**Mounisha Gotte Employer Details:**

**Email:** [**mounishag034@gmail.com**](mailto:mounishag034@gmail.com) **M. Srinath**

**Ph: 443-636-7123** 518 804 0007

srinathm@raytcs.com

* Over 10+ years of extensive experience working as a Data Engineer supporting AI Applications.
* Worked with both Google cloud service (**GCP**) and Amazon web Services (**AWS**).
* Responsibilities include experience with **python** production quality scripting, data wrangling, data visualization and showcasing results to business strategy team.
* Experience in **AWS Services like S3, Lambda, Glue, EMR, Step Function, IAM, CloudWatch, Athena, Redshift, etc.,**
* Design, code, and test solutions for a major, global re-platforming effort using **Apache Spark**
* Experience in **Bigdata** and **Hadoop** eco system, with **Apache Spark, Apache Kafka, Apache Flume** and **Apache Hive**
* Experience in **GCP services** like **BigQuery**, **DataProc**, **Compute Engine**, **Cloud Storage, Cloud RUN, Cloud Function** and **Cloud Composer**.
* As a **GCP Admin**, I have managed user access to GCP resources, including granting and revoking permissions, **setting up roles and permissions, and enforcing access policies**.
* I have monitored resource usage to ensure that resources are being used efficiently and cost-effectively.
* As a GCP Admin, I must **maintain** **up-to-date documentation** of the organization's GCP resources, **including configuration details, access policies, and usage logs.**
* Excellent interpersonal and communication skills, creative, research-minded, technically competent, and result-oriented with problem solving and leadership skills.
* Experience in **Extraction, Transformation and Loading (ETL)** data from various sources into Data Warehouses, as well as data processing like collecting, aggregating, and moving data from various sources **using Apache Flume, Kafka, PowerBI** and **Microsoft SSIS**.
* Design, code, and test solutions for a major, global re-platforming effort using **Apache Spark**
* Worked on implementing advanced procedures like text analytics using **Apache Spark** in **Scala**
* Orchestrated multiple **Hadoop** application jobs using **Oozie**.
* Managed Docker orchestration and **Docker** containerization using **Kubernetes**.
* Used **Kubernetes** to orchestrate the deployment, scaling, and management of **Docker** Containers.
* Skilled in monitoring and alerting using Prometheus.
* Experience in implementing optimization techniques in **Hive** and **Spark**.
* Exposure on **Big Data** ecosystems using **Hadoop, Pig, Hive, HDFS, MapReduce, Sqoop, Storm, Spark, Airflow, Snowflake, Teradata, Flume, Kafka, Yarn, Oozie**, and **Zookeeper**.
* Used IDEs like **Eclipse, IntelliJ IDE, PyCharm IDE, Notepad ++,** and **Visual Studio** for development.
* Worked on supporting projects using technologies like **Teradata** and **Informatica**.
* Worked in an Agile environment and used Rally and Jira tool to maintain the user stories and tasks.
* Proficient in Java for data processing and ETL (Extract, Transform, Load) tasks.
* Good experience in writing Spark applications using **Python** and **Scala**.
* Used Scala to develop Scala-coded spark projects and executed using spark-submit.
* Proficient in **SQL, PL/SQL**, and **Python** coding.
* Experience developing On-premises and Real-Time processes.
* Developed Dashboards and reports using Tableau and Power BI.

**Skill matrix:**

|  |  |
| --- | --- |
| **Cloud** | AWS and GCP |
| **Programming Languages** | SQL, Shell Script, Java ,Python and PySpark |
| **Big data Frameworks** | Hadoop, MapReduce, spark, Hive,Impala,Hue |
| **Databases** | MS-Access, Microsoft SQL Server, Oracle, DB2, PL SQL,  MYSQL, Mango DB, etc. |
| **AWS Services** | Lambda, S3, Athena, Glue, Step Function, EMR, EC2, CloudWatch, IAM, SNS, Sage Maker, Data Pipeline, etc. |
| **GCP Services** | Cloud Function, DataFlow, Composer, DataProc, BigQuery, Cloud Storage, Cloud Run, etc. |
| **Orchestration tools , Monitoring, and altering** | Kafka,Crontab, Airflow, Kubernetes, Docker and Prometheus. |
| **Devops** | Terraform and Jenkins |
| **Visualization tools** | Tableau, Power BI |

|  |
| --- |
| **Discover Financial Services, Chicago, IL DEC 2020 – TILL DATE**  **AWS Data Engineer/Data Analyst** |

