

**Name: - PAVITRA G**

Phone: - 609-606-1013

Email ID: - [vinay@centillioninfotech.com](mailto:vinay@centillioninfotech.com)

**PROFESSIONAL SUMMARY**:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* Overall, 8+ years of experience in the IT industry and played various roles of AWS cloud engineer, Azure cloud engineer while serving clients in various domains such as in **Retail, Banking, Healthcare, Information Technology Services.**
* Implementing Cloud Solutions using various AWS services, including EC2, Virtual Private Cloud (VPC) Glacier, EFS, Lambda, Cloud Formation, Elastic Cache, and RDS.
* Strong understanding of SDLC (Software Development Life Cycle) and SQA Methodology
* Strong experience developing end-to-end data pipelines, automating, and maintaining different data pipelines in production.
* Used Kubernetes to orchestrate the deployment, scaling and management of Docker Containers.
* Managed Docker orchestration and Docker containerization using Kubernetes.
* Strong experience with AWS (Amazon Web Services), Elastic Map Reduce (EMR), Storage S3, EC2 instances, Lambda, and kinesis.
* A solid grasp of distributed systems architecture and parallel processing frameworks like Spark and MapReduce.
* Experience in analyzing data using HiveQL, HBase, and custom Map Reduce programs in Python.
* Strong experience utilizing Spark RDD, Spark Data frames, Spark SQL, and Spark Streaming APIs extensively.
* Used AWS EMR to transform and transport huge amounts of data into and out of other AWS data stores and databases like Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB.
* I have used amazon ECS for deploying, managing, and scaling containerized applications.
* Working on implementing AWS services using ELB, RDS, VPC, Auto Scaling Groups.
* Excellent Knowledge of AWS Lambda for event-driven, serverless compute platform in AWS to manage and deploy resources.
* Strong experience building real-time streaming pipelines using kinesis and Spark Structured Streaming.
* Experienced with importing and exporting data using Sqoop from HDFS to RDBMS and vice-versa.
* Ability to design, build, manage and operate cloud infrastructure.
* Worked on building various data ingestion pipelines to pull data from various sources like S3 buckets, FTP servers, and Rest (Representational State Transfer) Applications.
* Used Kerberos, Azure AD, Sentry, and Ranger for maintaining the security
* Detailed exposure to Azure tools such as Azure Data Lake, Azure Data Bricks, Azure Data Factory, HDInsight and Azure SQL Server.
* Strong experience working with Spark for performing large scale data processing, data cleansing, data de-normalizations, data aggregations, etc.,
* Extract Transforming and Load data from Sources Systems to Azure Data services using a combination of Azure Data.
* Developed baseline VPC and Network design, including leveraging VPN connectivity and Direct Connect.
* Migrated legacy applications to AWS cloud environment
* Providing infrastructure solutions on AWS in a fast-paced, challenging, innovative company focused on reliability and customer service.
* Experience on transferring data from RDBMS to HDFS and HIVE table using Py spark.
* Proficient SQL experience in querying, data transformations/extraction, and developing queries for a wide range of applications.
* Having experience in Docker.
* Having knowledge in GIT, Ansible, Maven and Jenkins.
* Quick and adaptable to learn any new technologies and environments
* Ability to manage multiple tasks, Initiative, and Adaptable. Initiative-taking, organized collaborator with serious problem solving and analytical skills and total commitment to the organization's goals.
* Experience in giving users training, creating Technical/Functional documentation and users guide

**Certifications:**

* AWS Certified Developer - Associate (DVA)
* AWS Certified Cloud Practitioner (CLF)

**Technical Skills**:

|  |  |
| --- | --- |
| Big Data Tools | HDFS, Yarn, MapReduce, Spark, Kafka, Kafka Connect, Hive, Airflow, Stream Sets, Sqoop, HBase, Flume, Ambari, Nifi, Sentry, Ranger |
| Amazon AWS | EMR, EC2, EBS, RDS, S3, Athena, Glue, Elasticsearch, Lambda, SQS, DynamoDB, Redshift, ECS, Quick Sight, Kinesis. |
| Microsoft Azure | Databricks, Data Lake, Blob Storage, Azure Data Factory, SQL Database, SQL Data Warehouse, Cosmos DB, Azure Active Directory |
| Scripting Languages | Python, Scala, Shell Scripting, Pig Latin, HiveQL |
| NoSQL Database | HBase, DynamoDB |
| Database | MySQL, Oracle, Teradata, MS SQL SERVER, PostgreSQL, DB2 |
| ETL/BI | Snowflake, Informatica, Talend, SSIS, SSRS, SSAS, ER Studio, Tableau, Power BI |
| Version Control | Git, Bitbucket, Jenkins |

