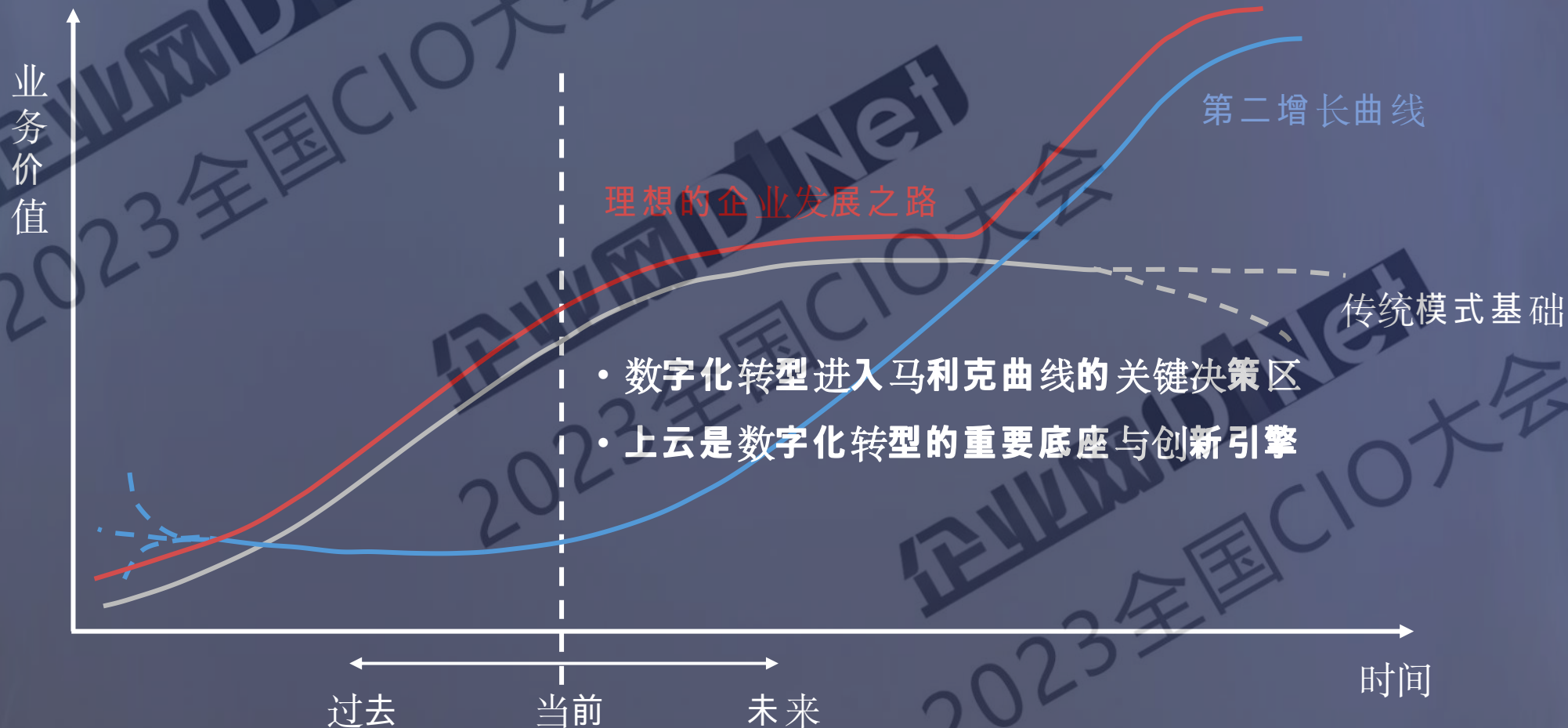


释放业务价值 打造增长引擎

火山引擎副总裁 张鑫

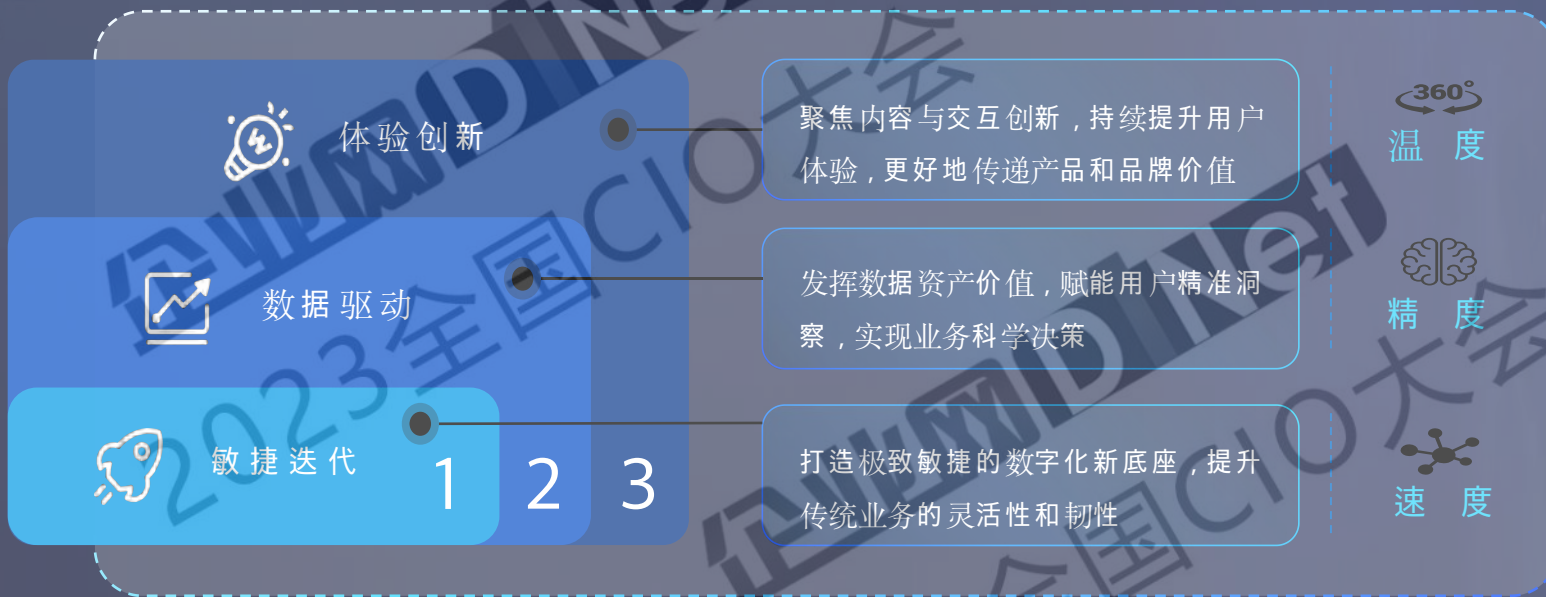
2023年5月

数字化转型进入马利克曲线的关键决策区



云上还需要“三要素”，拉动企业业务增长

“三要素”
汇集温度、精度和速度
帮助企业实现业务价值



云上“增长三要素”在字节跳动的实践

