Harish Chava

Hadoop Developer

**Experience Summary:**

Over 8+ years of professional experience in IT and Enterprise Application Development in multiple industries, which includes hands on experience in Big Data/Hadoop ecosystem and related technologies.

4+ years of comprehensive experience in Big Data Technologies.

* Excellent understanding / knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node,MapReduce,Pig,Hive,HBase,Sqoop, Oozie,and flume concepts.
* Wellversedininstallation,configuration,supportingandmanagingofBigDataandunderlyinginfrastructure of **Hadoop Cluster**along with **CDH3&4** clusters.
* Experienced in working with Hadoop/Bigdata storage and analytical frameworks over Amazon **AWS Cloud** using tools like SSH, Putty and **MindTerm**.
* Experience with leveraging Hadoop ecosystem components including Pig and Hive for data analysis, Sqoop for data migration, Oozie for scheduling, and HBase for NoSQL database.
* Experience in installing, configuring, and administrating Hadoop cluster for various distributionslikeCloudera, Horton works and Map R.
* Very good understanding on **NoSQL** databases like **MongoDB**, **HBase** and **Cassandra**.
* Experience in managing and reviewing Hadoop log files.
* Experience in using Pig, Hive,Sqoop, Impala and **ClouderaManager**.
* Experience in importing and exporting data using **Sqoop** from HDFS to Relational Database Systems and vice-versa.
* Hands on experience in RDBMS and LINUX shell scripting
* Extending **Hive** and **Pig** core functionality by writing custom**UDFs**.
* Experience in analyzing data using **HQL**, Pig Latin and Map Reduce.
* Wrote UDF to convert the time into proper format and used **SerDe** to process CSV file and text file formats.
* Developed MapReduce jobs to automate transfer of data from **HBase**.
* Knowledge in job work-flow scheduling and monitoring tools like **Oozie** and Zookeeper.
* Knowledge of data warehousing and ETL tools like **Informatica** and **Pentaho**.
* Experience with Eclipse/ RSA.
* Knowledge on coding skills **PYTHON**
* Expertise in RDBMS like MS SQL Server, MySQL, Greenplum and DB2.
* Extensive experience in using MVC architecture,Struts,hibernate for developing web
* Applications using Java, JSPs, JavaScript, HTML, jQuery, AJAX, XML and JSON.
* Excellent Java development skills using J2EE, Spring, J2SE, Servlets, JUnit, MRUnit, JSP, JDBC.
* Techno-functional responsibilities include interfacing with users, identifying functional and technical gaps, estimates, designing custom solutions, development, leading developers, producing documentation, and production support.
* Excellent interpersonal and communication skills, creative, research-minded, technically competent and result-oriented with problem solving and leadership skills.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Bigdata Ecosystem | Hadoop, MapReduce, HDFS, Hive, Pig, HBase, HCatalog, Sqoop, Oozie,Flume,Spark,Kafka,Storm |
| Java/J2EETechnologies | J2SE, J2EE, Servlets, JSP, JDBC, AJAX, SOAP, WSDL |
| SDLC Methodologies | Agile, UML, Waterfall |
| Enterprise Frameworks | Ajax, MVC, Struts 2/1, Hibernate 3, Spring 3/2.5/2 |
| Version and Source Control | Tortoise SVN,GitHub, Gerrit, Nexus |
| Programming Languages | C, C++,Java, Python, Shell Scripting, Scala, SQL and PLSQL |
| WebTechnologies | HTML, DHTML, XML, XSLT, JavaScript, CSS |
| NoSQL Databases | MongoDB, HBase, Cassandra |
| IDE Tools | Eclipse, NetBeans, WinSCP |
| WebServices | WebLogic, Web Sphere, JBoss |
| Databases | Oracle DB2, MS - SQL Server, MySQL, MS Access, Mongo DB |
| Build Tools | Maven, ANT, Jenkins |
| Operating Systems | Windows, UNIX, Linux |

