

V03.0云上科研平台方案分享

助力天文科研领域创新发展

李东/韩军

2023/4/20

- 01 国家天文台与阿里云合作历程 李东
- 02 阿里云能力介绍 李东
- 03 VO3.0科研云方案分享 韩军

01 国家天文台与阿里云合作历程

国家天文台与阿里云的合作历程



7月王坚博士首访国家天文台
9月国家天文台代表团回访阿里云
10月杭州云栖大会宣布双方战略合作
11月LAMOST数据上云

1月BASS 巡天数据上云
5月“天文数据挖掘”天池大赛举办

2月阿里云冠名博士后全部出站
3月“中国天眼”FAST向全球开放
12月阿里云助力青海冷湖天文选址

三期战略合作开启

2016

2017

2018

2020

2021

2022

2023

1月天文大数据联合研究中心成立
3月虚拟天文台主节点上云
4月联合研究中心科学技术指导委员会成立
4月宇宙驿站天文科普网站群100多个网站集体上云

6月二期战略合作启动

开启VO 3.0的技术探索



02 阿里云能力介绍

在云计算领域领先的市场地位：全球第三，亚太第一，中国第一



阿里云创立于2009年，是全球领先的云计算及人工智能科技公司。

2010年，成为阿里集团的基础设施。

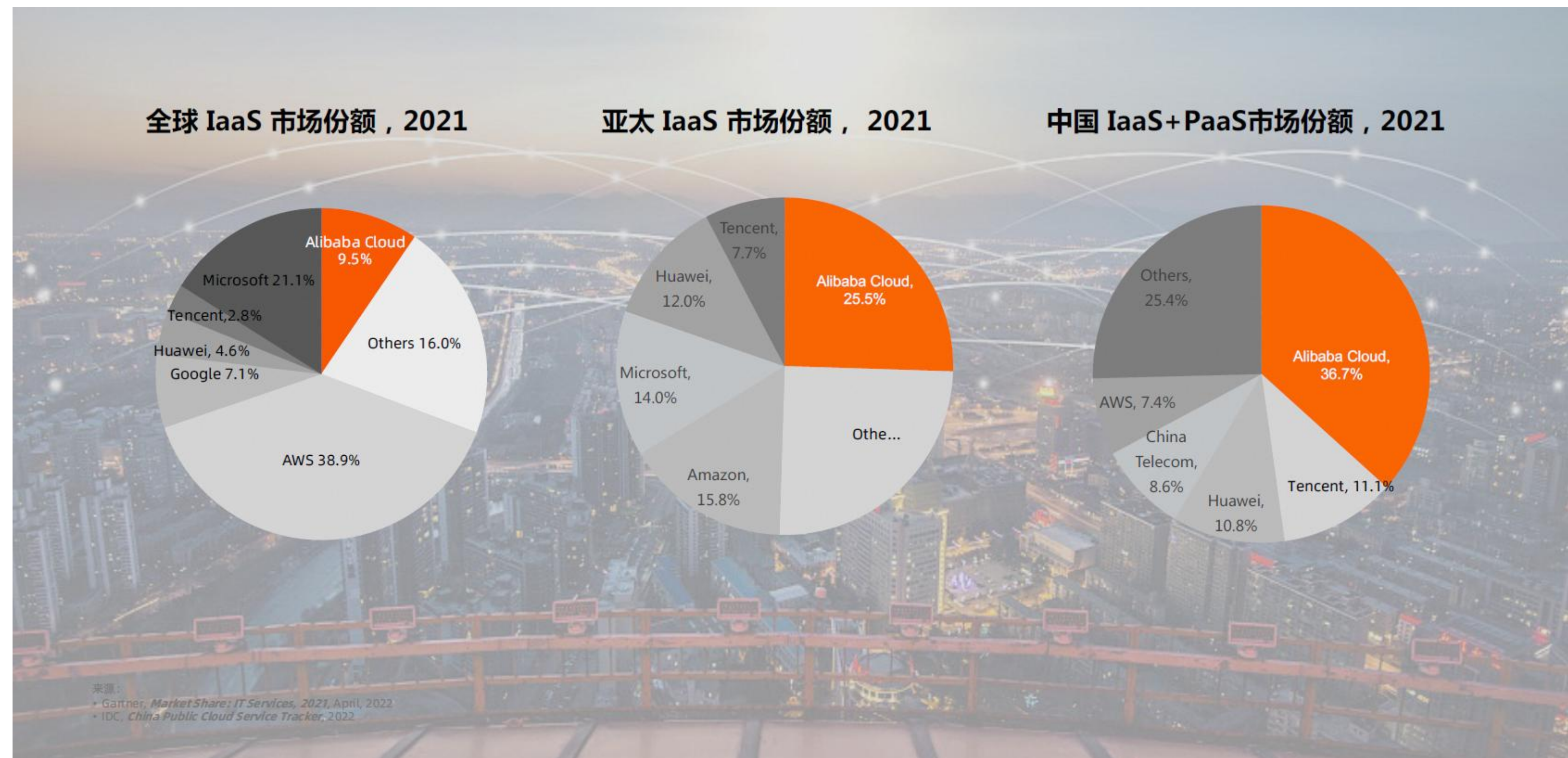
2013年，是第一家提供单集群5K规模的云计算服务厂商。

2014年，推出专有云，开始服务国家公安部。

2016年，发布世界第一个城市大脑，用人工智能进行城市治理。

2017年，奥运会全球指定云服务商。

2022年，推出全栈智能计算解决方案“飞天智算平台”，并启动张北、乌兰察布两座超大规模智算中心，通过公共云和专有云两种模式为科研、公共服务和企业机构提供强大的智能计算服务。张北智算中心建设规模为12 EFLOPS（每秒1200亿亿次浮点运算）AI算力，支撑AI预训练大模型、AI Earth、数字人等前沿AI应用，将成为全球最大的智算中心。



