**SANDEEP**[**krsbl2914@gmail.com**](mailto:krsbl2914@gmail.com) **+1 (347) 875-7725  
Senior Hadoop/Java Developer**



****

**PROFESSIONAL SUMMARY**

* Around **8 years** of IT experience in software development and support with experiences in developing strategic methods for deploying **Big Data Technologies** to efficiently solve Big Data processing requirements.
* Good hands on experience in Hadoop eco system components **HDFS/Map reduce, Hbase,Yarn, Pig, Spark, Sqoop, Spark SQL, Spark Streaming and Hive.**
* Experience in Installing, maintaining and configuring **Hadoop Cluster.**
* Efficient in processing **large sets** of structured, semi-structured and unstructured data and supporting systems application architecture.
* Capable of creating and monitoring Hadoop clusters on **Amazon EC2,Hortonworks Data Platform 2.1 & 2.2, VM,CDH3, CDH4Cloudera Manager on Linux, Ubuntu OS** etc.
* Experience in working with structured data using **Hive**, **Hive UDFs**, join operations, partitions, **bucketing and internal/external tables.**
* Hands on experience in different stages of big data applications such as data ingestion, data analytics and data visualization.
* Good experience on **Scala Programming language** and **Spark Core.**
* Hands on experience in Importing and Exporting the Data using SQOOP from HDFS to Relational Database management system.
* Experience in analyzing data with **Hive**.
* Well Experience in using **Hive Query** Language and data Analytics using **Hive Query Language**.
* Expertise in Managing and scheduling batch Jobs on a Hadoop Cluster using **Oozie**.
* Strong knowledge in **NOSQL** column oriented databases like **HBase, Cassandra, MongoDB**, and its integration with Hadoop cluster.
* Experience onsetting cluster in **Amazon EC2,Amazon EMR, Amazon RDS S3 Buckets, Dynamo DB, and RedShift**.
* Worked on **Oozie** and **Zookeeper** for managing **Hadoop** jobs.
* Capable to analyze data, interpret results, and convey findings in a concise and professional manner.
* Promote/ Simulatecomplete approach including request analysis, creating/pulling dataset, report creation and implementation and providing final analysis to the requestor.
* Very Good understanding of **SQL, ETL** and **Data Warehousing Technologies.**
* Strong experience in RDBMS technologies like MySQL, Oracle and Teradata.
* Strong database skills in **IBM- DB2, Oracle** andProficient in **database development**, including **Constraints, Indexes, Views, Stored Procedures, Triggers** and **Cursors**.
* Experience in developing Web-Services module for integration using **SOAP** and **REST**.
* Flexible with Unix/Linux and Windows Environments, working with Operating Systems like **Centos 5/6, Ubuntu 13/14, Cosmos**.
* Experience in working with structured data using Hive, Hive UDFs, join operations, partitions, bucketing and internal/external tables.
* Used Maven Deployment Descriptor Setting up build environment by writing Maven build XML, taking build, configuring and deploying of the application in all the servers
* Experience in build scripts using **Maven** and do continuous integrations systems like **Jenkins**.
* Intensive work experience in developing enterprise solutions using Java, J2EE, Servlets, JSP, JDBC, Struts, spring, Hibernate, JavaBeans, JSF, MVC.
* Proficient knowledge on **java virtual machines (JVM)** and multithreaded processing.
* Good understanding of **SQL, ETL** and **Data Warehousing Technologies.**
* Experience in working with job scheduler like **Autosys** and **Maestro**.
* Strong in databases like **Sybase, DB2, Oracle, MS SQL,Clickstream**.
* Loaded the dataset into **Hive** for **ETL** operation.
* Proficient in using various **IDE**s like RAD, **Eclipse**.
* Strong understanding of **Agile Scrum** and **Waterfall SDLC methodologies**.
* Excellent problem-solving analytical, communication, presentation and interpersonal skills that help me be a core member of any team.
* Strong communication, collaboration & team building skills with proficiency at grasping new technical concepts quickly and utilizing them in a productive manner.
* Experienced in provided training to team members as new per the project requirement.
* Experienced in creating **Product Documentation**&**Presentations**.

