**Srikanth K**

**PROFESSIONAL SUMMARY:**

* IT professional with around 10 years of experience as a Big Data Engineer with expertise in designing data Pipelines, utilizing a robust skill set that includes Python, SQL, Spark, Hadoop, Snowflake development, GCP and AWS Cloud technologies.
* Proficient in building robust data pipelines and integrating data from diverse sources for analytics and reporting.
* Hands-on experience with big data technologies like Spark, PySpark, Databricks, and Snowflake for advanced data processing and reporting.
* Skilled in leveraging AWS and Azure cloud platforms to create centralized data platforms, optimize ETL workflows, and ensure data security and compliance.
* Expertise in delivering interactive dashboards, predictive models, and KPIs to support data-driven decision-making.
* Proven ability to improve operational efficiency, reduce costs, and support sustainability goals through innovative data engineering solutions.
* Adept at working with cross-functional teams to implement data solutions tailored to business needs.
* Built and maintained scalable data pipelines with dbt, integrating data from diverse sources like APIs, SaaS platforms, and on-premises databases.
* Proficient in Python, SQL, Tableau, and data modeling techniques, with a strong foundation in machine learning and advanced analytics.
* Recognized and Awarded as the best employee for delivering data-driven insights that enhance decision-making and streamline processes across supply chain functions.
* Proficient in SQL across various dialects, including MySQL, PostgreSQL, Redshift, SQL Server, and Oracle.
* Familiarity with AWS storage solutions (S3, EFS, Glacier) for large-scale data storage, retrieval, and compliance with security and governance standards.
* Ability to align technical solutions with supply chain workflows to enhance efficiency.
* Strong focus on data security, privacy, and governance, ensuring compliance with industry regulations and best practices in cloud and on-premises environments.
* Proficient in leveraging modern cloud platforms like AWS and Azure, along with data engineering tools such as Databricks, Snowflake, Apache Spark, and PySpark to drive operational excellence and deliver actionable insights.
* Skilled in end-to-end data engineering, including ETL pipeline development, data modeling (dimensional, star schema), and big data processing to ensure optimal performance and efficiency.
* Expertise in integrating and transforming data from diverse sources into centralized systems for improved data reliability and governance.
* Proven ability to design, implement, and optimize data pipelines and scalable solutions on Azure platforms, ensuring data-driven decision-making across the organization.
* Hands-on experience with cloud-based technologies like Azure Data Factory, Synapse Analytics, and AWS Glue for scalable and secure data processing.
* Demonstrated ability to deliver actionable insights through the design and development of real-time data processing systems and advanced analytics frameworks.
* Building real-time streaming solutions using Databricks, Azure Stream Analytics, and Kafka for IoT and logistics data.
* Creating interactive dashboards and visualizations with Power BI and Tableau to monitor key performance indicators and support data-driven decision-making.
* Proficient in managing end-to-end ETL processes, leveraging Azure Data Factory, Databricks, and Synapse Analytics for efficient and reliable data workflows.
* Extensive experience implementing robust data governance frameworks, access controls, and compliance measures to ensure data integrity and security within Azure ecosystems.
* Collaborating in Agile/Scrum environments, ensuring timely delivery of scalable and innovative solutions.
* Expertise in data modelling and performance tuning, coupled with a strong ability to collaborate with cross-functional teams to meet business intelligence and analytical needs. Committed to driving data-driven decision-making through innovative and efficient data engineering practices.
* Proficient in utilizing Azure Databricks for big data analytics and transformation. Experienced in writing and optimizing Spark jobs, managing clusters, and performing complex data transformations and aggregations using Scala, Python, and SQL.
* Skilled in creating and optimizing scalable ETL pipelines and data workflows to efficiently process and integrate data from various sources, including SQL databases, Azure Blob Storage, Azure Data Lake Storage, S3 buckets, and NoSQL databases like Cosmos DB and MongoDB.
* Designed and implemented scalable data architectures and pipelines on Google Cloud Platform (GCP), ensuring high availability and performance for large-scale datasets.
* Skilled in developing and managing big data ecosystems, including Delta Lake, HDFS, and Data Lake, for efficient storage and processing of structured and unstructured data.
* Expertise in creating interactive dashboards and visualizations using Power BI, Tableau, and other BI tools to support data-driven decision-making.
* Experienced in optimizing machine learning workflows on platforms like Databricks and Azure ML, deploying predictive models for demand forecasting, route optimization, and operational improvements.
* Strong knowledge of cloud-native architectures, leveraging tools like Azure Resource Manager and AWS CloudFormation to design cost-effective and scalable solutions.
* Skilled in orchestrating data workflows using tools like Apache Airflow, Azure Data Factory, and AWS Step Functions for seamless integration and automation.
* Proven ability to work in Agile/Scrum environments, ensuring timely delivery of high-quality data engineering solutions while collaborating with cross-functional teams.
* Expertise in implementing data warehouse solutions like Azure Synapse Analytics and AWS Redshift for enterprise-scale analytics and reporting.
* Successfully optimized data storage and query performance using Snowflake's multi-cluster architecture, time travel, and cloning features.
* Demonstrated ability to reduce costs and enhance efficiency through innovative data-driven strategies and supply chain optimization.
* Hands-on experience in feature engineering and data preprocessing using tools like Python Pandas for predictive and prescriptive analytics

