**PRAVEEN ARASADA**

**Email Id:** [**Prawinarasada540@gmail.com**](mailto:Prawinarasada540@gmail.com)

**Phone No: 551-404-9508**

**Professional Summary:**

* Around 9 years of IT experience in various domains with Hadoop Ecosystems and Java J2EE technologies.
* Solid Mathematics, Probability and Statistics foundation and broad practical statistical and data mining techniques cultivated through various industry work and academic programs
* Involved in the Software Development Life Cycle (SDLC)phases which include Analysis, Design, Implementation, Testing and Maintenance.
* Strong technical, administration, and mentoring knowledge in Linux and Big Data/Hadoop technologies.
* Hands on experience on major components in Hadoop Ecosystem like Hadoop Map Reduce, HDFS, HIVE, PIG, Pentaho, Hbase, Zookeeper, Sqoop, Oozie, Cassandra, Flume and Avro.
* Solid understanding of RDD operations in Apache Spark i.e., Transformations &Actions, Persistence (Caching),Accumulators, Broadcast Variables, Optimising Broadcasts.
* In depth understanding of Apache spark job execution Components like DAG,lineage graph,Dag Scheduler, Taskscheduler, Stages and task.
* Experience in exposing Apache Spark as web services.
* Good understanding of Driver, Executor Spark web UI.
* Experience in submitting Apache Spark job and map reduce jobs to YARN.
* Experience in real time processing using Apache Spark and Kafka.
* Migrated Python Machine learning modules to scalable, high performance and fault-tolerant distributed systems like Apache Spark.
* Strong experience in Spark SQL UDFs,Hive UDFs, Spark SQL Performance, Performance Tuning.Hands on experience in working with input file formats like orc, parquet, json, avro.
* Good expertise in coding in Python,Scala and Java.
* Good understanding of the map reduce framework architectures (MRV1 & YARN Architecture).
* Good Knowledge and understanding of Hadoop Architecture and various components in Hadoop ecosystems - HDFS, Map Reduce, Pig, Sqoop and Hive.
* Handled importing of data from various data sources, performed transformations using Map Reduce, Spark and loaded data into HDFS.
* Experience in setting up of Hadoop cluster in cloud services like AWS and Azure.
* Manage and review Hadoop log files.
* Troubleshooting production support issues post-deployment and come up with solutions as required.
* Worked on data analysis and giving reports on daily basis.
* Check the registered logs in database whether the file status is properly updated or not.
* Handling the backup for input files in HDFS
* Worked with Avro Data Serialization system.
* Experience in writing shell scripts do dump the shared data from landing zones to HDFS.
* Experience in performance tuning the Hadoop cluster by gathering and analyzing the existing infrastructure.
* Expertise in Client Side designing and validations using HTML and Java Script.
* Excellent communication and inter-personal skills detail oriented, analytical, time bound, responsible team player and ability to coordinate in a team environment and possesses high degree of self-motivation and a quick learner.

**Education:**

Bachelors in Computer Science Engineering, VITS College of Engineering, Visakhapatnam, India.

**CERTIFICATIONS:**

* **Cloudera Certified Hadoop and Spark Developer(CCA-175**)

**Technical Skills:**

|  |  |
| --- | --- |
| **Big Data Frameworks** | Hadoop, Spark, Scala, Hive, Kafka, AWS, Cassandra, HBase, Flume, Pig, Sqoop, Map Reduce, Cloudera, Mongo DB. |
| **Big data distribution** | Cloudera, Amazon EMR |
| **Programming languages** | Core Java, Scala, Python, SQL, Shell Scripting |
| **Operating Systems** | Windows, Linux (Ubuntu) |
| **Databases** | Oracle, SQL Server |
| **Designing Tools** | Eclipse |
| **Java Technologies** | JSP, Servlets, Junit, Spring, Hibernate |
| **Web Technologies** | XML, HTML, JavaScript, JVM, JQuery, JSON |
| **Linux Experience** | System Administration Tools, Puppet, Apache |
| **Web Services** | Web Service (RESTful and SOAP) |
| **Frame Works** | Jakarta Struts 1.x, Spring 2.x |
| **Development methodologies** | Agile, Waterfall |
| **Logging Tools** | Log4j |
| **Application / Web Servers** | Cherrypy,Apache Tomcat, WebSphere |
| **Messaging Services** | ActiveMQ, Kafka, JMS |
| **Version Tools** | Git, SVN and CVS |
| **Analytics** | Tableau, SPSS, SAS EM and SAS JMP |

**Professional Experience:**

**Deutsche Bank, New York City, NY.**

**Jan 2016 – Present**

**Hadoop Developer**

**Description:** Deutsche Bank in New York City providing the financial services including credit card, debit card, mortgage and student loan. This project focus on implementing the real time risk detection system by using the advanced big data technologies.

