Kartik Girdhar

Hadoop Administrator

[Girdhar.k87@gmail.com](mailto:kgirdhar461@gmail.com), 832-975-9002

**SUMMARY**

* Over 7years of experience with emphasis on Big Data Technologies, administration, Development and Design of Java based enterprise applications.
* Excellent understanding / knowledge of Hadoop architecture and various components such as HDFS, Job Tracker, Task Tracker, NameNode, Data Node and MapReduce programming paradigm.
* Experience in installation, configuration, supporting and managing Hadoop Clusters using Apache, Cloudera (CDH3, CDH4, CDH5) distributions and on Amazon web services (AWS).
* Hands-on experience on major components in Hadoop Ecosystem including Hive, HBase, HBase & Hive Integration, Sqoop, Flume& knowledge of Mapper/Reduce/HDFS Framework.
* Set up standards and processes for Hadoop based application design and implementation.
* Worked on NoSQL databases including HBase, Cassandra and MongoDB.
* Experience on Horton works and Cloudera Hadoop environments.
* Talend administrator with hands on Big data ( Hadoop ) with Cloudera framework
* Setting up data in AWS using S3 bucket and configuring instance backups to S3 bucket.
* Good experience in analysis using Pig and Hiveand understanding of SQOOP and Puppet.
* Expertise in database performance tuningdata modeling.
* Good Experience in Talend DI Administration, Talend Data Quality and Talend Data Mapping
* Experience in designing, installing and configuring complete Hadoop ecosystem (components such as HDFS, Map reduce, pig, hive, Oozie, flume, zookeeper).
* Experience in managing the cluster resources by implementing fair scheduler and capacity scheduler.
* Experienced in developing MapReduceprograms using Apache Hadoop for working with Big Data.
* Experience in tools like puppet to automate Hadoop installation, configuration and monitoring.
* Used the Spark - Cassandra Connector to load data to and from Cassandra.
* Experience increating databases, users, tables, triggers, macros, views, stored procedures, functions, Packages, joins and hash indexes in Teradata database.
* Experience in developing ETL process using Hive, Pig, Sqoop and Map-Reduce Framework.
* Involved in log file management where the logs greater than 7 days old were removed from log folder and loaded into HDFS and stored for 3 months.
* Experienced in using Talend database components, File components and processing components based up on requirements.
* Expertise in development support activities including installation, configuration and successful deployment of changes across all environments.
* Loading the data into EMR from various sources S3 process it using Hive Scripts
* Familiarity and experience with data warehousing and ETL tools.
* Good working Knowledge in OOA&OOD using UML and designing use cases.
* Good understanding of Scrum methodologies, Test Driven Development and continuous integration.
* Experience in production support and application support by fixing bugs.
* Used HP Quality Center for logging test cases and defects.
* Major strengths are familiarity with multiple software systems, ability to learn quickly new technologies, adapt to new environments, self-motivated, team player, focused adaptive and quick learner with excellent interpersonal, technical and communication skills.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| Big Data | Hadoop, Hive, Sqoop, Pig, Puppet, Ambary, HBase, MongoDB, Cassandra, Power Pivot, Defamer, Pentaho, spark,Flume, SolrCloud ,Flume , Impala |
| Operating Systems | Windows, Ubuntu, Red Hat Linux, Linux, UNIX |
| Project Management | Plan View, MS-Project |
| Programming&Scripting Languages | Java, SQL, Unix Shell Scripting, C ,Python# |
| Modeling Tools | UML, Rational Rose |
| IDE/GUI | Eclipse |
| Framework | Struts, Hibernate |
| Database | MS-SQL, Oracle, MS-Access |
| Middleware | Web Sphere, TIBCO |
| ETL | Informatica, Pentaho |
| Business Intelligence | OBIEE, Business Objects |
| Testing | Quality Center, Win Runner, Load Runner, QTP |

