**Sundeep**

**Cloud Engineer**

**Ph.:469-598-1261 Email-Upendar.k@tekleaders.com**

**PROFESSIONAL SUMMARY:**

* 9 years of professional experience in the IT industry, involved in developing, implementing, configuring, and managing Big Data Components on the **Google Cloud Platform**.
* Certified Google Cloud Architect with expertise in cloud infrastructure, security design and deployment, cloud migration and integration, cloud automation and DevOps, as well as performance monitoring and optimization.
* Expertise in leveraging Google Cloud Platform (GCP) services to drive business transformation and innovation. Strong understanding of cloud architecture principles, **DevOps practices,** and **infrastructure automation.**
* Develop and implement complex data architectures using **BigQuery**, **Cloud Dataflow**, **Dataproc**, and **Cloud Pub/Sub** to handle large-scale data processing and analytics.
* Design and deploy **Virtual Machines** (VMs) on Compute Engine, selecting appropriate machine types, operating systems, and storage options.
* Deploying **GKE** (Google Kubernetes Engine) workloads using Helm charts from Git offers a streamlined and efficient approach to managing applications on **Google Kubernetes** clusters.
* **Use Dataproc** (Apache Spark and Hadoop) to design, implement, and manage data pipelines that handle batch and streaming data.
* Proficient in infrastructure-as-code using **Terraform**, automation with **Ansible**, and Continuous Integration (CI/CD) pipelines for rapid and reliable deployments.
* Automated ETL jobs and data operations using **Google Cloud Functions.**
* Expertise in creating data pipelines and notification services with the integration of **Cloud Pub/Sub** and Google **Cloud Functions**.
* Seamless data transfer between **GCS** and **Snowflake**, allowing to leverage Snowflake's powerful data warehousing and analytics capabilities with data stored in GCS.
* Strong technical background in Hadoop Administration (BIG DATA Administration) with Hadoop, Multi-Node cluster setup, Name node (**HDFS**) High Availability, HDFS Federation, **MapReduce** and **YARN** Framework, Apache Oozie workflow scheduler, Hadoop cluster implementation, Cloudera Manager, Ambari.
* Proficient in using **Zookeeper** and **ZKFC** to manage and configure NameNode failure scenarios, ensuring high availability and effective cluster service coordination.
* Hands-on experience in installing, configuring, and using Hadoop ecosystem components like **Hadoop Map Reduce (MR), HDFS, HBase, Oozie, Hive, Spark, Kerberos, Ranger and Ranger KMS.**
* Utilize **Kafka** and **Nifi** to gain real-time streaming data into one of the source systems for EPM.
* Expertise in managing data processing clusters to optimize performance, reduce data loss, mitigate hardware failures, ensure reliable data availability, and seamlessly scale by adding new worker nodes.
* Excellent command in creating **backups & recovery** and **disaster recovery** procedures and implementing backup and recovery strategies for off-line and on-line backups.
* Maintained **MySQL/PostgreSQL** databases, including creating and managing user access, and set up automated backup processes for cluster metadata databases using cron jobs.
* Setting up the cluster size and memory size based on the requirements, queues to run the jobs based on the capacities and node labels and enabling them for the job queues to run.
* With excellent communication and interpersonal skills, a strong track record in troubleshooting, problem-solving, coordination, and analysis, and the ability to adapt and prioritize effectively, I consistently deliver successful outcomes.

**Certifications :** Google Cloud Architect

**Technical Proficiency:**

**Cloud Platforms :** Google Cloud Platform (GCP)

**GCP Services :** Compute Engine, App Engine, Kubernetes Engine, Cloud Functions, Cloud Run, Cloud
 Storage, Cloud SQL, BigQuery, Dataflow, Pub/Sub, Composer, Cloud Monitoring, Cloud

 Logging, IAM

**DevOps Tools :** Terraform/Terragrunt, Ansible, Jenkins, Gitlab, Github, Docker, Kubernetes, CI/CD

**Hadoop Skills :** BigData Ecosystem HDFS, HBase, MapReduce, Zookeeper, Yarn, Kerberos, Ranger,

 Sentry, Hive, Pig, Sqoop, Flume, Oozie, Spark, Hortonworks, Kafka, Confluent Kafka,

