|  |  |
| --- | --- |
| Contact Number | (601)-675-2555 |
| Alternate Contact Number | (601)-675-2555 |
| Current location of the candidate (City/State) | Jacksonville, FL |
| Email Address | [Krishtej31@gmail.com](mailto:Krishtej31@gmail.com) |
| Skype ID | [Krishtej31@gmail.com](mailto:Krishtej31@gmail.com) |

**KrishnaTeja Sadhu**

**Professional Summary:**

* **9 years** of total IT experience which includes **Java** Application Development, **Database** Management & on **Big Data technologies using Hadoop Ecosystem.**
* **Over 4 years** of experience in working with different Hadoop ecosystem components such as **HDFS, MapReduce, HBase, Spark, Yarn, Kafka, Zookeeper, PIG, HIVE, Sqoop, Storm, Oozie, Impala** and **Flume**.
* Expertise in writing **Map-Reduce Jobs in Java** for processing large sets of structured, semi-structured and store them in HDFS.
* Good Knowledge on **Spark** framework for batch and real time data processing.
* Proficient in developing data transformation and other analytical applications in **Spark**, **Spark-SQL** using **Scala** programming language.
* Profound experience in creating real time data streaming solutions using **Apache Spark/Spark Streaming, Kafka.**
* Good knowledge on Apache Storm for processing **real time streaming data.**
* Performed importing and exporting data to/ from traditional RDBMS into HDFS and Hive using **Sqoop**
* Experience in developing **Custom UDFs** using Javato use in **Pig** and **Hive**.
* Have a great understanding on concepts like **Partitions**, **Bucketing** in **Hive** to optimize the performance and designed both Managed and External tables in Hive.
* Experience in job workflow scheduling and monitoring tools like **Oozie.**
* Proficient in designing and querying the NoSQL databases like **HBase, Cassandra.**
* Experience in designing and developing tables in HBase and storing aggregated data from Hive Table.
* Worked with **Java API, Rest API** to handle real time analytics on **HBase** data.
* Experience on streaming data using **Apache Flume** and **Kafka**
* Worked on various Hadoop distributions like **Cloudera, Hortonworks.**
* Hands on experience in installing, configuring, and using Hadoop ecosystem components like Hadoop **Map Reduce (MR), HDFS, HBase, Hive, Sqoop, Pig, and Flume.**
* Have worked on a 40 nodes live Hadoop cluster running on Cloudera CDH4.
* Familiar working on various formats of files like **delimited** text files, **click stream** log files, **Apache log** files, **Avro** files, **JSON** files, **XML** Files.
* Has good understanding of various compression techniques used in Hadoop processing like **Gzip**, **SNAPPY**, **LZO.**
* Experience in writing build scripts using **Maven, ANT** and **Gradle**.
* Experienced with different scripting language like **Python** and **shell scripts.**
* Good knowledge of BI Tools like **Tableau, Qlikview** and **Plotly**.
* Worked on version control tools like **Bit-Bucket**, **GIT, SVN**, **CVS**.
* Good experience in working with cloud environment like **Amazon Web Services (AWS) EMR, EC2, ES and S3.**
* Experienced in working with Machine learning libraries(spark-MLlib) and implementing ML algorithms for clustering, regression filtering and dimensional reduction.
* Experienced in writing spark jobs for Data clustering and data processing using spark-MLlib and cluster algorithms as per functional requirements along with data scientists.
* Proficient in creating **data ingestion pipelines, data transformations, data management and real time streaming** at an enterprise level.
* Involved in requirement analysis, reviews and working sessions to understand the requirements and system design.
* Have a good experience working in **AGILE** scrum process.
* Good experience with **SQL, PL/SQL** and database concepts.
* Good Experience as a Java Developer in Web, Client/Server technologies using **Java, J2EE, Servlets, JSP, EJB, JDBC.**
* Used Scala to write the code for all the use cases in spark
* Proactive and well organized with effective time management skills and problem solving skills
* Good Inter personnel skills and ability to work as part of a team.
* Exceptional ability to learn and master new technologies and to deliver outputs in short deadlines.

