Atherin
DATA ENGINEER
469 573 4105
Atherin00018@gmail.com
Green Card

PROFESSIONAL SUMMARY:

- Skilled Data Engineer with 4+ years of experience architecting, building, and optimizing large-scale data processing systems. Proficient in diverse programming languages, including Python and Scala Skilled in leveraging big data technologies such as Hadoop, Apache Spark, and Kafka for efficient data processing and streaming.
- Expertise in data integration and ETL processes using AWS Glue, Informatica, and SSIS. Hands-on experience with cloud platforms, including AWS, and proficient in services like Lambda, S3, EMR, and RDS.
- Experience in Data Engineering, Data Analysis/Business analysis, ETL Development, and Project Management.
- Programming languages: Experience working in Python, Scala, Java, C++, shell scripting on UNIX and LINUX (Ubuntu).
- Databases: Experience in Oracle, MySQL, Sql, Hive, Impala, Postgres & NoSQL databases like
- DynamoDB & MongoDB.
- Design and development of Java/Scala and Python based Spark applications.
- Experienced in Application development with Object oriented programming (OOPS) with Python, Java, C++.
- Developed Python APIs to retrieve, analyze and structure data from NoSQL platforms like HBASE.
- and DynamoDB. Developed "Business Intelligence" scripts for Data-analysis in HIVE.
- Deployment: Experience in CI/CD tools like Jenkins, Bitbucket, Ansible, Maven, Ant, Git, SVN.
- AWS: Experienced in deploying applications on AWS EMR, Glue, SNS,API Gateway, S3, Cloud watch, EC2, Lambda, Step functions, Athena, DynamoDB, Service catalogue, Glue , Redshift.
- Experienced with efficient data warehousing environments like Redshift.
- Developed the data ingestion tool with Spark-Scala, Py-spark and python scripts.
- Experienced in application development for Insurance, Telecommunications & Financial domains.
- Proven track record in designing and managing Snowflake and Amazon Redshift data warehouses, ensuring seamless data storage and retrieval. Extensive knowledge of databases, including MySQL, MongoDB, DynamoDB and Oracle. Proficient in real-time data streaming with AWS Kinesis and Apache Kafka.
- Adept at Infrastructure as Code using Terraform, strongly focusing on automation and scalability. Skilled in data processing tools such as Pandas, NumPy, PySpark, and Apache Airflow. Experienced in containerization and orchestration with Docker and Kubernetes.
- Strong background in monitoring and logging using CloudWatch and Splunk. Proficient in visualization and BI tools such as Tableau and Power BI, translating complex data sets into insightful visualizations. Well-versed in DevOps practices, utilizing Git, GitHub and Maven for version control and continuous integration.
- Collaborative team player experienced in Agile and Scrum methodologies, ensuring efficient project delivery through clear communication and effective task management using JIRA.
- Proficient in creating reliable, scalable, and high-performance applications, especially in the realm of web development and automation.

AREAS OF EXPERTISE

Programming Languages: Python, SQL, Scala,C++ and Ruby
Big Data Technologies: Hadoop, Apache Spark, Databricks,Data Modeling,Data Administration,Data Warehouse

Data Integration and ETL: AWS Glue, Informatica. SSIS, Talend	
Cloud Platforms: AWS, Lambda, S3, EMR, RDS, Step Functions, ECS,AI	
Data Warehousing: Snowflake, Amazon Redshift	
Databases: MySQL, MongoDB, DynamoDB, Oracle	
Real-Time Data Streaming: AWS Kinesis, Apache Kafka	
Data Processing Tools: Pandas, NumPy, PySpark, Apache Airflow	
Containerization and Orchestration: Docker, Kubernetes	
Monitoring and Logging: CloudWatch, Splunk	
Visualization and BI Tools: Tableau, Power BI	
DevOps and Infrastructure: Git and GitHub, Maven	
Methodologies and Collaboration Tools: Agile, Scrum, JIRA,SDLC	
Operating Systems: Windows, Linux	

