LINUXCON containercon CHINA 中国 CLOUDOPEN

THINK OPEN
开放性思维

阿里云弹性人工智能

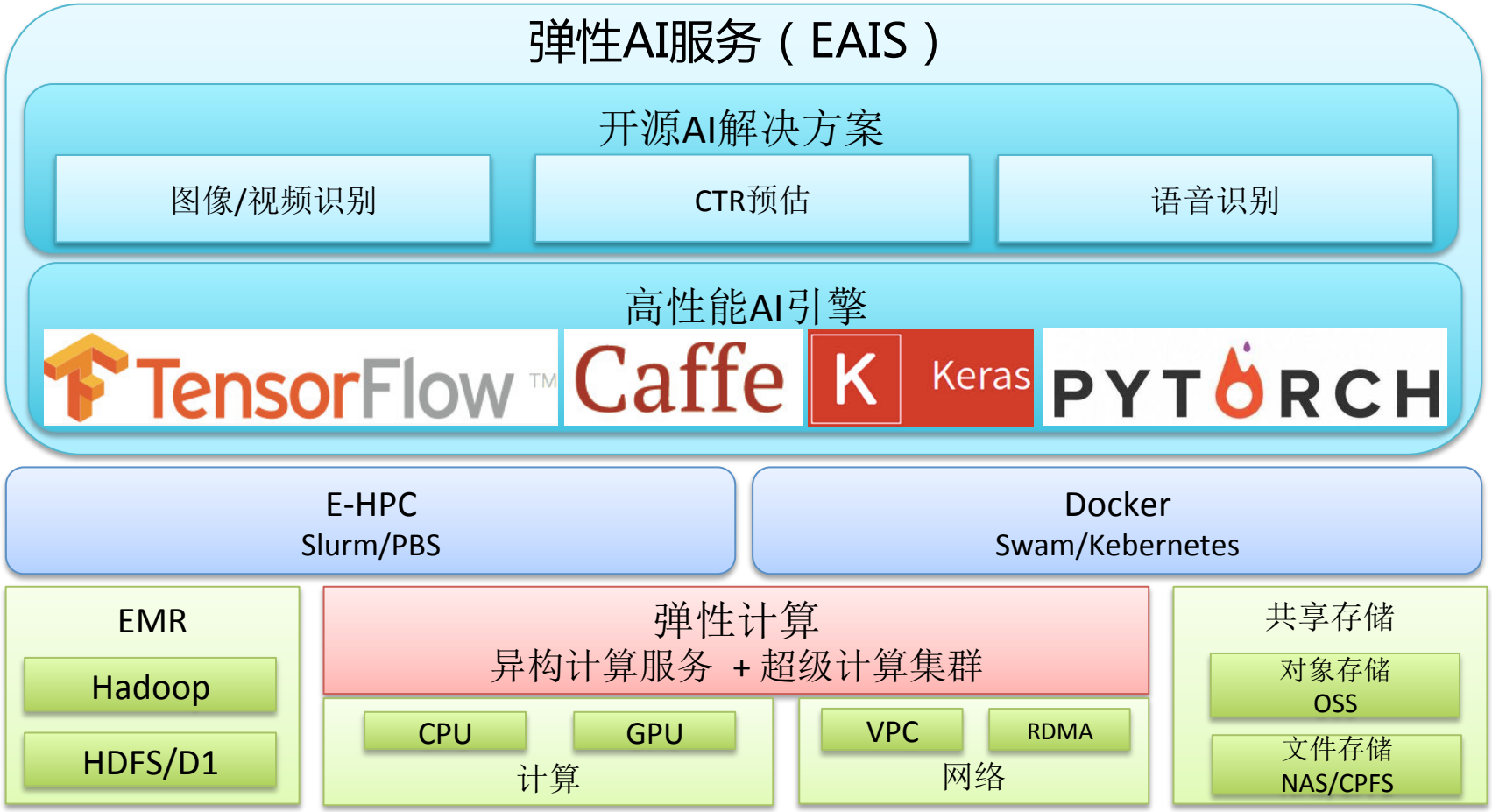
阿里云平台上深度优化分布式训练性能

游亮（昀龙）

LF ASIA, LLC

弹性 AI 服务 – Elastic AI Service

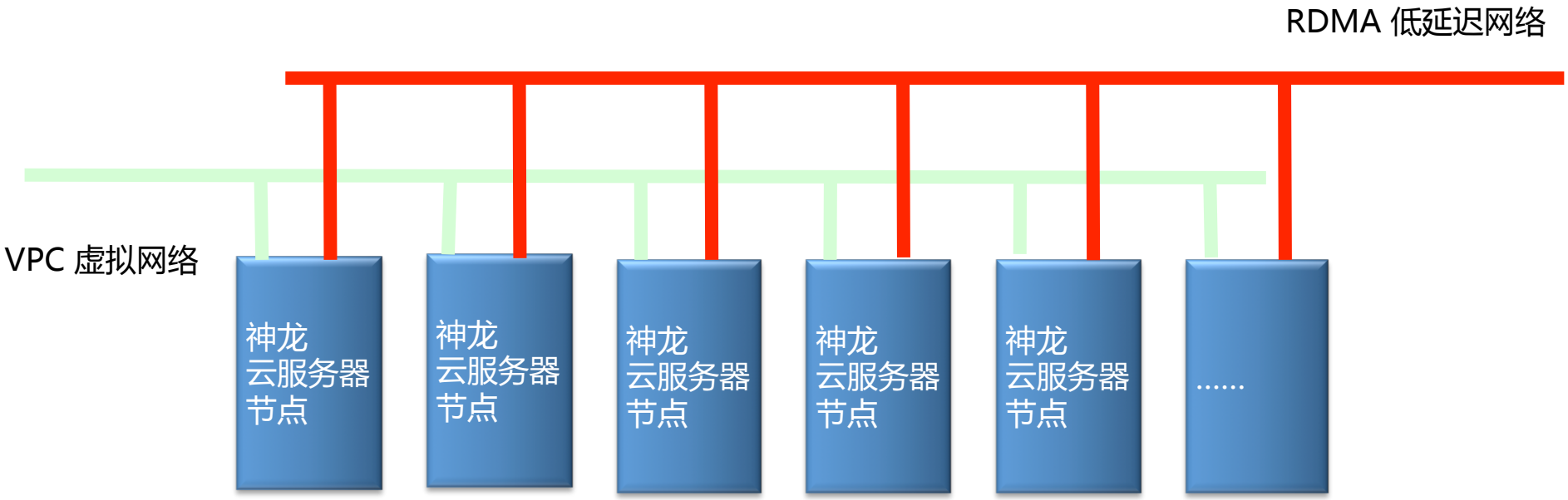
- 基于阿里云弹性基础资源，为用户提供深度性能优化的、一站式的、开源开放的人工智能解决方案



