**Bhanu Teja  
Sr. Data Engineer**

**Email:** [**tejab7879@gmail.com**](mailto:tejab7879@gmail.com) **| Phone: 407-476-6956**

**PROFESSIONAL SUMMARY**

* Over 10+ years of hands-on experience designing, implementing, and optimizing data engineering solutions.
* Extensive knowledge of cloud platforms, including AWS, Azure, and GCP, with a track record of successfully deploying scalable and robust data architectures.
* Demonstrated proficiency in Big Data technologies such as Hadoop, Spark, Hive, Sqoop, and CDH, applying them to tackle complex data processing challenges.
* Expertise in Data Integration and ETL processes using tools like AWS Glue, Talend, ADF, and Informatica, ensuring seamless data flow and transformation.
* Proven ability to build and manage data warehouses, leveraging platforms like Redshift, BigQuery, Teradata, and Snowflake for efficient storage and retrieval.
* Adept at scripting with Python and PowerShell, automating routine tasks, and ensuring the efficiency and reliability of data workflows.
* Proficient in version control systems like Git and GitHub. Skilled in IaC using CloudFormation and Terraform.
* Hands-on experience with Docker and Kubernetes, facilitating containerized deployments and orchestration for scalable applications.
* Expertise in managing various databases, including MySQL, DynamoDB, SQL Server, CosmosDB, Oracle, and DB2.
* Proficient in data analytics and manipulation using Pandas, NumPy, and Scikit-Learn, deriving valuable insights from large datasets.
* Applied machine learning techniques using TensorFlow, Mahout, and Scikit-Learn to enhance data processing capabilities.
* Implemented robust log management and monitoring solutions with Cloud Watch, ELK Stack, Grafana, and Splunk to ensure system health and performance.
* Implemented Continuous Integration and Deployment pipelines using Jenkins, Azure DevOps, and Maven, ensuring efficient and reliable software delivery.
* Proficient in creating meaningful visualizations using Power BI and Tableau to communicate insights effectively.
* Applied Agile, Scrum, and Kanban methodologies to streamline development processes and deliver high-quality solutions.
* Utilized collaboration tools like Jira, SharePoint, Service Now, and Bugzilla for effective project management and communication.
* Adept at fostering a collaborative environment, working seamlessly with cross-functional teams to understand business requirements and translate them into effective data solutions.
* Proven ability to communicate complex technical concepts to non-technical stakeholders, ensuring alignment between technical implementation and business goals.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Data Integration and ETL:** AWS Glue, Talend, ADF, Informatica  **Cloud Services:** AWS, Azure, GCP  **Big Data Technologies:** Hadoop, Spark, Hive, Sqoop, CDH  **Database Management:** MySQL, DynamoDB, SQL Server, CosmosDB, Oracle, DB2  **Data Warehousing:** Redshift, BigQuery, Teradata, Snowflake  **Data Analytics and Manipulation:** Pandas, NumPy, Scikit-Learn  **Machine Learning:** TensorFlow, Mahout, Scikit-Learn  **Scripting and Automation:** Python, PowerShell  **Version Control:** Git, GitHub | **Infrastructure as Code (IaC):** CloudFormation, Terraform  **Containerization and Orchestration: Docker, Kubernetes**  **Configuration Management:** Ansible  **Log Management and Monitoring:** Cloud Watch, ELK Stack, Grafana, Splunk  **Collaboration and Project Management:** Jira, SharePoint, Service Now, Bugzilla  **Continuous Integration and Deployment:** Jenkins, Azure DevOps, Maven  **Data Visualization:** Power BI, Tableau  **Methodologies:** Agile, Scrum, Kanban |

