****

**Sandeep Kumar Sarikonda**

**Certified Spark Developer**

**Summary:**

* Vibrant and self-motivated professional with more than 9 years of software development using **Spark, Hadoop, MapReduce, Java, Ruby on Rails.**
* Hortonworks Certified Spark Developer - [Badge](http://bcert.me/sasqyqxq).
* Good working experience in implementing end-to-end big data solutions with MapReduce, HDFS, Spark, Kafka, Hive, YARN, Flume, Sqoop for more than 4 years.
* Experienced in developing Java & Rails applications for over 5 years.
* In depth Knowledge of **Hadoop** Architecture & **YARN** in designing and developing big data applications.
* Excellent knowledge with distributed storages (HDFS) and distributed processing for real-time and batch processing (**Spark, SparkSQL, SparkStreaming, Hive**).
* Expertise on Hadoop components such has HDFS, Name Node, Job Tracker, Task Tracker, Data Node and Resource Manager(**YARN**).
* Hands on experience in writing **Ad-hoc Queries** for moving data from **HDFS** to **HIVE** and analyzing the data using **HIVE QL**.
* Adept in using **Zookeeper** and **Oozie** Operational Servicesfor coordinating the cluster and scheduling workflows.
* Solid understanding in using **Apache Flume** for collecting, aggregating and moving large amounts of data from application servers.
* Outstanding ability in writing transformations in **SparkSql** with **Dataset/Dataframe** API in Scala & Python.
* Good knowledge of RDMS technologies and SQL language.
* Hands on experience in Importing and exporting data from different databases like MySQL, Oracle, Teradata into Hadoop using **Sqoop**.
* Proficiency in manipulating/analyzing large datasets, deriving patterns or insights within structured and un structured data.
* Strong working knowledge with major distributions like HortonWorks, Cloudera.
* Good understanding of NoSQL databases like **HBase**, **Cassandra** & MongoDB.
* Experienced in writing and executing MapReduce programs that work with different file formats like text, sequence, xml, parquet and Avro.
* Experienced in working with Amazon Web Services (AWS) using EC2 for computing.
* Strong Requirements gathering experience using **JAD** Sessions and preparing functional documents like Use Cases, Software Requirements Specifications (SRS).
* Experienced in MVC and RESTful service oriented architectures.
* Self-motivated worker with strong focus on business goals and end-user experience.
* Worked on Integrated Development Environments like RubyMine, IntelliJ Idea, Eclipse, Text Mate, Net Beans.

**Technical Skills:**

|  |  |
| --- | --- |
| **Languages** | Ruby, Java, Scala, Python, Javascript. |
| **Hadoop Stack** | Hadoop, HDFS, MapReduce, Hive, SparkSQL, Spark Streaming, Yarn, Storm, Sqoop, Kafka, Flume, PIG. |
| **Hadoop Distributions** | MapR, Hortonworks, Cloudera. |
| **Query Languages** | HiveQL, SQL, PL/SQL. |
| **Databases** | HBase, MySQL, PostgreSQL, MongoDB, Cassandra. |
| **Build Tools** | AnthillPro, Jenkins, GIT, SVN. |
| **Frameworks** | Rails, Spring, Struts. |
| **Deployment Tools** | Amazon EC2, Heroku. |
| **Operating Systems** | Windows, Unix. |
| **IDEs** | Eclipse, RubyMine, IntelliJ, Subllime text. |
| **Other Tools** | MS Office (Excel, PowerPoint, Project 2013), Visual Studio 2013. |