阿里云弹性异构计算服务

- EGS: Elastic GPU Service
- FaaS: FPGA as a Service
- 异构计算
  - CPU + GPU/FPGA优势互补
- 云上大规模GPU/FPGA池
  - 短时间能够获取大量GPU/FPGA资源
  - 有效解决业务波峰、波谷的问题
  - 大大降低训练时间，提高模型迭代速
- 享受硬件升级的红利
- 和其他云产品深度整合

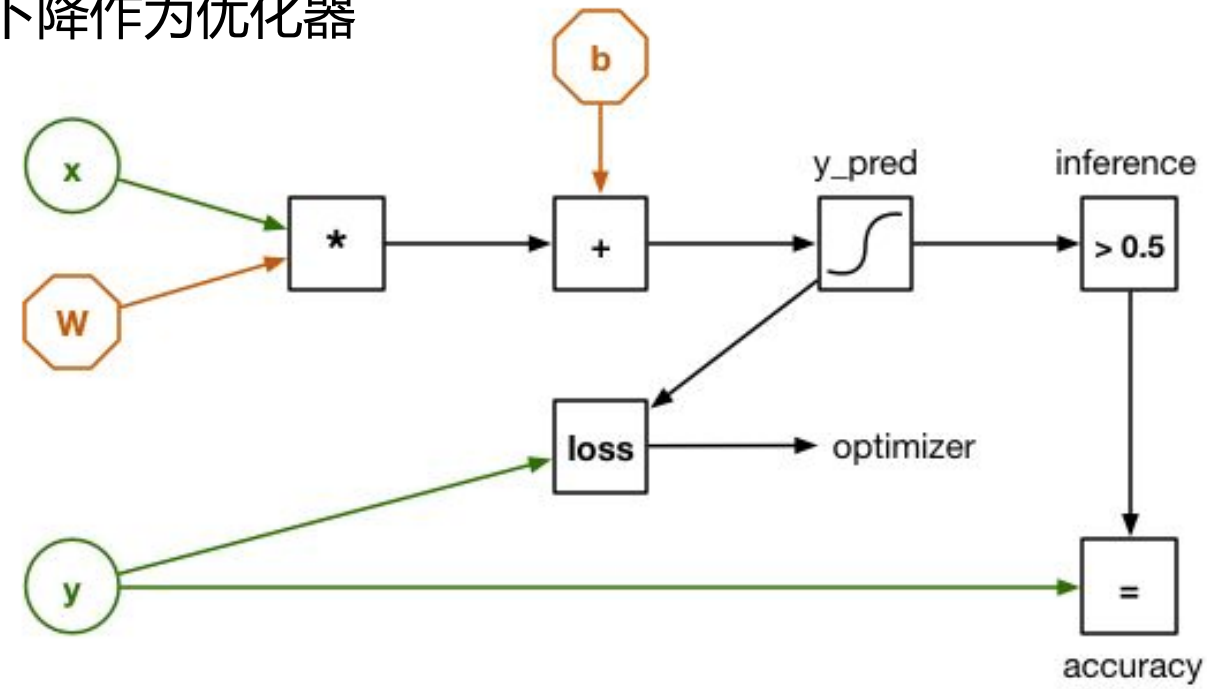
阿里云超级计算集群SCC



- 虚机的弹性 + 物理机的性能
- 支持2x25Gb(100Gb) RoCE RDMA网络
- 支持GPU Direct RDMA
- 适合大规模深度学习训练

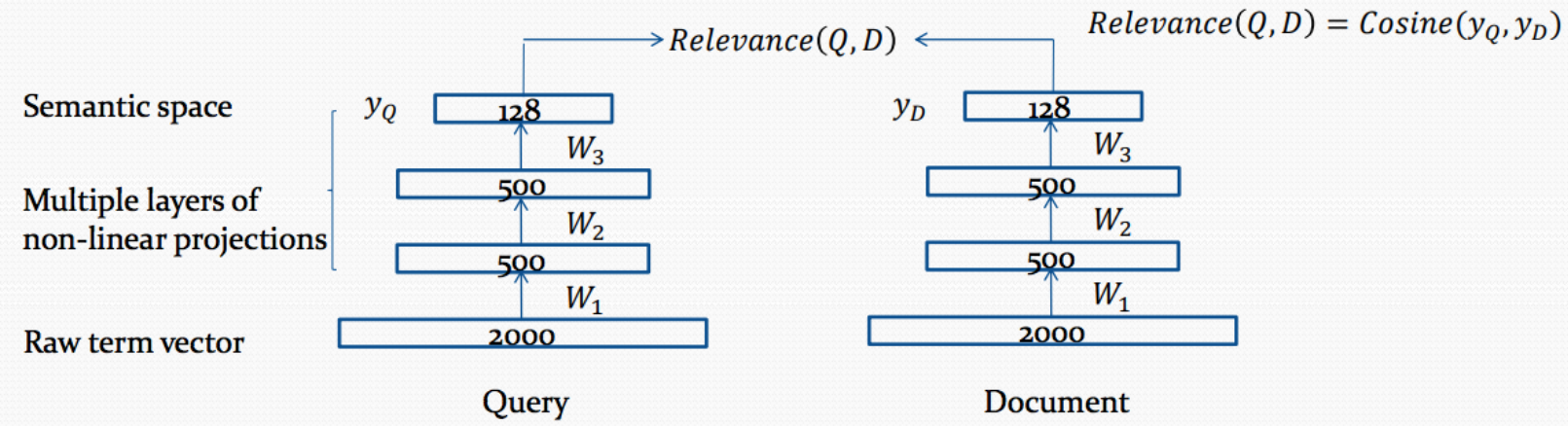
推荐算法I

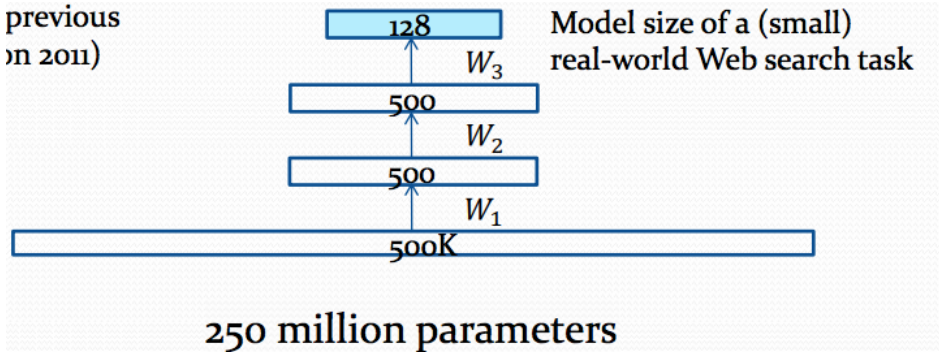
- 逻辑回归算法
  - 浅层模型
  - 需要大量特征工程
  - Sigmoid作为激活函数
  - Sigmoid交叉熵作为损失函数
  - 梯度下降作为优化器



