Nitish Kumar Rohira

Senior Data Engineer

✉ nitishrohira@gmail.com   📞 8769769505   📍 Bangalore, India

📅 10/07/1995

🪪 PROFILE

- Working as a big data engineer holding 6+ years of experience in developing end to end data pipelines using PySpark, Airflow, Kafka, S3, EMR, Lambda in AWS cloud native environment.
- Expertise in development/enhancement of converting legacy big data architecture to AWS serverless architecture minimizing the cost of the managed resources.
- Holds experience in setting up data and analytics platform on AWS for multiple downstream systems which acts as single source of truth for various project and operating teams doing advanced analytics on the data for business profitability.

💼 PROFESSIONAL EXPERIENCE

09/2023 – present
Noida, India

ShortHills AI
Senior Data Engineer
- Utilized Apache Airflow to orchestrate and schedule complex workflows, enhancing the automation and reliability of data processing tasks.
- Employed Kafka for real-time data streaming, enabling seamless communication and integration between various components of the data pipelines.
- Leveraged AWS EMR serverless architecture to execute PySpark jobs, optimizing resource utilization and reducing operational overhead.
- Ensured data quality and integrity by implementing data validation and error handling mechanisms within the pipelines.

03/2021 – 08/2023
Bangalore, India

Tekion India Pvt Ltd
Data Engineer
- Handling a team comprising of Data analysts, Backend and Frontend engineers to standardize all the key data being used across Tekion.
- Working on Chalice - Python Serverless Microframework for AWS to deploy backend APIs.
- Expertise in migrating legacy big data ecosystem to AWS Cloud from the scratch.
- Working extensively on Amazon EMR, PySpark, Lambda, SNS,SQS, S3, Athena, Glue, Snowflake and CloudWatch to migrate the legacy codes on AWS.
- Experience in writing PySpark scripts for data quality management and data sync.

01/2020 – 03/2021
Bangalore, India

o9 Solutions Pvt Ltd
Software Engineer
- Experience in setting up Data Lake on AWS from the scratch involving multiple source systems and building raw, refined, curated and outbound layers within the data lake.
- Expertise in fetching various logs leveraging ElasticSearch using beats and creating various reports and dashboards on Kibana.
- The data lake formation was streamlined by running transformations on transient EMR clusters leading to better cost optimisation.
- Expertise in writing PySpark scripts for the ingestion and transformation on huge raw data preparing custom data set for Advanced Analytics use cases.

| 07/2018 – 12/2020 | **Capgemini** |
| Bangalore, India | Senior Analyst |

- Experience in building large data pipelines using Spark with Scala and Python libraries in AWS cloud native environment.
- Experience in developing big data sqoop, hive, spark batch jobs leveraging AWS serverless fan out architecture using Lambda,SQS, SNS and transient EMR.
- Expertise in building python based Airflow DAGs for schedulingAWS Big data jobs in Apache Airflow and writing end to end workflow.xml and configuring property files to schedule jobs in oozie as and when required.

🎓 EDUCATION

08/2014 – 05/2018	**Bachelor of Technology**
Kurukshetra, India	NIT Kurukshetra
	CGPA : 8.7

| Kota, India | **H.S.C (12th)** |
| | Secured 87.8% |

| Kota, India | **S.S.C (10th)** |
| | Percentage: 88% |

🧠 SKILLS

- PySpark
- Python
- Kafka
- AWS Lambda
- Sqoop
- SNS

- Airflow
- Scala
- AWS EMR
- S3
- Hadoop
- SQS

- Snowflake
- SQL
- Hive
- Athena
- Cassandra
- CloudWatch

🌐 LANGUAGES

English   Hindi

📇 CERTIFICATES

- 2020 - Cloudera Certified :
  Spark and Hadoop Developer