**Sravan**

[**Rupa.d@solomonsint.com**](mailto:Rupa.d@solomonsint.com)

**2152973322**

**SUMMARY:**

* Over 6 years of overall experience as software developer in design, development, deploying and supporting large scale distributed systems.
* Over 3+ years of extensive experience as **HadoopDeveloper** and **Big DataAnalyst**.
* Primary technical skills in **HDFS**, **MapReduce**, **YARN, Pig, Hive, Sqoop, HBase, Flume, Oozie, Zookeeper.**
* In depth understanding/knowledge of Hadoop Architecture and various components such as **HDFS**, **Job Tracker, Task Tracker, Name Node, Data Node** and **MapReduce** concepts and experience in working with MapReduce programs using Apache Hadoop for working with Big Data to analyse large data sets efficiently.
* Hands on experience in working with Ecosystems like **Hive, Pig, Sqoop, Map Reduce, Flume, Oozie**. Strong knowledge of **Pig** and **Hive’s** analytical functions, extending Hive and Pig core functionality by writing custom UDFs.
* Experience in importing and exporting Tera bytes of data between HDFS and Relational Database Systems using **Sqoop**.
* Extensively used ETL methodology for performing Data Profiling, Data Migration, Extraction,transformation and Loading using Talend and designed data conversions from wide variety of source systems including Netezza, Oracle, DB2, SQL server, Teradata, Hive, Hanaand non-relational sources like flat files, XML and Mainframe Files.
* Experienced in Different Distributions like Cloudera, HortonWorks and MapR.
* Knowledge of job workflow scheduling and monitoring tools like Oozie and Ganglia, NoSQL databases such as **HBase**, **Cassandra**, **BigTable**, administrative tasks such as installing Hadoop, **Commissioning** and **decommissioning**, and its ecosystem components such as **Flume**, **Oozie**, **Hive** and **Pig**.
* Knowledge on importing and exporting data using Flume and kafka.
* Experienced in writing ETL codes
* Experienced in creating Triggers on TAC server to schedule Talend jobs to run on server.
* Experience in writing UNIX reusable scripts/codes for Informatica jobs that are used for data movement and transformation purpose.
* Experience designing and implementing fast and efficient data acquisition using Big Data processing techniques and tools.
* Good Knowledge and experience in performance tuning in the live systems for **ETL/ELT** jobs that are built on Informatica.
* Experience in writing database objects like **Stored Procedures, Triggers for Oracle, MS SQL**
* Creating the Data Frames handle in SPARK with Scala and python.
* Hands On experience on developing UDF , DATA Frames and SQL Queries in SPARK SQL
* Developed PIG Latin scripts,Python scripts and SPARKSQL scripts for handling data formation.
* Hands on experience on Real Time data tools like Kafka and Storm.
* Fast learner with good interpersonal skills, having strong **analytical** and **communication** skills and interested in problem solving and troubleshooting.
* Self-motivated, excellent **teamplayer**, with positive attitude and adhere to strict deadlines.

**TECHNICAL SKILLS:**

**Operating systems:** WINDOWS, LINUX (Fedora, CentOS), UNIX

**Languages and Technologies:** C, C++, Java, SQL, PLSQL

**Scripting Languages:** Shell scripting

**Databases:** Oracle, MySQL, Postgre SQL

**IDE:** Eclipse and Net Beans

**Application Servers:** Apache Tomcat server, Apache HTTP webserver

**Versioning Systems:** Git, SVN

**Hadoop Ecosystem:** Hadoop MapReduce, HDFS, Flume, Sqoop, Hive, Pig, Oozie, Cloudera

Manager Zookeeper. AWS EC2, Hortonworks HDP 2.3/2.4

**Apache Spark:** Spark, Spark SQL, Spark Streaming.SCALA, spark with python

**Cluster Mgmt.& Monitoring:** Cloudera Manager, Horton works Ambari, Ganglia and Nagios.

**Security:** Kerberos.

**No-SQL:** Cassandra,HBase,DataStax

**EDUCATION:**

**Masters in Computer Science New Orleans,LA**

**University of New Orleans**

**Bachelor of Technology in Computer Science Hyderabad, India**

*Jawaharlal Nehru Technological University*

**CERTIFICATIONS:**

* **CCAH – Cloudera Certified Administrator for Apache Hadoop**
* **CCDH – Cloudera Certified Developer for Apache Hadoop**
* **IBM Certified Database Associate(Db2)**
* **IBM Rational Application Developer v6**

**PROFESSIONAL EXPERIENCE**

**ADOBE Headquarters (SAN JOSE) July 2017-Jan 2018**

**Hadoop Developer**

**Project Description:** Adobe is an American multinational computer software company.The project involved attribute automation and development of required data which is used for executive dashboards.