推荐算法II

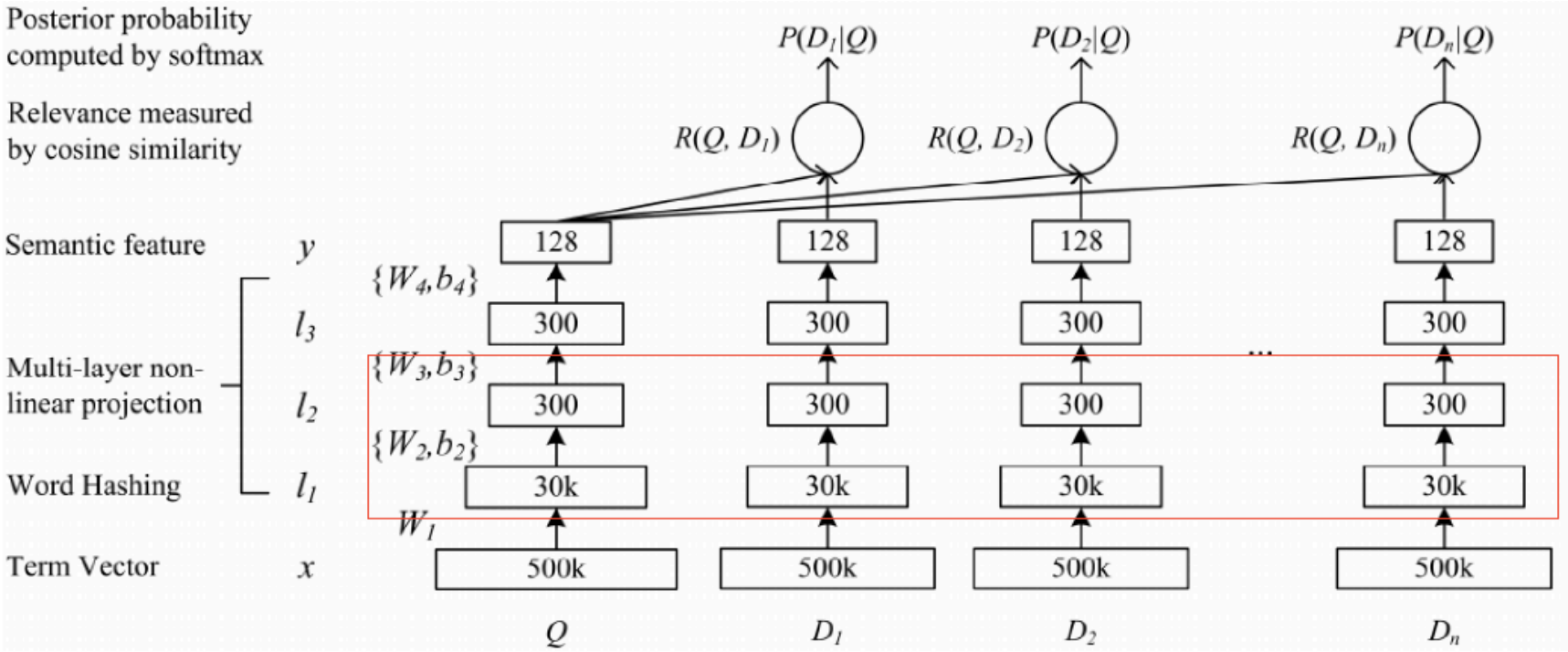
- Deep Auto Encoder[Hinton, 2011]
  - 把请求Q和文档D映射到语义空间
  - 用Cosine计算请求和文档的相关度
  - 缺点：真实生产中模型大小扩展很快 (2.5亿参数)





推荐算法III

- Deep Structured Semantic Model [Microsoft, 2013]
  - 输入层：将输入向量进行单词哈希
  - 使用多个非线性隐藏层抽取高级语义表达（DNN模型）
  - 使用点击信号来指导学习
  - 用Softmax最大化请求和点击过的文档的cosine相似度



推荐算法IV

- Wide and Deep [Google, 2016]
  - Wide输入层：原始输入特征 + 交叉特征；模型：FTRL
  - Deep模型的输入层：将稀疏特征转换成了稠密的embedding层；模型：DNN

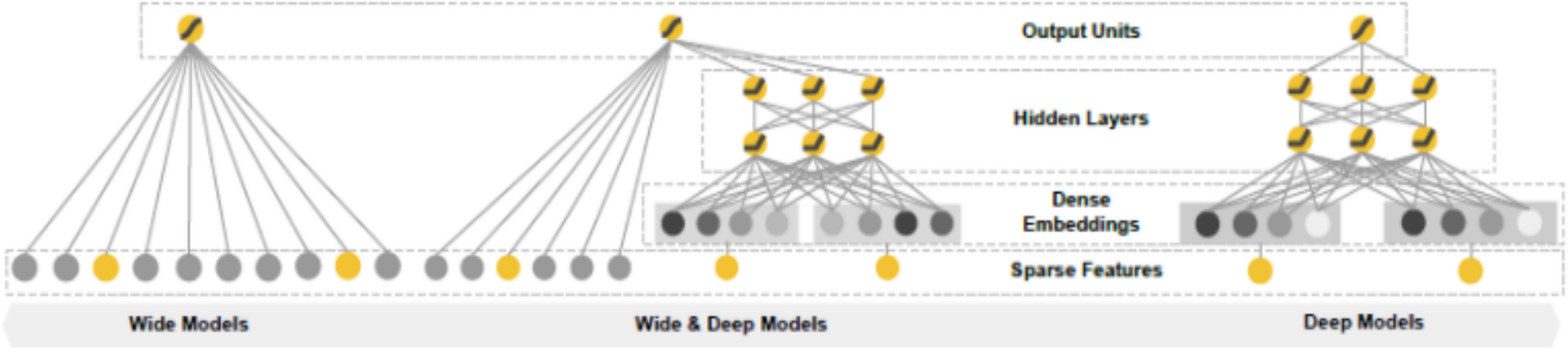


Figure 1: The spectrum of Wide & Deep models.

