Srilekha Kachakayala

+1 704 837 8374

Professional Summary

* Over 8+ years of experience in data engineering, specializing in big data implementations and Spark technologies.
* Proficient in various big data ecosystems such as Hadoop, Spark, Airflow, Snowflake, Teradata, Flume, and Kafka.
* Fluent in programming languages including Scala, Java, Python, SQL, T-SQL, and R.
* Successfully migrated on-premises applications to cloud platforms like AWS and Google Cloud.
* Extensive experience in cloud services such as EC2, S3, AWS Glue, Google Cloud Dataflow, and Azure Data Factory.
* Designed and implemented ETL processes using Informatica, Talend, SSIS, AWS Glue, and Azure Data Factory.
* Collaborated with cross functional teams to gather requirements, design ETL solutions, and deliver high-quality data integration solutions using Informatica.
* Proficient in analyzing both relational and non-relational databases, extracting insights, and providing valuable data-driven recommendations.
* Utilized scripting techniques to automate processes and optimize data workflows within Power BI, increasing efficiency and enabling seamless data visualization and analysis.
* Expert in developing comprehensive reports, KPIs, metrics, and attributes.
* Implemented data governance practices in the cloud to ensure data quality, integrity, and compliance.
* Proficient in data analysis and visualization using Tableau, Power BI, or Data Studio, with advanced analytics using Spark and Python.
* Worked in Agile environments, collaborating with cross-functional teams, and participating in integration testing and resolution.
* Experience in Big Data ecosystems using Hadoop, Pig, Hive, HDFS, MapReduce, Sqoop, Storm, Spark, Airflow, Snowflake, Teradata, Flume, Kafka, Yarn, PyCharm SQL server.
* Used Terraform for creating multiple EC2 instances & EBS volumes as per instance type in different availability zones & multiple regions in AWS.
* Developed and optimized stored procedures and scheduled jobs to automate data processing tasks, improving operational efficiency.
* Experience working within AWS cloud infrastructure and upgrades and modifications to the SaaS platform to minimize service.
* Implemented AWS S3 and GCP Cloud Storage for robust and reliable data storage solutions.
* Experience in developing and setting up Azure Logic Apps for workflow automation and integration activities.
* Demonstrated containerization techniques using Docker and Kubernetes, facilitating streamlined and efficient deployment processes for data transformation pipelines and applications
* Best practices ongoing support for internally developed processes to advance business, and a diverse array of Python packages.
* transferring code between different SQL tools, ensuring accuracy and efficiency in tasks across platforms.
* Proficient in utilizing statistical techniques, particularly regression modeling, to uncover valuable patterns from datasets.
* Proficient in regression analysis to inform decision-making by assessing variable impacts as a data analyst.
* Expert Troubleshooting Skills employing advanced techniques to ensure data accuracy and reliability. debugging Knowledge
* utilizing GraphQL to optimize data retrieval from servers, by query language and runtime for executing queries against APIs.
* Leveraged Kubernetes clusters in an AWS data engineering role to design, deploy, and manage containerized applications, ensuring scalability, availability, and security in dynamic environments.
* integrated Azure Cosmos DB and MongoDB, leveraging its NoSQL capabilities to support the storage and retrieval of various data types.
* detail-oriented monitoring techniques, including HTTP data tracking, to ensure data integrity and optimize system performance.
* Developed dashboards and visualizations to help business users analyze data as well as provide data insight.
* Skilled in Microsoft products like SQL Server Reporting Services (SSRS) SQL query, and Power BI.
* Experienced in leveraging data profiling tools and techniques to uncover patterns and trends, optimizing data quality and usability as a data analyst.
* Creating innovative data visualizations using Oracle Business Intelligence (OBI) to effectively convey insights and facilitate better decision-making processes.
* Utilized logical reasoning techniques to query, inspect, cleanse, and profile data, ensuring data quality and integrity across various formats.

