**Deepthi Vadladi**

**Sr. Data Engineer**

|  |  |  |
| --- | --- | --- |
| **CONTACT** | **PROFILE SUMMARY** | |
| **Contact:** 707-909-2269  **Email:** [deepthivad6@gmail.com](mailto:deepthivad6@gmail.com)  **TECHNICAL SKILLS**  **Data Processing:** Python, PySpark, Scala, SQL, Hadoop, Spark RDDs, Apache Kafka  **Cloud Platforms:** Azure, AWS, GCP  ETL Tools: SSIS, Informatica, Talend  **Data Warehouses:** Snowflake, Redshift, BigQuery  **Data Formats:** JSON, XML  **Database Management:** MySQL, MongoDB, Oracle, Cassandra, PostgreSQL, SQL Server, Cosmos DB  **Data Analysis:** Pandas, NumPy  **Infrastructure as Code:** AWS CloudFormation, Terraform, ARM templates  **Containerization & Orchestration:** Docker, Kubernetes  **Data Visualization:** Tableau, Power BI, Looker  **Logging and Monitoring:** Splunk, ELK, Azure Monitor  **Version Control & CI/CD:** Git, GitHub, Jenkins, Azure DevOps  **Project Management:** Jira, Agile, Scrum | * Around 9 years of experience in Data Engineering, Data Pipeline Design, Development, and Implementation as a senior data Engineer. * Proficient in data processing using Python, PySpark, Scala, and SQL, with a strong background in designing and implementing complex ETL pipelines. * Experienced in managing big data technologies such as Hadoop, Spark RDDs, and Apache Kafka for efficient data ingestion and processing. * Skilled in cloud platforms, including Azure, AWS, and GCP, focusing on designing scalable and cost-effective data solutions in the cloud. * Expertise in ETL tools such as SSIS, Informatica, and Talend for data integration and transformation, ensuring data quality and consistency. * Proficient in working with data warehouses like Snowflake, Redshift, and BigQuery, optimizing performance and data retrieval. * In-depth knowledge of various data formats, including JSON and XML, for seamless data exchange and integration. * Experienced in database management across multiple platforms, including MySQL, MongoDB, Oracle, Cassandra, PostgreSQL, SQL Server, and Cosmos DB. * Proficient in data analysis using Pandas and NumPy, enabling data-driven decision-making and insights generation. * Skilled in Infrastructure as Code (IaC) tools like AWS CloudFormation, Terraform, and ARM templates for automating infrastructure deployment. * Well-versed in containerization and orchestration technologies such as Docker and Kubernetes, ensuring portability and scalability of data solutions. * Proficient in data visualization tools like Tableau, Power BI, and Looker to create interactive and insightful data dashboards. * Experienced in setting up logging and monitoring solutions using Splunk, ELK, and Azure Monitor for real-time data tracking and issue resolution. * Strong background in version control and CI/CD practices using Git, GitHub, Jenkins, and Azure DevOps, ensuring code quality and continuous integration. * Familiar with project management methodologies like Jira, Agile, and Scrum, facilitating collaboration and project delivery within cross-functional teams. | |
| **WORK EXPERIENCE** | | |
| **Corteva, Indianapolis, IN**  **Sr. Data Engineer Roles & Responsibilities:** | | **Oct 2022 – Till Date** |
| * Analyze, design, and build Modern data solutions using Azure PaaS service to support visualization data. * Spearheaded the design and development of data pipelines using Python, PySpark, and Scala, ensuring efficient data extraction, transformation, and loading (ETL) processes. * Managed complex data integration tasks by leveraging SQL, Hadoop, and Spark RDDs, optimizing data processing performance for large-scale datasets. * Orchestrated data workflows in Azure Data Factory, Azure Databricks, and Apache NiFi, guaranteeing seamless data movement and transformation across cloud platforms. * Implemented real-time data streaming solutions using Apache Kafka, processing high-velocity data streams for immediate insights. * Developed and delivered applications with Adobe LiveCycle ES3/4 on a single J2EE Environment. * Implement One time Data Migration of Multistate level data from SQL server to Snowflake by using Python and SnowSQL * Contributed to Snowflake data warehousing tasks, including data loading and querying, and used PostgreSQL, ensuring data availability and accuracy. * Worked with data visualization tools