**Responsibilities:**

* Installed and configured HDFS, Hadoop Map Reduce, developed various Map Reduce jobs in Java for data cleaning and preprocessing.
* Analyzed various RDDS using Scala, Python with Spark.
* Performed complex mathematical, statistical and machine learning analysis using SparkMlib, Spark Streaming and GraphX.
* Performed data ingestion from various data sources.
* Worked with various types of databases like SQl, NOSQl and Relational for transferring data to and from HDFS.
* Used Impala for data processing on top of Hive.
* Understanding of data storage and retrieval techniques, ETL, and databases, to include graph stores, relational databases, tuple stores, NOSQL, Hadoop, PIG, MySQL and Oracle databases
* Experience in using Avro, Parquet, RCFile and JSON file formats and developed UDFs using Hive and Pig.
* Optimized MapReduce jobs to use HDFS efficiently by using Gzip, LZO, Snappy compression techniques.
* Imported and exported data into HDFS and Hive using Sqoop.
* Experience in loading and transforming huge sets of structured, semi structured and unstructured data.
* Worked on different file formats like XML files, Sequence files, JSON, CSV and Map files using Map Reduce Programs.
* This layer is the middle tier between SQL Azure and SharePoint online external content.
* Continuously monitored and managed Hadoop cluster using Cloudera Manager.
* Performed POC’s using latest technologies like spark, Kafka, scala.
* Created Hive tables, loaded them with data and wrote hive queries.
* Involved in collecting, aggregating and moving data from servers to HDFS using Apache Flume.
* Experience in managing and reviewing Hadoop log files.
* Executed test scripts to support test driven development and continuous integration.
* Created Pig Latin scripts to sort, group, join and filter the enterprise wise data.
* Worked on tuning the Pig queries performance.
* Installed Oozie workflow to run several MapReduce jobs.
* Extensive Working knowledge of partitioned table, UDFs, performance tuning, compression-related properties, thrift server in Hive.

**Environment:** Cloudera 5.x Hadoop, Linux, IBM DB2, HDFS, Yarn, Impala, Pig, Azure, Hive, Python, Sqoop, Spark, Scala, Hbase, MapReduce, Hadoop Datalake, Informatica BDM 10

**Verizon, Boston, MA**

**Nov 2014 – Dec 2015**

**Hadoop Developer**

**Description:**

This project is to store terabytes of log information generated by the ecommerce website and extract meaningful information out of it. The solution is based on the Hortonworks Hadoop distribution .The data is stored in Hadoop file system and processed using Map/Reduce jobs. Which intern includes getting the raw html data from the websites, Process the html to obtain product and pricing information, Extract various reports out of the product pricing information and Export the information for further processing. This project is mainly for the re plat forming of the current existing system which is running on Web Harvest a third party JAR and in Oracle DB to a new solution technology called Hadoop which can able to process large date sets (i.e. Tera bytes and Peta bytes of data) in order to meet the client requirements with the increasing completion from his retailers.

**Responsibilities:**

* Support all business areas of ADAC with critical data analysis that helps team members make profitable decisions as a forecast expert and business analyst and utilize tools for business optimization and analytics.
* Experience and talents to be a part of ground breaking thinking and visionary goals. As an Executive Analytics, we take the lead to Delivery analyses/ ad-hoc reports including data extraction and summarization using big data tool set.
* Experience with Unix and/or Linux.
* Conduct Trainings on Hadoop MapReduce, Pig and Hive. Demonstrates up-to-date expertise in Hadoop and applies this to the development, execution, and improvement.
* Ensures technology roadmaps are incorporated into data and database designs.
* Experience in extracting large data sets is a HUGE plus.
* Experience in data management and analysis technologies like Hadoop, HDFS.
* Create list and summary view reports.
* Create Cache Memory on Windows Azure to improve the performance of data transfer between SQL Azure and WCF services.
* Created Airflow Scheduling scripts in Python.
* Involved in collecting, aggregating and moving data from servers to HDFS using Apache Flume.
* Experience in managing and reviewing Hadoop log files.
* Handling and communicating with business and understanding the problems from business perspective rather than as a developer perspective.
* Used Hive Query language (HQL) to further analyze the data to identify issues and behavioral patterns.
* Used Apache Spark and Scala language to find patients with similar symptoms in the past and medications used for them to achieve best results.
* Configured Apache Sentry server to provide authentication to various users and provide authorization for accessing services that they were configured to use.
* Used Pig as ETL tool to do transformations, joins and pre-aggregations before loading data onto HDFS.
* Worked on large sets of structured, semi structured and unstructured data
* Preparing the Unit Test Plan and System Test Plan documents.
* Preparation & Execution of unit test cases and Troubleshooting and debugging.

