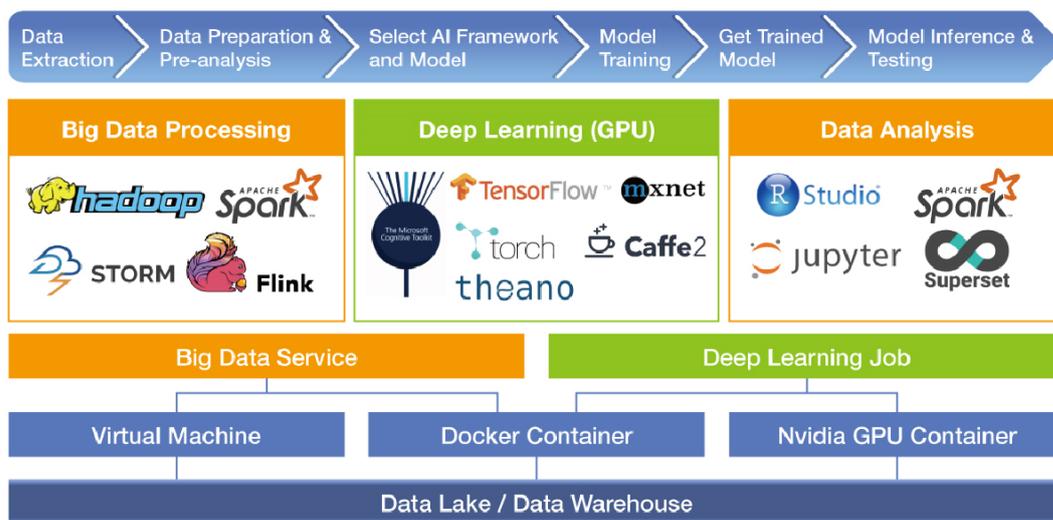


Gemini AI Console

A lightweight and efficient GPU management platform to optimize Big Data analysis and AI computing process for enterprise

The development of artificial intelligence (AI) undoubtedly brings unprecedented technological progress in human history. Gemini AI Console helps data engineers to manage single or multiple GPU servers easily, allowing valuable GPU resources to be used more efficiently, further reduces the cost of Big Data and AI development.



From big data preprocessing, AI model training, all the way to the complex process of AI application deployment, enterprises often suffer from heterogeneous environments, multiple computing architectures, and even cross-departmental collaborations required at different stages. Gemini AI console aims to facilitate enterprise and organization for cross-unit AI project collaboration, therefore help enterprises more efficiently extract better business opportunities from massive data.



Gemini GPU Management Node

Machine Learning Pipeline Management



GPU Server Farm

Features	Description
User and Roles Management	System administrator can create multiple projects and manage user accounts, roles and authorization. System administrator can use the Corporate LDAP or other authentication mechanisms to authorize users for specific services.
Jobs/Service Definition	Gemini AI Console supports two kinds of operation models: Cloud Service (pay-by-capacity) and Batch Jobs (pay-by-utility). Users and administrators can upload and use specific VM and container image files through our private Image Registry.
Resource Management and Workload Management	Gemini AI Console provides user self-provisioning for AI services which can run in heterogeneous environments (VM, Docker and GPU). In batch mode, jobs will be scheduled by the GOC workload manager. In addition, Gemini AI Console supports resource quota management by project and horizontal scale-out.
Resource Monitoring and Reporting	Through the user dashboard interface, the real-time usage statistics and historical data of all cloud services and jobs are presented, and the resource usage report can be provided for the billing basis.
Management Portal	Gemini AI Console provides a user friendly portal for both system administrator and users to manage resources and services, including creating new service or submit jobs with few clicks. This management portal also provides automatic alarm notification for abnormal events.

Advantages and Benefits

- ✓ **Simplify IT complexity and optimize GPU management**
 Manage the physical and virtual resources for multiple GPU and CPU Servers with a single platform. It can also optimize the utilization of GPU resources according to the needs of business organization and cross-unit AI projects.
- ✓ **Improve R&D efficiency and shorten development time**
 Make it easier to prepare the complex infrastructure environment with simple browser interface for the deployment of Big Data and AI computing tools, which helps scientists to focus on their AI algorithm development and training.
- ✓ **Support different computing architectures and heterogeneous environments, bringing perfect AI computing experience**
 A single platform can manage big data services and AI machine learning clusters at the same time. No matter using virtual machines (VMs) or containers (Docker) with GPU, users can manage easily through the management portal. Support both 'Cloud Service' and 'Batch Jobs', which can meet different usage scenarios and definitely enhance the AI development process experience.