|  |  |
| --- | --- |
| **Contact Number** | 512-766-4489 |
| **Alternate Contact Number** |  |
| **Current location of the candidate (City/State)** | Orlando, FL |
| **Email Address** | [shravan0620@gmail.com](mailto:shravan0620@gmail.com) |
| **Skype ID** |  |

**SHRAVAN KUMAR REPALA**

**PROFESSIONAL SUMMARY**

* Having 8 + years of IT experience, including 3+ years hands on experience in Hadoop/Big Data technologies and 4+ years of extensive experience in JAVA, J2EE Technologies and Database development
* Expertise in all components of Hadoop Ecosystem- Hive, Hue, Pig, Sqoop, HBase, Impala, Flume, Zookeeper, Oozie, Apache Nifi and Apache Spark.
* Good Experience in Linux Bash scripting and following PEP Guidelines in python.
* Collected data from different sources like web servers and **social media** for storing in HDFS and analyzing the data using other Hadoop technologies.
* Extensive experience in analyzing **social media** data and optimization of **social media** content.
* Excellent understanding of Hadoop architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, YARN and MapReduce programming paradigm.
* Experience in managing scalable Hadoop clusters including cluster designing, custom configurations, monitoring and maintaining of different Hadoop distributions like Cloudera CDH, Apache Hadoop, Hortonworks and Map Reduce.
* Experience in creating Nifi flow to streaming data between Kafka, PulsarDB, ElasticSearch, and FTP. And also deploying those Nifi flows both into staging and production.
* Experience in writing Pig Latin, HiveQL scripts and extended their functionality using User Defined Functions (UDF’s).
* Experience in analyzing data usingHiveQL, Pig Latin, HBase and custom MapReduce programs in Java.
* Experience in handling structured arrangement of data within certain limits (Data Layout’s) using Partitions and Bucketing in Hive.
* Experience in Hadoop Shell commands, writing Map Reduceprograms, verifying, managing and reviewing Hadoop log files.
* Knowledge on NOSQL databases like MongoDB, HBase, Cassandra and its Integration with Hive.
* Experience in importing and exporting data using Sqoop from HDFS to Relational Database Systems and vice-versa.
* Experience in collecting the log data from different sources like (webservers and **social media**) using Flume and storing in HDFS to perform MapReduce jobs.
* Integrated BI tool with Impala and analyzed the data.
* Experience in installation, configuration, supporting and managing- Cloudera'sHadoop platform along with CDH3 & CDH4 clusters.
* Experience with Cloudera Impala for ad-hoc querying and Apache sentry for role based authentication.
* Strong team player with good communication, analytical, presentation and inter-personal skills.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Programming Languages** | SQL, C, C++, Java, Core Java, J2EE, Python, Pig Latin, HiveQL and Unix shell scripting |
| **Big Data Technologies** | Hadoop, HDFS, Map Reduce, Hive, Pig, HBase, Impala, Hue, Sqoop, Storm, Kafka, Oozie, Flume, Zookeeper, Nifi, Cassandra, Spark 1.4, Cloudera CDH4, and Horton works. |
| **Databases** | Oracle, MySQL, SQL Server, DB2, Familiar with NoSQL (HBase, Cassandra, MongoDB) |
| **Scripting and 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, Linuxdistributions (Centos, Ubuntu), Mac OS |
| **Other Tools** | Eclipse, Informatica, Tableau 9.0, JUnit, QTP, JIRA, QC (Quality Center) |

**PROFESSIONAL EXPERIENCE:**

**Deloitte, Orlando, FL Jan’17-Present**

**Role: Hadoop Developer**

**Responsibilities:**

* Experience in developing solutions to analyze large data sets efficiently
* Developed Map Reduce application to find out the useful metrics from the data. Did a thorough
* testing in local mode and distributed mode found bugs with the code and ensured 100% issue free

delivery to production.

* Experienced in query tuning.
* Implemented Kafka to handle real time data.
* Worked on spark sql and spark streaming.
* Implemented multiple sqoop jobs.
* Developed algorithms for identifying influencers with in specified social network channels.
* Involved in fetching brands data from **social media** applications like Face book, twitter.
* Worked on POC Recommendation System for **social media** using Movie lens dataset.
* Written Ozzie work flows to automate jobs.
* Written multiple python scripts for data visualizations and ETL automations
* Expert level understanding of Map Reduce internals, including shuffling and partitioning.
* Created Hive external tables and managed tables, designed data models in hive.
* Implemented business logic using Pig and hive udf’s.
* Finding the right joins logics and creates valuable data sets for further data analysis.
* Responsible to develop custom udf’s in pig, hive.
* Developed multiple pig and hive jobs in Java for data cleaning and processing.
* Responsible for building scalable distributed data solutions using Hadoop.
* Worked hands on with ETL process.
* Handled importing of data from various data sources, performed transformations using Hive,

MapReduce, and loaded data into HDFS.

