**Aslam**

**Mobile: +1 248-667-8655**

**Email:**[aslammd0111@gmail.com](mailto:aslammd0111@gmail.com)**Sr. Big Data/Scala DEVELOPER**

**PROFESSIONAL SUMMARY:**

* Having 8**+ years** of experience in IT which includes Analysis, Design, Development of **Big Data using Hadoop,** design and development of web applications using **JAVA, J2EE, Python** and **data base** and **data warehousing Technologies**.
* Around 5+ years of work experience on **Big Data** Analytics with hands on experience in installing, configuring and using ecosystem components like Hadoop Map reduce, **HDFS, HBase, Zookeeper, Hive, Sqoop, Pig, Flume**, **Cassandra, Kafka and Spark, NiFi.**
* Good Understanding of Hadoop architecture and Hands on experience with **Hadoop** components such as Job Tracker, Task Tracker, Name Node, Data Node and **Map Reduce** concepts and **HDFS Framework.**
* Experience in using Cloudera Manager for installation and management of single-node and multi-node Hadoop cluster (CDH4&CDH5).
* Experience in pulling data from **Amazon S3** cloud to **HDFS.**
* Hands on experience in **VPN Putty and WinSCP.**
* Good expertise in using **AWS** services like **EMR**, **EC2** and **S3** to run apache spark development and production jobs.
* Experience in Data load management, importing & exporting data **using SQOOP & FLUME.**
* Experience in analyzing data using **Hive, Pig and custom MR programs in Java**.
* Experience in scheduling and monitoring jobs **using Oozie and Zookeeper.**
* Experienced in writing **Map Reduce** programs & UDF's for both Pig &**Hive in java.**
* Experience in dealing with log files to extract data and to copy into **HDFS using flume.**
* Developed Hadoop test classes **using MR unit** for checking Input and Output.
* Experience in integrating **Hive and Hbase** for effective operations.
* Implemented Scala scripts, UDF's using both Data frames/SQL and RDD/MapReduce in Spark for Data Aggregation, queries and writing data into HDFS through Sqoop.
* Worked on continuous Integration tools Jenkins and automated jar files at end of day.
* **Experience in Impala, Solr, MongoDB, HBase and Spark.**
* Hands on knowledge of writing code in **Scala.**
* Developed pipeline for constant information ingestion utilizing Kafka, Spark streaming.
* Expertise in Waterfall and Agile - SCRUM methodologies.
* Proficient in **Core Java, J2EE, JDBC**, Servlets, JSP, Exception Handling, Multithreading, **EJB, XML, HTML5, CSS3, JavaScript, AngularJS.**
* Used source debuggers and visual development environments.
* Experience in Testing and documenting software for client applications.
* Writing code to create single-threaded, multi-threaded or user interface event driven applications, either stand-alone and those which access servers or services.
* Good experience in using Data Modelling techniques to find the results based on **SQL and PL/SQL** queries.
* Good working knowledge on **Spring Framework**.
* Experience working with different databases, such as **Oracle, SQL Server, MySQL** and writing stored procedures, functions, joins, and triggers for different Data Models.
* Expertise in implementing Service Oriented Architectures (**SOA) with XML based Web Services (SOAP/REST).**

**TECHNICAL SKILLS**:

|  |  |
| --- | --- |
| Big Data Technologies | HDFS, Hive, MapReduce, Pig, Sqoop, Flume, Oozie, Hadoop distribution, and Hbase,Spark. |
| Programming languages | Java (5, 6, 7),Python,Scala |
| Databases | MySQL, SQL/PL-SQL, MS-SQL Server 20012/16, Oracle 10g/11g/12c |
| Scripting/Web Languages | JavaScript, HTML5, CSS3, XML, SQL, Shell, Perl. |
| NoSql Databases | Cassandra, HBASE, mongoDB ,ELASTIC SEARCH |
| Operating Systems | Linux, Windows XP/7/8,Mac. |
| Software Life Cycle | SDLC, Waterfall and Agile models |
| Cloud Technologies | Amazon EC2, S3, EMR, Dynamo DB. |
| Utilities/Tools | Eclipse, Tomcat, NetBeans, JUnit, SQL, SVN, Log4j, SOAP UI, ANT, Maven, Automation and MR-Unit, Erwin, Alteryx, Visio, Docker, Jenkins. |
| Data Visualization Tolls | Tableau, QlikView. |