including Tableau, Adobe Analytics Omniture, and matplotlib * Developed and fine-tuned Hive and Pig scripts for data transformation, enabling efficient query processing in a Hadoop ecosystem. * Developed time-based and event-based triggers for ADF pipelines. * Automated data ingestion from various sources into Azure Cosmos DB, guaranteeing a consistent and reliable data repository. * Monitored and optimized data pipelines using Azure Monitor, ensuring data quality and reliability while meeting SLAs. * Collaborated with cross-functional teams in an Agile environment, using JIRA and Azure DevOps for project management and version control with Git. * Containerized data applications using Docker and orchestrated container deployments with Kubernetes, ensuring scalability and resource efficiency. * Used AWS Glue for the data transformation, validate and data cleansing. * Hands-on experience with Snowflake utilities, SnowSQL, SnowPipe, Big Data model techniques using Python / Java. * Created and maintained Azure Resource Manager (ARM) templates for infrastructure as code, automating cloud resource provisioning. * Developed Python scripts to do file validations in Databricks and automated the process using ADF. * Created monitors, alarms, notifications and logs for Lambda functions, Glue Jobs, EC2 hosts using Cloudwatch * Designed data schemas in Avro and JSON formats, facilitating data interoperability and compatibility across systems. * Streamlined data extraction from on-premises systems into Azure using Sqoop, minimizing latency and ensuring data consistency. * Developed and maintained ETL processes using SSIS, ensuring data accuracy and consistency across the organization. * Created interactive data visualizations and reports using Tableau, providing actionable insights to stakeholders. * Collaborated with data architects to ensure data security, compliance, and governance by industry best practices.   **Environment**: Python, SQL, HDP, Hadoop, Spark RDDs, Hive, Pig, Sqoop, PySpark, Scala, PostgreSQL, Azure, Azure Data Factory, Azure Databricks, CosmosDB, Azure Monitor, Snowflake, Bash, ETL, SSIS, Tableau, ARM templates, Avro, JSON, Docker, Git, Azure DevOps, JIRA, Agile.  **GAP , Hyderabad , India** | | |
| **Data Engineer Roles & Responsibilities:** | | **May '19 - May '22** |
| * Designed and set Enterprise Data Lake to provide support for various use cases, including * Analytics, processing, storing, and Reporting of voluminous, rapidly changing data. * Developed and maintained ETL (Extract, Transform, Load) pipelines using Python to extract data from various sources, transform it, and load it into data storage systems. * Utilized Informatica to streamline and automate data integration processes for efficient data movement. * Managed data processing tasks on Elastic MapReduce (EMR) clusters in AWS, optimizing data workflows for scalability and performance. * Provisioned and configured AWS services such as EC2, S3, Lambda, and Step Functions to orchestrate data pipelines and ensure data availability and reliability. * Expertise Snowflake utilities such as SnowSQL, SnowPipe, Tasks, Streams, Time travel, Optimizer, Metadata Manager, data sharing, and stored procedures. * Worked on developing JSON Scripts for deploying the Pipeline in Azure Data Factory (ADF) that process the data. * Data Extraction, accumulations, and combination of Adobe information inside AWS Glue utilizing PySpark. * Used AWS glue catalog with crawler to get the data from S3 and perform sql query operations * Implemented IAM policies to manage access control and security for AWS resources. * Contributed to setting up and managing data lakes using AWS Lake Formation to centralize and catalog data assets. * Developed REST APIs for data retrieval and integration, handling data in JSON and XML formats. * Worked with Hadoop and HDFS to store and process large data efficiently. * Utilized PySpark for distributed data processing and analysis, optimizing data workflows for performance. * Collaborated on data modeling and query optimization using Hive and Pig to support data analytics and reporting. Worked with Spark for real-time data processing and analytics. * Build Integration between applications primarily Salesforce, Oracle CX, HubSpot, Marketo, Adobe Experience Cloud, and AWS