**Project Description:**

**Client: T-Mobile, Bellevue, WA Aug 2015 to Feb 2017**

**Role: Hadoop Developer**

**Responsibilities:**

* Analyze the assigned user stories in Rally (Agile software) and create design documents.
* Attend Daily stand ups and update the hours burned down in Rally.
* Design and develop data ingestion framework using Hadoop stacks.
* Expertise in analyzing the **logs** and diagnosis the issues.
* Actively involved in setting up coding standards, prepared low and high level documentation.
* Involved in preparing the S2TM document as per the business requirement and worked with Source system SME’s in understanding the source data behavior.
* Created Hive, **Phoenix**, HBase tables and **HBase** integrated Hive tables as per the design using **ORC file** format and **Snappy** compression.
* Indexing data into **Solr** for searching to implement **SCD1**.
* Imported required tables from RDBMS to HDFS using **Sqoop** and also used **Storm**/ Spark streaming and **Kafka** to get real time **streaming** data into HBase.
* Wrote different **pigscripts** to clean up the ingested data and created partitions for the daily data.
* Prepared **Data Profiling** and **Data Validation** rules as per the requirement specified in S2TM’s.
* Implemented Partitioning and bucketing in Hive based on the requirement.
* Wrote different UDF’s to convert the date format and to create hash value using **MD5 Algorithm** in Java and used various **UDF's** from **Piggybanks** and other sources.
* Prepared pig scripts and **Spark SQL** to handle all the transformations specified in the S2TM’s and to handle **SCD2** and **SCD1** scenarios.
* Involved in converting Hive/SQL queries into **Spark transformations** using **Spark SQL**, **Python** and **Scala**.
* Experienced in implementing **Spark RDD** transformations, actions to implement business analysis and Worked with **Spark** accumulators and broadcast variables.
* Created shell scripts to parameterize the Pig, Hive actions in Oozie workflow.
* Created tables in **Teradata** to export the data from HDFS using Sqoop after all the transformations and wrote **BTEQ** scripts to handle updates and inserts of the records.
* Created Teradata views on top of the table as per the business requirement.
* Worked closely with the **App Support** team in production deployment and setting up these jobs in **TIDAL/Control-M** for incremental data processing.
* Worked with **EQM and UAT** teams for fixing the defects immediately by understanding the issue.
* Involved in Unit level and Integration level testing and prepared supporting documents for proper deployment.

**Environment**: Hadoop Eco System (HDP 2.2.4, Hive, Pig, Sqoop, MapReduce, Kafka, Strom, HBase, Oozie, Ambari, HDFS, HUE),Teradata, Shell Script, Java, Oracle, MS SQL Server, GitHub, Gerrit, Jenkins, Nexus, Agile Scrum.

**Client: Thomson Reuters, Saint Paul, MN Mar 2014 to July 2015**

**Role: Hadoop Developer**

**Responsibilities:**

* Responsible for Build/develop/testing shared components that will be used across many modules.
* Imported all the Customer specific personal data to Hadoop using Sqoop component of Hadoop.
* Created tasks for **incrementalload** into staging tables, and schedule them to run.
* Used Apache **Kafka** as messaging system to load **log data**, data from UI applications into HDFS system.
* Implemented data Ingestion and handling clusters in real time processing using Kafka.
* Implemented POC’s using Kafka, Storm and Spark.
* Imported **Avro** files using **Kafka** and did some analytics using Spark.
* Written Map Reduce code that will take input as customer related flat file and parse the same data to extract the meaningful (domain specific) information for further processing.
* Created Hive External tables with partitioning to store the processed data from Map Reduce.
* Implemented Different **analytical algorithms** using map reduce programs to apply on top of HDFS data.
* Implemented Hive **optimized joins** to gather data from different sources and run ad-hoc queries on top of them.
* Wrote Hive Generic UDF's to perform business logic operations at record level.
* Implemented **Data classification algorithms** using map reduce design patterns.
* Extensively worked on creating **combiners**, **Partitioning, Distributed cache** to improve the performance of Map Reduce jobs.
* Experienced in migrating map reduce jobs into RDD(Resilient data distributions) and create **Sparkjobs** for better performance.
* Developed workflow in **Oozie** to automate the tasks of loading the data into HDFS and pre-processing with **Pig**.
* Involved in loading data from LINUX file system to HDFS.
* Used Pig to do data transformations, event joins, filter and some **pre-aggregations** before storing the data onto HDFS.
* Experienced in implementing test cases and writing test classes using MRUnit and mocking frameworks.
* Implemented test scripts to support test driven development and continuous integration.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.

