Architecture

Overview of Giza Platform's architecture

The architecture of the Giza Platform is designed to support complex workflows and machine learning operations within a scalable and secure environment. It integrates customer-written actions with a powerful backend infrastructure, allowing for seamless execution of verifiable inferences, scheduling, and monitoring of models. This section will provide an overview of the platform's architecture, with a deeper exploration of the robust and efficient actions system as well as the powerful ZKML inference serving service.

?

Custom Execution Environment

The Giza Platform empowers clients by allowing them to host and execute their actions in any environment of their choosing, rather than being confined to a cloud platform managed by Giza. This execution environment is a testament to the platform's adaptability, providing clients with the freedom and flexibility to:

- Choose Their Hosting
- : Clients can run their actions wherever they prefer, be it on-premises, in a private cloud, or through third-party cloud services.
- Maintain Control
- : By hosting their own actions, clients retain full control over their execution environment, ensuring compliance with their internal policies and security requirements.
- Leverage Existing Infrastructure
- : Clients can make the most of their established infrastructure investments and avoid potential vendor lock-in, which is crucial for those with specialized or regulated IT environments.
-

The Giza Platform is designed with privacy and efficiency in mind, operating on the principle of minimal data retention. It only stores metadata of action executions conducted within the Custom Execution Environment. This approach ensures that while the platform retains just enough information to provide valuable insights and maintain operational integrity, the actual execution data remains within the client's chosen environment, safeguarding the confidentiality and control over their proprietary processes.

User Workspaces

The Giza Platform serves as the central pillar for user operations, offering a hosted environment where individual workspaces are maintained. These workspaces are the sanctuaries where users orchestrate their workflows, manage deployments, and monitor the lifecycle of their actions. By centralizing these workspaces, Giza provides a unified and scalable infrastructure that supports the diverse demands of its user base.

Scalable and Verifiable ML Model Serving

Within these hosted workspaces, the Giza Platform excels at serving machine learning models at scale. The platform's MLOps stack is finely tuned for not just deploying and serving ML models but also for tracking and monitoring their performance. The following features characterize this aspect of the platform:

- Verifiable Inferences
- : Giza's infrastructure is uniquely capable of delivering verifiable inferences, ensuring that users can trust the outputs of their ML models and to be integrated in the blockchain through zero-knowledge proofs onchain verification.
- Scalability
- : Whether for small-scale experiments or enterprise-level operations, the platform can scale to meet the computational demands of verifiable ML models inferences supporting real-time and batch workloads.
- MLOps Integration
- : Giza's MLOps stack is integrated into the platform, providing users with tools for version control, model tracking, and performance monitoring, streamlining the model lifecycle management.
-

Streamlined MLOps for Model Tracking and Monitoring

The integration of MLOps practices is a testament to the platform's commitment to operational excellence in ML model management. Users benefit from:

- Continuous Monitoring
- : The platform offers tools to continuously monitor model performance, making it easier to identify and address issues in real-time.
- Model Tracking
- : Version control and tracking are built into the platform, allowing users to manage model iterations with ease and

transparency.

- Streamlined Operations
- : By unifying model deployment, tracking, and monitoring, Giza streamlines the entire ML workflow, from development to deployment and beyond.
-

In essence, the Giza Platform is the nexus where user workspaces are hosted, ML models are served and managed, and MLOps practices are embedded to ensure that machine learning operations are as verifiable and scalable as they are efficient and user-friendly.

Previous Introduction Next Modular Stack with Orion

Last updated2 months ago