**Professional Experience:**

**DST Systems Inc, Kansas City, MO April 2016 - Till date**

**Sr Big Data Developer**

**Roles & Responsibilities:**

* Responsible for building scalable distributed data solutions using Hadoop.
* Executed Job management using Fair scheduler and developed job processing scripts using Oozie workflow.
* Developed Scala scripts, UDFs using both Dataframes/SQL/Datasets and RDD in Spark 1.6 for Data Aggregation, queries and writing data back to storage system.
* Designed workflows and coordinators in **Oozie** to automate and parallelize Hive jobs on Apache Hadoop environment by **Hortonworks (HDP 2.4.3).**
* Involved in performance tuning of Spark Applications for correct level of Parallelism, Memory tuning and in reducing Data Shuffle.
* Worked on handling large datasets using Partitions, Spark in Memory capabilities, Broadcasts in Spark, Effective & efficient Joins, Transformations & actions during ingestion process itself.
* Worked on Cluster of size 130 nodes.
* Analyzed the SQL scripts and designed the solution to implement using Scala API.
* Involved in creating Hive tables, loading and analyzing data using Hive queries.
* Developed Hive queries to process the data and generate the data cubes for visualizing.
* Worked on Data Serialization formats for converting complex objects into sequence bits by using **AVRO, JSON, CSV** formats.
* Involved in loading and transforming large sets of structured data from relational databases into **HDFS** using **Sqoop** imports.
* Implemented Partitioning, Dynamic Partitions, and Buckets in **HIVE**.
* Integrated **Kafka** with Spark streaming for analyzing the real-time data.
* Collaborated with the infrastructure, network, database, application and BI teams to ensure data quality and availability.

**Environment**: HDP 2.4.3, Hadoop 2.7.1, Spark 1.6.2, Scala, Sqoop 1.4.6, Hive 1.2.1, HBase 1.1.2, Oozie 4.2.0, Flume, IntelliJ IDEA.

**United Health Group, Minneapolis, MN May 2014- March 2016**

**Sr Hadoop Developer**

**Roles & Responsibilities:**

* Involved in analyzing and cleansing raw data by performing **Hive queries** and running **Pig scripts** on data.
* Developed simple and complex **MapReduce** programs in Java for Data Analysis on different data formats.
* Involved in importing the real-time data to Hadoop using **Kafka** and implemented the **Oozie** job for daily imports.
* Developed custom aggregate functions using **Spark SQL** and performed interactive querying on a POC level.
* Implemented business logic by writing **Pig UDF's** in Java and used various **UDFs** from Piggybanks and other sources.
* Implemented **Spark** applications from existing **MapReduce** framework for better performance.
* Optimized **MapReduce** Jobs to use **HDFS** efficiently by using various compression mechanisms.
* Worked on partitioning **HIVE** tables and running the scripts in parallel to reduce run-time of the scripts.
* Worked on Data Serialization formats for converting Complex objects into sequence bits by using **AVRO, PARQUET, JSON, CSV** formats.
* Used cluster co-ordination services through **Zookeeper**.
* Created Hive tables, loaded data and wrote Hive queries that run within the map.
* Administration, installing, upgrading and managing distributions of **Hadoop, Hive, HBase**.
* Involved in moving all log files generated from various sources to **HDFS** for further processing through **Flume**.

**Environment:** MAPR Distribution (5.1), Hadoop 2.6.0, Hive, MapReduce, Spark 1.5.0, Scala, Java, Sqoop 1.4.6, HBase 1.1.0, ServiceNow.

**State Farm Insurance, Bloomington, IL Feb 2013 – Apr 2014**

**Hadoop Developer**

**Roles & Responsibilities:**

* Worked on writing transformer/mapping **Map-Reduce** pipelines using Java.
* Involved in creating **Hive Tables**, loading with data and writing **Hive queries** which will invoke and run **MapReduce** jobs in the backend.
* Designed and implemented **Incremental Imports** into **Hive** tables.
* Worked in Loading and transforming large sets of structured, semi structured and unstructured data
* Involved in collecting, aggregating and moving data from servers to **HDFS**.
* Written **Hive** jobs to parse the logs and structure them in tabular format to facilitate effective querying on the log data
* Experienced in managing and reviewing the **Hadoop** log files. Implemented the workflows using Apache **Oozie** framework to automate tasks
* Worked on different file formats like Sequence files, XML files and Map files using **MapReduce** Programs.
* Involved in Unit testing and delivered Unit test plans and results documents using **Junit** and **MRUnit**.
* Developed scripts and automated data management from end to end and sync up between all the clusters.
* Created and maintained Technical documentation for launching Hadoop Clusters and for executing **Pig** Scripts.
* Involved in Setup and benchmark of **Hadoop** /**HBase** clusters for internal use.