**EDUCATION:**

* Bachelor of Engineering in Information Technology, Jawaharlal Nehru Technological University, India 2006-10.

**PROFESSIONAL EXPERIENCE:**

**Ford, Dearborn, MI Nov 2017 – Present**

**Role: Sr. Hadoop Developer/ScalaEngineer.**

**Responsibilities:**

* Worked with **Spark** 2.x for converting unstructured data from various flat files and convert them to structured data along with respective file formats and place them in HDFS.
* Responsible for migrating the code base from Cloudera Platform to **Amazon EMR** and evaluated Amazon eco systems components like **RedShift**, Dynamo DB.
* Usage **of Spark Streaming** and **Spark SQL API** to process the files.
* Worked extensively with **Sqoop** for importing and exporting the data from **HDFS** to Relational Database systems/mainframe and vice-versa loading data into **HDFS.**
* Developed **Spark scripts** by using **Spark shell commands** as per the requirement.
* Stored data in **AWS S3** like HDFS and performed EMR programs on data stored in S3.
* Worked on **Big Data Hadoop** cluster implementation and data integration in developing large-scale system software.
* Experienced in performance tuning of **Spark** Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning
* Responsible in performing sort, join, aggregations, filter, and other transformations on the datasets using Spark to handle bad, null, partial records using **Spark SQL**.
* Developed Hadoop streaming Jobs using Spark**python** for integrating python API supported applications.
* Developed **Storm** topology to ingest data from various source into **Hadoop Data Lake.**
* Configured **ActiveMQ** for enterprise and resolved **ActiveMQ** issues
* Developed web application using **HBase and Hive API** to compare schema between **HBase and Hive tables**.
* Used **JVM** monitor to monitor threads and memory usage of **HBase and Hive** schema check web application.
* Developed **Python Script** to import data **SQLServer** into **HDFS**& created Hive views on data in **HDFS** using Spark.
* Created scripts to append data from temporary **HBase table to target HBase** table in **Spark.**
* Worked on **NOSQL Data**bases such as **HBase**, also used **SPARK** for real time streaming of data into the cluster.
* Developed complex and Multi-step data pipeline using **Spark**.
* Involved in converting **Hive/SQL** queries into **Spark** transformations using **Spark RDDS, Dataset, DataFrame and Scala**.
* Populated **HDFS** and Cassandra with huge amounts of data using **Apache Kafka.**
* Monitoring **YARN** applications. Troubleshoot and resolve cluster related system problems.
* Upgrading the **Hadoop Cluster from CDH3 to CDH4**, setting up High Availability Cluster and integrating **HIVE** with existing applications.
* Assessed existing and **EDW** (enterprise data warehouse) technologies and methods to ensure our **EDW/BI** architecture meet the needs of the business and enterprise and allows for business growth.
* Involved in loading the created **HFiles** into **Hbase** for faster access of large customer base without taking Performance hit.
* Using **Spark** - Cassandra connector to load data to and from Cassandra.
* Used various **spark** Transformations and Actions for cleansing the input data.
* Migrated Map reduce jobs to **Spark** Jobs to achieve better performance.
* Collecting and aggregating large amounts of log data using **Apache Flume** and staging data in **HDFS** for further analysis.
* Involved in creating ETL flow using **Pig**, loading with data and writing **Pig Latin** queries which will run internally in **Map Reduce way.**
* Involved in writing **Unix/Linux Shell** Scripting for scheduling jobs and for writing pig scripts and hive QL.
* Assisted in exporting data into **Cassandra** and writing column families to provide fast listing outputs.
* Used **Oozie** Scheduler systems to automate the pipeline workflow and orchestrate the map reduce jobs that extract
* Used **Zookeeper** for providing coordinating services to the **cluster.**
* Worked with **Hue UI** in scheduling jobs with ease and File browsing, Job browsing, Metastore management.
* Developed and designed system to collect data from multiple portal **using kafka and then process it using spark.**
* Handling large datasets using Partitions, **Spark in Memory capabilities**, **Broadcasts in Spark,** Effective & efficient Joins, Transformations and other during ingestion process itself.

