可信智能计算服务

开发指南

文档版本 02

发布日期 2025-06-24

华为技术有限公司

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：
https://www.huawei.com/cn/psirt/vul-response-process
如企业客户须获取漏洞信息，请参见如下网址：
https://securitybulletin.huawei.com/enterprise/cn/security-advisory

目 录

1 服务介绍 1
1.1 应用开发简介 1
1.2 常用概念 1
1.3 开发流程 2
1.4 开发规范 3
2 环境准备 5
2.1 开发环境简介 5
2.2 参考：获取认证信息 5
2.3 配置 CCE 服务 6
2.4 配置 IEF 服务 7
2.5 TICS 服务委托授权 8
2.6 配置 IEF 高可用节点 9
2.7 购买 TICS 服务 11
2.8 部署计算节点 17
2.9 创建连接器 23
3 使用场景 28
3.1 多方安全计算场景 28
3.1.1 场景描述 28
3.1.2 组合架构 29
3.1.3 可验证代码示例 29
3.1.3.1 数据准备 30
3.1.3.2 数据集发布 32
3.1.3.3 隐私规则防护 35
3.1.3.4 基本计算能力验证 36
3.1.3.5 基于 MPC 算法的高安全级别计算 38
3.1.3.6 统计型作业的差分隐私保护 39
3.2 纵向联邦建模场景 42
3.2.1 使用 TICS 多方安全计算进行联合样本分布统计 42
3.2.1.1 场景描述 42
3.2.1.2 准备数据 42
3.2.1.3 发布数据集 44
3.2.1.4 创建样本分布统计作业 45

3.2.1.5 执行样本分布联合统计 48
3.2.1.6 数据优化 48
3.2.2 使用 TICS 可信联邦学习进行联邦建模 49
3.2.2.1 场景描述 49
3.2.2.2 准备数据 49
3.2.2.3 发布数据集 51
3.2.2.4 创建可信联邦学习作业 52
3.2.2.5 选择数据 52
3.2.2.6 样本对齐 53
3.2.2.7 筛选特征 53
3.2.2.8 模型训练 54
3.2.2.9 模型评估 56
3.2.3 使用 TICS 联邦预测进行新数据离线预测 56
3.2.3.1 场景描述 56
3.2.3.2 准备数据 57
3.2.3.3 发布数据集 58
3.2.3.4 创建联邦预测作业 59
3.2.3.5 发起联邦预测 59
3.3 隐私求交黑名单共享场景 60
3.3.1 场景描述 61
3.3.2 准备数据 61
3.3.3 发布数据集 61
3.3.4 创建并运行隐私求交作业 62
3.3.5 查看求交结果 62
3.4 实时隐匿查询场景 63
3.4.1 外部数据共享 63
3.4.1.1 场景描述 64
3.4.1.2 准备数据 64
3.4.1.3 发布数据集 65
3.4.1.4 创建实时隐匿查询作业 65
3.4.1.5 执行实时隐匿查询作业 66
3.5 可信数据交换场景 67
3.5.1 场景描述 67
3.5.2 创建数据 67
3.5.3 申请使用数据 67
3.5.4 审批数据申请 69
3.5.5 创建合约 70
3.6 横向联邦学习场景 72
3.6.1 场景描述 72
3.6.2 测试步骤 73
3.6.2.1 数据准备 73
3.6.2.2 训练型横向联邦作业流程 75

3.6.2.3 评估型横向联邦作业流程 77
3.6.3 实验结果 78
3.6.3.1 乳腺癌数据集作业结果 78

1 服务介绍

1.1 应用开发简介

多方安全计算是可信智能计算服务（TICS）提供的关系型数据安全共享和分析功能。

您可以创建多方安全计算作业，根据合作方已提供的数据，编写相关SQL作业并获取您所需要的分析结果，能够在作业运行的同时保护数据使用方的数据查询和搜索条件，避免因查询和搜索请求造成的数据泄露。

1.2 常用概念

合作方、参与方：

空间成员，有权使用空间中的数据，或者将自有数据发布到空间，供其他合作方受限使用。

计算节点

部署在参与方侧，是可信智能计算与合作方侧数据的桥梁，保障数据按照合作方意愿受限使用。

计算节点是管理参与方数据的最小单位。部署计算节点时需要指定空间配置信息。在计算节点中支持配置连接器，注册数据集，任务执行，查看任务执行日志。

连接器（Connector）

连接器是可信智能计算节点内置的连接特定数据源所需的对象模板，目前支持连接MRS Hive、MySQL、RDS、DWS、ORACLE等多种连接器，并支持扩展增加新的连接器。

数据集（Data set）

数据集为计算节点获取并配置的合作方数据的元数据信息，以及附加其上的隐私策略。

作业（Job）

作业是指用户创建的分析、学习任务。

1.3 开发流程

图 1-1 开发流程

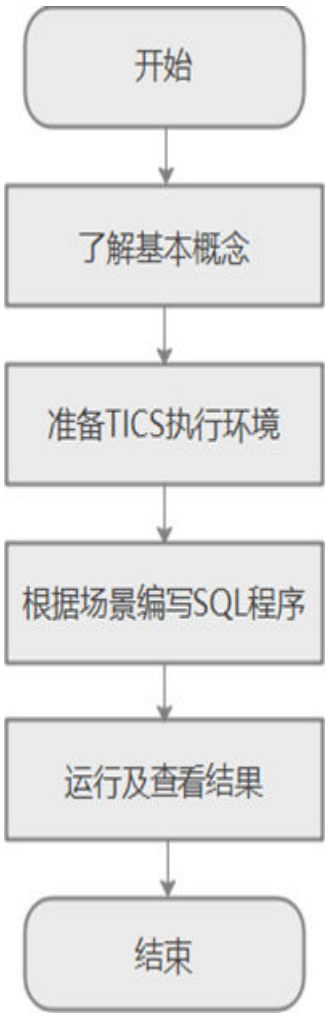


表 1-1 开发流程

阶段	说明	参考文档
了解基本概念	在开始开发前，需要了解多方安全计算的基本概念。	常用概念
准备TICS执行环境	TICS执行环境当前依赖TICS空间、计算节点和连接器。	环境准备

阶段	说明	参考文档
根据场景编写sql程序	当前多方安全计算支持通过编写sql语句，来构建多方安全计算业务场景的计算任务。	使用场景
运行程序及查看结果	指导用户将开发好的sql在计算节点控制台进行提交运行，并查看结果。	可验证代码示例

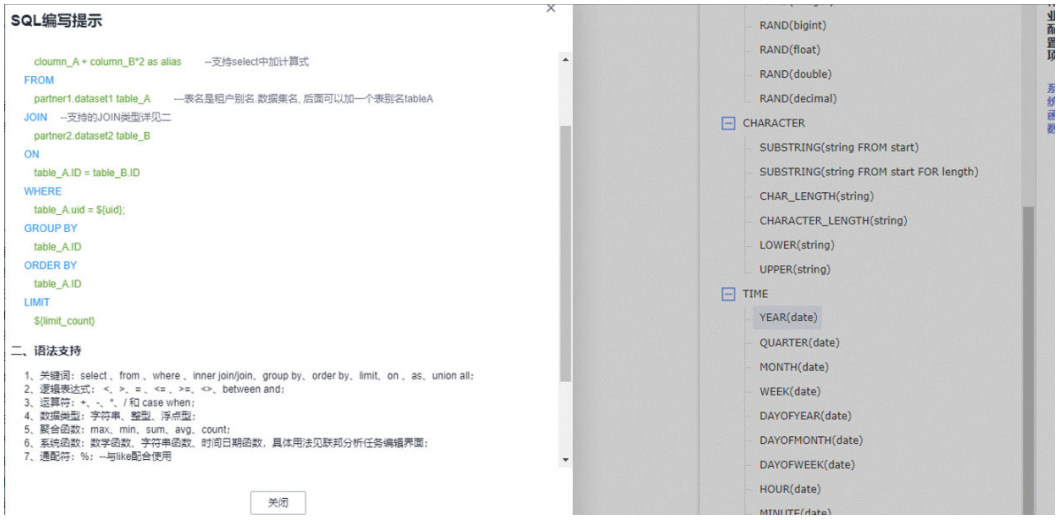
1.4 开发规范

规则

多方安全计算中，基础的sql语法都能够支持，但无法支持所有特殊语法。

语法规则如下：

图 1-2 语法规则



SQL Compilation Prompt

```
WHERE
  table_A.uid = ${uid};
GROUP BY
  table_A.ID
ORDER BY
  table_A.ID
LIMIT
  ${limit_count}
```

2. Supported Standard Syntax (Security MPC and Trusted Hardware)

1. Keywords: SELECT, FROM, WHERE, INNER JOIN/JOIN, GROUP BY, ORDER BY, LIMIT, ON, AS, UNION ALL
2. Logical expressions: <, >, =, <=, >=, <>, between and
3. Operators: +, -, *, /, CASE WHEN
4. Data types: string, integer, floating point
5. Aggregate functions: MAX, MIN, SUM, AVG, COUNT, MEDIAN, VARIANCE
6、System functions: mathematical, string, and time and date functions. For details, see the page for editing federated analysis tasks.
7、Wildcard: % (used with like)

3. Precautions

1. The data type that cannot be identified is considered as the string type.
2. A maximum of 100 results can be displayed on the page. Other results need to be queried in the work directory of the storage compute node.
3. If you set Privacy Protection Level to High, the sensitive fields of the multi-party computation will be encrypted in homomorphic mode. However, sensitive field values can only be integers or floating points, and multiplied by non-sensitive values or constant values.
4. If you set Privacy Protection Level to High, the join fields of the 2^n calculation use the PSI algorithm to output the collision ciphertext data, and only join conditions such as a=b are supported.

建议及示例

查询示例中两表join场景，建议将大表置于join左侧，小表置于join右侧，可借助初筛的能力，进行小表在大表端的加密过滤，提升性能。

- 建议示例：
  Select sum(l_tax+ s_acctbal) from **league_creator.lineitem_1000w b** join **league_partner1.supplier_1w a** on a.s_suppkey = b.l_suppkey
- 不建议示例：
  Select sum(l_tax+ s_acctbal) from **league_partner1.supplier_1w a** join **league_creator.lineitem_1000w b** on a.s_suppkey = b.l_suppkey

2 环境准备

2.1 开发环境简介

在进行多方安全计算应用开发时，要准备的环境如表1所示。

同时需要准备运行调测的Linux环境，用于验证应用程序运行正常。

表 2-1 准备项

准备项	说明
购买TICS服务	在TICS控制台通过下单建立数据空间，或者将租户加入已有的数据空间。
部署计算节点	在TICS控制台通过购买计算节点，支持接入数据空间进行操作。
创建连接器	在计算节点中，通过连接器连接数据源，用于后续的加密计算操作。
网络	确保计算节点能够与TICS空间部署节点互联互通。

2.2 参考：获取认证信息

在使用TICS时，您可能需要获取访问密钥、项目ID等信息，获取方式如下：

获取访问密钥

您可以通过如下方式获取访问密钥。

1. 登录控制台，在用户名下拉列表中选择“我的凭证”。
2. 进入“我的凭证”页面，选择“访问密钥 > 新增访问密钥”，如图2-1所示。

图 2-1 单击新增访问密钥

3. 单击“确定”，根据浏览器提示，保存密钥文件。密钥文件会直接保存到浏览器默认的下载文件夹中。打开名称为“credentials.csv”的文件，即可查看访问密钥（Access Key Id和Secret Access Key）。

说明

- 每个用户仅允许新增两个访问密钥。
- 为保证访问密钥的安全，访问密钥仅在初次生成时自动下载，后续不可再次通过管理控制台界面获取。请在生成后妥善保管。

获取项目 ID 和账号 ID

项目ID表示租户的资源，账号ID对应当前账号。用户可在对应页面下查看不同Region对应的项目ID和账号ID。

1. 注册并登录管理控制台。
2. 在用户名的下拉列表中单击“我的凭证”。
3. 在“API凭证”页面，查看账号名和账号ID，在项目列表中查看项目ID。

2.3 配置 CCE 服务

背景信息

如果您规划在购买TICS服务时选择基于“**云租户部署**”，则您在购买TICS服务前需要对CCE服务进行相关配置，避免影响TICS服务的正常使用。

注意

请自行关注部署节点的系统安全防护与配置加固，确保机器在安全的前提下进行隐私计算节点部署。

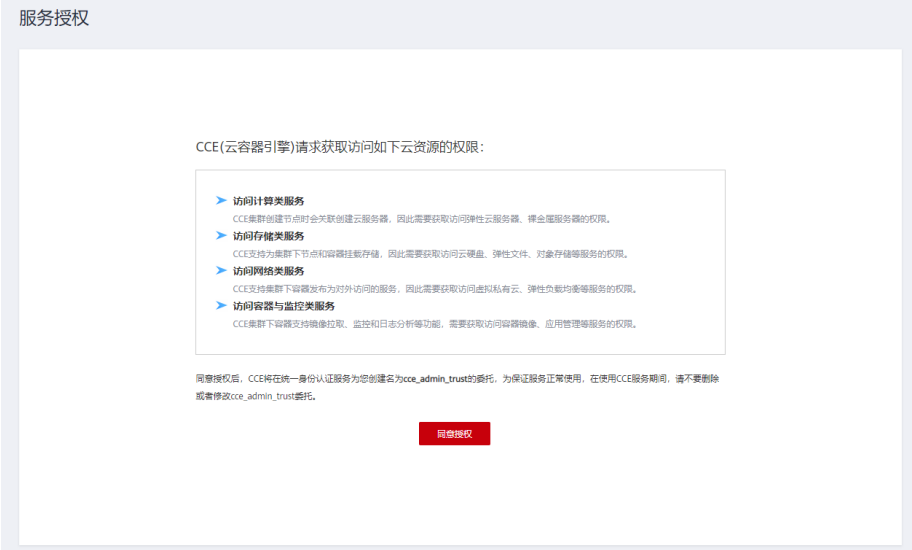
CCE 服务委托授权

由于CCE在运行中对计算、存储、网络以及监控等各类云服务资源都存在依赖关系，因此当您首次登录CCE控制台时，CCE将自动请求获取当前区域下的云资源权限，从而更好地为您提供服务。

说明

CCE的服务授权为全局配置，只要您所使用的账号在当前Region曾经进行过服务授权，则无需重新配置，可以跳过本节操作。

图 2-2 服务授权



当您同意授权后，CCE将在IAM中创建名为“cce_admin_trust”委托，统一使用系统账户“op_svc_cce”对您的其他云服务资源进行操作，cce_admin_trust委托具有Tenant Administrator权限。Tenant Administrator拥有除IAM管理外的全部云服务管理员权限，用于对CCE所依赖的其他云服务资源进行调用，且该授权仅在当前区域生效。关于资源委托详情，您可参考**委托**进行了解。

2.4 配置 IEF 服务

背景信息

如果您规划在购买TICS服务时选择基于“**边缘节点部署**”，则您在购买TICS服务前需要对IEF服务进行相关配置，避免影响TICS服务的正常使用。

⚠ 注意

请自行关注部署节点的系统安全防护与配置加固，确保机器在安全的前提下进行隐私计算节点部署。

IEF 服务委托授权

使用主账号访问IEF服务首页，单击“同意授权”，IEF将在统一身份认证服务为您创建名为ief_admin_trust的委托。

图 2-3 IEF 服务授权

ief_admin_trust委托具有Tenant Administrator权限。Tenant Administrator拥有除IAM管理外的全部云服务管理员权限，用于对IEF所依赖的其他云服务资源进行调用，且该授权仅在当前区域生效。

