10 月 27 日，FATE 开源社区第 14 期圆桌会圆满落幕。本次圆桌会，由 FATE 团队资深算法专家马国强老师、资深架构师曾纪策老师，为大家介绍 FATE v1.7 版本的内容。

接下来带大家回顾经典**问答环节**，为各位朋友答疑解惑。

#问答环节

left-Join 那部分，匹配 ID、样本 ID，是说训练过程不会泄露公共样本的真实 ID 吗?

所谓的真实 ID 其实是匹配 ID。纵向联邦算法需要做样本对齐，真实 ID 匹配这一步还是需要的。但是经过 PSI 模块之后，如果开启了匹配 ID 的功能，在实际建模的时候，得到的 ID 列只是样本 ID，类似 uuid 或者 index 这些无实际价值的字符串。

刚刚提到的 BloomFilter,过滤原理是什么，怎么快速得到的候选集?

BloomFilter 主要是用来做粗选，样本比较少的一方，用自身的 ID 去产生一个 BlommFilter，同时要保证有足够的误差。然后将结果发送给其它数据量大的一方，让其它方去过滤出可能的候选集，然后用候选集来进行正常的 PSI 流程。

如果是 1 亿对 1 亿的这种对称的、大量的 PSI，有什么优化建议吗?

1 亿对 1 亿对称的，我们的优化点之一 CRT 优化也是有好几倍的提升的，因为 A 方计算量是 r^e 和 da.= ya / r，这个计算是比较快的，rsa 协议里面，e 一般是 65537=2**16+1，而 d 是和 n 基本是同阶，耗时主要在 B 方（拥有 RSA 私钥的一方）。另外一亿这种数据一般是比较稳定的，该方的数据每次变动不大的话也可以应用离在线拆分的功能。毕竟 A 方即使是一亿，计算量也不大，也会有很大的提升。

1.7 版本里面加密机制，有用到差分隐私吗?

目前没有用到。因为联邦学习的应用，一般目标都是通过联邦建模去训练一个更好的模型，FATE 现阶段的模型效果和中心化机器学习相比，效果基本是无损的。但如果用了差分隐私，联邦情况下比较难去评估真实效果，另外，差分隐私需要在安全性和噪声的选择上的平衡，而更大的 noise 虽然带来了安全性上的提升，但建模效果的影响也会变大。所以 FATE 这个阶段还没有用到差分机制。

1.7 版本什么时候在 GitHub 开源 ？

当前已处于测试收尾阶段，预计 11 月开源。

FATE 可以直接做多方安全计算吗 ？

FATE 框架除了提供多种联邦模型外，也提供了多种多方安全计算协议供用户使用，是可以直接做多方安全计算的。

刚才提到 check point 预测，是可以把模型每轮迭代的模型都预测出来，方便进行对比吗?

是的，我们的架构师在会议上也分享了 checkpoint 的相关知识。用户可以通过 deploy 指令去生成不同迭代轮次的预测模型，然后进行预测。

横向模型训练好了可以迁移出来到其他系统使用吗?

对于横向模型，FATE 提供工具将模型转换成对应的 Scikit-Learn、TensorFlow、Pytorch、LightGBM 模型，然后用户可以部署到支持上述模型的在线推理系统，如 KFServing，以支持在线推理。

啥时候可以从页面新建训练任务?

FATEBoard 主要定位是可读性的可视化操作，我们会把这个功能放在另外的一个可视化的系统。

权限管理和数据追朔，主要是为了怎么做/谁来做审计？

商业化场景下，在构建一个的合作网络的时候，合作身份及权限管理是必不可少的，这是保护本方资源的一种手段，如数据集资源、计算资源等。数据追溯是指一份数据包括其衍生数据的完整使用记录，使用记录可以包含使用数据的作业、站点身份、角色、哪些算法组件等等，可以比较清晰的描述了数据流动的完整周期，用以满足各维度的安全审计。

