**Shiva S**

**Hadoop/Spark Developer**

**Shivas.bigdata@gmail.com**

**+1(424)-244-9572**

**PROFESSIONAL SUMMARY:**

* Big Data developer with over **8 years** of professional IT experience, which includes 4 years experience in the field of Big Data.
* Extensive experience in working with various distributions of **Hadoop** like enterprise versions of **Cloudera, Hortonworks** and good knowledge on **MAPR distribution** and **Amazon’s EMR.**
* In depth experience in using various Hadoop Ecosystem tools like **HDFS, MapReduce, Yarn, Pig, Hive,Sqoop, Spark, Storm, Kafka, Oozie, Elastic search, HBase,** and **Zookeeper.**
* Extensive knowledge of Hadoop architecture and its components.
* Good knowledge in installing, configuring, monitoring and troubleshooting Hadoop cluster and its eco-system components.
* Exposure to Data Lake Implementation using **Apache Spark.**
* Developed Data pipe lines and applied business logics using **Spark.**
* Well-versed in spark components like **Spark SQL, MLib, Spark streaming and GraphX.**
* Extensively worked on **Spark streaming and Apache Kafka** to fetch live stream data.
* Experience in installation and setup various **Kafka Producers** and **Consumers** along with the Kafka **Brokers** and **Topics.**
* Having experience on **RDD**architecture and implementing spark operations on **RDD** and optimizing Transformations and Actions in Spark.
* Used **Scala and Python** to convert **Hive/SQL** queries into **RDD** transformationsin **Apache Spark.**
* Experience in integrating Hive queries into Spark environment using **Spark SQL.**
* Experience with **Pyspark**for using Spark libraries by using python scripting for data analysis.
* Expertise in performing real time analytics on big data using **HBase and Cassandra**.
* Handled importing data from **RDBMS** into **HDFS** using **Sqoop** and vice-versa.
* Extensive experience in importing and exporting streaming data into **HDFS** using stream processing platforms like **Flume** and **Kafka**.
* Experience in developing data pipeline using **Pig, Sqoop, and Flume** to extract the data from **weblogs** and store in **HDFS**.
* Created User Defined Functions (UDFs), User Defined Aggregated Functions (UDAFs) in **PIG and Hive.**
* Hands-on experience in tools like **Oozie** and **Airflow** to orchestrate jobs.
* Proficient in **NoSQL** databases including **HBase, Cassandra, MongoDB** and its integration with **Hadoop cluster**.
* Involved in developing web services using **REST**, **HBase** Native **API**Client to query data from **HBase**.
* Expertise in **Cluster management** and **configuringCassandra** Database.
* Implemented **CRUD**operations using **CQL**on top of **Cassandra** file system.
* Set Solr for distributing indexing and search.
* Used **Solr**to enable indexing for enabling searching on Non primary key columns from Cassandra key spaces.
* Great familiarity with creating **Hive tables, Hive joins & HQL** for querying the databases eventually leading to complex **Hive UDFs.**
* Accomplished developing **Pig Latin Scripts** and using **Hive Query Language** for data analytics.
* Worked on different compression codecs (ZIO, SNAPPY, GZIP) and file formats (ORC, AVRO, TEXTFILE, PARQUET)
* Experience in practical implementation of cloud-specific AWS technologies including **IAM,** Amazon Cloud Services like **Elastic Compute Cloud (EC2), ElastiCache, Simple Storage Services (S3), Cloud Formation, Virtual Private Cloud (VPC), Route 53, Lambda, EBS.**
* Built AWS secured solutions by creating VPC with public and private subnets.
* Worked on data warehousing and ETL tools like **Informatica**, **Talend,** and **Pentaho**.
* Expertise working in JAVA J2EE, JDBC, ODBC, JSP, Java Eclipse, Java Beans, EJB, Servlets.
* Developed web page interfaces using JSP, Java Swings, and HTML scripting languages.
* Experience working with **Spring** and **Hibernate** frameworks for **JAVA**.
* Worked on various programming languages using IDEs like **Eclipse, NetBeans,** and **Intellij.**
* Excelled in using version control tools like **PVCS, SVN, VSS** and **GIT.**
* Used web-based UI development using **JavaScript, jquery UI, CSS, jquery, HTML, HTML5, XHTML** and **JavaScript.**
* Development experience in DBMS like **Oracle, MS SQL Server, Teradata,** and **MYSQL**.
* Developed stored procedures and queries using **PL/SQL**.
* Experience with best practices of Web services development and Integration (both **REST and SOAP**).
* Experienced in using build tools like **Ant, Gradle, SBT, Maven** to build and deploy applications into the server.
* Knowledge in **Unified Modeling Language (UML)** and expertise in **Object Oriented Analysis and Design** (OOAD) and knowledge
* Experience in complete **Software Development Life Cycle (SDLC)** in both **Waterfall** and **Agile** methodologies
* Knowledge in Creating dashboards and data visualizations using **Tableau** to provide business insights

