# Triveni

# Sr. Data Engineer

# 302-306-1681 | triveni25010@gmail.com

# SUMMARY

* Over 9+ years of extensive experience as a Sr. Data Engineer, proficient in Python T-SQL, with a strong foundation in building robust and scalable data solutions.
* Demonstrated expertise in big data technologies, including Hadoop, Hive, Pig, Hortonworks, HDInsight, Spark, PySpark, and Databricks.
* Proven track record in cloud platforms such as AWS, Azure, and Google Cloud Platform (GCP), with hands-on experience in deploying and managing data solutions in cloud environments.
* Proficient in continuous integration and deployment using Jenkins, AWS CodePipeline, Azure DevOps, and Maven.
* Skilled in working with various data warehouses, including RedShift, Snowflake, and BigQuery, ensuring optimal data storage, retrieval, and analysis.
* Expertise in managing diverse databases such as MySQL, DynamoDB, Amazon RDS (Aurora), MongoDB, Oracle, and SQL Server.
* Hands on experience on Google Cloud Platform (GCP) in all the bigdata products BigQuery, Cloud Data Proc, Google Cloud Storage, Composer (Air Flow as a service).
* Experienced working extensively on the Master Data Management (MDM) and application used for MDM
* In-depth knowledge of machine learning frameworks like Mahout, Pandas, NumPy, PyTorch, and SciPy, enabling the integration of advanced analytics into data pipelines.
* Proficient in implementing security measures using IAM, KMS, JWT, Azure Key Vault, and Azure AD.
* Extensive experience working with data serialization formats, including Avro, JSON, XML, and XSLT, as well as automation tools like Ansible for efficient data processing.
* Advanced proficiency in version control systems such as Git, BitBucket, GitHub, and data visualization tools including Looker, Tableau, and Power BI for comprehensive data analysis and reporting.
* Expertise in ETL tools like AWS Glue, Informatica, Apache Nifi, Apache Airflow, and Talend, facilitating seamless data extraction, transformation, and loading processes.
* Skilled in containerization and orchestration technologies like Docker, Kubernetes, and OpenShift.
* Hands-on experience with infrastructure-as-code tools such as AWS CloudFormation and Terraform.
* Proficient in monitoring and logging tools like ELK Stack, New Relic, Splunk, and AWS CloudWatch.
* Well-versed in Agile and Scrum methodologies, with expertise in utilizing Jira and Confluence for effective project management and collaboration.
* Excellent organizational and time management skills, managing multiple projects simultaneously and delivering high-quality results within deadlines.
* Team player with a collaborative mindset, fostering a positive work environment and contributing to the success of collective goals.

# TECHNICAL SKILLS

|  |  |
| --- | --- |
| **Cloud Platforms:** AWS, Azure, Google Cloud Platform (GCP)  **Programming Languages:** Python, T-SQL, SQL  **Database Management:** MySQL, DynamoDB, Amazon RDS (Aurora), MongoDB, Oracle, SQL Server  **Data Warehousing:** RedShift, Snowflake, BigQuery  **ETL Tools:** AWS Glue, Apache Nifi, Informatica, Apache Airflow, Talend  **Big Data Technologies:** Hadoop, Hive, Pig, Hortonworks, HDInsight, Spark, PySpark, Databricks  **Version Control:** Git, BitBucket, GitHub  **Continuous Integration/Continuous Deployment (CI/CD):** Jenkins, AWS CodePipeline, Azure DevOps, Maven  **Configuration Management:** Ansible | **Monitoring and Logging:** AWS CloudWatch, ELK Stack, New Relic, Splunk  **Data Visualization:** Looker, Tableau, Power BI, Looker  **Machine Learning Libraries:** Mahout, Pandas, NumPy, PyTorch, SciPy  **Security and Access Control:** IAM, KMS, JWT, Azure Key Vault, Azure AD  **Data Serialization:** Avro, JSON, XML, XSLT  **Infrastructure as Code (IaC):** AWS CloudFormation, Terraform  **Containerization:** Docker, Kubernetes, Openshift  **Collaboration and Project Management:** Jira, Confluence  **Operating Systems:** Linux, Windows, UNIX, macOS |

# EDUCATION

* Bachelors in computer science from Amrita School Of Engineering – 2012
* Masters in Information Assurance from Wilmington University - 2019

