**Surya P**

**PROFESSIONAL SUMMARY:**

* Over 5 Years of IT experience among which 4+ years with excellent knowledge on Cassandra Administration and Development (NoSQL)
* Good experience in Designing, Planning, Administering, Installation, Configuring, Troubleshooting, Performance monitoring and Fine-tuning of Cassandra Cluster.
* Superior knowledge on Cassandra architecture with better understanding of read and write processes including SSTable, Mem-table and Commit log.
* Good Understanding of Distributed Systems and Parallel Processing architecture.
* Excellent knowledge on CQL (Cassandra Query Language), for obtaining the data present in Cassandra by running queries in CQL.
* Experience in Cassandra stress tool for benchmarking Cassandra Cluster.
* Good Knowledge in Cassandra cluster topology and Virtual nodes.
* Experience in installing multi-data center and multi-rack Cassandra cluster.
* Two projects of experience in exporting data into DataStax Cassandra cluster from RDBMS using Java Driver or Sqoop tools.
* Experience in Cassandra data modeling along with managing and scheduling of the data backup and restore operations.
* Experience in Tuning the JVM.
* Good knowledge in implementation of DataStax Java driver to connect, load and retrieve data from Cassandra database.
* Experience in setting up alerts in OpsCenter.
* Excellent knowledge on tired and leveled sized compaction strategies.
* Involved in designing various stages of migrating data from RDBMS to Cassandra.
* Experience in deploying the Cassandra cluster in cloud, on premises, working on the data storage and disaster recovery for Cassandra.
* Experience in development Using Cassandra JAVA driver (2.X).
* Designing data models in Cassandra and working with Cassandra Query Language.
* In depth knowledge in Cassandra read and write paths and internal architecture.
* Implemented multi-data center and multi-rack Cassandra cluster.
* Good experience in design, development, maintenance and support of Big Data Analytics using Hadoop Ecosystem components like HDFS, Hive, Pig, Hbase, Sqoop, Flume, Zookeeper, MapReduce, and Oozie.
* Strong working experience with ingestion, storage, querying, processing and analysis of big data.
* Experience in installation, configuration, supporting and managing Hadoop clusters.
* Worked in Multiple Environments in installation and configuration of Hadoop Clusters
* Experience with SQL, PL/SQL and database concepts.
* Good understanding of NoSQL databases.
* Hands on experience with Amazon Web Services (AWS), Amazon EC2 and EMR
* Experience on creating databases, tables and views in HIVE, IMPALA
* Experience in working with different data sources like Flat files, XML files and Databases.
* Experience in database design, entity relationships, database analysis, programming SQL, stored procedure's PL/ SQL, packages and triggers in Oracle and MongoDB on Unix / Linux.
* Experience in various phases of Software Development Life Cycle (Analysis, Requirements gathering, Designing) with expertise in documenting various requirement specifications, functional specifications, Test Plans, Source to Target mappings, SQL Joins.
* Worked on different operating systems like UNIX/Linux, Windows XP and Windows 2K.
* Goal oriented self-starter, quick learner, team player and proficient in handling multiple projects simultaneously.

**EDUCATION:**

* Bachelor of Technology in Electronics and Communication Engineering, JNTU.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Cassandra | Cassandra DataStax Enterprise Cassandra, Open source Cassandra, Cluster Management Tools OpsCenter, ccm, Kafka, Spark, Stress tool, Sqoop, Cloudera Manager |
| Server Automation Tools | Chef, Puppet |
| Databases | Dynamo DB, Microsoft SQL Server, MySQL, Oracle |
| Languages | C, Java, Python, JavaScript, HTML, CSS |
| Operating Systems | Linux (Red Hat, CentOS, Ubuntu), Windows |
| Version Control, related Tools | Git, SVN, gerrit, jenkins |
| Bug Tracking Tools | QC, Bugzilla |
| Job Scheduling | Autosys, Cron |

**PROFESSIONAL EXPERIENCE:**

**Sr. Cassandra Adminstrator / Developer**

**ProV International, Tampa, FL Dec 2016 - Present**

This project involves installing Cassandra on AWS. Objective is to reduce total operational costs by replacing heavy licensing costs of products (Cost Optimization) and high-end machines for coherence with distributed open source platform and linearly scalable no SPOF database on cloud infrastructure. Migrating other data model to Cassandra to achieve and reduce the SLA from 10 minutes to 2 seconds and improve user experience by using spark and Cassandra.

