

数据来源：数据库产品上市商用时间



第十三届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2022

数据智能 价值创新



线上直播 | 2022/12/14-16



面向资效平衡的机器学习 平台的演进

高会军+小红书云原生负责人

大纲

- 背景
- 解决方案
- 成果展示
- 未来规划

背景

Inspire Lives, 分享和发现世界的精彩



● ● ● ● ●

每天百亿次笔记
曝光



● ● ● ● ●

超过2亿月活



● ● ● ● ●

国内优选的生活
分享社区

背景——遇到的问题

训练任务排队

部署发布效率低

新人上手困难

资源效率低下

开发环境配置复杂

k8s学习周期太长

花费大量时间在等待数据拉取

持续发布如何保证交付质量

如何快速定位线上问题

背景——MLOps发展趋势

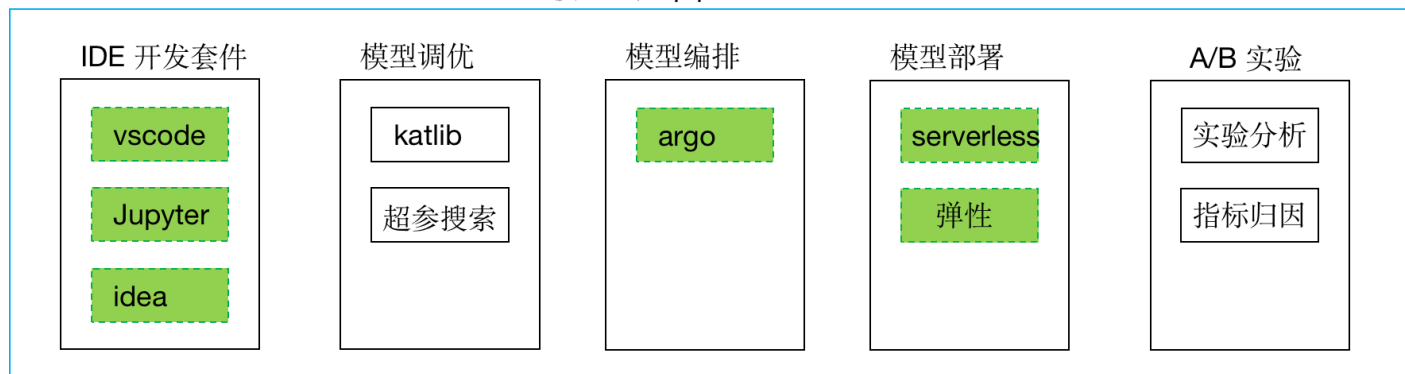


背景——建设资效平衡的机器学习平台

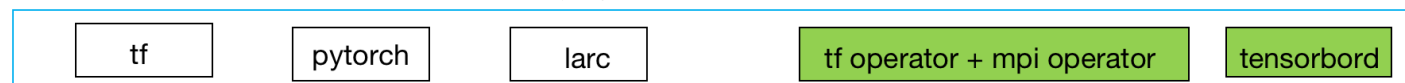


解决方案——整体建设情况

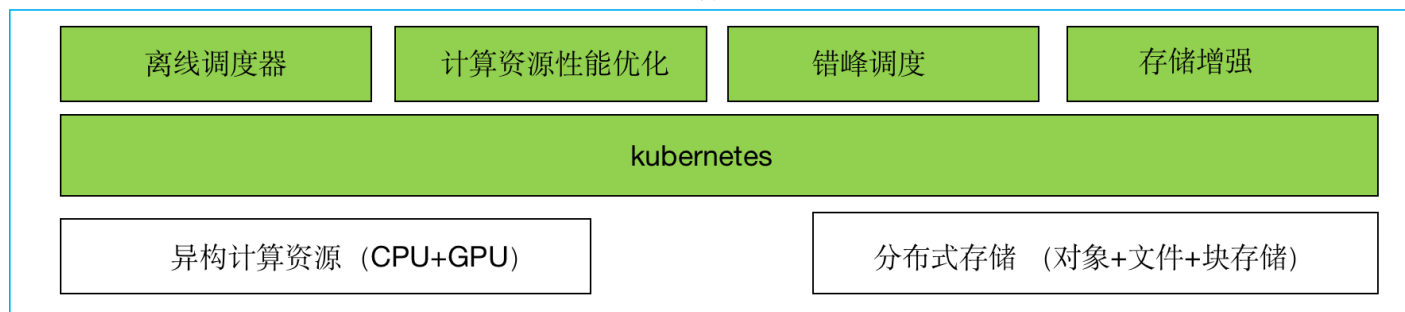
模型生产 pipeline



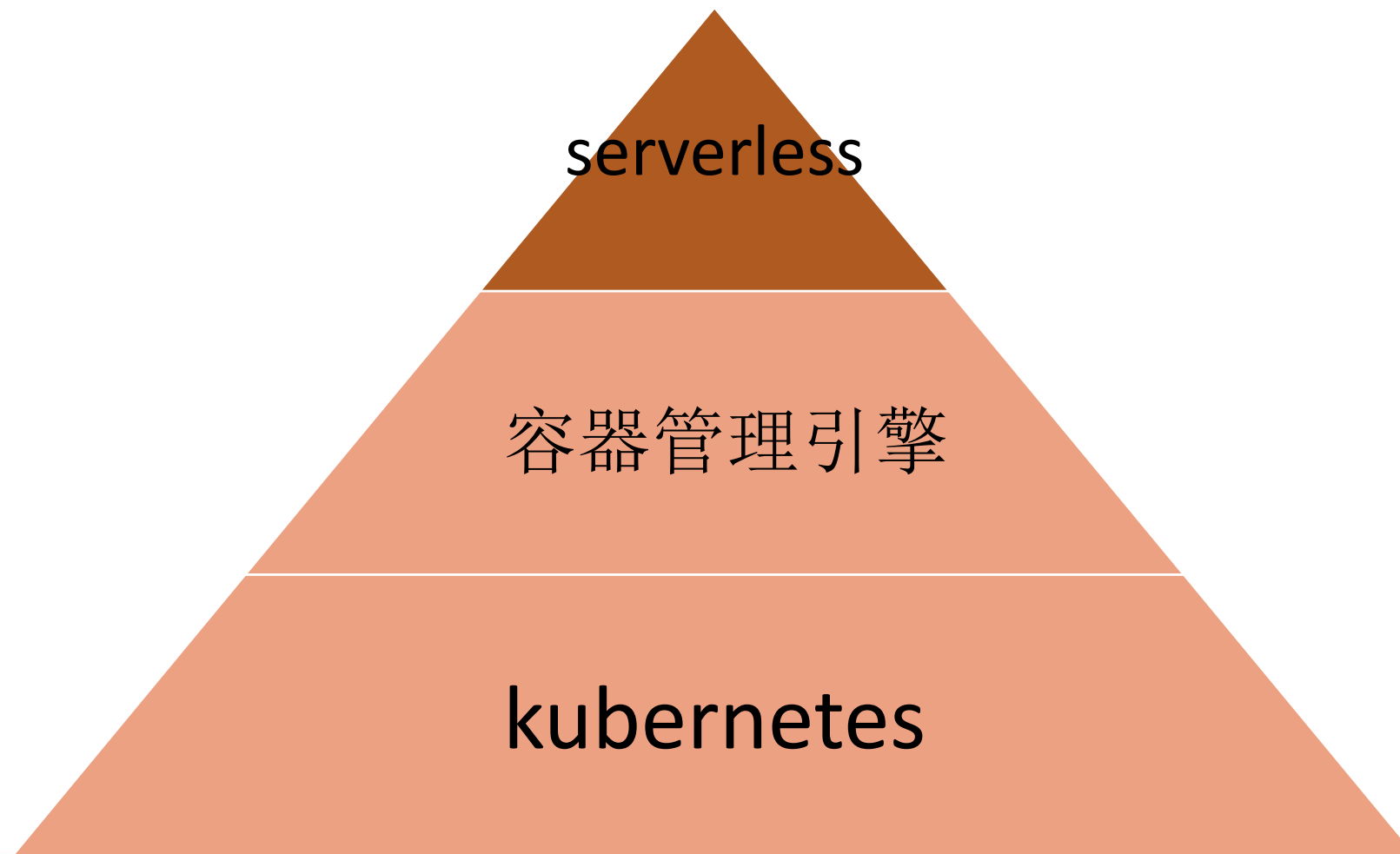
分布式训练框架



云原生资源管理



解决方案——模型部署

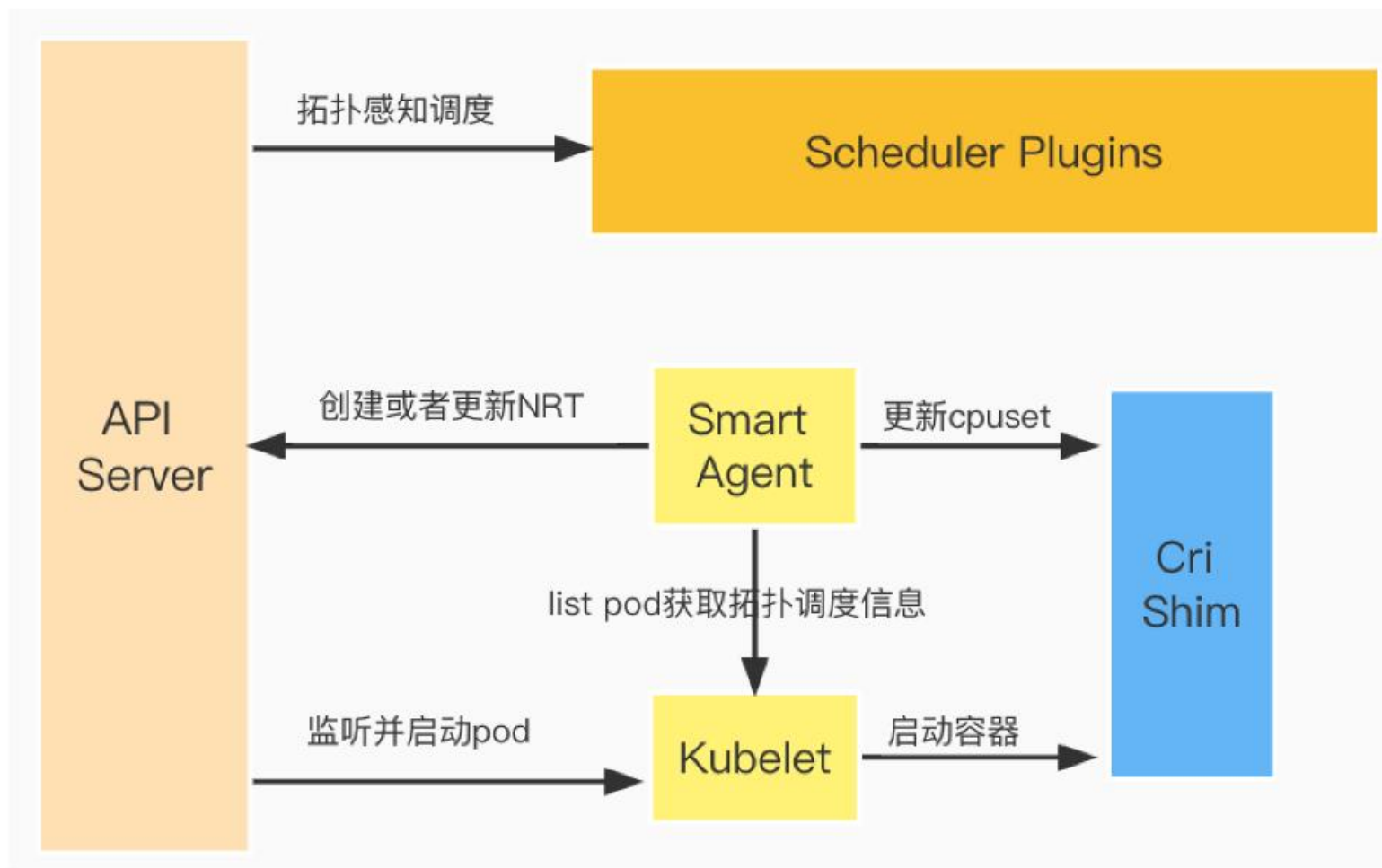