全球化布局：高性能、绿色低碳的全球云计算基础设施

28个地域 86个可用区 EB级别存储 百万级服务器总编排调度



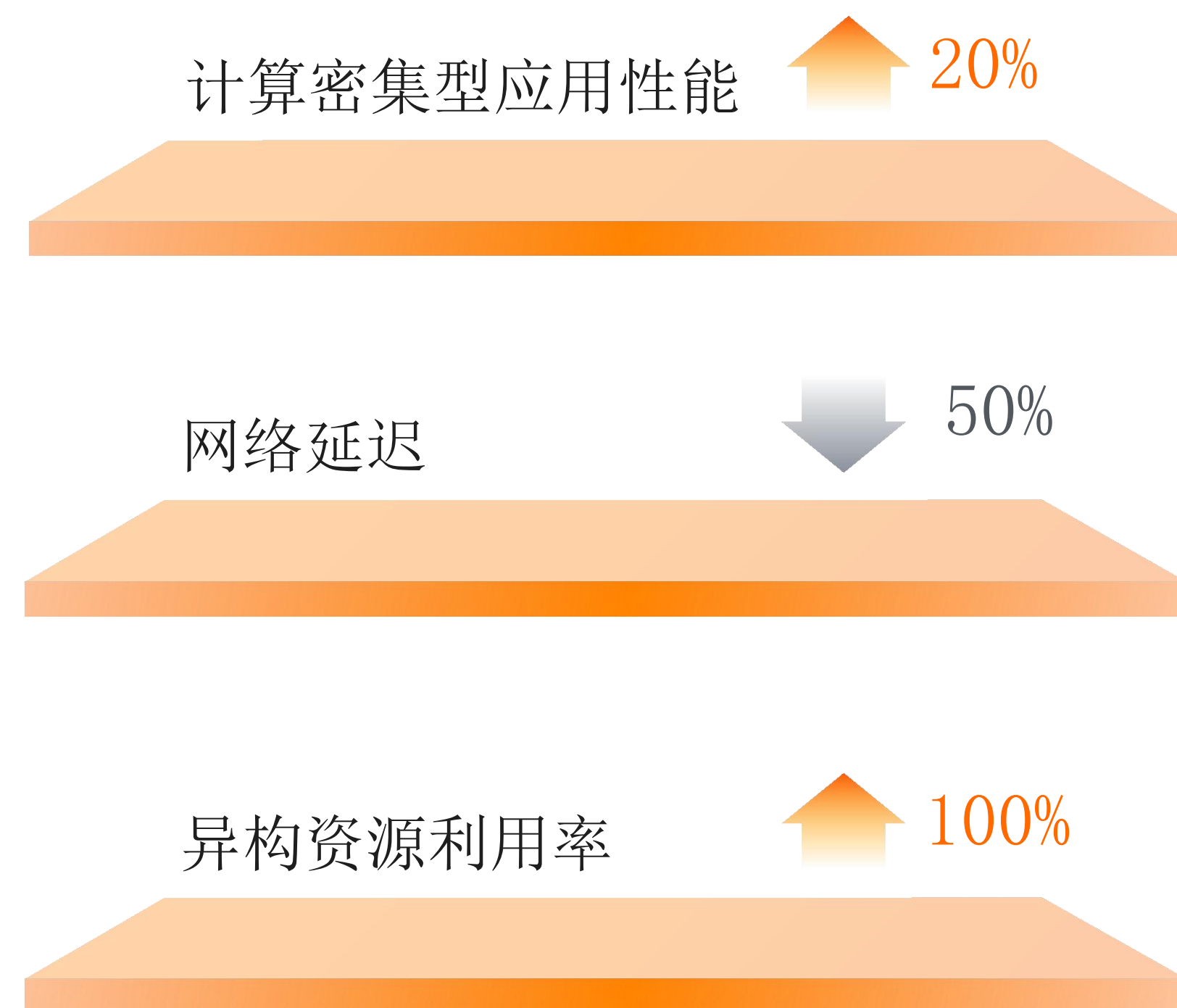
飞天云整体服务能力



云原生能力-容器服务

容器服务 ACK Anywhere: 计算无界、承载无限

在科研机构、企业任何需要云的地方，提供统一的基础设施



云原生能力-互联网中间件



互联网中间件：助力科研机构核心技术互联网化



消息队列

大规模，类型丰富，打造新一代“消息服务到云原生事件流”平台



应用 PaaS 与微服务

开放标准，开源增强，组件全托管，构建现代化应用

应用可观测

开源兼容，全栈可观测，全链路诊断

应用高可用

全链路压测，业务级多活容灾，打造数字安全生产体系

容器服务 ACK Anywhere

公共云

边缘云

专有云

数据库能力



大数据平台



大数据计算引擎



