Lakshmikanth Guduri

Dallas, Texas, United States   ✉ kanthguduri@gmail.com   📱 4043334414   🔗 in/kanth-g-15263628a

SUMMARY

A seasoned IT professional with 6 years 9 months of extensive experience in Bigdata technologies and Java J2EE. Proficient in deploying big data solutions and handling large sets of structured and unstructured data using Spark, Scala, Hive, Pig and more. Solid exposure to data operations in AWS, Azure, and GCP data services. Significant experience in streaming applications and proficient in languages like Scala, Python, and Java. Adept in data integration tools and well versed in developing key methods in established and emerging data technologies.

EXPERIENCE

Bigdata Developer

Comcast, Remote   June 2022 -December 2023

• Managed migration of an existing on-premises application to AWS using services like EC2 and S3 for data sets processing and storage
• Developed Spark programs using Scala and Java APIs, performed transformations and actions on RDDs
• Implemented data ingestion and handling clusters in real-time processing using Kafka
• Developed Spark applications for filtering JSON source data in AWS S3 and stored it into HDFS with partitions
• Developed and implemented spark jobs on Databricks to perform tasks like data cleansing, data validation, standardization and then applied transformations as per the use cases.
• Analyzed data from Azure data storages using Databricks for deriving insights using spark cluster capabilities.
• Worked in the AWS environment for the development and deployment of Custom Hadoop Applications
• Developed Automation Regressing Scripts for validation of ETL process between multiple databases like AWS Redshift, Oracle, Mongo DB, PL-SQL, and SQL Server using Python.

Bigdata Developer

Charter Communication, St. Louis, Missouri   December 2021 - May 2022

• Processed data into HDFS, analyzed the data using Hive, Spark and produced summary results for downstream systems
• Converted Talend jobs to support Snowake functionality
• Developed shell and python scripts to automate and provide control ow to data pipelines
• Migrated data from multiple source systems such as Teradata, DB2, SQL Server and Oracle to Hadoop distributed le system for data analysis
• As part of POC, worked on converting the legacy DataStage jobs into Airflow jobs
• Worked on Data Extraction, aggregations and consolidation of data within AWS Glue using PySpark.

Data Engineer

Walmart Inc, Bentonville, Arkansas   January 2020 - November 2021

• Performed source analysis on various databases for modelling and loading the customer data into Data Lake
• Developed Distcp framework shell script for copying the data across different clusters and environments
• Implemented pre-defined operators in spark such as map, at Map, filter, reduce By Key, group By Key, aggregate By Key and combine By Key etc.
• Developed customized Sqoop scripts for loading data from different databases like Oracle, MSSql, Teradata
• Extracted Real time feed using Kafka and Spark Streaming and converted it to RDD and processed data in the form of Data Frame and saved the data as Parquet format in HDFS.

Data Engineer

Synchrony Financial, Chicago, Illinois   August 2019 - December 2019

• Developed a PySpark application for creating the Cal report Hive table and made it available for tableau reporting
• Used Sqoop for loading data to HDFS collection layer
• Developed Pyspark application for creating Payfone reporting tables with different masking in both Hive and MySQL DB
• Used pre-defined Udf's for Protegrity tokenizing and AES encrypting the sensitive fields like Phone number, Country code, SSN and DOB etc., in the reporting tables
• Successfully migrated multiple Scala and Pyspark applications from old clusters to new LCM clusters.

Bigdata/Hadoop Developer

S&P Global,55 Water Street, New York   February 2019 - July 2019

• Analyzed Integrated data facility platform (IDF) and made recommendations to the Scala source code in storing large amount of data in parquet format on S3 buckets
• Analyzed Spears Automated Surveillance which is a rules-based monitoring system built on top of IDF platform
• Worked on several semi-structured formats to process datasets in S3 like JSON, XML, CSV
• Used Spark-SQL to load json data and create schema RDD and load it into Hive tables
• Developed Unix shell scripts to onboard the historical data from on-premise to S3.

Hadoop and Data Science Platform Engineer

Apple Inc, Sunnyvale, California   February 2018 - January 2019

• Performed benchmarking of federated queries in Spark and compared their performance by running the same queries on Presto
• Created partitions and buckets defined Hive tables for data analysis
• Developed Spark scripts by using Scala as per the requirement
• Worked on shell scripting to determine the status on various components in data science platform
• Performed data copying activities in a distributed environment using Ansible.

Java Developer

• Worked on Nifi data pipelines that are built for consuming data into Datalake
• Built Nifi flows for testing and debugging purposes.
• Managed end to end web applications to process large scale data using Spring MVC framework with Rest API
• Worked on Oracle database and developed SQL queries and stored procedures to integrate data from different tables
• Extensively followed Agile methodology like Sprint Planning, Story review/Story sizing and SCRUM meetings
• Successfully involved in test driven development model using Junit.

EDUCATION

Masters
Electrical Engineering • University of South Alabama • 2016

Bachelors
Electronics and Communication Engineering • SRM University • 2014

SKILLS

Bigdata Technologies (Sqoop, Hive, Pig, Spark, Scala, Flume, Kafka, Oozie, HBase)	Java J2EE Technologies	Cloud Services (AWS, Azure, GCP)	Data Analysis (HiveQL, PIG Latin)
Apache Spark Operations	Handling Large Datasets (Structured, Semi-structured, Unstructured)	NoSQL Databases (HBase, MongoDB)	Data Integration Tools (Apache Nifi, DataStage, Automic)
Programming Languages (Scala, Python, Java)		DevOps Tools (Maven, Jenkins, Ansible, Git)	