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Big Data Tools** | Spark Streaming, HBase, HDFS, MapReduce, Hive, Pig, Kafka. |
| **ETL/Data warehouse Tools** | Informatica, Talend, Snowflake, Azure Data Factory, Azure Data Bricks, Teradata |
| **Programming Languages** | SQL, Python, PL/SQL,Bash Shell, C |
| **Methodologies** | RAD, JAD, System Development Life Cycle (SDLC), Agile, Waterfall. |
| **Cloud Platform** | Azure, AWS, GCP, Snowflake |
| **Databases** | MySQL, Oracle, SQL Server, Hive, PostgreSQL, MongoDB, Teradata. |
| **Tools** | Tableau, Power BI, Jenkins CICD |

**PROFESSIONAL EXPERIENCE:**

**AWS Data Engineer / Snowflake developer | NLR, Inc | Windsor - CT (Remote)** **Jan 2024 – Till Date**

**Description:** At NLR, the project’s main aim is to use data to strengthen our position as leaders in the universal waste sector. As data engineers, I concentrate on promoting operational excellence through the development of strong data pipelines that facilitate cost-effective recycling solutions, customer service optimization, and transparency and efficiency in compliance tracking.

**Responsibilities:**

* Designed and implemented scalable data pipelines to process and analyze universal waste management data, ensuring transparency in compliance tracking.
* Developed data integration workflows to consolidate information from transportation, warehouse, and administrative systems for real-time reporting.
* Analyzed and re-engineered existing SQL queries, addressing performance bottlenecks and reducing query execution time by 30%.
* Developed efficient data pipelines to integrate data from multiple sources, leveraging tools like Apache Airflow, dbt, and Snowpipe to streamline ingestion and transformation.
* Created and managed dbt project documentation, including lineage graphs and model descriptions, to enhance transparency and collaboration across teams.
* Developed and optimized complex SQL processes using Teradata, improving query performance by 20% and resolving critical issues in data workflows.
* Optimized data transformations to improve efficiency in monitoring recycling operations and cost-effective solutions.
* Built and maintained a centralized data platform to support detailed compliance reporting and customer service analytics.
* Leveraged technologies such as Spark, PySpark, and Databricks for big data processing, enabling insights into operational performance.
* Utilized cloud-based storage and computing solutions (AWS, Azure) to ensure scalable and secure data management.
* Collaborated with cross-functional teams to develop dashboards and visualization tools for monitoring key performance indicators (KPIs) across recycling processes.
* Applied data modeling techniques, including dimensional modeling and schema design, to optimize reporting and analytics for waste industry operations.
* Focused on enhancing data reliability, implementing monitoring systems to ensure accurate and timely compliance reporting.
* Contributed to sustainability goals by delivering data-driven insights to improve operational efficiencies and support eco-friendly practices.
* Designed and implemented data models in Snowflake, optimizing storage and query performance for compliance tracking and recycling analytics.
* Built ETL pipelines to seamlessly load and transform data into Snowflake, integrating data from multiple sources, including transportation and warehouse systems.
* Leveraged Snowflake's multi-cluster architecture to ensure high availability and scalability for processing large volumes of universal waste data.
* Enforced data quality standards using dbt's built-in testing framework, validating schemas, uniqueness, and relationships across datasets.
* Utilized Snowflake's time travel and cloning features to enable historical data analysis and streamline compliance audits.
* Designed and optimized data sharing mechanisms in Snowflake to facilitate collaboration between teams and external partners.
* Integrated Snowflake with visualization tools like Tableauto deliver actionable insights for tracking recycling efficiency and compliance metrics.

**Environment** AWS, IAM, S3, SNS, SQS, EC2, VPC, EMR, Snowflake, LAMBDA, Redshift, Glue, Crawler, Athena, Quicksight, IAM, CLI, Databricks, Data Lake, Dynamo DB, Apache Spark, Spark SQL, Talend.