Excellent communication skills, interpersonal skills, problem-solving skills and very good team player along with a can do attitude and ability to effectively communicate with all levels of the organization such as technical, management and customers.

**Technical Skills:**

|  |  |
| --- | --- |
| Big Data Technologies | HDFS, Map Reduce, Yarn,HIVE, PIG, Pentaho, HBase,Oozie, ZooKeeper, Sqoop,Cassandra, Spark, Scala, Storm, Flume, Kafka and Avro,Parquet,Snappy |
| NO SQL Databases | HBase, Cassandra, MongoDB, Neo4j, Redis |
| Cloud Services | Amazon AWS, Google Cloud |
| Languages | C, C++, Java, Scala, Python, HTML, SQL, PL/SQL, Pig Latin, HiveQL, Unix, Java Script, Shell Scripting |
| ETL Tools | Informatica, IBM DataStage, Talend |
| Java & J2EE Technologies | Core Java, Servlets, Hibernate, Spring, Struts, JSP, JDBC, EJB |
| Application Servers | Web Logic, Web Sphere, JBoss,Tomcat. |
| Databases | Oracle,MySQL,DB2,Teradata,Microsoft SQL Server |
| Operating Systems | UNIX, Windows, iOS, LINUX |
| Build Tools | Jenkins, Maven, ANT |
| Business Intelligence Tools | Tableau, Splunk |
| Development Tools | Eclipse, IntelliJ, Microsoft SQL Studio, Toad, NetBeans |
| Methodologies | Agile, Waterfall |

**Work Experience :**

**Company: Allianz, Richmond, VA                                   Mar 2017 – Present**

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

**Responsibilities:**

* Involved in analyzing business requirements and prepared detailed specifications that follow project guidelines required for project development.
* Used **Sqoop** to import data from Relational Databases like MySQL, Oracle.
* Involved in importing structured and unstructured data into **HDFS**.
* Responsible for fetching real time data using **Kafka** and processing using **Spark and Scala**.
* Involved in loading data from rest endpoints to **Kafka Producers** and transferring the data to **Kafka Brokers.**
* Worked on **Kafka** toimport real time weblogs and ingested the data to **Spark Streaming**.
* Developed business logic using **Kafka Direct Stream in Spark Streaming** andimplemented business transformations.
* Worked on Building and implementing real-time streaming ETL pipeline using **Kafka Streams API.**
* Worked on **Hive** to implement Web Interfacing and stored the data in **Hive** tables**.**
* Migrated **Map Reduce** programs into **Spark** transformations using **Spark** and **Scala.**
* Configured **Spark streaming** to get ongoing information from the **Kafka** and store the stream information to HDFS.
* Experienced with **Spark Context, Spark-SQL, Spark YARN**.
* Implemented **Spark-SQL** with various data sources like **JSON, Parquet, ORC and Hive.**
* Implemented Spark Scripts using **Scala, Spark SQL** to access hive tables into spark for faster processing of data.
* Loaded the data into **Spark RDD** and do in memory data Computation to generate the Output response.
* Worked on loading AVRO/PARQUET/TXT files in **Spark Framework** using **Java/Scala** language and created **Spark Data frame and RDD to** process the data and save the file in **parquet** format in HDFS to load into fact table using ORC Reader.
* Good knowledge in setting up batch intervals, split intervals and window intervals in **Spark Streaming**.
* Implemented data quality checks using **Spark Streaming** and arranged passable and bad flags on the data.
* Implemented **Hive Partitioning** and **Bucketing** on the collected data in **HDFS**.
* Involved in Data Querying and Summarization using **Hive and Pig** and created **UDF’s, UDAF’s and UDTF’s**.
* Implemented **Sqoop** jobs for large data exchanges between **RDBMS and /Hive** clusters.
* Extensively used **Zookeeper** as a backup server and job scheduled for Spark Jobs.
* Knowledge on **MLLib** (Machine Learning Library) framework for auto suggestions.
* Developed **traits and case** classes etc in **Scala.**
* Developed Spark scripts using **Scala shell commands** as per the business requirement.
* Worked on **Cloudera** distribution and deployed on **AWS EC2** Instances.
* Experienced in loading the real-time data to NoSQL database like **Cassandra**.
* Experienced in using **Data StaxSpark Connector** which is used to store the data in **Cassandra** database from **Spark**.
* Involved in **NoSQL** (Datastax Cassandra) database design, integration, implementation, written scripts and invoked them using **CQLSH**.
* Well versed in using Data Manipulations, Compactions, tombstones in **Cassandra**.
* Experience in retrieving the data present in **Cassandra cluster** by running queries in **CQL** (**Cassandra Query Language**).
* Worked on connecting **Cassandra** database to the **Amazon EMR File System** for storing the database in **S3**.
* Implemented usage of **Amazon EMR** for processing Big Data across a Hadoop Cluster of virtual servers on **Amazon Elastic Compute Cloud (**[**EC2**](http://searchcloudcomputing.techtarget.com/definition/Amazon-Elastic-Compute-Cloud)**) and Amazon Simple Storage Service (S3).**
* Deployed the project on **Amazon EMR** with **S3** connectivity for setting a backup storage.
* Well versed in using of Elastic Load Balancer for Autoscaling **in EC2 servers**.
* Configured work flows that involves Hadoop actions using **Oozie.**
* Experienced with Faceted Reader search, Full Text SearchData querying using **Solr.**
* Used **Python** for pattern matching in **build logs** to format warnings and errors.
* Coordinated with **SCRUM** team in delivering agreed user stories on time for every **sprint**.