 Storm, Cloudera Manager, Ambari

**Databases :** MySQL, PostgreSQL, Oracle, Bigtable

**Operating Systems :** Linux (RHEL, Ubuntu, CentOS), Windows Server

**Methodologies :** Agile, Scrum, Jira, Waterfall

**Cloud Platforms :** Google Cloud Platform, AWS

**Infrastructure as Code (IaC):** Terraform, Ansible, CloudFormation, Packer, Chef, Puppet

**Programming & Scripting**: SQL, PL/SQL, XML, HTML, Perl, Shell, Bash, C/C++, python, YAML, JSON

**Networking & Security:** VPC Network, Load Balancing, IDS/IPS, DNS and Firewall, OAuth 2.0

**ETL Frameworks :** Hive, Spark,

**Monitoring Tools :** Catchpoint, Datadog, Opsview, Docker, elastic search (kibana), Dynatrace

**Protocols :** TCP/IP, FTP, SSH, SFTP, SCP, SSL

**RELEVANT EXPERIENCE:**

**Choreograph 03/22 – Current**

**Portland-OR**

**Cloud Engineer**

**Responsibilities:**

* Work on Google Cloud components include **Compute Engine**, **Google Kubernetes Engine (GKE)**, Composer, **BigQuery**, **Cloud Storage**, **Cloud Pub/Sub**, **and Cloud Functions**.
* Design, deploy, and manage containerized applications on **GKE**, defining pods, deployments, services, and ingress rules. Implement container orchestration strategies for scalability, resilience, and rolling updates.
* Configure GKE networking, including namespaces, network policies, and service meshes.
* Create and manage Cloud infrastructure using Terraform scripts. Use Git and CI/CD tools to manage Terraform code, automate deployments, and maintain version control.
* Design and implement data storage solutions using **Cloud Storage buckets**, selecting appropriate storage classes (Standard, Nearline, Coldline, Archive) based on access patterns and cost considerations.
* Configure bucket policies for data lifecycle management, versioning, and access control.
* Deploy and manage managed relational databases **(MySQL, PostgreSQL, and SQL Server)** on **CloudSQL**, configuring high availability, backups, and disaster recovery.
* Design, build, and deploy batch and stream data processing pipelines using **Dataflow**, leveraging its managed service capabilities for scalability and reliability.
* Deploy and manage Hadoop and Spark clusters on Google Cloud using **Dataproc**, configuring cluster sizes, optimizing job execution, and integrating with other GCP services.
* Build, schedule, and monitor complex data workflows using **Cloud Composer**, a managed Apache Airflow service, orchestrating tasks across multiple GCP services.
* Define and manage fine-grained access control policies for Google Cloud resources using **IAM** roles and permissions, ensuring the principle of least privilege.
* Use **Cloud Pub/Sub for** Real-time messaging service for event streaming and data ingestion. Publish and subscribe to messages between applications and services.
* Work on **Cloud Composer** to managed Apache Airflow service for orchestrating data pipelines. Define, schedule, and monitor complex data workflows visually.
* Design and implement data warehousing and analytics solutions using **BigQuery**, optimizing query performance and managing data loading processes.
* Managed and execute comprehensive cloud migration strategies to Google Cloud Platform (GCP) services such as Google Kubernetes Engine (GKE) and Cloud Functions, enhancing application performance, and ensuring efficient, secure, and scalable cloud-native solutions.
* Configure monitoring, logging, and tracing solutions to gain visibility into application performance, identify and debug issues, and ensure operational health.
* Responsible for designing, planning, and overseeing the implementation of cloud-based computing systems and ensure that cloud computing services are deployed effectively and securely.
* Automated repetitive tasks and processes using scripting languages, reducing manual intervention and errors.
* Understanding of cloud computing, data storage, network technologies, and security. Also able to communicate complex technical concepts to non-technical personnel.
* Design and implement secure and efficient **virtual private cloud** (VPC) networks, segmenting resources into subnets and configuring firewall rules.

**Environment:** GCP,Compute Engine, GKE, Cloud Functions, Cloud Run, Cloud Storage, Cloud SQL, BigQuery, Dataproc, Composer, Cloud VPC, IAM, Cloud Logging, pub/sub, Confluence, Gitlab, Catchpoint, Datadog