**Environment :**Cloudera Hadoop, Linux, HDFS, Maprduce, Hive, Pig, Azure, Sqoop, Python, Oracle, SQL Server, Eclise, Java and Oozie scheduler.

**OptumInsight, Inc.Eden Prairie, MN**

**Dec 2012 – Oct 2014**

**Hadoop Developer-Health Care**

**Description:**

OptumInsight provides software and information products, advisory consulting services and business process outsourcing to participants in the health care industry.

**Responsibilities:**

* Implemented Big Data platforms using Cloudera CDH4 as data storage, retrieval and processing systems.
* Involved in installing, configuring and managing Hadoop Ecosystem components like Hive, Pig, Sqoop and Flume.
* Responsible for loading data from Teradata database into a Hadoop Hive data warehousing layer, and performing data transformations using Hive.
* Partitioned the collected people's data by disease type and medication prescribed to the patient to improve the query performance.
* Used Hive Query language (HQL) to further analyze the data to identify issues and behavioral patterns.
* Used Apache Spark and Scala language to find patients with similar symptoms in the past and medications used for them to achieve best results.
* Configured Apache Sentry server to provide authentication to various users and provide authorization for accessing services that they were configured to use.
* Worked on Kafka while dealing with raw data, by transforming into new Kafka topics for further consumption.
* Installed and configured Hortonworks Sandbox as part of POC involving Kafka-Storm-HDFS data flow.
* Log data from webservers across the environments is pushed into associated Kafka topic partitions, SparkSQL is used to calculate the most prevalent diseases in each city from this data.
* Implemented Zookeeper for job synchronization
* Deployed NoSQL database HBase to store the outputs of the jobs.
* Supported operations team in Hadoop cluster maintenance activities including commissioning and decommissioning nodes and upgrades.
* Extended the core functionality of Hive language by writing UDF's and UDAF'S
* Worked on Oozie workflow engine for job scheduling.

**Environment:**Apache Hadoop, SQL, Sqoop, Hive, HBase, Pig, Oozie,Flume, Linux, Java 7, Eclipse 3.0, Tomcat 4.1, MySQL.

**Intermountain Healthcare (IHC), Salt lake City, UT**

**May 2011 – Nov 2012**

**Hadoop Developer**

**Description:**

Intermountain is the most diversified health care company in the United States and a leader worldwide in helping people live healthier lives and helping to make the health system work better for everyone. The particular project involves in assigning a personal physician for each and every customer respective to particular area. Categorizing the diseases encountered frequently based on the claims and on the geographical location. The particular project also analyses the effectiveness of campaigns conducted at various places.

**Responsibilities:**

* Developed Big Data Solutions that enabled the business and technology teams to make data-driven decisions on the best ways to acquire customers and provide them business solutions.
* Installed and configured Apache Hadoop, Hive, and HBase.
* Worked on Hortonworks cluster, which is responsible for providing open source platform based on Apache Hadoop for analysing, storing and managing big data.
* Developed simple and complex MapReduce programs in Java for Data Analysis on different data formats.
* Developed multiple map reduce jobs in java for data cleaning and pre-processing.
* Sqoop was used to pull data into Hadoop distributed file system from RDBMS and vice versa
* Defined workflows using Oozie.
* Used Hive to create partitions on hive tables and analyzes this data to compute various metrics for reporting.
* Created Data model for Hive tables
* Developed the LINUX shell scripts for creating the reports from Hive data.
* Good Experience in managing and reviewing Hadoop log files
* Used Pig as ETL tool to do transformations, joins and pre-aggregations before loading data onto HDFS.
* Worked on large sets of structured, semi structured and unstructured data
* Responsible to manage data coming from different sources
* Installed and configured Hive and also developed Hive UDFs to extend core functionality of hive
* Responsible for loading data from UNIX file systems to HDFS.

**Environment:**

Apache Hadoop, Hortonworks, MapReduce, HDFS, Hive, HBase, Pig, Oozie,Linux, Java, Eclipse 3.0, Tomcat 4.1, MySQL.

**Zions Bank, Salt Lake City, UT**

**Oct 2009 – Apr 2011**

**Java Developer**

**Description:**

Developed and maintained an online customer service system. This project involved implementing an online database for the users to enter technical problems and request service. The software allotted the requests to the technical staff depending on their area of expertise and recorded the amount of time required to service the requests. Online summary of the various requests and time taken for each were provided to facilitate faster response to future customer requests. The system allowed users to check the status of their requests using an online interface and provided notifications to the user via email once the request is satisfied.

