Data Engineer – Bioinformatics

Description

We are looking for a Data Engineer to help solve some of the key challenges around a programme of work at industrial scale with global significance. The successful Data Engineer will know how to communicate to and between technical and non-technical stakeholders as well as facilitate discussions within a multidisciplinary team including scientists, software engineers, product managers and other data engineers.

You will be contributing towards the delivery of data releases that will be used worldwide and will have experience with genetic data.

Our Future Health will be the UK's largest ever health research programme, bringing people together to develop new ways to detect, prevent and treat diseases. We are a charity, supported by the UK Government, in partnership with charities and industry. We work closely with the NHS and with public authorities across all nations and regions of the UK.

Our plan is to bring together 5 million volunteers from right across the UK who will be asked to contribute information to help build one of the most detailed pictures we have ever had of people's health. Researchers will be able to use this information to make new discoveries about human health and diseases. So future generations can live in good health for longer.

What you'll be doing

You'll be part of a multidisciplinary team that's creating pipelines that didn't exist before, owning them in production and improving them over time. Your key responsibilities will include but not be limited to:

- Supporting the build of data pipelines from data providers to our primary data store and trusted research environment.
- Producing logic for data transformation steps as code, which meets the requirements for our end users and builds well curated, accessible and quality controlled data for analysis.
- Developing prototypes for pipelines for complex transformations drawing on existing workflows developed in industry and academia.
- Keeping abreast of best practice in data engineering across industry, research and Government and facilitating the adoption of standards.
- Providing technical input into the upstream parts of the data pipeline, including the specification and transfer of data from data providers.
- Routine ad-hoc data curation activities requiring hands on development of bespoke ETL cleaning scripts using languages such as Python.
- Working with researchers to understand the data requirements and working with them to deliver the data needed for their projects.

Requirements

You will have a solid understanding and experience of bioinformatics, in particular tools and methods associated with genomic data. To succeed in this role, you will have some of the following skills:

Hiring organization
Candidate-1st

Employment Type
Full-time

Beginning of employment
asap

Job Location
London, England, United Kingdom

Working Hours
40

Base Salary
euro GBP 60K+

Date posted
May 24, 2024

- Experience of working in an agile development team following best practices like code review and pairing.
- Familiar with version control and Git/GitHub.
- Able to design, build and test pipelines using a range of different technologies. You will know how to create repeatable and reusable products.
- Experience with storing, searching and filtering large scale genomic data.
- Good understanding of cloud environments (ideally Azure), distributed computing and scaling workflows and pipelines.
- Proficient in Python.
- Experience of workflow management tools, e.g. Nextflow, WDL/Cromwell, Airflow, Prefect, Dagster).
- Understanding of common data transformation and storage formats, e.g. Apache Parquet.
- Experience of working with data lakes; experience with Spark, Databricks.
- Understanding of containerisation, e.g. Docker.
- Awareness of data standards such as GA4GH (https://www.ga4gh.org/) and FAIR (https://www.go-fair.org/fair-principles/).
- Understanding and working knowledge of information governance and data security approaches appropriate for sensitive health data.

Benefits

- Up to £60,000 per annum basic salary.
- Generous company pension package with employer contributions of up to 12%.
- 30 days annual leave (plus bank holidays).
- Continuous career development with regular appraisals and learning and development opportunities.
- A lovely new office in Holborn, Central London – we offer flexible and remote working arrangements.

Join us – let's prevent disease together.

How the process will look like

Your teammates will gather all requirements within our organization. Then, once priority has been discussed, you will decide as a team on the best solutions and architecture to meet these needs. In continuous increments and continuous communication between the team and stakeholders, you're part of making data play an even more important (and understood) part withing Brand New Day.

Job Benefits

GBP 60K+