Position: Senior Data Architect

About the California Policy Lab
The California Policy Lab (CPL) creates data-driven insights for the public good. Our mission is to improve the lives of Californians by generating evidence that transforms public policy. We do this by forming lasting partnerships between government and California’s flagship public universities to harness the power of research and administrative data. We work on California’s most urgent issues, including homelessness, poverty, criminal justice, and education inequality. We facilitate close working partnerships between policymakers and researchers at UC Berkeley and UCLA to help evaluate and improve public programs through rigorous empirical research.

CPL has also developed a revolutionary new research infrastructure that removes barriers to doing applied policy research and unlocks the power of administrative data in the world’s fifth largest economy. We are changing the landscape of research -- and more importantly, we are changing lives.

CPL operates as one Lab across two sites. Each site is led by an Executive Director, who is responsible for daily operations, and a Faculty Director, who provides strategic direction and also leads some Lab projects.

CPL recognizes the value of having a diverse staff at all levels of the organization. We are looking for equity-minded applicants who represent and understand the rich diversity of California and who demonstrate a sensitivity to and understanding of the diverse academic, socioeconomic, cultural, disability, gender identity, sexual orientation, and ethnic backgrounds present in California. When you join our team, you can expect to be part of an inclusive and equity-focused community.

The Position
The California Policy Lab is recruiting for a Senior Data Architect to be based at UC Berkeley and serve the Lab as a whole. This full-time position is open until filled, with an expected start date in Q2 2021.

The Senior Data Architect will join our growing team at a critical juncture in our development. We launched the Lab in January 2017 and now have nearly three dozen staff across our two sites. The Senior Data Architect will be the principal engineer of our back-end systems. We are seeking someone who can see the big picture and also get down and dirty
to help users with individual needs. The Senior Data Architect will build and manage all our databases (over 80 Terabytes in data, and growing!), and tune them for optimal performance. They will also work with our IT Manager to optimize our virtual enclave environments. They will also work with data users to optimize analyses and create new data schemas and data linkages. The ideal candidate is an experienced engineer with a strong background in designing and optimizing SQL databases and will also have the requisite expertise to guide in the potential adoption of new technologies, such as cluster computing, cloud alternatives, and the like.

**Responsibilities**

- Works with users to understand the data, clean it, work with it effectively and efficiently, and extract insights from it.
- Implements workflows to maximize security and efficiency, including developing ETL pipelines for data updates, data extract routines for users, and data cleaning routines in concert with CPL's other data analysts.
- Maintains and implements a data management plan for CPL's principal datasets. Maintains clear inventories and other organizational assets.
- Designs relational databases (PostgreSQL, MS SQL Server) for sensitive administrative datasets. Researches and adapts existing technologies for CPL's use case.
- Optimizes CPL's data infrastructure for maximum analytical capabilities, in collaboration with the IT Manager. Includes tuning of databases (PostgreSQL, MS SQL Server) and virtualization (VMware), resource allocation, and recommending appropriate scaling practices, including potentially cluster computing solutions such as Hadoop, Spark, or similar.
- Maintains metadata and documentation (using Confluence, DDI, and other technologies) and works with users to improve consistent documentation about CPL's data.

**Required Qualifications**

- Ability to prepare data models and database schemas unassisted.
- Deep experience with how to efficiently analyze large datasets (>10TB).
- Thorough knowledge of database design and database performance tuning, specifically with PostgreSQL or MS SQL Server.
- Experience with virtualized environments (VMware), computing resource allocation, and strategies for speeding up analysis of large datasets.
- Thorough knowledge of data management systems, practices and standards.
- Demonstrated ability to work with others from diverse backgrounds. Demonstrated effective communication and interpersonal skills. Demonstrated service orientation skills.
- Demonstrated ability to communicate technical information to technical and non-technical personnel at various levels in the organization, including offering technical support to analysts who are new to using SQL databases.
- Self-motivated and works independently and as part of a team.
- Demonstrated strong problem-solving skills. Able to learn effectively and meet deadlines.
• Strong organizational skills. Ability to understand and apply a complex compliance framework.
• Strong analytical and design skills, including the ability to abstract information requirements from real-world processes to understand information flows in computer systems.
• Ability to represent relevant information in abstract models. Critical thinking skills and attention to detail.
• Deep knowledge and facility with SQL and knowledge of, or the ability to learn, other programming languages (e.g., STATA, SAS, Python, R).
• Thorough knowledge of database permissioning and user-based access security.

Preferred Qualifications
• Knowledge of cloud and cluster-based big data solutions, such as Hadoop, Spark, and NoSQL platforms.
• Knowledge of entity resolution/record linkage methods, including privacy-preserving record linkages.

Salary & Benefits
This position is full-time, and will start as a two-year contract. Salary is commensurate with experience. Hiring range is $75,400 - $110,800/annually.

For information on the comprehensive benefits package offered by the University visit: https://ucnet.universityofcalifornia.edu/compensation-and-benefits/index.html

How to Apply
Go to https://jobs.berkeley.edu/job-listings and click “External Applicants” (or “Internal” if you’re a current UC Berkeley employee) and then search for keyword “14992”, which is the job ID. Use the system to submit your cover letter and resume as a single attachment.

This is a designated position requiring fingerprinting and a background check due to the nature of the job responsibilities. Berkeley does hire people with conviction histories and reviews information received in the context of the job responsibilities. The University reserves the right to make employment contingent upon successful completion of the background check.

The University of California is an Equal Opportunity/Affirmative Action Employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability, or protected veteran status. For more information about your rights as an applicant see: https://www.eeoc.gov/sites/default/files/migrated_files/employers/poster_screen_reader_optimized.pdf

For the complete University of California nondiscrimination and affirmative action policy see: http://policy.ucop.edu/doc/4000376/NondiscrimAffirmAct