**SOMASEKHAR K**

**Hadoop Developer**

**Mob: +919908660003**

**Email: somasekarak@gmail.com**

**Summary**

* Having 3+ years of IT experience, which includes experience in **Hadoop Development.**
* Hands on experience in installation, configuration, supporting and managing Hadoop Clusters using Apache Cloudera, Hortonworks distributions.
* Knowledge in Hadoop And it’s Eco System; HDFS, Spark, Pig, Hive, HBase, Oozie, Flume, Shell Script, Power BI and Sqoop.
* Experience in using Pig, Hive, Scoop, Oozie, Ambari and Cloudera Manager.
* Good Understanding with Hadoop architecture and its components like HDFS, Map Reduce, Job Tracker, Task Tracker, Name Node and Data Node.
* Good Knowledge in analyzing Log files for Hadoop and eco system services and finding root cause.
* Analyzed data on Cassandra cluster by running queries for searching, sorting and grouping
* Experience in analyzing data using HIVEQL and Pig Latin.
* Extending Hive and Pig core functionality by writing custom UDFs.
* Experience in analyzing data using HiveQL, Pig Latin and Map Reduce.
* Expert in importing and exporting data into HDFS and Hive using Sqoop.
* Experienced in loading data to Hive partitions and creating buckets in Hive.
* Experienced in configuring Flume to stream data into HDFS.
* Comprehensive understanding of client’s needs for the proper management and delivery of projects
* Knowledge of job workflow scheduling, monitoring and of administrative tasks such as installing Hadoop, Commissioning and decommissioning, and its ecosystem components such as Flume, Oozie, Hive and Pig.
* Knowledge in Software Development Life Cycle (Requirements Analysis, Design, Development, Testing, Deployment and Support).
* Experience in writing Complex SQL Queries involving multiple tables inner and outer joins.
* Experience in optimizing the queries by creating various clustered, non-clustered indexes and indexed views using and data modeling concepts.

**Experience**

* Working as Technical Specialist with Cargill in Bangalore, Since NOV 2017.
* Worked as Storage architect with DELL-EMC in Bangalore, From JAN 2016 to Nov 2017

**EduCATION qualifications**

* B.Tech from, Jawaharlal Nehru Technological University.

**Skill Summary**

* Worked on Cluster Setup using Cloudera & Hortonworks.
* Worked Hadoop Eco Systems like HDFS, Sqoop, Flume, Hive, PIG, Oozie, HBase, and Kafka.
* Worked on tools like Ambari, Nagios, SQL Plus, SQL Developer.
* Knowledge on Machine Learning, Python ,Scala ,Spark and R.
* Worked on the Dell EMC block devices like Install, configure and manage a highly available and scalable virtual infrastructure with VCE components.

**Cetifications**

* VCE Certified Converged Infrastructure Associate (VCE-CIA).
* EMCDSA (DELL-EMC Data Science Associate) Certification.

**project #1**

Client : Cargill

Role : Hadoop Developer

Team Size : 8

**Description:**

Cargill is world’s largest Food ingredients, agricultural, risk management, and commodity trading and processing company. The objective of the project was to monitor and maintain Hadoop cluster in Production.

**RESPONSIBILITIES**

* Responsible to manage data coming from different sources, involved in HDFS maintenance, and loading of structured and unstructured data.
* Developed data pipeline using Flume, Sqoop, Pig and Java MapReduce to ingest behavioral data into HDFS for analysis.
* Responsible for importing log files from various sources into HDFS using Flume.
* Imported large sets of data using Sqoop to load data from MySQL to HDFS on regular basis.
* Created Partitions, Buckets based on State to further process using Bucket based Hive joins.
* Compressed the data to fit into the infrastructure with minimal hardware requirements.
* Used Recursive queries using several levels of joins.
* Moved Relational Database data using Sqoop into Hive Dynamic partition tables using staging tables.
* Optimizing the Hive queries using Partitioning, Bucketing techniques with ACID properties for controlling the data distribution.
* Running Hive queries through different engines like Spark, MapReduce and TEZ.
* Worked with NoSQL database HBase to create tables and store data.
* Worked on custom Pig Loaders and storage classes to work with variety of data formats such as JSON and XML file formats.
* Used Pig as ETL tool to do Transformations, even joins and some pre-aggregations before storing the data onto HDFS.
* Set up Kerberos Authentication system using Ranger.
* Used Oozie workflow engine and coordinators to manage timed interdependent Hadoop jobs and to automate several types of Hadoop jobs such as Java map-reduce Hive and Sqoop.
* Implemented Various Optimization techniques in hive on TEZ for effective result sets.

**project #2**

Client : Dell - EMC

Role : Storage Analyst

Team Size : 12

**Description:**

Dell EMC: Converged Platforms & Solutions Division is focused on the next evolution of IT, the next generation data center and the future of cloud computing ,Dell EMC accelerates the adoption of converged infrastructure and cloud-based computing solutions that reduce IT costs while improving time to market.

**RESPONSIBILITIES**

* Install, configure and manage a highly available and scalable virtual infrastructure with VCE components.
* Supporting for storage products and hardware for problem resolution to EMC CX Celerra VNX File and Block Series Symmetrix VMAX and Vblock Series .
* Working on SRDF Mirrorview and Time Finder Snapview operations .
* Switched Fabric technology by Brocade and Cisco switches.
* Provide planning, installing, configuring, and upgrading support for vRealize Automation & vRealize Orchestrato.
* Experience in supporting Converged/Hyperconverged products such as vSAN, vBlock, VxRail nodes, VxRack, Dell Management & TOR Switches, vRealize, NSX-V, NSX-T, EMC Elastic Cloud Storage (ECS)
* Upgrading of VMware vRealize Automation 6.x, VMware vRealize Orchestrator 6.x.
* Virtualization of infrastructure, building of packages using VMware Pivotal Container Service with Kubernetes.
* Create mechanisms/architectures that enable rapid recovery, repair and cleanup of faulty migrations with good understanding of fault tolerance and failure domains.
* Identify opportunities to deliver self service capability for the most common infrastructure and application management tasks.
* This is for deep level SUPPORT role at an expert level. Coverage as per client time zone and contract/SOW requirements.
* Migration tools such as Open replicator VPLEX .