

**Veena Gourishetty**

**Data Engineer**

**Professional Summary**

**Data Engineer** with **13+ years** of IT experience in design, develop, support web, enterprise, and **cloud based** applications. Proven track record of optimizing **ETL** processes for efficiency and scalability while ensuring data quality and integrity. Skilled in **Agile** project execution processes, good exposure to version control tool **GitHub**, pipeline automation and orchestration using **Python**, **PySpark**, **Airflow**, **Snowflake** and **Databricks**. Proficient in utilizing the cloud automation (**IaC**) tools such as **Terraform** and **CloudFormation**. Very good experience in logging and monitoring tools and strong analytical and problem solving skills with collaborative mindset to work in a cross-functional team/project environments.

* Good Experience in **AWS** services such as **EC2**, **S3**, **VPC**, **IAM**, **RDS**, **Route 53**, **API Gateways**, **SNS**, **Cloud Watch**, **Cloud formation, Cloud Trail**, **Secret Manager**,**Athena**,**Lambda Functions**, **Step Functions**.
* Proficient in utilizing data engineering service such as **S3**, **AWS Glue**, **Matillion Glue Data Catalog**, **Lambda**, **AWS Event Bridge, Step Functions**, **Athena**, **Secret Manager**, **Kinesis**, **Kafka**, **Redshift**, **Snowflake** and **Databricks**.
* Very good experience in data data warehouse services such as **S3**, **redshift**, **Snowflake.**
* Implemented "Infrastructure as Code" (**IaC**) using **Terraform** and **Cloud Formation.**
* Extensive experience **data ingestions** from various sources such as **file based**, **database based**, **API based** and **stream-data/real-time data** based ingestions.
* Proficient in setting up and execute the **ETL** process for small to medium level datasets and large volumes of data using **serverless Lambda**, **Step Functions** and **EMR** cluster management services from AWS.
* Writing & optimizing ETL (**Extract** , **Transform** and **Load**) jobs in **AWS Glue** (autogenerated scripts) , **Python** , **PySpark** and other relevant tools.
* Data validation, cleansing, enrichment, data transformations handled very well using Python libraries such as **Pandas**, **NumPy**.
* Very good experience in ingesting and process real-time data / streaming data using AWS services such as **Kinesis** and **Apache service Kafka**.
* Extensive hands-on experience in data pipeline automation and orchestration using **AWS Glue**, **Apache Airflow** and **Databricks.**
* Used **PySpark** for large volumes of datasets and distributed data processing also clean, filter and reshape the data as per the required format. Understanding of **AWS security** practices & implementing **IAM**, **VPC**,**MFA**, **data base encryptions, SSL/TLS** for in transit data security **AWS KMS** data security for the data at rest.
* Understand and comply with industry-specific regulations (e.g., **GDPR**, **HIPAA**) and use AWS services like **AWS Artifact** for compliance reports and documentation.
* Familiar with version control systems like SVN, GitHub for collaborative development.
* Strong **SQL** skills for querying and manipulating data from the databases such as **MySQL**, **Oracle**, **Microsoft SQL Serever**, **PostgreSQL** and **TeraData.**
* Very good knowledge and understading in **No SQ**L databases such as **Ammazon DynamoDB**

and **Cassandra** while processing semi and unstructured data with huge data volumes.

* Experience working in **Agile** / **Scrum** project development / execution environments.
* Proficiency in languages like **Python**, **Pyspark** and **Java** for data processing and scripting tasks. Utilizing **AWS CloudWatch** and **AWS CloudTrail** tools to manage reliability, security, track system performance , cost & billing , audit and troubleshoot issues.
* Good exposure to logging and monitoring tools such as **AppDynamics** and **Splunk**
* Very good knowledge in monitor, analyze and visualize data & metrics using Amazon

**QuickSight** and other third party BI tools such as **Tableau**, **Microsoft Power BI**, **Looker**.

* Creating clear **documentation** for data pipelines, workflows processes and maintaining the documents in **Confluence** pages.

**Technical Skills**

* Cloud Platforms: Amazon Web Services (AWS), Microsoft Azure
* Cloud Services: EC2, S3, Data Factory, Databricks, AWS Lambda, Data Lake Storage, Glue, Cloud Formation, SNS, Athena, Athena, EMR ,Azure Functions
* IaC: Terraform , CloudFormation
* Serverless Architecture: AWS Lambda
* ETL Frameworks: Apache Airflow, Matillion, AWS Glue
* Real Time Data Processing: Kinesis, Kafka,Azure Stream Analytics
* Programming Languages: Python, PySpark
* Security Services: IAM, AAD, KMS, Secret Manager
* Data Warehousing: Redshift, Snowflake, Databricks
* Data Visualization: Tableau, Power BI, QuickSight
* Source Code Management: GitHub,Bidbucket
* Monitoring and Logging: CloudWatch, BMC Remedy, Splunk
* Issues & Tracking: Bug Zilla, JIRA,Confluence
* Database Systems: SQL, MySQL, Microsoft SQL Server ,Oracle
* Project Methodologies: SDLC, Agile/Scrum
* Operating Systems: Linux, Windows

**Certifications**

* AWS Certified Solutions Architect Associate

**Education**

* **Master of Computer Applications (M.C. A)** Osmania University, Hyderabad, India. (2010)
* **B. Sc Computer Science** Osmania University, Hyderabad, India. (2007)