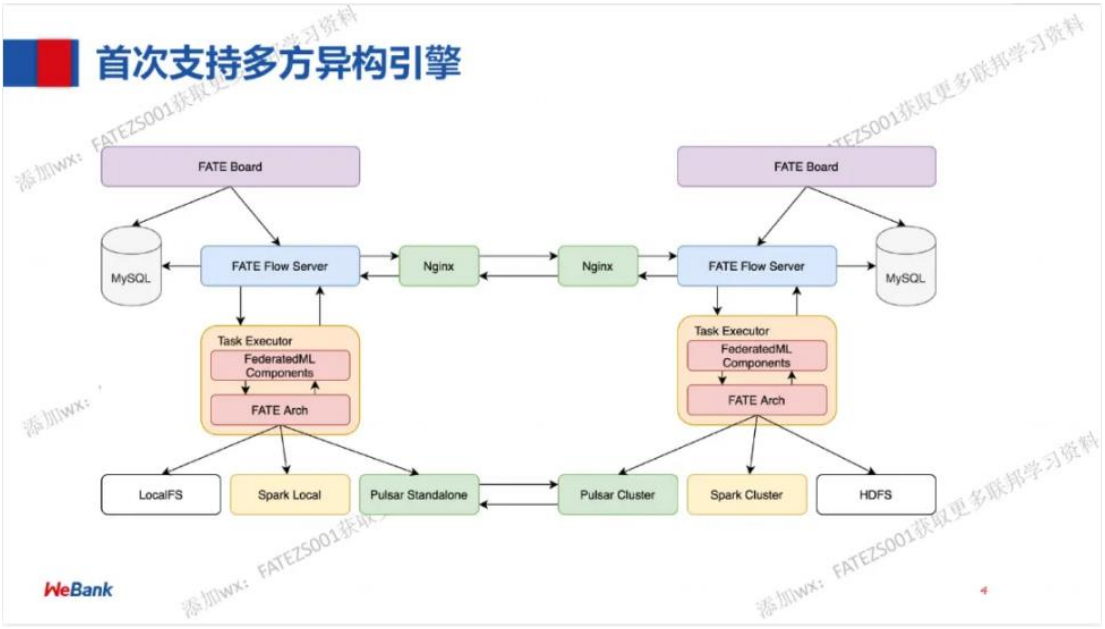
啥时候可以从页面新建训练任务？

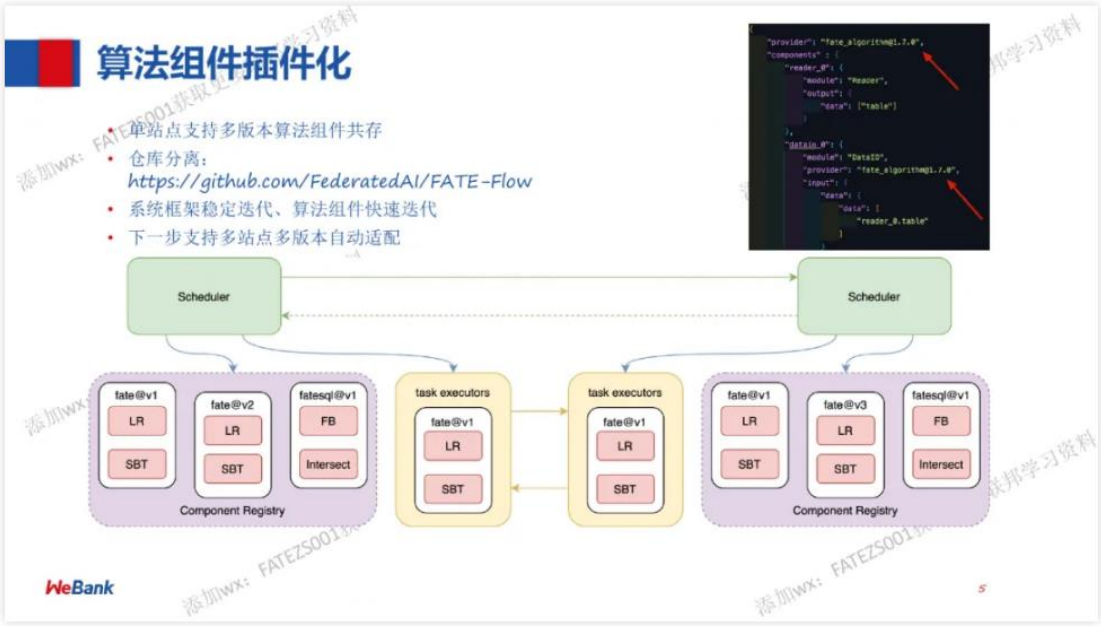
FATEBoard 主要定位是可读性的可视化操作，我们会把这个功能放在另外的一个可视化的系统。

