

Vikas D.

Greater New York City Area

Big Data (Hadoop) Engineer

Experience

Big Data Engineer at AT&T Labs, Inc.

Big Data Engineer at AT&T Labs, Inc.

September 2016 - Present (3 years 1 month)

- Hadoop installation, Configuration of multiple nodes using Cloudera platform.
- Installed and configured a Hortonworks HDP 2.2 using Ambari and manually through command line. Cluster maintenance as well as creation and removal of nodes using tools like Ambari, Cloudera Manager Enterprise and other tools.
- Handling the installation and configuration of a Hadoop cluster.
- Building and maintaining scalable data pipelines using the Hadoop ecosystem and other open source components like Hive and HBase.
- Worked on creating comprehensive MongoDB API and Document DB API using Storm into Azure Cosmos DB.
- Involved in developer activities of installation and configuring Hadoop MapReduce, HDFS, Developed multiple MapReduce jobs in java for data cleaning and pre-processing.
- Created instances in AWS as well as migrated data to AWS from data Center using snowball and AWS migration service
- Importing and exporting data into HDFS and Hive using Sqoop.
- Involved in Cluster Level Security, Security of perimeter (Authentication- Cloudera Manager, Active directory and Kerberos) Access (Authorization and permissions- Sentry) Visibility (Audit and Lineage - Navigator) Data (Data Encryption at Rest).
- Hands on experience in AWS Cloud in various AWS services such as Redshift cluster, Route 53 domain configuration.
- On demand secure Amazon EMR launcher with custom spark submit steps using S3 Event, SNS, KMS and Lambda function.
- Extensive knowledge of working on NiFi.
- Migrated an existing on-premises application to AWS.
- Used AWS services like EC2 and S3 for small data sets.
- Used Cloud watch logs to move application logs to S3 and create alarms based on a few exceptions raised by applications.
- Continuous coordination with QA team, production support team and deployment team.
- Implemented test scripts to support test driven development and continuous integration.

- Populated HDFS and Cassandra with huge amounts of data using Apache Kafka.

Hadoop Engineer at Pfizer Pharmaceuticals Inc

March 2013 - June 2016 (3 years 4 months)

- Installed and configured Hadoop Map Reduce, HDFS, developed multiple Map Reduce jobs in java for data cleaning and pre-processing.
- Worked on Installing and configuring the HDP Hortonworks 2.x Clusters in Dev and Production Environments.
- Worked on Capacity planning for the Production Cluster.
- Installed HUE Browser.
- Involved in loading data from UNIX file system to HDFS and creating Hive tables, loading with data and writing hive queries which will run internally in map reduce way.
- Experience in MapR, Cloudera, & EMR Hadoop distributions.
- Create a complete processing engine, based on Hortonworks distribution, enhanced to performance.
- Performed on cluster up gradation in Hadoop for cloudera.
- Ability to Configuring queues in capacity scheduler and taking Snapshot backups for HBase tables.
- Worked on fixing the cluster issues and Configuring High Availability for Name Node in HDP 2.1.
- Involved in Cluster Monitoring backup, restore and troubleshooting activities.
- Currently working as Hadoop administrator in MapR hadoop distribution for 5 clusters ranges from POC clusters to PROD clusters contains more than 1000 nodes.
- Implemented manifest files in puppet for automated orchestration of Hadoop and Cassandra clusters.
- Worked on installing cluster, commissioning & decommissioning of Data Nodes, Name Node recovery, capacity planning, Cassandra and slots configuration.
- Administration of Hbase, Hive, Sqoop, HDFS, and MapR.
- Importing and exporting data from different databases like MySQL, RDBMS into HDFS and HBASE using Sqoop.
- Worked on Configuring Kerberos Authentication in the cluster

Hadoop/Linux Admin at Microland Limited

January 2012 - February 2013 (1 year 2 months)

- Provide technical designs, architecture, Support automation, installation and configuration tasks and upgrades and planning system upgrades of Hadoop cluster.
- Design development and architecture of the Hadoop cluster, map reduce processes, Hbase system.
- Design and develop process framework and support data migration in Hadoop system.
- Involved in installation and configuration of Kerberos security setup on CDH5.5 cluster.
- Involved in upgrading Hadoop Cluster from HDP 1.3 to HDP 2.0.

- Implemented secondary sorting to sort reducer output globally in MapReduce.
- Involved in installation and configuration of LDAP server and integrated with kerberos on cluster.
- Worked with Sentry configuration to provide centralized security to hadoop services.
- Scheduled Oozie workflow engine to run multiple Hive and Pig jobs, which independently run with time and data availability.
- Work with network and Linux system engineers to define optimum network configurations, server hardware and operating system.
- Used Sqoop to import and export data from HDFS to RDBMS and vice-versa.
- Created HBase tables to store various data formats of data coming from different portfolios.
- Monitor critical services and provide on call support to the production team on various issues.
- Assist in Install and configuration of Hive, Pig, Sqoop, Flume, Oozie and HBase on the Hadoop cluster with latest patches.
- Involved in performance tuning of various hadoop ecosystem components like YARN, MRv2.
- Implemented the Kerberos security software to CDH cluster for user level as well as service level to provide strong security to the cluster.
- Troubleshooting, diagnosing, tuning, and solving Hadoop issues.
- Continuous monitoring and managing the Hadoop cluster using Cloudera Manager.

Environment: Hortonworks (HDP 2.2), Ambari, Map Reduce 2.0(Yarn), HDFS, Hive, Hbase, Pig, Oozie, Sqoop, Spark, Flume, Kerberos, Zookeeper, DB2, SQL Server 2014, CentOS, Linux, RHEL 6.x, 7.x.

Vikas D.

Greater New York City Area

Big Data (Hadoop) Engineer

Linked  **Recruiter**