# WORK EXPERIENCE

**OPTUM, Chicago, IL | Dec 2022 - Present**

**Sr. Data Engineer**

**Responsibilities:**

* Implemented and managed AWS services such as EC2, S3, RDS, and EMR for scalable and reliable data storage and computation.
* Utilized AWS Lambda for serverless computing, optimizing resource usage, and automating data processing tasks.
* Orchestrated AWS Glue data workflows, ensuring seamless extraction, transformation, and loading (ETL) processes.
* Implemented ETL processes using Spark, PySpark, and Scikit-Learn for efficient and scalable data processing.
* Managed MySQL and DynamoDB databases for data integrity, availability, and performance.
* Designed and maintained data warehouses on RedShift, optimizing data storage and retrieval for analytical purposes.
* Utilized AWS Step Functions to automate and coordinate complex data processing workflows.
* Implemented Oozie for workflow scheduling and coordination in Hadoop ecosystems.
* Developed multi cloud strategies in better using GCP (for its PAAS) and Azure (for its SAAS).
* Managed code repositories on BitBucket for version control and collaborative development.
* Implemented AWS CodePipeline for continuous integration and deployment, streamlining the development process.
* Collaborated on projects using Agile and Scrum methodologies, tracking tasks and progress using Jira.
* Implemented AWS CloudFormation for infrastructure as code (IaC) to efficiently provision and manage cloud resources.
* Completed enhancement for MDM (Master data management) and suggested the implementation for hybrid MDM (Master Data Management).
* Utilized Ansible for configuration management for consistency and efficiency in infrastructure.
* Implemented Erwin for data modeling for a clear and organized representation of data structures.
* Developed interactive reports and dashboards using Looker, providing stakeholders with actionable insights.
* Utilized XML and XSLT for data transformation and processing for compatibility with different systems.
* Integrated machine learning libraries like Mahout, Pandas, and NumPy for advanced analytics and predictive modeling.
* Build data pipelines in airflow in GCP for ETL related jobs using different airflow operators.
* Create external tables with partitions using Hive, AWS Athena and Redshift.
* Utilized Dataiku for collaborative and integrated data science and machine learning workflows.
* Implemented AWS Kinesis for real-time data streaming and processing.
* Managed content delivery through AWS CloudFront for low latency and high-performance access to data.
* Implemented IAM for secure access control for the principle of least privilege.
* Utilized KMS for data encryption at rest, enhancing data security measures.
* Set up and configured CloudWatch for real-time monitoring and logging of AWS resources.
* Ensured comprehensive logging and monitoring for infrastructure and application performance.

**Tech Stack:** AWS, AWS Glue, MySQL, DynamoDB, RedShift, Spark, MDM, PySpark, Scikit-Learn, Mahout, Pandas, NumPy, BitBucket, AWS CodePipeline, Jira, AWS CloudFormation, Athena, Ansible, Looker, IAM, KMS, CloudWatch, AWS Kinesis, AWS CloudFront, Agile, Scrum, BitBcuket.

**Corteva Agriscience, Indianapolis, IN | March 2021 – Nov 2022**

**Data Engineer**

**Responsibilities:**

* Implemented and optimized data processing workflows using Hadoop, Hive, and Pig for efficient extraction, transformation, and loading (ETL) of large datasets.
* Managed and configured Amazon EMR clusters for scalable and distributed data processing using Apache Spark and PySpark.
* Utilized AWS services such as EC2 and S3 for scalable and cost-effective data storage and computation.
* Managed relational databases on AWS RDS, including Aurora, for data integrity and availability.
* Implemented CI/CD pipelines using Jenkins for automated testing and deployment of data processing applications.
* Utilized Terraform for infrastructure as code (IaC) to provision and manage cloud resources efficiently.
* Implemented ELK (Elasticsearch, Logstash, Kibana) stack for centralized logging and monitoring of data processing workflows.
* Got involved in migrating on prem Hadoop system to using GCP (Google Cloud Platform).
* To analyze the data Vastly used Athena to run multiple queries on processed data from Glue ETL Jobs and then used Quick Sight to generate Reports for Business Intelligence.
* Ensured real-time performance monitoring and issue detection using monitoring tools like AWS CloudWatch.
* Designed and maintained data warehouse solutions on Snowflake, optimizing data storage and retrieval for analytics.
* Responsible for documenting, developing end-to end Data Quality (DQ) and Master Data Management (MDM) solutions from detailed and high-level specifications, and assisted in troubleshooting.
* Implemented Kafka for real-time data streaming and integration with various data sources.
* Managed and maintained MongoDB databases for efficient storage and retrieval of NoSQL data.
* Utilized Impala for high-performance querying and analysis of large datasets.
* Orchestrated containerized applications using Docker and Kubernetes for scalability and resource efficiency.
* Implemented AWS Step Functions to automate and coordinate data processing workflows.
* Utilized Informatica for data integration, seamless connectivity, and transformation of diverse data sources.
* Implemented ETL processes using PyTorch and SciPy for machine learning data preprocessing.
* Collaborated on projects using Agile methodologies, tracking tasks and progress using JIRA.
* Managed code repositories on BitBucket and utilized Git for version control.
* Developed interactive dashboards and visualizations using Tableau, providing stakeholders with actionable insights.
* Implemented JSON for data interchange, ensuring compatibility and ease of integration with various systems.
* Implemented JWT for secure authentication and authorization in data processing applications.
* Implemented AWS autoscaling to dynamically adjust computing capacity based on demand, optimizing resource utilization.

