**AHMED ANSARI**

**Sr. Data Engineer (Python)**

**+1 (317) 210-8090‬|**

**PROFESSIONAL SUMMARY**

* Senior Python Developer or Cloud Data Engineer with Around 9 years of impeccable experience working on multicloud environment like AWS, AZURE and GCP and involved in designing, developing, testing and implementing various stand-alone and client-server architectures based enterprise application software on different domains.
* Hands on Amazon Web Services (AWS) for creating and managing EC2, Elastic Map Reduce, Elastic Load-balancers, Elastic Container Service (Docker Containers), S3, Lambda, Elastic File system, RDS, Cloud Watch, Cloud Trail, IAM and Kinesis Streams.
* Conducted ad-hoc data analysis on large datasets from multiple data sources to provide data insights and actionable advice to support business leaders according to self-service BI goals.
* Experience in analyzing data using Python, R, SQL, Microsoft Excel,Hive, PySpark, Spark SQL for Data Mining, Data Cleansing, Data Munging and Machine Learning.
* Experience in working with Relational DB (RDBMS) like Snowflake, MYSQL, PostgreSQL, SQLite and No-SQL database MongoDB for database connectivity.
* Experienced with containerization and orchestration services like Docker, Kubernetes.
* Expertise in Build Automation and Continuous Integration tools such as Apache ANT, Maven, Jenkins.
* Strong experience in developing Web Services like SOAP, REST, Restful with Python programming language.
* Experience in using Docker and Ansible to fully automate the deployment and execution of the benchmark suite on a cluster of machines.
* Good Experience in Linux Bash scripting and following PEP-8 Guidelines in Python.
* Extensive Knowledge on developing Spark SQL jobs by developing Data Frames.

**TECHNICAL** SKILLS:

|  |  |
| --- | --- |
| **Programming Languages** | Python, Java, Scala, , Perl, C and C++. |
| **Database** | MySQL, PostgreSQL, Teradata, Snowflake, SQLite, MongoDB |
| **Cloud Computing** | AWS, Azure, OpenStack |
| **AWS** | Amazon EC2, S3, , EFS, RDS, EMR, Kinesis, ELB, IAM, EBS, Lambda. |
| **Web Technologies** | Java Script, JQuery, CSS, HTML, AngularJS |
| **Python Libraries/Packages** | PyTables, Data Frames, NumPy, PySide, Pandas, SQL, Data Frames, Matplotlib. |
| **ETL** | Informatica, Datastage, SSIS |
| **Version control tools** | Git, SVN, Bitbucket, CVS |
| **Automation tools** | Puppet, Chef, Ansible, Kickstart, Jumpstart, Terraform |
| **Frameworks** | Bootstrap, Hibernate and Django, |
| **Operating System** | UNIX, Linux, HPUX, Windows, Red hat Linux 5.x/6.x, Ubuntu |

**PROFESSIONAL EXPEREINCE:**

**Client : Bank of America Oct 2022 to Present**

**Role : Sr. Python / Data Engineer**

* Developed and implemented Print Big Data solutions in AWS cloud environment.
* Responsible for designing and onboarding new data source pipelines.
* Created, designed, and optimized ETL pipeline for different data products for printers’ data.
* Automated ETL pipeline using Airflow, Microsoft Azure DevOps, and Databricks.
* Used Python as a programming language to build the pipeline in Databricks and Airflow.
* Worked with Data Lake and Redshift table creation amendment, and updates.
* Developed and modified existing libraries for automating the ETL pipeline processing and
* performance tuning.
* Created real-time data streaming pipelines using GCP Pub/Sub and Dataflow for financial data.
* Experience in GCP Dataproc, GCS, Cloud functions, BigQuery and moving data between GCP and Azure using Azure Data Factory.
* Worked with GCP services like Cloud Storage, Compute Engine, App engine, Cloud SQL, Cloud Bigtable and Pub/Sub to process data for the downstream customers.
* Configured interactive clusters as per requirements for different jobs in Databricks.
* Utilized use JSON, Yaml, Terraform file formats and REST API calls for job processing.
* Working on GCP services like computer engine, cloud storage, cloud deployment manager.
* Working on to run all the jobs from Jenkins to GCP
* Actively involved in the deployment of the applications and data into existing systems and AWS
* cloud environments using CI/CD methodologies – Azure DevOps.
* Developed and Automated tools for creating and validating SQL scripts based on metadata files.
* Plotted Dashboards using Splunk for internal and external team monitoring.
* Actively involved in documentation, reporting, analysis, and knowledge transfer for new hires.
* Ensured timely completion of projects and recognized by team in first month of work for
* identifying root cause of complex bug before moving to production.

**Technologies Used:** Databricks, Redshift, Airflow, Spark, Python, Scala, Azure DevOps, Splunk, AWS, , Cloud run, BigQuery, Jenkins, Tableau, Kubernetes, Docker,GCP, Jira, Miro.

