DINESH H R

Phone: (+91) - 8970222928 Dineshhr23@gmail.com

PROFESSIONAL SUMMARY

- **Having 2** years of experience in designing and developing **Big Data** applications using the Hadoop Ecosystem technologies (**HDFS, Hive, Sqoop, Apache Spark, PySpark and AWS**).
- Experience with **Scala** spark and **Pyspark**.
- Real time experience in Hadoop/Big Data related technology experience in Storage, Querying, Processing and analysis of data.
- Experience in **Spark-SQL** query type data processing operations using **Scala & Pyspark**.
- Experience in **Spark-DSL** type data processing operations using **Scala & Pyspark**.
- Experience in **Spark Memory Tuning and Performance Tuning.**
- Skilled in processing large sets of **structured** and **semi-structured** data.
- Experience in working with different file formats like **JSON, CSV, AVRO, Parquet** data files, and text files.
- Proficient in developing codes and modules to address **complex data processing** (json, struct) requirements in Big Data.
- Experience in importing and exporting data using **Sqoop** from **HDFS to RDBMS** and vice versa.
- Hands-on experience in deploying Spark jobs over **EMR** cluster as a step execution.
- Skilled in performance testing of ETL processes to ensure scalability and efficiency.
- Deep knowledge in incremental imports, partitioning and bucketing concepts in Hive and Spark SQL needed for optimization.
- Experience in **Agile methodology** for efficient system development with IT and business teams.
- Worked on AWS Components like **RDS, S3, Athena, EMR and EC2**.
- Worked on AWS components like **S3 and data architecture including data ingestion** and pipeline design.
- Worked on RDBMS Tables import/export using SQOOP.

- Strong experience in Extraction, Transformation and Loading (ETL) data from various sources into Data Warehouses.
- Performed Spark Data Frame optimization techniques, such as predicate pushdown, column pruning, and vectorized execution, and their impact on query performance and resource utilization.
- Strong problem-solving and analytical skills with a passion for innovation.
- Experience in interpersonal and communication skills.
- Developed and maintained complex data pipelines using Apache Spark and PySpark, ensuring efficient data processing and analysis.
- Implemented data transformations, aggregations, and cleansing processes to improve data quality and accuracy.
- Utilized Hive for data warehousing, managing large datasets, and optimizing query performance.
- Monitored and optimized Spark jobs for performance, scalability, and resource utilization.
- Utilized AWS services such as S3 and EMR for data storage, orchestration, and processing.
- Conducted code reviews and mentored junior data engineers to enhance team productivity.

WORK EXPERIENCE

Cognizant – December 2021 – Till Date

Dedicated and results-driven Data Engineer with two years of hands-on experience in designing, developing, and maintaining data pipelines using Apache Spark, PySpark, Hive, and AWS services. Adept at optimizing data workflows to ensure data quality and reliability. Seeking a challenging position to contribute my expertise and drive innovation in data engineering.

TECHNICAL SKILLS

Programming Languages	:	Python, Scala, Java.
Data Eco System	:	Hadoop, Sqoop, Hive, Apache Spark, Pyspark, NIFI, Kafka.
Cloud Skills	:	AWS (EC2, S3, RDS, EMR).
Distribution	:	Cloudera.
Databases	:	MySQL, SQL Server, Cassandra.
Languages	:	Scala, Python, SQL.
Operating Systems	:	Linux and Windows.
IDE	:	Eclipse, Anaconda, EMR.

PROFESSIONAL EXPERIENCE

Project 1 : FlipKart

PROJECT ROLE: AWS CLOUD SPARK DEVELOPER

Flipkart is one of India's largest and most popular e-commerce companies. Flipkart is headquartered in Bengaluru, Karnataka, India. It operates as a marketplace platform, connecting sellers with buyers across various product categories. Flipkart has established a robust logistics and supply chain network to ensure efficient order fulfillment and timely delivery. It has built warehouses and fulfillment centers across the country, enabling smooth inventory management and last-mile delivery populay termed as Flipkart logistics.

