

Building Service 32BJ Benefit Funds

Job Description Posting

Health Fund

Job Title: Data Engineer
Grade: 12
Department: Health Services
Reports To: Executive Office
FLSA Status: Non-Union/Non-Management

Summary: As a Data Engineer you will get to play a key role in the delivery of powerful data-driven products that support a self-insured health fund for union members who work in residential and commercial buildings across the East Coast. The Data Engineer is responsible for data collection, movement, storage, transformation processing, and storage of large datasets. This individual works with both current ETL/Data Warehousing and future Data/Streaming/Pipeline architectures. The focus is on choosing optimal solutions to use for these purposes, then implementing, maintaining, and monitoring them.

Essential Duties and Responsibilities:

- Program in a variety of languages and platforms to automate the processing of patient-level healthcare transactions, third party data sources and aggregated public health data;
- Serve as the data wrangler and ETL expert for the company; Ingest, transform, cleanse and augment internal and external data assets;
- Build algorithms for fuzzy matching, de-duplication and rule-based de-identification; Fully indulge your love for math, statistics and logical problem solving;
- Leverage the main toolsets: Python, Anaconda stack (Jupyter Notebook, NumPy, Pandas, Matplotlib/Bokeh, SciPy, Scikit-Learn), SQL, Stata, Qlikview;
- Lead data modeling, database design and performance optimization; Write SQL for defining database objects and performing manipulations;
- Continuously learn by investigating and adopting new technologies;
- Facilitates data collection from a variety of different sources, getting it in the right formats, assuring that it adheres to data quality standards, and assuring that downstream users can get that data quickly and with a common standard interface;
- Ensures that data streams/pipelines are scalable, repeatable, and secure, and can serve multiple users within the Health Fund;
- Develops as a core member of an Agile team, using Agile tools and methodology; Work closely with other team members including Application Developers, Database Developers, and Data Scientists;
- Responsible for creating the infrastructure that provides insight from raw data and handles diverse sources of data seamlessly;
- Enables big data and batch/real-time analytical solutions that leverage emerging technologies;
- Additional responsibilities include developing prototypes and proof of concepts for the selected solutions, and implementing complex big data projects with a focus on collecting, parsing, and managing large sets of data using multiple platforms to allow for research and data science initiatives;
- Translates business requirements into modern data pipeline solutions; Create centralized documents and diagrams of all solutions;
- Creates a data catalog store of all metadata;
- Approaches all relationships with a world-class customer service approach; Maintains a customer-focused approach with users to provide solutions that are science/research-driven;
- Responsible for the integrity and security of data in all forms of storage throughout the Data Architecture;
- Works with other professionals throughout the Funds effectively; Comply with HIPAA to follow all applicable policies and procedures;
- Assists in the development of standards and procedures affecting data management, design and maintenance; Documents all standards and procedures;
- Provides presentations and training to other team members in the above;
- Possesses an extremely flexible attitude; Willing to work with multiple types of technologies and languages with an open mind and without technology bias; Continuous interest in updating skill sets and knowledge of trends in the Data Technology space; and
- Other duties as assigned.

Qualifications:

- Bachelor's degree in Computer Science or a related discipline; Advanced degree preferred;
- 5+ years of full-time experience or demonstrated accomplishments in relevant subject areas;
- 2 years of relevant professional development experience
- Demonstrated mastery with Python and its data science ecosystem;
- Stata is a must; Knowledge of other mathematics environments such as R,SAS, SPSS or Matlab is a plus;
- High level of expertise in SQL, relational database optimization, stored procedures and data modeling;
- Familiarity with Subversion or other Version Control Management Tools;
- Understanding of methods to ingest and process non-relational JSON and XML formatted data.
- Strong SQL and NoSQL Database Knowledge: Oracle, PostgreSQL/MYSQL, and Mongo DB (or similar);
- Experience with micro-services and SOA;
- Knowledge of Hadoop, Spark, Kafka and other big data technology stacks and streaming tools is a plus;
- Familiarity with and the ability to leverage a wide variety of open source technologies and tools;

- Experience with serverless computing, creating VMs, cloud security, and other cloud services is also a plus;
- Experience with scalable cloud data services; Azure preferred;
- Experience with Dynamics 365 cloud eco-system a plus;
- Familiarity with concepts used by ETL tools (such as SSIS, Informatica and Talend) is a plus, but an ability to create more purpose-built solutions by leveraging open source tools;
- Desire to be an expert in healthcare and passionate about making an impact in this field;
- Experience with healthcare claims and clinical data is beneficial; and
- Understanding of HIPAA compliance.

Candidates may submit their cover letter and resume to jobpostings@32bjfunds.com prior to closing date. Only those candidates under consideration will be contacted. No phone calls please.