**Chaitanya Kalidindi**  **Data Engineer**

**Professional Summary:**

* 10 years of industry experience in IT, specializing in Data Engineering, Data Warehousing, and Big Data Development.
* Expert in designing and implementing scalable data ingestion pipelines using AWS Glue, Apache Spark, and Snowflake.
* Proficient in leveraging AWS Glue and Spark for distributed data processing and transformation tasks.
* Skilled in ensuring data quality and integrity through validation, cleansing, and transformation operations.
* Adept at designing cloud-based data warehouse solutions using Snowflake on AWS, optimizing schemas, tables, and views for efficient data storage and retrieval.
* Utilized Snowflake Clone and Time Travel features for data management and maintenance.
* Actively involved in the development, improvement, and maintenance of Snowflake database applications.
* Built Logical and Physical data models for Snowflake, incorporating necessary changes and defining roles and privileges for database access.
* Designed **Multi-cluster Snowflake warehouses with dynamic scaling**, optimizing **performance and cost-efficiency**
* Possess in-depth knowledge of Snowflake Database, Schema, and Table structures.
* Strong expertise in optimizing Spark jobs and leveraging AWS Redshift for big data processing and analytics.
* Performed performance tuning on AWS Redshift workloads, optimizing resource utilization and query response times.
* Enhanced query performance through SQL query optimization, proper indexing, and partitioning tables for improved AWS Redshift efficiency.
* Developed CI/CD frameworks for data pipelines and collaborated with DevOps teams for automated deployment.
* Proficient in scripting languages such as Python and Scala.
* **Implemented scalable ETL workflows in Azure Synapse Analytics and Azure Data Factory**, optimizing **data ingestion, transformation, and orchestration** for high-volume datasets.
* Extensive experience with Databricks, Airflow, Python, and PySpark, successfully developing and managing large-scale data pipelines.
* Optimized Hive query performance using bucketing and partitioning techniques, with extensive hands-on experience in tuning Spark jobs.
* Leveraged **Adaptive Query Execution (AQE),** bucketing and partitioning techniques to dynamically optimize query plans, reducing **shuffle overhead and execution time.**
* Highly proficient in Agile methodologies, including JIRA for project management and reporting.
* Integrated **Apache Iceberg** and **Delta Lake** for **ACID-compliant, versioned data lakes**, improving schema evolution and rollback capabilities.
* Knowledge in Kafka installation & integration with Spark Streaming.
* Developed **CDC (Change Data Capture) pipelines** using **Debezium, Kafka, and Snowflake Streams**, enabling **incremental data processing**.
* Loaded and transformed large sets of structured, semi structured and unstructured data in various formats like text, zip, XML and JSON.

**Technical Skills:**

|  |  |
| --- | --- |
| Azure Services | Azure Data Factory (ADF), Azure Synapse Analytics, Azure Databricks, Azure SQL Database, Azure Data Lake Storage (ADLS), Azure Synsapse, Azure Analysis Services, Azure Cosmos DB, Azure DevOps, Functions, Azure Monitor, Azure API Management, Azure Blob Storage. |
| AWS Services | AWS S3, AWS Redshift, AWS Glue, AWS SQS, AWS Athena, Lambda, EC2, RDS, CloudFront, , AWS CloudWatch, IAM ,EMR, DMS, Data Pipeline, Step Functions. |
| Big Data Technologies | PySpark, Scala, Kafka, Spark streaming, Zookeeper, MapReduce, Hive, Databricks |
| Hadoop Distribution | Cloudera, Horton Works |
| Languages | SQL, PL/SQL, Python, HiveQL, Scala. |
| Web Technologies | HTML, CSS, JavaScript, XML, JSP, Restful, SOAP |
| Operating Systems | Windows, UNIX, LINUX. |
| Build Automation tools | Ant, Maven |
| Version Control | GIT, GitHub. |
| IDE &Build Tools, Design | Eclipse, Visual Studio. |
| Databases | MS SQL Server, PostgreSQL, Snowflake, Azure SQL DB, Azure Synapse, MS Excel, MS Access, Oracle |