**Tech Stack:** Python, Pandas, NumPy, SciPy, PyTorch, Hadoop, MDM, Hive, Pig, AWS, Spark, PySpark, Jenkins, Terraform, ELK, Snowflake, Kafka, Aurora, MongoDB, Informatica, Athena, Docker, Kubernetes, Impala, JSON, Agile, JIRA, BitBucket, Tableau, JWT, GIT

**Five9, San Ramon, CA | Dec 2019 – Feb 2021**

**Data Engineer**

**Responsibilities:**

* Employed Pandas and NumPy for efficient data manipulation and analysis in Python.
* Executed complex SQL queries in SQL Server using T-SQL for data extraction, transformation, and loading (ETL) processes.
* Implemented data solutions on Azure, utilizing Azure Data Factory (ADF) for orchestrating ETL workflows and managing data pipelines.
* Managed and optimized data storage on Azure Data Lake Storage for secure and scalable storage for diverse data sets.
* Leveraged Azure DevOps for Infrastructure as Code (IaC) and version control for efficient collaboration and code management.
* Orchestrated data workflows using Apache Nifi, maintaining a seamless and automated data pipeline.
* Implemented and managed Hortonworks and HDInsight clusters for Hadoop and Spark-based data processing.
* Utilized PySpark and Databricks for large-scale data processing and analytics for optimal performance and scalability.
* Integrated Azure Key Vault and Azure AD for secure storage and management of sensitive information and credentials.
* Implemented security measures in Openshift and Kubernetes for containerized applications and services.
* Implemented Avro for efficient data serialization for compatibility and performance in data exchange.
* Integrated data sources using Apache Nifi and orchestrated complex workflows to streamline data movement.
* Collaborated on projects using JIRA for efficient project management and tracking tasks and progress.
* Automated deployment and configuration management using Ansible, ensuring consistent and reliable infrastructure.
* Managed code repositories on GitHub for version control and collaborative development practices.
* Implemented monitoring solutions using New Relic for real-time performance monitoring and issue detection.
* Developed and maintained Power BI dashboards for visualizing and presenting data insights to stakeholders.

**Tech Stack:** Pandas, NumPy, SQL Server, T-SQL, Hortonworks, Azure DW,  ADF, Azure Azure, DevOps, Azure Python, PowerShell, Apache Nifi, Hadoop, Spark, PySpark, Azure AD, Openshift, Kubernetes, Avro, JIRA, Ansible, GitHub, New Relic, Power BI.

**Yugasa Software Labs, Gurugram, India | July 2015 – Sept 2017**

**Data Engineer**

**Responsibilities:**

* Implemented and maintained the Hadoop ecosystem, including HDFS, Hive, and Apache Spark, for efficient and scalable data processing.
* To build robust and scalable data pipelines, utilize Google Cloud Platform (GCP) services such as Cloud Storage, Cloud Pub/Sub, Dataproc, and Dataflow.
* Executed Hive queries for data retrieval and transformation for optimized performance in data processing.
* Proficient in handling Oracle databases and managing data extraction and integration processes using Sqoop.
* Implemented and optimized data workflows on GCP's BigQuery, leveraging its capabilities for high-performance analytics.
* Designed and implemented data integration tasks using Apache Airflow to orchestrate data workflows in the cloud seamlessly.
* Utilized Git for version control for collaborative development and tracking changes in the codebase.
* Collaborated with teams using Confluence to document project details, specifications, and progress updates.
* Implemented Apache Dataflow and Spark for efficient and distributed data processing, enabling faster insights from large datasets.
* Proficient in Dataprep for data wrangling and cleaning.
* Developed and maintained RESTful APIs for seamless communication between different data ecosystem components.
* Integrated external data sources through RESTful APIs to enrich and expand the scope of available data.
* Utilized Maven for project build and dependency management.
* Implemented log analysis using Splunk for real-time monitoring and troubleshooting of data processing workflows.
* Ensured effective XML data processing, transforming, and extracting relevant information from structured and unstructured XML data.

**Tech Stack:** Hadoop, HDFS, Hive, GCP, BigQuery, Sqoop, Spark, Git, Confluence, Apache Airflow, RESTful APIs,  Oracle, Hive queries, Dataprep, Maven, Splunk, XML.

**Ivy, Hyderabad, India | Aug 2012 – June 2015**

**Data Engineer**

**Responsibilities:**

* Developed and maintained ETL (Extract, Transform, and Load) processes using Python, Scala, and Hadoop for efficient data integration.
* Implemented data processing and analysis using Apache Spark, leveraging its distributed computing capabilities.
* Designed and optimized databases using SQL Server, ensuring data integrity and efficient retrieval.
* Utilized AWS services for building scalable and robust data solutions, including storage, computing, and analytics services.
* Created visualizations and reports using Tableau to communicate insights and trends from large datasets.
* Applied advanced data manipulation and analysis using NumPy, SciPy, and Pandas libraries in Python.
* Wrote and optimized PL/SQL queries for efficient data retrieval and manipulation within Oracle databases.
* Implemented data warehouse solutions with Hive, enabling structured querying for large datasets.
* Utilized Scala for developing scalable and high-performance applications within the Hadoop ecosystem.
* Managed data transfer between Hadoop and relational databases using Sqoop for seamless integration.
* Collaborated with cross-functional teams to gather and understand data requirements, ensuring alignment with business goals.

**Tech Stack:** Python, R, Hadoop, Spark, SQL Server, AWS, Tableau, NumPy, SciPy, Pandas, PL/SQL, Hive, Scala, Sqoop.