SUMMARY

Results-oriented Lead Data Engineer with over 10 years of experience in RDBMS, Java, Big Data, and Cloud ecosystems. Proven expertise in designing and implementing end-to-end data solutions, including cloud services like AWS, Azure, and Oracle Cloud. Demonstrated leadership in managing large-scale development projects, optimizing data processes, and ensuring data quality. Adept at leveraging advanced technologies such as Hadoop, Spark, and Kafka to drive innovation and streamline data workflows. Strong analytical and technical skills coupled with a track record of successful project delivery. Excited to bring a wealth of experience to lead transformative initiatives in data engineering.

RELEVANT PROFESSIONAL EXPERIENCE

***HEXstream, MA, USA Feb 2018 to Present***

***Role: Sr Data Engineer***

* Developed ELT/ETL process and end to end analytics solutions using Azure Cloud Services (PaaS & laaS), Azure Synapse Analytics, SQL Azure, Data Factory, Azure Analysis services, Application Insights, Azure Monitoring, Key Vault, Azure Data Lake, and PowerBI
* Implemented Solution in Azure DevOps for tracking and logging end to end software application.
* Designed and developed claim processing Azure data factory pipelines.
* Worked with cross application teams to solve the pipeline development requirements and issues.
* Designed a common real time spark application for the claim processing from different sources.
* Efficiently implemented the incremental load approach in the azure event driven pipelines.
* Created and deployed multiple pipelines for daily, monthly, and historical data loads.
* Improved performance of the queries on the Databricks engine.
* Implemented end to end process automation solution for clean water billing using Power Automate and PowerApps.
* Converted physical forms to digital forms using PowerApps to enable the automation.
* Designed the PowerApps application to save the data into SharePoint Online and Oracle database.
* Designed and developed the PowerApps application to work in the offline mode also.
* Created automation workflows to perform complex nested calculations using Power Automate.
* Developed streaming solution using Spark Streaming and Kafka connectors.
* Designed and implemented data lake solution for on-premises environment as well as cloud environment. Involved in finalizing the data lake architecture.
* Worked on extract, transform and load activities using Oracle Integration Cloud which moved the data into data lake solution.
* Performed transformations as well as analytics on the data lake.
* Integrated various on-premises applications with cloud using Oracle Integration Cloud.
* Worked on a solution to process handwritten documents using Python.
* Implemented invoice processing application using IQ Bot which is provided by Automation Anywhere.
* Installed and configured Kinetica GPU database on the on-premises server.
* Written python scripts for Ingestion as well as CRUD operations on Kinetica database.
* Created dashboards on Reveal to visualize the data present on Kinetica.
* Worked on Azure cloud and migrated On-premises databases to Azure.
* Provisioned Azure Cosmos database and performed benchmarking with Kinetica database.
* Worked ETL transformations using Spark on various source systems.
* Developed a replication tool which will copy the data from relational to Hadoop filesystem using Kafka cluster.
* Written Spark streaming code using Scala for data migration. Handled bulk load and real time data migration.
* Deployed Harmonia application to stream data from SAP HANA to MapR DB/HBase on the MapR cluster, Cloudera and Hortonworks single node clusters and thoroughly tested the data stream on all the three different clusters.
* Worked on configuring Cloudera multi node cluster using Cloudera Manager and configured Hortonworks cluster using Ambari.
* Developed a design for end-to-end data flow solution using Data virtualization.
* Worked on a proof of concept for Data virtualization using Denodo.

