Sai kiran Gandu

Hadoop Administration/developer

Contact      : 201-688-0542  
Email ID  : [saikiran7149@gmail.com](mailto:saikiran7149@gmail.com)

DOB(MM/DD)   : 04/14

**Professional Abridgment:**

* Over **7+** years of experience in IT industry as **Hadoop Administration/developer** using Cloudera and Hortonworks Distributions.
* A very good planning, time management, prioritization and organizational skills.
* Extremely high attention detail and ability to grasp things quickly.
* Exceptional written, verbal, and interpersonal communication skills.
* Positive, proactive and able to always exercise great judgment.
* Ability to perform business analysis, and desire to improve business processes.
* Ability to effectively work with a variety of organizations, management levels, cultures, and personalities.
* Demonstrated ability to work independently and autonomously.
* A quick learner having experience in working in multi-cultural and diverse background, self-motivation and team oriented and exhibits an honest work ethic with the ability to excel in fast paced, time-sensitive environments.
* Experienced in the complete life cycle including design, system architecture, development, testing, implementation, Migration and maintenance
* Good understanding of Hadoop architecture and different daemons of Hadoop clusters which include Name Node and Data Node, Task Tracker, Job Tracker.
* Experience in data ingestion process by using Sqoop and streaming tool Flume.
* Hands on experience in working with Cloudera CDH4 and CDH5.
* Wrote the MapReduce programs using Java for analyzing Big Data.
* Strong Knowledge in importing streaming data into HDFS using Flume sources, and Flume sinks and transforming the data using Flume interceptors.
* Loaded the dataset into Hive for ETL operation.
* Good knowledge/understanding of NoSQL data bases and hands on work experience in writing applications on NoSQL databases like Cassandra and HBase.

**Skills:**

|  |  |
| --- | --- |
| **Big Data Ecosystems** | HDFS, MapReduce, Hive, Pig, Sqoop, Oozie, Spark, Flume, Kafka and Yarn. |
| **Hadoop distributions** | Apache Hadoop, Yarn, Cloudera, Ambari. |
| **ETL/Data Integration tool** | Talend Open Studio. |
| **Databases** | MySQL, Hbase, Cassandra. |
| **Languages** | C, Java, Python, Scala, Pig Latin, Shell. |
| **BI/Analytical** | DQ Analyzer, Data Flux. |
| **Operating Systems** | Windows, Linux (Ubuntu, Centos), Mac OS. |
| **Other Skills** | Oyster, Microsoft Office Suite. |

**Experience:**

**Client : American Express, Phoenix AZ Aug 2016 – Feb 2017**

**Hadoop Developer**

**Responsibilities:**

* Created and assemble the Foundation layer script to stack the information from social databases

Sql, Oracle, Netezza to Hadoop Hive tables utilizing bash script.

* Used Sqoop to import the verifiable data into hive tables in text and Avro file formats.
* Developed Python Script to transfer the data into target databases Greenplum and Netezza, Hadoop from unified location.
* Implemented partitioning, Bucketing in Hive for the better association of the data.
* Implemented Adhoc, every day, week after week, month to month Batch jobs utilizing Control-M scheduler.
* Implemented Hadoop Hive Cluster using Cloudera Distributed Vendor (CDH4). Planning, installation and configuration of Hadoop cluster and its ecosystem components.
* Implemented Apache kafka for the data ingestion of real time data using Spark.
* Wrote the SQL queries and utilized as a part of scripting to computerize the email ready warnings to business clients at whatever point their information is stacked into target tables.
* Involved to build up the Python script to peruse the JSON objects from the document and store the comes about into TEXT and CSV records.
* Participated in development/implementation of Cloudera impala Hadoop environment
* Implemented Quality check controls for to prepare the information utilizing python/Bash script.
* Involved in the implementation of Hadoop cluster on AWS as a part of POC.
* Managed and reviewed Hadoop log files.
* Supported in setting up QA environment and updating configurations for implementing scripts with Pig and Sqoop.
* Working with Jenkins for any automation builds which are integrated with GIT, Docker as part of infra automation under continuous integration.
* **Environment:** Hadoop, MapReduce, HDFS, Hive, Sqoop, DB2, Docker, SQL, Elasticsearch, Tableau, Linux/Unix, Autosys, IBM Data studio, NDM, Quality Center, Cloudera distribution.