MaxCompute: 阿里巴巴最佳实践的大数据计算平台，EB级数据计算，10万台集群水平扩展

EMR: 基于开源Hadoop生态无缝对接，开箱即用，基于云的弹性计算最高节省50%以上成本，关键计算性能超过100%的提升

实时计算: Flink创始团队官方出品，自研Gemini存储引擎，关键性能2倍提升；最高节省成本80%

交互式分析: 新一代海量实时交互式分析引擎，兼容PostgreSQL，秒级返回

DataWorks: 一站式大数据研发平台，自带阿里巴巴最佳实践的数据治理体系，千万级任务调度。流批一体、AI、交互式查询和自定义引擎节点等跨引擎混编体验

GraphCompute图计算: 千亿规模图毫秒级查询，子图模式匹配、图查询/算法插件

搜索推荐引擎: 云上全托管ES服务。纯自主研发OpenSearch一站式内容智能搜索服务，为企业降低成本，提供高质量、免运维、高可用、灵活定制的云搜索服务

行业应用

城市治理

医疗健康

政务服务

电力能源

工业制造

交通运输

生物基因

大规模

机器学习PAI

全链路AI开发与工程化

丰富的AI开源生态

一体化大数据+AI平台

多模态大模型

“通义” 超大规模模型

M6

AliceMind

视觉

ModelScope魔搭社区

模型即服务，自主开源

300+领先模型

60+通用服务

产业智能OpenTrek

数字孪生

仿真推演

知识工程

决策优化

协同计算

人工智能基础研究

自然语言处理

智能对话 文档处理

机器翻译 NLP模型

视觉

图像识别 目标检测 OCR

编辑生成 视频增强/理解

语言

语音识别 语音合成

语音唤醒 语音增强

决策

优化求解器 黑盒优化

可解释AI 时序预测

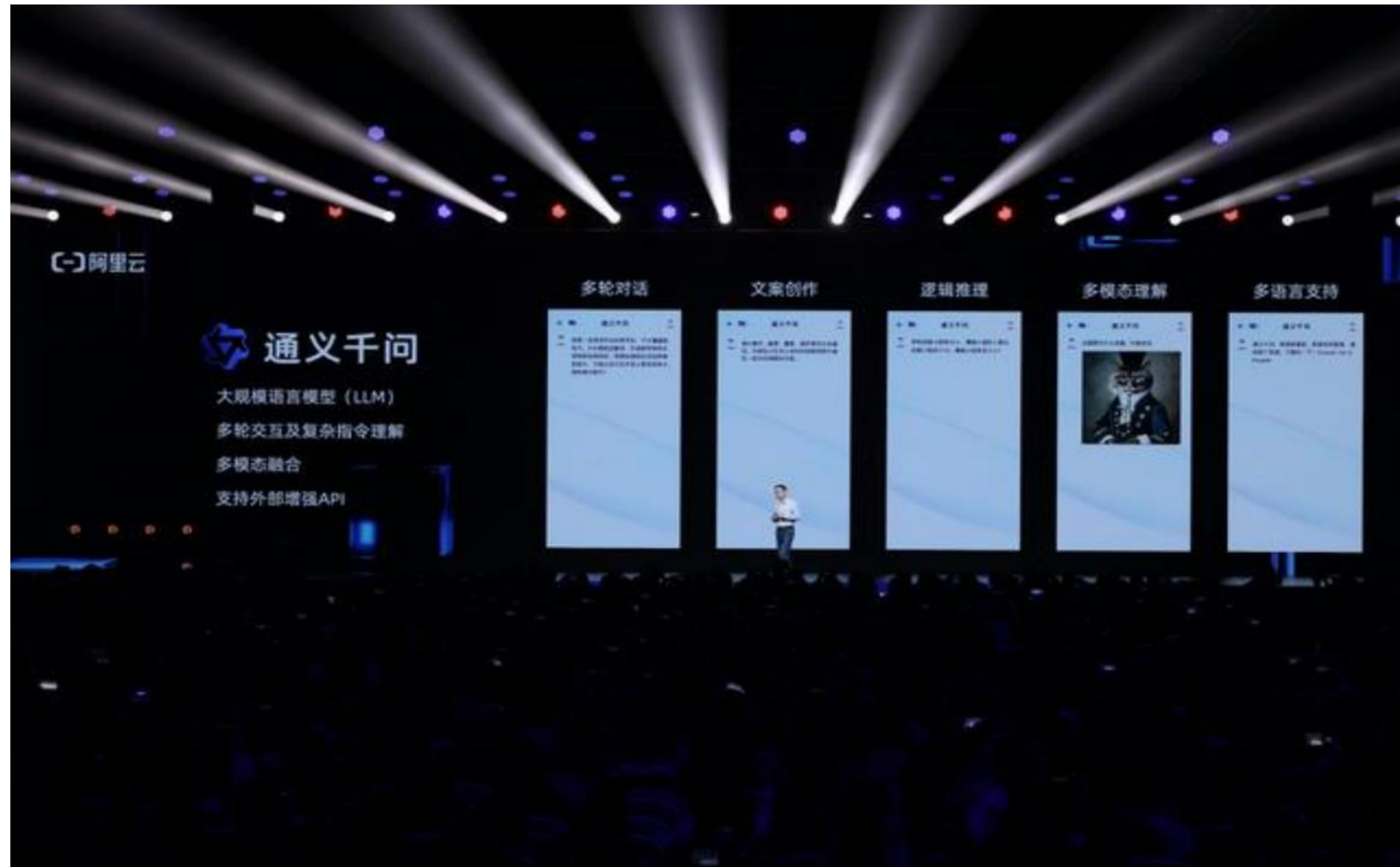
数据隐私

数据隐私保护 联邦学习

数据价值评估 数据共享平台

发布“通义千问”