**Environment**: CDH 4.x, Hadoop, MapReduce, Cloudera Manager, HDFS, Hive, Pig, Oozie, Sqoop, ApacheKafka, Spark,Java, Python, Linux, Shell, Chef, Junit, MRUnit

**Client: Amplify, Austin, TX Aug 2013 to Feb 2014**

**Role: Hadoop Developer**

**Responsibilities:**

* Experience in loading and transforming large amounts of structured, semi structured, and unstructured data.
* Experience importing raw data and exporting the analyzed data to the relational databases using **Sqoop** for data to be available for **data visualization** and generate reports for the BI team.
* Developed MapReduce programs to Validate Incoming Data and report invalid data.
* Automated testing of Map Reduce programs using **MRUnit**.
* Experience running Hadoop **Streaming** jobs for MapReduce programs to process data in **XML** format.
* Developed Pig Latin scripts to perform **Data Cleansing** and **business transformations**.
* Experience working with Hive Data Definition Language (DDL) and Hive Query Language (HQL); performed data transformations using Hive.
* Experience writing Hive and Pig UDFs to enable analytics for the business use case at hand.
* Worked on **Apache Solr** search engine server to do indexing and optimize the search.
* Experience in using **Avro, RC file** and **JSON** file formats.
* Performed data processing and pipelining using Apache Crunch
* Experience with **Cassandra** in optimizing it for writes and pre-computing aggregations to persist various statistics.
* Performed data modeling to connect data stored in Cassandra DB to the data processing layers and wrote queries in CQL.
* Designed and developed Oozie workflow for scheduling and orchestrating the automation of ETL process.

**Environment:** HortonWorks HDP1.0, HDP2.0, Ambari, SQL, Java, Shell, HDFS, Sqoop, Flume, MapReduce, Pig, Hive, HCatalog, Zookeeper, Oracle, Cassandra, Oozie.

**Client: Lease Plan USA - Alpharetta, GA Jan2013 to July 2013**

**Role: Hadoop Developer**

**Responsibilities:**

* Worked on importing and exporting data between **DB2** and **HDFS** using **Sqoop**.
* Used **Flume** to collect, aggregate, and store the web log data from different sources like web servers, mobile and network devices and pushed to HDFS.
* Developed MapReduce programs in Java to convert data from **JSON** format to **CSV** and TSV formats and perform analytics.
* Developed Pig Latin scripts for cleansing and analysis of semi structured data.
* Experience debugging MapReduce jobs and **PIG** scripts.
* Used Pig as **ETL tool** to do transformations, event joins and **pre**-**aggregations** before storing the data onto HDFS.
* Experience in creating Hive tables, loading with data and writing hive queries.
* Experience in migration of ETL processes from Relational databases to Hive to test the easy data manipulation.
* Experience using Hive to analyze the **partitioned** and **bucketed** data and compute various metrics for reporting.
* Experience writing Hive and Pig UDFs to perform aggregation to support the business use case.
* Performed MapReduce integration to import large amounts of data into HBase.
* Experience with performing CRUD operations using HBase Java Client API.
* Developed Shell scripts to automate MapReduce jobs and Pig scripts to process data.