**Environment:** Hadoop YARN, Spark SQL, Spark-Streaming, AWS S3, AWS EMR, Spark-SQL, GraphX, Scala, Python, Kafka, Hive, Pig, Sqoop, Solr, Cassandra, Cloudera, Oracle 10g, Linux.

**Company: Internet Brands, EI Segundo, CA**                           **Jan 2016 – Feb 2017**

**Role: Sr.Spark/HadoopDeveloper**

**Responsibilities:**

* Analyzed large and critical datasets using Cloudera, HDFS, HBase, MapReduce, Hive, Hive UDF, Pig, Sqoop, Zookeeper and Spark.
* Responsible to manage data coming from different sources.
* Developed Batch Processing jobs using Pig and Hive.
* Involved in gathering the business requirements from the Business Partners and Subject Matter Experts.
* Worked with different File Formats like **TEXTFILE**, **AVROFILE**, **ORC**, and **PARQUET** for **HIVE** querying and processing.
* Importing and exporting data into **HDFS** and **Hive** using **Sqoop**.
* Implemented **Elastic Search** on **Hive** data **warehouse platform**.
* Good experience in analyzing Hadoop cluster and different analytic tools like **Pig**, **Impala**.
* Experienced in managing and reviewing **Hadoop** log files.
* Extracted files from **CouchDB** through **Sqoop** and placed in HDFS and processed.
* Experienced in running Hadoop streaming jobs to process terabytes of **xml** format data.
* Experienced in working with **spark** eco system using **SparkSQL** and **Scala** queries on different formats like Text file, CSV file.
* Created concurrent access for **Hive** tables with shared and exclusive locking that can be enabled in **Hive** with the help of **Zookeeper** implementation in the cluster.
* Storing and loading the data from **HDFS** to **Amazon S3** and backing up the Namespace data into **NFS**.
* Implemented Name Node backup using **NFS**. This was done for High availability.
* Designed workflows and coordinators in Oozie to automate and parallelize Hive jobs on Apache Hadoop environment by Hortonworks (HDP 2.2).
* Responsible for building scalable distributed data solutions using Hadoop cluster environment with Horton works distribution.
* Integrated Hive server 2 with Tableau using Horton Works Hive ODBC driver, for auto

generation of Hive queries for non-technical business user.ni

* Troubleshooting, Manage and review data backups, Manage and review Hadoop log files.
Hortonworks Cluster.
* Used **PIG** to perform data validation on the data ingested using **Sqoop** and **Flume** and the cleansed data set is pushed into **MongoDB**.
* Ingested streaming data with Apache **NiFi** into **Kafka**.
* Worked with **Nifi** for managing the flow of data from sources through automated data flow.
* Designed and implemented the **MongoDB** schema.
* Wrote services to store and retrieve user data from the **MongoDB** for the application on devices.
* Used **Mongoose API** to access the **MongoDB** from NodeJS.
* Created and Implemented Business validation and coverage Price Gap Rules in **Talend** on **Hive**, using **Talend** Tool.
* Wrote shell scripts for rolling day-to-day processes and it is automated.
* Written the shell scripts to monitor the data of Hadoop daemon services and respond accordingly to any warning or failure conditions.

