SATHISH.V

Personal info
Email: infovssat@gmail.com
Phone: (+91) 9944020412

Career Objective:

Driven and innovative big data developer with a passion for leveraging cutting-edge technologies to extract actionable insights from complex data sets. Proficient in developing scalable data processing pipelines, optimizing query performance, and implementing advanced analytics solutions. Committed to continuous learning and staying abreast of emerging trends in big data technologies. Seeking to contribute expertise to a dynamic team dedicated to driving data-driven decision-making and innovation

Summary:

➢ Having over all 8.5+ years in IT. Accumulated 3 years of hands-on experience in the big data domain, encompassing proficient utilization of cutting-edge technologies and methodologies to extract, process, analyze, and visualize vast and complex datasets.

➢ Proficient in developing and implementing Spark RDD-based data processing workflows using Scala or Java programming languages.

➢ Strong understanding of Spark RDD integration with other big data technologies, such as Hadoop, Hive

➢ Expertise in using Spark Data Frame transformations and actions to process large-scale structured and semi-structured data sets, including filtering, mapping, reducing, grouping, and aggregating data.

➢ Proficient in processing serialized data in Spark using various formats, such as Avro, Parquet, ORC and their features and limitations.

➢ Skilled in working with binary and textual data formats in Spark, such as CSV, JSON, and XML, and their serialization and deserialization using Spark Data Frames and RDDs.

➢ Maintained and monitored Spark clusters on AWS EMR, ensuring high availability and fault tolerance.

➢ Designed and implemented data lake architectures on Amazon S3, leveraging partitioning and columnar formats such as Parquet to optimize query performance and minimize storage costs.

➢ Designed and implemented ETL processes using Spark.

➢ Worked with Spark Data Frame APIs for structured data analysis.

➢ Skilled in leveraging Hive partitioning, bucketing, indexing, and caching features to improve query performance and reduce data processing overhead.

➢ Experienced in importing and exporting large datasets between Hadoop and relational databases using Sqoop.

➢ Proficient in writing Sqoop commands to transfer data between Hadoop and various databases such as MySQL, and SQL Server

➢ Experienced in using Sqoop to import and export data from and to cloud-based data storage services such as Amazon S3

Educational Qualification:

➤ MCA from SVS college of Engineering, Anna University.
➤ B.Sc. Computers from SNR SONS College of Arts and Science, Bharathiyar University.

Work Experience:

➤ Working as an Big data developer in UST, Bangalore from OCT 2021 to till date.
➤ Working as an Automation Test Engineer in Infosys, Chennai from July 2017 to till 2021

➤ Worked as an Automation Test Engineer in ThoughtHub Solutions, Chennai from July 2015 to July 2017

Technical Skills:

Big Data Eco System	Hadoop, Sqoop, Apache Spark, Hive
Cloud Technologies	AWS
Databases	SQL Server, MySQL, Hive
OS	Mac OS, Windows, Linux
Build and CI Tool	Maven
Programming	Java, Scala, Python
Version Control	GIT

Projects:

#1Retail Project:Walmart
Role:Big Data Developer

- ➢ Developing and optimizing algorithms and workflows for processing, cleansing, and transforming large volumes of structured and unstructured data to extract valuable insights.
- ➢ Developed Spark applications for distributed data processing.
- ➢ Created and managed RDDs (Resilient Distributed Datasets) for data transformations.
- ➢ Utilized Data Frames for structured data manipulation and analysis.
- ➢ Designed and implemented Spark jobs using Scala or Python.
- ➢ Performed data cleansing and preprocessing using Spark transformations.
- ➢ Developed Spark Streaming applications for real-time data processing.
- ➢ Optimized Spark jobs for performance and resource utilization.
- ➢ Implemented Spark SQL queries for data querying and aggregation.
- ➢ Worked with Hive and HBase integration for data storage and retrieval.

#2Retail Project: Cloud Console
Role:Big Data Developer

- ➢ Wrote Spark applications for batch and stream processing.
- ➢ Created and managed Spark clusters for distributed computing.
- ➢ Implemented Spark partitioning and caching strategies.
- ➢ Monitored Spark jobs using cluster management tools like YARN or Mesos.
- ➢ Worked with Spark's data serialization formats (Avro, Parquet, JSON, etc.).
- ➢ Debugged and optimized Spark code for performance bottlenecks.Worked with Hive and HBase integration for data storage and retrieval.
- ➢ Integrated Spark with data lakes such as AWS S3 ,HDFS,EMR,EC2.
- ➢ Designed and implemented ETL processes using Spark.

#3Banking Project:
To develop Mobile application functional scripts using selenium and it will run in Mobile Labs cloud, Common Script for both Android and iPhone. Capable to run in Both local as well as cloud and should be individual in order to achieve the parallel operation.

Technologies used:

Java, Appium, Jenkins, Maven ,TestNG ,Jira, Mobile Labs, ALM,

#4 Retail Ticketing System:
EventZilla is product-based software, targeted as an online event registration, it helps us to host event anywhere in the country and sell tickets online and manage attendees from one integrated application and the reports are generated on weekly, monthly and yearly basis. It gives you all the information about sold tickets, number of user registered etc.,

Technologies used:

Java, Selenium, Jenkins, Maven, TestNG, Jira, Perfecto, ALM.