**Santhosh Kumar**

**Santhoshk7543@gmail.com**

**(347) 878-0172**

**PROFESSIONAL SUMMARY:**

* Over 8+ years of total IT Experience, including 4+ year’s hands-on experience in Big-data/Hadoop Technologies.
* 4+ years of extensive experience in, J2EE Technologies, Database development, ETL Tools, Data Analytics.
* Very good experience in the Application Development and Maintenance of SDLC projects using various technologies such as Java/J2EE, JavaScript, Data Structures and UNIX shell scripting.

**BIG DATA EXPERIENCE**

* Expertise in all components of Hadoop Ecosystem- Hive, Hue, Pig, Sqoop, HBase, Impala, Flume, Zookeeper, Oozie, and Apache Spark.
* Having sound experience in Big Data Hadoop Ecosystems experience in Ingestion, storage, querying, processing and analysis of big data.
* Excellent understanding/knowledge of Hadoop architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, YARN and MapReduce programming paradigm.
* Experience in importing and exporting data from different RDBMS like MySQL, Oracle and SQL Server into HDFS and Hive using Sqoop.
* Good understanding of NoSQL databases and hands on experience in writing applications on NoSQL databases like HBase.
* Experience in developing custom MapReduce programs using Apache Hadoop to perform Data Transformation and analysis as per requirement.
* Experience in writing PIG scripts and Hive Queries for processing and analyzing large volumes of data.
* Knowledge on NoSQL databases such as HBase and creating mapping Phoenix tables in HBase to query HBase using SQLs.
* Experience in deploying and managing the multi-node development and production Hadoop cluster with different Hadoop components (HIVE, PIG, SQOOP, OOZIE, FLUME, HCATALOG, HBASE, ZOOKEEPER) using Hortonworks Ambari.
* Hands on experience on Scala language features - Language fundamentals, Classes, Objects, Traits, Collections, Case Classes, High Order Functions, Pattern Matching, Extractors etc.
* Experience on Creating Internal and External tables and implementation of performance improvement techniques using partitioning tables, bucketing tables in Hive.
* Experience in analyzing data using HBase and custom Map Reduce programs in Java.
* Experience in creating PIG and HIVE UDFs using java in order to analyze data sets.
* Experience in collecting the log data from different sources like (web servers and social media) using Flume and storing on HDFS to perform MapReduce jobs.
* Experience in Spark Streaming in order to ingest data from multiple data sources into HDFS.
* Good Hands on Experience in Stream Processing including Storm and Spark Streaming.
* Experience in Design & Development, tuning and maintenance of NoSQL databases such as MongoDB, HBase, Cassandra and its Integration with Hive.
* Used Cassandra in accordance with PIG/Hive as and when required for real time low latency queries.
* Developed multiple POCs using Python Spark API and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL.
* Very good knowledge in advanced Big-Data technologies like Spark Ecosystem (Spark SQL, SparkR and Spark Streaming), Kafka and Predictive analytics.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Spark SQL and Scala.
* Worked on reading multiple data formats on HDFS using Python Spark API.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Programming Languages** | SQL, C, C++, Java, Core Java, Java 8, J2EE, Python, Scala, Pig Latin, HiveQL and Unix shell scripting |
| **Big Data Technologies** | Hadoop, HDFS, MapReduce, Hive, Pig, HBase, Hue, Sqoop, Storm, Kafka, Oozie, SparkSql, SparkR, Flume, Zookeeper, Cassandra, Spark,Hortonworks. |
| **Databases** | Oracle, MySQL, SQL Server, DB2, Familiar with NoSQL (HBase, Cassandra, MongoDB) |
| **Data wrangling tool** | Alteryx |
| **Scripting & Query Languages** | UNIX Shell scripting, SQL and PL/SQL. |
| **Web Technologies** | JSP, Servlets, JavaBeans, JDBC, AWT, Swing, JSF, XML, CSS, HTML, XHTML, JavaScript, AJAX. |
| **Hadoop Paradigms** | MapReduce, YARN, In-memory computing, High Availability, Real-time Streaming. |
| **Operating Systems** | Windows, UNIX, Linux distributions (Centos, Ubuntu), Mac OS |
| **Other Tools** | Eclipse, Tableau 9.0, JUnit, QTP, JIRA, QC (Quality Center) |