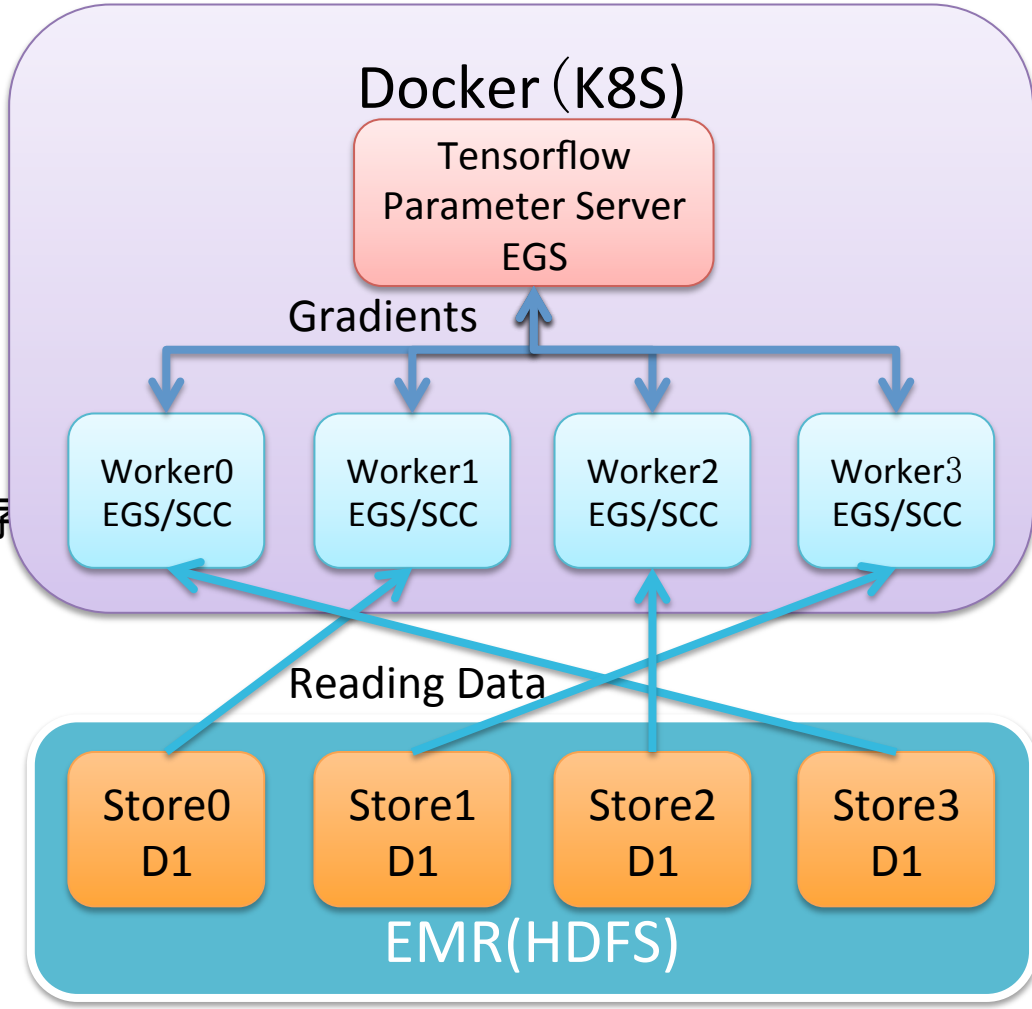
  - 在线效果提升3.9%

Table 1: Offline & online metrics of different models. Online Acquisition Gain is relative to the control.

Model	Offline AUC	Online Acquisition Gain
Wide (control)	0.726	0%
Deep	0.722	+2.9%
Wide & Deep	0.728	+3.9%

