Sindhu Yaramadha

Hadoop Developer

# SUMMARY

* 7+ years of overall experience in IT Industry which includes experience in **BigData**

**Ecosystem** related technologies.

* 3 years of comprehensive experience in **Big Data Technology Stack.**
* Good knowledge of **Hadoop Architecture** and various components such as **HDFS, Job Tracker, Task Tracker, Name Node, Data Node, and MapReduce concepts**.
* Wellversedininstallation,configuration,supportingandmanagingofBigDataandunderlyinginfrastructure of Hadoop Clusteralong with CDH3&4 clusters.
* Experience in installing, configuring, and administrating Hadoop cluster for various distributions like **Apache**, **Cloudera**, **Map R** and **Horton works**.
* Very good understanding on **NOSQL** databases like **HBase, Cassandra and MongoDB**,.
* Designed and implemented **a Cassandra NoSQL** based database and associated RESTful web service that persists high-volume user profile data for vertical teams.
* Experience in building large scale highly available Web Applications .Working knowledge of web services and other integration patterns.
* Experience in managing and reviewing **Hadoop** log files.
* Experience in using **Pig, Hive, Scoop** and **Cloudera Manager**.
* Experience in importing and exporting data using**Sqoop** from **HDFS** to Relational Database Systems and vice-versa.
* Generated Java APIs for retrieval and analysis on No-SQL database such as **HBase** and **Cassandra**.
* Analyzed the data on Cassandra cluster by running queries for searching, sorting and grouping.
* Hands-on experience with message brokers such as Apache Kafkaand RabbitMQ.
* Hands on experience in RDBMS, and Linux shell scripting.
* Extending Hive and Pig core functionality by writing **customUDF**s.
* Experience in analyzing data using **HiveQL**, **Pig Latin** and **Map Reduce**.
* 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, Kafka and Spark.**
* Expertise in **RDBMS** like **MSSQLServer**, **MySQL, Greenplum** and **DB2**.
* Knowledge of job workflow scheduling and monitoring tools like Oozie and Zookeeper, of NoSQL databases such as HBase, **Cassandra,** and of administrative tasks such as installing Hadoop, Commissioning and decommissioning, and its ecosystem components such as Flume, Oozie, Hive and Pig.
* Knowledge in Software Development Life Cycle (Requirements Analysis, Design, Development, Testing, Deployment and Support).
* 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:**

* **Programming Languages** **:** C, C++, Java, Shell Scripting, PL/SQL
* **J2EE Technologies**  : Spring, Servlets, JSP, JDBC, Hibernate.
* **BigData Ecosystem :** HDFS, HBase,MapReduce, Hive, Pig, Python,

Sqoop, Impala, Kafka, Cassandra, Oozie, Zookeeper,

Flume.

* **DBMS** **:** Oracle 11g, SQL Server, MySQL.
* **Modeling Tools :** UML on Rational Rose 4.0.
* **Web Technologies** **:** HTML, JavaScript, XML, jQuery, Ajax, CSS.
* **Web Services** : Web Logic, Web Sphere, Apache Cassandra ,Tomcat,
* **IDEs** **:**Eclipse, Netbeans, WinSCP.
* **Operating systems** **:** Windows,Unix, Linux (Ubuntu), Solaris, Centos.
* **Version and Source Control :** CVS, SVN.
* **Servers :** Apache Tomcat.
* **Frameworks**  **:** MVC, Struts, Log4J, Junit, Maven,ANT, WebServices.

|  |
| --- |
|  |

**Projects:**

**Client: Amerigroup– Virginia Beach, VA Dec 2013 to Current**

**Role: Hadoop Developer**

**Description:**

Amerigroup is a US health insurance and managed health care provider company headquartered in Virginia Beach, Virginia.