- 2023年4月11日，阿里巴巴董事局主席张勇，在阿里云峰会上正式发布了阿里的大语言模型工具“通义千问”。
- 阿里云智能CTO周靖人在北京云峰会现场宣布：“我们将开放通义千问的能力，帮助每家企业打造自己的专属大模型！”



持续服务好大家



微信二维码

03 VO3.0科研云方案分享

VO 3.0 一个数据不落地的科研云平台

•VO 1.0

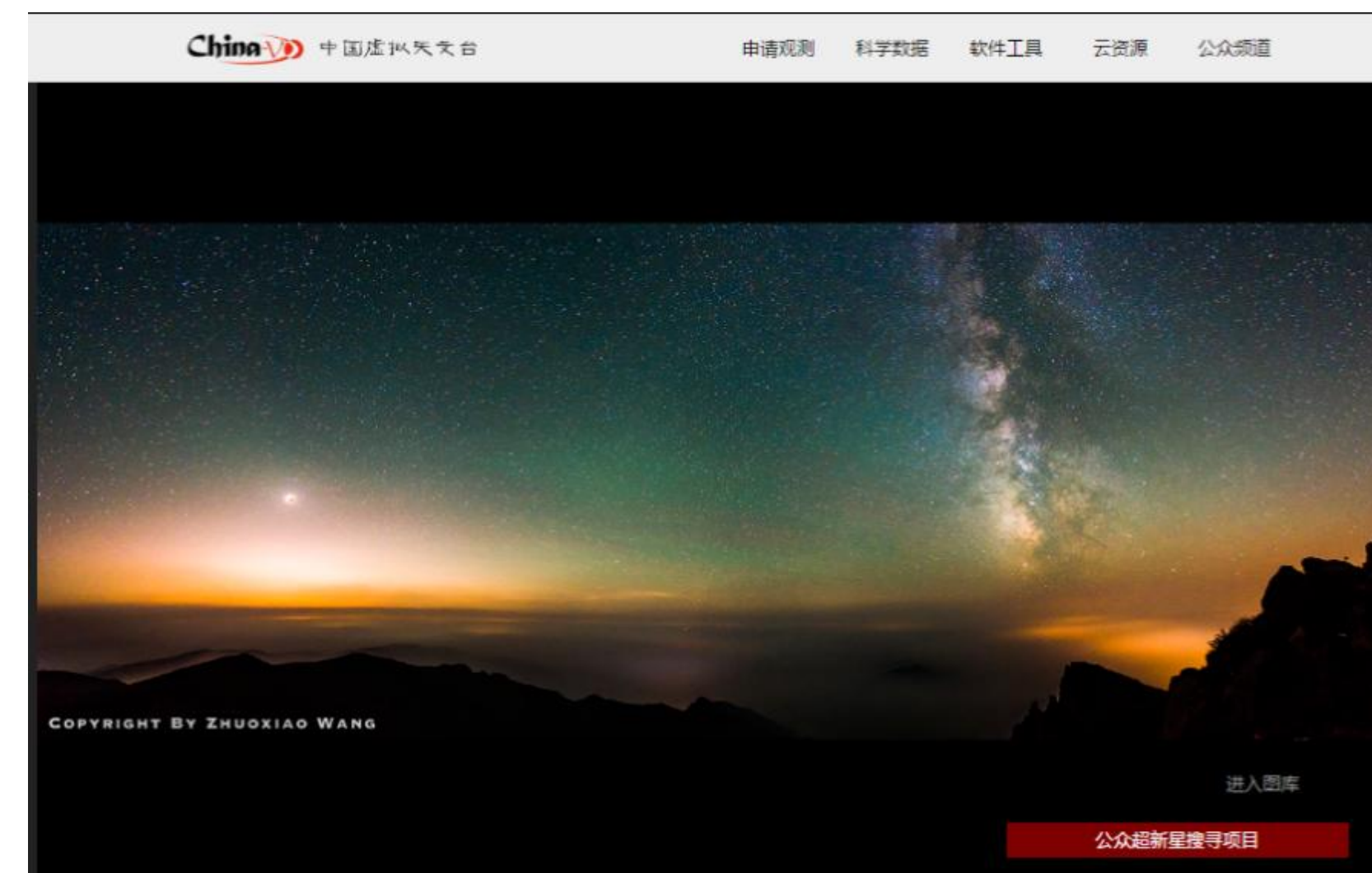