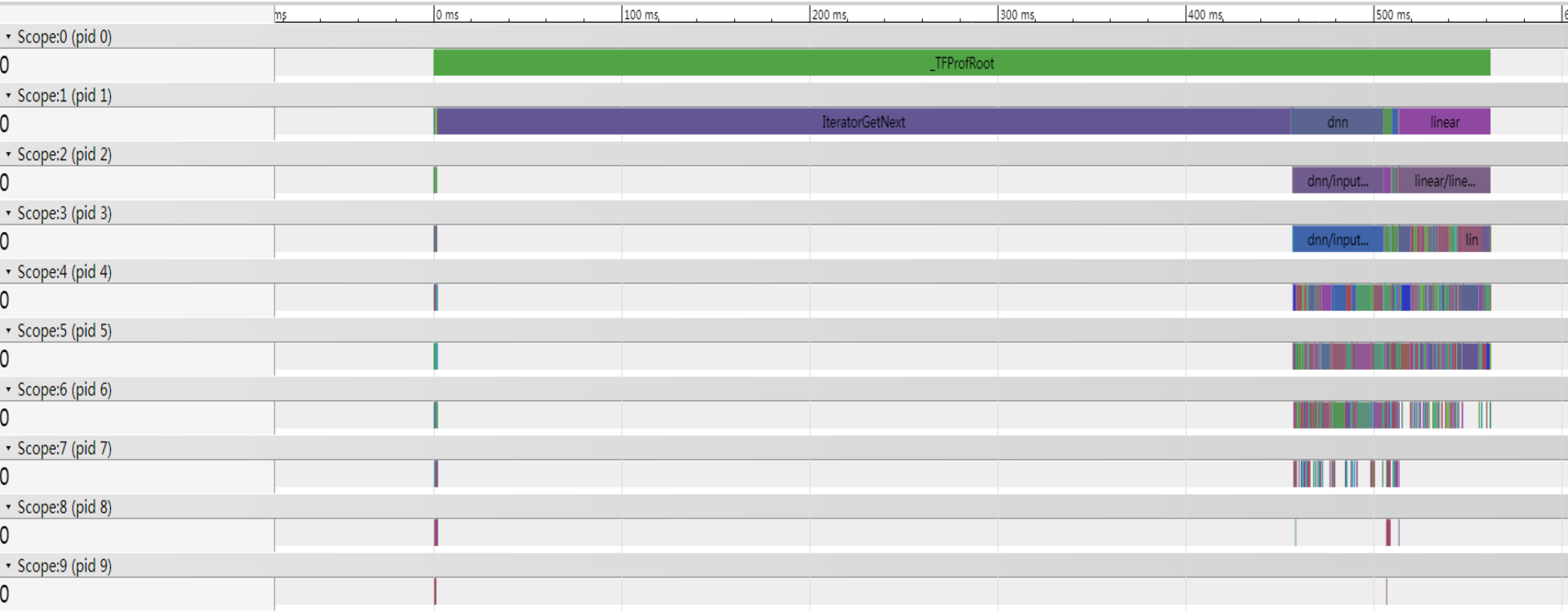
阿里云上推荐场景 – 技术架构

- 训练任务
  - 每天需要训练上千亿样本
- 行为预估
  - 根据用户的行为预估和推荐
- 算法
  - 原来使用逻辑回归和GBDT算法
  - 现在使用LR + DNN算法
- 配置
  - 双M40 GPU卡，56vcpu，96GB内存
  - 10Gb网络
- 预处理、存储：EMR
  - Hadoop做预处理
  - HDFS基于D1实例构建
- 分布式训练：K8S
  - 多GPU卡的分布式调度
  - Tensorflow PS模式调度
  - Tensorflow MPI模式调度