Glue. * Develop proof of concepts (POC) using Azure Cloud services such as Azure Data factory (ADF), Logic Apps, Data Bricks, Azure * Assisted in designing and managing Redshift data warehouses for data storage and analytics. * Managed and maintained relational databases such as MySQL and NoSQL databases like MongoDB. * Performed data analysis and manipulation using Pandas and NumPy libraries in Python. * Created and managed infrastructure as code using AWS CloudFormation for reproducible and scalable deployments. * Developed data visualizations and dashboards using Power BI for data-driven insights and reporting. Collaborated on version control and code management using GitHub. * Well versed with Big data on AWS cloud services i.e. EC2, S3, Glue, Anthena, DynamoDB and RedShift * Implementing One time Data Migration of Multistate level data from SQL server to Snowflake by using Python and SnowSQL. * Automated CI/CD (Continuous Integration/Continuous Deployment) pipelines using Jenkins for efficient code deployment. * Containerized applications and services using Docker and managed container orchestration with Kubernetes. Set up ELK (Elasticsearch, Logstash, Kibana) stack for log analysis and monitoring. * Actively participated in Agile and Scrum methodologies for project management and collaboration with cross-functional teams. * Managed tasks, issues, and project progress using Jira for effective project management and tracking.   **Environment:** Python, ETL, Informatica, Elastic MapReduce, AWS, EC2, S3, Lambda, Step Functions, IAM, AWS Lake Formation, REST API, JSON, XML, Hadoop, HDFS, PySpark, Hive, Pig, Spark, Redshift, MySQL, MongoDB, Pandas, NumPy, CloudFormation, Power BI, GitHub, Jenkins, Docker, Kubernetes, ELK, Agile, Scrum, Jira.  **Zapcom Solutions Pvt Ltd., Hyderabad, India** | | |
| **Data Engineer Roles & Responsibilities:** | | **July '14 - Apr '19** |
| * Involved in loading and transforming large sets of structured and semi-structured datasets and analyzed them by running Hive queries. * Developed custom Python program including CI/CD rules for Google Cloud data catalog for metadata management. * Designed and developed spark job with Scala to implement an end-to-end data pipeline for batch processing. * Supported data pipeline development using SQL Server, Cassandra, and Google Cloud Storage to ensure smooth data ingestion and storage. * Assisted in designing and implementing data processing workflows using Google Dataflow and Pub/Sub for real-time data streaming. * Contributed to cluster management and data processing tasks using CDH, Hadoop, and Spark RDDs to optimize data analysis. * Participated in the creation of data transformations and data modeling processes with Hive and Pig for efficient data processing. * Collaborated on developing REST APIs for data access and integration with various applications. * Worked on ETL (Extract, Transform, Load) processes using Talend to extract data from diverse sources and prepare it for analysis. * Assisted in managing and optimizing BigQuery for large-scale data querying and analysis. * Handled data in JSON format and used Ant for building automation to streamline data engineering tasks. * Supported log analysis and monitoring using Splunk for data troubleshooting and performance optimization. * Engaged in Agile and Scrum methodologies for effective project management and collaboration with cross-functional teams. * Contributed to continuous integration and deployment (CI/CD) pipelines with Jenkins to automate data pipeline deployments. * Assisted in infrastructure provisioning and management using Terraform to ensure scalability and reliability. * Collaborated with colleagues on version control using Git for code management and collaboration. * Worked with Looker to create and customize data dashboards and reports for data visualization. * Participated in issue tracking and project management using Jira to ensure project milestones were met.   **Environment:** Python, SQL Server, Cassandra, Google Cloud Storage, Dataflow, Pub/Sub, DataProc, Dataprep, CDH, Hadoop, Spark, Hive, Pig, REST API, ETL, Talend, BigQuery, JSON, Ant, Splunk, Agile, Scrum, Jenkins, Terraform, Git, Looker, Jira. | | |
| **EDUCATION DETAILS** | | |
| * Master's degree in computer science from JNTU Hyderabad. | | |