ML Ops Engineer

**Location:** Pune

# Job Description

ML Ops Engineer

Pune

What You’ll Do:

Data Infrastructure Reliability:

Build and maintain highly reliable and scalable data infrastructure, ensuring data availability and integrity across various AI/ML workflows.

Design and optimize data pipelines for real-time and batch processing that support training and inference for machine learning models.

Implement fault-tolerant and scalable data storage solutions, ensuring minimal downtime and performance bottlenecks.

Cloud and On-premises Infrastructure Management:

Oversee cloud-based (AWS, GCP, Azure) and on-premises infrastructure used for data storage, data processing, and machine learning workloads.

Ensure that data engineering platforms and tools (such as Apache Kafka, Spark, Hadoop, and similar technologies) are scalable and available to handle large-scale datasets and workloads.

Manage and monitor data storage solutions such as S3, HDFS, BigQuery, Redshift, or similar technologies for performance and cost optimization.

Monitoring and Automation:

Implement comprehensive monitoring and alerting systems for all data pipelines, storage, and processing environments.

Automate operational tasks to reduce manual intervention and improve efficiency, including pipeline orchestration and data flow management using tools like Apache Airflow, Kubeflow, or similar.

Develop automated health checks and system recovery protocols to minimize downtime and system failures.

Collaboration with Data Science & Engineering Teams:

Work closely with data engineers and ML teams to ensure seamless integration of data sources, models, and computational resources.

Help deploy and maintain machine learning models in production environments by ensuring that necessary data transformations, model inputs, and outputs are handled reliably.

Ensure data is accessible in the required formats, structures, and latencies for machine learning training and inference.

Incident Response and Troubleshooting:

Lead incident response efforts to quickly resolve issues affecting the availability and performance of data pipelines and machine learning services.

Identify root causes of system failures and implement measures to prevent recurrence, driving improvements in system design and operational processes.

Conduct post-mortem analyses of incidents, create actionable reports, and implement preventative measures.

Security and Compliance:

Ensure compliance with data privacy regulations (GDPR, CCPA, HIPAA, etc.) and secure management of sensitive data throughout its lifecycle.

Implement and enforce security best practices, including encryption, access controls, and audit trails, to protect data in transit and at rest.

Optimization and Cost  Efficiency:

Identify opportunities to optimize data processing and storage costs while maintaining high performance and availability.

Leverage cloud cost management tools to optimize resource usage and ensure that data infrastructure is both cost-effective and performant.

On your first day, we'll expect you to have:

3+ years of experience as an SRE, DevOps Engineer, Data Engineer, or similar role, with a focus on managing large-scale data pipelines and infrastructure.

Experience with data engineering tools and technologies such as Apache Kafka, Apache Spark, Hadoop, Airflow, and Kubernetes.

Proven experience managing cloud platforms (AWS, GCP, Azure) and data storage technologies such as S3, BigQuery, Redshift, and HDFS.

Hands-on experience with building and optimizing data pipelines and real-time streaming solutions for machine learning applications.

Expertise in programming and scripting languages such as Python, Bash, Go, or similar.

Strong understanding of distributed computing, parallel processing, and storage architectures.

Experience in infrastructure automation tools (e.g., Terraform, Ansible, CloudFormation).

Proficiency in containerization and orchestration technologies (Docker, Kubernetes).

Knowledge of monitoring tools (e.g., Prometheus, Grafana, Datadog) and observability practices.

Excellent troubleshooting and problem-solving skills, especially in high-pressure situations.

Strong communication skills and the ability to collaborate effectively with cross-functional teams, including data science, ML engineers, and infrastructure teams.

A proactive mindset with a focus on continuous improvement, automation, and scalability.

Bachelor’s degree in Computer Science, Engineering, Data Science, or a related field, or equivalent practical experience.

We'd be super excited if you have:

Experience with machine learning lifecycle management tools (e.g., MLflow, Kubeflow).

Familiarity with serverless computing frameworks and data engineering tools.

Expertise in working with GPU and distributed computing frameworks (e.g., TensorFlow on GPUs).

Experience with performance tuning and optimization for big data and AI/ML workloads.

Benefits:

At Calfus, we value our employees and offer a strong benefits package. This includes medical, Group, and parental insurance, coupled with gratuity and provident fund. Further, we support employee wellness and provide birthday leave as a valued benefit.

Calfus Inc. is an Equal Opportunity Employer.

We believe diversity drives innovation. We’re committed to creating an inclusive workplace where everyone—regardless of background, identity, or experience—has the opportunity to thrive. We welcome all applicants!