**EDUCATIONAL QUALIFICATION:**

* Bachelor of Engineering from Jawaharlal Nehru Technological University of Hyderabad, India.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Hadoop/Big Data Technologies** | HDFS, Map Reduce, Sqoop, Flume, Pig, Hive, Oozie, impala, Spark, Zookeeper and Cloudera Manager. |
| **NO SQL Database** | HBase, Cassandra |
| **Monitoring and Reporting** | Tableau, Custom shell scripts, |
| **Hadoop Distribution** | Horton Works, Cloudera, MapR |
| **Build Tools** | Maven, SQL Developer |
| **Programming & Scripting** | JAVA, C, SQL, Shell Scripting, Python, Scala |
| **Java Technologies** | Servlets, JavaBeans, JDBC, Spring, Hibernate, SOAP/Rest services |
| **Databases** | Oracle, MY SQL, MS SQL server, Teradata |
| **Web Dev. Technologies** | HTML, XML, JSON, CSS, JQUERY, JavaScript, angular JS |
| **Version Control** | SVN, CVS, GIT |
| **Operating Systems** | Linux, Unix, Mac OS-X, Cen OS, Windows10, Windows 8, Windows 7, Windows Server 2008/2003 |

**PROFESSIONAL EXPERIENCE:**

**Ani labs Inc.Pleasanton, CA Dec 2015 - Till Date**

**Senior Hadoop Developer**

**Description:** Ani labs provides services in healthcare, finance, insurance, telecom, and manufacturing. While dealing with healthcare, maintained database related to insurer, policyholder, subscriber, claims, flexible spending accounts and master savings accounts details using Hadoop. Enabled Online Renewals and capture request for Endorsements.This project mainly deals with pulling the data from various structured and unstructured data sources into Hadoop platform and standardized all data through a series of master data management processes to reach client goal to become efficient and effective within different geographic location in attracting more customers and improve the user experience which helps in developing the market.