2.5 TICS 服务委托授权

背景信息

为保证正常创建TICS服务，需要先设置服务委托。

前提条件

- 服务授权需要主账号或者admin用户组中的子账号进行操作。
- 授权委托需查看IAM委托列表，如果存在名为tics_admin_trust的委托和tics_role_trust的权限，需要先删除。

服务授权操作

步骤1 进入TICS服务控制台，为保证正常创建TICS服务，需要先设置服务委托。

步骤2 进入计算节点购买页面，在“部署配置”区域，设置部署方式为“边缘节点部署”，在弹出的对话框单击“同意授权”。

同意授权后，TICS将在统一身份认证服务IAM下为您创建名为tics_admin_trust的委托，委托绑定的权限名为tics_role_trust。授权成功后，可以进入委托列表查看。

图 2-4 授权访问权限名

说明

委托tics_admin_trust和权限tics_role_trust创建成功后，请勿删除。

表 2-2 TICS 委托权限列表

权限名	详细信息	备注
tics_role_trust	TICS服务计算节点依赖IEF作为底层资源，因此需要tics_role_trust角色来部署应用。	由于云服务缓存需要时间，该权限3分钟左右才能生效。

----结束

2.6 配置 IEF 高可用节点

IEF高可用节点实现该功能要手动操作，使用rsync命令在多台虚机间定时同步文件，操作步骤如下：

说明

以下教程适用于ECS机器系统为Centos 7.5。操作前需要购买两台同网段同文件系统的ecs节点A与节点B。

步骤1 在两台虚机上安装rsync及corntab服务，已安装则跳过（HCS底座发行的系统镜像是默认安装的；客户提供的机器，需要客户运维侧保障）。

步骤2 参照**如何在两个节点间免密ssh登录**完成节点免密设置。

步骤3 在节点A任意目录下创建该脚本sync_tics.sh，建议放在 /opt/tics目录下，确保脚本文件具备可执行权限。

```
#!/bin/bash
if [[ -n $(docker ps | grep k8s_db) ]];then
```

```
  echo "has install postgres"
  rsync -avzrog --exclude=postmaster.pid /var/lib/tics_db/ 对端节点ip:/var/lib/tics_db/
fi
  chmod 755 /home/tics/
  rsync -avzrog /home/tics/ 对端节点ip:/home/tics/
```

步骤4 在节点A上执行如下命令启动定时同步任务。

```
crontab -e
```

在弹出的编辑框中输入。

```
*/1 * * * *  /opt/tics/sync_tics.sh
```

保存后退出。

步骤5 在节点B上重复**步骤3~步骤4**操作，注意**步骤3**中脚本内容应替换为对端节点A的ip。

说明

使用tail /var/log/cron 可以查看定时命令执行情况，务必保证同步命令执行正常。

----结束

如何在两个节点间免密 ssh 登录

步骤1 登录机器A，执行如下命令

```
ssh-keygen
```

```
[root@ecs-9662 ~]# ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/root/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /root/.ssh/id_rsa
Your public key has been saved in /root/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:JKe58nBb03xRKFlHZegOxISRxeKgWs1cLmEW4i4npBs root@ecs-9662
The key's randomart image is:
+---[RSA 3072]----+
|       . ...O+.ooo|
|      . .= =o+o.. |
|     . oBo*oo...  |
|    o .o*= o....  |
|   E oo= S.  .o   |
|    o.+ . o   ..  |
|   . o o o o .    |
|      = o . .     |
|       o          |
+----[SHA256]-----+
```

遇到需要Overwrite(y/n)时输入y，其他提示均回车即可

步骤2 在机器A上继续执行如下命令，按照提示输入B的登录密码即可

```
ssh-copy-id -i 图中红框部分 root@机器B的ip
```

注：以上操作为节点采用密钥登录，无密码的场景下

若所建节点采用密钥对登录的形式，可手动复制公钥文件id_rsa.pub到对端节点的指定用户的home路径下（root用户的路径为/root）

在对端节点下操作：

查看指定用户home目录下有无.ssh文件夹，没有的话创建一个，复制中的id_rsa.pub的内容到authorized_keys文件

```
[root@yuancheng ~]# cd .ssh
[root@yuancheng .ssh]# cat ../id_rsa.pub | tee -a authorized_keys
ssh-rsa
AAAAB3NzaC1yc2EAAAADAQABAAABAQDINuohcfbWG8DMHY7mwnAIkp7jglJczOrk1ie5stdSF9GLroot@yuan
cheng
[root@yuancheng .ssh]# ll
total 12
-rw-r--r-- 1 root root  408 Aug 10 09:58 authorized_keys
```

设置authorized_keys文件的权限为600

步骤3 在机器B上执行1、2步骤。

步骤4 接下来两台机器，即可相互直接ssh不需要输入密码

----结束

2.7 购买 TICS 服务

前提条件

购买TICS服务前，已完成**配置CCE服务**、**配置IEF服务**和**TICS服务委托授权**。

购买 TICS 服务并进入控制台

购买TICS服务即创建空间。一个空间的成员包括组织方和合作方。用户参与的空间情况，可以在“空间管理”中查看。

1. 以主账号登录管理控制台。在控制台左上方，单击“服务列表”按钮 ☰ ，选择“EI企业智能 > 可信智能计算服务 TICS”，进入TICS控制台。

   如果需要以IAM子账号购买TICS服务，则需要先授予IAM子账号相应权限，详情请参见**CCE服务委托授权**和**IEF服务委托授权**。
2. 在TICS控制台页面，单击“购买可信智能计算服务”。

   配置购买参数，各参数说明如**表2-3**和**表2-4**所示。

表 2-3 空间信息配置参数

参数名称	样例	说明
区域	华北-北京四	选择服务的区域，不同区域的资源之间内网不互通。
项目	cn-north-4	选择该区域内的项目。
空间名称	TICS-test	由用户自定义，用以区分各个空间。要求：空间名称不允许重复，名称不能以空白字符开头或结尾，也不能包含下列特殊字符：\ / : * ? " < > |，长度要求在1~128之间。
区块链存证	打开	若您希望空间启用区块链服务（BCS）来审计任务信息，请打开此选项。 使用前需要按照**启用区块链审计服务（可选）**章节的描述完成准备工作。

参数名称	样例	说明
BCS服务实例	-	选择BCS空间链。
通道	-	选择邀空间链邀请租户时选择的通道。
组织	-	选择链代码部署的组织。
区块链签名证书	-	上传签名证书文件（选择按照**启用区块链审计服务（可选）**章节步骤七描述中保存至本地的证书文件“/msp/signcert/xxx.pem”）。
区块链私钥文件	-	上传私钥证书文件（选择按照**启用区块链审计服务（可选）**章节步骤七描述中保存至本地的证书文件“/msp/keystore/xxx_sk”）

表 2-4 计算节点配置参数

参数名	样例	参数描述
计费方式	包年/包月	当前仅支持“包年/包月”。
购买时长	1个月	支持按月或按年购买。
自动续费	-	支持自动续费。 ● 按月购买时，自动续费周期为1个月。 ● 按年购买时，自动续费周期为1年。
版本类型	企业版	当前可选版本只包含企业版
计算节点配置相关参数		
计算节点名称	-	计算节点别名，由用户自定义，用以区分部署的各个计算节点。要求：名称不能以空白字符开头或结尾，也不能包含下列特殊字符：\ / : * ? " < > |，长度要求在1~128之间。
访问密钥ID（AK）	-	用户的身份标识，需要用户去IAM服务自行下载。文件获取方式请参考**获取访问密钥**章节。AK、SK需与用户当前所在的项目保持一致。 **说明** ● 如果访问密钥泄露，会带来数据泄露风险。 ● 每个访问密钥只能下载一次，为了账号安全性，建议定期更换访问密钥并妥善保存。
加密密钥（SK）	-	
计算节点登录名称	-	登录计算节点控制台的用户名。用户可通过“计算节点登录名称”和“登录密码”进入计算节点控制台，建立连接器，发布数据。
登录密码	-	登录计算节点控制台的密码。
确认密码	-	与“登录密码”保持一致即可。

参数名	样例	参数描述
支持国密	否	若选择是，则登录计算节点必须使用国密浏览器（如奇安信浏览器）。
指定开放端口	-	计算节点控制台系统的网络端口
部署配置相关参数		

参数名	样例	参数描述
部署方式	-	当前版本支持云租户部署和边缘节点部署。 • **云租户部署**：数据上云的用户可以选择“云租户部署”，可信计算节点部署在华为云租户的虚拟私有云VPC中，可信计算节点组件部署在基于华为云CCE服务的容器中。关于CCE集群的更多信息可参考**CCE**。 当前仅支持直接创建CCE集群，不支持选择已有集群。您需要配置CCE集群的部署规格、虚拟私有云、子网、节点密码、弹性IP等信息。 **说明** ◦ CCE集群的部署规格根据您的业务量自行选择。 ◦ 所创建CCE集群的虚拟私有云、子网，应与数据源所在云服务（如MRS Hive、DWS等）的虚拟私有云、子网保持一致，以确保网络互通。 ◦ 自动创建的CCE集群费用不需要单独结算，当前TICS费用已包含CCE集群费用。 • **边缘节点部署**：数据不上云的用户可以选择“边缘节点部署”，数据不需要上传到华为云上，通过纳管节点的方式，即可参与多方安全计算或者联邦学习任务，关于IEF边缘节点的更多信息可参考**IEF**。 您可参考**纳管节点**来纳管节点，注意：纳管节点防火墙需要开通30000-65535端口，且需要建立消息端点和消息路由，步骤如下： 1. 登录IEF服务，选择左侧“边云消息”列，选择“消息端点”。 2. 创建消息端点，填写相关参数。 “消息端点类型”选择“边缘端点（ServiceBus）”； “消息端点名称”参数值为“tics-agent”； “服务端口”参数值为“30000”。 3. 选择左侧“边云消息”列，单击“消息路由”，勾选“专业版服务实例”，填写相关参数。 **“消息路由名称”**参数值为“tics-agent-route”； **“源端点”**参数值为“SystemREST”； **“源端点资源”**参数值为“/tics-agent”； **“目的端点”**参数值为“tics-agent”； **“目的端点资源”**参数值为“/”。
云租户部署参数		

参数名	样例	参数描述
部署规格	中规格	• 中规格：适用百万级别数据多方安全计算，五十万内对齐样本联邦建模 • 大规格：适用千万级别数据多方安全计算，百万级别对齐样本联邦建模
虚拟私有云	-	选择合适的VPC
子网	-	选择合适的子网地址
NAT网关	-	选择子网下NAT网关，若子网下不存在NAT网关，默认新建。
弹性IP	-	选择NAT网关已关联的弹性公网IP。若NAT网关无关联弹性公网IP，默认新建。 弹性公网IP提供外网访问能力，可以灵活绑定及解绑，随时修改带宽。未绑定弹性公网IP的云服务器无法直接访问外网，无法直接对外进行互相通信。
存储方式	-	提供OBS存储和极速文件存储两种持久化存储卷的选择。
OBS存储	-	存储方式选择obs存储时，可以选择自动创建OBS桶，也可以通过下拉框的搜索功能寻找已有的OBS桶。选择已有的OBS桶时，需要确认OBS桶的访问权限中包含读取权限和写入权限，否则其上的联邦作业将会失败。
卷名	-	存储方式选择极速文件存储时，默认选取已有的极速文件存储，也可手动填写SFS ID。
挂载路径	-	存储方式选择极速文件存储时需填写。默认根路径，若自定义路径，请确保该路径在极速文件存储上存在。
开启AOM日志监控	-	开启后可收集可信计算节点日志，推荐开启。 对接AOM之后，相应的日志存储在AOM平台上，平台每月提供500M的免费空间，超出则计费。具体的计费规则参见**计费概述**。
节点密码	-	设置可信计算节点宿主机的登录密码。
确认密码	-	与“节点密码”保持一致即可。
边缘节点部署参数		
AI加速卡	-	• 不启用：部署常规的CPU规格计算节点 • 启用：启用边缘节点的AI加速卡，可以大幅减少联邦建模的耗时。通过IEF边缘节点部署时，请确保计算节点的AI加速卡相关功能可用，如需帮助请联系客服或技术支持人员。

参数名	样例	参数描述
纳管节点	-	用户选择边缘节点部署计算节点时呈现此参数。用户通过IEF服务纳管用户侧的边缘节点，用于部署计算节点。使用边缘节点部署方式，请先参考**纳管节点**执行纳管节点操作。
主机docker IP	-	请前往ief纳管节点，执行命令ifconfig docker0 | grep inet | grep -v 127.0.0.1 | grep -v inet6 | awk '{print $2}' | tr -d "addr:" 填入所得的ip地址
proxy配置（选填）	-	用户选择IEF部署计算节点时，可根据实际情况选填该参数。如果纳管节点使用了网络计算节点，请按照实际情况配置proxy信息，也可在部署成功后，通过配置变更项进行修改，具体操作可参考**变更计算节点配置**。
存储方式	-	选择采用外部文件挂载容器目录的方式。IEF当前仅支持“主机存储”。 • **主机存储**：该方式将计算节点所在的集群节点的主机路径，挂载到计算节点容器的目录上。用户需要选择集群中的节点（对应“纳管节点”下拉选）作为挂载节点，此时，部署的计算节点容器会运行到该节点上。同时，用户需要输入“主机路径”，设置该节点的主机挂载目录。计算节点成功部署后，用户可登录集群该节点，访问输入的“主机路径”来进行文件的上传。
主机路径	-	“存储方式”选择“主机存储”时呈现此参数，计算节点成功部署后通过输入的“主机路径”来进行文件的上传。 例如：“192.168.0.61/tmp”，如何在后台查找该路径请参考**登录节点**的相关描述。 **说明** 请确保选择的主机路径具有1000:1000属组权限，否则会影响部分功能使用。
资源分配策略		
CPU(Cores)	-	用户可根据返回资源剩余规格，按照分析与学习需求，灵活分配核数。
内存(GiB)	-	用户可根据返回资源剩余规格，按照分析与学习需求，灵活分配内存。容器预留部分管理资源，作业可用内存最大值设置为内存数值的0.6倍，且向下取整。

3. 确认无误后单击下一步并提交订单。
4. 付款成功，显示空间部署和可信计算节点部署，二者部署成功后即可在首页看到已创建的空间。

2.8 部署计算节点

同一个空间中的用户，在使用可信计算服务时（多方安全计算和可信联邦学习），需要部署计算节点，将数据上传，作为可信计算服务的输入，通过执行多方安全计算和可信联邦学习作业后，最终拿到结果。

计算节点以容器的形式部署，支持云容器引擎（CCE，Cloud Container Engine）服务和智能边缘平台（IEF，Intelligent EdgeFabric）服务部署，用户可根据数据上云的实际需求，采用合适的计算节点部署方案。

