

DATA ENGINEER

Job Description

Contribute to the business value of Data-oriented products based on on-premise Data lake or on cloud environments, by implementing end-to-end data processing chains, from ingestion to API exposure and data visualization

General responsibility: Quality of data transformed in the Data lake, proper functioning of data processing chains and optimization of the use of resources of on-premise or cloud clusters by data processing chains

General skills: Experience in the implementation of end-to-end data processing chains and Big data architectures, mastery of languages and frameworks for the processing of massive data in particular in Streaming Mode (**Spark / Scala**). Practice agile methods.

Role

You will set up end-to-end data processing chains in cloud environments and in a DevOps culture, you will work on brand new products, for a wide variety of functional areas (Engineering, connected vehicle, Manufacturing, IoT, Commerce, Quality, Finance), with a solid team to support you.

Main responsibilities

- During the definition of the project
- Design of data ingestion chains
- Design of data preparation chains
- Design of basic ML algorithms
- Data product design
- Design of NOSQL data models
- Data visualization design
- Participation in the selection of services / solutions to be used according to usage
- Participation in the development of a data toolbox

During the iterative realization phase

- Implementation of data ingestion chains
- Implementation of data preparation chains
- Implementation of basic ML algorithms
- Implementation of data visualizations
- Use of ML framework
- Implementation of data products
- Exhibition of data products
- Configuration of NOSQL databases
- Distributed processing implementation
- Use of functional languages
- Debugging distributed processing and algorithms
- Identification and cataloging of reusable items
- Contribution to the evolution of work standards

We are hiring brilliant minds.



- Contribution and advice on data processing problems

During integration and deployment

- Participation in problem solving

During serial life

- Participation in the monitoring of Operations
- Participation in problem solving

Required Skills

- Expertise in the implementation of end-to-end data processing chains
- Mastery of distributed development
- Basic knowledge and interest in the development of ML algorithms
- Knowledge of ingestion frameworks
- Knowledge of Spark and its different modules
- Mastery of Scala
- Knowledge of the use of Solace
- Knowledge of the ecosystem of NOSQL databases
- Knowledge in building data product APIs
- Knowledge of Data visualization tools and libraries
- Ease in debugging Spark and distributed systems
- Popularization of complex systems
- Control of the use of data notebooks
- Expertise in data testing strategies
- Strong problem-solving skills, intelligence, initiative and ability to resist pressure
- Excellent interpersonal skills and great communication skills (ability to go into detail)
- For the future architecture from next year:
 - ***Nice to have knowledge of the GCP ecosystem DataProc, DataFlow, BigQuery, Pub-Sub, PostgreSQL/Composer, Cloud Functions, StackDriver***
 - ***Nice to have knowledge of Beam and its different execution modes on DataFlow***
 - ***Nice to have knowledge of Java***

Please apply with your most actual CV to:

Crystal System Group Ltd.

talents@crystal-system.eu

www.crystal-system.eu/careers