

# Name: Nagraj Jeurkar

Position: Aws Data Engineer

Current Company: Allsafe IT Services Pvt Ltd

Total Experience : 4.1 Years

+91 8767446338

Nagrajjeurkar108@gmail.com

linkedin.com/in/Nagraj-Jeurkar-44b871151

## INTRODUCTION

---

A Data Engineer with almost 4+ years of experience working with a variety of tools and technologies and handling data at scale. I am passionate about using data to drive business decisions and enjoy working with cross-functional teams to deliver high-quality data solutions.

## PROFESSIONAL SYNOPSIS

---

- As an AWS Data Engineer, I specialize in designing, implementing, and optimizing data solutions on the AWS cloud platform. With expertise in Python, SQL, PySpark and AWS services like AWS S3, EMR, Glue, Redshift, Athena, Lambda and Apache Airflow, I excel at creating scalable and efficient data pipelines and analytics systems.
- I have command of programming languages like Python and SQL, and use them to build data models, optimize queries, and ensure high-performance data processing
- Extensive experience in big data processing using AWS EMR, effectively handling large-scale data processing and analysis tasks. Proven track record in extracting, validating, cleansing and transforming data from S3 using PySpark on AWS EMR.
- have a good knowledge AWS Glue for data Integration and ETL processes
- Utilized Amazon Athena, a serverless query service, which enables fast and cost-effective querying of data stored in s3 which allow to explore and derive insights from data in self- service manner.
- By combining Airflow with AWS EMR, I can effectively orchestrate and schedule data workflows, ensuring timely and reliable data processing.
- Have knowledge of Data Lake Architecture, in implementing data lake architecture using AWS S3 as the storage foundation by organizing data in a centralized data lake
- On-Premises Data Migration, I have experience in migrating data from on-premises Oracle databases to the AWS cloud. I am skilled in using Oracle SQL to extract, transform, and load data from the on-premises environment.
- Experience in dealing with various file formats like CSV, Json, and Parquet.
- I am committed to working in collaborative environments, partnering with cross-functional teams to understand business requirements and translate them into robust and scalable data solutions.

## TECHNICAL SKILLS

---

- **AWS Service** : AWS S3, EMR, Redshift, Lambda, Glue, Athena, Cloud Watch.
- **Programming** : Python, Pyspark, SQL
- **Automation Tools** : Airflow
- **Versioning tools** : GitHub
- **Operating Systems** : Windows, Linux
- **Ticketing Tool** : JIRA
- **IDE** : Spyder, Jupyter, Vscod

## WORK EXPERIENCE

---

Allsafe IT Services Pvt Ltd

Software Engineer (August 2019 – Present)

## EDUCATION QUALIFICATION

---

### Bachelor of Engineering

Rajiv Gandhi College of Engineering, Chandrapur

## PROJECT UNDERTAKEN

---

### Project (Current):

**Client:** Aramex

Description: Aramex is a Dubai-based multinational courier company with Jordanian origins. Founded in 1982 by Fadi Ghandour and Bill Kingson, it became the first Arab-based company listed on NASDAQ. It expanded globally, forming key partnerships and alliances, including the Global Distribution Alliance.

### Roles & Responsibilities:

- Developed end-to-end data pipelines using Amazon S3, EMR, and Redshift for seamless data flow.
- Transformed raw data using EMR, ensuring accuracy through cleansing and normalization.
- Designed scalable data lakes on Amazon S3, optimizing organization and partitioning.
- Enhanced ETL processes on EMR for faster processing through parallelization.
- On-Premises Data Migration, from on-premises Oracle databases to the AWS cloud.
- Automated workflow scheduling and monitoring via Apache Airflow.
- Implemented data quality checks, ensuring high integrity throughout pipelines.
- Managed Redshift warehouse, optimizing schema, indexing, and query performance.
- Created a comprehensive data catalog, documenting metadata and lineage.
- Monitored pipeline health and performance, enabling proactive issue resolution.
- Scaled solutions to accommodate growing data volumes and user demands.
- Collaborated with teams, delivering solutions and maintaining detailed documentation.

### Project (2<sup>nd</sup>):

**Client:** Marks & Spencer

Description: Marks and Spencer plc (M&S) is a prominent British multinational headquartered in London, specializing in clothing, beauty, home products, and food items. Founded in 1884 by Michael Marks and Thomas Spencer in Leeds, it currently operates 959 stores across the UK, including 615 food-only stores.

### Roles & Responsibilities:

- Created end-to-end data pipelines using S3, EMR, and Redshift for transformations and warehousing.
- Transformed raw data on S3 through EMR to facilitate structured analysis.
- Designed and implemented efficient Redshift data models for optimal querying.
- Integrated diverse data sources into S3 and Redshift, ensuring data accuracy.
- Enforced data quality checks to maintain integrity across S3, EMR, and Redshift.
- Optimized EMR clusters and Redshift database to enhance overall performance.
- Automated workflow scheduling and monitoring via Apache Airflow.
- Successfully led data migration projects, minimizing downtime impact.
- Proactively monitored and resolved pipeline bottlenecks and failures.
- Collaborated and documented effectively to meet diverse data requirements.

**Project (1<sup>st</sup>):****Client:** eMAG

Description: eMAG is a leading company headquartered in Romania, with a significant presence in Eastern Europe. It offers a wide array of products spanning electronics, fashion, home appliances, and more. The Company has expanded internationally, operates physical showrooms, and has diversified through acquisitions like Fashion Days and PC Garage.

**Roles & Responsibilities:**

- Gather and analyse data requirements in the eMAG.
- Write and optimize SQL queries for data extraction and analysis.
- Utilize PySpark for data integration, transformation, and analysis tasks.
- Design and implement data models for efficient storage and retrieval.
- Store processed data in AWS S3 for scalable and centralized storage.
- Use Git for managing and tracking changes to SQL and PySpark code.