**Responsibilities:**

* Design, deploy, Manage cluster nodes for our data platform operations (racking/stacking)
* Install and configure cluster. Setting up puppet for centralized configuration management.
* Monitoring Cluster using various tools to see how the nodes are performing.
* Expertise in cluster task like Adding Nodes, Removing Nodes without any effect to running jobs and data.
* Write scripts to automate application deployments and configurations. Monitoring YARN applications. Troubleshoot and resolve cluster related system problems.
* Wrote map reduce programs to clean and pre-process the data coming from different sources.
* Implemented various output formats like Sequence file and parquet format in Map reduce programs. Also, implemented multiple output formats in the same program to match the use cases.
* Developed Hadoop streaming Map/Reduce works using Python.
* Performed benchmarking of the No-SQL databases, Cassandra and HBase.
* Created data model for structuring and storing the data efficiently. Implemented partitioning and bucketing of tables in Cassandra.
* Implemented test scripts to support test driven development and continuous integration.
* Converted text files into Avro then to parquet format for the file to be used with other Hadoop eco system tools.
* Experienced on loading and transforming of large sets of structured, semi structured and unstructured data.
* Exported the analyzed data to HBase using Sqoop and to generate reports for the BI team.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.
* Participate in requirement gathering and analysis phase of the project in documenting the business requirements by conducting workshops/meetings with various business users.
* POC work is going on using Spark and Kafka for real time processing.
* Developed a data pipeline using Kafka and Storm to store data into HDFS.
* Populated HDFS and Cassandra with huge amounts of data using Apache **Kafka**.
* POC work is going on comparing the Cassandra and HBaseNoSQL databases.

**Environment:**MapReduce, HDFS, Hive, Pig, Hue, Oozie, Core Java, Eclipse, Hbase, Flume, Spark, Kafka, Cloudera Manager, Cassandra,Python,Greenplum DB, IDMS, VSAM, SQL\*PLUS, Toad, Putty, Windows NT, UNIX Shell Scripting, Pentaho,Talend,Bigdata, YARN

|  |
| --- |
|  |

**Client: First Data Corp – Omaha, NEDec 2011 to Nov 2013**

**Role: Hadoop Developer**

**Description:**

First Data Corporation is a global [payment](http://en.wikipedia.org/wiki/Payment) processing company. First Data is a provider of electronic commerce and payment solutions. The company's portfolio includes merchant transaction processing services; [credit](http://en.wikipedia.org/wiki/Credit_Card), [debit](http://en.wikipedia.org/wiki/Debit_Card), [private-label](http://en.wikipedia.org/wiki/Stored-value_card), [gift](http://en.wikipedia.org/wiki/Gift_card), payroll and other [prepaid](http://en.wikipedia.org/wiki/Prepayment_for_service) card offerings; fraud protection and authentication solutions, credit reporting agency services (through First Data Merchant Services Corporation); [electronic check](http://en.wikipedia.org/wiki/Electronic_check) acceptance services through Tele Check; as well as Internet commerce and [mobile payment](http://en.wikipedia.org/wiki/Mobile_payment) solutions. The company's [STAR interbank network](http://en.wikipedia.org/wiki/STAR_(interbank_network)) offers PIN-secured debit acceptance at [ATM](http://en.wikipedia.org/wiki/Automated_teller_machine) and retail locations.

**Responsibilities:**

•DevelopedtheapplicationusingStrutsFrameworkthatleveragesclassicalModelViewController(MVC) architecture.

• Extensively worked on User Interface for few modules using JSPs, JavaScript and Ajax

• Created Business Logic using Servlets, POJO's and deployed them on Web logic server

•Involved in running Hadoop jobs for processing millions of records of text data.

• Responsible for Cluster maintenance, adding and removing cluster nodes, Cluster Monitoring and Troubleshooting, manage and review data backups and log files.

•WorkedwithsystemsengineeringteamtoplananddeploynewHadoopenvironmentsandexpandexisting Hadoop Clusters.

• Monitored multiple Hadoop clusters environments using Ganglia.

• Managing and scheduling Jobs on a Hadoop cluster.

• Involved in defining job flows, managing and reviewing log files.

• Monitored workload, job performance and capacity planning using Cloud era Manager.

• Installed Oozie workflow engine to run multiple Map Reduce, Hive and Pig jobs.

• Implemented Map Reduce programs on log data to transform into structured way to find user information.

• Responsible for loading and transforming large sets of structured, semi structured and unstructured data.

• Collected the log data from web servers and integrated into HDFS using Flume.

• Responsible to manage data coming from different sources.