**PROFESSIONAL EXPERIENCE**

**The Capital Group Companies, Irvine, CA March 2015 to Present**

**Hadoop Admin**

* Worked on setting up Hadoop cluster for the Production Environment.
* Install Hadoop2/ Yarn, spark, scala IDE, JAVA JRE on three machines. Configure these machines as a cluster, and set one Name node and two Data nodes
* Developed data pipeline using Flume, Sqoop, Pig and Java map reduce to ingest customer behavioral data and financial histories into HDFS for analysis.
* Implemented AWS solutions using EC2, S3 and load balancers.
* Installed application on AWS EC2 instances and also configured the storage on S3 buckets.
* Storing and loading the data from HDFS to Amazon S3 and backing up the Namespace data.
* Worked on Hadoop clusters capacity planning and management.
* Monitored and Debugged Hadoop jobs/Applications running in production.
* Setup and monitored Cloudera CDH4 cluster with Hadoop2/ YARN running to read data from the Cluster
* Involved in implementing High Availability and automatic failover infrastructure to overcome single point of failure for Namenode utilizing zookeeper services.
* Experienced in using Talend Big Data components to create connections to various third-party tools used for transferring, storing or analyzing big data, such as Sqoop, MongoDB and BigQuery to quickly load, extract, transform and process large and diverse data sets.
* Expert in processing bulk amount data into data warehouse using complex SQL and Talend components.
* Troubleshooted, Managed and reviewed data backups, Manage & review Hadoop log files.
* Worked with data delivery teams to setup new Hadoop users. This job includes setting up Linux users, setting up Kerberos principals and testing HDFS, Hive.
* Involved in collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis.
* Performed maintenance, monitoring, deployments, and upgrades across infrastructure that supports all our Hadoop clusters.
* Used Ganglia to Monitor and Nagios to send alerts about the cluster around the clock.
* Created concurrent access for hive tables with shared and exclusive locking that can be enabled in hive with the help of Zookeeper implementation in the cluster.
* Worked on importing and exporting data from Oracle and DB2 into HDFS and HIVE using Sqoop for analysis, visualization and to generate reports.
* Implemented PySpark and Spark SQL for faster testing and processing of data.
* Developed multiple MapReduce jobs in java for data cleaning.
* Ran many performance tests using the Cassandra-stress tool in order to measure and improve the read and write performance of the cluster.
* Worked on migrating MapReduce programs into PySpark transformation.
* Built wrapper shell scripts to hold Oozieworkflow.
* Experienced in using debug mode of talend to debug a job to fix errors.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Python and Scala.
* Provided ad-hoc queries and data metrics to the Business Users using Hive, Pig.
* Useded Hive to analyze the partitioned and bucketed data and compute various metrics for reporting.
* Wrote the shell scripts to monitor the health check of Hadoop daemon services and respond accordingly to any warning or failure conditions.
* Wrote shell scripts for rolling day-to-day processes and it is automated.
* Document and manage failure/recovery.
* Involved in story-driven agile development methodology and actively participated in daily scrum meetings.

**Environment**: Hadoop, Hadoop 2, MapReduce, Hive, HDFS,Cassandra,PIG, Sqoop, Oozie, Cloudera, Flume, HBase, ZooKeeper, CDH5, MongoDB, Cassandra, Talend ,Oracle, NoSQL and Unix/Linux, Kafka, Amazon web services.

**Baker Hughes , Houston ,Tx April 2013to Feb 2015**

**Hadoop Admin**

* Experienced with Cloudera ,MapR and Horton works distribution of Hadoop
* Analyzed the clients existing Hadoop infrastructure and understanding the performance bottlenecks and provide the performance tuning accordingly
* Experienced with Installing Hadoop in new servers and rebuild existing servers
* Experienced in setting up automated 24x7 on monitoring and escalation infrastructure for Hadoopcluster using Nagios and Ganglia
* Expert in processing bulk amount data into data warehouse using complex SQL and Talendcomponents
* Worked in ETL tools Talend to simplify Map Reduce jobs from the front end.
* Expertise in using Oozie for configuring job flows
* Installed/Configured/Maintained Apache Hadoop clusters for application development and Hadoop tools like Hive, Pig, HBase, Zookeeper and Sqoop.
* Deployed Hadoop Cluster in the following nodes.
* Manage Hadoop clusters : Monitor , maintain , setup.
* Strong troubleshooting and performance tuning skills
* Configured High Availability for Control services like Namenode and Job tracker.
* Performed a upgrade in development environment from CDH 4.2 to CDH 4.6.
* Involved in analyzing system failures, identifying root causes and recommended course of actions.
* Worked on Hive for exposing data for further analysis and for generating transforming files from different analytical formats to text files.
* Worked on analyzing Hadoop cluster and different big data analytic tools including Pig, Hbase database and Sqoop.
* Involved in creating Hadoop streaming jobs using Python.
* Handled importing of data from various data sources, performed transformations using Hive, Pig and Spark and loaded data into HDFS.
* Troubleshooting, Manage and review data backups, Manage & review Hadoop log files.
* Wrote queries to create, alter, insert and delete elements from lists, sets and maps in Datastax Cassandra.
* Developed multiple MapReduce jobs in java for data cleaning and accessing.
* Importing and exporting data into HDFS and Hive using Sqoop
* Expert in Talend job migration and deployment to different environment and successfully scheduled job in TAC.
* Implemented NameNode backup using NFS. This was done for High availability.
* Monitored workload, job performance and capacity planning using Cloudera Manager..
* Managed and reviewed Log files as a part of administration for troubleshooting purposes. Communicate and escalate issues appropriately.
* Documented and managed failure/recovery.
* Implemented Fair schedulers on the Job tracker to share the resources of the Cluster for the Map Reduce jobs given by the users.