**Environment:** HDP, Hadoop, Pig, MapReduce, HBase, HDFS, Apache Solr, Java, Eclipse.

**Raymond James, St. Petersberg, FL Feb 2011- Jan 2013**

**RoR Developer**

**Roles & Responsibilities:**

* **Rails Framework** is used to design the UI part of the report. **Prawn Gem** is used for the implementation of PDF version of the report.
* Designed and developed the front-end of the application using RAILS, **AJAX, CSS, JSON** and **jQuery. Active Records** is used for the back-end of the application.
* Used **SCRUM** and **Agile** Development Methodology for the project.
* Used **Cucumber** for Test-Driven Development.
* Implemented the Query part of the reports using Ruby, **ActiveRecords** and ran the rake tasks for every report weekly, dumping the SQL data to **CSV file**.
* Fixed bugs related to old features and worked on new features in a space of 3 weeks.
* Gave daily updates to clients through teleconferences and formal status updates.
* Actively involved in fine tuning the application.
* Interacted with the design team to add extra new features to the project.
* Took care of bug fixes and code reviews.
* Coordinated with the Business Analyst to get the details of each of the reports.

**Environment**: Ruby 1.9, Ruby Gems, Rails 2.3, MySQL 5.0, Ubuntu 10.04, Webrick and Phusion Passenger server, Git Repository, Aptana RadRails IDE.

**GeekLabs, India Sep 2009 – Nov 2011**

**Java Developer**

**Roles & Responsibilities:**

* Worked on developing a web application based on **Java**.
* Worked on different design patterns.
* Responsible for requirement gathering along with business analyst.
* Responsible for creating and updating applications related understanding documentsl
* Actively involved in writing **Test scripts** and doing System testing for **JRE** and **IE8** browser compatibility for my applications.
* Actively participated in weekly and monthly status & business user meetings.
* Involved in conducting meetings with Pfizer business users for understanding gaps in the CTO applications.
* Creating programs to use **JMS** and message queuing to process requests.
* Involved in the Development and Deployment of Stateless Session beans.
* Generated deployment descriptors for **EJBs** using **XML**.
* Developed **GUI** related changes using **JSP**, **HTML** and client validations using Java script.
* Used **JDBC** to communicate with database.
* Involved in writing Managed **Beans/Controller** logic for assigned modules.
* Responsible for developing the functionalities as per use case documents.
* Responsible for writing navigation-rules and configuring managed beans in Faces-**config.xml file**.
* Implemented Java documentation for complete application.
* Deployed the application on Tomcat server at client locations.

**Environment**: Java 1.4, JSP, HTML, Java Script, Struts, Springs, Apache Tomcat, Eclipse, MySQL.

**Netwin Soft Solutions, India Jun 2008 – Aug 2009**

**Java Developer**

**Roles & Responsibilities:**

* Analyzed the requirements based on business data and user requirements.
* Responsible for leading all phases of Java EE web product development and web product quality.
* Handle the tasks of scheduling all project activities and managing risks throughout development cycle.
* Perform responsibilities of reviewing and clarifying business requirements for new application development.
* Handle the tasks of drafting technical specification for business stakeholders.
* Responsible for reviewing technical solutions/designs prepared by junior engineers.
* Perform the tasks of providing directions to technical resources in providing technical solutions.
* Played a lead role in designing projects as well as reviewed proposed designs to ensure application integrity.
* Wrote server-side programs by using Servlets and JSP**.**
* Created prototype in **HTML** to understand requirement.
* Discussions with Project Manager and business analyst to understand requirement.
* Implement requirement by creating interface in **Java server faces** and server-side technologies **like EJB, seam 2.0, JPA**.
* Used **agile methodology** for development of application.

**Environment:** Jdbc, Servlets, Java, EJB.

**Education:** Bachelor of Technology in Mechanical Engineering, Kakatiya University, India (2000 – 2004).