**Environment**: HDFS, Hive, Core Java, Sqoop, Spark, Cloudera CDH4, Oracle, Elastic search, Kerberos, SFTP, Impala, Jira, Wiki, Alteryx, Teradata, Shell/Perl Scripting, Kafka, Python, YARN, Zookeeper, Docker, Jenkins.

**General Motors, Detroit, MI Dec 2016 – Oct 2017**

**Role: Sr. Hadoop/Scala Developer**

**Responsibilities:**

* Involved in converting **Hive/SQL** queries into Spark transformations **using Spark RDDs and Scala.**
* Developed multiple POCs using **Scala** and deployed on the Yarn cluster, compared the performance of **Spark, with Hive and SQL/Teradata.**
* **Analysed the SQL scripts and designed the solution to implement using Scala.**
* Developed analytical component using **Scala, Spark and Spark Stream**.
* Developing UDFs in java for hive and pig and worked on reading multiple data formats on **HDFS using Scala.**
* **Used Scala collection framework** to store and process the complex consumer information.
* Used **Scala functional programming concepts to develop business logic**.
* Designed and implemented **Apache Spark** Application (Cloudera)
* Importing and exporting data into **HDFS Sqoop and Flume and Kafka.**
* **Troubleshoot and debug Hadoop ecosystem run-time issues.**
* **Used Oozie Scheduler** systems to automate the pipeline workflow and orchestrate the map reduce jobs that extract
* Worked with **Scala/SparkSQL** for data cleansing and generating Data Frames to transform them into row Data Frame's to populate the aggregate tables in Cassandra.
* Wrote **python** scripts to process semi-structured data in formats like JSON.
* Implementations were done using the spark API's and SparkSQL written in **Python**.
* Worked with Hue GUI in scheduling jobs with ease and File browsing, Job browsing, Metastore management.
* Worked with **BI** team in the area of **Big Data Hadoop** cluster implementation and data integration in developing large-scale system software.
* Installed/Configured/Maintained Apache Hadoop clusters for application development and Hadoop tools like **Hive, Pig, HBase, Flume, Oozie Zookeeper and Sqoop.**
* Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and troubleshoot managing and reviewing data backups and **Hadoop log files.**
* Continuous monitoring and managing the **Hadoop cluster through Cloudera Manager.**
* Extensively involved in Installation and configuration of Cloudera distribution **Hadoop, NameNode, Secondary NameNode, JobTracker, TaskTrackers and DataNodes.**
* Load D-Stream data into **Spark RDD** and do in memory data Computation to generate Output response.
* Created **POC to store Server Log data in MongoDB** to identify System Alert Metrics.
* Monitored **Hadoop cluster** job performance, performed capacity planning and managed nodes on **Hadoop cluster**.
* Used **Zookeeper** operational services for coordinating cluster and scheduling workflows.
* Worked with application teams to install operating system, Hadoop updates, patches, version upgrades as required.
* Loaded data into the cluster from dynamically generated files using Flume and from relational database management systems using Sqoop.
* Performed analysis on the unused user navigation data by loading into HDFS and writing MapReduce jobs. The analysis provided inputs to the new APM front end developers and lucent team.
* Wrote MapReduce jobs using Java **API and Pig Latin.**
* Wrote **Pig scripts** to run **ETL** jobs on the data in HDFS and further do testing.
* Used **Hive** to do analysis on the data and identify different correlations.
* Involved in **HDFS** maintenance and administering it through **Hadoop-Java API.**
* Written Hive queries for data analysis to meet the business requirements.
* Automated all the jobs, for pulling data from **FTP** server to load data into Hive tables, using **Oozie workflows.**
* Involved in creating **Hive tables**& working on them using **HiveQ**L and perform data analysis using **Hive and Pig.**
* Used **Qlikview** and D3 for visualization of query required by BI team.
* Defined **UDFs** using **PIG and Hive** in order to capture customer behavior.
* Design and implement MapReduce jobs to support distributed processing using **java, Hive and Apache Pig.**
* Create **Hive** external tables on the MapReduce output before partitioning, bucketing is applied on it.
* Loaded the load ready files from mainframes to **Hadoop** and files were converted to ASCII format.
* Configured **Hive Server (HS2**) to enable analytical tools like **Tableau, Qlikview and SAS to interact with Hive tables.**