- 云容器引擎（CCE，Cloud Container Engine）提供高可靠高性能的企业级容器应用管理服务，支持Kubernetes社区原生应用和工具，简化云上自动化容器运行环境搭建。
- 智能边缘平台（Intelligent EdgeFabric）通过纳管您的边缘节点，提供将云上应用延伸到边缘的能力，联动边缘和云端的数据，满足客户对边缘计算资源的远程管控、数据处理、分析决策、智能化的诉求。同时，在云端提供统一的设备/应用监控、日志采集等运维能力，为企业提供完整的边缘和云协同的一体化服务的边缘计算解决方案。

前提条件

1. 本地存在下载好的空间信息和证书文件，下载方式参考**下载计算节点配置信息**。
2. 若需将执行过程记录审计至区块链，请确保当前加入的空间已开启区块链审计服务，同时完成**启用区块链审计服务（可选）**中对应角色（发起方/参与方）的准备工作，保证当前各参与方均处于区块空间链中。
3. 根据实际情况选择部署方式，参考**计算节点部署方式**，并执行相关操作。

约束限制

- IEF边缘节点部署计算节点：

  纳管节点只负责运行TICS的计算节点服务；每个纳管节点，只能运行一个计算节点。

  IEF边缘节点上部署的计算节点不支持创建DWS类型的连接器。

  IEF边缘节点服务器上的dokcer版本需要大于或等于20.10.10。

计算节点部署方式

云租户部署：

数据上云的用户可以选择“云租户部署”。可信计算节点部署在华为云租户的虚拟私有云VPC中，可信计算节点组件部署在基于华为云CCE服务的容器中。

当前仅支持直接创建CCE集群，不支持选择已有集群。您需要配置CCE集群的部署规格、虚拟私有云、子网、节点密码、弹性IP等信息。

关于CCE集群的更多信息可参考**CCE**。

选择边缘节点部署计算节点：

数据不上云的用户可以选择“边缘节点部署”。数据不需要上传到云上，通过纳管节点的方式，即可参与多方安全计算或者联邦学习任务，关于IEF边缘节点的更多信息可参考**IEF**。

您可参考**纳管节点**来纳管节点，注意：纳管节点防火墙需要开通30000-65535端口，且需要建立消息端点和消息路由，步骤如下：

1. 登录IEF服务，选择左侧“边云消息”列，选择“消息端点”。
2. 创建消息端点，填写相关参数。
   “消息端点类型”选择“边缘端点（ServiceBus）”；
   “消息端点名称”参数值为“tics-agent”；
   “服务端口”参数值为“30000”。
3. 选择左侧“边云消息”列，单击“消息路由”，勾选“专业版服务实例”，填写相关参数。
   “消息路由名称”参数值为“tics-agent-route”；
   “源端点”参数值为“SystemREST”；
   “源端点资源”参数值为“/tics-agent”；
   “目的端点”参数值为“tics-agent”；
   “目的端点资源”参数值为“/”。

部署计算节点

步骤1 用户登录TICS控制台。

步骤2 进入TICS控制台后，单击页面左侧“通知管理”，进入通知管理页面。

步骤3 浏览通知信息，在对应空间通知处单击“前往购买计算节点”，在弹出的页面配置参数。

图 2-5 部署计算节点

表 2-5 参数配置说明

参数名	参数描述
计算节点位置相关参数	
区域	下拉选择用户将计算节点部署在哪个区域。
项目	下拉选择用户将计算节点部署在区域下的哪一个项目内。
计费方式	选择包年/包月。
购买时长	支持按月或按年购买。
自动续费	支持自动续费。 ● 按月购买时，自动续费周期为1个月。 ● 按年购买时，自动续费周期为1年。
版本类型	当前可选版本只包含企业版。
空间配置相关参数	

参数名	参数描述
导入空间配置（可选）	用户从“前往购买计算节点”进入部署页面则无需该操作。 其它情况下需在TICS“通知管理”页面，单击“下载计算节点配置”，得到agentConfig.zip文件，本地解压后，导入json文件，空间配置信息将会自动填充到“区域”（league_region_name）、“空间名称”（league_name）、“空间ID”（league_id）。
空间区域	导入配置文件会自动填充，若未导入下拉选择空间所在的区域即可。可通过在TICS“通知管理”页面，单击“下载计算节点配置”，得到agentConfig.zip文件，本地解压后，打开json文件，查看参数“league_region_name”。
空间名称	通过“计算节点配置”文件agentConfig.zip中的json文件获取，参数名为“league_name”。
空间ID	通过“计算节点配置”文件agentConfig.zip中的json文件获取，参数名为“league_id”。
计算节点配置相关参数	
计算节点名称	计算节点别名，由用户自定义，用以区分部署的各个计算节点。要求：名称不能以空白字符开头或结尾，也不能包含下列特殊字符：\ / : * ? " < > |，长度要求在1~128之间。
访问密钥ID（AK）	用户的身份标识，需要用户去IAM服务自行下载。文件获取方式请参考**参考：获取访问密钥**章节。 **说明** • 如果访问密钥泄露，会带来数据泄露风险。 • 每个访问密钥只能下载一次，为了账号安全性，建议定期更换访问密钥并妥善保存。
加密密钥（SK）	
计算节点登录名称	登录计算节点控制台的用户名。用户可通过“计算节点登录名称”和“登录密码”进入计算节点控制台，建立连接器，发布数据。
登录密码	登录计算节点控制台的密码。
确认密码	与“登录密码”保持一致即可。
支持国密	若选择是，则登录计算节点必须使用国密浏览器（如奇安信浏览器）。
指定开放端口	计算节点控制台系统的网络端口。
部署配置相关参数	

参数名	参数描述
部署方式	当前版本支持云租户部署和边缘节点部署。 • **云租户部署**：数据上云的用户可以选择“云租户部署”，可信计算节点部署在华为云租户的虚拟私有云VPC中，可信计算节点组件部署在基于华为云CCE服务的容器中。关于CCE集群的更多信息可参考**CCE**。 当前仅支持直接创建CCE集群，不支持选择已有集群。您需要配置CCE集群的部署规格、虚拟私有云、子网、节点密码、弹性IP等信息。 **说明** – CCE集群的部署规格根据您的业务量自行选择。 – 所创建CCE集群的虚拟私有云、子网，应与数据源所在云服务（如MRS Hive、DWS等）的虚拟私有云、子网保持一致，以确保网络互通。 – 自动创建的CCE集群费用不需要单独结算，当前TICS费用已包含CCE集群费用。 • **边缘节点部署**：数据不上云的用户可以选择“边缘节点部署”，数据不需要上传到华为云上，通过纳管节点的方式，即可参与多方安全计算或者联邦学习任务，关于IEF边缘节点的更多信息可参考**IEF**。 您可参考**纳管节点**来纳管节点，注意：纳管节点防火墙需要开通30000-65535端口，且需要建立消息端点和消息路由，步骤如下： 1. 登录IEF服务，选择左侧“边云消息”列，选择“消息端点”。 2. 创建消息端点，填写相关参数。 “消息端点类型”选择“边缘端点（ServiceBus）”； “消息端点名称”参数值为“tics-agent”； “服务端口”参数值为“30000”。 3. 选择左侧“边云消息”列，单击“消息路由”，勾选“专业版服务实例”，填写相关参数。 **“消息路由名称”**参数值为“tics-agent-route”； **“源端点”**参数值为“SystemREST”； **“源端点资源”**参数值为“/tics-agent”； **“目的端点”**参数值为“tics-agent”； **“目的端点资源”**参数值为“/”。
云租户部署参数	
部署规格	• 中规格：适用百万级别数据多方安全计算，五十万内对齐样本联邦建模 • 大规格：适用千万级别数据多方安全计算，百万级别对齐样本联邦建模
虚拟私有云	选择合适的VPC
子网	选择合适的子网地址
NAT网关	选择子网下NAT网关，若子网下不存在NAT网关，默认新建。

参数名	参数描述
弹性IP	选择NAT网关已关联的弹性公网IP。若NAT网关无关联弹性公网IP，默认新建。 弹性公网IP提供外网访问能力，可以灵活绑定及解绑，随时修改带宽。未绑定弹性公网IP的云服务器无法直接访问外网，无法直接对外进行互相通信。
存储方式	提供OBS存储和极速文件存储两种持久化存储卷的选择。
OBS存储	存储方式选择obs存储时，可以选择自动创建OBS桶，也可以通过下拉框的搜索功能寻找已有的OBS桶。选择已有的OBS桶时，需要确认OBS桶的访问权限中包含读取权限和写入权限，否则其上的联邦作业将会失败。
卷名	存储方式选择极速文件存储时，默认选取已有的极速文件存储，也可手动填写SFS ID。
挂载路径	存储方式选择极速文件存储时需填写。默认根路径，若自定义路径，请确保该路径在极速文件存储上存在。
开启AOM日志监控	开启后可收集可信计算节点日志，推荐开启。 对接AOM之后，相应的日志存储在AOM平台上，平台每月提供500M的免费空间，超出则计费。具体的计费规则参见**计费概述**。
节点密码	设置可信计算节点宿主机的登录密码。
确认密码	与“节点密码”保持一致即可。
边缘节点部署参数	
AI加速卡	• 不启用：部署常规的CPU规格计算节点 • 启用：启用边缘节点的AI加速卡，可以大幅减少联邦建模的耗时。通过IEF边缘节点部署时，请确保计算节点的AI加速卡相关功能可用，如需帮助请联系客服或技术支持人员。
纳管节点	用户选择边缘节点部署计算节点时呈现此参数。用户通过IEF服务纳管用户侧的边缘节点，用于部署计算节点。使用边缘节点部署方式，请先参考**纳管节点**执行纳管节点操作。
主机docker IP	请前往ief纳管节点，执行命令ifconfig docker0 | grep inet | grep -v 127.0.0.1 | grep -v inet6 | awk '{print $2}' | tr -d "addr:" 填入所得的ip地址
proxy配置（选填）	用户选择IEF部署计算节点时，可根据实际情况选填该参数。如果纳管节点使用了网络计算节点，请按照实际情况配置proxy信息，也可在部署成功后，通过配置变更项进行修改，具体操作可参考**变更计算节点配置**。

参数名	参数描述
存储方式	选择采用外部文件挂载容器目录的方式。IEF当前仅支持“主机存储”。 • **主机存储**：该方式将计算节点所在的集群节点的主机路径，挂载到计算节点容器的目录上。用户需要选择集群中的节点（对应“纳管节点”下拉选）作为挂载节点，此时，部署的计算节点容器会运行到该节点上。同时，用户需要输入“主机路径”，设置该节点的主机挂载目录。计算节点成功部署后，用户可登录集群该节点，访问输入的“主机路径”来进行文件的上传。
主机路径	“存储方式”选择“主机存储”时呈现此参数，计算节点成功部署后通过输入的“主机路径”来进行文件的上传。 例如：“192.168.0.61/tmp”，如何在后台查找该路径请参考**登录节点**的相关描述。 **说明** 请确保选择的主机路径具有1000:1000属组权限，否则会影响部分功能使用。
资源分配策略	
CPU(Cores)	用户可根据返回资源剩余规格，按照分析与学习需求，灵活分配核数。
内存(GiB)	用户可根据返回资源剩余规格，按照分析与学习需求，灵活分配内存。容器预留部分管理资源，作业可用内存最大值设置为内存数值的0.6倍，且向下取整。
区块链配置	
是否开启区块链审计	勾选该项表示启用区块链审计服务，使用前需要按照“准备工作 > 启用区块链审计服务（可选）”章节的描述完成准备工作。
BCS服务实例	选择BCS空间链。
通道	选择邀空间链邀请租户时选择的通道。
组织	选择链代码部署的组织。

步骤4 单击下一步并提交订单，完成计算节点部署。

说明

- 计算节点在不同时刻有以下7种状态：部署中，部署失败，启动中，运行中，删除中，删除失败，重启中。
- 可以在“？”标识处，查看部署计算节点的概要事件信息。
- 计算节点在部署完成后会向外访问如下地址，发送节点状态信息，用作心跳监测以及执行联邦作业操作命令。

  1.tics.****.myhuaweicloud.com（地址信息以空间所在region为准）。

  2.聚合器ip（空间创建时自动申请的聚合器公网ip）。

步骤5 给CCE类型计算节点的最终租户增加CCE命名空间运维权限。

图 2-6 添加运维权限-入口

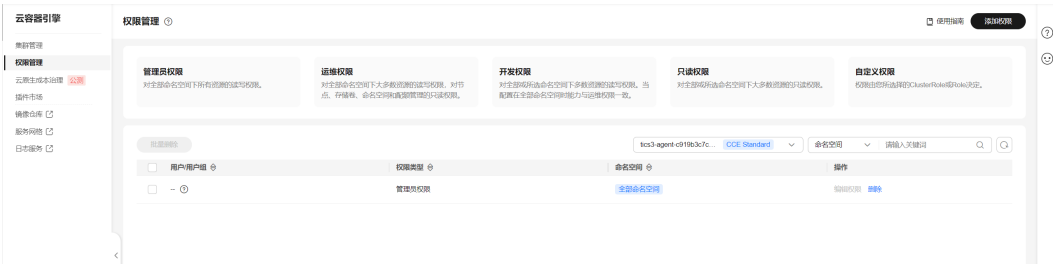


图 2-7 添加运维权限-类型

----结束

2.9 创建连接器

连接器用来快速连接到用户名下的各类资源服务。

前提条件

- 计算节点处于运行中，且所在空间信息的“认证状态”为“已认证”。
- 建议使用者提前了解MapReduce服务（MRS Hive）集群。
- “连接器类型”选择MapReduce服务（MRS Hive）时，选择的MRS集群需与当前计算节点部署CCE在同一VPC。填写的用户名，需具有Hive的读写权限。“集群名称”为用户所需要使用的MRS Hive数据源所在的MRS集群。“用户名”为MRS集群中拥有Hive权限的集群用户。

注意事项

- IEF上部署的计算节点不支持创建MRS Hive、ModelArts和DWS类型的连接器。
- MRS Hive、MySQL、DWS、RDS、ORACLE连接器当前只支持在多方安全计算作业中使用。
- API连接器当前只支持在实时预测作业和实时隐匿查询中使用。

创建连接器

步骤1 用户登录TICS控制台。

步骤2 进入TICS控制台后，单击页面左侧“计算节点管理”。

步骤3 在“计算节点管理”页面，查找需要发布数据的计算节点名称，单击“计算节点名称”进入计算节点详情页。

图 2-8 选择计算节点

步骤4 在“计算节点详情”页，单击“前往计算节点”，在登录页正确输入部署计算节点时设置的“登录用户名”和“密码”。

图 2-9 前往计算节点

步骤5 登录成功后，进入到计算节点界面，选择左侧导航栏中“连接器管理”，单击“创建”，在弹出的界面配置创建连接器的参数，配置完成后单击“确定”。

说明

测试功能为数据源连通性及密码正确性的检查测试。

图 2-10 创建连接器（以 RDS 服务为例）

表 2-6 参数说明

参数名	描述
连接器类型	• “连接器类型”选择Hive连接时，需要选择Hive版本，当前仅支持MRS2.x和MRS3.x版本，选择的MRS集群需与当前计算节点部署CCE或IEF（非云上IEF节点不支持接入Hive）在同一VPC。“用户名”为MRS集群中拥有Hive权限的集群用户，“用户认证凭据”需要上传对应用户的认证凭据，请在MapReduce服务的**下载用户认证文件**中获取。 • “连接器类型”选择RDS服务时，所选择的RDS服务实例需与计算节点在同一VPC下，且端口开放。填写的用户名，需具有数据库的读写权限（参考**修改权限**）。“密码”为该用户登录RDS实例的密码。 • “连接器类型”选择MySql时，需保证计算节点与数据库所在虚机的连通性，“驱动文件”需与目标MySQL数据库版本一致。驱动类名com.mysql.cj.jdbc.Driver，仅支持mysql-connector-java-5.x以后的版本，驱动文件请在**Mysql驱动下载地址**中获取。 • “连接器类型” 选择DWS连接时，填写的用户名，需具有数据库的读写权限（参考**权限管理**）。“密码”为该用户登录DWS实例的密码。 • “连接器类型” 选择ORACLE连接时，需保证计算节点与数据库的连通性，当前仅支持ORACLE 12c和19c版本。驱动文件需与目标ORACLE数据库版本一致，请在**ORACLE驱动下载地址**中获取。 • “连接器类型” 选择API连接时，需保证计算节点与api接口的连通性，当前仅支持基础认证方式。
连接器名称	根据实际情况设置即可。

