

JOB DESCRIPTION

Title of position	Data Engineer
Contract type	Permanent
Job Location	Work from Anywhere in India
Department	Data CoE, APAC
Reporting to	Head of Data CoE, APAC
Job description date	25 th May 2022
Position opening start	25 th May 2022

Position Summary

Natixis is a French multinational financial services firm specialized in asset & wealth management, corporate & investment banking, insurance and payments, and a subsidiary of Groupe BPCE, the second-largest banking group in France.

In APAC Data CoE, we build enterprise-grade tech products to contribute to business growth and improve operational efficiency regionally as well as globally for Natixis CIB. Partnering with Business Lines and Support Functions, we dive into each domain and strive to deeply understand the pain-points and thereafter deliver bespoke data solutions with measurable business benefits.

As a Data Engineer, you will work closely with a fully decentralized international team (spread across APAC) of Architects, Data Scientists, UX Engineers, Product Managers, but also with business stakeholders and Fintech partners. You will bring in industry best-practices around design and implementation of data architecture on on-prem / cloud / hybrid platforms. Most importantly, as the subject matter expert you will continue adopting the latest technology and bringing the best solutions to Business.

Main responsibilities and duties of the role

- Build and maintain robust data pipelines for complex data projects with / without AI component:
 - programmatically ingesting data from different static and real-time sources (incl. web scraping)
 - define optimized schemas and data models for scalable implementation of a solution
 - log granular usage of web / mobile apps for comprehensive product analytics
- Perform ETL on large and complex datasets for AI applications - work closely with data scientists on performance optimization of large-scale ML/DL model training
- Build data tools to facilitate fast data cleaning and statistical analysis
- Industrialize ML / DL solutions and deploy and manage production services; proactively handle data issues arising on live apps
- Resolve issues escalated from Business and Functional areas on data quality, accuracy, and availability
- Work closely with APAC CDO and coordinate with a fully decentralized team across different locations in APAC and global HQ (Paris).

Technical skill requirements

- Expert in Python programming especially handling large volumes of data both in traditional and distributed environments - pandas, numpy, scikitlearn, pySpark
- Well versed with structured (SQL) and unstructured (noSQL) database in traditional and Big data environments (Hive, Spark)
- Well versed with concepts like containerization (Docker, Kubernetes, Openshift), workflow management (AirFlow), streaming data (Kafka or Spark Streaming), CI/CD pipeline (Jenkins)
- Experienced in shell scripting and writing schedulers
- Some experience with Cloud deployment (AWS / Azure / GCP) would be a big plus
- Strong knowledge of data security best practices
- And most importantly, you must be a passionate coder who really cares about building apps that can help us do things better, smarter, and faster

Job Qualification (education, years of experience and other soft skills)

- Masters or PhD degree in Computer Science or related areas from a Tier-1 university
- 1 to 2 years' experience in data engineering role
- Ability to take ownership and deliver within timelines
- Effective written and verbal communication skills