**WORK EXPERIENCE**

**JPMC, New York, NY | June 2022 - Present**

**Sr. Data Engineer**

* Involved in the entire Software Development Life Cycle (SDLC) for data projects, ensuring seamless integration of various technologies.
* Spearheaded the development of end-to-end ETL pipelines, leveraging AWS Glue and Python scripting, ensuring seamless extraction, transformation, and loading of diverse data sources into a centralized AWS data lake.
* Implemented Extract, Transform, Load (ETL) processes using AWS Glue to move and transform data across diverse sources efficiently.
* Utilized AWS EMR with spark to perform batch processing operations on various Big Data sources.
* Proficiently leveraged AWS services, including EC2, S3, RDS, Lambda, and DynamoDB, to create robust and scalable data solutions.
* Designed and managed Data Pipelines using AWS Data Pipeline, incorporating Step Functions for workflow orchestration.
* Create Athena on data source on S3 buckets for adhoc querying and business dashboarding using Quicksight and Tableau reporting tools.
* Utilized Redshift for high-performance data warehousing, optimizing queries for enhanced analytics capabilities.
* Demonstrated data manipulation and analysis expertise using Pandas and NumPy libraries within Python.
* Developed proficiency in handling XML data, employing XSLT transformations for structured data representation.
* Implemented large-scale data processing using Apache Spark, embracing the MapReduce paradigm for distributed computing.
* Applied advanced data querying techniques with Hive and HBase in conjunction with Hadoop ecosystem technologies.
* Employed Pig and Sqoop for efficient data ingestion and transfer between Hadoop and relational databases.
* Implemented machine learning solutions using Mahout for recommendation systems and data clustering.
* Harnessed the power of serverless computing with AWS Lambda, integrating with SQS and SNS for an event-driven architecture.
* Developed Python scripts to automate the ETL process using Apache Airflow and CRON scripts in the UNIX operating system as well.
* Designed and developed the data warehouse models using Snowflake and Star schema.
* Managed source code repositories and version control using Bitbucket for collaborative development.
* Optimized content delivery through CloudFront, ensuring low latency and high-performance data distribution.
* Established continuous integration and deployment pipelines using Jenkins, enhancing the development workflow.
* Collaborated within an Agile and Scrum framework, ensuring iterative development and timely project deliveries.
* Employed Infrastructure as Code (IaC) principles with CloudFormation for automated resource provisioning.
* Containerized applications using Docker, orchestrating deployments with Kubernetes for scalability and portability.
* Implemented configuration management with Ansible, streamlining application deployment and updates.
* Monitored and managed cloud infrastructure using CloudWatch, ensuring optimal performance and resource utilization.

**Environment:** ETL, AWS Glue, AWS, MySQL, DynamoDB, Redshift, Pandas, NumPy, MapReduce, Hive, Hadoop, HBase, Pig, Sqoop, Mahout, AWS Lambda, SQS, SNS, Bitbucket, CloudFront, Jenkins, Jira, Agile, Scrum, CloudFormation, Docker, Kubernetes, Ansible, CloudWatch, Qlik

**Horizon BCBS, Newark, NJ | Oct 2020 - May 2022**

**Data Engineer**

* Designed and implemented scalable data processing solutions using Hadoop, Spark, and MapReduce to efficiently handle large volumes of data.
* Implemented and optimized AWS infrastructure, including EC2, S3, Lambda, and RDS, utilizing Autoscaling for efficient resource management.
* Orchestrated complex data workflows using AWS Step Functions and leveraged Amazon EMR for large-scale data processing.
* Developed data transition programs from DynamoDB to AWS Redshift (ETL Process) using AWS Lambda by creating functions in Python for certain events based on use cases.
* Created Stacks using Amazon Cloud Formation templates to launch AWS Infrastructure and resources.
* Worked on AWS Redshift and RDS for implementing models and data on RDS and Redshift.
* Developed step functions to design custom workflows that integrated with AWS Glue and AWS Lambda Designed and executed data integration processes using Informatica to ensure seamless connectivity between diverse data sources.
* Designed and implemented ETL pipelines between from various Relational Data Bases to the Data Warehouse using Apache Airflow.
* Defined and managed infrastructure as code using Terraform, automating the provisioning and configuration of resources.
* Developed and maintained ETL processes using Python, PySpark, and Scala, ensuring seamless extraction, transformation, and data loading.
* Set up and managed Kafka for real-time data streaming, facilitating quick data processing and analysis.
* Implemented ELK stack for log management and monitoring, ensuring comprehensive visibility into system performance.
* Cleaned and manipulated complex datasets to create the data foundation for further analysis and the development of key insights (MS SQL server, R, Tableau, Excel)
* Contributed to Agile development methodologies, participating in JIRA for project management and collaboration.
* Collaborated with cross-functional teams to deliver high-quality solutions in alignment with Agile principles.
* Ensured adherence to Agile best practices, conducting regular sprint reviews and retrospectives for continuous improvement.
* Provided technical expertise and support, fostering a culture of innovation and excellence within the data engineering team.