PROFESSIONAL EXPERIENCE

Client: Nuvance Health-CT
Role: Data Engineer
Duration: Nov 2022 - Present
Roles & Responsibilities:

- Involved in writing Spark applications using Python to perform various data cleansing, validation, transformation and summarization activities according to the requirement.
- Proficient in creating reliable, scalable, and high-performance applications, especially in the realm of web development and automation.
- Created various spark applications using Scala to perform various enrichment of these click stream data combined with enterprise data of the users.
- Design & Development of spark applications for data extraction, transformation, and aggregation from multiple sources.
- Performance tuning of Spark application for setting right batch interval time, correct level of parallelism and memory tuning.
- Application design for integration with REST API's, Merchant UI and custom python libraries.
- Building framework for Data migration and developing spark jobs.
- Translate business and data requirements into Logical data models in support of Enterprise Data Models, OLAP, OLTP and Analytical systems.
- Visualization of store layout with adjacency, left/right, opposite & perpendicular mapping.
- Infrastructure design, deployment and configuration management on GCP cloud.
- Automation of ETL processes to wrangle the data, posting recommendation data store by store.
- Automation of store data analysis by department and validation.
- Built Horizontally, vertically scalable distributed data solutions with Python multiprocessing, REST API, and GCP.
- Implemented data integration solutions with Sqoop and Kafka, ensuring smooth data transfer between various systems.
- Designed and optimized ETL processes using tools like Informatica, transforming raw data into meaningful insights.
- Worked with cloud-based data warehousing solutions like Amazon Redshift and DynamoDB to store and retrieve structured and unstructured data.
- Created interactive and responsive web applications using HTML, CSS, and JavaScript to visualize data and enhance user experience.
- Developed data-driven applications in Scala, integrating them with Gradle and Jenkins for continuous integration and deployment.
- Collaborated on agile development teams, employed Agile and Scrum methodologies, and managed tasks and workflows using Jira.

- Created interactive dashboards and reports using Tableau, ensuring stakeholders had real-time access to key performance indicators.
- Implemented centralized logging and monitoring systems with ELK stack (Elasticsearch, Logstash, Kibana) for comprehensive system analysis.
- Containerized applications and services using Docker, enabling seamless deployment across different environments.
- Orchestrated containerized applications using Kubernetes, ensuring efficient resource management and scalability.
- Version controlled and collaborated on code repositories using GitHub, facilitating team collaboration and code versioning.
- Developed and maintained scalable data pipelines using Python, Pandas, and NumPy for efficient data processing.
- Implemented end-to-end data pipelines, ensuring data accuracy, reliability, and timeliness in processing and analysis.
- Big data technologies such as Hadoop, Spark, Hive, and Pig were used to process and analyze vast datasets.
- Troubleshoot and optimize data processes, identify bottlenecks and implement solutions for improved performance.
- Worked on Relational Data base and Monitered Relational Data Base.

Environment: Python, Pandas, NumPy, Hadoop, AI ,Spark, Hive, Pig, Sqoop, Kafka, ETL, Informatica, Redshift, HTML, CSS, JavaScript, Oracle, DynamoDB, Scala, Scripting, Gradle,Data Modeling,Data Warehouse, Jenkins, GitHub, Data Pipelines, Agile, Scrum, Jira, Tableau, ELK, Docker, Kubernetes, Ruby.

Client: Trigyn Technologies, India
Role: Data Engineer
Duration:Oct 2019-Sep 2022
Roles & Responsibilities:

- Troubleshooting Spark applications for improved error tolerance.
- Fine-tuning spark applications/jobs to improve the efficiency and overall processing time for the pipelines.
- Development of PySpark application to extract data from multiple TPAs with Python & Shell scripting.
- Automating the backup jobs on a daily basis with AWS Cloud watch & Lambda services.
- Development of spark jobs to classify the data on industry type & category.
- Designed the ETL data pipeline to transform the data from multiple sources to Master tables.
- Development of Python APIs to dump the array structures in the Processor at the failure point for debugging.
- Build deployment on AWS ecosystem with EMR, S3, Cloud watch, Step functions and Lambda.
- Deployment of application for all Dev, Stage & Prod environments.
- Creating Databricks notebooks using Pyspark, Python, SQL and automated notebooks using jobs.
- Responsible for estimating the cluster size, monitoring and troubleshooting of the Spark Databricks cluster.
- Proficient in creating reliable, scalable, and high-performance applications, especially in the realm of web development and automation.
- Developed the scripts to create Hive table DDL and Analyze table from Pyspark jobs.
- Development of Tableau dashboard with the views for weekly data category wise.
- Used AWS EMR for all the data processing activities and Athena for all the data analysis.
- Implemented batch processing of jobs using Spark
- Worked with Apache NiFi to automate the data flow between the systems and managed flow of information between systems.

- Developed Sqoop scripts to import/export data from Oracle to S3 and into Hive tables.
- Involved building and managing NoSQL Database models using HBase.
- Worked in Spark to read the data from Hive and write it to HBase.
- Optimized the Hive tables using optimization techniques like partitions and bucketing to provide better performance with Hive QL queries.
- Working experience in Agile methodology
- Worked on Relational Data Base funtions like create alter delete and updating the records.

Environment: Pyspark, Python, Hive, Tableau, Scripting AWS Lambda, EMR, Glue, step function, Cloud watch, Data Modeling, Data Ware House ,Shell- scripting, Linux,AI, Jenkins, Bitbucket, Ruby.

Client: Spruce InfoTech, Inc, India
Role: Python Developer
Duration: April 2018-Sep 2019
Roles & Responsibilities:

- Developed Python Django forms to record data and the Login module page for users.
- Designed email marketing campaigns and created interactive forms that saved data into database using
  Django Framework.
- Developed the automated process for processing the payer acknowledgements that helped in resubmitting the medical claims automatically using PERL and DBI modules.
- Worked on Technologies: QT, JAVASCRIPT, QML, and unit test, C++, QNX, UML API, JavaScript and Json.
- Used collections in Python for manipulating and looping through different your defined objects and worked in Test driven development with Behave in Python.
- Also have rich experience on testing REST Full API like Facebook and Twitter.
- Wrote Python routines to log into the websites and fetch data for selected options.
- Improving the performance while processing data by modifying functions, queries, cursors, triggers and stored procedures for MySQL database.
- Extracted data from XML files using PERL -XML modules.
- Developed the consumer-friendly front-end with an easy-to-use API UI and fast access RESTFUL.
- Worked with JSON based REST Web services.
- Use of API, Open NLP & Stanford NLP for Natural Language Processing and sentiment analysis
- Embedded AJAX in UI to update small portions of the web page avoiding the need to reload the entire page.
- Used Amazon Web Services (AWS) for improved efficiency of storage and fast access.
- Added support for Amazon AWS S3 and RDS to host static/media files and the database into Amazon Cloud.
- Participated in the complete SDLC process and used PHP to develop website functionality.
- Designed and developed the UI of the website using HTML, XHTML, AJAX, unittest, GCP CSS, JAVASCRIPT and JavaScript.
- Developed entire frontend and backend modules using python on Django Web Framework.
- Used Django APIs for database access and built application logic using python.
- Used python to extract information from XML files.
- Expertise in Service Oriented Architecture (SOA) and its related technologies like Web Services, BPEL, WSDLs, AWS, SOAP, API, XML, XSD, XSLT etc.
- Worked on development of SQL and stored procedures on MYSQL and designed Cassandra schema for the APIs.

Environment: Python, Django, API, Behave, Puppet Rspec, Jenkins, unit test, Grafana Graphite, MySQL, Linux, HTML, CSS, Bootstrap, RDS J Query, Scripting,API, AWS, JavaScript, GCP, Apache, Linux, Git, C++, API, SOAP, Data Modeling, REST, Perl,AI, Cassandra, Azure Automation cloud services, Ruby.