EMPLOYMENT OPPORTUNITY Be a part of something BIG.

DATA ENGINEER

JOB OPENING ID 342961

JOB DESCRIPTION

IPUMS is a leader in the field of quantitative social science research and the largest disseminator of census and demographic data to the world's academic research community. Or, to put it another way - we're on a mission to gather, process, link and publish billions of records spanning hundreds of years and more than 100 countries so that demographers, historians, economists, environmental scientists, journalists, policymakers, and others around the globe can use the data to do amazing research and make the world a better place.

The IPUMS IT group supports this mission by using leading open source tools to solve complex data and computation challenges and build reliable, scalable data dissemination systems. Your work will be highly visible and will contribute directly to the overall success of our organization. IPUMS is part of the University's Institute for Social Research and Data Innovation. IPUMS and its affiliated units support the work-life balance of our staff with 40-hour work weeks, flexible work hours, partial remote work schedules, and generous vacation and sick leave benefits. The University also offers excellent health insurance, tuition assistance, and retirement benefits. IPUMS IT has a robust professional development fund for staff training and development.

Diversity and inclusion are core values of our organization; we are committed to creating a work environment that celebrates and promotes diversity, and advances a culture of inclusion for persons from historically excluded and underrepresented groups. We strongly encourage members of underrepresented groups to apply.

RESPONSIBILITIES

We are currently seeking a data engineer to join our Data Team. The Data Team is responsible for creating software that manages large scale data manipulation within complex data structures. You will be working in close collaboration with expert historians, demographers, and data scientists. We use Python for general data processing, along with R, Rust, Apache Spark, and C++.

Some examples of software your team will be responsible for are:

- The data conversion program, which transforms thousands of census and survey datasets from over a hundred different countries into well documented and easily comparable forms

- The historical record linkage program, which links hundreds of millions of U.S. census records across time using machine learning to match on fuzzy characteristics such as names and dates of birth
- The workspaces program which version controls terabytes of data and metadata across 13 distinct projects

This role will work throughout the project life cycle, from architecture and design through implementation and testing to deployment and support. We practice Agile and collaborative software development, which means you will have ample opportunity to work alongside both developers and researchers. Specific duties include:

20% Software Architecture and Design. Working in partnership with researchers and data scientists defining business needs

50% Software Implementation and Analysis. Coding, refactoring, testing, executing and analyzing data pipelines in a cross-functional, agile team environment.

20% Deployment and Support. Working with the operations team to build out high performance infrastructure to support new data pipelines. Developing deployment processes to continuous integration, internal, and production environments. Providing user support to the team and our researchers.

10% Other duties as assigned. Professional development activities, participation in IT working groups, and other tasks as assigned.

This position will have an annual starting salary of $80,000+, commensurate with experience.

QUALIFICATIONS

Required Qualifications

- BA/BS degree plus two years of work experience in the areas of software development with a related BA/BS degree OR related master's degree.
- Excellent ability to plan and manage complex tasks/projects
- Demonstrated excellent oral and written communication skills for both technical and non-technical academic audiences.
- This role requires a desire to:
    - Work in close collaboration with multiple team members
    - Work on large dataset manipulation and complex data models
    - Work on high performance computing

Candidates with some of the following skills and/or experience are preferred:

- Developing software for high performance and/or distributed computing
- Databases and SQL
- Large dataset manipulation and complex data models
- A variety of data serializations
- Modern data science techniques and approaches
- Python, R, Rust, or C/C++
- Apache Spark or the Hadoop ecosystem
- Prior work experience in a research and/or higher education setting
- Four years of work experience in the areas of software development with a non-related BA/BS degree.

APPLICATION PROCEDURE

Please apply using the University of Minnesota's online employment system. External applicants can apply directly here: https://hr.myu.umn.edu/jobs/ext/342961. Current University of Minnesota employees can apply directly here: https://hr.myu.umn.edu/jobs/int/342961. Application requirements include a resume, and a cover letter describing your interest and qualifications in the position. Questions concerning the application process may be addressed to Alex Lunde, HR & Operations Manager, at isrdi-jobs@umn.edu.

Any offer of employment is contingent upon the successful completion of a background check. Our presumption is that prospective employees are eligible to work here. Criminal convictions do not automatically disqualify finalists from employment.