**Client:Target Minneapolis MN Jul 2014 to Aug 2016**

**Role: Hadoop Developer**

**Project: BLAZE - BRMS**

**Responsibilities:**

* Analyze, develop and implementation of Importing and exporting data into HDFS using Hive, Sqoop.
* Designed and developed loading data process from Oracle database into HDFS using Sqoop queries
* Created Hive queries and tables to help LOB identify current credit card trends and take decisions.
* Apply strategies on historical data before promoting them to production.
* Developed Pig scripts to parse the raw data, populate staging tables and store the refined data in partitioned DB2 tables for Business analysis.
* Review Hadoop Log files, code reviews and walkthroughs.
* Developed Pig UDF'S for manipulating the data according to Business Requirements and also worked on developing custom Pig Loaders.
* Participate in Requirement gathering, architecture, design, development, test and implementation meetings.
* Understand business requirements, collaborate with stakeholders, discuss on source-target data mapping, design, signoff and implementations.
* Create implementation plan and Detailed Task Schedules

**Environment:** Hadoop, MapReduce, HDFS, Cloudera, HBase, Hive, Pig, Sqoop, DB2, SQL, Linux, Autosys, IBM Data studio, NDM, Quality Center, Windows & Microsoft Office.

**JC Penny, Plano, TX Jan 2013 – July 2014**

**Hadoop Consultant**

**Responsibilities:**

* Handle the installation and configuration of a Hadoop cluster.
* Monitored health of all the Processes related to Name Node HA, HDFS, YARN, HIVE, HBASE, Sqoop.
* Monitored disk, Memory, Heap, CPU utilization on all Master and Slave machines and took necessary measures to keep the cluster up and running on 24/7 basis.
* Configured Capacity Scheduler to provide service-level agreements for multiple users of a cluster.
* Used Hive and created Hive tables and loaded data from Local file system and HDFS.
* Created Hive internal and external tables defined with appropriate static and dynamic partitions, intended for efficiency.
* Worked on installing cluster, commissioning & decommissioning of data nodes.
* Involved in implementing High Availability and automatic failover infrastructure to overcome single point of failure for Namenode utilizing zookeeper services.
* Involved in Installing Kerberos and Configuring Server, Client Systems to enable Hadoop Security.
* Imported logs from web servers with Flume to ingest the data into HDFS.
* Involved in collecting metrics for Hadoop clusters using Ganglia and Ambari.
* Interacted with developers when we had to deploy new jobs, Jobs throwing exceptions, and data related issues.
* Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig.
* Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team.
* Responsible for cluster maintenance, adding and removing cluster nodes, cluster monitoring and troubleshooting, manage and review data backups, manage and review Hadoop log files.

**Environment:** Hadoop, Sqoop, Flume, Oozie, Mapreduce, Hortonworls, HDFS, Pig, Hive, Hbase, Java, Oracle 10g, MySQL, Ubuntu

**Union Bank of Switzerland (UBS), NJ Nov 2011-Dec 2013**

**Hadoop Administrator**

**Responsibilities:**

* Worked on setting up the Hadoop cluster for the dev, test and prod Environment.
* Worked on pulling the data from oracle databases into the hadoop cluster using the sqoop import.
* Worked with flume to import the log data from the reaper logs, sys log’s into the Hadoop cluster.
* Configured Fair Scheduler to provide fair resources to all the applications across the cluster.
* Worked with application teams to install Hadoop updates, patches, version upgrades as required.
* Involved in Cluster coordination services through Zookeeper.
* Manage and review data backups and log files.
* Configured services with the web interface Hue.
* Implemented the hadoop Namenode HA services to make the Hadoop services highly available.
* Upgrading the hadoop cluster from cdh3 to cdh4.
* Effectively used Sqoop to transfer data from databases (MySql, Oracle) to HDFS, Hive.
* Created Hive Managed and External tables defined with static and dynamic partitions.
* Used Ganglia to monitor the cluster around the clock.
* Supported Data Analysts in running Map Reduce Programs.
* Contribute to the creation and maintenance of system documentation

**Environment:** Java (JDK 1.7), Mapreduce, HDFS, Pig, Hive, Hbase, Flume, Sqoop, Oozie, Cloudera, Unix.

**Virtusa, Hyderabad, India Aug 2009 to Oct 2011**

**Java/Hadoop Developer**

**Responsibilities:**

* Developed simple to complex MapReduce jobs using Hive and Pig.
* Analyzed large data sets by running Hive queries and Pig scripts.
* Job management using Fair scheduler.
* Worked with Sqoop for importing metadata form Database.
* Load and transform large sets of structures, semi structured and unstructured data.
* Responsible to manage data coming from different sources.
* Assisted in exporting analyzed data to relational databases using Sqoop.

**Educational Qualifications:**

* Bachelor of Technology 2005-2009
* *Jawaharlal Nehru technological university, Hyderabad.*