- Skilled in handling **semi structured/serialized data processing** using Pyspark. (AVRO, PAQUET, JSON, CSV)
- Experienced in working as part of a data ingestion team where we sourced data from **RDBMS** and utilized **Sqoop** to transfer it to **HDFS**.
- Familiar with running multiple **Sqoop** jobs to process the data as there were numerous **RDBMS** tables involved.
- Strong understanding of Spark RDD integration with other big data technologies, such as Hadoop, Hive, and their impact on data processing workflows and performance.
- Proficient in processing serialized data in Spark using various formats, such **as Avro, Parquet,** with their features and limitations.
- Skilled in working with data formats in Spark, such as CSV, JSON, and XML, and their serialization and deserialization using Spark DataFrames and RDDs.
- Processed web URL data using scala and converted it to data frames for further transformations.
- Generated complex JSON data after all the transformations for easy storage and access as per client requirements.
- As per the business requirement storing the spark processed data in HDFS/S3 with appropriate file formats.
- Improved performance and optimized existing algorithms in **Hadoop** by utilizing **Spark Context, Spark-SQL, Data Frame.**
- Created a pre-processing job utilizing **Spark Data Frames** to transform **JSON** documents into **Data Frames**.

Technologies: Pyspark, Spark/Scala, Python, HDFS, Hive, Sqoop & SQL

Project 2: HEWLETT PACKARD

Hewlett Packard is a full-service, technology-driven and product development company. It developed and provided a wide variety of hardware components as well as software and related services to consumers, small- and medium-sized businesses (SMBs) and large enterprises, including customers in the government, health and education sectors.

PROJECT ROLE: AWS CLOUD DATA ENGINEER

RESPONSIBILITIES

- Analyzed the requirements and performed Impact Analysis based on the requirements
- Involved in applying transformation with SQL based on business Logic in the Mapping sheet
- Involved in modifying existing Procedures and ETL workflows according to the new business needs using Microsoft SQL server Management studio.
- Created EC2 instances and EMR clusters for development and testing.
- Performed step execution in EMR clusters for the job deployment as per requirements.
- Familiarity with Spark RDD-based data processing libraries and frameworks, such as Apache Spark SQL and their features.
- Proficient in developing and implementing Spark DataFrame-based data processing workflows using Scala, or Python programming languages.
- Experienced in optimizing Spark DataFrame performance by tuning various configuration settings, such as memory allocation, caching, and serialization.
- Expertise in using Spark DataFrame transformations and actions to process large-scale structured and semi-structured data sets, including filtering, mapping, reducing, grouping, and aggregating data.
- Familiarity with Spark DataFrame schema and data type operations, such as adding, renaming, and dropping columns, casting data types, and handling null values.
- Knowledge of Spark DataFrame optimization techniques, such as predicate pushdown, column pruning, and vectorized execution, and their impact on query performance and resource utilization.
- Ability to troubleshoot common issues with Spark DataFrame, such as data processing errors, performance bottlenecks, and scalability limitations.

- Maintained and monitored Spark clusters on AWS EMR, ensuring high availability and fault tolerance
- Involved in working on the Data Analysis, Data Quality and data profiling for handling the business that helped the Business team.
- Loaded and transformed large sets of semi structured data.

Technologies: Pyspark, Spark/Scala, Python, HDFS, Hive, Sqoop & SQL

EDUCATION

Institute/University/Board	Duration	Percentage Obtained
Bachelor of Engineering /*EPCET/ CSE*	**2017-2021**	**7.38**
PUC / *Karnataka*	**2017**	**82.16**

PERSONAL DETAILS

Name : DINESH H R

DOB : 11/05/1999

Gender : Male

Nationality : Indian

Languages : English, Kannada, Telugu

Address : Bangalore.

DECLARATION

I am here by declare that the above mentioned information is correct up to my best of knowledge and I bear the responsibility for the correctness of the above mentioned particular.

Place: **Bangalore** **DINESH H R**