**Professional summary:**

* 7+ years of programming experience involved in all phases of Software Development Life Cycle (**SDLC**).
* Around 4 Years of **Big data** related architecture experience developing **Hadoop applications.**
* Hands on experience with the Hadoop stack **(MapReduce, Pig, Hive, Sqoop, HBase, Flume, Oozie).**
* Proven Expertise in performing analytics on Big Data using **Map Reduce, Hive** and **Pig**.
* Implemented POC to migrate map reduce jobs into **Spark RDD** transformations using **Scala.**
* Developed **Apache Spark** jobs using **Scala** in test environment for faster data processing and used **Spark SQL** for querying.
* Experienced with performing real time analytics on **NoSQL** data bases like **HBase** and **Cassandra**.
* Good knowledge in working with **Impala, Storm and Kafka.**
* Experienced with Dimensional modeling, Data migration, Data cleansing, Data profiling, and ETL Processes features for data warehouses.
* Worked with **Oozie work flow** engine to **schedule** time based jobs to perform multiple actions.
* Experienced in importing and exporting data from RDBMS into HDFS using **Sqoop.**
* Analyzed large amounts of data sets using **Pig scripts** and **Hive scripts.**
* Hands on experience in working with database like **Oracle, MySQL** and **PL/SQL**.
* Experience in high level ETL architecture for overall data transfer from the OLTP to OLAP
* Good working knowledge on processing **Batch** applications.
* Experienced in writing MapReduce programs and UDFs for both Hive and Pig in Java.
* Used **Flume** to channel data from different sources to **HDFS**.
* Experienced with **Maven**, **Jenkins** and **GIT.**
* Experience with configuration of Hadoop Ecosystem components: **Hive, HBase, Pig, Sqoop and Flume**.
* Good experience in **Hive** partitioning, bucketing and perform different types of joins on Hive tables and implementing Hive **SerDe** like **JSON** and **Avro**.
* Experience in using different file formats **– Avro, Sequence Files, ORC, JSON** and **Parquet.**
* Experience in Performance Tuning, Optimization and Customization.
* Experience with Unit Testing Map Reduce programs using **MRUnit, JUnit**.
* Experience in **Active Development** as well as onsite coordination activities in web-based, client/server and distributed architecture using Java, **J2EE** which includes **Web services, Spring, Struts, Hibernate and JSP/Servlets along with incorporating MVC architecture.**
* Good working knowledge on servers like **Tomcat**, **Web Logic 8.0.**
* Ability to work in teams as well as an individual, quick learner and able to meet deadlines.

**Technical Skills**

|  |  |
| --- | --- |
| **Hadoop 2.0** | Hadoop, CDH5.3.2, MapReduce, YARN, Spark 1.4, Sqoop, Hive, Oozie, PIG, HDFS, Flume, Impala, Storm 0.9.5, Apace Kafka 0.9.0.0 |
| **Programming Languages** | C, Java 6.0, Scala 2.11 MySQL, PL/SQL, PIG Latin, HiveQL and Python |
| **Web Technologies** | HTML, CSS, JavaScript, Ajax, XML |
| **Java &J2EE Technologies** | Core Java, Servlets, JSP, JDBC, EJB’s |
| **Databases** | Hive QL, HBase, MongoDB , Cassandra 2.2.0, |
| **IDE** | Eclipse |
| **Application Server** | Apache Tomcat 7.0, Apache Tomcat 6.0, WebSphere |
| **Process Automation Tools** | SVN, JUnit |
| **Design Patterns** | MVC, Singleton, Factory |
| **Frameworks** | Spring 3.0.5, Hibernate 3.5.1, Struts 1.3.10, EJB, JUnit, MRUnit |

**Professional Experience:**

**Norfolk Southern, Atlanta, GA Feb 2015 – Till Date**

**Hadoop Developer**

**Project Abstract:**

Norfolk Southern is one of the nation's premier transportation companies specializing in freight railroading. It operates approximately 21,000 route miles in 22 states and provides efficient connections to other rail carriers. As a hadoop developer, my responsibility was to maintain ETL pipelines to support data and analytical needs and solve problems at the core execution layer of the Spark engine. In addition, my liability was to work with the clients to understand their needs on emerging use cases like stream processing and build them into our product

