complex SQL queries to query ERP system for da Experience in google cloud platform (GCP) cloud by provisioning compute engine, cloud load balancing, cloud storage, cloud SQL, stack driver monitoring components using the Terraform GCP Foundation modules analysis purpose.
* Migrated ETL code from Talend to Informatica. Involved in development, testing, and post-production for the entire migration project.
* Tuned ETL jobs in the new environment after fully understanding the existing code.
* Maintained Talend admin console and provided quick assistance on production jobs.
* Involved in designing Power BI reports and dashboards.
* Built ad-hoc reports using stand-alone tables in Power BI.
* Created publications which split into various reports based on specific vendor using Power BI.
* Wrote custom SQL for some complex reports in Power BI.
* Worked with business partners internal and external during requirement gathering.
* Worked closely with Business Analyst and report developers in writing the source to target specifications for Data warehouse tables based on the business requirement needs.
* Exported data into Excel for business meetings, which made the discussions easier while looking at the data.
* Performed analysis after requirements gathering and walked the team through major impacts.
* Provided and debugged crucial reports for finance teams during the month-end period using Power BI.
* Addressed issues reported by Business Users in standard reports by identifying the root cause.
* Get the reporting issues resolved by identifying whether it is report-related issue or source-related issue.
* Creating Ad hoc reports as per user needs using Power BI.
* Investigating and analysing any discrepancy found in data and then resolving it.

**Environment**: Informatica PowerCenter 10.x, Azure Synapse, GCP, Cloud SQL, Analytics, Talend Integration suite, Power BI, Oracle 12c, SQL Server 2019, Azure Logic Apps, JIRA.

|  |
| --- |
| **Client: In Rhythm Solutions, India Aug 2017 – Oct 2019**  **Role: Data Engineer** |

**Responsibilities:**

* Gathered business requirements and prepared technical design documents, target to source mapping document, mapping specification document.
* Extensively worked on ETL tools such as Informatica Power centre or Talend to extract, transform, and load data from various sources into data warehouses.
* Developed and maintained ETL workflows and mappings for data integration and data quality.
* Optimized query performance by working with indexing, partitioning, and other database tuning techniques.
* Developed complex data transformations using SQL, PL/SQL, and other programming languages.
* Utilized cloud-based technologies such as AWS S3, Redshift, and Snowflake for data storage and processing.
* Designed and implemented data quality checks and data cleansing routines to ensure accuracy and consistency of data.
* Developed and maintained shell scripts for file transfer automation and scheduling ETL workflows using tools such as Control-M or Airflow.
* Worked with business users and data analysts to understand data requirements and provide data solutions that meet their needs.
* Performed testing and debugging of ETL workflows and provided necessary reports to the business users.

**Environment**: ETL tools such as Informatica Power centre, Talend, AWS S3, Redshift, Snowflake, SQL, PL/SQL, Control-M, Airflow, and various database technologies.

|  |
| --- |
| **Client: Smart Bridge, Hyderabad, India April 2015 – July 2017**  **Role: ETL Developer/ Data ware House** |

**Responsibilities:**

* Collaborated with business stakeholders to gather requirements and documented them for project development.
* Conducted design reviews and ETL code reviews with teammates.
* Developed ETL mappings using Informatica to extract, transform and load data from various sources such as Relational databases, Flat files, XML files, and APIs into the target Data Warehouse.
* Implemented complex Informatica transformations to transform data as per business rules and requirements.
* Created data mappings in Informatica to extract data from Sequential files and other non-database sources.
* Extensively worked on UNIX Shell Scripting and Python scripting for file transfer automation, error logging, and other automation tasks.
* Scheduled processes using job scheduling tools like Autosys, Control-M, and Airflow.
* Performed various types of testing, including Unit, Integration, System, and Acceptance testing of various ETL jobs.
* Optimized ETL jobs and SQL queries for performance tuning and data processing.
* Provided production support and troubleshooting for ETL issues and worked with other teams to resolve them.
* Perform day-to-day operation and troubleshooting of VMs and Docker swarm in GCP.
* Designed and developed data models and dimensional models using 3NF, Star and Snowflake schemas for OLAP and Operational data store (ODS) applications.
* Implemented ETL solutions for cloud platforms such as AWS, Azure and Google Cloud using cloud-based ETL tools like AWS Glue, Azure Data Factory, and Google Cloud Dataflow.
* Extensively worked with Big Data technologies such as Hadoop, Spark, Hive, and Pig for analysing and processing large volumes of data.
* Worked with containerization technologies such as Docker and Kubernetes for deploying ETL jobs and other data processing tasks.
* Collaborated with data scientists and machine learning engineers to design and develop data pipelines for machine learning and predictive analytics models.
* Provided technical leadership and guidance to junior team members and mentored them on best practices and industry standards for ETL development and data warehousing.

**Environment**: Informatica Power Centre, Oracle, SQL Server, GCP, Python, Google Cloud, UNIX Shell Scripting, Job Scheduling Tools (Autosys, Control-M, Airflow), Data Warehousing concepts.