**Wunderman 01/19 – 03/22**

**Portland-OR**

**Hadoop Engineer**

**Responsibilities:**

* Proactively analyze Hadoop cluster and driving performance, security, and innovation within the big data ecosystem using tools like **Ambari, Cloudera Manager**, or open-source utilities.
* Identify and resolve performance bottlenecks in Hadoop components (**HDFS, YARN, MapReduce**). Optimize job configurations, data storage formats, and cluster resources for efficiency.
* Design, develop, and maintain robust and scalable data pipelines for ingesting, processing, and storing data using Hadoop ecosystem tools like Spark, Hive, Sqoop, and Flume.
* Implement and manage CI/CD pipelines for Hadoop cluster deployments and application releases.
* Automate infrastructure provisioning and configuration management for Hadoop clusters using tools like Terraform or Ansible.
* Analyze **HDFS** read/write patterns, optimize block size, replication factor, and data locality for specific workloads. Implement caching mechanisms for frequently accessed data.
* Configure HDFS permissions and access control lists (ACLs) to restrict access to sensitive data at the file and directory levels. Integrate with **Kerberos** for secure authentication.
* Analyze **YARN** resource utilization patterns, forecast future needs, and adjust queue configurations to ensure optimal resource allocation for different workloads (batch, interactive, real-time).
* Fine-tune YARN parameters (e.g., container sizes, scheduler settings) to optimize job execution times and resource utilization. Implement resource preemption strategies for high-priority workloads.
* Architecture design and implementation of deployment, configuration management, backup, and disaster recovery systems and procedures. Hand on experience on cluster up gradation and patch upgrade without any data loss and with proper backup plans.
* Tune **HBase** regions for optimal performance, manage HBase compactions, and configure replication and failover mechanisms for high availability.
* Provided security and authentication with ranger where ranger admin provides administration and user sync adds the new users to the cluster.
* Analyze **MapReduce** job execution plans, optimize data serialization formats, and tune map and reduce tasks for efficiency. Leverage combiners to reduce data shuffling.
* Developed Map Reduce programs to cleanse the data in HDFS obtained from heterogeneous data sources to make it suitable for ingestion into **Hive** schema for analysis.
* Analyze MapReduce job execution plans, optimize data serialization formats, and tune map and reduce tasks for efficiency. Leverage combiners to reduce data shuffling.
* Integrate Kafka with Hadoop for real-time data ingestion, manage **Kafka** topics and partitions, and ensure data durability and consistency.
* Optimize Hive queries for performance, manage Hive metastore configurations, and implement security policies for data access control.
* Working experience on maintaining **PostgreSQL** databases creation and setting up the users and maintain the backup of cluster metadata databases with corn jobs.
* Setting up PostgreSQL master and slave replications and helping business applications to maintain their data in Postgres Servers.
* Maintaining **Microsoft SQL Server** databases to ensure their availability, security, and performance through indexing, query optimization, and resource allocation.
* Regularly update and improve existing code to enhance performance, fix bugs, or adapt to new requirements. Ensure the functionality of the software by writing unit tests, integration tests, and performing debugging.
* Worked with systems engineering team to plan and deploy new environments and expand existing clusters.
* Monitored multiple clusters environments using AMBRI Alerts, Metrics and Nagios. Improved efficiency in choreograph work by developing robust **database automation** systems
* Responsible for Cluster maintenance, Monitoring, commissioning and decommissioning Data nodes, Troubleshooting, Manage and review data backups, Manage &review log files.

**Environment**: Cloudera(CDP), Hadoop Hdfs, Map reduce, Hive, Pig, Spark, Kafka, Oozie, Kerberos, Ranger, Ranger KMS, Docker, Streamsets, Payara, Hortonworks, Ambari, Opsview, Kibana, elastic search.

**Nordstrom 06/18 – 01/19**

**Seattle, WA**

**Hadoop Engineer**

**Responsibilities:**

* Handles a variety of complex tasks to ensure the efficient operation, security, and optimization of a Hadoop ecosystem. Develop complex cluster architectures tailored to specific business needs, including multi-tenant environments and hybrid cloud setups.
* Installed/Configured/Maintained Apache Hadoop clusters for application development and Hadoop tools like Hive, Pig, HBase, HDFS and Sqoop.
* Performed system administration activities on Linux, CentOS & Ubuntu. Developed Java Map/Reduce job for Trip Calibration, Trip summarization and data filtering.
* Build and manage CI/CD pipelines for data pipelines using tools like Jenkins, GitLab CI. Automate testing, deployment, and version control of data processing jobs.
* Implement and manage dynamic configuration systems to adjust cluster settings in real-time based on workload demands.
* Create Backups & Recovery and Disaster recovery procedures and Implementing BACKUP and RECOVERY strategies for off-line and on-line Backups.  Develop and test disaster recovery plans to ensure data integrity and availability in case of failures.
* Worked with Infrastructure teams to install operating system, Hadoop updates, patches, version upgrades to Hadoop components while minimizing downtime.
* Handled importing of data from various data sources, performed transformations using Hive, Map Reduce, and loaded data into HDFS. Also, ensure data is stored efficiently with proper replication and block size configurations.
* Extensively used out of the box Kafka processors available in Nifi to consume data from Apache Kafka specifically built against the Kafka consumer API.
* Provision, configure, monitor, and manage **EMR clusters** for Hadoop workloads. Handle cluster scaling, upgrades, and lifecycle management.
* Design and manage **Virtual Private Cloud (VPC)**, subnets, security groups, and network ACLs to secure **EMR** clusters.
* Analyzed the data by performing Hive queries and running Pig scripts to know user behavior like shopping Enthusiasts, travelers, music lovers etc.
* Fine-tune MapReduce, YARN, and Spark jobs for better performance. Adjust YARN ResourceManager settings to balance load and ensure fair resource allocation.
* Installed Oozie workflow engine to run multiple Hive. Monitoring workload, job performance and capacity planning using Cloudera Manager. Developed Hive queries to process the data and generate the data cubes for visualizing
* Configure Kerberos for authentication and set up ACLs (Access Control Lists) for data security. Writing shell scripts for manipulating data. Experienced in defining job flows.
* Managed and optimized log management systems throughout the process to monitor performance, troubleshoot issues, and enhance CI/CD pipelines.
* Responsible to manage data coming from various sources. Used Oozie tool for job scheduling.
* Installing and configuring EC2 CLI tools on EC2 instances and deploying customs scripts for taking backup/snapshot of instances associated with the account.