**Environment:** Hadoop 1x, Hive, Pig, HBASE, Sqoop, Flume, Zookeeper, Talend ,Pig, HDFS, Ambary, Cassandra, Oracle, CDH4, HPD 2.2

**Universal American, Dallas, TX May 2011 to March 2013**

**Hadoop Developer**

* Exported data from DB2 to HDFS using Sqoop.
* Developed MapReduce jobs using Java API.
* Installed and configured Pig and also wrote Pig Latin scripts.
* Wrote MapReduce jobs using Pig Latin.
* Developed workflow using Oozie for running MapReduce jobs and Hive Queries.
* Worked on Cluster coordination services through Zookeeper.
* Worked on loading log data directly into HDFS using Flume.
* Involved in loading data from LINUX file system to HDFS.
* Responsible for managing data from multiple sources.
* Experienced in running Hadoop streaming jobs to process terabytes of xml format data.
* Responsible to manage data coming from different sources.
* Assisted in exporting analyzed data to relational databases using Sqoop.
* Implemented JMS for asynchronous auditing purposes.
* Created and maintained Technical documentation for launching Cloudera Hadoop Clusters and for executing Hive queries and Pig Scripts
* Experience with CDH distribution and Cloudera Manager to manage and monitor Hadoop clusters
* Experience in defining, designing and developing Java applications, specially using Hadoop [Map/Reduce] by leveraging frameworks such as Cascading and Hive.
* Experience in Develop monitoring and performance metrics for Hadoop clusters.
* Experience in Document designs and procedures for building and managing Hadoop clusters.
* Strong Experience in troubleshooting the operating system, maintaining the cluster issues and also java related bugs.
* Experienced import/export data into HDFS/Hive from relational database and Teradata using Sqoop.
* Successfully loaded files to Hive and HDFS from Mongo DB Solar.
* Experience in Automate deployment, management and self-serve troubleshooting applications.
* Define and evolve existing architecture to scale with growth data volume, users and usage.
* Design and develop JAVA API (Commerce API) which provides functionality to connect to the Cassandra through Java services.
* Responsible for cluster availability and available 24x7 on call support.
* Experience with disaster recovery and business continuity practice in hadoop stack
* Installed and configured Hive and also written Hive UDFs.
* Experience in managing the CVS and migrating into Subversion.
* Experience in managing development time, bug tracking, project releases, development speed, release forecast, scheduling and many more.

**Environment:** Hadoop, HDFS, Hive, Flume, Sqoop, HBase, PIG, MySQL and Ubuntu, Zookeeper, Java (JDK 1.6)

**Varshtech, India March 2009 to Feb 2011**

**Java Developer**

* Created the Database, User, Environment, Activity, and Class diagram for the project (UML).
* Implement the Database using Oracle database engine.
* Designed and developed a fully functional generic n-tiered J2EE application platform the environment was Oracle technology driven. The entire infrastructure application was developed using Oracle JDeveloper in conjunction with Oracle ADF-BC and Oracle ADF- Rich Faces.
* Created an entity object (business rules and policy, validation logic, default value logic, security)
* Created View objects, View Links, Association Objects, Application modules with data validation rules, LOV, dropdown, value defaulting, transaction management features.
* Web application development using J2EE: JSP, Servlets, JDBC, Java Beans, Struts, Ajax, JSF, JSTL, Custom Tags, EJB, JNDI, Hibernate, ANT, JUnit and Apache Log4J, Web Services, Message Queue (MQ).
* Designing GUI prototype using ADF 11G GUI component before finalizing it for development.
* Create Reusable Component (ADF Library and ADF Task Flow)
* Experience using Version controls such as CVS, PVCS, and Rational Clear Case.
* Creating Modules Using Task Flow with Bounded and Unbounded
* Generating WSDL (Web Services) And Create Work Flow Using BPEL
* Handel the AJAX functions (partial trigger, partial Submit, auto Submit)
* Created the Skin for the layout.

**Environment:** Java core, Servlet, JSF, ADF Rich client UI Framework ADF-BC (BC4J) 11g, web services Using Oracle SOA (Bell), Oracle WebLogic.

**Education:** Bachelor’s in Computer Science and Engineering, IP University, India