**Environment:** Apache Flume, Hive, Pig, HDFS, Zookeeper, Sqoop, RDBMS, AWS, MongoDB, Talend, Shell Scripts, Eclipse, WinSCP, Hortonworks.

**Company: Delek, Brentwood, TN                      April2015- Oct 2015**

**Role: Hadoop Developer**

**Responsibilities:**

* Launching and Setup of **Hadoop** Cluster which includes configuring different components of **Hadoop**.
* Hands on experience in loading data from UNIX file system to **HDFS**.
* Wrote the Map Reduce jobs to parse the web logs which are stored in **HDFS**.
* Managing the Hadoop distribution with **ClouderaManager**, Cloudera Navigator, Hue.
* Developed Simple to complex MapReduce Jobs using **Hive** and **Pig**.
* Developed multiple Map Reduce jobs in **PIG** and **Hive** for data cleaning and pre-processing.
* Cluster coordination services through **Zookeeper**.
* Designed and implemented **Hive queries** and functions for evaluation, filtering, loading and storing of data.
* Installed **Oozie workflow** engine to run multiple Hive and Pig jobs which run independently with time and data availability.
* Expertise in Partitions, bucketing concepts in **Hive** and analyzed the data using the **HiveQL**
* Installed and configured **Flume**, **Hive**, **PIG**, **Sqoop** and **Oozie** on the Hadoop cluster.
* Involved in creating Hive tables, loading data and running Hive queries in those data.
* Extensive Working knowledge of partitioned table, UDFs, performance tuning, compression-related properties, thrift server in **Hive**.
* Used **Hive** to analyze data ingested into HBase by using Hive-HBase integration and compute various metrics for reporting on the dashboard.
* Loading the data to **HBase** Using Pig, **Hive** and Java API's.
* Worked with **NoSQL** databases like **HBase** in creating HBase tables to load large sets of semi-structured.
* Experienced with performing **CURD** operations in HBase.
* Collected and aggregated large amounts of web log data from different sources such as web servers, mobile using **Apache Flume** and stored the data into **HDFS/HBase** for analysis.
* Developed **Flume ETL** job for handling data from **HTTP Source** and **Sink** as HDFS.
* Involved in writing optimized **PIGScript** along with involved in developing and testing PIG Latin Scripts.
* Created Map Reduce programs for some refined queries on big data.
* Working knowledge in writing **PIG'sLoad** and Store functions.

**Environment:** Apache Hadoop 1.0.1, MapReduce, HDFS, CentOS, Zookeeper, Sqoop, Cassandra, Hive, PIG, Oozie, Java, Eclipse, Amazon EC2, JSP, Servlets.

**Company: RLI Corp, Peoria, IL Dec 2013- Feb 2015**

**Role: Java/Hadoop Developer**

**Responsibilities:**

* Developed the application using **StrutsFramework** that leverages classical **Model View Layer** (MVC) architecture **UML** diagrams like use cases, class diagrams, interaction diagrams, and activity diagrams were used.
* Participated in requirement gathering and converting the requirements into technical specifications.
* Extensively worked on User Interface for few modules using **JSPs, JavaScript and Ajax.**
* Created Business Logic using **Servlets, Session beans** and deployed them on **Web logic server**.
* Wrote complex **SQL queries** and **stored procedures**.
* Developed the **XML Schema** and **Amazon Web services** for the data maintenance and structures.
* Worked on analyzing, writing Hadoop MapReduce jobs using JavaAPI, Pig and hive.
* Selecting the appropriate AWS service based upon data, compute, system requirements.
* Implemented the **Web Service client** for the login authentication, credit reports and applicant information using **Apache Axis 2 Web Service**.
* Experience in creating integration between Hive and HBase for effective usage and performed MR Unit testing for the Map Reduce jobs.
* Got good experience with **NOSQL** database like MongoDB.
* Involved in loading data from **UNIX** file system to **HDFS**.
* Installed and configured **Hive**and also written **Hive UDFs.**
* Designed the logical and physical data model, generated **DDL scripts**, and wrote **DML scripts** for **Oracle 9i database**.
* Used Hibernate ORM framework with spring framework for data persistence and transaction management.
* Used struts validation framework for form level validation.
* Wrote test cases in JUnit for unit testing of classes.
* Involved in templates and screens in HTML and JavaScript.
* Involved in integrating Web Services using WSDL and UDDI.
* Suggested latest upgrades for Hadoop clusters.
* Created HBase tables to load large sets of data coming from UNIX and NoSQL
* Built and deployed Java applications into multiple Unix based environments and produced both unit and functional test results along with release notes