* Extracted the data from multiple data sources into HDFS using Sqoop.
* Developed Hive queries to process the data and generate the data cubes for visualizing.

**Environment:** Linux, MapReduce, HDFS, Hive, Ooozie, Java (jdk1.6), eclipse, HBase (NoSQL), Sqoop, pig.

**HGST- San Jose, CA Mar’15-Dec’16**

**Role: Big Data –Hadoop Developer**

**Responsibilities:**

* Involved in Installing, Configuring Hadoop components using CDH 5.2 Distribution.
* Responsible for analyzing large data sets and derive customer usage patterns by developing new MapReduce programs.
* Written MapReduce code to parse the data from various sources and storing parsed data into Hbase and Hive.
* Worked on creating combiners, partitions, and distributed cache to improve the performance of MapReduce jobs.
* Developed Shell Script to perform data profiling on the ingested data with the help of HIVE Bucketing.
* Responsible for understanding the scope of the project and developed algorithm for identifying the influencers per brand across various social networking sites.
* Developed and generated insights based on brand conversations, which in turn helpful for effectively driving brand awareness, engagement and traffic to **social media** pages.
* Migrate infrastructure services to new data center, e.g. Elastic search, Hbase, Hive.
* Migrated Existing MapReduce programs to Spark Models using python.
* Leveraged spark(Pyspark) to manipulate unstructured data and apply text mining on user's table utilization data.
* Use to collect, conduct and curate real time data by moving it from source to destination by Apache Nifi.
* Experience in sequence data pre-processing, extraction, model fitting and validation using ML pipelines.
* Designed Data Quality Framework to perform schema validation and data profiling on spark(Pyspark).
* Responsible for debug, optimization of Hive scripts and also implementing DE duplication logic in Hive using a rank key function (UDF).
* Experienced in writing Hive validation scripts that are used in validation framework (for daily analysis through graphs and presented to business users).
* Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting.
* Developed workflow in Oozie to automate the tasks of loading data into HDFS and pre-processing with Pig and Hive.
* Used Pig as ETL tool to do transformations, joins and some pre-aggregations before storing the data into HDFS.
* Imported all the customer specific personal data to Hadoop using Sqoop component from various relational databases like Netezza and Oracle.
* Used Impala to read, write and query the Hadoop data in HDFS and Hbase.
* Experienced in running query using Impala and used BI tools to run ad-hoc queries directly on Hadoop.
* Worked with BI teams in generating the reports and designing ETL workflows on Tableau.
* Set up Logstash ELK stack and created dashboards to monitor marketing emails status.
* Develop testing scripts in Python and prepare test procedures, analyze test results data and suggest improvements of the system and software.
* Experience in streaming log data using Flume and data analytics using Hive.
* Developed a data pipeline using Kafka and Storm to store data into HDFS.
* Extracted the data from RDBMS (Oracle, MySQL & Teradata) to HDFS using Sqoop.

**Environment:** Hadoop, MapReduce, HDFS, Pig, Hive QL, HBase, Zookeeper, Oozie, Flume, Impala, Cloudera 5.2, Storm, MySQL, UNIX Shell Scripting, Nifi, Tableau, Python, Spark.

**DELL, Austin, TX Dec'13– Feb’15**

**Role: Hadoop Developer**

**Responsibilities:**

* Responsible for loading the customer’s data and event logs from Oracle database, Teradata into HDFS.
* Developed MapReduce jobs for Log Analysis, Analytics and to generate reports for the number of activities created on particular day.
* Involved in implementing complex MapReduce programs to perform joins on the Map side using distributed cache in Java.
* Used Spark DataFrame API over Cloudera platform to perform analytics on hive data.
* Implemented Apache Spark data processing project to handle data from RDBMS and streaming sources.
* Involved in creating Hive tables, loading with data and writing Hive queries, which will run internally in MapReduce way.
* Complete end to end design and development of Apache Nifi flow which acts as the agent between middleware team and EBI team and executes all the actions mentioned above.
* Implemented different analytical algorithms using MapReduce programs to apply on top of HDFS data.
* Implemented Partitioning, Dynamic Partitions and Buckets in HIVE for efficient data access.
* Utilized Apache Hadoop ecosystem tools like HDFS, Hive and Pig for large datasets analysis.
* Developed Pig and Hive UDF's to analyze the complex data to find specific user behavior.
* Designed a data warehouse using Hive and created Partitioned tables in Hive.
* Experienced in using Pig for data cleansing and developed Pig Latin scripts to extract the data from web server output files to load into HDFS.
* Load and transform large data sets of structured, semi structured and unstructured data.
* Designed and modified database tables and used HBASE Queries to insert and fetch data from tables.
* Developed code in Scala & Python to use Map Reduce framework by Hadoop streaming.
* Worked on Hive by creating external and internal tables, loading it with data and written Hive queries.
* Created HBase tables to store data from different sources.
* Worked with various Hadoop file formats, including Text Files, Sequence File, RCFile.
* Create/Modify shell scripts for scheduling various data cleansing scripts and ETL load process.
* Configured Zookeeper for Cluster co-ordination services.

