**Data Engineer**

**Location: Remote – US based**

**Sponsorship: This role does not offer visa sponsorship, but any type of work visa is required (US Citizen, Greengard/Permanent Resident, H4B, J2, etc.)**

**Overview:**

Powerlytics is a cutting edge, well-funded, venture-backed company that assists their Financial Services customers to provide a frictionless experience for their clients while maximizing revenue, minimizing risk, and discovering new market opportunities. If working in a fast paced, innovative atmosphere, where creativity is a driving theme, come join us as a Principal Data Engineer. Our client base includes top 5 banks, insurance companies, asset managers as well as alternative lenders, marketing firms and global consulting firms, among others. You have the opportunity to work on one of the most unique and proprietary data sets available in the US. With a portfolio of products and solutions underpinned by predictive analytics powered by proprietary databases of the anonymized tax returns of all households (150 million+) and for-profit businesses (30 million+), you will have an opportunity to work with big data and large-scale analytics systems for our team’s innovative data creation and analytics needs.

This is an outstanding opportunity to join a focused team and work collaboratively to make a significant impact on our organization and those we serve as well as substantially grow your own skills. You will report directly into our Chief Data and Analytics Officer, Narahara Chari Dingari, PhD. You will also work with our team of PhD Data Scientists and our CTO. Chari has mentored and led numerous Data Engineers and Data Scientists in Banking, Financial and Insurance domains. He published several high-impact journal articles and has been granted patents on machine learning techniques.

You will contribute to a large-scale data platform and provide end-to-end analytics solutions to transform rich data into actionable insights. Your data sets will help us determine important features, monitor data product launches and understand data usage and quality in detail. We are looking for Data Engineers who love to build end-to-end analytics solutions for their customers and can produce high quality data artifacts at scale. The ideal candidate has experience implementing extendable frameworks on top of Spark and understands the inner workings of Spark execution.

**Key Qualifications:**

* 6+ years of deep experience with large scale distributed big data systems, pipelines, and data processing in an industry setting.
* Practical hands-on programming and engineering experience in Python and R.
* Experience with Python modules such as Airflow, Pandas, PySpark, H2O, scikit-learn, XGBoost.
* Proven experience using distributed computer frameworks, such as Apache Spark
* Able to setup large scale data pipelines and data monitoring systems to make sure overall pipeline is healthy.
* Willing to take ownership of pipeline and can communicate concisely and persuasively to a varied audience including data provider, engineers, and analysts.
* Ability to identify, prioritize, and answer the most critical areas where analytics and modeling will have a material impact.
* Experience in stream data processing and real time analytics of data generated from user interaction with applications is a plus.
* Prior experience with Linux, cloud data engineering and machine learning platforms, preferably AWS SageMaker, AWS EMR, or AWS Lambda.
* Aptitude to independently learn new technologies.
* Excellent verbal and written communication skills is required.

**Description:**

As a Data Engineer, you will own significant responsibility in crafting, developing, and maintaining our large-scale ETL pipelines, storage, and processing services. The team is looking for a self-driven data engineer to help design and build data pipelines that allow our team to develop high quality, scalable data assets. In addition to the design and implementation of this infrastructure, you will be responsible for communicating with data scientists and other team members to determine the most effective models to improve data access, promote econometrics research, and eventually ship groundbreaking features to our customers.

**Education & Experience:**

BS or MS in Computer Science, Engineering or equivalent. Master's degree preferred.

**Additional Requirements:**

* Passion for building trustworthy, reliable, stable, and fast data products that serve customer needs.
* Software engineering experience and discipline in design, test, source code management and CI/CD practices.
* Experience in data modeling and developing SQL database solutions.
* Deep understanding of key algorithms and tools for developing high efficiency data processing systems.
* Experience in Financial Service sectors such as Banking, Insurance or Wealth Management is a plus.

**Compensation:**

Position offers a very competitive compensation package consisting of a base salary, bonus and equity as well as a benefits package.

**To Apply:**

Please send a message with your resume to [employment@powerlytics.com](mailto:employment@powerlytics.com).  Briefly tell us why you are interested in this role and how you meet the requirements listed above. Please include the number of years of Spark and other programming experience from an industry setting you bring as well.  Please also confirm that you are legally authorized to work in the US and do not require a work visa now, or in the future.