**Professional Experience**

**Client: Maryland Department of Information Technologies (MDTHINK) Aug 2022 – Present Role: Data Engineer**

**Responsibilities:**

* + Collaborate with business stakeholders, relevant departments to understand the specific data requirements, data sources, data formats, and business rules.
  + Identify and connect to source systems using python code using its available packages.
  + Designed and implemented data models in SQL-based databases, including S3, Snowflake, PostgreSQL and Oracle.
  + Designed and implemented AWS data pipelines using Matillion, Airflow, Python and PySpark to process and transform the structured and unstructured data.
  + Developed serverless architecturesusing AWS Lambda and Step Functions and real-timedataand event-driven workflows using Kinesis and Kafka.
  + Conducted data cleansing, enrichment using python and PySpark scripting and also the transformation and business logic implementations.
  + Orchestrated and automated data pipelines using Apache Airflow to schedule, monitor and manage ETL/ELTprocesses across various platforms.
  + Design and implement data loading processes to move transformed data to a data warehouse or data lake , Snowflake using services like Amazon Redshift, Amazon S3 and Amazon RDS.
  + Utilized Snowflake and databricks as cloud-based services for managing medium to large-scale data sets for end to end data processings.
  + Optimized loading processes for performance, reliability, and data consistency.
  + Implemented distributed log processing system using Kinesis and Apache Spark to enable real- time analysis and monitoring of application logs.
  + Set up appropriate access controls and security using IAM , MFA, Secret Manager and Bearer Tokens while accessing API based data and encryption mechanisms for data at rest and in transit.
  + Implemented ETL/ELT processes and data transformations using AWS Glue and DBT ensuring high- quality data for downstream analytics and reporting.
  + Monitor and optimize ETL processes for performance, scalability and efficiency using CloudWatch and Splunk.
  + Identify and resolve data quality issues, ETL process failures, and performance bottlenecks.
  + Document ETL processes, data mappings, transformations and workflows for reference and future maintenance using Confluence.

**Environment:** AWS S3, Cloudformation,Terraform, Matillion, Airflow, Python, PySpark, Service Now, Redshift, Lambda, Snowflake, SQL, Datalake, Data Pipeline and Databricks, Confluence and Splunk.

**Client : General Dynamics Information Technology (GDIT) Jul 2019 – Aug 2022 Role: Data Engineer**

**Responsibilities:**

* + Designed and implemented scalable data pipelines using Python and PySpark to process and transform large volumes of structured and unstructured data.
  + Designed and implemented data models in SQL-based databases, including Snowflake, PostgreSQL, and Oracle.
  + Developed serverless architecturesusing AWS Lambdaand Step Functionsfor real-timedata processing and event-driven workflows.
  + Orchestrated automated data pipelines using Apache Airflow to schedule, monitor and manage ETL/ELTprocesses across various platforms.
  + Utilized Snowflake as a cloud-based data warehousing solution, managing large-scale data sets and performingcomplex data manipulations.
  + Implemented a distributed log processing system using Apache Spark, enabling real-time analysis and monitoring of application logs
  + Collaborated with data analysts, data scientists, and business stakeholders to understand data requirements andprovide optimized solutions.
  + Implemented ETL/ELT processes and data transformations using AWS Glue and DBT ensuring high- quality data for downstream analytics and reporting..

**Environment:** AWS EC2, S3,Python, PySpark, AWS Glue, Airflow, Athena, SNS, SQS, EMR, Athena, Redshift, Lambda, DBT, Snowflake, SQL, Datalake and Data Pipeline.

**VBLP Technologies, Hyderabad, India Jun 2014 – Jul 2019 Role: Application Engineer**

**Responsibilities:**

* + Application Development: Design, develop, test, and deploy Java-based enterprise applications using J2EE technologies such as Servlets, JavaServer Pages (JSP), JavaServer Faces (JSF), Enterprise JavaBeans (EJB), and others.
  + Deployment and Configuration: Prepare deployment packages, configure application servers (e.g., Apache Tomcat, Wild Fly, WebSphere), and deploy applications to various environments.
  + Coordinate with architecture teams to design new software technologies for clients
  + Test and implement constant updates while identifying improvement opportunities
  + Automation the deployment of applications using cloud formation.

**Environment:** Java, RESTful Services, Maven, Eclipse, JavaScript, Junit, Log4j, Jenkins, Java, J2ee, HTML, CSS, JSP, Servlets, Java Script, Struts, Windows

**Ageis Limited Jun 2011 – May 2014**

**Role: Systems Engineer**

**Responsibilities:**

* + An excellent team player and should be able to take responsibility for the tickets assigned.
  + Proven hands-on Software/Application support with core Java
  + Object-Oriented Analysis and debug using common design patterns.
  + Profound insight of core Java and JEE internals (Class loading, Memory Management, Transaction management, etc.) will be added advantage)
  + Experience with test-driven development
  + Excellent debugging skills, Strong problem-solving skills
  + Experience with Version Control Systems like SVN, GIT, TFS, etc.
  + Excellent client-facing skills along with excellent written and verbal communication skills
  + Installed and administered Artifactory repository to deploy the artifacts generated by Maven and to store the dependent jars which are used during the build.
  + Involved in setting up JIRA as defect tracking system and configured various workflows, customizations, and plugins for the JIRA bug/issue tracker.
  + Mentor Business Areas on Subversion Branching, Merging strategies concepts
  + Resolved update, merge and password authentication issues in Bamboo and JIRA
  + Involved in writing and organizing Shell and Perl scripting for building complex software systems.
  + Migrated source control tool from Clear Case to Subversion (SVN) and build tool from Ant to Maven.
  + Used Maven as a build tool on Java projects for the development of the build artifacts on the source code.
  + Involved partially in deploying WARs/EARs (backend) through WebLogic Application Server console.
  + Performed setup of Clustered environment with WebLogic Application Server.

**Environment:** Java, J2ee, HTML, CSS, JSP, Servlets, Java Script, Struts, XML, Oracle, Web services, BMC Remedy, Clear Quest,SVN