Sr. Hadoop/Spark Developer

Sr. Hadoop/Spark Developer Sr. Hadoop/Spark Developer - Astra Zeneca Wilmington, DE

Authorized to work in the US for any employer

Work Experience

Sr. Hadoop/Spark Developer

Astra Zeneca - Wilmington, DE

March 2018 to Present

Roles & Responsibilities:

- ? Ingested gigabytes of click stream data from external servers such as FTP server and S3 buckets on daily basis using custom Input Adapters.
- ? Created Sqoop scripts to import/export user profile data from RDBMS to S3 data lake.
- ? Developed various spark applications using Scala to perform various enrichments of user behavioral data (click stream data) merged with user profile data.
- ? Involved in data cleansing, event enrichment, data aggregation, de-normalization and data preparation needed for downstream model learning and reporting.
- ? Utilized Spark Scala API to implement batch processing of jobs
- ? Trouble Shooting Spark applications for improved error tolerance.
- ? Fine-tuning spark applications/jobs to improve the efficiency and overall processing time for the pipelines
- ? Created Kafka producer API to send live-stream data into various Kafka topics.
- ? Developed Spark-Streaming applications to consume the data from Kafka topics and to insert the processed streams to HBase.
- ? Utilized Spark in Memory capabilities, to handle large datasets.
- ? Used Broadcast variables in Spark, effective & efficient Joins, transformations and other capabilities for data processing.
- ? Worked on EMR cluster and S3 in AWS cloud.
- ? Creating Hive tables, loading and analyzing data using hive scripts. Implemented Partitioning, Dynamic Partitions, Buckets in HIVE.
- ? Involved in continuous Integration of application using Jenkins.
- ? Interacted with the infrastructure, network, database, application and BA teams to ensure data quality and availability

Environment: AWS EMR, Spark, Hive, HDFS, Sqoop, Kafka, Oozie, HBase, Scala, MapReduce

Spark Developer

AT&T - Ledgewood, NJ

March 2016 to January 2018

Roles & Responsibilities:

- ? Extensively worked on migrating data from traditional RDBMS to HDFS.
- ? Ingested data into HDFS from Teradata, MySQL using Sqoop.
- ? Involved in developing spark application to perform ELT kind of operations on the data.
- ? Modified existing MapReduce jobs to Spark transformations and actions by utilizing Spark RDDs, Dataframe and Spark SQL API's
- ? Utilized Hive partitioning, Bucketing and performed various kinds of joins on Hive tables
- ? Involved in creating Hive external tables to perform ETL on data that is produced on

daily basis ? Validated the data being ingested into HIVE for further filtering and cleansing. ? Developed Sqoop jobs for performing incremental loads from RDBMS into HDFS and further applied Spark transformations ? Loaded data into hive tables from spark and used Parquet columnar format. ? Created Oozie workflows to automate and productionize the data pipelines ? Migrating Map Reduce code into Spark transformations using Spark and Scala. ? Collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis. ? Designed, documented operational problems by following standards and procedures using JIRA Environment: HDP (Hortonworks Data Platform), Spark, Scala, Sqoop, Oozie, Hive, Cent OS, MySQL, Oracle DB, Flume Hadoop Developer Life Lock - Tempe, AZ September 2015 to February 2016 Roles & Responsibilities: ? Involved in writing Spark applications using Scala to perform various data cleansing, validation, transformation and summarization activities according to the requirement. ? Load the data into Spark RDD and Perform in-memory data computation to generate the output as per the requirements. ? Developed data pipelines using Spark, Hive and Sqoop to ingest, transform and analyze operational data. ? Developed Spark jobs, Hive jobs to summarize and transform data. ? Worked on performance tuning of Spark application to improve performance. ? Performance tuning the Spark jobs by changing the configuration properties and using broadcast variables. ? Real time streaming the data using Spark with Kafka. Responsible for handling Streaming data from web server console logs. ? Worked on different file formats like Text, Sequence files, Avro, Parquet, JSON, XML files and Flat files using Map Reduce Programs. ? Developed daily process to do incremental import of data from DB2 and Teradata into Hive tables using Sqoop. ? Wrote Pig Scripts to generate Map Reduce jobs and performed ETL procedures on the data in HDFS. ? Analyzed the SQL scripts and designed the solution to implement using Scala. ? Solved performance issues in Hive and Pig scripts with understanding of Joins, Group and Aggregation and how does it translate to MR jobs. ? Work with cross functional consulting teams within the data science and analytics team to design, develop and execute solutions to derive business insights and solve clients operational and strategic problems. ? Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI

team. ? Extensively used Hive/HQL or Hive queries to query data in Hive Tables and loaded data into HBase tables. ? Extensively worked with Partitions, Dynamic Partitioning, bucketing tables in Hive, designed both Managed and External tables, also worked on optimization of Hive queries. ? Involved in collecting and aggregating large amounts of log data using Flume and staging data in HDFS for further analysis. ? Assisted analytics team by writing Pig and Hive scripts to perform further detailed analysis of the data. ? Designed Oozie workflows for job scheduling and batch processing. Environment: Java, Scala, Apache Spark, MySQL, CDH, IntelliJ IDEA, Hive, HDFS, YARN, Map Reduce, Sqoop, PIG, Flume, Unix Shell Scripting, Python, Apache Kafka Big Data (Hadoop Developer) Zyno Medical - Natick, MA November 2013 to August 2015 Roles & Responsibilities: ? Coordinated with business customers to gather business requirements and interacted with other technical peers to derive Technical requirements and delivered the BRD and TDD documents ? Involved in validating the aggregate table based on the rollup process documented in the data mapping. ? Designed and Modified Database tables and used HBASE Queries to insert and fetch data from tables. ? Involved in loading and transforming large sets of structured, semi structured and unstructured data from relational databases into HDFS using Sqoop imports. ? Responsible for analyzing and cleansing raw data by performing Hive queries and running Pig scripts on data. ? Developed Pig Latin scripts to extract the data from the web server output files to load into HDFS. ? Created Hive tables, loaded data and wrote Hive queries that run within the map. ? Used OOZIE Operational Services for batch processing and scheduling workflows dynamically and created UDF's to store specialized data structures in HBase and Cassandra. ? Involved in application development using RDBMS, and Linux shell scripting. ? Developed and updated social media analytics dashboards on regular basis. ? Created a complete processing engine, based on Hortonworks distribution. ? Manage and review Hadoop log files. ? Developed and generated insights based on brand conversations, which in turn helpful for effectively driving brand awareness, engagement and traffic to social media pages. ? Involved in the identifying, analyzing defects, questionable function error and inconsistencies in output. Environment: Hadoop, MapReduce, Yarn, Hive, Pig, HBase, Oozie, Sqoop, Oracle 11g, Scala, HDFS, Eclipse Java

Developer Mariner Finance - Laurel, MD October 2012 to October 2013 Roles & Responsibilities: ? Gathered requirements from end users and create functional requirements. ? Worked on process flow analyzing the functional requirements ? Development of Graphical user interface for user self-service screen ? Implemented four eyes principle and created quality check process -reusable across all workflow on overall platform level ? Development of UI models using HTML, JSP, JavaScript, Web Link and CSS. ? Developed Struts Action classes and Validation classes using Struts controller component and Struts validation framework. ? Support in end user training, testing and documentation. ? Implemented Backing beans for handling UI components and stores its state in a scope. ? Worked on implementing EJB Stateless sessions for communicating with Controller. ? Implemented database integration using Hibernate and utilized spring with Hibernate for mapping with Oracle database. ? Worked on Oracle PL/SQL queries to Select, Update and Delete data. ? Worked on MAVEN for build automation. Used GIT for version control Environment: Java, J2EE, JSP, Maven, Linux, CSS, GIT Oracle, XML, SAX, Rational Rose, UML Java Developer Acute Soft Solutions India Private Limited - Hyderabad, Telangana July 2011 to August 2012 Roles & Responsibilities: ? Involved in developing the application using Java/J2EE platform. Implemented the Model View Control (MVC) structure using Struts. ? Responsible to enhance the Portal UI using HTML, JavaScript, XML, JSP, Java, CSS as per the requirements and providing the client-side Java script validations and Server-side bean Validation Framework (JSR 303). ? Used Spring Core Annotations for Dependency Injection. ? Used Hibernate as persistence framework mapping the ORM objects to table using Hibernate annotations. ? Responsible to write the different service classes and utility API which will be used across the framework. ? Used Axis to implementing Web Services for integration of different systems. ? Developed Web services component using XML, WSDL and SOAP with DOM parser to transfer and transform data between applications. ? Exposed various capabilities as Web Services using SOAP/WSDL. ? Used SOAP UI for testing the Restful Web Services by sending and SOAP request. ? Used AJAX framework for server communication and seamless user experience. ? Created test framework on Selenium and executed Web testing in Chrome, IE and Mozilla through Web driver. ? Used client-side java scripting: JQUERY for