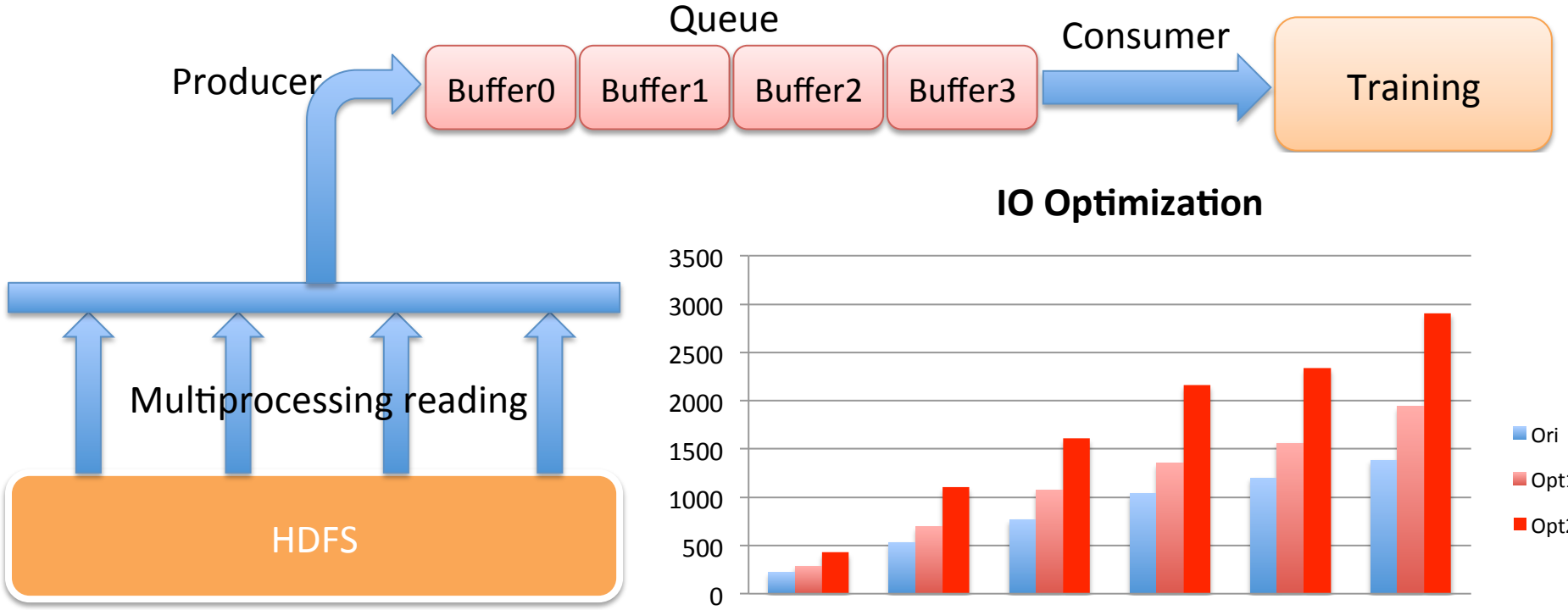
阿里云上推荐场景 – 性能优化I

- 定位性能瓶颈
  - Tensorflow profiler
  - 瓶颈1: 文件读取和解析
  - 瓶颈2: 分布式多机通信



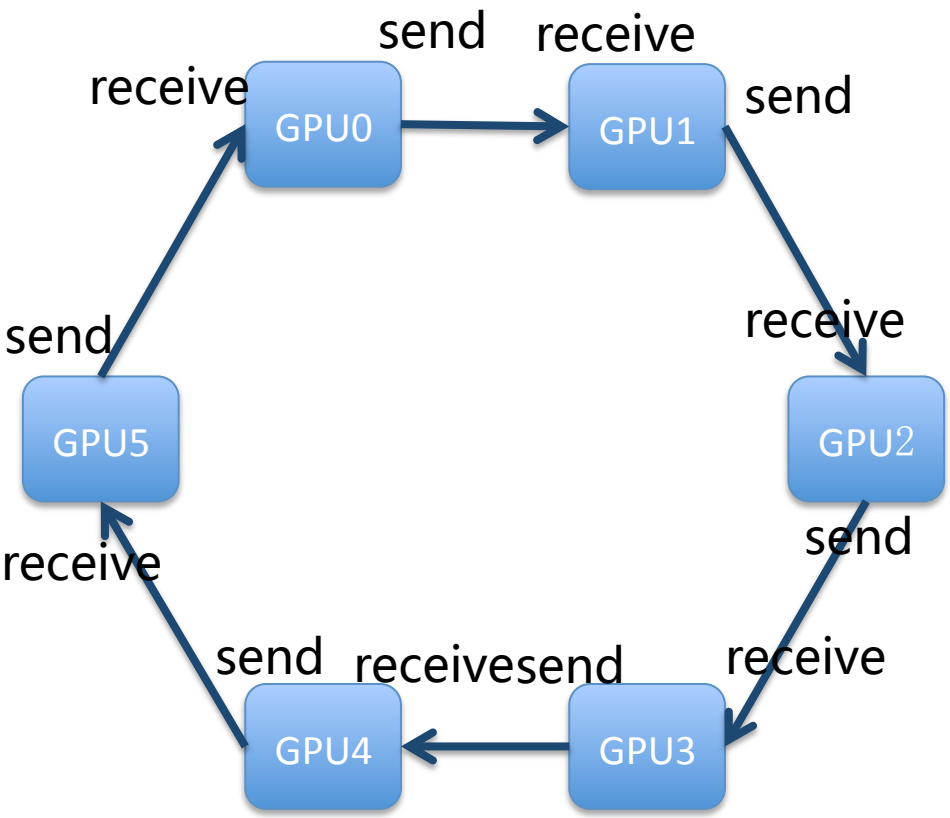
阿里云上推荐场景 – 性能优化II

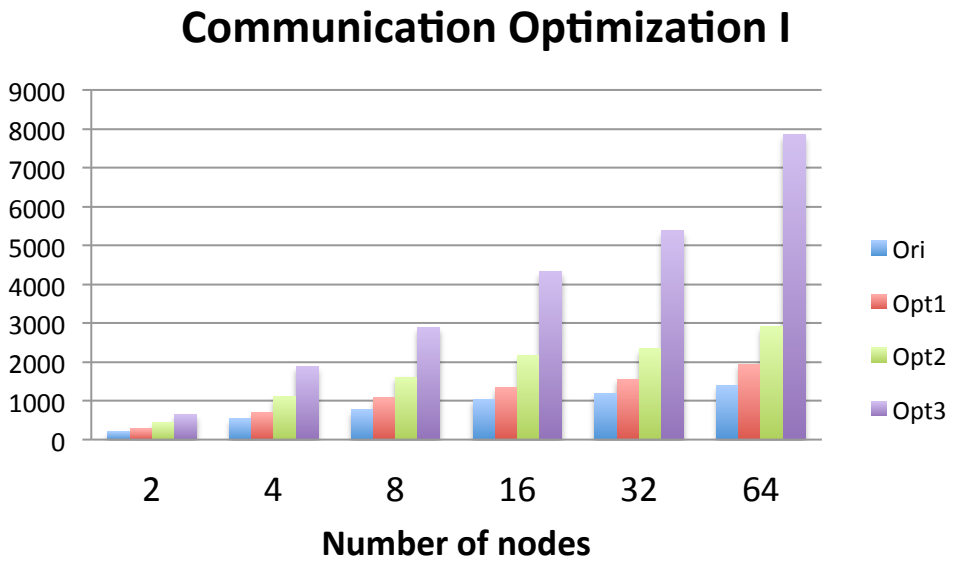
- 优化IO性能
  - 多进程同时从HDFS文件系统里读取大量文件
  - 多缓存队列：让文件读取和计算重叠
  - 64节点比原始性能提升2.1倍



阿里云上推荐场景 – 性能优化III

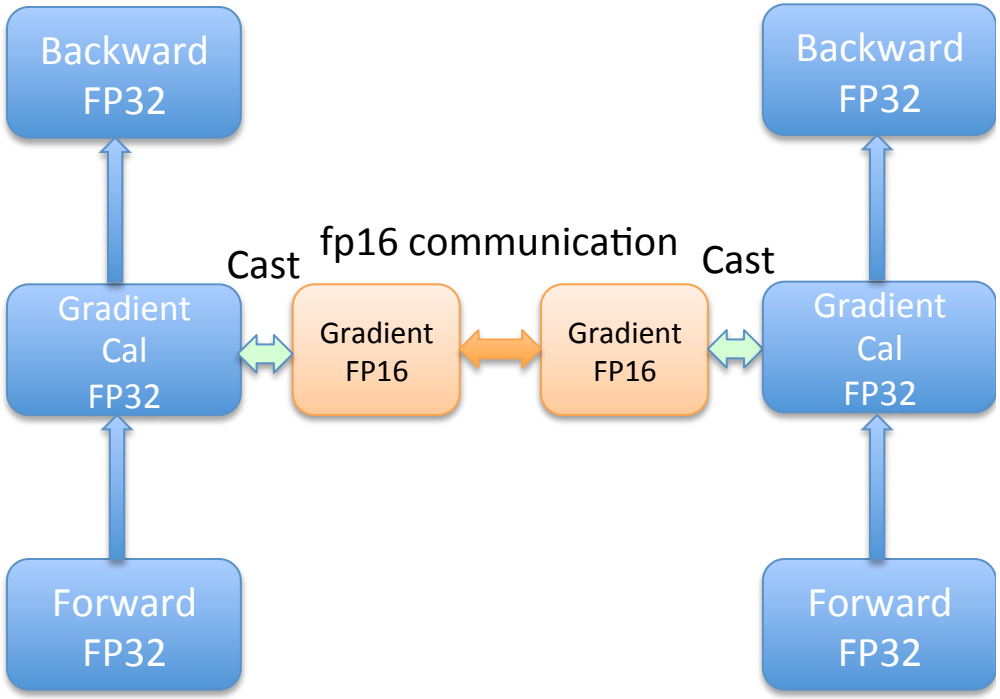
- 优化通信性能
  - 使用MPI通信代替了gRPC通信
  - 使用allreduce环形通信
  - 64节点比原始性能提升2.7倍，累计提升5.7倍