- 2015年天文科技领域云
- 握手云计算

•VO 2.0

- 2019年NADC
- 建章立制

•VO 3.0

- 2023年 EP、CSST牵引
- 微服务、容器化、AI/ML

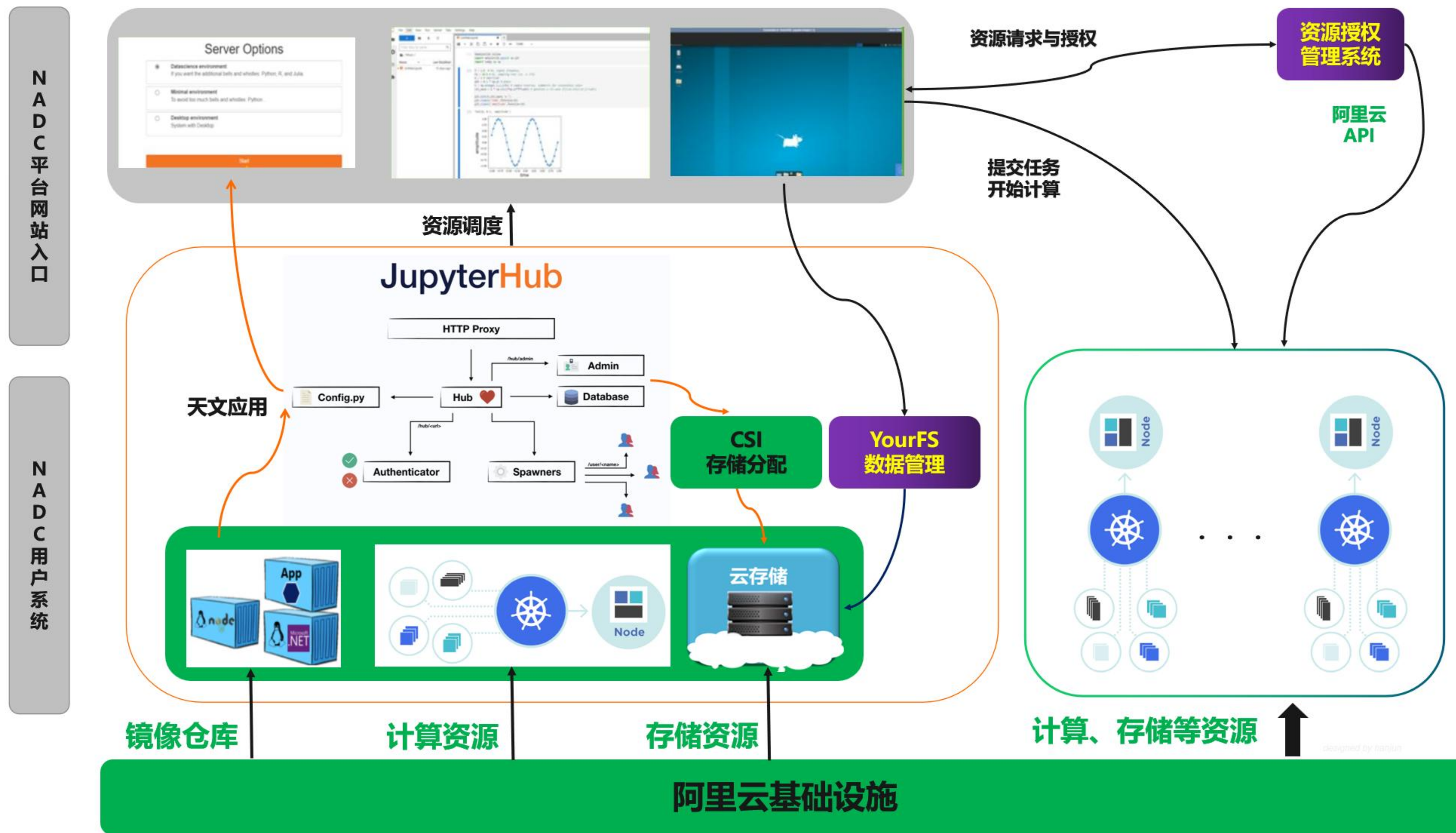


一切基于云上，采用开源堆栈 (K8S+Container) 进行容器化，基于微服务架构提高灵活性和可维护性，借助敏捷方法、DevOps支