***AT&T, TX, USA Mar 2017 to Jan2018***

***Role: Sr Big Data Developer***

* Worked on an APM application based on Glowroot using Java and Cassandra.
* Developed Java Rest APIs to prevent agents connecting to wrong cluster.
* Worked on analyzing and organizing logs for vmstat, netstat, psdata, threaddump and thread gc logs.
* Deployed Elasticsearch, logstash and Kibana for dev, test, and prod environments.
* Written shell scripts to install the packages and deploy them on a given hostname.
* Configured yaml files for filebeat and ELK technologies.
* Written logstash configuration files by using filters like grok and ruby.
* Installed Elasticdump tool to handle dumping of elastic search indices from one host to another.
* Created various visualizations and dashboards for the log data on Kibana.
* Worked on template file to create string analyzed and string not analyzed fields.
* Created scripted fields on Kibana to accommodate various aggregations.
* Upgraded ELK components from version 2.3 to 5.2.1, which is an 8-node cluster.
* Installed cerebro plugin to monitor and manage ES data. Implemented various queries on Kibana.
* Performed data backups as part of upgrade process and created scripts, which reduces the downtime.
* Performed reindex operation to make old data compatible with new version of Elasticsearch.
* Improved performance of reindex operation and completed the entire process in a day.
* Implemented authorizations and permissions for ELK cluster using searchguard.
* Implemented shell scripts to reduce manual efforts in configuring ELK components.
* Good Working knowledge of Tableau.
* Performed advanced procedures like text analytics and processing, using in-memory capabilities of Spark using Scala.
* Perform data quality checks like schema validation, record count validation and timestamp check on the raw data received.

***United Health Group, NJ, USA Dec 2016 to Mar 2017***

***Role: Sr Big Data Developer***

* Worked on MapR ecosystem to manage and scale the cluster of 30 nodes.
* Worked on MapR to access data easily across the cluster and stored the data in variety of warehouse environments.
* Stored logs data on MapR file system that are generated from various machines
* Worked on integrating MapR DB with MapR ES for real time data flow that enables real time data ingestion and stream processing.
* Created streams and topics in MapR-ES and processed the data to multiple consumers
* Performed a proof of concept for streaming data from Cassandra to MapR-DB using Spark Streaming. Written Scala scripts which can handle both bulk data load and streaming data.
* Configured drill to do analytics on MapR data.
* Written partitions and buckets in Hive for speedy query processing.
* Created HBase tables to store daily snapshots on top of Hive.
* Ingested data from various relational data sources using Sqoop to Hive.
* Solved performance issues in Hive and Pig scripts with understanding of Joins and aggregation and how does it translate to MapReduce jobs.
* Worked on a migration project which handles archival and purge functionalities from Java 1.3 to Java 1.6.
* Have created performance monitoring shell script which identifies how many requests were served each second.
* Also worked on log analysis script which captures http request and responses.

***Client: AT&T, TX, USA Sep 2016 to Nov 2016***

***Role: Sr Big Data Developer***

* Working on analyzing and organizing logs for vmstat, netstat, psdata, threaddump and thread gc logs.
* Configured Elasticsearch, logstash and Kibana for dev, test, and prod environments.
* Created a deployment script, which deploys ELK instance on a given hostname.
* Worked on yaml files for filebeat and ELK technologies.
* Written logstash conf files by using filters like grok and ruby.
* Installed Elasticdump tool to handle dumping of elastic search indices from one host to another.
* Created various visualizations and dashboards for the log data on Kibana.
* Worked on template file to create string analyzed and string not analyzed fields.
* Created scripted fields on Kibana to accommodate various aggregations.
* Installed sense and elasticsearch head plugin to manage ES data.
* Implemented various queries on Kibana.
* Written shell scripts to install the packages and deploy them.
* Maintained code through GIT.

***Ventois Inc June 2016 to August 2016***

***Role: Sr Hadoop Developer***

* Configured and deployed instances on AWS cloud environment and data centers.
* Created and managed S3 buckets and policies for storage and backup purposes.
* Worked on transition workflows to Continuous Integration and Delivery model and implemented best practices for utilizing AWS
* Configured and implemented Secure Elastic Load Balancers for high availability of web and application servers.
* Performed user, group and application access and permissions in AWS using the Identity Access Manager (IAM).
* Used Jenkins AWS CodeDeploy plugin to deploy to AWS
* Responsible for code build, release, and configuration on Amazon EC2
* Created monitors, alarms, and notifications for EC2 hosts using Cloudwatch.
* Migrated applications to the AWS cloud and automated administration tasks in AWS using Jenkins, GIT, and shell scripting.
* Worked on ingesting data using Talend Jobs and Transform data to various sources using SQL Server Integration Service and Talend Open Studio.
* Moved data from Oracle data sources to Hive tables.
* Broad design, development and testing experience with Talend Integration Suite and knowledge in Performance Tuning of mappings.
* Created joblets in Talend for the processes that can be used in most of the jobs in a project like to Start job and Commit job.
* Troubleshooting, debugging & fixing Talend specific issues, while maintaining the health and performance of the ETL environment.