审计功能可以再详细介绍下吗？有提供可视化过程和报表吗？

审计是 FATE 会逐步加强的一个重要功能，当前 1.7 版本做了一些基础的信息收集工作，例如记录每个作业用到的数据集、数据集的来源父数据集等。

以下为本次圆桌会的部分内容介绍，添加小助手（FATEZS001）可获取详细资料：





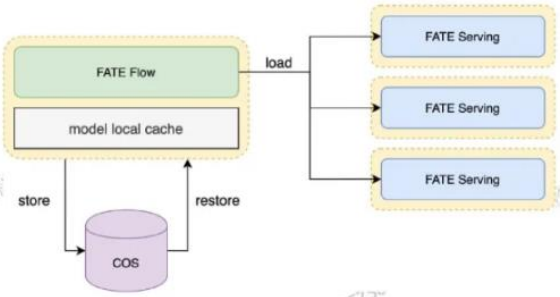
模型&数据复用能力，加速训练

- Reader组件，支持输入已有作业某组件输出数据
- Model Loader新组件，输入已有作业生成的模型
- Cache Manager，新增缓存管理模块
  - 组件新增输出Cache，目前支持交集组件
  - 使用Cache Loader新组件，输入已有作业生产的数据输出

容错能力增强，生产友好

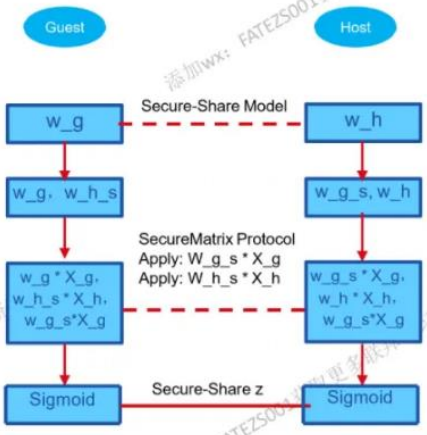
- 组件任务失败自动重试：auto_retris + warm_start
- 修改更新组件运行参数，并从指定组件重新运行：update_parameters + rerun
- 自动推送上线模型到远端可靠存储：load + store
- 推送上线模型自动从远端可靠存储恢复：restore + load

```
{
  "job_id": "20210806124916233456O",
  "component_parameters": {
    "common": {
      "hetero_lr_0": {
        "alpha": 0.02,
        "batch_size": 320,
        "learning_rate": 0.15
      }
    }
  }
}
```