参数名	描述
数据库版本	“连接器类型”选择MySql和ORACLE时，呈现此参数。根据实际情况设置即可。
数据库名称	“连接器类型”选择ORACLE时，呈现此参数。根据实际情况设置即可。
数据库服务器	“连接器类型”选择ORACLE时，呈现此参数。用户根据实际情况设置。
端口	“连接器类型”选择ORACLE时，呈现此参数。用户根据实际情况设置。
实例名称	“连接器类型”选择RDS或DWS服务时，呈现此参数。下拉选择实例即可。
数据库	“连接器类型”选择DWS服务时，呈现此参数。可手动输入DWS服务里面购买的数据库名称。
用户名	用户根据实际情况设置。
密码	用户根据实际情况设置。
驱动类名	“连接器类型”选择MySql和ORACLE时，呈现此参数。根据实际情况设置，注意驱动类名com.mysql.cj.jdbc.Driver仅支持mysql-connector-java-5.x以后的版本。
JDBC URL	“连接器类型”选择MySql时，呈现此参数。JDBC访问端口。取值样例：198.0.0.1：3306。
驱动文件	“连接器类型”选择MySql和ORACLE时，呈现此参数。JDBC驱动。
其他属性	“连接器类型”选择MySql时，呈现此参数。用户根据实际情况设置任务所需的Key和Value。

----结束

管理连接器

步骤1 在“计算节点详情”页，单击“前往计算节点”，在登录页正确输入部署计算节点时设置的“登录用户名”和“密码”。

图 2-11 前往计算节点

步骤2 登录成功后，单击左侧导航栏中“连接器管理”，在操作栏中通过单击编辑、删除，来进行连接器管理操作。

图 2-12 连接器管理

----结束

3 使用场景

3.1 多方安全计算场景

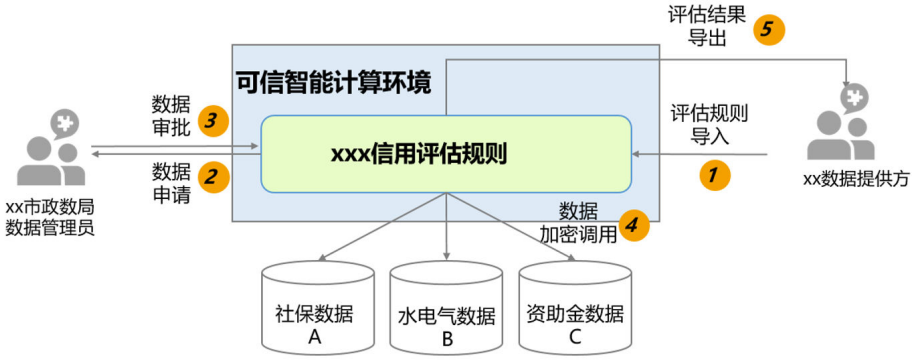
3.1.1 场景描述

本章节以“小微企业信用评分”场景为例。

背景信息

社保、水电气和资助金等数据统一存储在某政务云，由不同的局进行管理，机构想单独申请进行企业相关评分的计算会非常困难。 因此可以由市政数局出面，统一制定隐私规则，审批数据提供方的数据使用申请， 并通过华为TICS可信智能计算平台进行安全计算。

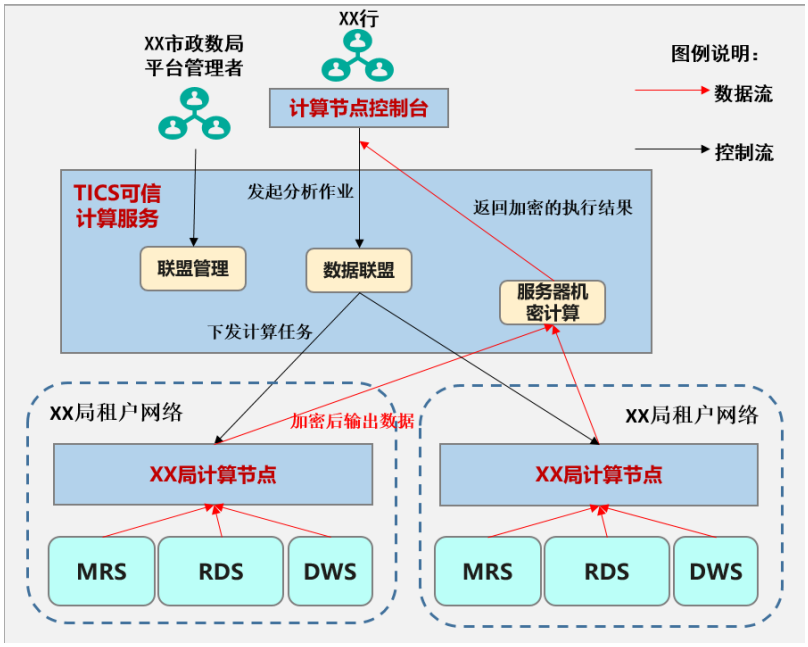
图 3-1 企业信用评估应用场景示意图



3.1.2 组合架构

架构说明

图 3-2 架构图例



1. 作业发起方通过计算节点提供的控制台页面，发起多方安全计算作业。
2. 多方安全计算作业在TICS中进行解析和任务计划构建，并下发任务给各个数据参与方所在的计算节点。
3. 参与方计算节点从租户侧网络内的数据中获取数据，并使用安全算法进行加密输出。
4. 数据在TICS提供的服务器中进行机密计算。
5. 最终将计算完成的结果加密返回给作业发起方。
6. 空间的整体配置通过空间管理员进行统一管理。

3.1.3 可验证代码示例

3.1.3.1 数据准备

数据准备

⚠ 注意

- 以下数据和表结构是根据场景进行模拟的数据，并非真实数据。
- 以下数据需要提前存导入到MySQL\Hive\Oracle等用户所属数据源中，TICS本身不会持有这些数据，这些数据会通过用户购买的计算节点进行加密计算，保障数据安全。
- 政府信息提供方的数据tax和support，在用户计算节点agent_gov上发布。
- 能源信息提供方的数据power，在用户计算节点agent_power上发布。

表 3-1 企业税收和资助金情况表 tax

列名	含义	字段分类
Id	企业id	唯一标识
tax_bal	税收	敏感
Industry	行业类型	不敏感

表 3-2 企业政府资助金数据表 support

列名	含义	字段分类
Id	企业id	唯一标识
supp_bal	资助金的金额	敏感
Industry	行业类型	不敏感

表 3-3 企业水电情况表 power

列名	含义	字段分类
Id	企业id	唯一标识
electric_bal	电费	敏感
water_bal	水费	敏感

从业务角度考虑，安排五个阶段，来对TICS系统进行验证和测试。本章重点讲述如何端到端实现一个该场景下的隐私计算作业完整执行流程。

导入数据

步骤1 在第一个合作方Partner1的MySQL数据源中，通过如下的SQL语句创建数据表：

```
CREATE TABLE tax (
        id integer COMMENT '企业id' ,
        tax_bal integer COMMENT '税收金额' ,
        industry varchar(150) COMMENT '行业'
);
CREATE TABLE support (
        id integer COMMENT '企业id' ,
        supp_bal integer COMMENT '资助金额',
        industry varchar(150) COMMENT '行业'
);
```

步骤2 在第二个合作方Partner2的MySQL数据源中，通过如下的SQL语句创建数据表：

```
CREATE TABLE power (
        id integer COMMENT '企业id' ,
        electric_bal integer COMMENT '电费',
        water_bal integer COMMENT '水费'
);
```

步骤3 将下面的数据分别导入csv文件并上传到MySQL数据源所在服务器。

- Tax表的数据如下：

  ```
  id,tax_bal,industry
  123400999,745,互联网
  123400998,324,其他
  123400997,664,其他
  123400996,243,金融
  123400995,715,互联网
  123400994,475,通讯
  123400993,526,其他
  123400992,272,互联网
  123400991,646,金融
  123400990,510,其他
  ```

- Support表的数据如下：

  ```
  id,supp_bal,industry
  123400999,314,互联网
  123400998,405,其他
  123400997,371,其他
  123400996,484,金融
  123400995,381,互联网
  123400994,405,通讯
  123400993,292,其他
  123400992,503,互联网
  123400991,303,金融
  123400990,412,其他
  ```

- Power表的数据如下：

  ```
  id,electric_bal,water_bal
  123400999,79,48
  123400998,57,70
  123400997,69,37
  123400996,50,57
  123400995,66,50
  123400994,56,55
  123400993,63,53
  123400992,45,76
  123400991,80,36
  123400990,39,63
  ```

步骤4 执行如下SQL语句，将csv文件内的数据导入创建的数据表。

```
LOAD DATA INFILE 'csv数据文件名' INTO TABLE 表名
```

或者执行如下的插入语句：

- Tax表：


```
insert into tax values
(123400999,745,'互联网'),
(123400998,324,'其他' ),
(123400997,664,'其他' ),
(123400996,243,'金融' ),
(123400995,715,'互联网' ),
(123400994,475,'通讯' ),
(123400993,526,'其他'),
(123400992,272,'互联网' ),
(123400991,646,'金融' ),
(123400990,510,'其他');
```

- Support表：

```
insert into support values
(123400999,314,'互联网' ),
(123400998,405,'其他' ),
(123400997,371,'其他' ),
(123400996,484,'金融' ),
(123400995,381,'互联网' ),
(123400994,405,'通讯' ),
(123400993,292,'其他' ),
(123400992,503,'互联网' ),
(123400991,303,'金融' ),
(123400990,412,'其他');
```

- Power表：

```
insert into power values
(123400999,79,48),
(123400998,57,70),
(123400997,69,37),
(123400996,50,57),
(123400995,66,50 ),
(123400994,56,55),
(123400993,63,53),
(123400992,45,76),
(123400991,80,36),
(123400990,39,63);
```

----结束

3.1.3.2 数据集发布

前提条件

完成**数据准备**工作。

操作步骤

步骤1 进入TICS服务控制台。

步骤2 在计算节点管理中，找到购买的计算节点，通过登录地址，进入计算节点控制台。

图 3-3 前往计算节点

步骤3 登录计算节点后，在下图所述位置新建连接器。

图 3-4 新建连接器

步骤4 输入正确的连接信息，建立数据源和计算节点之间的安全连接。

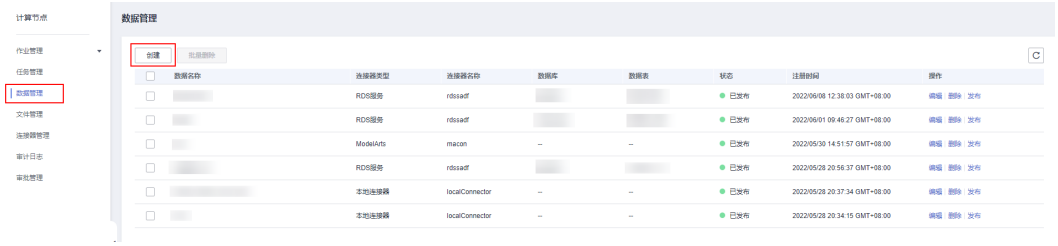
图 3-5 输入信息

步骤5 建立完成后，连接器显示正常说明连接正常。

图 3-6 连接正常

步骤6 进入数据管理，进行数据集发布。

图 3-7 新建数据管理



步骤7 填写参数信息。

图 3-8 填写参数

----结束

重复步骤1~7，发布tax税务表和power_data能源表。

说明

数据发布的过程并不会直接从数据源中导出用户数据，仅从数据源处获取了数据集相关的元数据信息，用于任务的解析、验证等。

3.1.3.3 隐私规则防护

使用TICS的隐私规则防护能力确保数据安全。

前提条件

完成**数据集发布**。

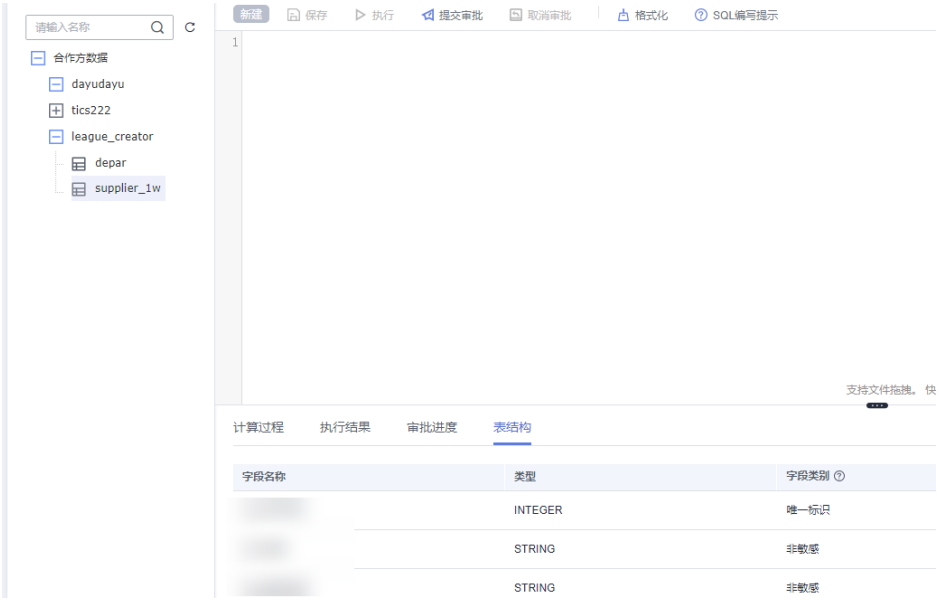
操作步骤

步骤1 进入多方安全计算的作业执行界面，单击创建。

图 3-9 创建作业

步骤2 在作业界面中，按照**示例一**和**示例二**提供的案例和SQL语句进行作业测试。

图 3-10 作业界面



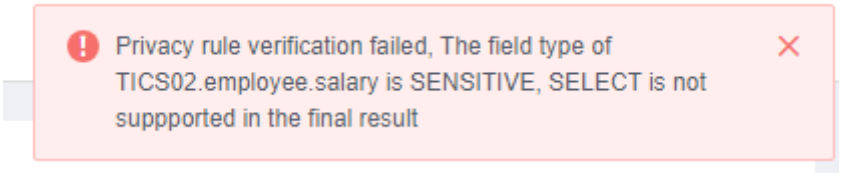
- 示例一：

  假设有人输入以下代码试图直接查询敏感数据。

```
select
  tax_bal, id
from
  league_creator.tax
```

  系统提示不支持进行敏感数据的SELECT操作。

图 3-11 不支持敏感操作



- 示例二：

  若试图在敏感数据中追加自己的数据， 从结果倒推敏感数据，即求原数据。

```
Select
  tax_bal + electric_bal
from
  LEAGUE_CREATOR.tax a
  join ZZZZZZ.power_data b on a.id = b.id
```