**Education:**

* Bachelors in Computer Science and Engineering, SCSVMV University, TN, India. 2008

**Technical Skills:**

|  |  |
| --- | --- |
| **Big Data Ecosystems** | Hadoop, Map Reduce, Spark, HDFS, HBase, Pig, Hive, Sqoop, Oozie, Storm, Kafka and Flume. |
| **Spark Streaming Technologies** | Spark Streaming, Spark SQL, RDD, Storm |
| **Scripting Languages** | Python and shell scripting |
| **Programming Languages** | C, Java, Scala, python, SQL, PL/SQL |
| **JAVA/J2EE Technologies:** | Servlets, JSP, JSF, JDBC, Java Beans, RMI & Web services (SOAP, RESTful). |
| **Databases** | Oracle 12c, MySQL and Microsoft SQL Server |
| **NoSQL Technologies** | Cassandra, MongoDB, HBase |
| **Web Technologies** | HTML, CSS, XML, JavaScript, jQuery, AngularJS, AJAX, SOAP, and REST. |
| **Development Tools** | Eclipse, Net Beans, SVN, Git, ANT, Maven, Jenkins, Bamboo, SOAP UI, QC, Jira, Bugzilla |
| **Methodologies** | Agile /Scrum , Waterfall |
| **Operating Systems:** | Windows X/7/8/10, UNIX, LINUX. |

**Professional Experience**

**Citi Bank, Jacksonville, FL Jan 2016 - Till Date**

**Hadoop /Spark Developer**

Citi Bank is one of the leading banking sectors in America providing Banking, Mortgage, Credit Card services and many commercial financial servicesto its clients. Scope of this project is to develop spark applications to access the data of their customers and developed several customer usage patterns.

**Responsibilities:**

* Worked on cloud platform which was built with a scalable distributed data solution using **Hadoop** on a 40**-node cluster using AWS cloud** to run analysis on **25+ Terabytes** of customer usage data.
* Involved in creating end to end **spark applications** for various data transformation activities.
* Performed series of ingestion jobs using Sqoop, Kafka and custom Input adapter to move data from various sources to HDFS.
* Developed a data pipeline using Kafka, **Spark** and Hive to ingest, transform and analyzing customer behavioral data.
* Created **Spark** jobs to see trends in data usage by users.
* Real time streaming the data using **Spark** with **Kafka**
* Implemented **Spark** using **Scala** and utilizing **Data frames** and **Spark SQL** API for faster processing of data.
* Used **Spark** for interactive queries, processing of streaming data and integration with popular **NoSQL** database for huge volume of data.
* Real time streaming the data using **Spark** with **Kafka**
* Configured **Kafka** to read and write messages from external programs.
* Converted Hive queries into Spark transformations using Spark RDDs.
* Exploring with the Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, Pair RDD's, Spark YARN.
* Implemented various machine learning techniques like Random Forest, K-Means, Logistic Regression for predictions and pattern identification using Spark-MLib.
* Worked extensively with importing metadata into Hive using Scala and migrated existing tables and applications to work on **AWS cloud(S3).**
* Data processing and hosting Real time data streaming using **Spark** with **Kafka.**
* Collecting and aggregating large amounts of log data using **Flume** and staging data in **HDFS** for further analysis.
* Involved in performing the Linear Regression using Scala API and Spark
* Working on custom portions and configuration to cluster nodes with the **Zookeeper.**
* Created **HBase** tables to store user data and Written automated **HBase** test cases for data quality check coming from different portfolios Data processing using **SPARK**.
* Developed Hive scripts in **spark SQL** to de-normalize and aggregating the data.
* Working on NoSQL databases like **HBase, Cassandra** for **internal data storage** and **test validations**.
* Used **Hive** to analyze the partitioned and bucketed data and compute various metrics for reporting.
* Developed Hive scripts in **Hive QL** to de-normalize and aggregate the data.
* Created **HBase** tables and column families to store the user event data.
* Written automated **HBase** test cases for data quality checks using **HBase** command line tools.
* Scheduled and executed workflows in **Oozie** to run Hive and Spark jobs.
* Have a hands-on experience working with **Airflow (**replace the work of oozie**).**
* Used to monitor and manage the Hadoop cluster using Cloudera Manager.
* Developed interactive shell scripts for scheduling various data cleansing and data loading process.

**Environment:**

Hadoop Distribution of Cloudera, AWS Service (Clusters on cloud), HDFS, Map Reduce, Sqoop, Spark, SparkSQL, Hive, HBase, LINUX, Java, Scala, Eclipse, JIRA, GIT, Oracle, Toad 9.6, Tableau, UNIX Shell Scripting, Putty.