体验创新：为用户触点构建卓越体验

10亿级月活用户

极致优化的音视频技术

基于万亿级特征的实时推荐算法

敏捷迭代：为业务发展推动产品创新

日均上线变更20000+

线上应用总数超过60000+

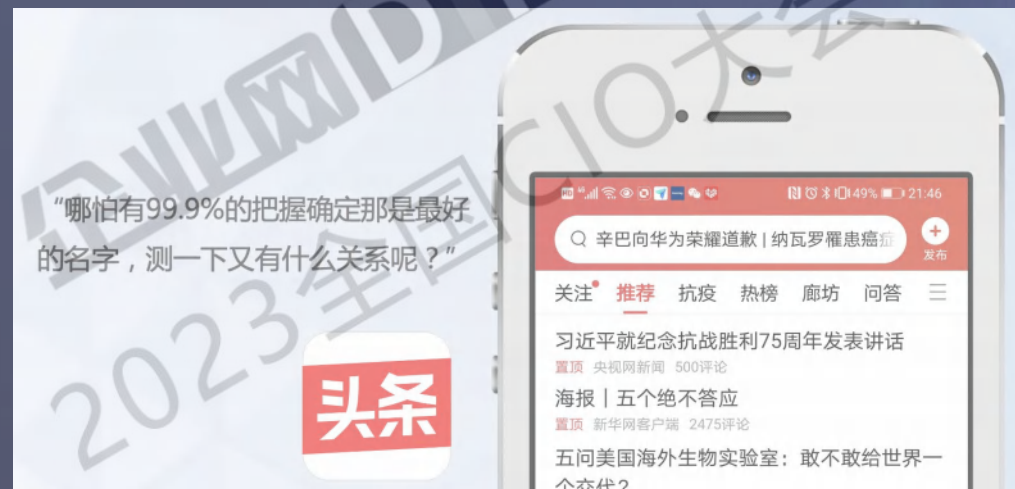
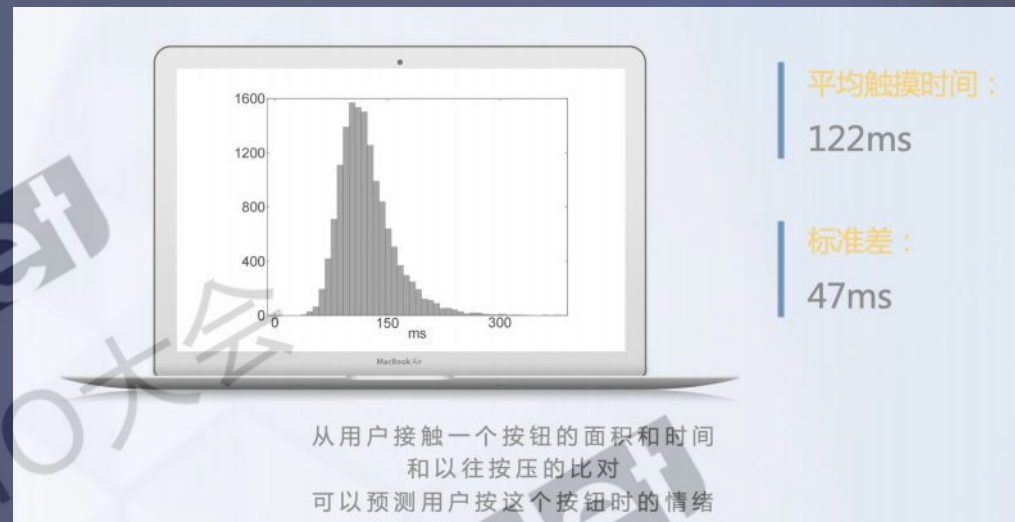
1000万+线上容器