**Responsibilities:**

* Participated in the discussions with business experts to understand Business requirements and translate them into technical requirements towards development.
* Designed concepts for frameworks using Spring and Hibernateand assisted with development environment configuration.
* Prepared the proof of concept by configuring the Spring MVC and Hibernate for various modules.
* Designed and developed functionality with excellent understanding of design patterns like singleton, List Iterator, Command, Factoryetc.
* Used HTTP Request and SOAP based Web services to post XML data to the End client.
* Exposed web services to the client applications by sharing the WSDL.
* Used Spring Frameworkto develop beans from already developed parent bean.
* Used Dependency Injection feature of spring framework and O/R mapping tool Hibernate for rapid development and ease of maintenance.
* Involved in designing, developing and deploying reports in MS SQL Server environment using SSRS-2008 and SSIS in Business Intelligence Development Studio (BIDS).
* Used ETL (SSIS) to develop jobs for extracting, cleaning, transforming and loading data into data warehouse.
* Worked with Cassandra Query Language (CQL) to execute queries on the data persisting in the Cassandra cluster.
* Developed database objects in SQL Server 2005 and used SQL to interact with the database during to troubleshoot the issues.
* Updated and saved the required data in the DB2 database using JDBC, corresponding to actions performed in the struts class.
* Involved in bug fixing and resolving issues with the QA.
* Developed SQL scripts to store data validation rules in Oracle database.
* Configured Log4j for logging activity at various levels and written test cases using JUnit.
* Involved in developing Ant build scripts for automating deployment on WebSphere test environment.
* Addressing high severity production issues on regular basis by researching and proposing quick fix or design change as required

**Environment:**

JAVA 1.6, J2EE1.6, Servlets, JDBC, Spring, Hibernate3.0, JSTL, JSP2, JMS, Oracle10g, Web Services, SOAP, Restful, Maven, Apache AXIS, SOAP UI, XML1.0, JAXB2.1, JAXP, HTML, JavaScript, CSS3, AJAX, JUnit, Eclipse, WebLogic10.3, SVN, Shell Script

**Smart Cards, Hyderabad, India**

**Jun 2008 – Sep 2009**

**Java Developer**

**Description:**

LM Benefits Client GUI an Intranet web application designed and developed to create a workspace for Highmark Business users to create Insurance Claim codes and associate various health diagnosis categories to each code. These claim codes are in turn used by Highmark Insurance Customers to claim their Benefits while claiming their insurance. Each Diagnosis code (or Product Service Component) is composed of several health related diagnosis attributes that are fetched from database (pre-loaded). The main functionalities involves creating new PSCOs, Searching PSCOs, PSCo Copy, Combining two or more PSCOs based on class/Type modifiers and displaying attributes for PSCo creation from Database, based on Business Rules.

**Responsibilities:**

* Understanding Use requirements participating in design discussions, implementation feasibility analysis both at front-end and backend level, documenting requirements.
* Using RUP and Rational Rose, developed Use Cases, created Class, Sequence and UML diagrams.
* Application Modeling, developing Class diagrams, Sequence Diagrams, Architecture / Deployment diagrams using IBM Rational Software Modeler and publishing them to web perspective with Java Doc.
* Participation did in Design Review sessions for development / implementation discussions.
* Designed & coded Presentation (GUI) JSP’s with Struts tag libraries for Creating Product Service Components (Health Care Codes) using RAD.
* Developing Test Cases and unit testing using JUnit
* Coded Action classes, Java Beans, Service layers, Business delegates, to implement business logic with latest features of JDK1.5 such as Annotations and Generics.
* Extensive use of AJAX and JavaScript for front-end validations, and JavaScript based component development using EXT JS Framework with cross browser support.
* Appropriate use of Session handling, data Scope levels within the application.
* Designed and developed DAO layer with Hibernate3.0 standards, to access data from IBM DB2 database through JPA(Java Persistence API) layer creating Object-Relational Mappings and writing PL/SQL procedures and functions
* Integrating Spring injections for DAOs to achieve Inversion of Control, updating Spring Configurations for managing Java objects using callbacks
* Application integration with Spring Web Services to fetch data from external Benefits application using SOA architecture, configuring WSDL based on SOAP specifications and marshalling and un-marshalling using JAXB
* Prepared and executed JUNIT test cases to test the application service layer operations before DAO integration
* Creating test environments with WAS for local testing using test profile. And interacting with Software Quality Assurance (SQA) end to report and fix defects using Rational Clear Quest.
* Creating views and updating code to IBM Rational Clear case for Source code control.
* Solving QA defects, scheduling fixes, providing support to production application.

**Environment:**Java: JDK 1.5, JSP, JSP Custom Tag libraries, JavaScript, EXT JS, AJAX, XSLT, XML, DOM4J 1.6, EJB, DHTML, Web Services, SOA, WSDL, SOAP, JAXB, IBM RAD, IBM WebSphere Application server, IBM DB2 8.1, UNIX, UML, IBM Rational Clear case, JMS, Spring Framework, Hibernate 3.0, PL/SQL, JUNIT 3.8, log4j 1.2, Ant 2.7.