**Liberty Mutual Insurance, MA June 2014 – Dec 2015**

**Hadoop Developer**

Liberty Mutual Insurance Company is a 100 billion capital worth insurance company. It is well known for its Auto, Home and life insurance products in USA. This company offers financial services (fixed, indexed, and variable annuities) and life insurance products through banks, financial planners, regional dealers and independent agents. The BRP Solution Center provides services and project support for a wide variety of business partners and applications.

**Responsibilities:**

* Understanding business needs, analyzing functional specifications and map those to develop and designing end to end data transformation pipelines.
* Created Hive Tables, loaded data from Teradata using **Sqoop**.
* Importing and exporting data into **HDFS** from **Relational Databases** and vice versa using **Sqoop.**
* Worked extensively with importing metadata into **Hive** and migrated existing tables and applications to work on **Hive** and **AWS cloud**.
* Developed **MR** jobs for cleaning, validating and transforming the data.
* Implemented **Hive Generic UDF’s** to incorporate business logic into Hive Queries.
* Extensively worked on **HiveQL**, join operations, writing custom **UDF’s** and having good experience in optimizing Hive Queries.
* Written **Hive** jobs to parse the logs and structure them in tabular format to facilitate effective querying on the log data.
* Worked on debugging, performance tuning **PIG** and **HIVE** scripts by understanding the joins, group and aggregation between them.
* Wrote **Pig** scripts to transform raw data from several data sources.
* Experience of using different columnar file formats like **RCFile**, **Parquet and ORC formats.**
* Monitored workload, job performance and capacity planning using **Cloudera Manager**.
* Involved in build applications using **Maven** and integrated with Continuous Integration servers like Jenkins to build jobs.
* Performing data migration from Legacy Databases RDBMS to HDFS using **Sqoop**.
* Got good experience with **NOSQL** database **HBase, MongoDB and** Hybrid implementations.
* Hands on experience on whole ETL (Extract Transformation & Load) process.
* Worked with BI teams in generating the reports and designing **ETL workflows** on Tableau
* ETL development to normalize this data and publish it in **IMPALA**
* Used **IMPALA** to analyze data ingested into HBase and compute various metrics for reporting on the dashboard.
* Designed and Maintained **Oozie workflows** to manage the flow of jobs in the cluster.
* Worked with the testing teams to fix bugs and ensure smooth and error-free code.
* Involved in **Agile** methodologies, daily Scrum meetings, Sprint planning.

**Environment**: Hadoop, MapReduce, HDFS, Hive, Java, SQL, Cloudera Manager, Pig, Sqoop, Oozie, HBase, Zookeeper, PL/SQL, MySQL, DB2, Teradata

**Xerox, Jersey City, NJ May 2013 – Jun 2014**

**Hadoop Developer**

Xerox is the enterprise for business process and document management. Its administration, innovation, and aptitude empower working environments from small organization to substantial worldwide undertaking improved the way work completes so they work more adequately. The idea of Integrated Data Ware House Project is to ingest data from different multiple sources to Hadoop Data Lake, perform transformations on it as per business requirements and exporting the data to external systems.

**Responsibilities:**

* Involved in design low level design documents for functional and nonfunctional requirements.
* Developed MapReduce programs in Java and **Sqoop** the data from ORACLE database.
* Involved in writing complex MapReduce programs that work with different file formats like **Text, Sequence, Xml and Avro.**
* Have solid understanding of **REST** architecture style and its application to well performing web sites for global usage.
* Experienced in creating data model and implement queries to handle time series data with **HBase** data.
* Integrated **HBase** with **Map Reduce** to move bulk amount of data into HBase.
* Imported data using **Sqoop** to load data from **Oracle** to HDFS on regular basis.
* Developed multiple MapReduce jobs in **Pig** for data cleaning and processing.
* Developed **Pig scripts** and **UDFs** extensively for Value added Processing (VAPS).
* Used **Avro** Storage to use in **Pig Latin** to load and Store data.
* Installed and configured **Hive** and wrote **Hive UDFs** to implement the business requirement.
* Developed **Pig scripts** to convert the data from **Avro** to Text file format.
* Worked on partitioning the **HIVE** table and running the scripts in parallel to reduce the run time of the scripts.
* Written **Hive queries** for data analysis to meet the business requirements.
* Creating **Hive tables** and working on them using **HiveQL**.
* Developing Scripts and Batch Job to schedule various Hadoop Program using **Oozie**.
* Importing and exporting data into **HDFS** from **Oracle Database** and vice versa using **Sqoop.**
* Involved in preparation of docs like Functional Specification document and Deployment Instruction documents.
* Fix defects as needed during the QA phase, support QA testing, troubleshoot defects and identify the source of defects.