***Mayo Clinic August 2015 to Jun 2016***

***Role: Java/Hadoop Developer***

* Developed Java Restful web-services to make it work with HBase
* Implemented HTTP protocol methods to support the user requests
* Designed the architecture and working on flow diagrams.
* Developed and designing code, based on user stories provided in TFS agile model
* Worked on Elastic Search and created indexes for many resources.
* Worked closely with UI team, business analysts and medical practitioners to understand the requirements clearly and make the code efficient.
* Created tables in HBase as per user requirements and performing queries on tables.
* Involved in preparing application maintenance manual.

***United Health Group, MN, USA June 2015 to July 2015***

***Role: Hadoop Developer***

* Worked on setting up MapR cluster for ingestion and worked on MapReduce jobs for input files while being ingested.
* Worked on Hive tables to store the daily and monthly data and created partitions on hive tables.
* Worked on pig scripts and ran the jobs using Apache Falcon
* Involved in running various Falcon jobs which get data from different external sources.
* Involved in data loading from Linux file system to HDFS.

***Ventois Inc., MA, USA March 2015 to May 2015***

***Role: Hadoop Developer***

* Involved in gathering of functional and non-functional requirements.
* Implemented MapReduce jobs to process the data.
* Imported and exported data using Sqoop into HDFS and Hive data warehouse.
* Created and monitored jobs using Oozie.
* Ran Hadoop streaming jobs to process terabytes of xml data
* Experienced in processing of load and transform the large data sets of structured, unstructured, and semi structured data.
* Implemented Map-side and Reduce-side joins according to the requirements.
* Handled data copying from Linux operating systems to HDFS.
* Experienced in data summarization and analysis using Hive queries.
* Performed dynamic partitions in Hive and wrote UDF’s.
* Done performance tuning by optimizing various joins.
* Performed Unit testing using MRUnit.

***Tata Consultancy Services Limited, Chennai, India***

***Project: AC Nielsen Mar 2011 to Sep 2013***

***Role: Systems Engineer***

* Responsible for gathering technical and functional requirements
* Prepared application maintenance, Low level, and high-level design documents
* Worked on SQL Server and developed Joins, triggers, stored procedures, and functions.
* Had good grasp on pros and cons of clustered and non-clustered indexes.
* Worked on performance tuning of stored procedures
* Developed a script to archive and purge the historic data using shell scripting.
* Expertise in scheduler tools like Autosys and Cron.
* Expert in maintaining of DB scripts in Serena PVCS version controller.

QUALIFICATIONS

Microsoft Power BI Data Analyst exam

Microsoft certification Azure Fundamentals

Oracle Cloud Infrastructure 2023 Application Integration Professional

Oracle Cloud Infrastructure 2023 Foundations Associate

Automation Anywhere Certified Master RPA Professional (Automation 360)

Automation Anywhere Certificate for Automation Anywhere Certified Advanced RPA Professional (A2019)

Certified Automation Hero Automation Developer

AWS Certified SysOps Administrator

Oracle Big Data 2017 Certified Implementation Specialist

Certified in Programming in HTML5 with java script and css3.

**Southern New Hampshire University, Manchester, NH - 2014**

Masters in computer and information sciences

**Jawaharlal Nehru Technological University, Telangana, India – 2010**

Bachelor of Technology (Electronics and Communications Engineering)

TECHNICAL COMPETENCIES

Cloudera, HortonWorks, MapR, Oracle Integration cloud, Hadoop, Map Reduce, MapR-FS, MapR-DB, HDFS, Apache Spark, Cassandra, HBase, Hive, Pig , Elastic Search 5.2.1, LogStash 5.2.1, Kibana 5.2.1, Filebeat, AWS, Azure, EC2, Java, Groovy, IntelliJ Idea, JFrog, XL deploy, Linux, Shell Scripting, Git, SVN, DB2, IBM WebSphere 8.5.1, DB Visualizer, , Falcon, YARN scheduler, MySQL, Talend Open Studio, Tableau, Rest Assured, Jenkins, Maven, Jetty, SQL Server, JIRA, TFS, Rally and Agile environment