

DINESH H R

Data Engineer

With 2.4 of experience as a seasoned data engineer, I possess a strong foundation in data processing and management, and am adept at leveraging tools such as Spark, PySpark, and Sqoop to maintain seamless data pipelines. I am proficient in ETL processes, data modeling, and ensuring data quality assurance by implementing best practices. I am proficient on working on complex data as struct, array. My expertise spans the entire data lifecycle, from ETL processes to data warehousing. I have a proven track record of collaborating with cross-functional teams to understand and address data requirements, facilitating data-driven decision-making. I am now seeking opportunities to further apply and expand my skills in a dynamic organization.

Contact

Phone : 8970222928
Email : dineshr23@gmail.com
Address : Bangalore, India.

Education

2021
Bachelor's in Computer Science Engineering.
East Point College of Engineering and
Technology, Bangalore

2017
2nd PUC (10+2)
SRI Bhagawan Mahaveer Jain College, KGF.

Technical Skills

Python, Java, Scala, Hadoop, Sqoop, Hive,
Apache Spark, PySpark, Nifi, Kafka,
MySQL, Sql Server, Cassandra, Snowflake,
AWS EMR, S3, RDS, EC2

Language

English Kannada Telugu

Certifications

Microsoft Certified:
• Azure Fundamentals.
• Azure Data Fundamentals.
• Azure Data Engineer Associate.

AWS Certified Cloud Practitioner.

Experience

2021 - Present | Cognizant | Bangalore

Jr. Data Engineer

- Experience with Scala spark and Pyspark.
- Experience in Spark-SQL query type data processing operations using Scala Spark & Pyspark.
- 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.
- 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 and Spark DataFrames 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.
- 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, RDS, Snowflake 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.
- 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.
- Experienced in optimizing Spark DataFrame performance by tuning various configuration settings, such as memory allocation, caching, and serialization.
- Familiarity with Spark DataFrame schema and data type operations, such as adding, renaming, and dropping columns, casting data types, and handling null values.
- Ability to troubleshoot common issues with Spark DataFrame, such as data processing errors, performance bottlenecks, and scalability limitations.
- Experience in Agile methodology for efficient system development with IT and business teams.

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