**Environment**: CDH3, Cloudera Manager, Java, Shell, SQL, Hadoop, HDFS, Sqoop, Flume, MapReduce (MRV1), Pig, Hive, Oracle, MongoDB, TDD, JUnit.

**Client: InfoTech Enterprises (CYIENT), Hyderabad Aug 2011 to Nov 2012**

**Role: Java/J2EE Developer**

**Responsibilities:**

* Involved in various phases of Software Development Life Cycle (SDLC) as design development and unit testing.
* Leading a team of 4 developers and point of contact for onshore / offshore communication.
* Developed and deployed UI layer logics of sites using JSP, XML, JavaScript, HTML/DHTML, and Ajax.
* Used CSS and JavaScript to build rich internet pages.
* Designed different design specifications for application development that includes front-end, back-end using design patterns.
* Developed proto-type test screens in HTML and JavaScript.
* Involved in developing JSP for client data presentation and, data validation on the client side with in the forms.
* Developed the application by using the Spring MVC framework.
* Collection framework used to transfer objects between the different layers of the application.
* Developed data mapping to create a communication bridge between various application interfaces using XML, and XSL.
* Spring IOC being used to inject the parameter values for the Dynamic parameters.
* Developed JUnit testing framework for Unit level testing.
* Actively involved in code review and bug fixing for improving the performance.
* Documented application for its functionality and its enhanced features.
* Created connection through JDBC and used JDBC statements to call stored procedures.

**Environment:**SpringMVC,Oracle10g,J2EE,Java,JDBC,Servlets,JSP,XML,DesignPatterns,CSS,HTML, JavaScript 1.2, JUnit, Apache Tomcat, My SQL Server 2008.

**Client: Xcel Energy, India Oct 2009 to June 2011**

**Role: Java Developer**

**Responsibilities:**

* Involved in software development life cycle (SDLC) of the project (analysis, design, development, testing and implementation).
* Involved in various Software Development Life Cycle (SDLC) phases of the project.
* Experience in Application Development in Core Java using Multi-Threading and Collections extensively.
* Experience developing in Core Spring Framework for Dependency Injections of components.
* Developed Controller classes for handling requests and processing form submissions using Spring MVC framework.
* Experience developing in Hibernate for data access layer to access and update information in database.
* Experience developing in JSP, Servlets and custom JSP tags for Presentation Layer.
* Used Apache Tomcat web server to run Servlets and JSP.
* Responsible for migrating code between various environments (Development, Testing and Production).
* Solved the long transactions jobs issues with server configurations in web servers.
* Used Eclipse IDE for development, JIRA for defect tracking and CVS for version control.

**Environment:**Core Java, J2EE, JDBC, Spring, Hibernate, Servlets, JSP, Apache Tomcat, Eclipse IDE, JIRA, CVS, HTML, CSS, XML, JavaScript, REST, Oracle 10g, Windows

**Sonata Software, Hyderabad, India July 2008 to Aug 2009**

**Role: Java Developer**

**(Enterprise Transaction Scanning in AML (Anti Money Laundering))**

**Responsibilities:**

* Provide L3 application support as primary on call.
* Involved in the development of Report Generation module which includes volume statistics report, Sanctions Monitoring Metrics report, and TPS report.
* Implemented Online List Management (OLM) and FMM module using spring and Hibernate.
* Wrote various SQL, PL/SQL queries and stored procedures for data retrieval.
* Created Configuration files for the application using Spring framework.
* Used HTML, CSS, XML and JavaScript to design a page.
* Successfully migrated legacy application written in VB6.0 to VB.Net.
* Developed Web Services to get data from the external system in terms of .txt file to load into the database.
* Developed DTS/SSIS packages to load employee details into row Money tables of the SQL server for further processing.

**Environment:** Core Java, J2EE, JSP, Servlets, JQuery, JavaScript, CSS, HTML, SQL.

**EDUCATION:**

Bachelor of Technology, ECE, Vignan University.