**Work Experience:**

**Client: Orchestro.ai, Dublin, CA July 2024 – Till Date**

**Role: Data Engineer
Responsibilities:**

* Optimized route and cost analysis by evaluating delivery routes, parcel movement, rate cards, cost per parcel, and network coverage, resulting in reduced turnaround time and increased revenue through data-driven business decisions.
* Developed and deployed data-driven tools using Pyspark, SQL, and Snowflake to analyze and transfer carrier and shipper data, uncovering hidden potential from raw files and driving increased revenue through service and pricing comparisons.
* Converted business and technical requirements into efficient data models, implementing DDL, roles, and views to enable seamless data loading and cost card generation.
* Designed and optimized scalable data architectures, including star schemas, data marts, and advanced data models, leveraging REST APIs and cloud-based integrations for seamless external and internal data interoperability.
* Designed and automated ETL pipelines using AWS Glue, S3, and Lambda to integrate carrier and supply chain data, enabling real-time tracking of shipments, operational KPIs reducing data refresh times by 30%
* Implemented Apache Airflow to orchestrate data workflows for files landing in AWS S3, ensuring schema identification, validation, and automated processing.
* Developed custom transformations and backend operations to standardize and enrich sales and delivery data. Designed an interactive UI dashboard to visualize cost savings insights, empowering customers to take data-driven actions.
* Streamlined contract management processes by leveraging SQL and Python to automate amendments, ensure timely database updates, analyze carrier performance metrics, and implement data-driven adjustments, enabling informed decision-making and securing more favorable terms.
* Automated customer inquiry responses and streamlined internal processes by implementing a custom-built small-scale LLM and API-based data dispatching, reducing response times and client onboarding time
* Developed a user-friendly web portal for data uploads, integrating backend processes with optimized database models to generate rate cards, enhance business decision-making to dynamically adjust rates and seize opportunities faster.
* Ensured data quality and compliance by implementing governance practices such as automated data validation checks, schema enforcement, and role-based access controls (RBAC).
* Optimized data pipelines by incorporating error-handling mechanisms and monitoring tools, ensuring clean and accurate data for reporting and analytics

**Environment:** AWS Glue, AWS Redshift, Amazon S3, Snowflake (Transient, Temporary, Persistent Tables), PySpark, Apache Hive, SQL, Power BI, Python, AWS Glue, Lambda, AI, ML Models

**Client: Peritus software, Remote, CA Mar 2023 – July 2024**

**Role: Data Engineer**

**Responsibilities:**

* Led the design and optimization of data warehouses on Amazon Redshift and S3 to support energy data analytics, reducing query times by 35% and enabling faster insights into energy usage patterns, enhancing data -driven decision- making for the client.
* Developed and prioritized a product roadmap for energy data infrastructure projects on AWS, aligning project milestones with the client’s business goals and implementing automated batch processes that improved data refresh speeds, enabling near-real-time analytics for energy monitoring.
* Implemented secure data governance protocols using AWS IAM and data lineage tracking to safeguard sensitive energy usage data, reducing data breach risks and ensuring compliance with industry standards for energy data privacy.
* Reduced data storage costs on S3 and Redshift through optimized data compression and lifecycle management policies, supporting the client's need for cost-effective, scalable storage of large energy datasets.
* Automated ETL processes using AWS Glue and Lambda to efficiently ingest, transform, and load large energy datasets, reducing manual processing time and enabling timely delivery of insights critical for energy usage forecasting and demand planning.
* Utilized Git for version control in a collaborative, multi-team environment, ensuring streamlined deployment of new features and improved management of code for energy data pipelines, accelerating feature delivery and maintaining code.

**Environment:** AWS Glue, AWS Redshift, Amazon S3, Snowflake (Transient, Temporary, Persistent Tables), PySpark, PowerBI.

**Client: Enterprise Minds, CA. Feb 2022 – Feb 2023 Role: Data Engineer**

**Responsibilities:**

* Successfully migrated 100+ complex Pricing, Supply Chain and logistics operation workflows from Alteryx to Azure using Pyspark notebooks and scheduled pipelines on Azure Synapse Analytics and Databricks.
* Migrated data from onsite systems into Azure via Azure Data Factory and built workflows to transform and clean the data, resulting in an increase in data processing efficiency and reduction in manual data entry errors. Designed end-to-end workflows to automatically update Tableau dashboards for sales and finance teams.
* Collaborated on the transfer of 50 million sensitive data records from SAP systems, optimizing data models and custom ETL scripts.
* Implemented snapshot and incremental load frameworks, reducing runtime through efficient record updates and new record insertions, enabling faster data retrieval and storage in cloud warehouses.
* Initiated product development lifecycle by proactively gathering comprehensive business requirements from new clients and converted them into actionable insights with flawless delivery of product and services that contributed to business growth.
* Conceptualized and implemented data transformation algorithms that aligned with business logic, facilitating cohesive reporting and analytics across diverse teams.
* Ensured continuous data flow across systems with timely robust pipelines.
* Played key role in setting up cloud infrastructure, optimizing storage solutions on Azure blob and implemented efficient role- based access controls and implemented robust data encryption measures to enhance data security and compliance.
* **Monitored and optimized data pipeline performance** by implementing **Azure Monitor and Databricks performance tuning**, reducing **latency** and ensuring **high availability and system stability**. Developed **automated anomaly detection scripts** for real-time error handling, minimizing failures and reducing manual debugging efforts.
* Led and mentored a team of engineers, establishing Azure DevOps best practices and CI/CD pipelines to streamline deployments. Collaborated with cross-functional teams to improve data modeling standards and governance policies.
* Assisted new clients in onboarding to Azure, setting up secure environments with RBAC and encryption, and providing hands-on training and technical demos to ensure a smooth transition.
* Enhanced scalability using distributed data processing frameworks i.e Apache Spark and optimized via adaptive query execution, shuffle partitions, and caching mechanisms improving pipeline efficiency. Designed auto-scaling workflows in Databricks, enabling seamless expansion as data volumes grew by 60% without performance degradation.