**Roles:**

* Converted existing SQL queries into Hive QL queries.
* Implemented UDFs, UDAFs, UDTFs in java for hive to process the data that can’t be performed usingHive inbuilt functions
* Effectively used Oozie to develop automatic workflows of Sqoop, Mapreduce and Hive jobs.
* Exported the analyzed data into relational databases using Sqoop for visualization and to generate reports for the BI team.
* Able to integrate state-of-the-art Big Data technologies into the overall architecture and lead a team of developers through the construction, testing and implementation phase.
* Experience in all phases of Data warehouse development from requirements gathering for the data warehouse to develop the code, Unit Testing and Documenting.
* Parsed high-level design specification to simple ELT coding and mapping standards.
* Utilized **Agile Scrum Methodology** to help manage and organize a team of 4 developers with regular code review sessions.
* Weekly meetings with technical collaborators and active participation in ELT code review sessions with senior and junior developers.
* Automated jobs using cisco’s Tidal Scheduler
* Written Shell scripts with appropriate logging features to automate few jobs which were done manually and scheduled them in tidal
* Performed regression testing for integral code releases
* Worked along with the kyos team to build the cube required for tableu and Power BI dashboard view

**Environment:** Cloudera 5.8, Microsoft Azure, Hue, Shell, Kyos cube, Cisco Tidal, Oozie, hive, Hbase

**Direct Energy(Houston) Feb 2015-June 2017**

**Big Data Consultant**

**Roles:**

* Used Sqoop to transfer data between RDBMS and HDFS.
* Involved in collecting and aggregating large amounts of streaming data into HDFS using Flume and defined channel selectors to multiplex data into different sinks.
* Implemented complex map reduce programs to perform map side joins using distributed cache
* Worked on migrating MapReduce python programs into Spark transformations using Spark and Scala.
* Set up the Cloudera clusters on AWS
* Accesed hive from python using pyhive library
* Intergrated Kafka with spark and improved the performance of data ingestion.
* Consulted with business partners and made recommendations to improve the effectiveness of Big Data systems, descriptive analytics systems, and prescriptive analytics systems.
* Designed and implemented custom writable, custom input formats, custom partitions and custom comparators in MapReduce.
* PyMySQL and MySQLdb are both database connectors for Python, libraries to enable Python programs to talk to a MySQL server
* Configured Spark Streaming to receive real time data from the Apache Kafka and store the stream data to HDFS using Scala.
* Thoroughly tested MapReduce programs using MRUnit and JUnit testing frameworks.
* Responsible for troubleshooting issues in the execution of MapReduce jobs by inspecting and reviewing log files
* Exploring with the Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark Context,Scala,Phython, Spark-SQL, Data Frame, and Pair RDD's
* Connected mysql to hdfs using python pymysql lib
* Converted existing SQL queries into Hive QL queries.
* Implemented UDFs, UDAFs, UDTFs in java for hive to process the data that can’t be performed usingHive inbuilt functions
* Effectively used Oozie to develop automatic workflows of Sqoop, Mapreduce and Hive jobs.
* Exported the analyzed data into relational databases using Sqoop for visualization and to generate reports for the BI team.
* Able to integrate state-of-the-art Big Data technologies into the overall architecture and lead a team of developers through the construction, testing and implementation phase.
* Experience in all phases of Data warehouse development from requirements gathering for the data warehouse to develop the code, Unit Testing and Documenting.
* Parsed high-level design specification to simple ELT coding and mapping standards.
* Using Kafka on publish-subscribe messaging as a distributed commit log, have experienced in its fast, scalable and durability.
* Utilized **Agile Scrum Methodology** to help manage and organize a team of 4 developers with regular code review sessions.
* Weekly meetings with technical collaborators and active participation in ELT code review sessions with senior and junior developers.
* Loaded and analyzed **Omniture logs** generated by different web applications.
* Loaded and transformed large sets of structured, semi structured and unstructured data in various formats like text, zip, XML and JSON.
* Refined the Website clickstream data from Omniture logs and moved it into Hive.
* Written multiple **Map Reduce** programs to power data for extraction, transformation and aggregation from multiple file formats including XML, JSON, CSV & other compressed file formats.
* Defined job flows and developed simple to complex Map Reduce jobs as per the requirement.
* Optimized Map/Reduce Jobs to use HDFS efficiently by using various compression mechanisms.
* Developed **PIG UDF**s for manipulating the data according to Business Requirements and also worked on developing custom PIG Loaders.
* **PIG UDF** was required to extract the information of the area from the huge data which we get from the sensors
* Responsible for creating **Hive** tables based on business requirements
* Implemented Partitioning, Dynamic Partitions and Buckets in **HIVE** for efficient data access.
* Involved in **NoSQL** database design, integration and implementation.
* Loaded data into **NoSQL** database **HBase**

**Environment: Amazon web services (AWS EC2),Sqoop, MapReduce, Hive, Oozie, Pig Latin,Kafka, HBase,Cloudera Distribution**

**Spark InfoTech June 2012- July 2014**

**Junior Java Developer**

**Roles:**

* Involved in design phase meetings for Business Analysis and Requirements gathering.
* Worked with business functional lead to review and finalize requirements and data profiling analysis.
* Worked on MVC framework
* Worked on entry level Java programming assignments.
* Responsible for gathering the requirements, designing and developing the applications.
* Worked on UML diagrams for the project use case.
* Worked with Java String manipulations, to parse CSV data for applications.
* Connected Java applications to Java database to read, write data.
* Developed static and dynamic Web Pages using JSP, HTML and CSS.
* Processed the webservices data with spring framework
* Worked on JavaScript for data validation on client side.
* Involved in structuring Wiki and Forums for product documentation
* Maintained the customer support portal.

**Environment**: Java, Servlets,Spring, JSP, JavaScript, HTML, PHP, CSS, Eclipse