



JOB OPENING

Data Engineer

Do what you love. Create the future you want. Explore the Data Engineer role at Nebulaworks.



About Nebulaworks

Nebulaworks is a consulting and engineering firm founded, built, and managed by engineers, for engineers. Our mission is to be the best in the world at creating high-performance engineering teams where members are inspired to collaborate openly, incentivized to gather new knowledge and skills, and value simplicity when solving difficult problems.

We believe technological and cultural change starts when the focus is on developing the team's skillset, collaboration, and process to deliver software. The tools come and go, but the right team is invaluable. They possess the foundational knowledge and expertise to reason about any problem, propose numerous reasonable solutions, and identify a low-risk path to proving their hypothesis.

Job Purpose

As a Data Engineer at Nebulaworks, you will be pivotal in managing and analyzing large datasets crucial for advancements in life sciences. This role involves collaborating with cloud engineers and scientists, providing essential data management, machine learning, and data lifecycle expertise to support our cutting-edge research and discovery projects.

You Will

- Build distributed, scalable, and reliable data pipelines that ingest and process data at scale and in real time.
- Collaborate with other teams to design, develop, and deploy data tools that support both operations and product use cases.
- Perform offline analysis of large data sets using components of a big data software ecosystem. Evaluate and advise on technical aspects of open work requests in the product backlog with the product lead.
- Own product features from the development phase through to production deployment. Evaluate big data technologies and prototype solutions to improve data processing architecture.
- Develop and maintain specialist knowledge of database concepts, object and data modeling techniques, and design principles, and a detailed knowledge of database architectures, software, and facilities.
- Apply Agile methodologies in project management.
- Deploy and maintain database storage Infrastructures in AWS cloud.

- Troubleshoot and determine the root cause of complex data provenance, metadata issues, and engineering questions that may involve interfacing with various technical staff in multiple organizations and with differing levels of expertise.
- Investigate, evaluate, test, and recommend technical solutions for future systems.
- Develop tools and procedures to monitor and automate system tasks on servers and clusters.
- Creates data environments and/or data sets to serve a wide range of data users, including but not limited to Data Scientists, Data Analysts, Business Analysts etc
- Code/Build/Deploy - git, jenkins, AWS
- Handle large, complex data sets to meet functional and non-functional business requirements.
- Implement methods for enhancing data reliability, efficiency, and quality.
- Use programming languages such as Python (must) and R for data tasks.
- Expertise in RDBM, NoSQL, JSON, and other database technologies.
- Develop and maintain machine learning models aligned with business objectives.
- Collaborate with data scientists in implementing advanced machine learning algorithms.
- Adapt to dynamic working environments, managing multiple priorities.

Job Purpose

As a Data Engineer at Nebulaworks, you will be pivotal in managing and analyzing large datasets crucial for advancements in life sciences. This role involves collaborating with cloud engineers and scientists, providing essential data management, machine learning, and data lifecycle expertise to support our cutting-edge research and discovery projects.

Requirements

- You value driving all changes through version control
- You value all team members being hands-on and directly contributing to business outcomes
- Proven experience as a Data Engineer or in a similar role.
- Strong understanding of machine learning principles and application in data processes.
- Experience in building and deploying machine learning models.
- Expertise in Python, Object Oriented Programming and SQL database management.
- Familiarity with AWS cloud-based solutions.

- Excellent problem-solving, communication, and analytical skills.
- Bachelor's degree in Data Science, Computer Science, IT, or related field; Master's is a plus.

Nice to Haves

- Expertise in scientific data, especially within life sciences, is a significant plus.
- Data engineering or machine learning-related certification
- Mastered Continuous Integration, Continuous Delivery, and Continuous Deployment approaches and processes
- Experience with functional programming languages (Haskell, etc.)
- Experience with containers and container platforms (Kubernetes (EKS, GKE, AKS), docker, etc.)
- Experience across multiple major cloud platforms (AWS, Azure, GCP)
- Expertise in advanced deployment patterns (blue/green, feature flags, canary, etc.)
- Experience with Twelve-Factor application architectures
- Experience with managing the large-scale infrastructure supporting numerous teams and applications

Why Apply:

- Hybrid work: Irvine office and work from home
- Competitive salary and compensation package
- 401K with match
- Employee Incentive Plan (EIP)
- Health and dental, medical coverage

Posting Statement:

- All applicants must be legally permitted to work in the United States without a visa; we cannot provide new or continuing visa sponsorship opportunities at this time.
- Nebulaworks is an Equal Employment Opportunity employer, and all qualified applicants will receive consideration for employment without regard to race, color, religion, sex, age, national origin, protected veteran status, disability status, sexual orientation, gender identity or

expression, marital status, genetic information, or any other characteristic protected by law.

- We request candidates to scrub such information from their resumes before applying to help make Nebulaworks an inclusive environment for everyone.

To apply:

```
$ ssh -p 23234 <github_user>@sshapply.nebulaworks.com  
$ scp -P 23234 <resume>.pdf <github_user>@sshapply.nebulaworks.com:resume.pdf
```

Note: You must use an SSH key that's associated with the same user in github