**Responsibilities:**

* Used Agile software development methodology in defining the problem, gathering requirements, development iterations, business modeling, and communicating with the technical team for the development of the system.
* Expertise in implementing and managing robust and scalable SQL and NoSQL databases such as MySQL, MongoDB, and SQL servers.
* Obtained the data from the AWS, RDS, MYSQL of the client to perform the ETL operations by using Informatica.
* Involved with project team members to deliver data models to meet data requirements and create all documentation and deliverables in accordance with enterprise guidelines and standards. Work on Metadata transfer amongst various proprietary systems.
* Build Self-service data pipelines using **AWS Services** like **SNS, StepFunction, Lambda, Glue, EMR, EC2, Athena, Sage Maker, QuickSight, Redshift,** etc.
* Moved large amounts of data from AWS S3 buckets to **AWS Redshift** using **Glue** and **EMR**.
* Analyzed large and critical datasets using **EMR, Glue,** and **Spark**.
* Experienced in writing live Real-time Processing and core jobs using **Spark** Streaming with **Kafka** as a data pipeline system.
* Managed datasets using Panda data frames and MYSQL. Queried the database queries using Python-MySQL connector and retrieved information using MySQL db.
* Coordinated daily, weekly, bi-weekly, monthly, and quarterly oversight monitoring activities using **AWS QuickSight.**
* Developed ETL pipelines in and out of data warehouse using tools like **Python** and **AWS Glue.**
* Implemented Spark using **Scala** and Spark **SQL** for faster testing and processing of data.
* Consumed the data from **Kafka** using **Apache spark.**
* Used **Spark** and **Spark-SQL** to read the parquet data and create the tables in hive using the **Scala API**.
* Building models in Spark using python in **Databricks** and **Sage Maker Notebooks**.
* Worked on migrating **SQL** scripts from **Redshif**t and **Athena**.
* Led the design and implementation of data pipelines using **Hadoop** ecosystem tools, including **Impala** and **Hue**, resulting in a 30% improvement in data processing efficiency.
* Conducted performance tuning and optimization of **Impala** queries, reducing query execution times by 40%.
* Utilized **Hue** for managing and monitoring **Hadoop** clusters, simplifying cluster management and improving operational efficiency.
* Experience in building CICD pipelines for testing and production environments using **Terraform**.
* Proficient with container systems like **Docker** and container orchestration like **EC2Container Service**, **Kubernetes**, worked with **Terraform**.
* Implemented monitoring solutions using Prometheus to ensure the reliability and performance of data processing applications.
* Worked on POC’s with **Apache Spark** using **Scala** to implement **spark** in project.
* Used **AWS EMR** to **transform** and move large amounts of data into and out of other **AWS** data stores and databases, such as **Amazon Simple Storage Service** (**Amazon S3**) and **Amazon DynamoDB**.
* Designed and implemented **data architectures on Snowflake** that enable efficient and reliable **data processing, storage, and retrieval**.
* Built and maintained data pipelines using tools such as **Apache Spark**, **Python, and Snow SQL**
* Developed and managed data models on **Snowflake to transform raw data into a structured format** that is suitable for analysis.
* worked with large volumes of data from various sources, ensuring that the data is cleansed, transformed, and loaded into the data warehouse promptly and efficiently.
* I also have experience in **Tableau**, which is a powerful data visualization tool used for creating **interactive dashboards and reports**. I am responsible for designing and developing **Tableau dashboards** that enable end-users to explore and analyze healthcare data.
* **Experience with Tableau**, including designing and **developing dashboards and reports**. Strong problem-solving skills and attention to detail.
* Worked with healthcare data analysts and scientists to understand their data needs and developed solutions that meet those needs.
* Collaborating with other members of the healthcare project team to ensure that data pipelines are integrated with other systems and applications.
* up to date with the latest industry trends in data engineering and **Snowflake** to ensure that the healthcare project is using the **most efficient and effective solutions**.

|  |
| --- |
| **Fonteva, Arlington, VA MAY 2019 – NOV 2020**  **AWS Cloud Data Engineer** |