**PROFESSIONAL EXPERIENCE**

**Centene Corporation, St. Louis, MO April 2018 – Present**

**Role: Big Data/Hadoop Developer**

**Responsibilities:**

* Involved in Analysis, Design, System architectural design, Process interfaces design, design documentation.
* Responsible for developing prototypes of the selected solutions and implementing complex big data projects with a focus on collecting, parsing, managing, analyzing and visualizing large sets of data using multiple platforms.
* Involved in Installing, Configuring Hadoop Eco System, Cloudera Manager using Cloudera Distribution.
* Understand how to apply technologies to solve bigdata problems and to develop innovative big data solutions.
* Developed various data loading strategies and performed various transformations for analyzing the datasets by using Cloudera Distribution for Hadoop ecosystem.
* Worked extensively with importing metadata into Hive and migrated existing tables and applications to work on Hive and Spark.
* Utilized Built-in Maven repositories for Cloudera distributions.
* Experienced in Monitoring Cluster using Cloudera manager.
* Experience in using HDFS and My SQL and deployed HBase integration to perform OLAP operations on HBase data. Creating HBase tables for random read/writes by the map reduce programs.
* Importing and exporting data from different databases like DB2, Teradata into HDFS and HBASE using Sqoop.
* Implemented of POC ETL solution using Spark, Cassandra, Alteryx and tableau.
* Created sessions, tasks and workflows using Alteryx.
* Experience with data wrangling and creating workable datasets.
* Utilized Alteryx Software and SQL query engines as a means of data cleansing, preparation and blending to ensure consistency and quality
* Performed data wrangling to clean, transform and reshape the data.
* Involved in converting Hive QL queries into Spark transformations using Spark RDDs, and Scala.
* Developed Spark code using Scala and Spark-SQL/Streaming for faster testing and processing of data.
* Utilized Apache Phoenix to hold all the metadata of the processes and for memory management of MR and Spark Jobs.
* Responsible for loading Data pipelines from web servers and Teradata using Sqoop with Kafka and Spark Streaming API.
* Used Spark for interactive queries, processing of streaming data and integration with popular NoSQL database for huge volume of data.
* Performance tuning the Spark jobs by changing the configuration properties and using broadcast variables.
* Worked on HBase Java API to populate operational HBase table with Key value. Worked with Hadoop eco system covering HDFS, HBase and YARN .
* Worked on performing transformations & actions on RDDs and Spark Streaming data.
* Responsible in handling Streaming data from web server console logs
* Developed multiple POCs using Spark and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata.
* Implementing Sqoop jobs, PIG and Hive scripts for data ingestion from relational databases to compare with historical data.
* Implemented scripts in Spark to import and export data from Cassandra, Teradata, Hadoop.
* Developed RDD's/Data Frames in Spark using Scala and applied several transformation logics to load data from Hadoop Data Lake to HBase.
* Worked on various file formats and compressions Text, Avro, Parquet file formats, snappy, bz2, gzip compression.
* Used Spark API over Hadoop YARN to perform analytics on data in Hive.
* Created tasks for incremental load 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 Hive optimized joins to gather data from different sources and run ad-hoc queries on top of them.
* Wrote Hive Generic UDFs to perform business logic operations at record level.
* Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig and used Zookeeper to coordinate the clusters.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.
* Troubleshooting the cluster by reviewing Hadoop LOG files.
* Worked on reading multiple data formats on HDFS using Python Spark API.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Python and Scala.
* Used Zookeeper for various types of centralized configurations.

**Environment*:*** Hadoop, Spark core, Scala, Spark, Cloudera distribution, Apache Phoenix, Python, Spark SQL, Alteryx  ,Spark Streaming, Hive, HBase, Pig, Sqoop, Kafka, Zookeeper, Java 8, and UNIX Shell Scripting.