数据驱动：为内核驱动提供敏捷运营

日新增数十PB数据

每日创建上千次AB实验

6大标签主题，10万+标签类型



火山引擎为企业提供云上增长新动力



今日头条



抖音



抖音火山版



西瓜视频



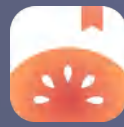
懂车帝



住小帮



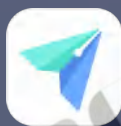
幸福里



番茄小说



皮皮虾



飞书



FaceU 激萌



剪映



轻颜相机

“三要素”助力产业升级

行业解决方案

能源

零售

医疗

汽车

金融

制造

互联网

智能应用

数据增长

智能内容

个性化推荐

智能体验

智能创意

技术引擎

视频架构

敏捷中台

人工智能

数据中台

云基础服务

计算

存储

数据库& 中间件

网络

边缘计算

案例：“三要素”的核心是为业务服务，伴随并推动业务成长



速度篇：敏捷迭代

企业网DINet
2023全国CIO大会

企业网DINet
2023全国CIO大会

企业网DINet
2023全国CIO大会

敏捷基础设施：多云、全栈云原生@字节跳动

在线微服务：10w+ Pod：1000w+ 接入 Mesh 容器数：300w+

日变更数：2w+ 峰值请求：2亿+ 总存储量：数十EB 日新增存储量：数十PB



海量服务支撑

配置中心
全局网络与安全
Docker -> Containerd
FaaS安全沙箱
对接 DevOps

大规模调度

单集群5000+节点
镜像预热
集群联邦
弱网边缘计算
AML GPU调度

在线/离线混部

资源弹性调度，削峰填谷
K8s+yarn
离线任务：训练、视频转码
超售+服务弹性伸缩
Quota计量

集群管理

K8S集群

服务容器化

异构计算

边缘计算

多活调度

资源域Quota

计算资源

存储资源

网络资源

云原生 PaaS 平台

多云基础设施

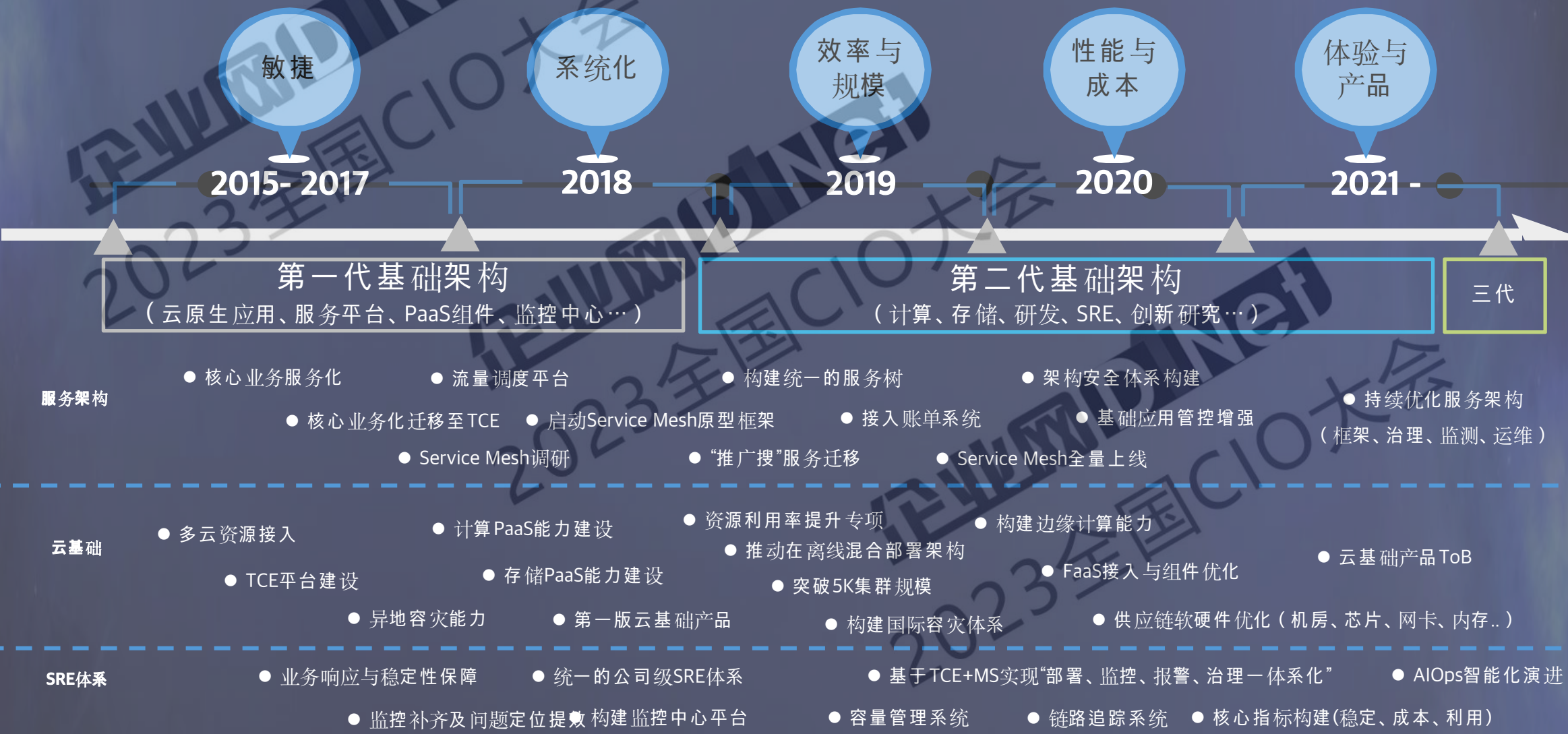
Region 1

Region 2

公有云1

公有云2

罗马非一日建成，技术随着业务发展敏捷迭代



敏捷业务迭代：打造“行业操作系统”，加速业务创新

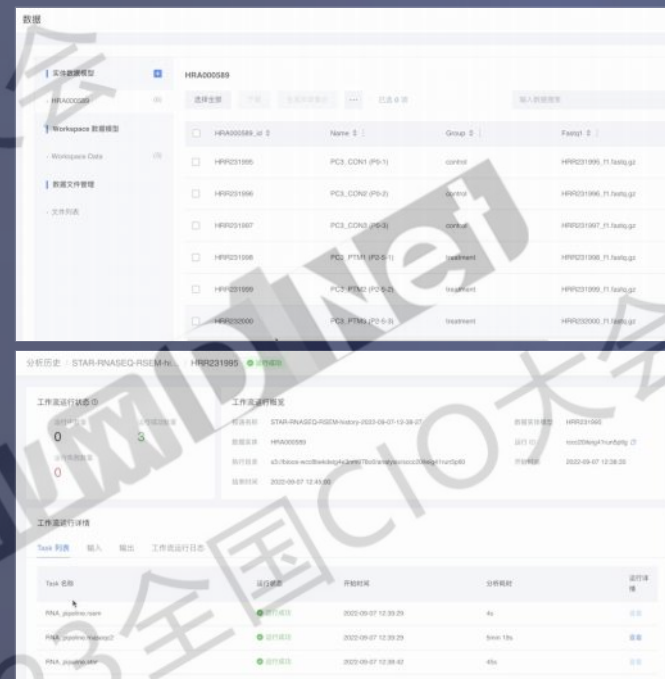


行业操作系统提供Workspace, 打造业务迭代一站式工作空间

研发环境孤立

多团队各自为政, 无协同平台