**Professional Work Experience**:

**Client: Dollar Tree Stores, Chesapeake, VA Jan-2021 – Till Present**

**Role: AWS Cloud Engineer**

**Responsibilities**:

* Automate manual data ELT/ETL processes for project managers and business analysts using Airflow and Python scripts to process, RedShift and S3 to store, and Power Bi to visualize the data
* Creating Lambda functions with Boto3 to deregister the unused AMIs in all application regions to reduce the cost of EC2 resources.
* Designed and set up Enterprise Data Lake to provide support for various use cases, including Analytics, storing, processing, and Reporting of voluminous, rapidly changing data.
* Implemented the machine learning algorithms using Python to predict the quantity a user might want to order for a specific item so we can automatically suggest using kinesis firehose and S3 data lake.
* Used Spark SQL for amp and Scala, Python interface that automatically converts RDD case classes to schema RDD.
* Stored data in AWS S3 and performed EMR programs on data stored in S3.
* Performed end-to-end Architecture & implementation assessment of various AWS services like Amazon EMR, Redshift, and S3.
* Experience on Deployment Automation & Containerization (Docker, Kubernetes).
* Changing the AWS infrastructure Elastic Beanstalk to Docker with Kubernetes.
* Developed Kibana Dashboards based on the Integrated different source Log stash data and target systems into Elasticsearch for near real-time log analysis of monitoring End to End transactions
* Used Spark to process the data before ingesting the data into the HBase.
* Creating Lambda functions with Boto3 to deregister unused AMIs in all application regions to reduce the cost of EC2 resources.
* Experienced on AWS Systems Manager (Run command, Automation, and maintenance window)
* Used AWS Athena extensively to ingest structured data from S3 into other systems such as RedShift or to produce reports.
* Implemented Apache Airflow for authoring, monitoring, scheduling, and Data Pipelines
* Ingested sensor data from 4 different sources using S3, AWS EMR, Apache Airflow, and Spark (Py Spark).
* Worked with JSON-based REST Web services.
* Using Apache Airflow for workload scheduling with Directed Acyclic Graphs.
* Hive, as the primary query engine of EMR, we have built external table schemas for the data being processed.
* Set up scalability for application servers using the command-line interface for Setting up and administering DNS system in AWS using Route53.
* Experience in coding on various AWS services like Cloud Formation Templates, Cloud Watch, Cloud Trail, encryption, logging, and Lambda.
* Utilized Python libraries such as boto3, pandas and NumPy to read data from CSV files and aggregate and update data.
* Used Python IDE PyCharm for developing the code and performing the unit test.
* Involved in building database models, APIs, and Views utilizing Python technologies to build web-based applications.
* Complete software development was implemented and designed using scrum process and Agile Methodology.
* Good Understanding of Data ingestion, Airflow Operators for Data Orchestration, and other related python libraries.

Environment: AWS EMR, S3, Map Reduce, RDS, Redshift, Lambda, Boto3, DynamoDB, Apache Spark, HBase, Snowflake, Python, SSRS, Tableau, pandas, NumPy, xml, Django, Apache airflow.

**Client: Atlantic Union Bank, Richmond, VA Dec -2018 – Dec- 2020**

**Role: AWS Cloud Engineer**

**Responsibilities**:

* Worked with data transfer from on-premises SQL servers to cloud databases (Azure Synapse Analytics (DW) & Azure SQL DB).
* Implementing Data masking in conjunction with Data Transparent Encryption (TDE) to encrypt SQL Server Backup files.
* Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics.
* Involved in Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In Azure Databricks
* Used Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform, and load data from different sources like Azure SQL, Blob storage, Azure.
* Implemented Copy activity, get metadata activity, Lookup activity, Custom Azure Data Factory Pipeline Activities
* Experience in data analysis, data modeling and implementation of enterprise class systems spanning big data, data Integration, Object Oriented programming.
* Responsible for estimating the cluster size, monitoring, and troubleshooting of the Spark data bricks cluster.
* Dealt with Storage services named data Lake Storage Gen 1 and Gen 2, Storage Explorer for hosting CSV, JSON and Parquet files and managing access across storage accounts.
* Built ETL solutions using Databricks by executing code in Notebooks against data in data Lake using Delta Lake and loading data into Azure DW.
* Used data Flow debug for effectively building ADF data flow pipelines. Improved performance by using optimization options by effectively using partitions during various transformations.
* Implemented data bricks delta tables to expose the data from global lake into data Lake
* Extract Transform and Load data from Sources Systems to Azure data Storage services using a combination of Azure data Factory, T-SQL, Spark SQL, and U-SQL Azure data Lake Analytics.
* Experienced in performance tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.
* Worked on implementing source control on Azure Databricks environment.
* Implemented Spark Scripts using Scala, Spark SQL to access hive tables into spark for faster processing of data.