**Environment:** Hadoop, Map Reduce, HDFS, Hive, Pig, HBase, Linux, XML, Java, Eclipse, Oracle, JIRA, , GIT Hub, CDH4. Rest API.

**Fifth Third Bank, Charlotte, NC Nov 2012 – Apr 2013**

**Sr. Java Developer**

Online Account Opening (OAO) portal which will be used by applicants to open new Deposit Accounts (Checking, Saving, Money Market) And Customer Care representatives and branch personnel shall be provided access to allow them to complete an application on behalf of a customer, as well as search for and finish incomplete applications.

**Responsibilities:**

* Participated in requirement gathering and converting the requirements into technical specifications.
* Developed UI using HTML, JavaScript, and JSP, and developed Business Logic and Interfacing components using Business Objects, XML, and JDBC.
* Designed user-interface and checking validations using JavaScript.
* Involved in design of JSP’s and Servlets for navigation among the modules.
* Developed various EJBs for handling business logic and data manipulations from database.
* Managed connectivity using JDBC for querying/inserting & data management including triggers and stored procedures.
* Developed SQL queries and Stored Procedures using PL/SQL to retrieve and insert into multiple database schemas.
* Developed the XML Schema and Web services for the data maintenance and structures Wrote test cases in JUnit for unit testing of classes.
* Provided Technical support for production environments resolving the issues, analysing the defects, providing and implementing the solution defects.
* Built and deployed Java applications into multiple UNIX based environments and produced both unit and functional test results along with release notes.
* Developed the presentation layer using CSS and HTML taken from bootstrap to develop for browsers.

**Environment:** Java, Spring, JSP, Hibernate, XML, HTML, JavaScript, JDBC, CSS, SOAP Web services, JIRA, SVN.

**Magnaquest - Hyderabad, India July 2010 – Aug 2012**

**Java Developer**

Magnaquest Technologies Limited (Magnaquest) is an enterprise solutions and products company, with offices in India, Malaysia and USA, and partners spread all over the globe. The project is to build a retail applications related by using Java and J2EE.

**Responsibilities:**

* Assisted in designing and programming for the system, which includes development of Process Flow Diagram, Entity Relationship Diagram, Data Flow Diagram and Database Design.
* Designed front end components using **JSF**.
* Involved in developing Java APIs, which communicates with the Java Beans.
* Implemented **MVC** architecture using Java, Custom and JSTL tag libraries.
* Involved in development of **POJO** classes and writing **Hibernate** query language (HQL) queries.
* Implemented MVC architecture and DAO design pattern for maximum abstraction of the application and code reusability.
* Created Stored Procedures using **SQL/PL-SQL** for data modification.
* Used **XML, XSL** for Data presentation, Report generation and customer feedback documents.
* Used **Java Beans** to automate the generation of Dynamic Reports and for customer transactions.
* Developed **JUnit** test cases for regression testing and integrated with **ANT** build.
* Implemented Logging framework using **Log4J**.
* Involved in code review and documentation review of technical artifacts.

**Environment:** Java, Spring, JSP, Hibernate, XML, HTML, JavaScript, JDBC, CSS, SOAP Web services, CVS, JIRA .

**Apollo Microsystems, Hyderabad June 2008 – Jun 2010 Java Developer**

Developed the front-end featuring rich web interface using HTML4/5, CSS2/3, JavaScript, jQuery and implemented the Responsive, browser compatible, table-less and w3c compliant standards

**Responsibilities:**

* Worked on gathering the requirements and for the development of the code for the entire project.
* Designed UI using JSP pages, HTML, CSS.
* Implemented the object-oriented concepts in programming.
* Implemented validations using java script.
* Implemented the application using JAVA and JDBC connectivity was used to integrate with the database.
* Used various design patterns like Singleton, Facade, Command, Factory, DAO.
* Utilize Struts (MVC2) framework and developed JSP pages, Action Servlets and XML based

action-mapping files for web tier.

**Environment:** Eclipse, JIRA, SVN version control, MYSQL, HTML, CSS, JAVASCRIPT, JAVA, JQuery, MVC architecture.