**Environment:** Hadoop, MapReduce, HDFS, Pig, Hive, HBase, Zookeeper, Nifi, Oozie, Horton works, Python, Scala, Oracle 11g/10g, Windows NT,UNIX Shell Scripting.

**Intel, Santa Clara, CA Oct’12- Nov'13**

**Role: Sr. Java/J2EE Developer**

**Responsibilities:**

* Involved in Full Life Cycle Development in Distributed Environment using Java and J2EE framework.
* Designed the application by implementing Struts Framework based on MVC Architecture.
* Designed and developed the front end using JSP, HTML and JavaScript and JQuery.
* Implemented the Web Service client for the login authentication, credit reports and applicant information Apache Axis 2 Web Service.
* Extensively worked on User Interface for few modules using JSPs, JavaScript and Ajax.
* Developed framework for data processing using Design patterns, Java, XML.
* Used the lightweight container of the Spring Framework to provide architectural flexibility for Inversion of Controller(IOC).
* Used Hibernate ORM framework with Spring framework for data persistence and transaction management.
* Designed and developed Session beans to implement the Business logic.
* Developed EJB components that are deployed on Web logic Application Server.
* Written unit tests using JUnit Framework and Logging is done using Log4J Framework.
* Designed and developed various configuration files for Hibernate mappings.
* Designed and documented REST/HTTP APIs, including JSON data formats and API versioning strategy.
* Developed Web Services for sending and getting data from different applications using SOAP messages.
* Actively involved in code reviews and bug fixing.
* Applied CSS(Cascading style Sheets) for entire site for standardization of the site.
* Assisted QA Team in defining and implementing a defect resolution process including defect priority, and severity.

**Environment:** Java 5.0, Struts, Spring 2.0, Hibernate 3.2, WebLogic7.0, Eclipse 3.3, Oracle 10g, JUnit 4.2,Maven, Windows XP, HTML, CSS, JavaScript, and XML.

**Info com- Hyderabad/India Oct'10 - Sep'12**

**Role: Java Developer**

**Responsibilities:**

* Involved in the analysis & design of the application using Rational Rose.
* Developed and implemented the MVC Architectural pattern using Struts Framework including JSP, Servlets, EJB, From Bean and Action classes.
* Object Oriented Analysis and Design using UML include development of class diagrams, Sequence diagrams and State diagrams and implemented these diagrams in Microsoft Visio.
* Designed and created Java Objects, JSP pages, JSF, JavaBeans and Servlets to achieve various business functionalities. Created validation methods using JavaScript and Backing Beans.
* Involved in writing client side validations using JavaScript, CSS.
* Involved in the design of the Referential Data Service module to interface with various databases using JDBC.
* Designed and developed the UI using Struts view components JSP, HTML, CSS and JavaScript.
* Implemented Java/J2EE Design Patterns like Business Delegate and Data Transfer Object (DTO), Data Access Object and service locator.
* Developed JMS API using J2EE package.
* Used Oracle as Database and used Toad for queries execution and also involved in writing SQL scripts, PL/SQL code for procedures and functions.
* Involved in designing test plans, test cases and overall Unit testing of the system.
* Prepared documentation and participated in preparing user's manual for the application.

**Environment:** Java, Rational Rose, Spring 2.0, Hibernate 3.2, WebLogic, Eclipse, JSP, Windows XP, HTML, CSS, JavaScript, and XML.

**Knot Solutions- Hyderabad/ India Jun’09 -Sep '10**

**Role: Programmer Analyst/ SQL Developer**

**Responsibilities:**

* Gathered requirements from client, analyzed and prepared the requirement specification document.
* Working Knowledge of relational database concepts like entities, attributes, relationships, data integrity.
* Involved in writing SQL queries.
* Generated database SQL Scripts and deployed databases including installation and configuration.
* Developed SQL Scripts to Insert/Update and Delete data in MS SQL database tables.
* Experience in writing PL/SQL and in developing and implementing Stored Procedures, Packages and Triggers.
* Experience on modeling principles, database design and programming, creating E-R diagrams and data relationships to design a database.
* Build data connection to the database using MS SQL Server.
* Used different joins, sub queries and nested querying.
* Worked with different sources such as Oracle, SQL and Flat files.
* Worked on project to extract data from xml file to SQL table and generate data file reporting using SQL Server 2008.

**Environment:** My SQL, SQL Server 2008(SSRS & SSIS), Visual studio 2000/2005, MS Excel.

**Academics**: Bachelor of Engineering in Computer Science and Engineering-JNTU – Hyderabad, India -2008.