  TICS会识别并提示。

图 3-12 执行失败告警

说明

上述隐私规则，均为TICS系统提供的默认规则。

----结束

3.1.3.4 基本计算能力验证

验证TICS的基础计算能力，以计算各企业在2021年的价值评分，用于评估信贷能力，其中的公式仅为简单的参考计算式。

操作步骤

步骤1 执行如下的sql作业。

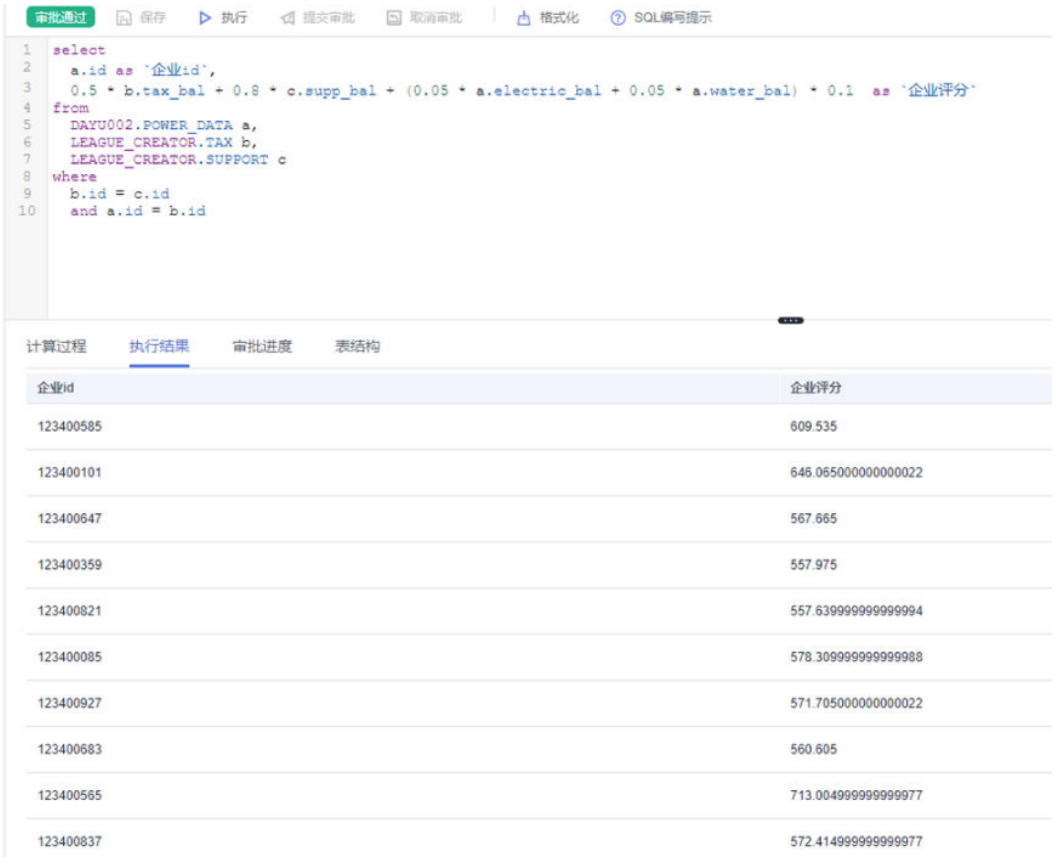
```
select
  c.id as `企业id`,
  0.5 * a.tax_bal + 0.8 * b.supp_bal + (0.05 * c.electric_bal + 0.05 * c.water_bal) * 0.1  as `企业评分`
from
  Partner1.TAX a,
  Partner1.SUPPORT b,
  Partner2.POWER_DATA c
where
  b.id = c.id
  and a.id = b.id
```

步骤2 审批时可以看到如下的信息，涉及关联字段较多，其使用方式都能够在审批界面中展示出来。

图 3-13 基础信息

步骤3 执行结果如下。

图 3-14 执行结果



步骤4 结果显示，TICS支持大量基础的SQL语法。

图 3-15 SQL 编写提醒

----结束

3.1.3.5 基于 MPC 算法的高安全级别计算

完成demo验证阶段，为提升数据保护级别，接入以**纯密文的状态做计算**的更高安全级别的数据，可以通过开启高隐私级别开关，提升空间安全级别。

图 3-16 高隐私级别开关

再次单击作业，审批进行的同时敏感数据被进行了同态加密。DAG图显示了“psi + 同态”的全过程流向，基本符合业界已公开的PSI算法流程和同态加密流程。

图 3-17 加密流程

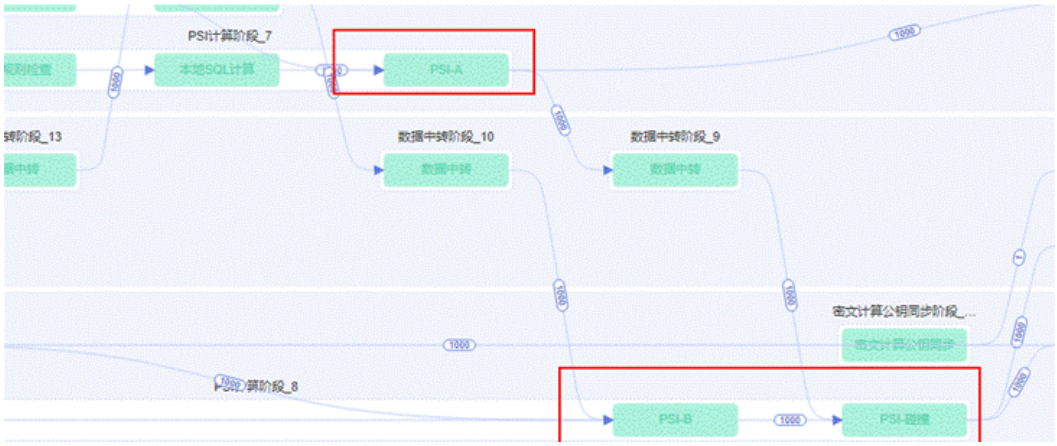
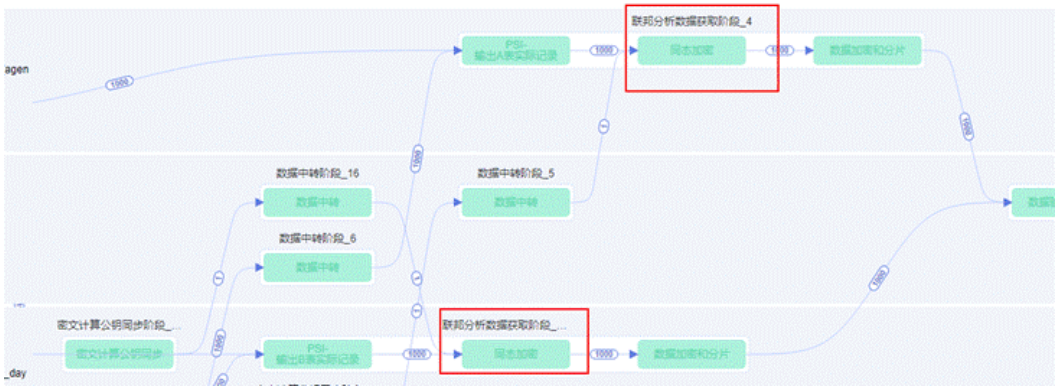


图 3-18 加密流程



3.1.3.6 统计型作业的差分隐私保护

本示例作业，以统计各行业的“企业税收总和”与“用电量总和”，进行统计分析：

```
Select
  industry,
  sum(tax_bal),
```



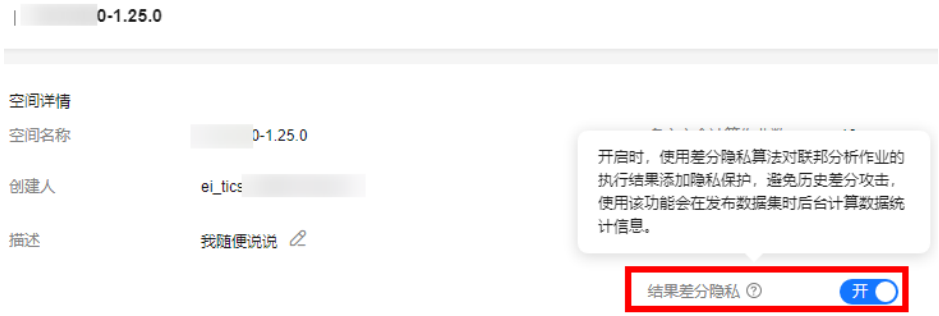
```
  sum(electric_bal)
from
  LEAGUE_CREATOR.tax a join
  dayu002.power_data b
  on a.id = b.id
group by
  industry
```

⚠ 注意

统计分析型的作业，可能被作业执行方通过增删某个碰撞的id，得到两次作业之间的差值，从而推算出实际taxpay和water_fee。

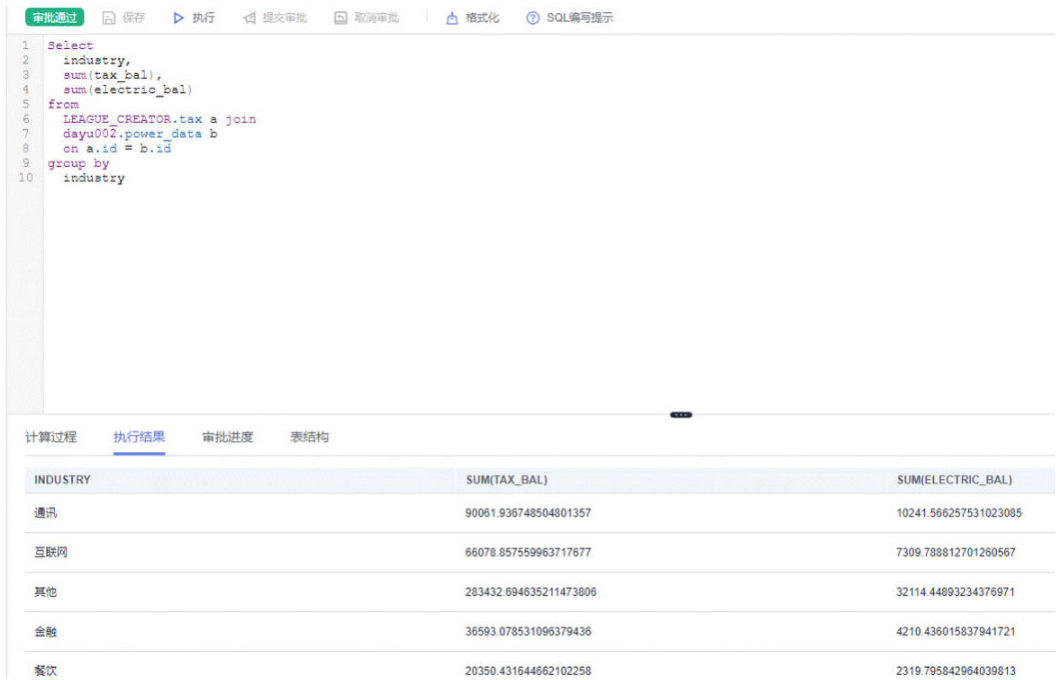
开启空间中的差分隐私开关保护敏感数据，符合差分隐私条件的统计作业，会自动应用差分隐私算法对计算结果进行加噪保护， 在一定误差范围内保证数据无法被恶意偷取。

图 3-19 差分隐私开关



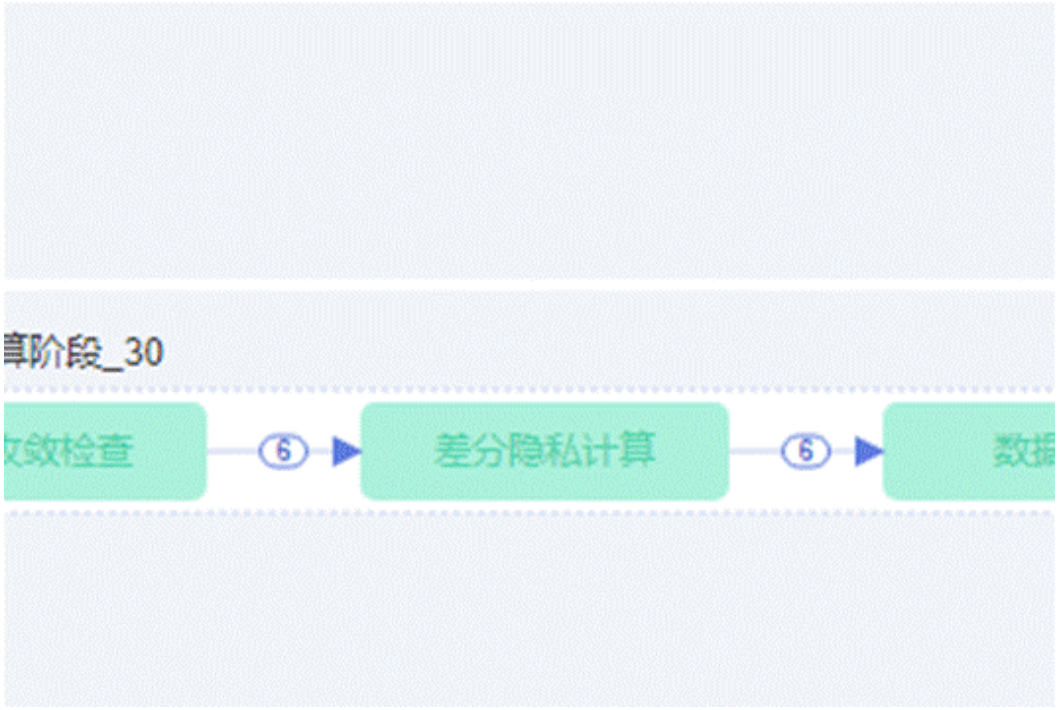
第一次执行作业的结果如下：

图 3-20 作业结果



在返回最终统计结果前，增加了一个差分隐私计算的任务节点，如图3所示。

图 3-21 差分隐私计算任务节点



再执行如下sql，sql中过滤掉了某个企业，试图用差值去计算这个企业的税收值。

```
Select
  industry,
  sum(tax_bal),
  sum(electric_bal)
from
  LEAGUE_CREATOR.tax a join
  dayu002.power_data b
  on a.id = b.id
  where a.id <> '123400558'
group by
  industry
```