解决方案——离线调度

新增调度能力：大套餐优先调度、queue资源保障、CVM和公有云容器实例服务混合调度等，大幅度提升装箱率。



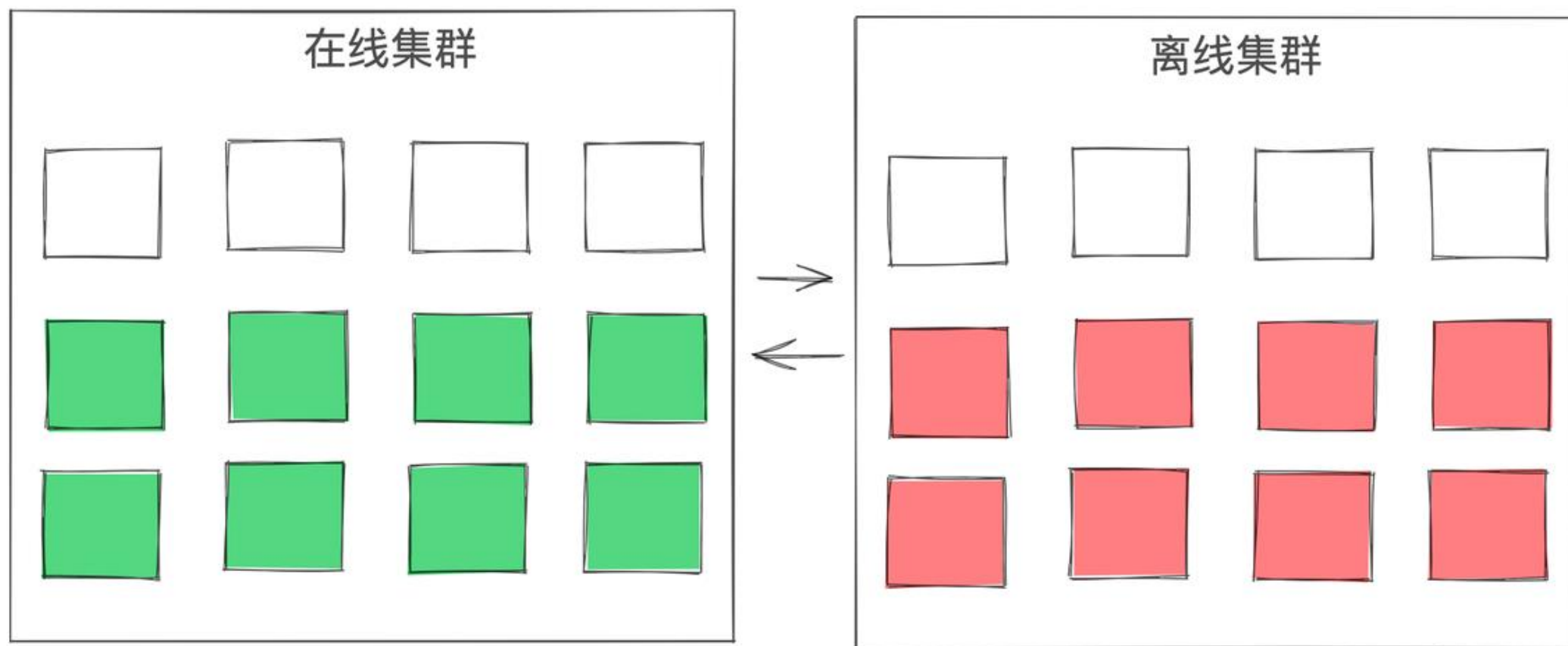
解决方案——计算资源性能优化



解决方案——数据加速



解决方案——资源管理策略



成果展示

- 降低用户使用门槛，实现了低代码模型开发
- 模型训练效率提升 100%
- 提供一整套覆盖从推理到训练的云原生高性能异构计算解决方案，更好地支持批处理作业和高性能服务，进一步提升资源使用效率和降低运维管理复杂性

未来规划

- 协同社区，进一步探究如何在云原生场景下更好地支持批处理作业和高性能作业
- 实现无差别混部

THANKS

SQL Server
vertica
D B 2
G B a s e
O r a c l e
达梦数据库
神舟通用
KingbaseES

2010

2014

2018

openGauss
OceanBase
ArkDB
RASESQL
HotDB
StellarDB
QianBase xTP
云树Shard
GoldenDB
DolphinDB
MatrixDB
DynamoDB
SinoDB
FastData
Galaxybase
KunDB
GDB
GaussDB
PolarDB
KunDB
Spacture
SequoiaDB
OushuDB
ArgoDB
开务数据库
GreatDB
MongoDB
TDSQL
TiDB
Tapdata
StarRocks
UbiSQL