**Client:DukeEnergy  
Role: Data Architect 2 August 22 to Oct 22  
  
Responsibilities:   
AMIDA :** AMIDA pipeline is responsible for ingesting the meter voltage, usage data from all the Duke Energy meters.

As part of this pipeline I have been part of the below :

* Integration the AMIDA pipeline with 3rd party client to enable their EV vehicle's meter enrollement.
* To enable NRT enrollment of EV meter I implemented AWS Lambda to connect to 3rd party's Azure EventHub and consume their enrollement requests, also leveraged DynamoDB for offset management
* Built a batch analytics pipeline to calculate the voltage and usage metrics to bucketize them and store it in DynamoDB for quick retrieval
* Used step functions to schedule and trigger the batch pipeline
* Used EventBridge for triggering the AWS Lambda consumer for NRT requirement of Azure EventHub Consumer
* Used Terraform for deploying of all the above pipeline

**Tech Used :**

* Python
* AWS Lambda
* AWS EMR
* DynamoDB
* AWS Step Functions
* AWS EventBridge
* Terraform

**Client: AIG, New York, NY**

**Role: Python Developer/ Cloud Data Engineer Jan 2020 – July 2022**

**Responsibilities:**

* Created S3 buckets maintained them and utilized the policy management of S3 buckets and Glacier for storage and backup in AWS.
* Used Docker and Ansible to fully automate the deployment and execution of the benchmark suite on a cluster of machines.
* Worked on Amazon Redshift and AWS a solution to load data, create data models and run BI on it.
* Worked on AWS S3 bucket integration for application and development projects.
* Worked on ETL pipeline to source these tables and to deliver this calculated ratio data from AWS to Datamart (SQL Server) & Credit Edge server
* Scheduled Airflow DAGs to run multiple Hive and Pig jobs, which independently run with time and data availability
* Developed Automated Framework for Data Extraction from all kinds of Data sources using Python/Flask connecting Snowflake, PostgreSQL as Data Warehouse.
* Leveraged spark (Pyspark) to manipulate unstructured data and apply text mining on user's table utilization data.
* Analyzed SQL scripts and designed the solutions to implement using PySpark.
* Implemented Hadoop clusters on processing big data pipelines using Amazon EMR and Cloudera whereas it depended on Apache Spark for fast processing and for the integration of APIs
* Troubleshooting pipelines submitted using Apache Spark and Hadoop services.
* Used JSON and XML SerDe's for serialization and de-serialization to load JSON and XML data into Hive tables.
* Used SparkSQL to load JSON data and create Schema RDD and loaded it into Hive Tables and handled structured data using SparkSQL.
* Using Jenkins pipelines to drive all Microservices builds out to the Docker registry and then deployed to Kubernetes, Created Pods and managed using Kubernetes.
* Developed a fully automated continuous integration system using Git, Jenkins, MySQL and custom tools developed in Python and Bash
* Developed entire frontend and backend modules using Python on Django Web Framework.
* Developed Python code for instantiating multi-threaded application and running with other applications and Designed and developed Database management using POSTGRESQL.
* using collections in Python for manipulating and looping through different user defined objects.
* Maintaining the scripts using the GIT version Control and Maintaining customer database using MS EXCEL.
* Automated resulting scripts and workflow using Apache Airflow and shell scripting to ensure daily execution in production.
* Install and configure Apache Airflow for S3 bucket and Snowflake data warehouse and created dags to run the Airflow.

**Client: Capital One, McLean, VA July 2018 –Dec 2019**

**Role: Cloud Python Developer/ Data Engineer**

**Responsibilities:**

* Involved in development of Python APIs to dump the array structures in the Processor at the failure point for debugging. Using Chef, deployed and configured Elasticsearch, Logstash and Kibana (ELK) for log analytics, full text search, application monitoring in integration with AWS Lambda and CloudWatch.
* Importing and exporting data jobs, to perform operations like copying data from HDFS and to HDFS using Sqoop and developed Spark code and Spark-SQL/Streaming for faster testing and processing of data.
* Developing data processing tasks using PySpark such as reading data from external sources, merge data, perform data enrichment and load in to target data destinations.
* Designed Data Quality Framework to perform schema validation and data profiling on Spark (Pyspark)
* Wrote and executed various MySQL database queries from Python using Python-MySQL connector and MySQL dB package. Implemented user interface guidelines and standards throughout the development and maintenance of the website using the CSS, HTML, JavaScript, and jQuery.
* Used Pandas as API to put the data as time series and tabular format for manipulation and retrieval of data.
* Developed Python Framework and integrated with AWS S3, SQS, RDS, Snowflake for continuous Extraction and loading data from several sources. Hosted Application on Elastic Beanstalk with Auto Scaling.
* Added support for Amazon AWS S3 and RDS to host static/media files and the database into Amazon Cloud.
* Worked in development of applications especially in LINUX environment and familiar with all its commands and worked on Jenkins continuous integration tool for deployment of project and deployed the project into Jenkins using GIT version control system
* Managed the imported data from different data sources, performed transformation using Hive, Pig and Map- Reduce and loaded data in HDFS.