**Technical Skills**

|  |  |
| --- | --- |
| **Databases** | SQL Server 2016, SQL Server 2018, SQL Developer |
| **Big Data Eco Systems** | MapReduce, YARN, Kafka, Spark, Airflow, MongoDB, Cassandra, HBase, Hive, Spark, Terraform, Snowflake |
| **Data Visualization Tools** | Power BI, Tableau, Data Studio |
| **Version Control** | Jira, Git Hub, ServiceNow, SAP |
| **Languages** | T-SQL, PL-SQL, Python, XML HTML C, C++ Java, SQL, JavaScript |
| **Others** | Power shell script Beginner, Data Base Admin Beginner. Kibana, MS access,  PyTorch, and TensorFlow |
| **Tools** | SQL Developer, SQL Server Management Studio, Informatica Power Center, Autosys SSIS |
| **Cloud Platforms** | AWS, (EC2, S3, RDS, Red Shift, Glue) Azure, ( ADF, DATA LAKE STORAGE GEN2, DATABRICKS), GCP, pubsub Big Query. |
| **Soft Skills** | time management, problem-solving, communication skills statistics Knowledge, Accounting. |

# Work Experience

**Client Wells Fargo bank Charlotte NC United States June 2023 – Present**

# Role: Business Data Analyst/Management

* Version control and configuration management of code provided by all the vendor dev teams.
* Creating queries for Adding, Updating, and Deleting the data from the Database as per Business requirements. Learning the things that go under daily Scrum calls.
* Developed comprehensive reports, KPIs, metrics, and attributes proficiently.
* Collaborated closely with the Data Management team to monitor and enhance data quality.
* Designing and optimizing data storage solutions, ensuring efficient access and retrieval of critical information in data warehousing.
* Unit Test Python library was used for testing many programs on Python and other codes.
* Created Data Mapping for ETL and application support, DFD, ERD, mapping documents, metadata, and DDLL as required.
* Updated and manipulated content and files by using Python scripts.
* Conducted test strategies, and test cases and performed Unit Testing to identify and resolve functional and Test Plan
* Used several Python libraries like Python, NumPy, and Matplotlib.
* Demonstrated ability to manage loan portfolios within LOANIQ, including loans.
* Deploying code in the Production environment and live production run
* Demonstrated effective verbal and written communication skills, coupled with compelling storytelling abilities.
* Engaging in DQ issue remediation efforts, including root cause analysis, and confirming the accuracy of data assets.
* Deploying code in different environments and doing smoke runs and end-to-end runs of builds.
* Leveraged AWS EMR and GCP Dataproc for distributed processing of big data workloads.
* Leveraged Aurora, DynamoDB, Lambda, and Kinesis to architect and implement scalable data solutions, ensuring efficient processing and storage.
* Developed analytics solutions capable of processing both streaming and batch-based data sets, delivering comprehensive reporting capabilities to stakeholders.
* Utilized AWS Glue and GCP Dataflow for scalable data processing in data engineering tasks.
* Coordinating with dev and QA teams to mitigate issues. Administrative activities related to the server.
* Writing shell scripts and code for automation of the build process. Working with the quantitative modeling team.
* Designed and implemented database solutions in SQL Data Warehouse, and Analysis.
* Major and Minor enhancements and Converting Oracle packages to SQL server. Using Jira as an issue management studio Performing testing on UAT, SIT, and PROD. attention to detail.
* For business growth Utilized data profiling and analysis skills to extract actionable insights, enhancing decision-making processes.
* Exhibited strong problem-solving, analytical, and research capabilities.
* Developed a visionary roadmap for the future of data utilization, focusing on optimizing stakeholder interactions and streamlining analytical processes for enhanced effectiveness.
* Led initiatives to identify and rectify data discrepancies, ensuring data consistency and integrity throughout the organization's datasets.

**Environmental Skills:** – business strategies Documentation, Data Modelling Tool, Oracle Service, SQL, T-SQL, Teradata Repository Manager, Workflow Manager, Teradata coding, Excel, Oracle, PL/SQL, communication skills, analysis, data strategy, strategic Methods, machine learning. analytical skills, finance knowledge, self-motivation, Python scripting, and Problem-solving skills, Microsoft Excel, pivot table.