Environment: Cosmos DB, Azure Logic Apps, Azure Data Factory, Azure Databricks, Azure data lake Storage Gen2, Azure Logic Apps, Azure Windows Virtual Desktop, Azure SQL Database, Azure SQL Data warehouse, Azure Log Analytics, Azure Event Hub, Python, Spark, Scala, Py Spark, U-SQL, T-SQL, RDBMS, Oracle, Microsoft SQL Server

**Client: Vangent Inc., Arlington, VA Nov- 2016 – Nov- 2018**

**Role: AWS Cloud Engineer**

**Responsibilities**:

* Worked on building centralized Data Lake on AWS Cloud utilizing primary services like S3, EMR, Redshift, and Athena.
* Worked on migrating datasets and ETL workloads from On-prem (Map R Cluster) to AWS Cloud
* Built a series of Spark Applications and Hive scripts to produce various analytical datasets needed for the digital marketing team
* Worked extensively on PYSPARK to build Big Data flow.
* Build, test, and deploy predictive models through batch & API in an AWS ecosystem (e.g., Sage maker).
* This entire pipeline was orchestrated using a state machine from AWS Step Functions.
* Worked extensively on building and automating data ingestion pipelines and moving terabytes of data from existing data warehouses to the cloud.
* Created Airflow Scheduling scripts in Python. Created job flow using Airflow in Python language and automated jobs.
* Written AWS Lambda code in Python for nested JSON files, converting, comparing, sorting, etc.
* Developed Apache flows dealing with various kinds of data formats such as XML, JSON, and Avro.
* Involved in deploying and designing multi-tier applications using all the AWS services (EC2, Route53, S3, RDS, Dynamo DB, SNS, SQS) IAM focusing on high availability, fault tolerance, and auto-scaling in AWS Cloud Formation
* Participated in all aspects of Software Development Life Cycle (SDLC) and Production troubleshooting, Software testing using Standard Test Tool.
* Involved in writing Spark applications using Scala to perform various data cleansing, validation, transformation, and summarization activities according to the requirement.
* Explored with Spark to improve the performance and optimization of the existing algorithms in using Spark-SQL, Spark Context, Data Frame, and pair RDD.
* Performed the migration of Map reduce and Hive Jobs from on-premises Map R to AWS cloud using EMR.
* Worked on NoSQL support enterprise production and loading data into HBase using Impala and Sqoop.
* Handled import of data from various data sources, performed transformations using Hive and loaded data into aws.
* Implemented the AWS lambda functions using Python to improve the performance of the file upload and merge functionality to AWS S3 buckets within the Amazon Cloud environment.
* Transferred the data from AWS S3 to AWS Redshift.
* Worked on HBase databases which differ from classic relational databases.
* Involved in Agile methodologies, daily scrum meetings, and spring planning.

Environment: Apache NiFi 1.4, ETL, Hive 2.3, HBase 1.4, Flume 1.8, Airflow, Tableau, GIT, MapReduce, JSON, AVRO, Teradata, Maven, SOAP (Simple Object Access Protocol), Aws s3, EMR, EC2, EBS, RDS, S3, Athena, Lambda, SQS, DynamoDB, Redshift, ECS, Quick Sight, Kinesis.

**Client: TCS, India Jan-2015 – Sep-2015**

**Role: Software Engineer**

**Responsibilities**:

* Design and developed web app that manage different data. Use J2EE, Hibernate, Spring, REST, JSON
* Developed GeoTome’s data service store use spring boot, spring cloud, CI, CD for distributed, real-time service.
* Use HTML5, BootStrap3, AngularJs, CSS3 in the DataViewr web app to and reporting data in different devices.
* Implement responsive web design principal to design and developed high degree reusable UI components
* Worked on database layer, designed data schema with Oracle, SQL server, MySQL, Nosql
* Use core java, Java8 and c# developed service layer, provide service for MiVu, GeoThrust
* Consultant with client and Implement business logic to process data, handle and notify event flow in java
* Maintain, debug and test the exiting software to support and communicate with client to provide service.
* Participated architecture design with new technology and research in new technology.
* Developed Android mobile application for company new software, use UI Fragments RxJava, RxAndroid, Retrofit
* Utilized JPA2, Hibernate and ORM in data storage layer for the Drilling Report software.
* Developed WEB APPLICATION user interface with JAVA SCRIPT, JSP, CSS, HTML, XSL, JSTL and JSF

**Client: HCL Technologies Limited, Hyderabad, India Apr-2012 – Dec-2014**

**Role: Software Engineer**

**Responsibilities**:

* Testing numerous web and mobile software apps developed by company clients in Agile environment
* Reviewed technical documentation to identify incomplete and ambiguous requirements
* Performed testing using emulator/simulator as well as Android and iOS devices
* Created and executed test cases using Charles Proxy breakpoints to validate application behavior with different balance points
* Executed test cases and documented test results, reported defects and tracked the status of issues
* Involved in functional positive and negative testing, cross-browser and cross-platform compatibility testing, recoverability testing, end-to-end testing, usability testing, performance, smoke, and regression testing
* Interacted with product management in the review and analysis of functional requirements
* Executed SQL queries to ensure the data integrity throughout the back-end and front-end
* Installed and configured multiple test environments using Virtual Machines
* Defined areas suitable for automation testing, prepared test data
* Tested major workflow on Android with emulator to verify usability of the app main functions
* Tested UI components using AWS and Samsung Device Farm to ensure that the application functions similarly on different Android versions and devices
* Followed Agile Scrum methodology: took part in Sprint Planning, Daily Stand-Up, Sprint Review
* Participated in Bug Triage meetings to discuss and prioritize bugs