**Data Engineer | Pilot Company Feb 2023 to Dec 2023**

**Description:** As a Data Engineer Led the design and management of end-to-end data pipelines at Pilot Company, optimizing data storage and management with diverse file formats and building robust systems to optimize fuel delivery, streamline services for motorists and professional drivers, and support community initiatives with actionable insights.

**Responsibilities:**

* Designed and built comprehensive end-to-end data pipelines using Google Cloud Dataflow, orchestrating data movement and transformation across multiple sources to streamline data integration and processing.
* Integrated on-premises data sources with Google Cloud Storage (GCS), BigQuery, Cloud SQL, and Cloud Spanner to centralize data for analysis and reporting.
* Managed and optimized data storage solutions with BigQuery, Firestore, and Cloud Storage, implementing best practices for data partitioning, clustering, and table optimization to enhance performance and scalability.
* Applied advanced data partitioning and clustering techniques within BigQuery to improve query performance and storage efficiency, supporting high-speed analytics and reporting.
* Implemented robust data security measures using Google Identity and Access Management (IAM) and VPC Service Controls to manage identities and secure data access, ensuring compliance with organizational and regulatory standards.
* Ensured adherence to data governance policies by deploying Data Loss Prevention (DLP), Cloud Key Management (KMS), and audit logs within GCP environments to protect sensitive information and maintain data integrity.
* Leveraged Dataproc and Cloud Data Fusion for big data processing and real-time analytics, writing and optimizing Apache Spark jobs, managing clusters, and creating interactive workflows to facilitate data-driven insights and machine learning workflows.
* Utilized Pub/Sub and Dataflow for real-time data ingestion and processing, enabling timely and actionable insights from streaming data sources.
* Monitored and tuned data pipelines and storage performance using Cloud Monitoring, Cloud Logging, and Error Reporting, proactively addressing issues and ensuring high operational efficiency.
* Implemented cost-saving strategies by leveraging sustained-use discounts, committed-use discounts, and dynamically scaling resources based on demand using Cloud Scheduler and Cloud Functions.
* Collaborated with cross-functional teams to gather and analyze requirements, delivering data solutions that align with business objectives and support strategic decision-making.
* Documented technical designs, data flows, and standard operating procedures to ensure knowledge sharing, process consistency, and effective project management.
* Designed and implemented an end-to-end data pipeline using Cloud Composer (managed Apache Airflow) for orchestrating and transforming data, seamlessly integrating various data sources for comprehensive retail sales analytics.
* Orchestrated the movement and integration of data from multiple sources, including Cloud SQL, Cloud Storage, BigQuery, and SFTP servers, through BigQuery Data Transfer Service to centralize and unify retail sales data for detailed analysis.
* Employed Cloud Build and Artifact Registry for CI/CD automation, streamlining the deployment of data pipelines and reducing manual interventions to ensure consistent and reliable deployments across development, staging, and production environments.
* Developed and managed data ingestion processes to collect and store raw sales data, ensuring data consistency and reliability across the entire pipeline, enabling accurate and timely analytics.
* Leveraged Dataproc to create and manage data transformation workflows, utilizing Spark for big data processing. Designed interactive workflows and Spark jobs to clean, aggregate, and transform retail sales data into valuable insights.

**Environment:** PySpark, Dataproc, BigQuery SQL, Looker Studio, Google Cloud Deployment Manager, Google Cloud Pub/Sub, Apache Spark, Spark SQL, Python, Google Cloud Dataflow, Google Sheets.

