

Ragini Gopidi
Raginigopidi67@gmail.com
945-217-3164
Data Engineer

Professional Summary:

- Over 4 years of experience as a Data Engineer, specializing in designing, developing, and managing data pipelines and architectures.
- Advanced proficiency in SQL for querying, managing, and manipulating relational databases (MySQL, PostgreSQL, Oracle).
- Experience with NoSQL databases such as MongoDB, Cassandra, and DynamoDB for handling unstructured data.
- Extensive experience with cloud platforms including AWS, Azure, and Google Cloud for deploying and managing data solutions.
- Expertise in utilizing SQL and NoSQL databases, data warehousing solutions, and big data frameworks.
- Designed scalable and robust data pipelines using tools like Apache NiFi, Apache Kafka, and AWS Glue.
- Deployed and managed Kubernetes Operators (IKOs) on GKE to optimize workloads and enhance cluster efficiency.
- Automated deployment pipelines for containerized applications using Kubernetes and CI/CD tools, ensuring rapid feature delivery.
- Architected and implemented secure, resilient Kubernetes clusters on Google Cloud Platform (GCP) using Google Kubernetes Engine (GKE).
- Developed and maintained data warehouses using Amazon Redshift, Snowflake, and Google BigQuery.
- Proficient in programming languages such as Python, Java, and Scala for data engineering tasks and automation.
- Developed real-time data processing solutions using Apache Flink and Apache Storm.
- Skilled in version control and collaboration using Git, GitHub, and Bitbucket for managing code and data pipelines.
- Strong understanding of data governance principles and ensured compliance with GDPR and CCPA regulations.
- Used monitoring and logging tools like Prometheus, Grafana, and ELK Stack for system health and troubleshooting.
- Experienced with containerization technologies like Docker and Kubernetes for deploying and scaling data applications.
- Collaborated with stakeholders to identify data needs and deliver effective data solutions.
- Developed and managed data APIs for seamless integration with external applications and services.
- Experience with data visualization tools like Tableau, Power BI, and Looker for creating insightful dashboards and reports.
- Proficient in using Jupyter Notebooks for data exploration, prototyping, and sharing analyses with teams.
- Stayed updated on the latest data engineering trends and best practices in big data technologies.
- Developed automated CI/CD pipelines using Jenkins and Azure DevOps to streamline data pipeline deployment.
- Specialized in implementing and managing InterSystems IRIS for Health, integrating it within healthcare environments.
- Led deployment and configuration of IRIS for Health systems, optimizing performance and scalability for healthcare applications.
- Proficient in utilizing Azure Data Lake for scalable data storage and analytics.
- Designed and managed data pipelines on Azure using services like Azure Databricks and Azure Data Factory.
- Proficient in designing and implementing CI/CD pipelines with tools like Jenkins, GitLab CI, and GitHub Actions to automate build, test, and deployment processes.

Education:

- Masters in **Computer Science** from **University of Texas at Arlington**

Technical Skills:

- **Programming & Scripting:** Python, SQL, T-SQL, PL/SQL, Java, Scala
- **ETL & Data Integration:** Apache NiFi, SSIS, Talend, Informatica, Azure Data Factory
- **Data Warehousing:** Snowflake, Amazon Redshift, Google BigQuery, Azure Synapse Analytics
- **Big Data Technologies:** Apache Hadoop, Spark, Hive, HBase, Pig
- **Real-Time Data Streaming:** Apache Kafka, Spark Streaming, Apache Flink
- **Cloud Platforms:** AWS (S3, Glue, Redshift, Lambda, EMR), Azure (Data Factory, Synapse, Data Lake, Databricks), Google Cloud Platform (BigQuery, Dataflow, Pub/Sub)
- **Orchestration & Workflow Management:** Apache Airflow, Luigi, Prefect, Oozie
- **Databases:** PostgreSQL, MySQL, SQL Server, Oracle, MongoDB, Cassandra, DynamoDB
- **Containerization & CI/CD:** Docker, Kubernetes, Jenkins, GitLab CI/CD, Terraform, CloudFormation
- **Monitoring & Logging:** Prometheus, Grafana, ELK Stack, Datadog, AWS CloudWatch
- **Data Governance & Security:** Azure Purview, Apache Ranger, Apache Atlas, RBAC, data encryption, masking, GDPR compliance
- **Methodologies & Collaboration:** Agile/Scrum, cross-functional team collaboration, technical documentation, stakeholder communication

Professional Experience:

Client : GE HealthCare

July 2022 – Present

Role : Data Engineer

Responsibilities :