**Environment**: Hive, YARN, Pig, HBase Apache Nifi, PL/SQL, Hive, Java, Unix Shell scripting, Sqoop, Kafka EMR, Splunk, HBase, Datadog, elastic search (Kibana) and Cloudera Manager.

**Galaxy Soft Solutions, 02/17 – 05/18**

**Canton, MI**

**Hadoop/Linux Administrator**

**Responsibilities:**

* Installed and configured a Horton Works HDP 2.2 and Hadoop 2.6 using AMBARI.
* Worked on analyzing Hadoop cluster and different big data analytic tools including Pig, Hbase database and Sqoop.
* Responsible for building scalable distributed data solutions using Hadoop
* Worked on installing cluster, commissioning & decommissioning of datanode, name node recovery, capacity planning, and slots configuration.
* Created HBase tables to store variable data formats of PII data coming from different portfolios.
* Managing and reviewing Hadoop log files and debugging failed jobs.
* Implemented Kerberos Security Authentication protocol for production cluster.
* Implemented test scripts to support test driven development and continuous integration. Also, worked on tuning the performance Pig queries.
* Creating Backups & Recovery and Disaster recovery procedures and Implementing BACKUP and RECOVERY strategies for off-line and on-line Backups.
* Worked with Infrastructure teams to install operating system, Hadoop updates, patches, version upgrades as required.
* Monitoring and controlling local file system disk space usage, log files, cleaning log files with automated scripts.
* Automated all the jobs for pulling data from FTP server to load data into Hive tables, using Oozie workflows.
* Backed up data on regular basis to a remote cluster using distcp. And Responsible to manage data coming from various sources.
* Configure and optimize advanced system features such as RAID arrays, LVM (Logical Volume Manager), and kernel parameters. Implement high-availability and clustering solutions (e.g., Pacemaker, Corosync).
* Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team.
* Perform in-depth performance analysis using tools like perf, dtrace, or sar. Plan for system capacity and scalability to accommodate growth and optimize resource usage.
* Involved in data analysis projects using Elastic Map Reduce on the Amazon Web Services (AWS) cloud.
* Cluster coordination services through Zookeeper. Loaded the dataset into Hive for ETL Operations.
* Monitor and tune the performance of critical applications and services. Collaborate with application developers to optimize application performance on the Linux/Unix environment.

**Environment:** LinuxOS, RHEL, CDH4.X, HDFS, HUE, Oozie, HIVE, Sqoop, Zookeeper, Spark Unix scripts, YARN, Fair Scheduler, HBase Kerberos, Oracle, MySQL, Ganglia, Nagios.

**Source One 07/13 – 12/14**

**Hyderabad-TS**

**Associate Engineer**

**Responsibilities:**

* Install and configure Linux/Unix operating systems.Set up and manage system hardware, software, and network configurations.
* Configured volume groups and logical volumes, extended logical volumes for file system growth needs using Logical Volume Manager (LVM) commands.
* Create and manage user accounts and groups. Set permissions and access controls for files and directories
* Building Linux servers by using VMware ESX 3.x and also involved in channels and pushing the packages in Redhat Satellite Server. Installing, configuring and maintaining the VMwareESXi 4.1/5.0 and Upgradation
* Diagnose and resolve hardware, software, and network issues. Provide technical support to users and other IT staff.
* Ensure systems comply with organizational policies and industry regulations. Conduct periodic audits and reviews to ensure compliance.
* Checking the Virtual Center for tuning and Performance. Working on File system and Swap configurations
* Updating YUM Repository and Redhat Package Manager (RPM).
* Monitor system performance and resource usage. Tune system parameters to optimize performance. Use tools like top, vmstat, iostat, and netstat to monitor system health.

**Environment:** Solaris 9/10, Redhat Linux 4/5/6, HP-UX 11i, AIX, Sun Enterprise Servers, Sun Fire, Sun T Series, Sun M Series, HP 9000K, L, N class Server, HP & Dell blade servers, IBM RS/6000, p Series servers, VMware ESXi Server, Oracle.

**Education: 01/15- 01/17**

* Master of Science in Electrical & Computer Engineering from Lawrence Technological University, Southfield-Michigan, USA.
* Bachelor of Technology in Electrical and computer Engineering, SCSVMV University, Chennai-TN, India.