**Data Engineer** | **Amazon India (HYD13) | Hyderabad, India** **Aug 2020 – Aug 2022**  **Responsibilities**:

* Collaborated with Business Analysts and Engineers across departments to gather business requirements and identify workable items for development.
* Selected and generated data into CSV files and stored them into AWS S3 by using EC2 and then structured and stored in AWS Redshift.
* Developed pre-validation logic frameworks to streamline points redemption processes, ensuring compliance with data governance standards.
* Adoptability of best practices in reporting and analysis: data integrity, test design, analysis, validation.
* Hands on experience working with AWS EMR, EC2, S3, Redshift, DynamoDB, lambda, Athena, and Glue.
* Subscribing the Kafka topic with Kafka consumer client and processing the events in real time using spark.
* Collected data using Spark Streaming from AWS S3 bucket in near-real-time and performs necessary Transformations and Aggregation on the fly to build the common learner data model and persists the data in HDFS. Converting Hive/SQL queries into Spark transformations using Spark RDDs and Pyspark.
* Designed and implemented scalable ETL pipelines to integrated supply chain data from multiple sources, including inventory, North American logistics, and procurement systems.
* Developed custom reporting tools using SQL to visualize supply chain KPIs for business stakeholders.
* Engineered solutions to track and predict costs across supply chain operations, mitigating risks and enhancing profitability. Implemented cost-cutting strategies which resulted in saving around 30%.
* Ensured accurate tracking of compliance data (e.g., customs regulations, trade restrictions) in supply chain workflows.
* Developed machine learning models using Python and SQL for accurate demand forecasting, reducing overstock and stockout rates by 25%.
* Worked on Databricks and with delta tables. Used python to write a service which is event based using AWS Lambda to achieve real time data to One-Lake (A Data Lake solution in Cap-One Enterprise).
* Designed predictive models to identify and mitigate risks in supply chain operations, such as disruptions due to natural disasters or supplier delays.
* Collaborated with analytics teams to align dbt transformations with business requirements, ensuring data models meet stakeholder needs.
* Worked on blockchain-based supply chain solutions to enhance traceability and transparency of goods across logistics networks.
* Built data models to analyze shipping routes and costs, reducing transportation expenses by 20%.
* Skilled in data ingestion from multiple sources (API, cloud storage, databases) and integrating them into the CDP
* Performed data preprocessing and feature engineering for further predictive analytics using Python Pandas.
* Generated report on predictive analytics using Python and Tableau including visualizing model performance and prediction results.
* Utilized Agile and Scrum methodology for team and project management.
* Ensured accurate tracking of compliance data (e.g., customs regulations, trade restrictions) in supply chain workflows.
* Hands-on experience with platforms like Salesforce CDP, Adobe Experience Platform, or Segment.

**Environment:** Jira dashboard, Linux, AWS lambda, Python, GIT, Spark, Scala, MySQL, PostgreSQL, Snowflake data warehouse, AWS S3, AWS kinesis, Snow SQL, Microsoft Word/Excel, Flask, Snowflake, DynamoDB, Athena, Lambda, MongoDB, Pig, Sqoop, Tableau.

**Data Analyst | LI Blocks Pvt Ltd.| Hyderabad, India June 2016 to July 2020**

**Responsibilities:**

* Built real-time streaming applications using PySpark on Dataproc and Pub/Sub, replacing Kafka as a managed messaging service on GCP.
* Partnered with data scientists to design and optimize workflows using Vertex AI and BigQuery for machine learning model deployment and data analysis.
* Extracted and integrated data from databases, flat files, APIs, and other sources using Cloud Data Fusion, BigQuery Data Transfer Service, and custom workflows on Cloud Functions.
* Migrated SSIS packages to Cloud Dataflow and Cloud Composer (Apache Airflow) for scalable ETL workflows, implementing tasks such as conditional splits and derived column transformations.
* Used BigQuery ML and Vertex AI for AI-driven insights to forecast trends and improve decision-making based on historical data.
* Ensured data accuracy and quality by leveraging Cloud Data Loss Prevention for validation and BigQuery for schema enforcement.
* Analyzed customer profitability and lifetime value using BigQuery SQL for segmentation and prioritization of high-value users.
* Processed healthcare data workflows (EHR, HIE, and claims) using Cloud Healthcare API with support for HL7/FHIR standards.
* Ensured compliance with HIPAA and HITECH regulations by using Cloud KMS for encryption and Cloud IAM for role-based access control.
* Utilized Dataprep and Dataflow to clean and transform healthcare datasets, leveraging Python and R on Vertex AI Notebooks for statistical analysis.
* Built predictive models with BigQuery ML and Vertex AI to forecast healthcare utilization, identify at-risk patients, and optimize resource allocation.
* Integrated marketing data with BigQuery and Looker Studio to analyze channel contributions and optimize budgets.
* Architected data infrastructures tailored for Salesforce applications using BigQuery and orchestrated ETL workflows via Cloud Composer.
* Delivered OLTP/OLAP models on BigQuery for streamlined analytics, ensuring efficient reporting and querying.
* Generated interactive dashboards with Looker Studio and BigQuery, visualizing customer journeys and identifying bottlenecks.
* Applied statistical methods to patient data with BigQuery and Vertex AI, optimizing care quality and reducing costs.
* Coordinated with clinicians, IT, and business stakeholders using Google Workspace and GCP collaboration tools to deliver data-driven solutions.
* Extended reporting capabilities by linking BigQuery data with Google Sheets and Looker Studio for advanced analysis.

**EDUCATION**

* Bachelor of Technology - TKR College of Engineering and Technology, India
* Master of Science – Southern Illinois University Edwardsville, IL, USA

**CERTIFICATIONS**

* Microsoft Certified: Azure Data Engineer Associate
* AWS Certified Data Engineer – Associate
* Databricks Lake house