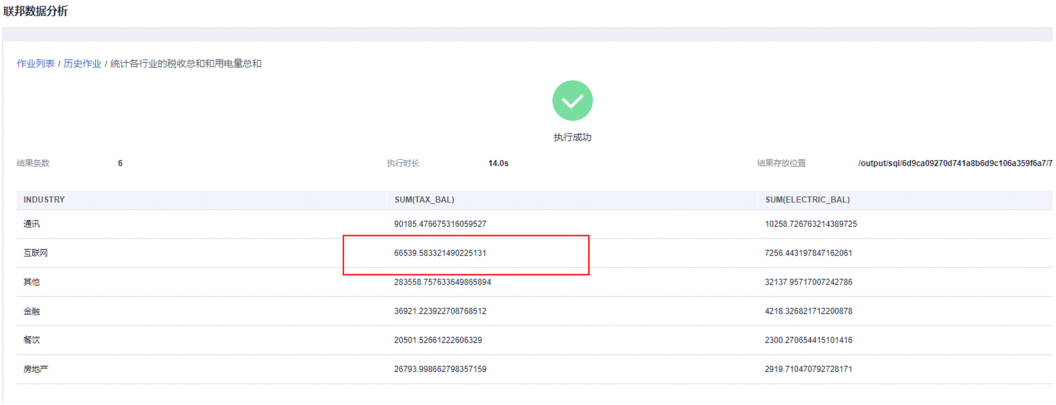
这个企业的实际tax为274：

图 3-22 tax

```
id,tax_bal,industry
123400558,274,互联网
```

得到新的结果如下：

图 3-23 新结果



经过计算，66539.583321490225131-66078.857559963717677=461，

通过差分隐私算法保护聚合操作的安全性，使开启算法保护的计算差值与预期得到的实际差值274不同，避免真实数据被窃取。

3.2 纵向联邦建模场景

3.2.1 使用 TICS 多方安全计算进行联合样本分布统计

3.2.1.1 场景描述

某企业A在进行新客户营销时的成本过高，想要通过引入外部数据的方式提高营销的效果，降低营销成本。

因此企业A希望与某大数据厂商B展开一项合作，基于双方共有的数据进行联邦建模，使用训练出的联邦模型对新数据进行联邦预测，筛选出高价值的潜在客户，再针对这些客户进行定向营销，达成提高营销效果、降低营销成本的业务诉求。

本文主要介绍在进行建模之前，双方需要对已有的数据进行碰撞求交，找到双方共有的数据，了解数据的分布情况并为后续的建模做好调整。

3.2.1.2 准备数据

首先，企业A和大数据厂商B需要商议确定要提供的数据范围及对应的元数据信息，例如双方初始决定使用最近三个月的已有用户转化数据作为联邦训练的训练集和评估集。

表 3-4 企业 A 的数据

字段名称	字段类型	描述
id	string	hash过后的手机号字符串
col0-col4	float	企业A数据特征
label	int	企业A对用户的标签属性

industry1.csv

```
id,col0,col1,col2,col3,col4,label
19581e27de7ced00ff1ce50b2047e7a567c76b1cbaebabe5ef03f7c3017bb5b7,-0.823913755,0.787712038,0.429635596,-1.315646486,-1.652321611,1
2c624232cdd221771294dfbb310aca000a0df6ac8b66b696d90ef06fdefb64a3,3.041881096,-0.651684341,3.661649955,0.035548734,3.477873904,0
3fdba35f04dc8c462986c992bcf875546257113072a909c162f7e470e581e278,-1.847252571,0.496981447,1.654416521,-1.945006902,0.394151993,1
4523540f1504cd17100c4835e85b7eefd49911580f8efff0599a8f283be6b9e3,-0.593556893,-0.351750558,0.964512256,-0.017390132,0.092562565,1
4a44dc15364204a80fe80e9039455cc1608281820fe2b24f1e5233ade6af1dd5,0.241505219,-0.219114719,1.51438745,-0.665234511,0.178575706,0
4b227777d4dd1fc61c6f884f48641d02b4d121d3fd328cb08b5531fcacdabf8a,0.372607556,-0.29194018,0.080862655,0.391501604,-0.012276428,1
4e07408562bedb8b60ce05c1decfe3ad16b72230967de01f640b7e4729b49fce,1.544282251,-0.203027285,3.076050022,-0.530666302,2.156693386,0
4ec9599fc203d176a301536c2e091a19bc852759b255bd6818810a42c5fed14a,1.006651366,-0.972403786,1.314115256,0.363296291,5.171128738,0
4fc82b26aecb47d2868c4efbe3581732a3e7cbcc6c2efb32062c08170a05eeb8,-2.859681221,-1.465959913,-0.930994729,-0.773533542,-3.673734138,0
5feceb66ffc86f38d952786c6d696c79c2dbc239dd4e91b46729d73a27fb57e9,-1.409250598,-0.589367921,-4.467396693,1.370376188,-1.2368325,1
```

大数据厂商B的数据如下，一共有10条记录。

表 3-5 大数据厂商 B 的数据

字段名称	字段类型	描述
id	string	hash过后的手机号字符串
f0-f4	float	大数据厂商数据特征

bigdata1.csv

```
id,f0,f1,f2,f3,f4
2c624232cdd221771294dfbb310aca000a0df6ac8b66b696d90ef06fdefb64a3,0.390064223,0.664175034,3.20228741,0.380574513,0.017733811
3fdba35f04dc8c462986c992bcf875546257113072a909c162f7e470e581e278,-0.483250226,0.616586578,3.001851708,2.407914633,0.856369412
4a44dc15364204a80fe80e9039455cc1608281820fe2b24f1e5233ade6af1dd5,-0.070919538,-2.219653517,1.461645551,1.66185096,0.778770954
4ec9599fc203d176a301536c2e091a19bc852759b255bd6818810a42c5fed14a,0.024227451,-1.087235302,3.67470964,-2.420729037,-3.132456573
4fc82b26aecb47d2868c4efbe3581732a3e7cbcc6c2efb32062c08170a05eeb8,-0.771151327,-1.184821181,-0.674077615,-0.379858223,0.158957184
6b51d431df5d7f141cbececcf79edf3dd861c3b4069f0b11661a3eefacbba918,-0.738091802,-1.474822882,2.93475295,-3.763763721,-1.817301398
6b86b273ff34fce19d6b804eff5a3f5747ada4eaa22f1d49c01e52ddb7875b4b,-1.216062821,-1.093614452,-1.632396806,0.887601314,-4.40930101
8527a891e224136950ff32ca212b45bc93f69fbb801c3b1ebedac52775f99e61,-0.789268594,1.071733834,3.763254446,-3.760298263,0.49776472
e7f6c011776e8db7cd330b54174fd76f7d0216b612387a5ffcfb81e6f0919683,-2.759963795,0.405262468,1.264947591,1.027350049,1.293868423
```

其中为了保证数据安全，企业A和大数据厂商B通过讨论决定使用hash过后的手机号作为已有数据的唯一标识id字段，并将唯一标识作为数据对齐的依据。

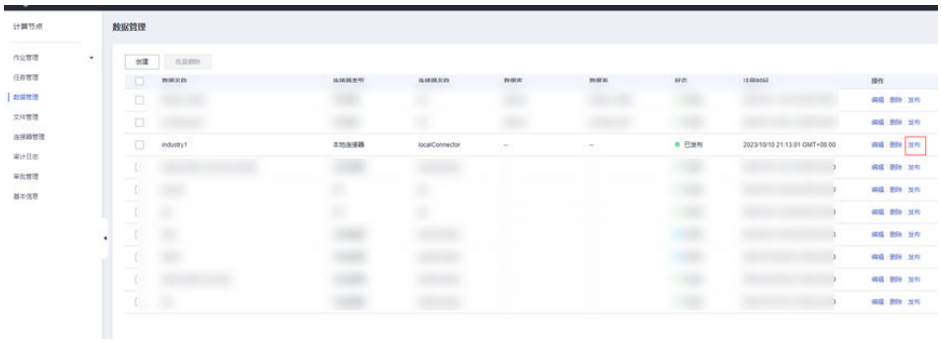
3.2.1.3 发布数据集

企业A和大数据厂商B分别将自己的csv数据文件上传到自己的计算节点上，通过“数据管理”模块创建各自的数据集。

企业A的数据集如下：

大数据厂商B的数据集如下：

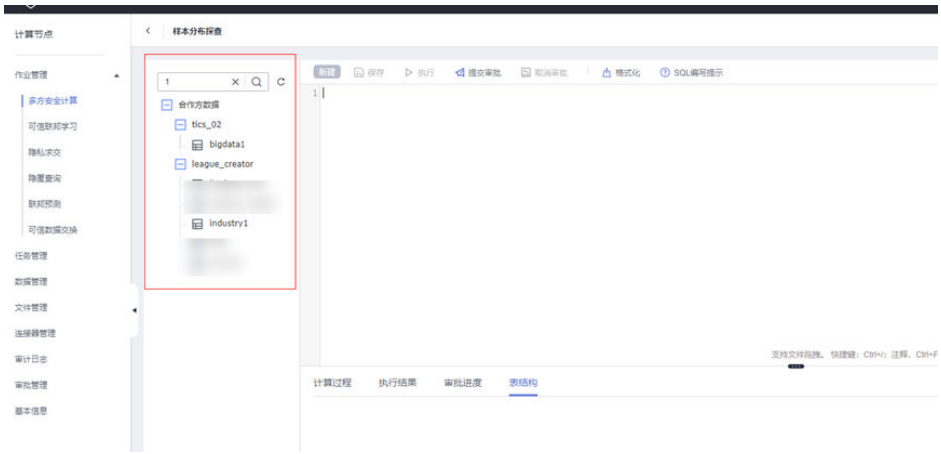
创建数据集后单击“发布”按钮即可将数据的元数据信息发布到tics空间侧，供其他合作方参考。



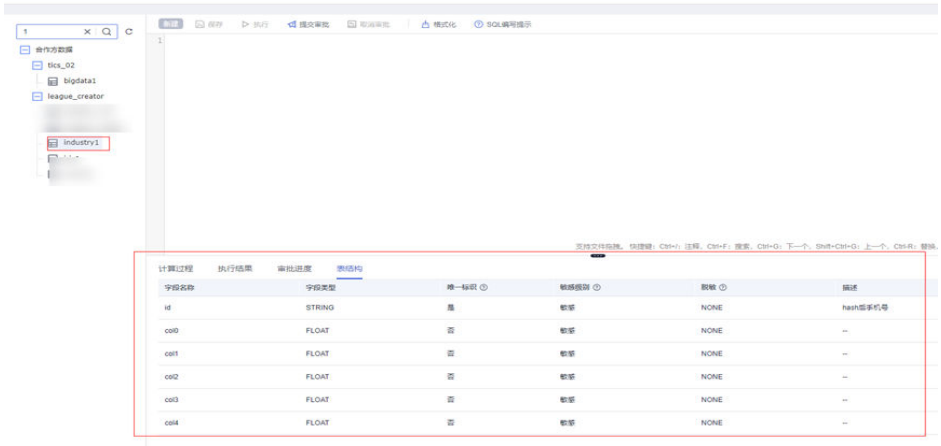
3.2.1.4 创建样本分布统计作业

创建样本分布统计作业步骤如下：

步骤1 在“作业管理 > 多方安全计算”页面单击创建，进入sql开发页面，展开左侧的“合作方数据”可以看到企业A、大数据厂商B发布的不同数据集。



步骤2 单击某一个数据集可以看到数据集的表结构信息。



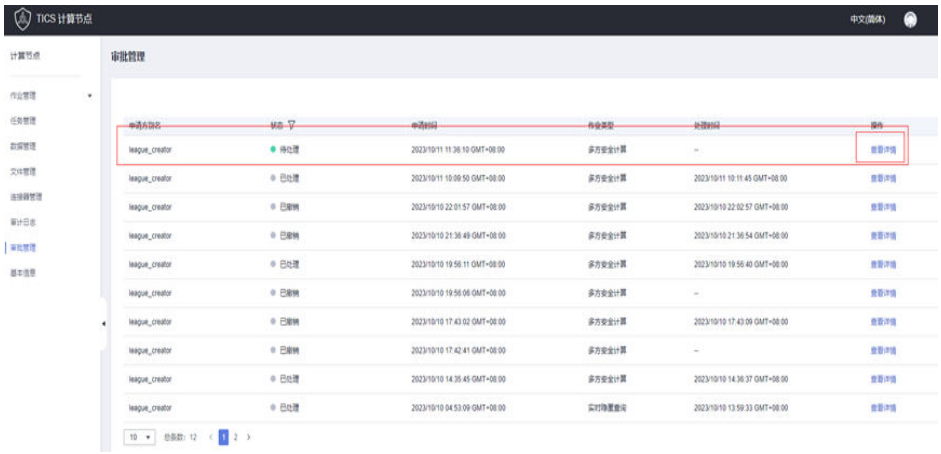
此时企业A可以编写如下的sql语句统计双方的数据碰撞后的正负样本总数，正负样本总数相加即为双方共有数据的总数。

```
select
  sum(
    case
      when i.label > 0 then 1
      else 0
    end
  ) as positive_count,
  sum(
    case
      when i.label <= 0 then 1
      else 0
    end
  ) as negtive_count
from
  tics_02.bigdata1 b
  join league_creator.industry1 i on b.id = i.id;
```

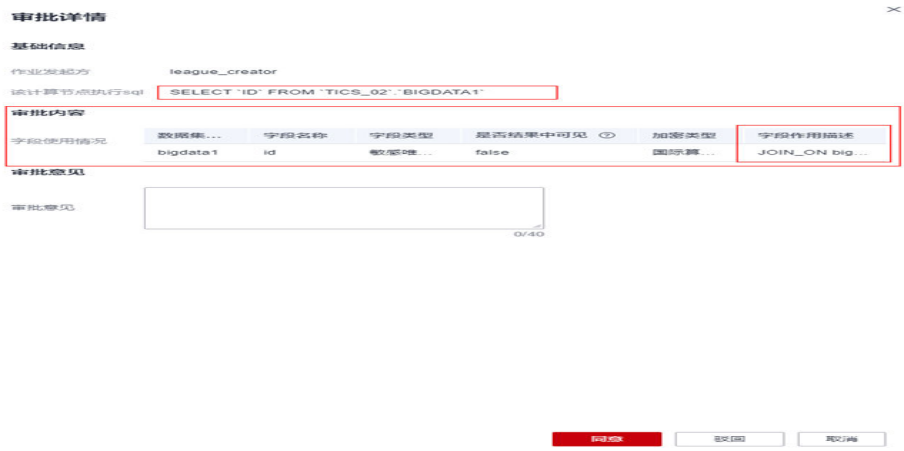
编写完成后单击“保存”和“提交审批”，由于这条sql使用到了大数据厂商B的数据集，为保证数据安全和参与方的知情权，tics服务会自动解析sql语句将大数据厂商B需要执行的sql语句发送到大数据厂商B的计算节点上，当大数据厂商B同意审批之后才可以执行该条sql。