**Environment:** Hadoop, Spark, MapReduce, Snowflake, SQL Server, CosmosDB, JSON, Databricks, Pandas, NumPy, TensorFlow, PowerShell, GitHub, AWS, Informatica, Terraform, Kafka, ELK, Power BI, JIRA, Agile.

**Iris Software Pvt. Ltd, India | Oct 2017 - Sep 2020**

**Data Engineer**

* Orchestrated data processing workflows using Google Cloud services, including Pub/Sub for messaging, Dataprep for data preparation, and Dataflow for scalable data processing.
* Implemented Hadoop and Hive for distributed storage and querying, ensuring efficient handling of large-scale datasets.
* Utilized Cloudera Distribution for Hadoop (CDH), Sqoop, and Apache Spark for seamless integration and processing of diverse data sources.
* Executed data analytics and reporting with BigQuery, harnessing the power of Google Cloud's fully managed, serverless data warehouse.
* Managed Oracle databases, ensuring data integrity, performance optimization, and efficient querying.
* Integrated REST APIs to facilitate data exchange and communication between different systems.
* Implement One-time Data Migration of Multistate level data from SQL server to Snowflake using Python and SnowSQL.
* Applied advanced data manipulation and analysis using NumPy and Pandas in Python and Scala environments.
* Employed NumPy and Pandas libraries for efficient data manipulation and transformation, optimizing ETL processes.
* Implemented machine learning models and algorithms using Scikit-Learn, enhancing data-driven decision-making processes.
* Collaborated with Git for version control and GitLab CI for continuous integration, ensuring code quality and reliability.
* Utilized Maven for project management and build automation, streamlining development processes.
* Managed data-related incidents and requests through Service Now, ensuring timely resolution and efficient ticket handling.
* Created visualizations and dashboards in Tableau for effective data presentation and insights.
* Implemented Grafana to monitor and visualize data pipeline performance and system health.
* Worked in an Agile and Kanban environment, participating in iterative development cycles and ensuring rapid and flexible responses to changing requirements.

**Environment:** GCP, Hadoop, Hive, CDH, Sqoop, Apache Spark, Bigquery, Oracle, REST APIs, NumPy, Pandas, Python, Scala, Scikit-Learn, Git, GitLab CI, Maven, Service Now, Tableau, Grafana, Agile, Kanban.

**Byteridge Software Private Limited, India. | June 2014 - Sept 2017**

**Data Engineer**

* Designed and implemented Extract, Transform, Load (ETL) processes using Talend for efficient data integration.
* Utilized Python, NumPy, and Pandas for data manipulation and analysis, ensuring data accuracy and reliability.
* Managed data storage and retrieval on AWS, leveraging EC2 instances, S3 for object storage, and RDS for relational databases.
* Developed serverless applications with AWS Lambda, optimizing data processing workflows for scalability.
* Developed data transition programs from DynamoDB to AWS Redshift (ETL Process) using AWS Lambda by creating functions in Python for certain events based on use cases.
* Wrote Lambda functions in python for AWS's EMR, Lambda which invokes python scripts to perform various transformations and analytics on large data sets to various formats
* Proficient in PL/SQL, Teradata, and DB2 for database management and SQL for querying relational databases.
* Implemented Big Data solutions with Hadoop, Spark, Hive, and Sqoop, handling large-scale data processing efficiently.
* Conducted version control and collaboration using GIT, ensuring seamless teamwork and code management.
* Administered SharePoint for effective document sharing and collaboration within the project team.
* Demonstrated expertise in SQL for database querying and data retrieval from various sources.
* Utilized Excel for data analysis, reporting, and visualization.
* Adopted Agile methodology for iterative and collaborative project development.
* Leveraged Splunk for log analysis, monitoring, and troubleshooting within the data engineering ecosystem.
* Tracked bugs and resolve issues using Bugzilla, ensuring the reliability of data pipelines and systems.

**Environment:** Python, ETL, Talend, NumPy, Pandas, AWS (EC2, S3, RDS, Lambda, EMR), PL/SQL, Teradata, DB2, Hadoop, Spark, Hive, Sqoop, GIT, SharePoint, SQL, Excel, Agile, Splunk, Bugzilla.

**EDUCATION:**

* Bachelors in in Bankatlal Badruka College of Information Technology.