持持续迭代和运维自动化，利用云平台设施实现弹性伸缩、动态调度、优化资源利用率。



基于云原生的科研云平台技术框架

借助阿里云平台整合优势，以JupyterHub为入口，提供科学数据分析环境、提供数据私人定制、打通数据分析环境与高性能计算集群的融合，从而实现数据不落地的科研云平台系统。



新型文件系统实现数据不落地

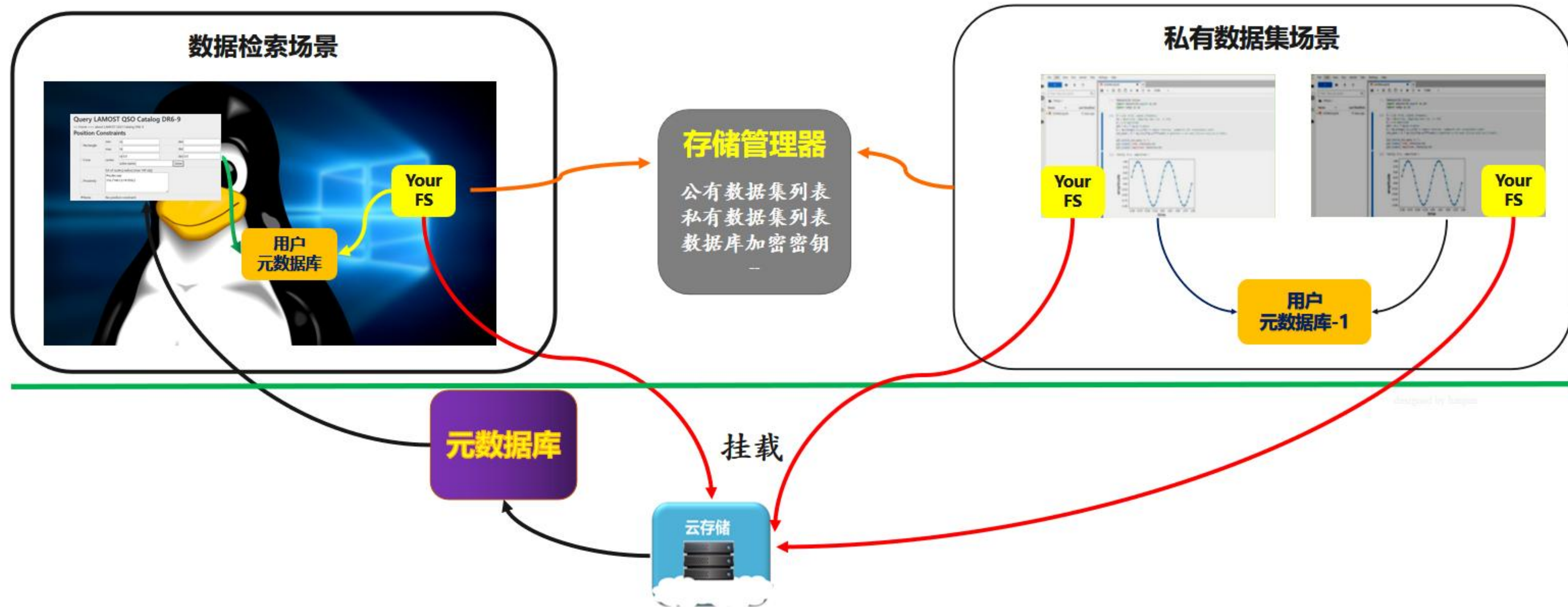
需要面对的问题:

数据检索 -- 下载 -- 处理
数据权限管理



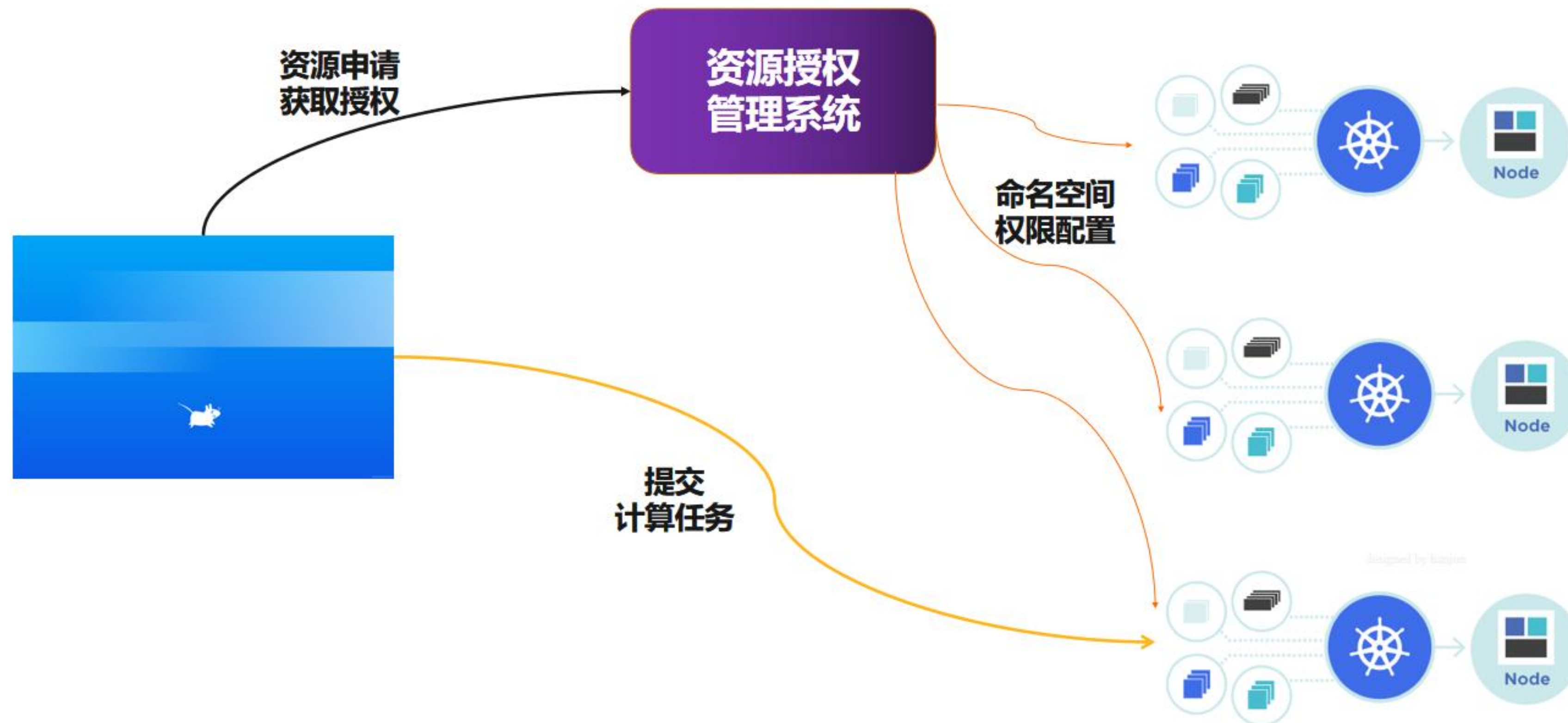
YourFS

私人定制，所见即所得
一个实体，所想即所得



JupyterHub与高性能集群融合

打通个人工作环境与超算环境的融合，实现数据不落地开展大规模的计算、应用测试等工作。





阿里云