**Responsibilities:**

* Maintained a Multi-Datacenter Cassandra cluster.
* Experience in performance tuning a Cassandra cluster to optimize writes and reads.
* Involved in the process of data modeling Cassandra Schema.
* Installed and Configured DataStax OpsCenter for Cassandra Cluster maintenance and alerts.
* Benchmarked Cassandra cluster based on the expected traffic for the use case and optimized for low latency.
* Built Cassandra Cluster on both the physical machines and on AWS.
* Automated Cassandra Builds/installation/monitoring etc.
* Involved in requirements gathering and capacity planning for multi data center (four) Cassandra cluster.
* Administered and maintained multi rack Cassandra cluster using OpsCenter based on the use case implemented consistency level for reads and writes.
* Automated and deployed Cassandra environments using Chief recipes.
* Optimized the Cassandra cluster by making changes in Cassandra configuration file and Linux OS configurations.
* Setup, upgrade and maintain Cassandra DSE clusters.
* Tune databases and provide design changes and support stress tests to proactively fix problems.
* Working as Cassandra Admin (Datastax DSE-DevOps-NoSQL DB) on 39 node cluster.
* Created tables in HIVE and one-time data load into HIVE Environment.
* Converting existing SQL Queries/Stored Procedures into Hive environment.
* Used Apache KAFKA for Asynchronous exchange of information between Different Business Applications.
* Worked with Docker Containers for Testing.
* Administration and maintenance of the cluster using OpsCenter, Devcenter, Linux, Node tool etc.
* Data migration from Teradata to Cassandra using Teradata Fexport & Cassandra loader.
* Installing, configuring Cassandra on AWS platform.
* Developed Spark code using Scala and Spark-SQL/Streaming for faster testing and processing of data.
* Developed Spark Streaming applications for Real Time Processing.
* Working on OpsCenter(monitoring), Devcenter and Node tool.
* Worked on Cassandra Data modelling, NoSQL Architecture, DSE Cassandra Database administration. Key space creation, Table creation, Secondary and Solr index creation, User creation & access administration.
* Working closely with Datastax to resolve issues on cluster using ticketing mechanism.
* Node tool repair, Compaction, Secondary index issues resolution.
* Query tuning & performance tuning on cluster& suggesting best practice for developers.
* Working closely with Cassandra loading activity on history load and incremental loads from Teradata and Oracle Databases and resolving loading issues and tuning the loader for optimal performance.

**Environment:**Cassandra 2.1, AWS, Tera Data, SOL, Datastax 4.7, DevCenter, Cqlsh, OpsCenter, Shell Scripting, HIVE, APACHE KAFKA, DOCKER

**Sr. Cassandra Admin / Developer**

**Singular Analysts Inc., Dallas, TX Aug 2015 – Nov 2016**

The main goal of this project is to create a central, integrated data platform which can handle Big data. Data from different systems is pushed to HDFS, Transformations are applied on the data available in HDFS using pig, hive and Map Reduce. Here, we Used different Hadoop ecosystem tools to integrate the large volume of data. Extract load it for Business Use.

**Responsibilities:**

* Responsible for building scalable distributed data solutions using Datastax Cassandra.
* Involved in business requirement gathering and proof of concept creation.
* Created data models in CQL for customer data.
* Involved in Hardware installation and capacity planning for cluster setup.
* Involved in the hardware decisions like CPU, RAM and disk types and quantities.
* Used the Spark - Cassandra Connector to load data to and from Cassandra.
* Worked with the Data architect and the Linux admin team to set up, configure, initialize and troubleshoot an experimental cluster of 12 nodes with 3 TB of RAM and 60 TB of disk space.
* Ran many performance tests using the Cassandra-stress tool in order to measure and improve the read and write performance of the cluster.
* Wrote and modified YAML scripts to set the configuration properties like node addresses, replication factors, client storage space, memTable size and flush times etc.
* Used the Datastax Opscenter for maintenance operations and Keyspace and table management.
* Loaded and transformed large sets of structured, semi structured and unstructured data in various formats like text, zip, XML, YAML and JSON.
* Created data-models for customer data using the Cassandra Query Language.
* Used collections like lists, sets and maps to create data models highly optimized for reads and writes.
* Created User defined types to store specialized data structures in Cassandra.
* Developed PIG UDFs for manipulating the data and extracting useful information according to Business
* Requirements and implemented them using the Datastax Pig functionality.
* Responsible for creating Hive tables based on business requirements
* Implemented advanced procedures like text analytics and processing using the in-memory computing capabilities like Spark.
* Enhanced and optimized product Spark code to aggregate, group and run data mining tasks using the Spark framework.
* Implemented the clustering algorithms in Mahout to cluster consumer by location of purchase and general category of purchase in order to create specialized and targeted credit and foreign exchange products.
* Implemented a distributed messaging queue to integrate with Cassandra using Apache Kafka and Zookeeper.
* Involved in a POC to implement a failsafe distributed data storage and computation system using Apache YARN.
* Involved in the implementation of a POC using the OpenStack Cloud Computing Framework.
* Tuned and recorded performance of Cassandra clusters by altering the JVM parameters like -Xmx and -Xms. Changed garbage collection cycles to place them in tune with backups/compactions so as to mitigate disk contention.
* Queried and analyzed data from Datastax Cassandra for quick searching, sorting and grouping.
* Implemented Partitioning, Dynamic Partitions and Buckets in HIVE for efficient data access.
* Participated in NoSQL database integration and implementation.
* Exported the analyzed data into relational databases using Sqoop for visualization and to generate reports.
* Gathered the business requirements from the Business Partners and Subject Matter Experts like Data Scientists.