**Responsibilities:**

* Installed/Configured/Maintained Apache Hadoop clusters for Analytics, application development and Hadoop tools like **Hive, HSQL Pig, HBase,OLAP, Zookeeper, Avro, parquet** and **Sqoop** on Linux **ARCH**.
* Wrote the shell scripts to monitor the health check of Hadoop daemon services and respond accordingly to any warning or failure conditions.
* Having experience working with **Devops**.
* Installed and configured **Hadoop, MapReduce, HDFS** (Hadoop Distributed File System), developed multiple MapReduce jobs in java for data cleaning.
* Having experience in doing structured modelling on unstructured data models.
* Developed data pipeline using **Flume, Sqoop, Pig** and **Java MapReduce** to ingest customer behavioral data and financial histories into HDFS for analysis.
* Involved in collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis.
* Worked on installing cluster, commissioning & decommissioning of Data Nodes, Name Node recovery, capacity planning, and slots configuration.
* Developed **PIG Latin** scripts to extract the data from the web server output files to load into HDFS.
* Worked on **Hortonworks Data Platform (HDP).**
* Worked with **SPLUNK** to analyze and visualize data.
* Worked on **Mesos cluster** and **Marathon.**
* Used Pig as **ETL** tool to do transformations, event joins and some pre-aggregations before storing the data onto **HDFS**.
* Worked with Orchestration tools like **Airflow**.
* Write test cases, analyze and reporting test results to product teams.
* Good experience on **Clojure, Kafka** and **Storm.**
* Worked with **AWS data pipeline**.
* Worked with **Elastic Search, Postgres, Apache NIFI.**
* Hadoop workflow management using **Oozie, Azkaban, Hamake**.
* Responsible for developing data pipeline using **Azure HDInsight, flume, Sqoop** and **pig** to extract the data from weblogs and store in HDFS.
* Installed **Oozie** workflow engine to run multiple **Hive** and **Pig** Jobs, used Sqoop to import and export data from HDFS to RDBMS and vice-versa for visualization and to generate reports.
* Involved in migration of **ETL** processes from Oracle to Hive to test the easy data manipulation.
* Worked in functional, system, and regression testing activities with agile methodology.
* Worked on Python plugin on **MySQL** workbench to upload CSV files.
* Used Hive to analyze the partitioned and bucketed data and compute various metrics for reporting.
* Worked with HDFS Storage Formats like **Avro, Orc**.
* Worked with **Accumulo** to Modify server side Key Value pairs.
* Working experience with **shiny** and **R.**
* Working experience with**Vertica, QilkSense, QilkView** and **SAP BOE.**
* Worked with NoSQL databases like **HBase, Cassandra, DynamoDB**
* Worked with **AWS** based data ingestion and transformations.
* Good experience with **Python, Pig, Sqoop, Oozie, Hadoop Streaming, Hive and Phoenix.**
* Worked on importing and exporting data from **Oracle** and **DB2** into **HDFS** and **HIVE using Sqoop**.
* Responsible for building scalable distributed data solutions using Hadoop like cluster maintenance, adding and removing cluster nodes, Cluster Monitoring, and troubleshooting, Manage and review data backups and log files.
* Developed several new MapReduce programs to analyze and transform the data to uncover insights into the customer usage patterns.
* Worked extensively with importing metadata into Hive using Sqoop and migrated existing tables and applications to work on **Hive**.
* Extract, Load, and Transfer data through **Talend**.
* Responsible for running Hadoop streaming jobs to process terabytes of xml's data, utilized cluster co-ordination services through **Zookeeper**.
* Extensive experience in using the MOM with Active MQ**, Apache storm, Apache Spark & Kafka Maven,** and **Zookeeper.**
* Worked on the core and Spark SQL modules of Spark extensively.
* Worked on Descriptive statistics Using **R**.
* Developed Kafka producer and consumers**, HBase clients, Spark,shark, Streams** and **Hadoop MapReduce** jobs along with components on **HDFS, Hive**.
* Strong Working experience in **snowflake, Clickstream.**
* Worked on Hadoop **EMC Greenplum, Gemstone, Gemfire**.
* Analyzed the SQL scripts and designed the solution to implement using **PySpark**.
* Experience using **Spark** with **Neo4J** where acquiring the interrelated graphical information of the insurer and to query the data from the stored graphs.
* Experience in writing batch processing huge Scala programs.
* Load and transform large sets of structured, semi structured, and unstructured data using **Hadoop/Big Data** concepts.
* Responsible for creating Hive External tables and loaded the data in to tables and query data using HQL.
* Handled importing data from various data sources, performed transformations using **Hive, MapReduce**, and loaded data into HDFS.

**Environment:Hadoop Cluster, HDFS, Hive, Pig, Sqoop,OLAP, data modelling, Linux, Hadoop Map Reduce, HBase, Shell Scripting, MongoDB, and Cassandra, Apache Spark,Neo4J**.

**ACT Inc., Iowa City, IA Apr2015– Nov2015**

**Big Data Software Developer**

**Description:**It’s a fact that more than 1.8 million graduates have taken the ACT each year making ACT the leading US college admissions testing company. ACT also serve millions of job seekers, businesses, government agencies, and policymakers around the world. The purpose of the project is to store terabytes of log information generated by the websiteand extract meaningful information out of it. The solution is based on the open source Big Data s/w Hadoop. Which in turn includes getting the raw html data from the websites, Process the html to obtain product and pricing information, extract various reports out of the product pricing information and Export the information for further processing.