除此之外，tics服务会基于数据集的安全隐私策略自动校验sql语句中字段的使用方式，如有违反字段隐私配置规则的语句会被明确拒绝。

步骤3 大数据厂商B在自己的计算节点单击“审批管理”模块，找到“待处理”的审批请求单击“查看详情”，可以看到企业A是如何使用自己的数据集的。



步骤4 确认无误后再单击“同意”即允许企业A使用己方的数据集进行联合统计。

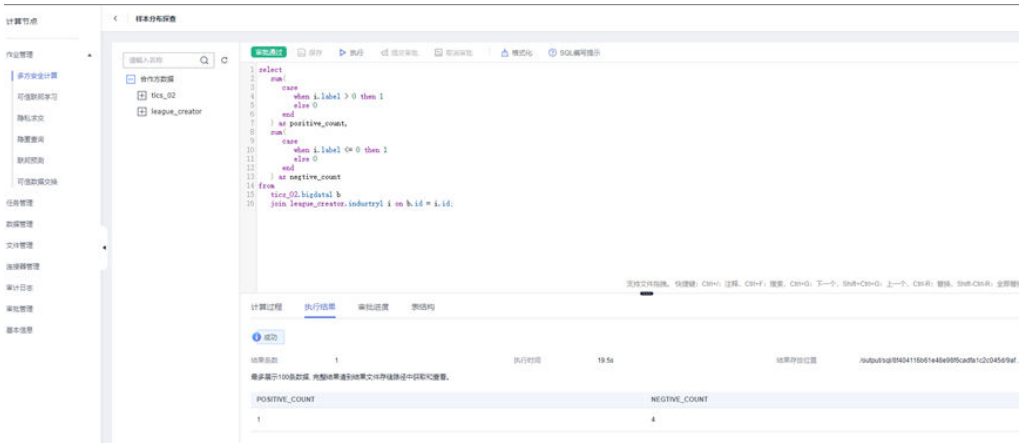


此时企业A在自己的计算节点上可以看到这个样本分布联合统计作业的状态已经变为了审批通过，“执行”按钮已经亮起。

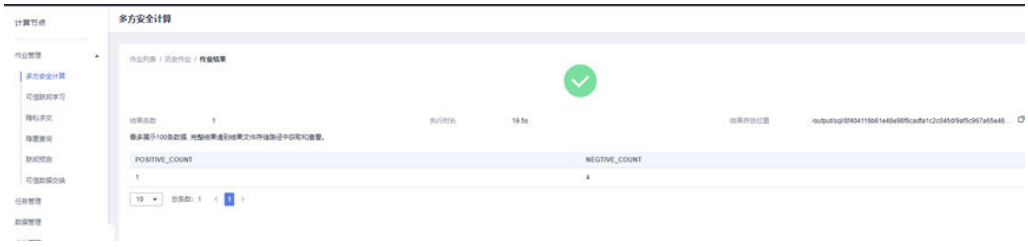
----结束

3.2.1.5 执行样本分布联合统计

企业A单击“执行”并等待一段时间之后，可以在页面下方“执行结果”看到sql的运行结果。



也可以通过“作业管理 > 多方安全计算 > 历史作业 > 查看结果”查看对应的结果。



3.2.1.6 数据优化

根据统计结果，双方可能会发现存在以下两个问题：

1. 碰撞后的数据总数比较小。

2. 碰撞后的数据分布不太均衡，负样本的比例过高。

这种情况下双方可以重复2-5的步骤更新自己提供的数据，多次执行样本分布统计直至达到比较满意的碰撞结果和分布结果。

至此联邦建模的数据准备阶段完成，接下来就是使用准备好的数据进行联邦建模。

3.2.2 使用 TICS 可信联邦学习进行联邦建模

3.2.2.1 场景描述

某企业A在进行新客户营销时的成本过高，想要通过引入外部数据的方式提高营销的效果，降低营销成本。

因此企业A希望与某大数据厂商B展开一项合作，基于双方共有的数据进行联邦建模，使用训练出的联邦模型对新数据进行联邦预测，筛选出高价值的潜在客户，再针对这些客户进行定向营销，达成提高营销效果、降低营销成本的业务诉求。

基于多方安全计算功能准备好合适的数据，本文主要介绍双方对已有的数据进行样本对齐、特征筛选和联邦建模，并对产生的模型进行评估。

3.2.2.2 准备数据

首先，企业A和大数据厂商B需要商议确定要提供的数据范围及对应的元数据信息，双方初始决定使用最近三个月的已有用户转化数据作为联邦训练的训练集和评估集，之后使用每周产生的新数据作为联邦预测的预测集。

表 3-6 企业 A 的数据

字段名称	字段类型	描述
id	string	hash过后的手机号字符串
col0-col4 label	float int	企业A数据特征 企业A对用户的标签属性

industry_all.csv

```
id,col0,col1,col2,col3,col4,label
5feceb66ffc86f38d952786c6d696c79c2dbc239dd4e91b46729d73a27fb57e9,-1.4092505981594734,-0.5893679205612337,-4.467396692737264,1.370376187747878,-1.236832500268279,1
6b86b273ff34fce19d6b804eff5a3f5747ada4eaa22f1d49c01e52ddb7875b4b,-1.5143756509526236,-1.9007475942180778,-5.617412558508785,2.2624690030531363,0.2886799132470795,0
d4735e3a265e16eee03f59718b9b5d03019c07d8b6c51f90da3a666eec13ab35,-1.768367116508903,1.2721845837988317,1.1497337351126178,-1.3322677230347135,0.9716103319957519,1
4b227777d4dd1fc61c6f884f48641d02b4d121d3fd328cb08b5531fcacdabf8a,0.37260755643902965,-0.2919401803207504,0.08086265459068624,0.3915016044811785,-0.01227642831882032,1
ef2d127de37b942baad06145e54b0c619a1f22327b2ebbcfbec78f5564afe39d,-2.963183239713765,0.15113195842028704,-3.8749664899828824,1.0598464836794779,-4.400883309764479,1
e7f6c011776e8db7cd330b54174fd76f7d0216b612387a5ffcfb81e6f0919683,-0.35120767987472346,1.8018318746365054,1.4431627055321963,0.33307198119824927,0.8626132267902704,0
7902699be42c8a8e46fbbb4501726517e86b22c56a189f7625a6da49081b2451,-2.6642415757243825,0.8836647864509011,-1.2340786744195096,-1.4945873871135977,-2.6999504889710626,1
2c624232cdd221771294dfbb310aca000a0df6ac8b66b696d90ef06fdefb64a3,3.0418810956792526,-0.6516843409674193,3.6616499550343105,0.035548733627266224,3.477873903864847,0
19581e27de7ced00ff1ce50b2047e7a567c76b1cbaebabe5ef03f7c3017bb5b7,-0.8239137547429756,0.7877120377027675,0.4296355963569869,-1.315646485980162,-1.652321610851379,1
```

4a44dc15364204a80fe80e9039455cc1608281820fe2b24f1e5233ade6af1dd5,0.24150521920304757,-0.21911471888817458,1.5143874504690156,-0.6652345113435701,0.17857570592695637,0
6b51d431df5d7f141cbececcf79edf3dd861c3b4069f0b11661a3eefacbba918,0.9669487046029339,-1.5427187535294289,2.490658334326762,0.4233920429380765,2.972622142213776,0
3fdba35f04dc8c462986c992bcf875546257113072a909c162f7e470e581e278,-1.847252571492643,0.4969814473631169,1.6544165211185982,-1.9450069019776826,0.39415199332185435,1
8527a891e224136950ff32ca212b45bc93f69fbb801c3b1ebedac52775f99e61,0.1622108420432964,0.17716762081899 43,4.55368226430978,-1.1032207991089722,2.375621631048501,0
e629fa6598d732768f7c726b4b621285f9c3b85303900aa912017db7617d8bdb,4.0527809556953,1.2053939486734313,3.260708709473611,1.1400990661834884,5.025657734758696,0
b17ef6d19c7a5b1ee83b907c595526dcb1eb06db8227d650d5dda0a9f4ce8cd9,-0.21563539406333465,0.5231489445682316,-2.639937297036372,2.3738020768486425,0.34341393069722226,1
4523540f1504cd17100c4835e85b7eefd49911580f8efff0599a8f283be6b9e3,-0.5935568930535046,-0.35175055806960276,0.9645122559090376,-0.017390131639078914,0.09256256476781644,1
4ec9599fc203d176a301536c2e091a19bc852759b255bd6818810a42c5fed14a,1.0066513658973761,-0.9724037855292317,1.3141152564284 94,0.363296291355055,5.171128738363806,0
9400f1b21cb527d7fa3d3eabba93557a18ebe7a2ca4e471cfe5e4c5b4ca7f767,0.1406977237605178,-1.455646778048175,-0.7223212422509906,1.265951206785454,-0.5504387433588089,1

表 3-7 大数据厂商 B 的数据

字段名称	字段类型	描述
id	string	hash过后的手机号字符串
f0-f4	float	大数据厂商数据特征

bigdata_all.csv

id,f0,f1,f2,f3,f4
5feceb66ffc86f38d952786c6d696c79c2dbc239dd4e91b46729d73a27fb57e9,-0.246852445,-1.761531756,-2.840375975,-0.562750693,-2.23499737
6b86b273ff34fce19d6b804eff5a3f5747ada4eaa22f1d49c01e52ddb7875b4b,-1.216062821,-1.093614452,-1.632396806,0.887601314,-4.40930101
4e07408562bedb8b60ce05c1decfe3ad16b72230967de01f640b7e4729b49fce,-0.150047899,-1.323266508,3.01679156,1.728583156,0.656158732
4b227777d4dd1fc61c6f884f48641d02b4d121d3fd328cb08b5531fcacdabf8a,-0.333871414,-1.21968931,-0.082894791,0.020390259,-0.076884947
ef2d127de37b942baad06145e54b0c619a1f22327b2ebbcfbec78f5564afe39d,-2.438861166,0.111880807,-3.51428545,1.123004835,0.228893969
e7f6c011776e8db7cd330b54174fd76f7d0216b612387a5ffcfb81e6f0919683,-2.759963795,0.405262468,1.264947591,1.027350049,1.293868423
7902699be42c8a8e46fbbb4501726517e86b22c56a189f7625a6da49081b2451,0.189352371,-0.607297495,-0.808339321,2.048455567,1.303872778
2c624232cdd221771294dfbb310aca000a0df6ac8b66b696d90ef06fdefb64a3,0.390064223,0.664175034,3.20228741,0.380574513,0.017733811
19581e27de7ced00ff1ce50b2047e7a567c76b1cbaebabe5ef03f7c3017bb5b7,0.379250902,1.962293246,0.066277661,3.083228267,1.952626328
4a44dc15364204a80fe80e9039455cc1608281820fe2b24f1e5233ade6af1dd5,-0.070919538,-2.219653517,1.461645551,1.66185096,0.778770954
4fc82b26aecb47d2868c4efbe3581732a3e7cbcc6c2efb32062c08170a05eeb8,-0.771151327,-1.184821181,-0.674077615,-0.379858223,0.158957184
6b51d431df5d7f141cbececcf79edf3dd861c3b4069f0b11661a3eefacbba918,-0.738091802,-1.474822882,2.93475295,-3.763763721,-1.817301398
3fdba35f04dc8c462986c992bcf875546257113072a909c162f7e470e581e278,-0.483250226,0.616586578,3.001851708,2.407914633,0.856369412
8527a891e224136950ff32ca212b45bc93f69fbb801c3b1ebedac52775f99e61,-0.789268594,1.071733834,3.763254446,-3.760298263,0.49776472
e629fa6598d732768f7c726b4b621285f9c3b85303900aa912017db7617d8bdb,-0.372531118,1.559382514,2.403559204,-0.041093457,0.169341125
b17ef6d19c7a5b1ee83b907c595526dcb1eb06db8227d650d5dda0a9f4ce8cd9,-2.773477116,-1.137653133,-1.50133841,0.82842642,-1.25476711
4523540f1504cd17100c4835e85b7eefd49911580f8efff0599a8f283be6b9e3,-1.542814756,1.019110477,1.395515599,0.539956076,0.100325065

4ec9599fc203d176a301536c2e091a19bc852759b255bd6818810a42c5fed14a,0.024227451,-1.087235302,3.67470964,-2.420729037,-3.132456573

其中为了保证数据安全，企业A和大数据厂商B通过讨论决定使用hash过后的手机号作为已有数据的唯一标识id字段，并将唯一标识作为数据对齐的依据。

3.2.2.3 发布数据集

企业A和大数据厂商B分别将自己的csv数据文件上传到自己的计算节点上，通过“数据管理”模块创建各自的数据集。

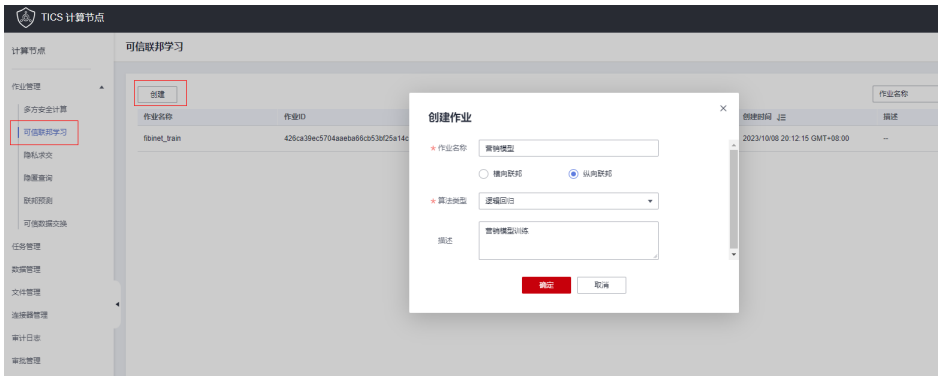
企业A的数据集如下：

大数据厂商B的数据集如下：

创建数据集后单击“发布”按钮即可将数据的元数据信息发布到tics空间侧，供其他合作方参考。

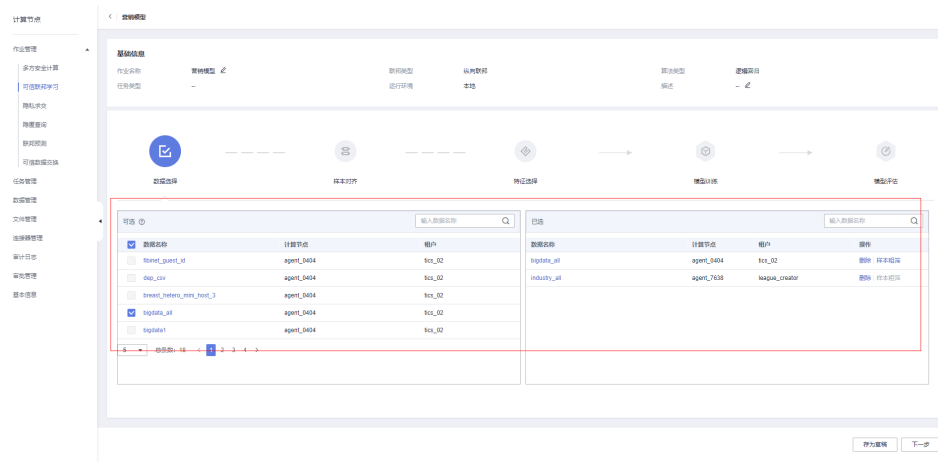
3.2.2.4 创建可信联邦学习作业

联邦建模的过程由企业A来操作，在“作业管理 > 可信联邦学习”页面单击“创建”，填写作业名称并选择算法类型后单击确定即进入联邦建模作业界面。本文逻辑回归算法为例。