**Environment:** Hadoop, MapReduce, YARN, HDFS, Hive, Java, SQL, Cloudera Manager, Pig, Sqoop, ZooKeeper, Teradata, PL/SQL, MySQL, Hbase, Python, Spark, Jira, Jenkins, Docker.

**Broadcom, San jose, CA Sep 2015 – Nov 2016**

**Role: Bigdata Developer**

**Description**: This project involves setting up a data repository and uses it for search processing for analytical and research purposes. This application provides the capability for large batch processing using Hadoop map reduce jobs using Java runtime environment as well as real time search capabilities using Solr cloud environment

**Responsibilities:**

* **Written Map-Reduce code** to process all the log files with rules defined in HDFS(as log files generated by different devices have different xml rules).
* Developed and designed application to **process data using Spark**.
* Developed MapReduce jobs, **Hive & PIG scripts** for Data warehouse migration project.
* **Developed and designed system to collect data from multiple portal using kafka and then process it using spark**.
* Developing MapReduce jobs, **Hive & PIG scripts** for Risk & Fraud Analytics platform.
* Developed **Data ingestion platform** using **Sqoop and Flume** to ingest Twitter and Facebook data for Marketing & Offers platform.
* Developed and designed **automate process using shell scripting** for data movement and purging.
* Installation & Configuration Management of a small multi node Hadoop cluster.
* Installation and configuration of other open source software like **Pig, Hive, Flume, Sqoop.**
* **Developed programs in JAVA, Scala-Spark for data reformation after extraction from HDFS for analysis**.
* Written **Hive** jobs to parse the logs and structure them in tabular format to facilitate effective querying on the log data.
* Developed **Spark** scripts by using **Scala** shell commands as per the requirement.
* Importing and exporting data into **Impala, HDFS and Hive using Sqoop.**
* Responsible to manage data coming from different sources.
* Implemented Partitioning, Dynamic Partitions and Buckets in **HIVE** for efficient data access.
* Developed **Hive** tables to transform, analyze the data in HDFS.
* Involved in creating **Hive tables,** loading with data and writing hive queries which will run internally in map way.
* Develop code in **Python** utilizing Pandas (DataFrame) to read data from the excel for creating Measure Rule objects dynamically.
* **Developed Simple to Complex Map Reduce Jobs using Hive and Pig.**
* Involved in running Hadoop Jobs for processing millions of records of text data.
* Developed the application by using the **Struts framework.**
* Created connection through **JDBC and used JDBC** statements to call stored procedures.
* Developed **Pig Latin scripts** to extract the data from the web server output files to load into **HDFS.**
* Developed the **Pig UDF’S to pre-process the data for analysis**.
* Implemented multiple **Map Reduce Jobs in java for data cleansing and pre-processing.**
* Moved all RDBMS data into flat files generated from various channels to HDFS for further processing.
* Developed job workflows in **Oozie to automate the tasks of loading** the data into HDFS.
* Handled importing of data from various data sources, performed transformations using **Hive, MapReduce**, loaded data into HDFS and extracted data from Teradata into **HDFS using Sqoop.**
* Writing the script files for processing data and loading to HDFS.