**Responsibilities:**

* Used Agile software development methodology in defining the problem, gathering requirements, development iterations, business modeling and communicating with the technical team for development of the system.
* Built Real-time and batch pipelines using **GCP services.**
* Modeled and developed a new data warehouse from scratch and then migrated that data warehouse to **BigQuery**. Automated data pipelines and quality control checks.
* Architected and implemented a solution to migrate the data platform from **Hadoop** to **Google cloud Platform.**
* Experienced in designing and developing applications using relational databases like Oracle, and MySQL.
* Worked with different file formats like Parquet files and **Impala** using PySpark for accessing the data and performed Spark Streaming with RDDs and Data Frames.
* Worked on **Impala** query engine for data store on distributed cluster for various business user cases on top of Hive table.
* Involved in working with data extracted from two different sources like MYSQL, Web Servers.
* Extracted of data from MySQL is performed by SQOOP on to HDFS.
* Queried both Managed and External tables created by **Hive** using **Impala**.
* Worked on developing continuous integration (CI), continuous delivery (CD), and continuous training (CT) for the ML system using **Cloud Build.**
* Worked heavily with the business to determine reporting requirements and to explore the data with each other.
* Deployed various **Hadoop** applications in EMR - Hadoop, Hive, Spark, HBase, **Hue**, **Glue**, Oozie and Presto etc. based on the needs.
* Assisted in the development of data pipelines, gaining hands-on experience with Hadoop, Impala, and Hue.
* Used Cloud functions, Cloud **pub-sub,** and **BigQuery** to build streamlined dashboard to monitor services.
* Worked on processing the data and testing using **Spark** **Python** and on real-time processing by Spark Streaming and **Kafka** using Hadoop.
* Scripted using **Python** and PowerShell for setting up baselines, branching, merging, and automation processes across the process using GIT.
* Worked on scripting with Python in **Spark** for transforming the data from various files like Text files, CSV, and JSON.
* Created jobs using **Cloud Composer** (**Airflow DAG**) to migrate data from **Data Lake** (cloud storage) to transform it using DataProc and ingest in **BigQuery** for further analysis.
* Member of the Enterprise Architecture Committee and presented multiple times to this group and other business leaders.
* Engineered ETL packages using SSIS to extract data from diverse sources and utilized **SQL** Server to maintain data integrity.
* Migrated tasks adeptly from Oracle apps to **PostgreSQL** databases, showcasing versatility in diverse data landscapes.
* Built batch and streaming jobs using **GCP** Services like **BigQuery**, **Pub/Sub**, **DataProc**, **Dataflow**, **Cloud Run**, **Compute engine** and **Cloud Composer**.
* Helped to implement **PowerBI** in the organization. Developed PowerBI reports for every department in the company.
* Utilizing analytical, statistical, and programming skills to collect, analyze and interpret large data sets to develop data-driven and technical solutions to difficult business problems using tools such as **SQL**, and **Python**.
* Worked on Streaming data transfer, data from different data sources into **HDFS**, **NoSQL** databases.
* Developed **PySpark** scripts for ETL using **DataProc**.
* Good experience with **GCP Stack driver** Trace, Profiler, Logging and Error Reporting, Monitoring
* Setup and Implement Continuous Integration and Continuous Delivery (CI & CD) Process stack using   **GIT**, **Jenkins**
* Experience in developing APIs in cloud.
* Good Knowledge of **IAM** rules and **Cloud Security.**
* Designed **PowerBI** data visualizations using cross tabs, maps, scatter plots, pie charts etc.

|  |
| --- |
| **SWIFT, Manassas, VA NOV 2017 – APR 2019**  **AWS Big Data Engineer** |

**Responsibilities:**

* Built smart data pipelines to read data from multiple sources like **Docker** and **Kubernetes** applications and load data into **HDFS** and **AWS S3**.
* Used **AWS services** like **AWS Services** like, **SNS**, **StepFunction**, **Lambda**, **Glue**, **EMR**, **EC2**, **Athena**, etc.
* Developed and managed data models on **Athena** to **transform raw data** into a structured format that is suitable for analysis.
* Ensured that data pipelines are **scalable, reliable, and secure**, and that they adhere to best practices for data governance and compliance.
* Developed scripts using pandas to easily perform read/write operations to CSV files, manipulate and compare data by columns.
* Cleaned data and processed third-party spending data into maneuverable deliverables within specific formats with Excel macros and **Python** libraries.
* Gathering business requirements by organizing and managing meetings with business stakeholders and IT analysts on a scheduled basis.
* Analyzed the business requirements by dividing them into subject areas and understood the data flow.
* Conducted one-on-one sessions with business users to gather data warehouse requirements.
* Created and updated views with **python** view controller and template language to create new functionalities to the website.
* Built development environment using bug-tracking tools like **Jira**, **Confluence**, and version controls such as **Git**, and **SVN**.
* Provided consultation as required throughout the project life cycle.
* Designed, deployed, maintained, and lead the implementation of Cloud solutions and underlying technologies.
* Involved in migrating tables from **RDBMS** into Hive tables using **SQOOP** and later generated data visualizations using **Tableau**.
* Experienced in writing Spark Applications in **Scala** and **Python**.
* Data analysis on the BI side includes customer profitability, sales trend, inventory turnover, variance between standard and actual product cost.
* Implemented **Spark** using **Scala** and **Spark SQ**L for faster testing and processing of data.
* Implemented **Spark** using **Scala** and utilizing **Data frames** and **SparkSQL API** for faster processing of data.
* Building/Maintaining Docker container clusters managed by **Kubernetes Linux**, **Bash**, **GIT**, **Docker**, on **GCP (Google Cloud Platform).** Utilized Kubernetes and **Docker** for the runtime environment of the **CI/CD** system to build, test deploy.
* Worked on Metadata transfer among various proprietary systems using XML.