**Responsibilities:**

* Performed benchmarking of HDFS and Resource manager using Test DFSIO and Tera Sort.
* Worked on **SQOOP** to import data from various relational data sources.
* Working with Flume in bringing click stream data from front facing application logs
* Worked on strategizing **SQOOP** jobs to parallelize data loads from source systems
* Participated in providing inputs for design of the ingestion patterns.
* Participated in strategizing loads without impacting front facing applications.
* Worked in agile environment using **Jira, Git.**
* Worked on design on **Hive, ANSI** data store to store the data from various data sources.
* Involved in brainstorming sessions for sizing the Hadoop cluster.
* Involved in providing inputs to analyst team for functional testing.
* Worked with source system load testing teams to perform loads while ingestion jobs are in progress.
* Worked with Continuous Integration and related tools (i.e. **Jenkins, Maven**).
* Worked on performing data standardization using **PIG scripts**.
* Worked with query engines **Tez, Apache Phoenix.**
* Worked on installation and configuration Horton works cluster ground up.
* Managed various groups for users with different queue configurations.
* Worked on building analytical data stores for data science team’s model development.
* Worked on design and development of **Oozie** works flows to perform orchestration of **PIG** and **HIVE** jobs**.**
* Worked on performance tuning of **HIVE** queries with partitioning and bucketing process.
* Worked on the core and Spark SQL modules of Spark extensively.
* Developed **Kafka** producer and consumers**, HBase** clients, **Spark,** and **Hadoop MapReduce** jobs along with components on **HDFS, Hive**.
* Worked with big data tools like **Apache Phoenix, Apache Kylin, Atscale, Apache Hue**.
* Worked with securities like **Knox, Ranger, Atlas.**
* Worked with **BI Concepts-Dataguru,Talend.**
* **Worked with** Source Code Management Tools GitHUB, Clearcase SVN, CVS,
* Working experience with Testing tools **JUNIT / SOAPUI.**
* Experienced in analyzing the SQL scripts and designed the solution to implement using **PySpark**.
* Worked with Code Quality Governance related tools (**Sonar, PMD, FindBugs, Emma, Cobertura**,
* Analyzed the **SQL scripts** and designed the solution to implement using **PySpark**.

**Environment:Hadoop, HDFS, Map Reduce, Flume, Pig, Sqoop, Hive, Pig, Sqoop, Oozie, Ganglia, HBase, Shell Scripting, Apache Spark**.

**Accolite, Madison, WI Oct 2013– Dec 2014**

**Hadoop Developer**

Accolite is a global IT Services company recognized for providing cutting edge information technology services with a proven track record of being totally client focused,This project is mainly for the replat forming of the current existing system which is running on Web Harvest a third-party JAR and in MySQL DB to a new cloud solution technology called Hadoop which can able to process large data sets (i.e. Tera bytes and Peta bytes of data) to meet the client requirements with the increasing competition from the retailers.In the finance domain, worked for claiming credit for capturing settlement details. Gets Reports on matched, un- matched and disputed credits.Data is imported from Teradata to Hadoop, pre-processing of the data is completed in Hadoop.Provided Online Premium Calculator for Nonregistered/Registered Users Provided Online Customer Support like Chat, Agent Locators, Branch Locators, FAQs, Best Plan Selector, to increase the likelihood of sale.

**Responsibilities:**

* Worked on Distributed/Cloud Computing (**Map Reduce/Hadoop, Hive, Pig, HBase, Sqoop, Spark AVRO, Zookeeper** etc.), Cloudera distributed Hadoop (CDH4).
* Installed and configured Hadoop **MapReduce, HDFS**, developed multiple Map Reduce jobs in java for data cleaning and processing.
* Involved in installing Hadoop Ecosystem components.
* Importing and exporting data into **HDFS, Pig, Hive and HBase** using **SQOOP**.
* Responsible to manage data coming from different sources.
* Flume and from relational database management systems using **SQOOP**.
* Responsible to manage data coming from different data sources.
* Involved in gathering the requirements, designing, development and testing.
* Worked on loading and transformation of large sets of structured, semi structured data into Hadoop system.
* Developed simple and complex **MapReduce** programs in Java for Data Analysis.
* Load data from various data sources into HDFS using **Flume**.
* Developed the **Pig UDF'S** to pre-process the data for analysis.
* Worked on Hue interface for querying the data.
* Created Hive tables to store the processed results in a tabular format.
* Developed Hive Scripts for implementing dynamic Partitions.
* Developed Pig scripts for data analysis and extended its functionality by developing custom UDF's.
* Extensive knowledge on **PIG** scripts using bags and tuples.
* Experience in managing and reviewing Hadoop log files.
* Developed workflow in **Oozie** to automate the tasks of loading the data into HDFS and pre-processing with Pig.
* Exported analyzed data to relational databases using **SQOOP** for visualization to generate reports for the BI team.

