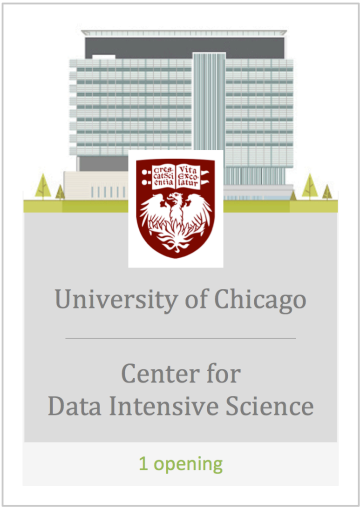
Lead Software Engineer for the
NIH/NCI Genomic Data Commons Project

The University of Chicago, Center for Data Intensive Science

Chicago, IL (relocation offered) (visa sponsorship offered)



Job Description

The Center for Data Intensive Science (CDIS) is looking for a Lead Software Engineer for the National Cancer Institute (NCI) Genomic Data Commons (GDC) project. Alongside our team of experts, the Lead Software Engineer will develop the scalable and interoperable software stack for the GDC, supporting upwards of 5 petabytes of centralized and harmonized cancer genomic data.

The GDC will provide an open source, scalable, modern informatics framework that uses community standards to make raw and processed genomic data broadly accessible. The GDC will harmonize and centralize existing petabyte-scale cancer genomics datasets in a large scale computing infrastructure. The GDC will also enable previously infeasible collaborative efforts between scientists.

The GDC serves as a key step toward the development of precision medicine, targeted treatments that are tailored to individual patients. Once fully developed, it will provide an interactive system for researchers and clinicians to upload genomics data and use it to identify the molecular subtype of cancer and potential therapeutic targets. Genetic data will be linked to extensive clinical information from patients and their response to treatment.

The Lead Software Engineer for this project manages all aspects of programming projects, including requirements, design, implementation, deployment/delivery, and support. Leads team efforts and oversees the work of other software engineers. Provides technical oversight and develops standards, guidelines, and processes for applications. Reviews the design and code development of key architectural components. Contributes to decisions on project and infrastructure needs, including the evaluation of server technologies, languages, platforms, and frameworks. Develops timelines, technical diagrams, project plans, and resources allocation in an agile methodology. Works with cloud computing infrastructures including OpenStack, Amazon AWS and Google Cloud to design, develop, maintain, and evaluate software applications to meet business and technical requirements. Works in Linux-based systems primarily with Python and C/C++, with the ability to work with other programming languages as the need arises. Oversees code testing and ensures appropriate standards are met. Works with users, collaborators, and technical staff to resolve problems and respond to feedback regarding potential improvements and enhancements. Ensures appropriate documentation. Serves as a liaison with internal and external collaborators on multiple research projects.

Research includes the full stack from systems to algorithms to user interfaces. Research projects span management, sharing, and provenance of large data sets; resource allocation and scheduling for cloud computing, large scale pipelining of next-generation sequence analysis, transfer programs/protocols for high-speed networks and resource visualization.

Skills & Requirements

Education

- Bachelor's degree in computer science, mathematics, statistics, engineering, or a related field required.
- Advanced degree in mathematics, computer science, engineering, or a related field preferred.

Experience

- Minimum five (5) years of relevant programming experience required.
- Experience with Python, C/C++, Java, or Ruby required.
- Unix/Linux experience required.
- Version control experience required.
- Experience with full design life cycle required.
- High performance/cloud computing experience preferred.
- Unix/Linux programming or system administration experience preferred.
- UX/UI experience preferred.
- Git version control experience preferred.
- Experience with genomics preferred.
- Experience creating development specifications, use cases, and other development related documentation preferred.

Competencies

- High performance/ cloud computing experience preferred.
- Unix/Linux programming or system administration experience preferred.
- UX/UI experience preferred.
- Git version control experience preferred.
- Experience with genomics preferred.
- Experience creating development specifications, use cases, and other development related documentation preferred

Apply for Requisition#098542 at https://jobopportunities.uchicago.edu.

Join us in transforming cancer research

The Center for Data Intensive Science is developing the emerging field of data science with a focus on applications to problems in biology, medicine, and health care. Our vision is a world in which researchers have ready access to the data and tools required to make discoveries that lead to deeper understanding and improved quality of life. We democratize access, speed discovery, create new knowledge and foster innovation through implementation using data at scale. Our scientific data clouds and commons include:

The Genomic Data Commons (GDC) is a comprehensive computational facility to centralize and harmonize cancer genomic data generated from NCI-funded programs. The GDC is the foundation for a genomic precision medicine platform and will enable the development of a knowledge system for cancer.

The Bionimbus Protected Data Cloud (PDC) is the first open-source cloud-based computational platform that allows researchers authorized by NIH to compute over human genomic data in a secure and compliant fashion. Bionimbus and related cloud-based infrastructure are used by researchers working on cancer, diabetes and neuropsychiatric disorders.

The Open Science Data Cloud (OSDC) provides the scientific community with resources for storing, sharing, and analyzing terabyte and petabyte-scale scientific datasets. The OSDC is a data science ecosystem in which researchers can house and share their own scientific data, access complementary public datasets, build and share customized virtual machines with whatever tools necessary to analyze their data, and perform the analysis to answer their research questions.

The Biomedical Data Commons (BDC) is cloud-based infrastructure that we are developing for a consortium of medical research centers and commercial partners that provides secure, compliant cloud services for managing and analyzing genomic data, electronic medical records (EMR), medical images, and other PHI data. It provides resources to researchers so they can more easily make discoveries from large complex controlled access datasets.

We have excellent benefits

Health & Welfare Benefits

- 5 Medical Plans
- 2 Dental Plans
- Vision Plan
- Healthcare Flexible Spending
- Health Savings Account
- Life and Disability Insurance
- Long Term Care Insurance
- On-Site Childcare Program
- Staff and Faculty Assistance Program
- Childcare Referral Services
- Elder Care Referral Service
- Retiree Medical Insurance

Tuition Benefits

- Professional Development
- Staff Education Assistance Program
- Dependent Children Education Assistance Program

Retirement & Financial Benefits

- Retirement Income Program
- Supplemental Retirement Program
- Staff Pension Plan
- Employer-Assisted Housing Program
- Dependent Care Flexible Spending
- Qualified Transportation Benefits
- Perks and Discounts

Vacation, Holidays & Time Off Benefits

- 8 Paid Holidays
- 3 Weeks Vacation Time
- 5 Personal Holidays
- 2 Weeks Sick Time
- Leave of Absence Programs
- Short & Long Term Disability

Apply for Requisition#098542 at https://jobopportunities.uchicago.edu.

This position is grant supported and longevity of the position is dependent upon future funding.
The University of Chicago is an Affirmative Action / Equal Opportunity / Disabled / Veterans Employer.

THE UNIVERSITY OF CHICAGO | Center for Data Intensive Science