**Responsibilities:**

* Assisted in the development of high quality analytical reports of weather and GIS data.
* Used **Talend** to generate optimized code to load, transform, enrich, and cleanse data inside Hadoop.
* Imported unstructured data like logs from different web servers to HDFS using **Flume** and developed **MapReduce** jobs for log analysis, recommendations and analytics.
* Involved in real-time data processing using **Storm**.
* Expertise in real-time analytics, machine learning and continuous monitoring of operations using **Storm**.
* Converted and loaded local data files into HDFS through the **Unix shell.**
* Developed Pig Latin scripts to extract the data from the web server output files to load into **HDFS.**
* Created **HBase** tables to store variable data formats coming from different portfolios.
* Implemented **HBase** Co-processors to notify Support team when inserting data into **HBase** Tables.
* Used **Kafka** along with **HBase** to render data streaming.
* Implemented **Spark** for fast interactive data analysis of datasets loaded in RDD.
* Analyzed the data by performing Hive queries (**HiveQL**), ran **Pig scripts, Spark SQL** and **Spark streaming.**
* Used **Hive** for data warehousing and summarization.
* Used **Pig** for aggregation, cleaning and incremental **ETL** functions and developing **UDFs** for filtering.
* Developed Pig UDFs to specifically preprocess and filter data sets for analysis.
* Continuous monitoring and managing the Hadoop cluster using Cloudera Manager.
* Wrote custom MapReduce jobs for cleaning and reorganizing data, accounting for and removing statistical anomalies.
* Used **Teradata** database management system to manage the warehousing operations and parallel processing.
* Validated data sets graphically with Excel and did touch ups in Photoshop.
* Designed & scheduled workflows for updating system reports using **Oozie.**

Client: **Arizona State University Phoenix, AZ** Oct’13– Jan 2015

Role: **Hadoop Developer**

**Responsibilities:**

* Experience with Hortonworks 2.2 distribution.
* Worked extensively on **HIVE**, **SQOOP**, **PIG** and **PYTHON**.
* Implemented Partitioning, Dynamic Partitions, Buckets in Hive.
* Used **AXWAY** to **FTP** the **OPTIM** files to move to Hadoop and create tables on top of the data.
* Scheduled **CRONJOB** to schedule the shell scripts.
* Worked with moving the **VSAM** files in mainframes to the Hadoop.
* The**INFORMATICA**is used to handle Occurs and Redefines of the **VSAM** files and loaded the data into Oracle.
* Developed **PIG** predefined functions to convert the fixed width file to delimited file.
* Migrated Complex map reduce programs into in memory spark processing using Transformations and actions.
* Used **Scala** to store streaming data to HDFS and to implement Spark for faster processing of data.
* Creating hive table on parquet schema using scala and spark.
* Used python to read the **AVRO** file
* Experience in developing**shell scripts** to perform the **incremental** loads.
* Used **HIVE** join queries to join multiple tables of a source system and load them into Elastic Search Tables.
* Developed **Pig Latin** scripts to extract data from the web server output files to load into **HDFS**
* Worked on moving data from **MULTIMEMBER’S** to Hadoop.
* Involved in data migration from one cluster to another cluster.
* Worked on **Talend** for Loading and extracting data from **Oracle** and **SQL**.
* The**HUE**is used for data validation and running hive queries.
* Experienced in managing Hadoop Jobs and logs of all the scripts.
* Used **Spark** for predictive analysis and tumor simulator.
* Experienced in creating hive tables by getting the source field names and data types.
* Experienced in Data Validations and gathering the requirements from the business.

**Wells Fargo Bank - San Francisco, CA July 2012-Sep 2013**

**Big Data/Hadoop Developer**

**Project Abstract:**

This project is to fulfill the needs within Whole sale Finance teams to understand the customers in order to fulfill customers’ expectations when profitability takes into the picture. Here as a hadoop developer, my responsibility was to analyze and relate all the customers from different data sources like Loan data, Insurance data, Auto dealership data, Municipal Product data, etc. In addition, my liability was to analyze customers based on the attributes that were calculated using transaction and profitability levels. The analysis was performed on 3 years of data and on a 300-node Hadoop cluster.

**Responsibilities:**

* Created Hive tables and working on them using **Hive QL**.
* Involved in installing Hadoop Ecosystem components.
* Validated Name node, Data node status in a **HDFS** cluster.
* Importing and exporting data from HDFS to RDBMS and vice-versa using **SQOOP.**
* Experienced in developing **HIVE** Queries on different data formats like Text file, CSV file.
* Developed multiple MapReduce jobs in java for data cleaning and preprocessing.
* Collecting and aggregating large amounts of log data using Apache **Flume** and staging data in **HDFS** for further analysis
* Installed and configured Hadoop cluster in Test and Production environments
* Performed both major and minor upgrades to the existing **CDH cluster**
* Code review as per the customer coding standards.
* Testing and providing the valid test data to users as per requirement.
* Weekly meetings with technical collaborators and active participation in code review sessions with senior and junior developers.
* Responsible to manage data coming from different sources.
* Supporting Hbase Architecture Design with the Hadoop Architect team to develop a Database Design in HDFS.
* Involved in **HDFS** maintenance and loading of structured and unstructured data.
* Wrote Hive queries for data analysis to meet the business requirements.
* Installed and configured Pig and also written **Pig Latin** scripts.
* Developed **UDFs** for Pig Data Analysis.
* Involved in managing and reviewing **Hadoop** log files.
* Developed Scripts and Batch Job to schedule various Hadoop Program.
* Utilized Agile Scrum **Methodology** to help manage and organize a team of 4 developers with regular code review sessions.
* Handled importing of data from various data sources, performed transformations using **Hive**, **MapReduce**.
* Analyzed the data by performing **Hive** queries and running **Pig** scripts to know user behavior.
* Continuous monitoring and managing the Hadoop cluster through Cloudera Manager.
* Installed **Oozie** workflow engine to run multiple Hive and Pig jobs.
* Developed Hive queries to process the data and generate the data cubes for visualizing.