• Extracted files from Couch DB and placed into HDFS using Sqoop and pre-process the data for analysis.

• Gained experience with NoSQL database.

* Unit tested and tuned SQLs and ETL Code for better performance.
* Performed major role in understanding the business requirements and designing and loading the data into data warehouse (ETL).

•CreatedandmaintainedTechnicaldocumentationforlaunchingHADOOPClustersandforexecutingHive queries and Pig Scripts.

**Environment:**JDK1.5,J2EE1.4,Struts1.3,JSP, Spring,Servlets2.5,WebSphere6.1,HTML,XML,JavaScript,Informatica Power Center 9.1 Hadoop, HDFS, Pig, Hive, MapReduce, HBase, Sqoop, Python,Oozie, Ganglia and Flume.

|  |
| --- |
|  |

**Client: Lease Plan USA - Alpharetta, GAJuly 2010 to Nov 2011**

**Role: Java/J2EE developer**

(Development focus of Java/J2EE based applications.)

**Description:**

Vehicle Acquisition is a Web based application for Lease Plan USA to manage their Fleet, Contacts, Dealers andOrders.UserswithRolesSupervisorsorManagerscanlogintotheapplicationtoviewtheListofOrders, Vehicles,Dealers,Contactsetc.andtheirdetails.Basedontheirroletheywillalsohavetheauthoritytomodify the Dealer and Order Details.

**Responsibilities:**

•DesignedanddevelopedStrutslikeMVC2Webframeworkusingthefront-controllerdesignpattern,which is used successfully in a number of production systems.

•Spearheadedthe"QuickWins"projectbyworkingverycloselywiththebusinessandenduserstoimprove the current website's ranking from being 23rd to 6th in just 3 months.

• Normalized Oracle database, conforming to design concepts and best practices.

•Resolvedproductcomplicationsatcustomersitesandfunneledtheinsightstothedevelopmentand deployment teams to adopt long term product development strategy with minimal roadblocks.

• Convinced business users and analysts with alternative solutions that are more robust and simpler to implement from technical perspective while satisfying the functional requirements from the business perspective.

• Applied design patterns and OO design concepts to improve the existing Java/JEE based code base.

•Identifiedandfixedtransactionalissuesduetoincorrectexceptionhandlingandconcurrencyissuesdueto unsynchronized block of code.

**Environment:**Java1.2/1.3,Swing,Applet,Servlet,JSP,customtags,JNDI,JDBC,XML,XSL,DTD,HTML, CSS, Java Script, Oracle, DB2, PL/SQL, Web logic , JUnit, Log4J and CVS.

|  |
| --- |
|  |

**Client: Tate & Lyle Ingredients America Inc. -Warren, NJ SEPT 2008 – JUN 2010Role: Java/J2EE Interface Developer**

**Description:**

Michigan Millers offers a wide range of products and services from personal auto, homeowners, and umbrella coverage to specialized protection for your business.

**Responsibilities:**

* Created Use case, Sequence diagrams, functional specifications and User Interface diagrams using Star UML.
* Involved in complete requirement analysis, design, coding and testing phases of the project.
* Participated in JAD meetings to gather the requirements and understand the End Users System.
* Developed user interfaces using JSP, HTML, XML and JavaScript.
* Generated XML Schemas and used XML Beans to parse XML files.
* Created Stored Procedures & Functions. Used JDBC to process database calls for DB2/AS400 and SQL Server databases.
* Developed the code which will create XML files and Flat files with the data retrieved from Databases and XML files.
* Created Data sources and Helper classes which will be utilized by all the interfaces to access the data and manipulate the data.
* Developed web application called iHUB (integration hub) to initiate all the interface processes using Struts Framework, JSP and HTML.
* Heavily used bug trackers Pivotal, jira, Bugzilla and Target Process
* Developed the interfaces using Eclipse 3.1.1 and JBoss 4.1 Involved in integrated testing, Bug fixing and in Production Support.

**Environment:**Java 1.3, Servlets, JSPs, Java Mail API, Java Script, HTML, MySQL 2.1, Java Web Server 2.0, JBoss 2.0, RMI, Rational Rose, Red Hat Linux 7.1.