|  |
| --- |
| **ICICI Lombard General Insurance Company, MUM, IND JUNE 2014 – APR 2017**  **ETL Developer/Data Engineer** |

**Responsibilities:**

* Installed, Configured and Maintained Apache Hadoop clusters for application development and **Hadoop** tools like **Hive, Pig, Zookeeper, Kafka** and **Sqoop**.
* Integrated HDP clusters with Active Directory and enabled Kerberos for Authentication.
* Installed and Configured **Sqoop** to import and export the data into Hive from Relational databases.
* Administering large **Hadoop** environments build and support cluster set up, performance tuning and monitoring in an enterprise environment.
* Monitoring the **Hadoop** cluster functioning through **MCS** and worked on NoSQL databases including **HBase**.
* Used Hive and created **Hive** tables and involved in data loading and writing Hive UDFs and worked with **Linux** server admin team in administering the server hardware and operating system.
* Configured **Spark Streaming** to receive real time data from the **Kafka** and store the stream data to **HDFS**.
* Designed and Developed data mapping procedures ETL-Data Extraction, Data Analysis and Loading process for integrating data using R programming.
* Implemented Partitioning, **Dynamic Partitions**, **Buckets** in **HIVE**.
* Developed story telling dashboards in **Tableau** Desktop and published them on to **Tableau** **Server** which allowed end users to understand the data on the fly with the usage of quick filters for on demand needed information.
* Close monitoring and analysis of the **MapReduce** job executions on cluster at task level and optimized **Hadoop** clusters components to achieve high performance.

**Technologies:**

|  |
| --- |
| Apache Hadoop, HDFS, Kafka, Flume, Airflow, Hive, Docker, Hive, MapReduce, SQL, Python, Spark, SparkSQL etc., |

|  |
| --- |
| **Myntra, BLR, India MAY 2012 – MAY 2014**  **SQL Developer** |

**Responsibilities:**

* Involved with **Business Analysts** team in requirements gathering and in preparing functional specifications and changing them into technical specifications.
* Created incremental refreshes for data sources on **Tableau** server.
* Responsible for resolving errors, generating client payment data, transmitting data to clients, and ensuring conflict resolution utilizing **MS Access, MySQL, Perl** scripting and other various tools.
* Wrote **PL/SQL** statement and stored procedures in DB2 for extracting as well as writing data.
* As a PL/SQL developer, I have migrated the legacy **SQL stored procedures** to Oracle. I have a solid understanding of **both SQL and PL/SQL**, as well as experience in migrating data from one platform to another.
* To begin with, I have analysed the existing **SQL stored procedures** and identify any compatibility issues with Oracle. I also rewrote some of the code to make it compatible with the **Oracle database.**
* I also have experience with data visualization tools and creating dashboards using Power BI. I have experience with **Power BI,** and a good understanding of **data visualization principles**.
* Worked extensively on SQL querying using **Joins, Alias, Functions, Triggers**, and **Indexes**.
* Developed mappings in IBM Infosphere DataStage to load the data from various sources includes SQL server, DB2, Oracle, Flat files into the Data Warehouse.
* Cleansed, Extracted, and analyzed business data on regular basis and generated ad-hoc analytical reports using Excel and **T-SQL.**
* Responsible for database creation logins & user’s creation, granting object level permission.
* Worked together with Implementation team to assure a smooth transition from the design to implementation phase.