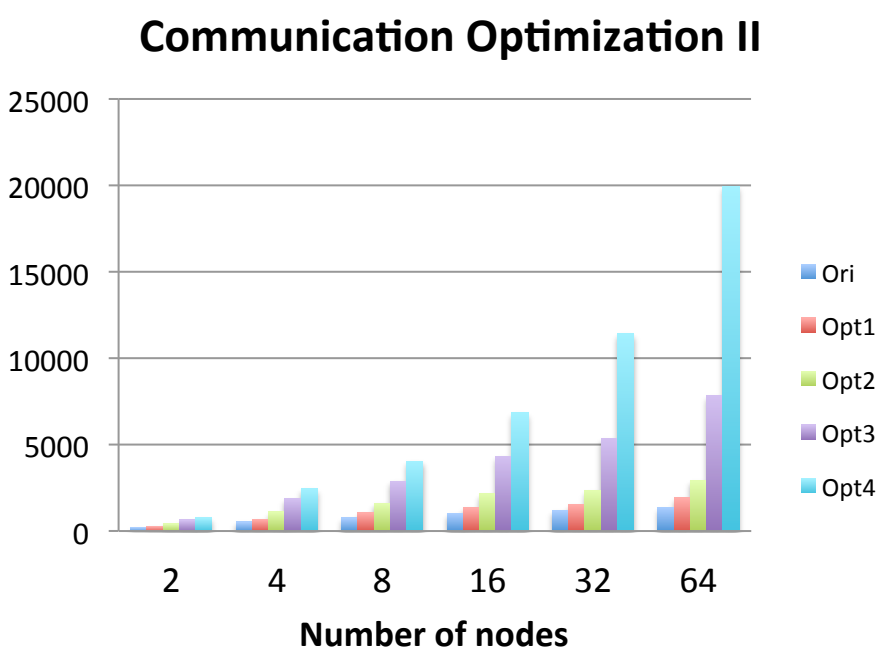




阿里云上推荐场景 – 性能优化IV

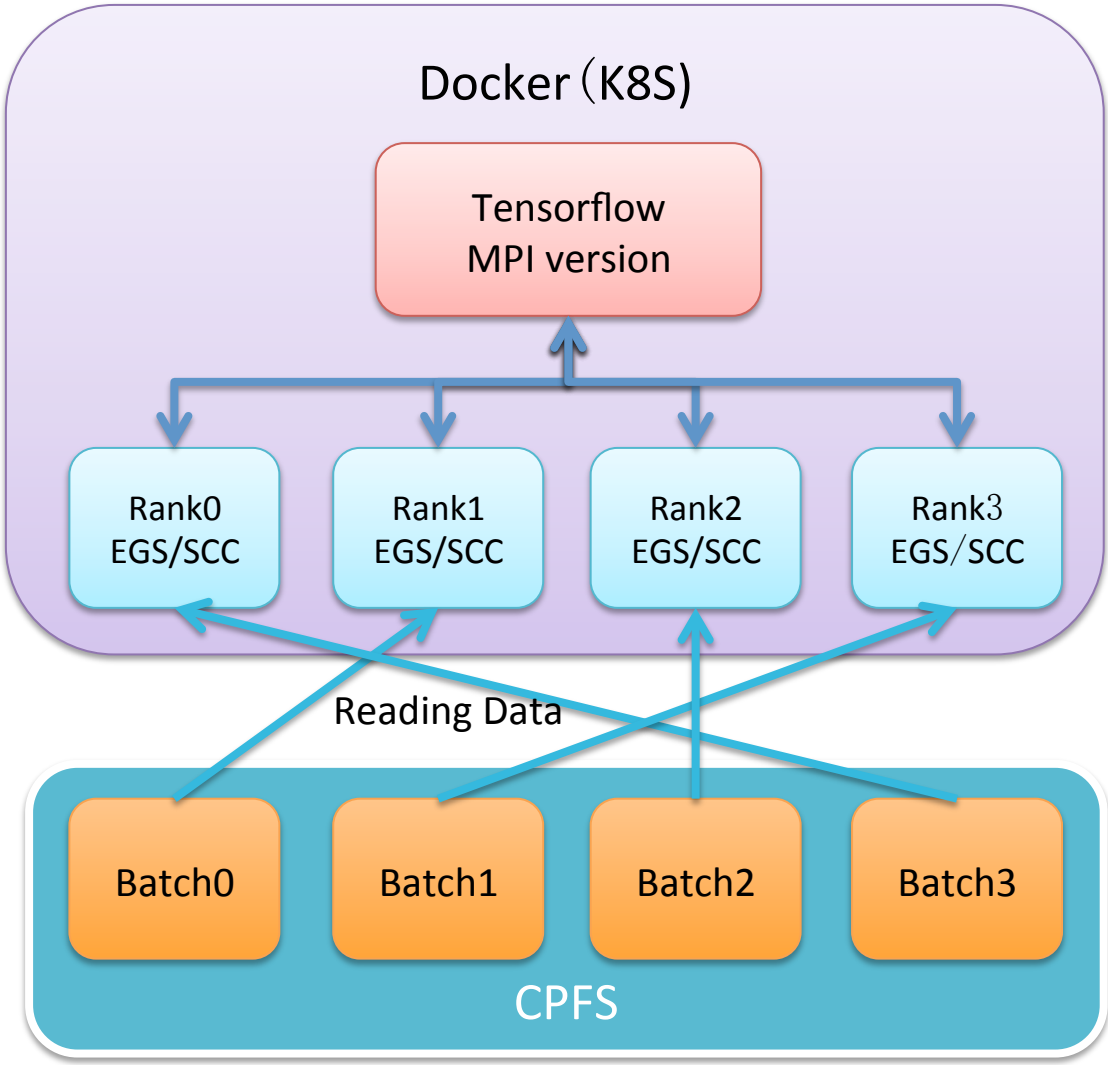
- 优化通信性能
  - 使用FP16通信，带宽压力降低一倍，使用FP32做计算
  - 选择合适的 scaling 值避免下溢
  - 64节点比原始性能提升2.5倍，累计提升14倍





阿里云上图像识别场景 – 技术架构

- 训练任务
  - 一小时内需要训练128万*90epoch=1.15亿张图片
- 算法模型
  - ResNet-50
- 配置
  - 8xP100 GPU卡，56vcpu，480 GB内存
  - 25Gb网络
- 分布式调度：K8S
  - 多GPU卡的分布式调度
  - Tensorflow PS模式调度
  - Tensorflow MPI模式调度
- 存储：
  - CPFS(Lustre on 阿里云)