最佳实践难沉淀和传承



数据管理

工作流管理

...

研发环境统一标准化管理

科研协同效率提升

结果可复现, 成果可共享

精度篇：数据驱动

企业网DINet

2023全国CIO大会

企业网DINet

2023全国CIO大会

企业网DINet

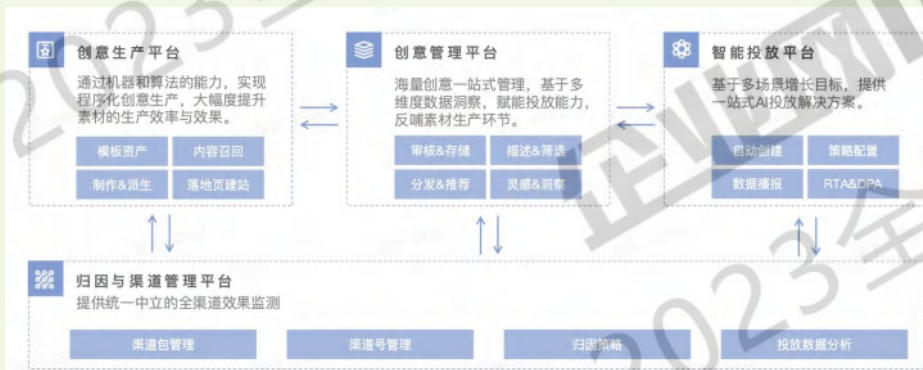
2023全国CIO大会

“数据飞轮”推动业务科学决策

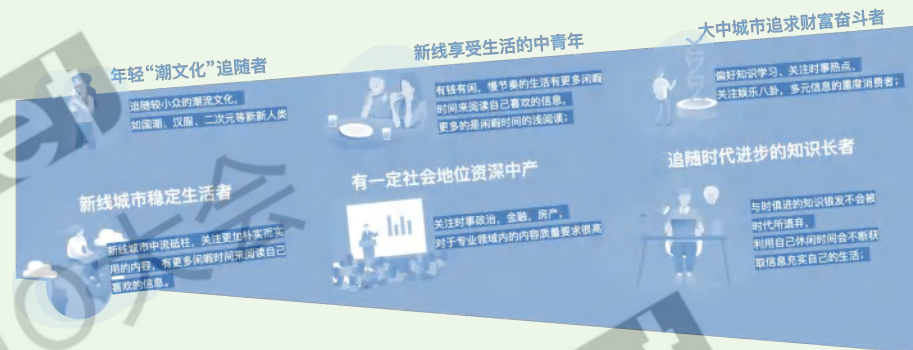


案例：数据驱动的精细化新用户获取

“大家都在做新增获客，怎样才能做得好？”



- 强大的创意生产和管理，让最优质、最吸引人的内容得到最多曝光
- 准确归因、精细化追踪和效果监控：确保每一分钱都能价值最大化，放大好的，控制差的



- 基于强大数据的用户定向：找到对产品最高净值的用户（聚焦在长期的留存活跃/GMV）



- 相信数据和实验，A/B test+高速自动化迭代，每次比对手高10%，经过10次迭代就积累成259%的优势能力

案例：数据驱动落地在消费行业，应是什么样的业务场景？

增量 业务场景

第二增长 曲线

渠道增长

产品增长

用户增长

品类增长

... ..