**Anthem, Indianapolis, Indiana Jan 2017 – Mar 2018**

**Role: Hadoop/Spark Developer**

**Responsibilities:**

* Developed data pipeline using Spark, Hive and HBase to ingest customer behavioral data and financial histories into Hadoop cluster for analysis.
* Hands on experience in designing, developing, and maintaining software solutions in Hadoop cluster.
* Exploring with the Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark SQL, Data Frame, Spark Yarn.
* Build scalable distributed data solutions using Hadoop Cloudera Distribution. Ensure data accuracy and create reports.
* Wrote multiple queries to pull data from HBase .
* Created and optimized Hive queries to analyze the customer frustration level which helped reduce the call traffic and ensured better customer satisfaction.
* Involved in creating Hive tables, loading structured data and writing hive queries which will run internally in MapReduce way.
* Adding/installation of new components and removal of them through Ambari.
* Experienced with Spark Streaming to ingest data into Spark Engine.
* Designed the ETL runs performance tracking sheet in different phases of the project and shared with Production team.
* Involved in converting Hive/SQL queries into Spark Transformations using Spark RDDs and Scala.
* Involved in using SQOOP for importing and exporting data between RDBMS and HDFS.
* Used Hive to analyze the Partitioned and Bucketed data and compute various metrics for reporting.
* Involved in developing Hive DDLS to create, alter and drop Hive tables.
* Involved in loading data from Linux file system to HDFS.
* Created PIG scripts to load, transform and store the data from various sources into HIVE metastore.
* Developed and maintained ETL (Extract, Transformation and Loading) mappings to extract the data from multiple source systems like Oracle, SQL server and Flat files and loaded into Teradata.
* Importing and exporting data into HDFS and Hive using Sqoop.
* Experienced in managing and reviewing Hadoop log files.
* Involved in data warehousing and Business Intelligence systems.
* Responsible for System performance management, Systems change / configuration management and Business requirements management.
* Primary contributor in designing, coding, testing, debugging, documenting and supporting all types of applications consistent with established specifications and business requirements to deliver business value. Identify and design most efficient and cost-effective solution through research and evaluation of alternatives.
* Demonstrated Hadoop practices and broad knowledge of technical solutions, design patterns, and code for medium/complex applications deployed in Hadoop production.
* Ingested semi structured data using Flume and transformed it using Pig.

**Environment*:*** Hadoop, Spark, Hive, Pig, Spark SQL, Spark Streaming, HBase, Sqoop, Python, Kafka, Cloudera , Scala, Scala IDE(Eclipse), Maven, HDFS.

**State Farm Insurance, Boston, MA April 2015 – Dec 2016  
Role: Java Developer**   
 **Responsibilities:**

* Involved in Various Stages of Software Development Life Cycle (SDLC) deliverables of the project using the AGILE Software development methodology.
* Attended Daily scum meeting, sprint grooming/review and demo with management and other teams.
* Implemented the application using Spring MVC Framework and handled the security using spring security.
* Developed Controller Classes using spring MVC, spring AOP frame work.
* Designed and Developed End to End customer self service module using annotation-based Spring MVC, Hibernate, Java Beans and jQuery.
* Developed the User Interface using JSP, jQuery, HTML5, CSS3, Node JS, Bootstrap JS and Angular JS.
* Implement functionality such as searching, filtering, sorting, categories, validating using Angular framework.
* Used Angular directives, working on attribute level, element level and class level directives.
* Implemented Bean classes and configured in spring configuration file for dependency injection.
* Implemented persistence layer using Hibernate that use the POJOs to represent the persistence database.
* Created mappings among the relations and written named HQL queries using Hibernate.
* Design common framework for REST API consumption using Spring Rest Templates.
* Used Design Patterns like Facade, Data Transfer Object (DTO), MVC, Singleton and Data Access Object (DAO).
* Written SQL and PL/SQL queries like stored procedures, triggers, indexes, views.
* Used Log4j, Junit for logging and Testing.
* Documented all the SQL queries for future testing purpose.
* Prepared test case scenarios and internal documentation for validation and reporting.
* Coordinating with the QA team, and resolving the QA defects.
* Wrote services to store and retrieve user data from the Mongo DB.
* Worked with WebSphere application server that handles various requests from Client.
* Deploying fixes and updates using IBM WebSphere application server.
* Experience in developing automated unit testing using Junit, Mockito frameworks.
* Used Git controls to track and maintain the different version of the project.