designing TABS and DIALOGBOX. ? Created UNIX shell scripts to automate the build process, to perform regular jobs like file transfers between different hosts. ? Used Log4j for the logging the output to the files. ? Used JUnit/ Eclipse for the unit testing of various modules. ? Involved in production support, monitoring server and error logs and foreseeing the Potential issues and escalating to the higher levels. Environment: Java, J2EE, JSP, Servlets, Spring, Servlets, Custom Tags, Java Beans, JMS, Hibernate, IBM MQ Series, AJAX, Junit, Log4j, JNDI, Oracle, XML, SAX, Rational Rose, UML Education Bachelor's Degree in Computer Science in Computer Science Acharya Nagarjuna University 2011 Skills Cassandra, Hdfs, Mapreduce, Oozie, Sqoop, Hbase, Db2, Mongodb, Nosql, Teradata, Hbase, Hive, Mapreduce, Pig, Python, Database, Sql server, Mysql, Oracle, Sql Additional Information ? Result-driven IT Professional with referable 8+ experience in Software development with 5+ years of recent expertise in BigData technologies including Hadoop and Spark. ? Professional Java developer with strong expertise in data engineering and big data technologies. ? Highly skilled on Spark, Hive, Pig, MapReduce, Sqoop, Kafka, Oozie, HBase, Impala and Yarn. ? Hands on experience in programming using Java, Python, Scala and SQL. ? Sound knowledge of architecture of Distributed Systems and parallel processing frameworks. ? Experience with Hadoop distributions both on premise (CDH, HDP) and in cloud (AWS). ? Good experience working with various data analytics and big data services in AWS Cloud like EMR, Redshift, S3, Athena, Glue etc. ? Expert in developing production ready spark application using Spark RDD API's, Data frames, Spark-SQL and Spark-Streaming API's. ? Strong experience in using Spark Streaming, Spark SQL and other components of spark like accumulators, Broadcast variables, different levels of caching and optimization techniques for spark jobs ? Proficient in importing/exporting data from RDBMS to HDFS using Sqoop. ? Solid experience in working various data formats like Parquet, Orc, Avro, JSon etc., ? Experience automating end-to-end data pipelines with strong resilience and recoverability. ? Strong knowledge of NoSQL databases and worked with HBase, Cassandra and Mongo DB. ? Expert in SQL, extensively worked RDBMSs like Oracle, SQL Server, DB2, MySQL and Teradata ? Proficient and Worked with GIT, Jenkins and Maven. ? Good understanding and Experience with Agile and Waterfall methodologies of Software Development

Life Cycle (SDLC). ? Highly motivated, self-learner with a positive attitude, willingness to learn new concepts and accepts challenges.   Technical competencies: Big Data Ecosystems: HDFS, MapReduce, YARN, Hive, Sqoop, Pig, Spark HBase, Oozie Programming Languages: Java, Scala, Python, SQL AWS technologies: S3, EMR, Redshift, Athena, Glue Database: SQL Server, MySQL, Oracle, DB2, Teradata NoSQL Databases: HBase, MongoDB, Cassandra IDE's & Utilities: Eclipse, IntelliJ Development Methodologies: Agile, Waterfall Model

Name: Mary Johnston

Email: gary91@example.org

Phone: 001-519-507-8373x9440