**Client Huntington Bank – Columbus Ohio United States FEB’2022 – May’2023**

# Role:: GCP Data Engineer

* working with product teams to create various store-level and supporting data pipelines written in GCP's big data stack.
* built data marts in big query for analytical reporting for the sales and products team.
* built data pipelines in airflow as a service using Bash and Bigquery operators.
* Built a program with Python with Apache beam framework and executed it in Cloud Dataflow to stream pub/sub-messages into big query tables.
* used Python and Apache Beam to execute it in cloud Dataflow and to run Data validation jobs between raw source files and big query tables.
* Used Cloud shell SDK in GCP to configure the services Data Proc, Storage, and Big Query.
* implemented Spark-based data processing solutions on the Google Cloud Platform, enhancing data analytics capabilities.
* loaded and transformed large sets of tech-structured, semi-structured datasets and analyzed them by running Hive queries.
* Using Cloud Functions to support the data migration from Big Query to the downstream applications.
* Conducted data cleaning processes, enhancing data accuracy by 15%.
* Implemented AWS S3 and GCP Cloud Storage for robust and reliable data storage solutions.
* Developed scripts using Spark to push the data from GCP to the third-party vendors using their API framework.
* Created Big Query jobs for loading the data into Big Query tables from data files stored in Google Cloud storage daily.
* Developed reports using Tableau, tracked the dashboards published and found potential future clients in the organization.
* Designed star schema in Big Query and architecting Staging layers of Data Lake.
* Used rest API with Python to ingest Data from and some other sites to BIGQUERY.
* Designed and optimized SQL queries and ETL operations, cutting down operation times by 35%.
* optimized data processing workflows with Google Cloud Data Fusion, complex ETL processes across heterogeneous data environments.
* Used and maintained Cosmos DB for storing catalog data and for event sourcing to process Streaming pipelines.
* Developed an understanding of SDLC for model production.
* Used Informatica Intelligent Cloud Services (IICS) in conjunction with SQL to design and implement data integration solutions, ensuring efficient data processing.
* Performed detailed analysis of raw data sources for data quality problems and model development needs.
* Implemented real-time data streaming pipelines using Google Cloud Pub/Sub, ensuring reliable and scalable ingestion of data from diverse sources.
* implemented scalable data pipelines on Snowflake, optimizing data ingestion, transformation, and processing for efficient analytics.

**Environmental Skills:** MYSQL, MS SQL Server, T-SOL, PL/SOL, Google Cloud Platform: CP Cloud Storage, Big Query, Composer, Cloud Datapost, Cloud SOL, Cloud Functions, Cloud Pub/Sub, Spark, Azure Storage, Power BI, Data Studio, Tableau Python, Pandas, NumPy, SciPy, Matplotlib, Shell/Bash, C#, R, Visio, Big Query GCP Cloud Storage, critical thinking.

**Client F P L, Juno Beach, FL United States OCT’2020- JAN’2022**

# Role: : Data Engineer

* Implemented end-to-end Azure data solutions, including migration strategies and traditional system transformation.
* Developed and executed complex data transformations and manipulations using Azure Data Factory and PySpark with Databricks.
* Utilized Kubernetes for managing deployment and scaling of data pipelines built with Apache Airflow or Luigi.
* Leveraged Azure Data Factory for seamless data ingestion from diverse source systems and orchestration of data integration cross functional upstream and downstream systems.
* Effectively utilized Git for version control and JIRA for issue management and project workflows, ensuring efficient coordination and task resolution
* Used and maintained Cosmos DB for storing catalog data and for event sourcing in order processing pipelines.
* Integrated Azure Active Directory authentication into every Cosmos DB request and showcased the
* Designed and implemented ETL pipeline in Azure cloud using Glue jobs to fetch customer data from API and process it into Azure SQL Data Brick.
* Implemented database solutions in Azure SQL Data Warehouse, Azure SQL, storage, and Azure Data Factory, SSIS, PowerShell.
* Developed applications for loading data from Azure storage blob and web API to Azure SQL, on daily loads.
* Implemented optimized data storage with ORC, Parquet, and text file formats, ensuring efficient compression and retrieval in large-scale data processing workflows.
* Developed dashboards and visualizations using SQL Server Reporting Services (SSRS) and Power BI for data analysis and insights for business users.
* Wrote optimized Hive queries for data analysis, meeting specified business requirements by creating and working with Hive tables.
* Proficiently performed ETL using Azure Databricks and migrated on-premises ETL processes to Azure Synapse Analytics.
* Implemented CI/CD pipelines for building and deploying projects in the Hadoop environment to ensure a streamlined development workflow.