Imported data from AWS S3 into Spark RDDs to perform transformations and actions on those RDDs.

* Worked on ETL Migration services by developing and deploying AWS Lambda functions for generating a serverless data pipeline which can be written to Glue Catalog and can be queried from Athena.
* Used AWS data pipeline for Data Extraction, Transformation and Loading from homogeneous or heterogeneous data sources and built various graphs for business decision-making using Python matplot library
* Implemented Apache Airflow for authoring, scheduling and monitoring Data Pipelines
* Worked on big data (Hadoop) environment along with exposure to HIVE, Spark, Cassandra, SQL and ETL components.

**Client: Comcast, Philadelphia, PA Dec 2016 –June 2018**

**Role: Python Data Engineer/AWS Cloud Engineer**

**Responsibilities:**

* Analyzed large and critical datasets using Cloudera, HDFS, MapReduce, Hive, Hive UDF, Pig, Sqoop and Spark.
* Developed Spark Applications by using Scala and Implemented Apache Spark data processing project to handle data from various RDBMS and Streaming sources.
* Worked with the Spark for improving performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Spark MLlib, Data Frame, Pair RDD's, Spark YARN.
* Used Spark Streaming APIs to perform transformations and actions on the fly for building common learner data model which gets the data from Kafka in Near real time and persist it to Cassandra.
* Worked and learned a great deal from AWS Cloud services like EC2, S3, EBS, RDS.
* Used AWS services like EC2 and S3 for small data sets processing and storage, Experienced in Maintaining the Hadoop cluster on AWS EMR.
* Worked on importing and exporting data from snowflake, Oracle and DB2 into HDFS and HIVE using Sqoop for analysis, visualization and to generate reports.
* Involved in file movements between HDFS and AWS S3 and extensively worked with S3 bucket in AWS
* Used EMR (Elastic Map Reducing) to perform bigdata operations in AWS.
* Worked on Apache spark writing python applications to convert txt, xls files and parse.
* Design and maintain databases using Python and developed Python based API (RESTful Web Service) using Flask, SQL Alchemy and PostgreSQL.
* Design and manage API system deployment using fast http server and Amazon AWS architecture.
* Used Python and Django creating graphics, XML processing, data exchange and business logic implementation
* Installed the application on AWS EC2 instances and configured the storage on S3 buckets.
* Scheduled Oozie workflow engine to run multiple Hive and Pig jobs, which independently run with time and data availability.

**Client: JPMorgan Chase, Sep 2015 –Nov 2016**

**Role: Python Developer/ Data Engineer**

**Responsibilities:**

* Use Python unit and functional testing modules such as unit test, unittest2, mock, and custom frameworks in-line with Agile Software Development methodologies.
* Installed Hadoop, Map Reduce, HDFS, AWS and developed multiple MapReduce jobs in PIG and Hive for data cleaning and pre-processing.
* Used EC2 Container Service (ECS) to support Docker containers to easily run applications on a managed cluster of Amazon EC2 instances.
* Unit Test Python library was used for testing many programs on python and other codes.
* Wrote Python scripts to parse JSON documents and load the data in database.
* Implemented SQL Alchemy which is a python library for complete access over SQL.
* Created APIs, database Model and Views Utilization Python in order to build responsive web page application. Performed troubleshooting, fixed and deployed many Python bug fixes of the two main applications that were a main source of data for both customers and internal customer service team.
* Used Pandas API to put the data as time series and tabular format for east timestamp data manipulation and retrieval.
* Build SQL queries for performing various CRUD operations like create, update, read and delete and worked with team of developers on Python applications for RISK management.
* Write Python scripts to parse JSON documents and load the data in database.

**Client:, Global Tele Systems, India Feb 2014 –Aug 2015**

**Role: Jr. Python Developer/ Software Developer**

**Responsibilities:**

* Developed Cloud Formation templates, also launched AWS Elastic Beanstalk for deploying, monitoring and scaling web applications using different platforms like Docker, Python etc.
* Used Amazon Web Services (AWS) for improved efficiency of storage and fast access.
* Developing applications using RESTFUL architecture using Node.js and PHP as backend languages.
* Designed and maintained databases using Python and developed Python based API (RESTful Web Service) using Flask, SQLAlchemy and PostgreSQL.
* Implemented monitoring and established best practices around using elastic search.
* Followed AGILE development methodology to develop the application.
* Developed web-based applications using Python, Django, XML, CSS, HTML, JavaScript, AngularJS, jQuery and Bootstrap.
* Responsible for gathering requirements, system analysis, design, development, testing and deployment.
* Wrote program to fetch data and from amazon cloud and send the client according their requirement.
* Passed query and wrote script to automate machine alarm data with timestamp to engineer.
* Used Python and Pandas library, built data analysis graph for documentation and record.