**Environment**: Hadoop, MapReduce, HDFS, Pig, Hive, Java (jdk1.7), Oracle 11g/10g, PL/SQL, SQL\*PLUS, Windows NT, Sqoop, Spark, Jira, Docker.

**Wells Fargo, St. Louis, MO Feb 2015 to Aug 2015**

**Role: Hadoop Developer**

**Description:** Designed and developed big data solutions involving Terabytes of data. The big data solution consists of collecting large amounts of log data from distributed sources, transformations and standardizations analysis, statistics, aggregations and reporting etc. Built an ondemand elastic Hadoop cluster infrastructure to cater the needs of various Big Data projects, automated various Big Data workflows to process and extracts analytics out of the data using MapReduce, Pig and Hive.

**Responsibilities:**

* Processed data into **HDFS** by developing solutions.
* Analyzed the data using **Map Reduce, Pig, Hive** and produce summary results from **Hadoop** to downstream systems.
* **Used Pig as ETL tool** to do transformations, event joins and some pre-aggregations before storing the data onto **HDFS.**
* Developed data pipeline **using flume, Sqoop** and pig to extract the data from weblogs and store in **HDFS.**
* Used **Sqoop** to import and export data from **HDFS to RDBMS** and vice-versa.
* Created Hive tables and involved in data loading and writing **Hive UDFs.**
* Exported the analyzed data to the relational database **MySQL** using Sqoop for visualization and to generate reports.
* **Created HBase tables** to load large sets of structured data.
* Managed and reviewed **Hadoop log files.**
* Involved in providing inputs for estimate preparation for the new proposal.
* Worked extensively with **HIVE DDLs and Hive Query language (HQLs).**
* Developed **UDF, UDAF, UDTF** functions and implemented it in **HIVE Queries**. Implemented **SQOOP for large dataset** transfer between **Hadoop and RDBMs**.
* Involved in converting **Cassandra/Hive/SQL** queries into Spark transformations using Spark RDD's in Scala and Python.
* Developed Preprocessing job using **Spark** Data frames to flatten JSON documents to flat file.
* Created **Map Reduce Jobs** to convert the periodic of **XML** messages into a partition avro Data.
* **Used Sqoop** widely in order to import data from various systems/sources (like **MySQL) into HDFS.**
* Created components like **Hive UDFs** for **missing functionality in HIVE for analytics.**
* Developing Scripts and Batch Job to schedule a bundle (group of coordinators) which consists of various.
* Used different file formats like **Text files, Sequence Files, Avro.**
* Cluster co-ordination services through **Zookeeper.**
* Assisted in creating and maintaining Technical documentation to launching **HADOOP Clusters** and even for executing **Hive queries and Pig Scripts.**
* Assisted in **Cluster maintenance, cluster monitoring**, adding and removing cluster nodes and
* Installed and configured **Hadoop, Map Reduce, HDFS**, Developed multiple Map Reduce jobs in **java** for data cleaning and pre-processing.

**Environment:** Hadoop, HDFS, Map Reduce, Hive, Pig, Sqoop, HBase, Shell Scripting, Oozie, Oracle 11g, Spark, ServiceNow.