**Environmental Skills:** Azure Cloud, Azure Data Factory (ADF v2), Azure functions Apps, Azure Data Lake, BLOB Storage, SQL server, Teradata Utilities, Windows remote desktop, UNIX Shell Scripting, AZURE PowerShell, Data bricks, Python, Erwin Data Modelling Tool, Azure Cosmos DB, Azure Stream Analytics, Azure Event Hub, Azure Database, Azure Data Factory, Azure Analysis.

**Client T-Mobile, Virginia United States JUL'2018 - SEP’2020**

# Role :AWS Data Engineer

* Conducted Gap Analysis to assess existing system infrastructure compatibility with new business requirements.
* Used MS Visio for business flow diagrams and defined the workflows.
* Performed data analysis for the existing data warehouse and changed the internal schema for performance.
* Used the user stories approach for describing the requirements using the user story template.
* application under development confines to the business requirement. Performed data mapping from source to target.
* Generated Use Case of business flow diagrams, and sequence diagrams to elucidate the business process from various stakeholder perspectives.
* Utilized AWS services including Aurora, DynamoDB, Lambda, and Kinesis to design and develop robust data pipelines, enabling real-time data processing and analysis
* Utilized SQL for database access and manipulation, engaging in design with SMEs to establish the business architecture.
* Collaborated with the QA team and developers to resolve issues and provide feedback for bug resolution.
* Performed end-to-end Architecture & implementation assessment of various AWS services like Amazon EMR, Redshift, and S3.
* Performed data blending and preparation with Alteryx and SQL, facilitating Tableau consumption, and publishing data sources to the Tableau server.
* Conducted troubleshooting and problem resolution for issues related to Proxies, Security Rules, and other QMC configurations.
* Developed and managed Proxies within the Qlik Management Console (QMC), ensuring secure access to Qlik Sense and Qlik Analytics Platform applications.
* Used Spark SQL for Scala & amp, a Python interface that automatically converts RDD case classes to schema RDD.
* Coded Teradata BTEQ scripts for data loading, transformation, and defect fixing including SCD 2 date chaining and cleanup.
* Imported data from various sources like HDFS/HBase into Spark RDD, utilizing PySpark for computations to generate the output response.
* Utilized AWS services Lambdas, EC2, S3, and Postgres RDS to build and manage scalable infrastructure, ensuring performance of data systems.
* Spearheaded collaborative efforts with stakeholders and executives to establish a comprehensive set of KPIs and reporting metrics, ensuring alignment with organizational objectives.
* Utilized coding skills with proficiency in Python, Java, and C/C++
* Developed reports, KPIs, metrics, and attributes proficiently provided to the Consumers.
* Maintained a close and strong working relationship with teammates and management staff to achieve expected results for the project team.
* Experienced in leadership roles and adept at fostering collaboration within teams to drive project success.
* Wrote manual Test Cases for checking the application. Used Agile methodology for repeated testing. Involved in Manual Testing by checking various validations Created, maintained, and executed manual test scripts.

**Environmental Skills:** AWS (EC2, S3, EBS, ELB, RDS, SNS, SQS, VPC, Redshift, Cloud information, CloudWatch, ELK Stack), Jenkins, Python, shell scripting, PowerShell, GIT, Microservice, Jira, Kubernetes, Docker, Web Logic, Maven, Web sphere, Unix/Linux, Nagios, Splunk, AWS Glue, Data Brick, understanding of AI/ML, Model Development and Deployment Life Cycle