**Environment:** Java, Hadoop, MapReduce, HDFS, Hive, Pig, Sqoop, Linux, XML, Eclipse, Cloudera

**Chubb Group of Insurance Companies, Warren, NJ Mar 2011-June 2012**

**Java Developer**

**Project Abstract:**

The Agent Portal project involves the design and development of a Web based application to provide online insurance services. The web application provides functionalities such as policy management, profile management, and claim reporting. As a Java developer, my responsibility was to develop enterprise integration solutions utilizing WebSphere Application Server/Java. In addition, designing Web-based applications, Web services and MQ/JMS services solutions based on business need and maintaining appropriate systems documentation was also my liability.

**Responsibilities:**

* Involved in study of User Requirement Specification, Communicated with Business Analysts to resolve ambiguity in Requirements document.
* Worked in Agile Scrum Methodology.
* Deployed and Developed **Java**/**J2EE** based application
* Developed application using **Spring MVC** and **Hibernate** as the ORM tool.
* Developed user interface with **HTML**, **CSS**, **JavaScript**, **JQuery**, **AJAX** and **JSP**
* Connected with database using **MYSQL** and **PL/SQL**
* Created **SOAP** based webservice to provide the service of Location Look up for the agents.
* Created a Subscriber-Publisher Messaging Topics using **JMS** to send and receive messages.
* Created **Daemon threads** for certain long running reporting process.
* Scheduled a certain tasks using **Java Timer** Class.
* Interfaced with **Oracle** back-end using **Hibernate** Framework.
* Implemented Test Driven Development using **JUnit**.
* Deployed applications on **WebSphere** during development
* Wrote shell scripts to schedule standalone Java programs.
* The message transport mechanism supported various data formats like **XML** and **JSON**.

**Environment-** Java, Spring, Hibernate, JSP, HTML, CSS, XML, JavaScript, JQuery, JUnit, AJAX, Multi-Threading, Oracle, Web Service – SOAP, WebSphere, MYSQL**.**

Client: **CIBIL Hyderabad, India May’09 -Feb' 11**

Role**: Java Developer**

**Responsibilities:**

* Involved in design and development phases of **Software Development Life Cycle (SDLC)**
* Involved in designing **UML** Use case diagrams, Class diagrams, and Sequence diagrams using Rational Rose.
* Followed agile methodology and **SCRUM** meetings to track, optimize and tailored features to customer needs.
* Developed user interface using JSP, JSP Tag libraries, and Java Script to simplify the complexities of the application.
* Implemented **Model View Controller (MVC)** architecture using Jakarta Struts frameworks at presentation tier.
* Developed a Dojo based front end including forms and controls and programmed event handling.
* Implemented SOA architecture with web services using **JAX-RS (REST)** and **JAX-WS (SOAP)**
* Developed various Enterprise Java Bean components to fulfill the business functionality.
* Created Action Classes which route submittals to appropriate EJB components and render retrieved information.
* Validated all forms using Struts validation framework and implemented Tiles framework in the presentation layer.
* Used Core java and object oriented concepts.
* Extensively used Hibernate in data access layer to access and update information in the database.
* Used Spring Framework for Dependency injection and integrated it with the **Struts Framework and Hibernate.**
* Used JDBC to connect to backend databases, **Oracle and SQL Server 2005**.
* Proficient in writing SQL queries, stored procedures for multiple databases, Oracle and SQL Server 2005.
* Used multithreading in programming to improve overall performance.
* Identified proper Exception Handling and logging Mechanism.
* Developed various helper classes needed following Core java multi-threaded programming and Collection classes.
* Wrote Stored Procedures using **PL/SQL**. Performed query optimization to achieve faster indexing and making the system more scalable.
* Deployed application on windows using IBM Web Sphere Application Server.
* Used Java Messaging Services (JMS) for reliable and asynchronous exchange of important information such as payment status report.
* Used Web Services - **WSDL** and **REST** for getting credit card information from third party and used **SAX** and **DOMXML** parsers for data retrieval.
* Implemented SOA architecture with web services using Web Services like **JAX-WS**.
* Used ANT scripts to build the application and deployed on Web Sphere Application Server.
* Developed JUnit test cases to validate the services