**Environment:** Java, Spring, Hibernate, Web Services, Html 5, CSS3, Java Script, Angular JS, Bootstrap, MAVEN, WebSphere, Eclipse, IBM ODM, JUnit, Mockito, jQuery, log4j, Mongo DB, Windows, SOAP UI, REST Client, AWS, Git.

**Quomatic Solutions Pvt Ltd – Bangalore, INDIA Jan 2013 - Dec 2014**

**Role: Java Developer**

**Responsibilities:**

* Reviewed the functional, non-functional requirements and high-level designs.
* Worked with Agile Methodologies.
* Developed application using Struts MVC, Spring Dependency Injection and iBATIS as the ORM tool.
* Worked extensively in the backend Oracle database.
* Created logging system for development stage using Apache Commons Logging.
* Developed web pages using HTML, CSS, JavaScript, Angular Js and Free marker.
* Produced SOAP Web Service using Apache Axis. Generated WSDL files for the Service that served XML data.
* For logging and other functions used Spring AOP and Spring IOC Framework to implement them.
* Soap UI is used for web services testing.
* Used multithreading in programming to improve overall performance.
* Developed Multi-Threaded/Concurrent applications and Unix Shell Scripting.
* Implemented Collections for managing the dynamic data and other Core Java APIs.
* Worked on Garbage Collection for memory issues to improve the performance.
* Created functions and stored procedures in Oracle SQL/PL-SQL.
* Worked with XPath when need to identify objects in the XML.
* Developed lookup tables and views to speed up queries.
* Provide critical support to current order management systems being used as and when required.
* Used Jenkins as build management tool. Selenium was used to keep a track of bugs.
* Worked in Quality analysis and testing.

**Environment:** Java, EJB, Hibernate, Struts, Free Marker, Selenium, Oracle SQL, PL/SQL, XML, XPath, JUnit, Eclipse IDE, HTML, CSS, JavaScript, SOAP, JAX-B, JAX-RPC, JAX-WS, JAX-RS, Apache Tomcat, Apache Commons Logging, SOAP, Multi-Threaded/Concurrent, Collections, Stored procedures.

**EPAM – Hyderabad, India Mar 2010 - Dec 2012 Role: Software Developer**

**Responsibilities:**

* Core member of the one of the main products of the company
* The core project was developed using Core Java, Multithreading and JDBC
* Analyzed and understood the business requirements.
* Involved in various SDLC phases of the project.
* Connecting to MySQL database using JDBC
* Developed the presentation tier using Struts Framework.
* Used Hibernate as ORM and also involved in configuring hibernate to access database for various modules.
* Designed the database tables and wrote the respective mapping files in hibernate and POJO classes.
* Used log4j for logging for the application.
* Build the Ant Scripts for building and deploying the application.
* Automated the settlement process with the application and Banks and sends the reports to the client.
* Developed the loyalty program applications to enhance the client business.
* Developed a protocol between the client and server using XML and build the parse to communicate between both ends.
* Developed a router application for the ICICI Bank to route the transactions.
* Implementation of PGP in Java to transfer and receive the confidential files between our servers and clients.
* Developed test cases and performed unit testing using JUnit.
* Used CVS to maintain the versions of the source code of the project.
* Implemented many file parses and XML parses in the application to communicate between the server and client.
* Developed the UI using HTML, JSP, jQuery, AJAX, JavaScript.
* Certification of the product with various Banks and meet the standards of the industry.

**Environment:** Core Java/J2EE, Multithreading, JDBC, Struts, Java Mail, JDBC, MySQL, MSSQL, Ant, Apache Tomcat, WSDL, SOAP, XML, Log4j, Hibernate, JSP, Ajax, jQuery, SFTP, PGP, Linux.