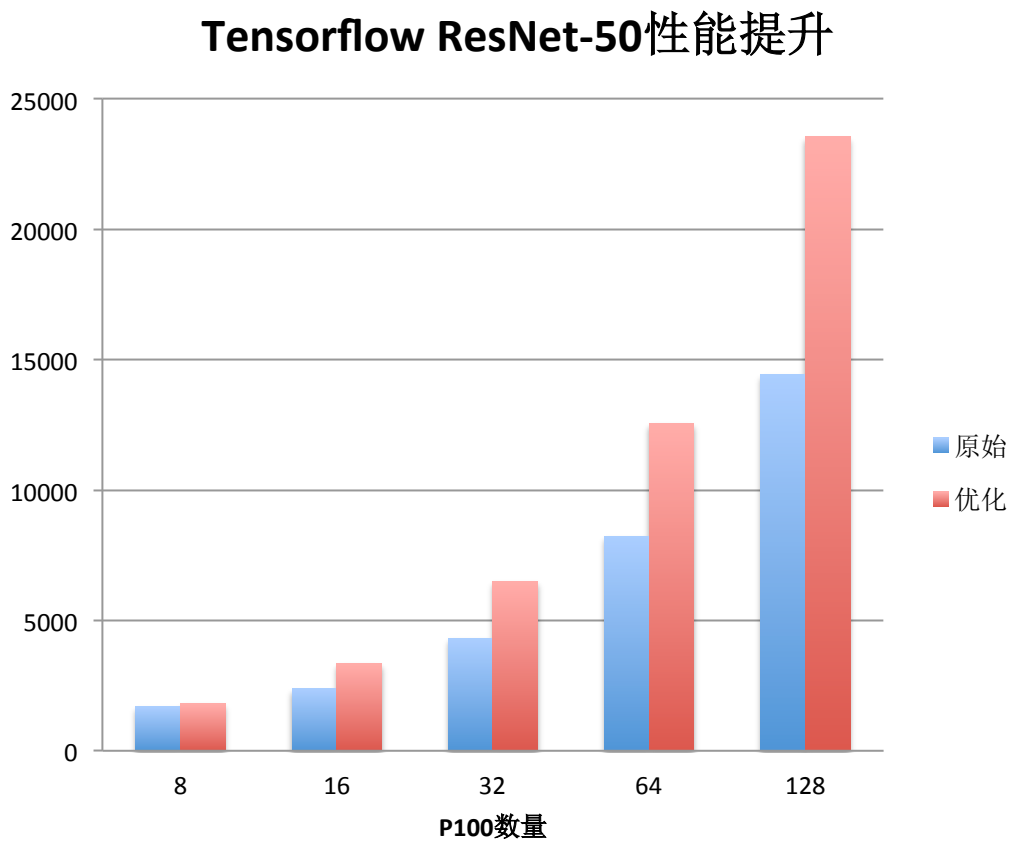


阿里云上图像识别场景 – 性能优化

- MPI + ring allreduce性能优化
  - •移除多对1的通信瓶颈
  - •分散带宽压力
- 计算和通信重叠优化
  - •多层权值融合成一个block
  - •自动调节block大小达到最佳性能
- 性能提升
  - •ResNet-50 128卡性能提升63%

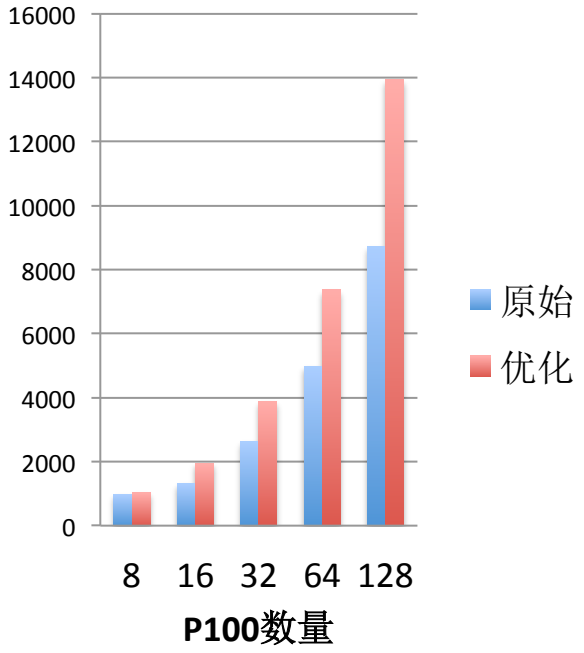
•精度调节

- •0.1;5;3.2;30;0.32;60;0.032;80;0.0032
- •训练Top-1精度为75%，Top5为92.4%



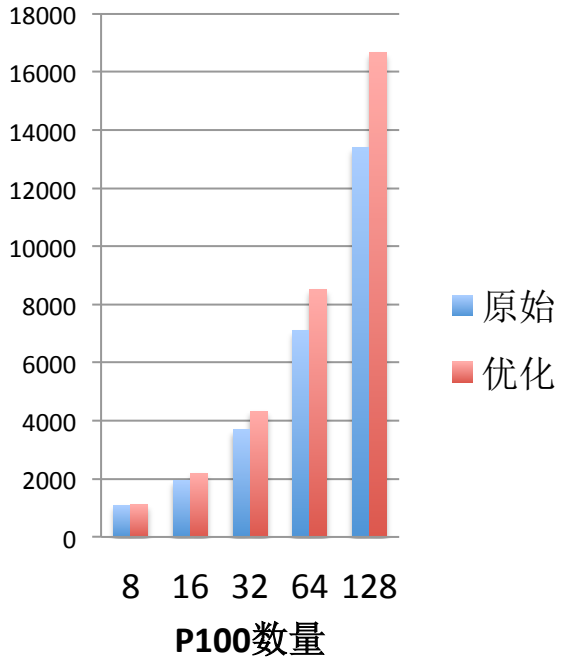
阿里云上图像识别场景 –更多模型性能提升

Tensorflow ResNet-101性能提升



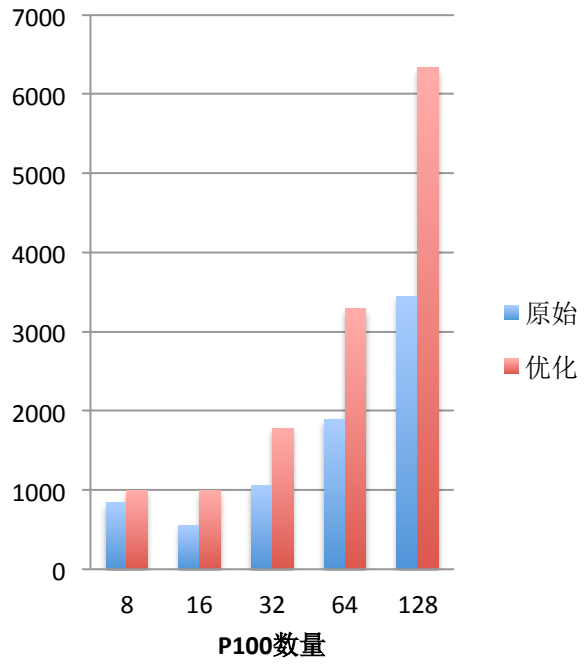
ResNet-101 128卡性能提升70%

Tensorflow Inception-v3性能提升



Inception-v3 128卡性能提升50%

Tensorflow VGG16性能提升



VGG16 128卡性能提升80%

欢迎加入阿里云弹性人工智能团队

当人工智能遇上云计算，一切皆有可能

愿景：加速阿里云上人工智能企业的发展


LINUXCON
containercon
CLOUDOPEN
CHINA 中国
THINK OPEN
开放性思维
LF ASIA, LLC