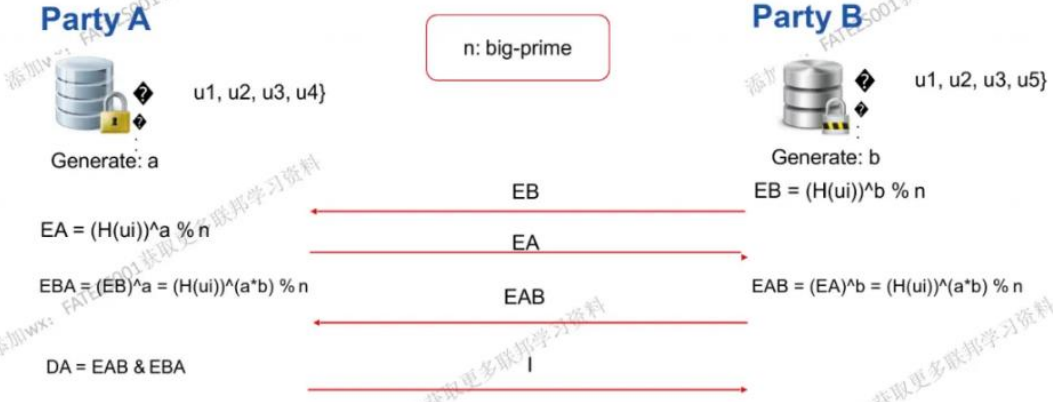


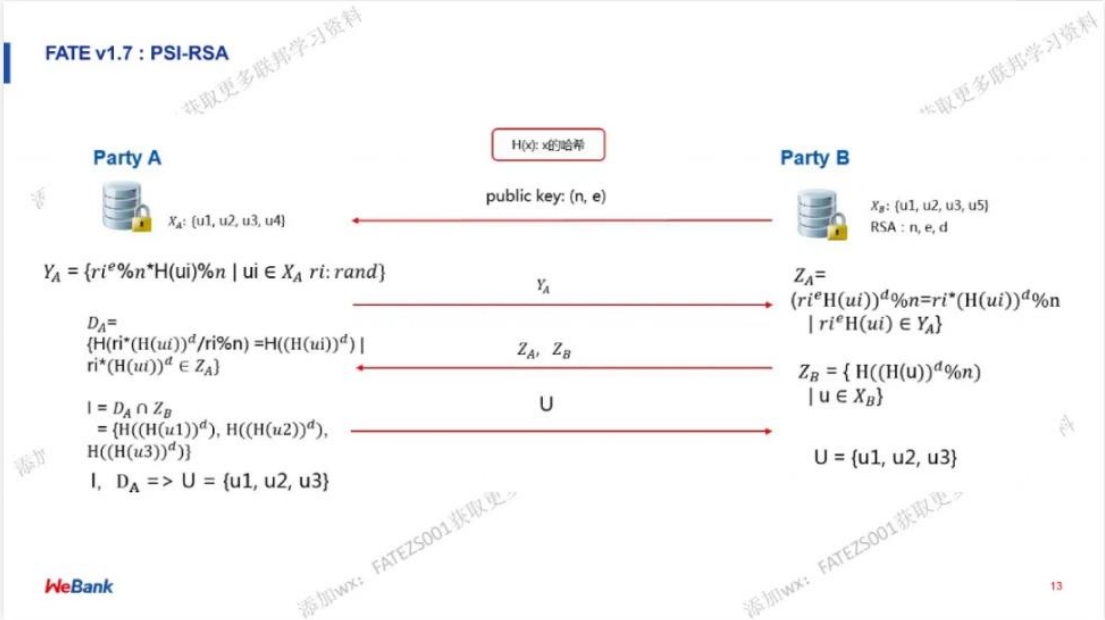
FATE v1.7 - SSHE LR

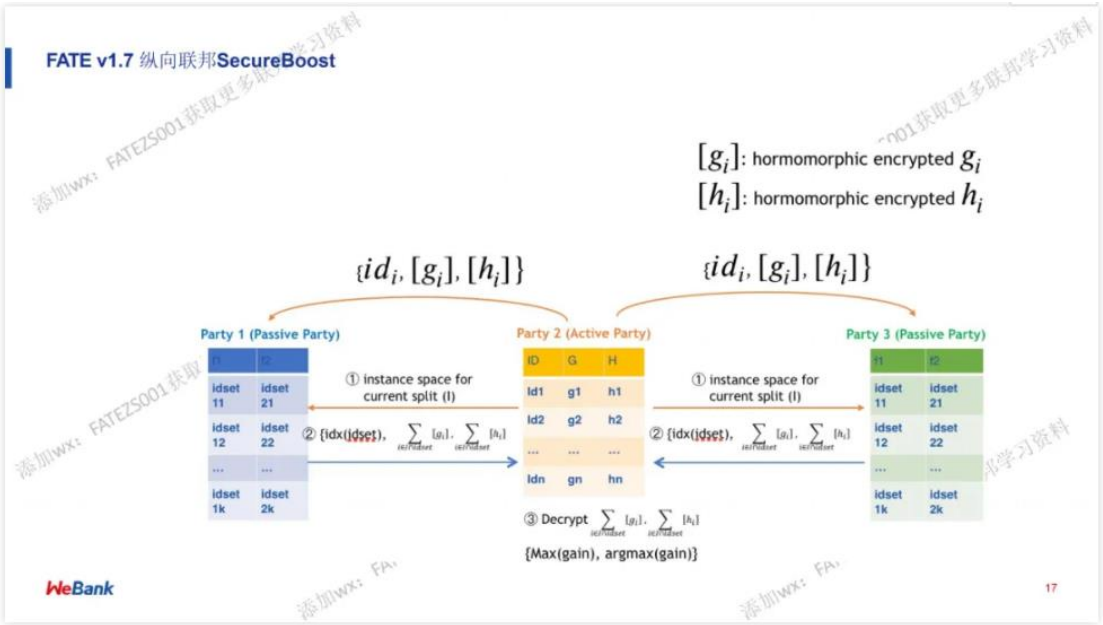
- **SSHE-LR**
  - 无可信第三方
  - 使用**HE + SecretShare**新混合安全协议
    - 两种迭代模式：
      - 每轮**w**重构
      - **w**最终轮重构
  - 支持模型在**Guest**端重构，**host**端使用加密模型预测



FATE v1.7 - PSI-DH







获取会议 PPT，或对圆桌会还有别的疑问？欢迎联系 FATE 开源社区助手获得帮助。

原文链接：https://mp.weixin.qq.com/s/dDkZS-wyc_Btk1ATBbLqXA