



Role: AI

About AI71:

AI71 is an industry leader in artificial intelligence, delivering innovative solutions that empower developers, businesses and governments to solve complex challenges. AI71 builds secure, enterprise-ready applications powered by cutting-edge technology—tailored for knowledge workers and sector-specific needs. AI71 bridges the gap between advanced AI and real-world impact. Guided by a strong commitment to research and responsibility, we create transformative solutions that drive progress and empower communities.

The Role:

As a Senior Data Engineer, you will be responsible for designing, developing, and maintaining advanced, scalable data systems that power critical business decisions. You will lead the development of robust data pipelines, ensure data quality and governance, and collaborate across cross-functional teams to deliver high-performance data platforms in production environments. This role requires a deep understanding of modern data engineering practices, real-time processing, and cloud-native solutions.

What You'll Do:

- **Data Pipeline Development & Management:**
 - Design, implement, and maintain scalable and reliable data pipelines to ingest, transform, and load structured, unstructured, and real-time data feeds from diverse sources.
 - Manage data pipelines for analytics and operational use, ensuring data integrity, timeliness, and accuracy across systems.

- Implement data quality tools and validation frameworks within transformation pipelines.
- Data Processing & Optimization:
 - Build efficient, high-performance systems by leveraging techniques like data denormalization, partitioning, caching, and parallel processing.
 - Develop stream-processing applications using Apache Kafka and optimize performance for large-scale datasets.
 - Enable data enrichment and correlation across primary, secondary, and tertiary sources.
- Cloud, Infrastructure, and Platform Engineering:
 - Develop and deploy data workflows on AWS or GCP, using services such as S3, Redshift, Pub/Sub, or BigQuery.
 - Containerize data processing tasks using Docker, orchestrate with Kubernetes, and ensure production-grade deployment.
 - Collaborate with platform teams to ensure scalability, resilience, and observability of data pipelines.
- Database Engineering:
 - Write and optimize complex SQL queries on relational (Redshift, PostgreSQL) and NoSQL (MongoDB) databases.
 - Work with ELK stack (Elasticsearch, Logstash, Kibana) for search, logging, and real-time analytics.
 - Support Lakehouse architectures and hybrid data storage models for unified access and processing.
- Data Governance & Stewardship:
 - Implement robust data governance, access control, and stewardship policies aligned with compliance and security best practices.
 - Establish metadata management, data lineage, and auditability across pipelines and environments.
- Machine Learning & Advanced Analytics Enablement:
 - Collaborate with data scientists to prepare and serve features for ML models.
 - Maintain awareness of ML pipeline integration and ensure data readiness for experimentation and deployment.
- Documentation & Continuous Improvement:

- Maintain thorough documentation including technical specifications, data flow diagrams, and operational procedures.
- Continuously evaluate and improve the data engineering stack by adopting new technologies and automation strategies

What You'll Bring:

- 8+ years of experience in data engineering within a production environment.
- Advanced knowledge of Python and Linux shell scripting for data manipulation and automation.
- Strong expertise in SQL/NoSQL databases such as PostgreSQL and MongoDB.
- Experience building stream processing systems using Apache Kafka.
- Proficiency with Docker and Kubernetes in deploying containerized data workflows.
- Good understanding of cloud services (AWS or Azure).
- Hands-on experience with ELK stack (Elasticsearch, Logstash, Kibana) for scalable search and logging.
- Familiarity with AI models supporting data management.
- Experience working with Lakehouse systems, data denormalization, and data labeling practices.
- Bachelor's or Master's degree in Computer Science, Engineering, Data Science, or a related field.
- Demonstrated success in designing, scaling, and operating data systems in cloud-native and distributed environments.
- Proven ability to work collaboratively with cross-functional teams including product managers, data scientists, and DevOps.

Great Pluses / Preferred Experience

- Working knowledge of data quality tools, lineage tracking, and data observability solutions.
- Experience in data correlation, enrichment from external sources, and managing data integrity at scale.
- Understanding of data governance frameworks and enterprise compliance protocols.

- Exposure to CI/CD pipelines for data deployments and infrastructure-as-code.

Why AI71:

- **Mission-Driven Work:** Work on cutting-edge AI applications with a talented and passionate team, solving real-world challenges in critical sectors.
- **Unparalleled Opportunity:** This is a chance to innovate and solve real-world challenges using AI at a company with unique access to world-leading models and resources.
- **Career Growth:** We offer competitive compensation, benefits, and significant career growth opportunities as a foundational member of the team.
- **World-Class Environment:** Enjoy a flexible working environment and the latest tools & technologies needed to do your best work.