**Wipro Technologies, Hyderabad INDIA Jun 2012 to Dec 2014**

**Role: Java Developer**

**Responsibilities:**

* Excellent **JAVA, J2EE** application development skills with strong experience in Object Oriented Analysis, Extensively involved throughout Software Development Life Cycle (SDLC
* Implemented various **J2EE** standards and **MVC** framework involving the usage of Struts, **JSP, AJAX and servlets for UI design.**
* Used **SOAP/ REST** for the data exchange between the backend and user interface.
* **Utilized Java and MySQL** from day to day to debug and fix issues with client processes.
* Developed, tested, and implemented financial-services application to bring multiple clients into standard database format.
* Created web service components using **SOAP, XML and WSDL to receive XML** messages and for the application of business logic.
* Involved in configuring **web sphere variables**, queues, **DSs, servers** and deploying EAR into Servers.
* Involved in developing the business Logic using Plain Old Java Objects (POJOs) and **Session EJBs.**
* Developed authentication through **LDAP by JNDI.**
* Developed and debugged the application using **Eclipse IDE**.
* **Involved in Hibernate mappings**, configuration properties set up, creating sessions, transactions and second level cache set up.
* Involved in backing up database & in creating dump files. And also creating DB schemas from dump files. Wrote developer test cases & executed. Prepared corresponding scope & traceability matrix.
* Implemented **JUnit and JAD** for debugging and to develop test cases for all the modules.
* Hands-on experience of Sun One Application Server, Web logic Application Server, Web Sphere Application Server, **Web Sphere Portal Server, and J2EE** application deployment technology.

**Environment:** Java multithreading, JDBC, Hibernate, Struts, Collections, Maven, Subversion, JUnit, SQL language, Struts, JSP, SOAP, Servlets, Spring, Hibernate, Junit, Oracle, XML, Putty and Eclipse.

**JK Infotech, HYD, India Apr 2010 to Jun 2012**

**Role: Java Developer**

**Responsibilities:**

* Involved in analysis and design phase of Software Development Life cycle (SDLC).
* **Used JMS** to pass messages as payload to track statuses, milestones and states in the workflows.
* Involved in reading & generating pdf documents using **ITEXT**. And also merge the pdfs dynamically.
* Involved in the software development life cycle coding, testing, and implementation.
* Involved in Using **Java Message Service** (JMS) for loosely coupled, reliable and asynchronous exchange of patient treatment information among **J2EE** components and legacy system
* Developed MDBs using JMS to exchange messages between different applications using **MQ Series.**
* Involved in working with **J2EE** Design patterns (Singleton, Factory, DAO, and Business Delegate) and Model View Controller Architecture with **JSF** and **Spring DI**.
* Involved in Content Management using **XML**.
* Developed a standalone module transforming XML 837 module to database using **SAX** parser.
* Installed, Configured and administered **WebSphere ESB** v6.x
* Worked on Performance tuning of **WebSphere ESB** in different environments on different platforms.
* Configured and Implemented web services specifications in collaboration with offshore team.
* Involved in Creating dash board charts (business charts) using fusion charts.
* Involved in creating reports for the most of the business criteria.
* Involved in the configurations set for **Web logic servers**, **DSs, JMS** queues and the deployment.
* Involved in creating queues, **MDB**, Worker to accommodate the messaging to track the workflows
* Created **Hibernate** mapping files, sessions, transactions, Query and Criteria’s to fetch the data from DB.
* Enhanced the design of an application by utilizing **SOA.**
* Generating **Unit Test** cases with the help of internal tools.
* **Used JNDI for connection pooling**.
* Developed **ANT scripts** to build and deploy projects onto the application server.
* Involved in implementation of continuous build tool as Cruise control using **Ant**
* Used Star Team as version controller.

**Environment:** JAVA/J2EE, HTML, JS, AJAX, Servlets, JSP, XML, XSLT, XPATH, XQuery, WSDL, SOAP, REST, JAX-RS, JERSEY, JAX-WS, Web Logic server 10.3.3, JMS, ITEXT, Eclipse, JUNIT, Star Team, JNDI, Spring framework – DI, AOP, Batch, Hibernate.