**Client FEPOC, Virginia United States. JAN'2017- JUN'2018**

# Role: Data Analyst

* Performed data profiling in the source systems that are required for the RML System
* Documented the complete process flow to describe program development, testing, implementation, integration, and coding.
* Worked with project team representatives to ensure that logical and physical data models were developed in line with corporate standards and guidelines.
* Responsible for defining the functional requirement documents for each source to target interface.
* development and testing team. Reverse-engineered all the Source databases using Erwin.
* Successfully managed complex integrated landscapes and applications, ensuring master data configuration and data quality.
* Skilled in creating OLAP Cubes and data visualization with SSAS, Power BI
* Written complex SQL data mining.
* Worked in ETL methodology, including Informatica.
* Experienced in error handling and reporting with SSIS and MDX/SQL.
* Strong in data reconciliation, reporting, and metrics generation.
* Developed SQL and SSIS for database management and integration.
* Proficient in workflow management, backend testing, and SQL scripting.
* Governed the utilization of master data policies and procedures within established domains globally, including Customer, Material, Vendor, Cost, Price, and Discounts
* Designed and implemented data integration modules for Extract/Transform/Load (EL) functions.
* Involved in data warehouse design.
* Experience with various ETL, and data warehousing tools and concepts.
* Worked with project team representatives to ensure that logical and physical ER/Studio data models were developed in line with corporate standards and guidelines.
* Used data analysis techniques to validate business rules and identify missing data in the existing Amgen enterprise data warehouse.

**Environmental Skills**: Basic Hadoop, Python, HDFS, Hive, Scala, MapReduce, Agile, Cassandra, Kafka, AWS, Spark, ETL, Teradata, NoSQL, Java, Cassandra, Talend, LINUX, Kibana, HBase.MS SQL Server 2012 Microsoft Management Studio, Microsoft Visual Studio, MS SOL Server Integration Services (SSIS), 2012/2008R2/2008, R, SQL Server Reporting Services (SSRS), Power BI, Tableau, T-SQL, MDX, OLAP Cubes, DAX decision-making.

**Client – Genems Systems, Hyderabad. India JAN’2015-AUG'2016**

# Role: BI Data Designer /Analyst

* Responsible for providing data analysis that focuses on improving user experience and optimizing user retention.
* designed, and maintained MySQL databases, and developed pipelines with user-defined functions and stored procedures for daily reporting tasks.
* Designed and implemented metadata models, frameworks, dimensions, and reporting objects within Cognos Framework Manager, enhancing reporting capabilities.
* I crafted intricate ad-hoc MySQL queries featuring correlated subqueries and window functions.
* Developed dashboards in Tableau and presented data to track KPIs and product performance using data.
* from various sources such as MS Excel, AWS S3, AWS RDS, AWS Redshift, JSON, and XML.
* Spotted and analyzed trends in user activity data and identified user report insight to stakeholders.
* Performed statistical analysis such as hypothesis testing causal inference and Bayesian analysis.
* Communicated between the business, product manager, and data engineers to ensure the data quality.
* Built ETL pipelines to retrieve data from NoSQL databases and load aggregated data into the analytical platform.
* Managed data storing and processing using big data tools such as Hadoop, HDFS, Hive, and Spark.
* Automated data validation and cleaning processes, including deduplication and consistency checks, using Python scripts leveraging Pandas and Apache Airflow.
* Analyzed large-scale user log data and generated features for classification models using Spark SQL in Spark.
* Implemented scalable machine learning models Random Forest using Spark ML and Python to predict customers.
* Applied advanced classification models such as XG boost, SVM, and Neural Network to train data using Python packages such as Scikit-learn.
* Defined metrics to estimate the impact of new features and give recommendations for business decisions based on data analysis.

**Environmental Skills** MySQL, PostgreSQL, MS SQL Server, MongoDB, Hive, Presto, AWS Redis,

Big Query Tableau, Hadoop, Hive, Apache Airflow, Apache Spark, Flask, Apache Kafka, Jupyter, Notebook, Excel,

Jira, Git, Docker, Kubernetes.

**Education**

**Bachelor’s in computer science May 2014**

Jawaharlal Nehru Technological University Konda Gattu. (JNTUK)2014  **GPA**:**7.9/10**

Subjects**:** Advanced data structures, DBMS, Cloud computing, Algorithms Design Python, Java Data Analytics.

**Certifications**

* **Scrum Foundation Professional** Certified which includes knowledge of key definitions, roles, Scrum values, and principles.
* Completed **Database Design and Programming with SQL** which includes SQL stored procedures, HIVE, Hadoop