**Environment:** JDK 1.5, J2EE 1.4, Struts 1.3, JSP, Servlets 2.5, WebSphere 6.1, HTML, XML, ANT 1.6, JavaScript, Node.js, JUnit 3.8, HDFS, MongoDB, Hive, HBase UNIX, AWS

**Company: Metlife, India                                Aug 2011 - Oct 2013**

**Role: Java Developer**

**Responsibilities:**

* Developing rules based on different state policy using **SpringMVC, iBatis ORM, spring web flow, JSP, JSTL, Oracle, MSSQL, SOA, XML, XSD, JSON, AJAX, Log4j**
* Gathered requirements, developed, implemented, tested and deployed enterprise integration patterns (EIP) based applications using Apache **Camel, JBoss Fuse**
* Developed service classes, domain/DAOs, and controllers using **JAVA/J2EE** technologies
* Designed and developed using web service framework – **Apache CX**
* Worked on Active **MQ messaging** service for integration
* Worked with **SQL** queries to store and retrieve the data in MS SQL server
* Performed unit testing using **JUnit**
* Worked on continuous integration using **Jenkins/Hudson**
* Participated in all phases of development life cycle including analysis, design, development, testing, code reviews and documentations as needed.
* Involved in configuring Struts, Tiles and developing the configuration files
* Used **ECLIPSE** as **IDE, MAVEN** for build management, **JIRA** for issue tracking, **CONFLUENCE** for documentation purpose, **GIT** for version control, **ARC (Advanced Rest Client)** for endpoint testing, **CRUCIBLE** for code review and **SQLDeveloper** as DB client.

**Environment**: Spring Framework, Spring MVC, spring web flow, JSP, JSTL, SOAP UI, rating Engine, IBM Rational Team, Oracle 11g, XML, JSON, Ajax, HTML, CSS, IBM WebSphere Application Server, RAD with sub-eclipse, jenkins, maven, SOA, SonarQube, Log4j, Java, JUnit

**Company: Cadsys, India  May 2010–June2011**

**Role: Java Developer**

**Responsibilities:**

* Involved in gathering business requirements, analyzing the project and created UML diagrams such as **Use Cases, Class Diagrams, Sequence Diagrams** and **flowcharts** for the optimization Module using **Microsoft Visio**
* Configured **faces-config.xml** for the page navigation rules and created managed and backing beans for the Optimization module.
* Developing Enterprise Application using **SpringMVC, JSP, MySql**
* Working on developing client-side **Web Services** components using **Jax-Ws** technologies
* Extensively worked on **JUnit** for testing the application code of server-client data transferring
* Developed and enhanced products in design and in alignment with business objectives
* Used **SVN** as a repository for managing/deploying application code
* Involved in the system integration and user acceptance tests successfully
* Developed front end using **JSTL, JSP, HTML,** and **Java Script**
* Used **XML** to maintain the **Queries, JSP page mapping, Bean Mapping** etc.
* Used Oracle 10g as the backend database and written **PL/SQL** scripts.
* Maintained and modified system based on user feedbacks using the **OO** concepts
* Implemented database transactions using Spring **AOP & Java EE CDI** capability
* Enriched organization reputation via fulfilling requests and exploring opportunities
* Business Analysis, Reporting Service and Integrate to Sage Accpac (ERP)
* Developing new and maintaining existing functionality using **SPRING MVC, Hibernate**
* Developed test cases for integration testing using **JUnit**
* Creating new and maintaining existing web pages build in **JSP, Servlet**.

**Environment:** Java, SpringMVC, Hibernate, MSSQL, JSP, Servlet, JDBC, ODBC, JSF, Servlet, NetBeans, GlassFish, Spring, Oracle, MySQL, Sybase, Eclipse, Tomcat, WebLogic Server.