**Environment:** Apache Hadoop 2.2.0, Cloudera 4.5, HDP 1.2, Apache Kafka, Cassandra, MapReduce, Spark, Hive 0.12, Pig 0.11, HBase, Linux, XML.

**Cassandra Admin**

**Adepto Geoinformatics Pvt. Ltd., India Nov 2013 – May 2015**

I worked as a Cassandra Administrator involved in administrating the cluster, building and securing the similar environments globally, measuring gaps in the Architecture for simple and effective solution. It was focused on automating the Catalogue data that powers the Product Information Page and Product Listing Page as an initial step to Big Data. These are the two main web pages that encounter maximum hit per second, we could measure 50K hit/sec way ahead of the expected SLA with just a five-node cluster with proper sizing and configuration.

**Responsibilities:**

* Responsible for the build out, day-to-day managing and supporting Cassandra clusters.
* Configure backup, alerts, repairs and monitoring of Cassandra clusters using Opscenter.
* Troubleshoot the performance issues.
* Troubleshoot read/write latency and timeout issues using nodetool cfstats, tpstats, and cfhistogram.
* Involved in migrating data from Oracle to Cassandra.
* Created the upgrade plans for DSE upgrades.
* Designed and developed an API for rider's preferences with all CRUD capabilities.
* Installed Datastax Cassandra 4.5.1 in Production, Testing environments as per best practices.
* Installed Datastax OpsCenter for monitoring purposes.
* Administered, monitored and maintained multi data-center Cassandra cluster using OpsCenter and Nagios in production.
* Involved closely with developers for choosing right compaction strategies and consistency levels.
* Involved in Cassandra Cluster environment administration that includes commissioning and decommissioning nodes, cluster capacity planning, performance tuning, cluster Monitoring and Troubleshooting.
* Performed daily administrative tasks of Cluster health check, balancing, and name node metadata backup.
* Performed back up, added libraries and jars successfully migrate from the existing infrastructure to latest releases.

**Environment:** Datastax 4.7, Cassandra 2.1, DevCenter, Cqlsh, OpsCenter, Shell Scripting, Oracle11g, Eclipse, SQL, windows7, Log4J, GIT, AWS.

**Big data Consultant/ Administrator**

**Value labs, India Jan 2013 – Oct 2013**

Valuelabs is a perioperative data management system that creates leading data edge capture and file sharing platform which seamlessly integrate with all EMR systems on the market to give new level of efficiency, accuracy in healthcare industry.

**Responsibilities:**

* Installed and Configured Apache Hadoop clusters for application development and Hadoop tools like Hive, Pig, Oozie, Zookeeper, Hbase, Flume and Sqoop.
* Implemented multiple Map Reduce Jobs in java for data cleansing and pre-processing.
* Responsible for Cluster maintenance, adding and removing cluster nodes, Cluster Monitoring
* Hands on experience in writing Linux/Unix Shellscripting and python Scripting
* Developed a data cleaner for IMDB data files in python for a research project.
* Written python scripts for internal testing which pushes the data reading form a file into Kafka queue which in turn is consumed by the Storm application.
* Troubleshooting, manage and review data backups and log files.
* Responsible to manage data coming from various sources.
* Worked with python to create UDF's and used them as part of Pig Scripts, and used python for creating graphs for data analysis.
* Involved in importing the real-time data to Hadoop using Kafka and implemented the Oozie job for daily imports.
* Managed and scheduled Jobs on a Hadoop cluster.
* REST APIs were developed in python for Video streaming interfaces.
* Involved in defining job flows, managing and reviewing log files.
* Developed analytics data store in python, MongoDB for data analysis
* Installed Oozie workflow engine to run multiple Map Reduce, HiveHQL and Pig jobs.
* Implemented a script to transmit sysprin information from Oracle to Hbase using Sqoop.
* Participated in requirement gathering form the Experts and Business Partners and converting the requirements into technical specifications.
* Created Hive tables and impala to store the processed results in a tabular format.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala.
* Learning Supported and maintained HDFS architecture
* Collaborated with application teams to install operating system and Hadoop updates, patches, version upgrades when required.

**Environment:** Hadoop, MapReduce, HDFS, Java, SQOOP, Flume, Kafka, LINUX, OOZIE, Python, Pig, AWS, Scala, ETL, MySQL, JIRA, Hive, Jenkins, HBASE, Oracle.