**Environment:Hadoop (CDH4), UNIX, Eclipse, HDFS, Java, MapReduce, Apache Pig, Hive, HBase, Oozie, SQOOP and MySQL.**

**Google India Pvt Ltd., Hyderabad, India. Feb 2009 – Sep 2013**

**Senior System Analyst/ JAVA Developer**

**Description:** Google India Private Limited operates as an online search engine. The company provides access to online information. The project was about implementing the Geo Atlas and Geo Pushpin tool undergoing the process Google GTL (Ground Truth Local) for devolving Maps on google. Enables user for easy identification and access of locations effectively and efficiently. Additional Functionality like Geo Local, Sprouts, and Mapmaker was developed to this system for accessing the status of business locations. It is implemented on a java platform. It also serves as a platform for users to use as vehicles for social networking, video engagement, mailing and blogging.

**Responsibilities:**

* Used Web Sphere for developing use cases, sequence diagrams and preliminary class diagrams for the system in **UML**.
* Interacting with business users to identify information needs and initiate process changes. Major part in project planning and scoping documents. Adopt waterfall methodology.
* Perform large scale data analysis and modeling to identify opportunities for improvement based on the impact and feedback.
* Provide professional quality project artifacts, including Business requirement documents (BRD’s), requirement plans, models, tractability matrix, use cases and issue logs.
* Preparing user ‘as-is’ workflow and ‘to-be’ business process.
* Worked with QA testing team, creating test plans and test cases and actively participated in user training for user Acceptance testing (UAT) with business users.
* Extensively used Web Sphere Studio Application Developer for building, testing, and deploying **Pushpin Tool**.
* Real-time experience in database management, BA, and software development life cycle.
* Gained knowledge on analyzing stastical data.
* Creating standard operating procedures (SOP’s). Performed SQL queries on database using SQL server.
* Hands on experience on training and mentoring. Used IBM COGNOS for reports.
* Used **Spring Framework** based on **(MVC) Model View Controller**, designed GUI screens by using HTML, JSP.
* Developed the presentation layer and GUI framework in **HTML, JSP** and Client-Side validations were done.
* Involved in Java code, which generated XML document, which in turn used XSLT to translate the content into **HTML** to present to **GUI**.
* Implemented XQuery and XPath for querying and node selection based on the client input XML files to create Java Objects.
* Used Web Sphere to develop the Entity Beans where transaction persistence is required and JDBC was used to connect to the **MySQL** database.
* Developed user interface using the JSP pages and **DHTML** to design the dynamic **HTML** pages.
* Developed Session Beans on Web Sphere for the transactions in the application.
* Utilized **WSAD** to create **JSP**, Servlets, and EJB that pulled information from a DB2 database and sent to a front-end GUI for end users.
* In the database end, responsibilities included creation of tables, triggers, stored procedures, sub-queries, joins, integrity constraints and views.
* Worked on MQ Series with J2EE technologies (**EJB, Java Mail, JMS**, etc.) on Web Sphere server.

**Environment:  Java, EJB, IBM Web Sphere Application server, Spring, JSP, Servlets, JUnit, JDBC, XML, XSLT, CSS, DOM, HTML, MySQL, JavaScript, Oracle, UML, Clear Case, ANT**.