要实践Sco²pe框架，消费零售企业可以试试“六字诀”



案例：数据驱动的消费云3+1体系，助力零售企业“提效降本”

3朵“提效云”



1朵“降本云”



温度篇：体验创新

企业网DINet
2023全国CIO大会

企业网DINet
2023全国CIO大会

企业网DINet
2023全国CIO大会

持续升级外部用户与内部办公双重体验

面向终端消费者

持续升级用户体验



智能创作



互动营销



端上体验



内容运营

面向企业内部

重塑员工办公体验



身份、网络、终端
一站式 IT 管理



云、网、端
一站式办公组网



体系化
办公安全能力

2022 抖音世界杯直播

累计直播
观看人次达

106 亿

抖音画质



用户直播
总互动

130 亿

原始画质



决赛直播
观看人数达

2.30

主会场



最高同时
在线人数达

3706

狂欢夜



Source: 抖音《世界杯观赛报告》

开放字节跳动的混合办公基础设施

员工日常使用

10万+

办公终端覆盖

25万+

业务应用接入

1000+

覆盖客户100万 终端设备

案例：字节跳动的全球用户体验提升实践



案例：基于AIGC提升产品的潮趣体验



动态照片



婚纱照



三屏人生



赛博朋克



蝴蝶消散

将AI的能力开放出来，助力企业产品体验创新

应用场景

泛互联网

电商零售

金融

汽车

文旅

开放能力

图像识别

文字识别

语音唤醒

语音合成

文本分类

关键词抽取

分割抠图

卡证识别

语音识别

音乐检索

情感分析

机器写作

人脸融合

人像漫画风

音乐副歌检测

语音质检

AI同传

新闻摘要

图像编辑

视觉搜索

音频事件检测

歌曲合成

视频翻译

文本纠错

图像风格化

图片转视频

音频指纹对比

歌唱合成

图片翻译

文档聚类

图像质量优化

视频内容摘要

音视频字幕识别

音乐卡点

文本翻译

个性化文案生成

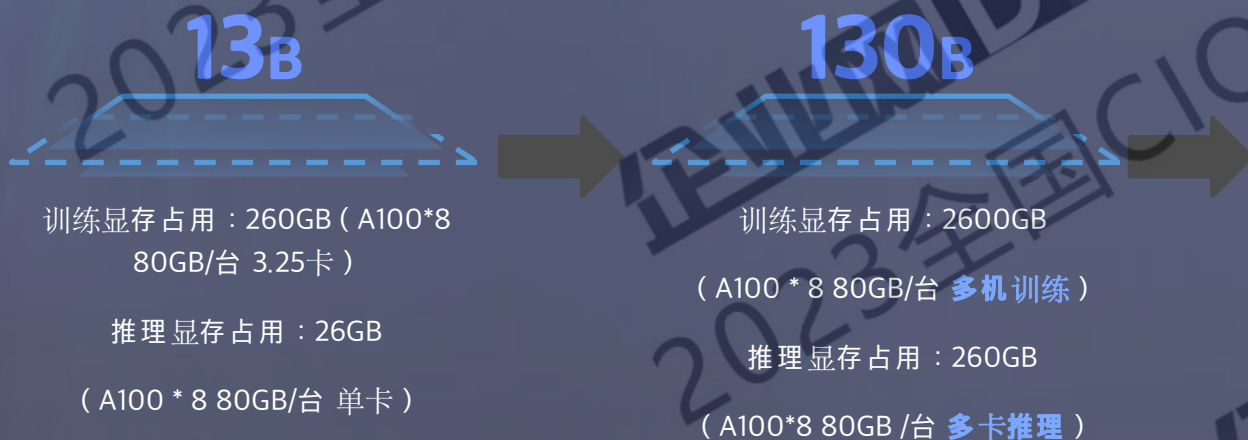
AI能力栈

视觉智能 (200+能力)