- Diagnosed and resolved production issues in the IRIS ecosystem, significantly reducing downtime and enhancing system reliability.
- Partnered with the L2 operations team to provide technical expertise, troubleshooting complex incidents and accelerating resolution times.
- Improved overall data processing performance through advanced SQL query optimization, indexing, partitioning, and in-memory processing.
- Utilized Hadoop, Spark, and Kafka to process and analyze large-scale datasets in distributed environments.
- Engineered scalable, cloud-based data infrastructure leveraging AWS, Azure, and Google Cloud Platform to meet evolving business needs.
- Automated recurring data workflows using Python and Bash, and orchestrated tasks with tools like Apache Airflow and Luigi.
- Collaborated cross-functionally with data scientists, analysts, and business stakeholders to gather requirements and deliver robust data solutions.
- Managed version-controlled updates to data pipelines, scripts, and infrastructure using Git, ensuring traceability and collaboration.
- Designed and implemented scalable ETL pipelines using Azure Synapse and Azure Data Factory, boosting data processing efficiency by 30%.
- Authored and maintained thorough documentation for ETL pipelines, data models, workflows, and system configurations.
- Led data migration initiatives, including system upgrades, cross-platform transfers, and cloud transitions with minimal disruption.
- Performed regular data audits to ensure data integrity, quality, and compliance with organizational and regulatory standards.
- Proactively researched and tested emerging data engineering tools and technologies to drive innovation and continuous improvement.

- Forecasted storage and processing requirements, aligning infrastructure with future scalability needs.
- Built real-time data streaming solutions using Azure Stream Analytics, enabling seamless integration into Azure Synapse for immediate insights.
- Fine-tuned and debugged production environments within the IRIS platform, ensuring optimal system performance and stability.
- Developed and enforced disaster recovery strategies to protect data integrity and ensure business continuity.
- Created and maintained RESTful APIs to streamline data integration and enable seamless connectivity between systems and applications.

Environment: *Python, Jupyter Notebook, Apache Spark, Hadoop, Hive, Apache Kafka, AWS (S3, Redshift, Glue), Azure Data Lake, Google BigQuery, Snowflake, SQL, PostgreSQL, MongoDB, Docker, Kubernetes, Databricks, Pandas, NumPy, Scikit-learn, TensorFlow, dbt, Tableau, Power BI, Git, Jenkins, Linux, Shell Scripting.*

Client : Accenture, Remote

May 2020- Dec 2021

Role : Data Engineer

Responsibilities:

- Worked on projects involving the analysis, management, and reporting of fixed income securities such as government and corporate bonds, municipal bonds, and mortgage-backed securities.
- Managed and developed systems for trading, risk management, and reporting of derivatives, including options, futures, forwards, and swaps.
- Monitor and manage costs associated with data storage and processing, optimizing for budget constraints.
- Work with various teams, including IT, software engineering, and business units, to align data infrastructure with organizational goals.
- Collaborated with healthcare providers and IT teams to ensure compliance with regulatory standards and interoperability requirements in data exchange processes.
- Experienced in version control and collaboration using Git, including branching strategies, pull requests, and code reviews, ensuring smooth and efficient development workflows.
- Built and managed data Lakehouse architectures using Databricks and Delta Lake, facilitating seamless data integration and consistency.
- Integrated CI/CD pipelines with containerization platforms like Docker and Kubernetes, enabling continuous deployment of microservices and cloud-native applications.
- Integrate data from multiple sources such as databases, APIs, and third-party services into a unified data warehouse or lake.
- Implement data validation and cleaning processes to ensure the accuracy, completeness, and reliability of data.
- Ensure data infrastructure can scale to handle increasing volumes and velocities of data.
- Automate data ingestion, transformation, and loading processes to improve efficiency and reduce manual intervention.
- Work closely with data scientists to understand data requirements and deliver datasets that support advanced analytics and machine learning models.
- Designed and implemented SDA (Structured Data Architecture) and DTL (Data Transformation Language) models for seamless data conversion between HL7, CCD, and FHIR formats.
- Implement security measures to protect sensitive data, including encryption, access controls, and compliance with data protection regulations.
- Monitor data pipelines and infrastructure for performance issues and troubleshoot problems as they arise.
- Maintain comprehensive documentation of data workflows, pipeline configurations, and infrastructure architecture.
- Use version control systems to manage changes to data pipelines and infrastructure configurations.
- Assess and recommend data engineering tools and technologies to improve efficiency and capability.
- Develop and manage ETL jobs to automate data processing tasks.
- Utilize big data technologies such as Hadoop, Spark, and Kafka for handling large-scale data processing.

Environment: *Python, Apache Spark, Hadoop, Hive, Apache Kafka, Apache Airflow, AWS (S3, Redshift, Glue, EMR), Azure Data Lake, Google BigQuery, Snowflake, SQL, PostgreSQL, MySQL, MongoDB, Cassandra, Terraform, Docker, Kubernetes, Databricks, Pandas, NumPy, SciPy, Scikit-learn, TensorFlow, PyTorch, dbt, Looker, Tableau, Power BI, Git, Jenkins, CI/CD, Linux, Shell Scripting.*