**Environment:** Azure Databricks, Azure Data Factory, Snowflake (Time Travel, Cloning), Azure Logic Apps, MS SQL, Python, PySpark, PostgreSQL, shell, GIT, JIRA, Kafka, Power BI, Azure DevOps, JSON

**Client: Carvana, Tempe, AZ. June 2021 – Jan 2022**

**Role: Data Analyst**

**Responsibilities:**

* Leveraged SQL and Tableau to optimize relational database design and data models, improving operational efficiency by implementing stored procedures, partitioning, indexing, and views for faster data extraction and live reporting, driving actionable insights.
* Identified and resolved historical data quality issues and inconsistencies in the Snowflake data warehouse by developing robust testing frameworks, ensuring accurate decision-making.
* Created and optimized Indexes, Views, Constraints, Stored Procedures, User -Defined Functions, CTEs, and DDL/ DML Triggers to enhance data manipulation, retrieval, and consistency across systems.
* Addressed operational bottlenecks within supply chain operations by reorganizing resources and collaborating with cross-functional teams, resulting in increase in process efficiency at key auction locations.
* Led the design and implementation of Snowflake data warehouse architecture, including schema design, data modeling, and query optimization for performance and scalability.

 Integrated APIs to enable data transfer across connected systems, ensuring real-time updates for sales and auction data.

**Technologies:** SQL, Python, Snowflake, Excel, Github, Streaming, Tableau, Azure,

**Client: LexisNexis Risk solutions, King of Prussia, PA Jun 2017 – May 2021**

**Role: Data Warehouse Developer**

**Responsibilities:**

* Conducted in-depth research on product lines to inform strategic development plans, driving future product functionality.
* Collected, cleaned, and transformed raw data into actionable insights, uncovering data inconsistencies and developing custom algorithms to resolve hidden issues.
* Re-engineered and optimized existing data models using big data tools, improving run -time efficiency by 40% and enhancing data quality through regular validation checks.
* Extracted data from legacy systems and migrated it to a big data platform via robust, automated pipelines, reducing manual efforts and ensuring data consistency across platforms.
* Assisted in developing a data quality framework that enhanced overall data accuracy and reliability through automated validation checks.
* Built statistical models and interactive dashboards that provided key metrics, driving data driven decision-making for stakeholders.

**Environment:** MySQL, Excel, GIT, GitHub, Informatica, ECL, Python, PowerBI, HPCC, Azure, Tableau, Pandas, Numpy, Seaborn.

**Client: TravelUR, Hyderabad, India Jan 2015 – Dec 2016**

**Role: Data Analyst**

**Responsibilities:**

* Conducted **user research, information architecture, visual design, and user testing** **across web and mobile platforms,** optimizing user experience and increasing **conversion rates while reducing bounce rates.**
* Developed interactive prototypes and wireframes using Figma and Adobe XD, facilitating cross-functional collaboration and streamlining the design-to-development handoff.
* Spearheaded a mobile-first design approach, enhancing accessibility and responsiveness, resulting in improvement in mobile user engagement.
* Collaborated with product managers and developers to ensure seamless integration of UX principles, reducing development iterations and accelerating product launch timelines.
* Implemented **A/B testing, usability testing, and heatmap analysis,** leading to **data-driven design improvements** and enhanced **navigation flow** for better user engagement.
* Presented **UX/UI insights and data-driven recommendations** to stakeholders, advising **alternative design strategies** that improved **customer retention and increased profits.**

**Environment:** MySQL, Excel, Google Analytics, HTML, CSS3, Javascript, SQL, Dashboarding, Talend, Adobe,Python