语音智能 (50+能力)

语义智能 (50+能力)

在大模型AI时代，显存、计算和通信面临三重压力



新的挑战

大模型任务工程稳定性差

- 频繁的随机硬件故障
- 算法造成的OOM
- 多机多卡分布式通讯阻塞
- 业务代码问题

计算&存储资源需求膨胀：千卡规模

- 千卡训练对存储带宽要求更高

项目成本高昂：千万级成本

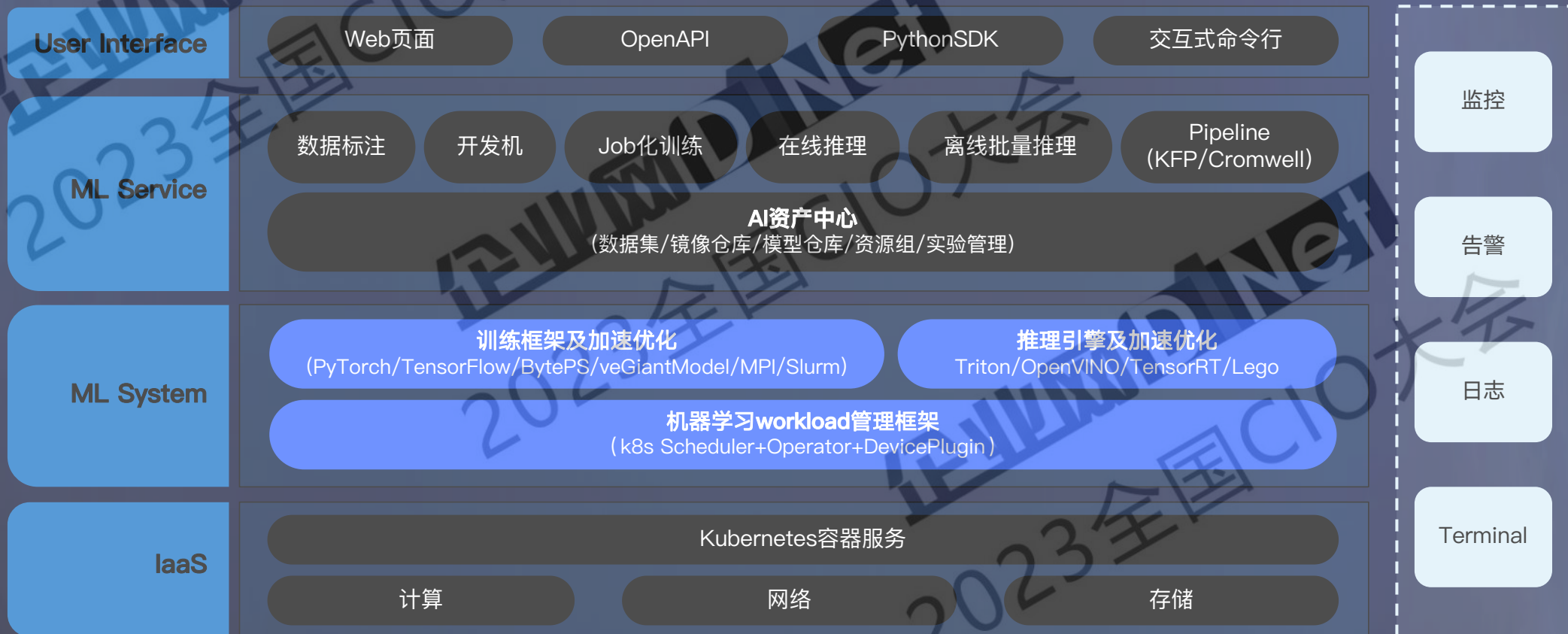
- 千卡A100自建：仅GPU卡需1500万美元
- 云服务：200~500万美元/4个月/52台A100 (Bloom)

业务部署依赖多组件

- 搭建完整的业务pipeline，除了大模型推理本身，还依赖多种PaaS/SaaS服务；多云之间数据传输造成额外成本和业务延时。

面向AGI与大模型的智算基础设施，加速AI创新

服务于专业算法工程师，整合云原生的工具+算力（GPU、CPU云服务器），进行一站式AI算法开发和迭代的平台



充分释放业务价值，助力打造增长引擎

THANKS