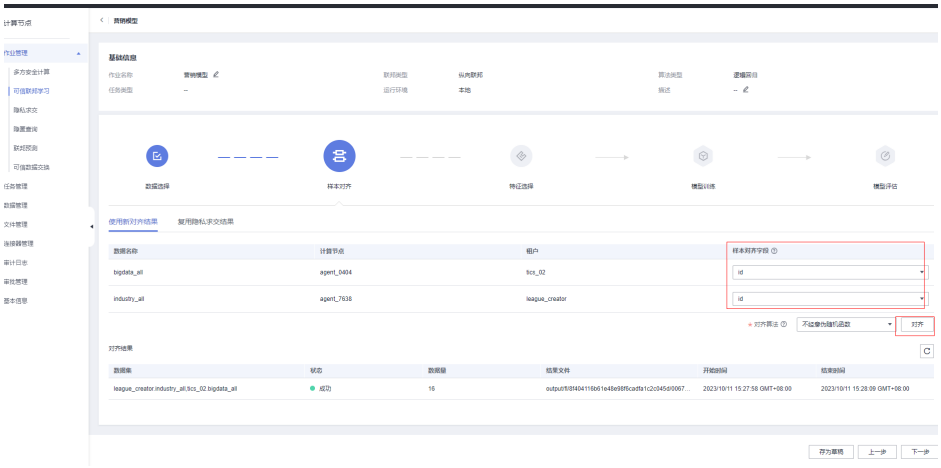
3.2.2.5 选择数据

首先企业A要在“数据选择”页面选择双方发布的数据集，已选择的数据集会出现在右侧，所选的数据集会用于后续的步骤。



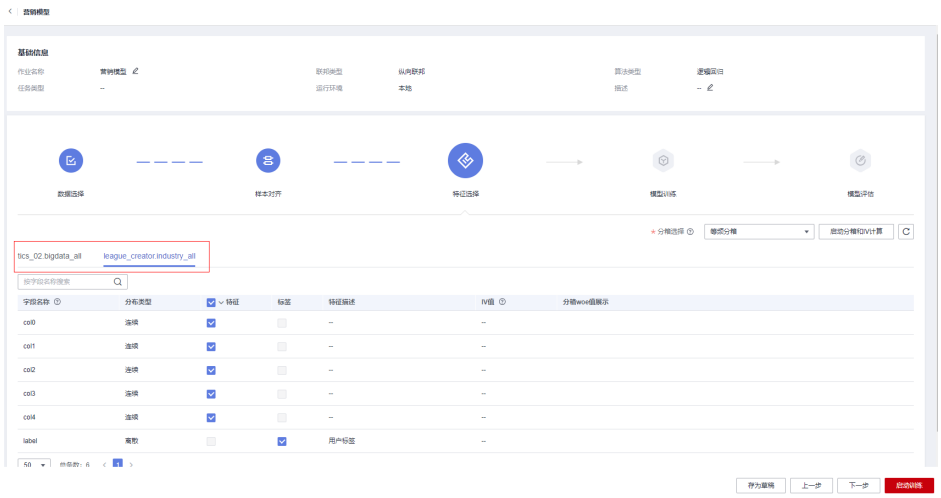
3.2.2.6 样本对齐

单击右下角的下一步进入“样本对齐”页面，这一步是为了进行样本的碰撞，过滤出共有的数据交集，作为后续步骤的输入。企业A需要选择双方的样本对齐字段，并单击“对齐”按钮执行样本对齐。执行完成后会在下方展示对齐后的数据量及对齐结果路径。



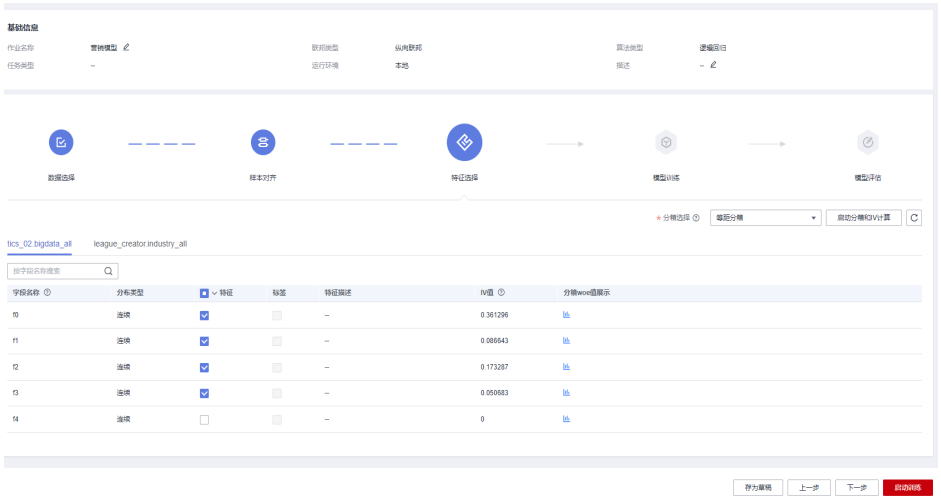
3.2.2.7 筛选特征

样本对齐执行完成后单击下一步进入“特征选择”页面，这一步企业A需要选出企业A自己和大数据厂商B的特征及标签用于后续的训练。



企业A可以选择特征及标签后“启动分箱和IV计算”，通过联邦的统计算法计算出所选特征的iv值，一般而言iv值较高的特征更有区分性，应该作为首选的训练特征；过低的iv值没有区分性会造成训练资源的浪费，过高的iv值又过于突出可能会过度影响训练出来的模型。

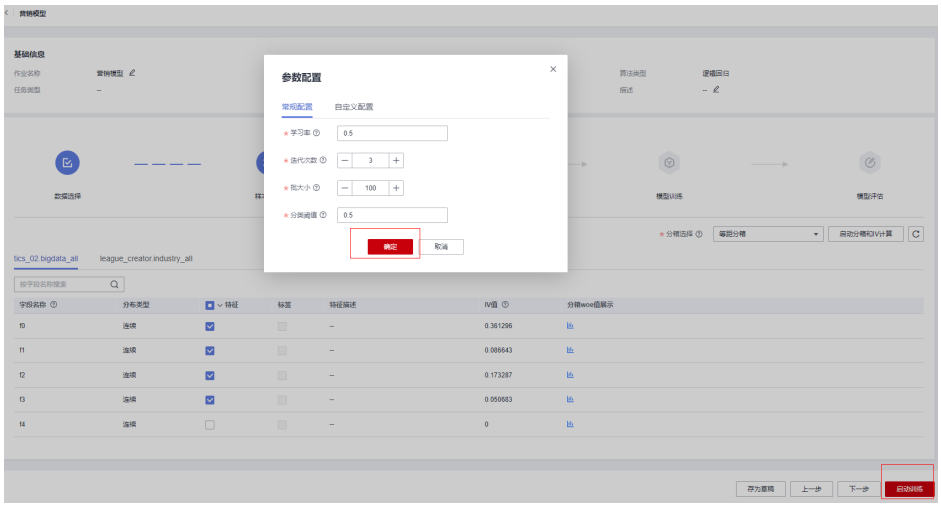
例如这里大数据厂商提供的f4特征iv值是0，说明这个特征对于标签的识别没有区分度，可以不选用；而f0、f2特征的iv值中等，适合作为模型的训练特征。

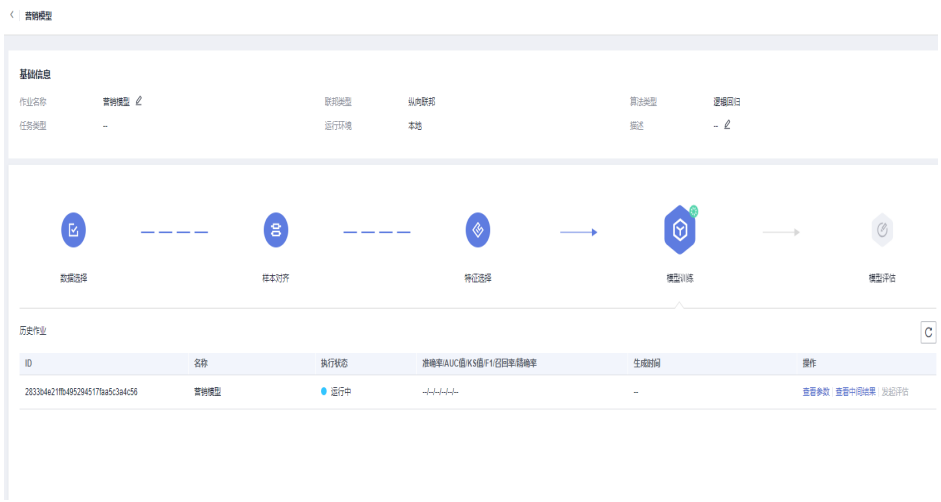


根据计算得出的iv值，企业A调整了训练使用的特征，没有选用双方提供的特征全集，去掉了部分iv值较低的特征，减少了无用的计算消耗。

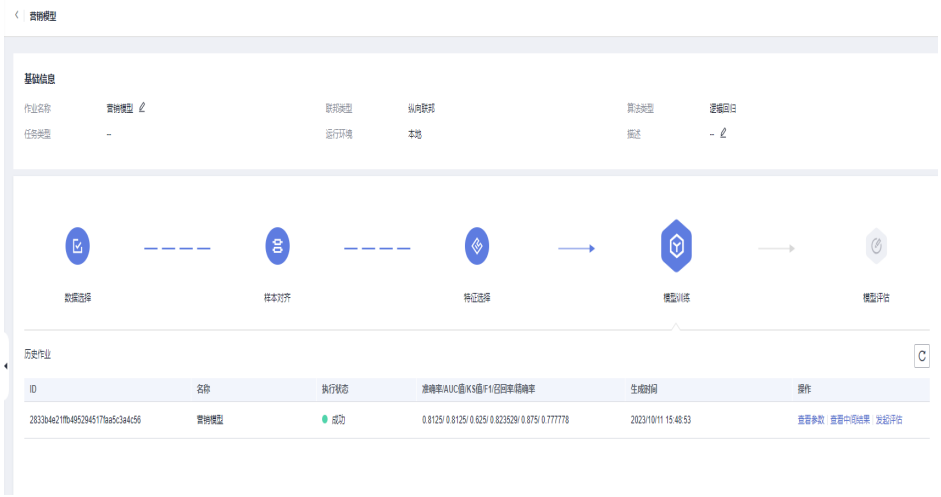
3.2.2.8 模型训练

企业A在完成特征选择后，可以单击右下角的“启动训练”按钮，配置训练的超参数并开始训练。





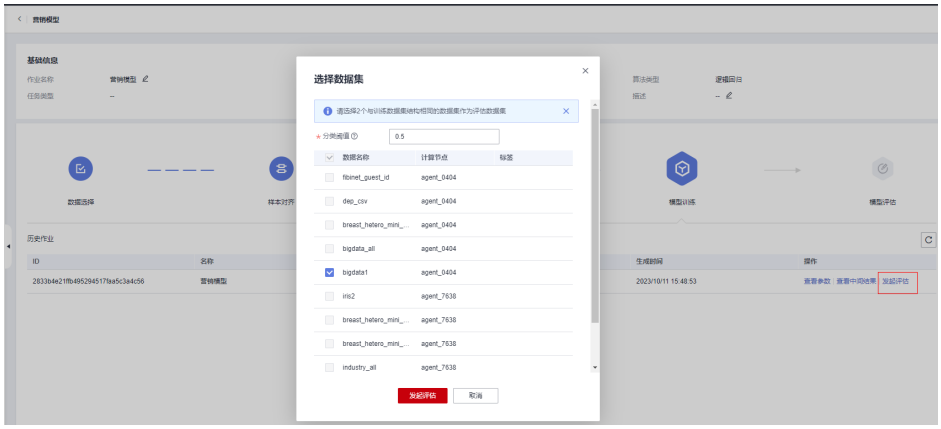
等待训练完成后就可以看到训练出的模型指标。

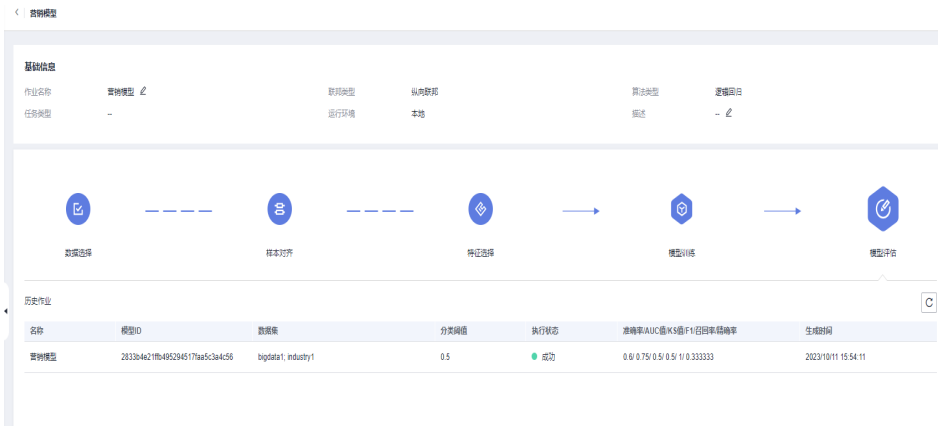


模型训练完成后如果指标不理想可以重复调整7、8两步的所选特征和超参数，直至训练出满意的模型。

3.2.2.9 模型评估

训练时的评估指标是用训练的数据集中随机采样的记录计算的，完成训练后企业A也可以使用其他的数据集对同一个模型进行多次的评估。单击“发起评估”选择训练参与方不同的数据集即可发起模型评估。





至此使用可信联邦学习进行联邦建模的过程已经完成，企业A已经训练出了一个符合自己要求的算法模型，后续文档会介绍如何使用已有的算法模型对新的数据进行预测。

3.2.3 使用 TICS 联邦预测进行新数据离线预测

3.2.3.1 场景描述

某企业A在进行新客户营销时的成本过高，想要通过引入外部数据的方式提高营销的效果，降低营销成本。

因此企业A希望与某大数据厂商B展开一项合作，基于双方共有的数据进行联邦建模，使用训练出的联邦模型对新数据进行联邦预测，筛选出高价值的潜在客户，再针对这些客户进行定向营销，达成提高营销效果、降低营销成本的业务诉求。

根据前一篇文章，企业A已经通过可信联邦学习功能训练出了一个预测客户时候是高价值用户的模型。

本文主要介绍企业A和大数据厂商B如何通过已有的模型对新的业务数据进行预测。

3.2.3.2 准备数据

企业A和大数据厂商B需要按照训练模型使用的特征，提供用于预测的数据集，要求预测的数据集特征必须包含训练时使用的特征。

表 3-8 企业 A 的数据

字段名称	字段类型	描述
id	string	hash过后的手机号字符串
col0-col4	float	企业A数据特征

industry_predict.csv

```
id,col0,col1,col2,col3,col4
4e07408562bedb8b60ce05c1decfe3ad16b72230967de01f640b7e4729b49fce,-0.823913755,0.787712038,0.429635596,-1.315646486,-1.652321611
2c624232cdd221771294dfbb310aca000a0df6ac8b66b696d90ef06fdefb64a3,3.041881096,-0.651684341,3.661649955,0.035548734,3.477873904
8527a891e224136950ff32ca212b45bc93f69fbb801c3b1ebedac52775f99e61,0.162210842,0.177167621,4.553682264,-1.103220799,2.375621631
```

注意由于这是新产生的业务数据，企业A并不知道这些用户是否是高价值用户，因此没有label用户标签字段。

表 3-9 大数据厂商 B 的数据

字段名称	字段类型	描述
id	string	hash过后的手机号字符串
f0-f4	float	大数据厂商数据特征

bigdata_all.csv

```
id,f0,f1,f2,f3,f4
5feceb66ffc86f38d952786c6d696c79c2dbc239dd4e91b46729d73a27fb57e9,-0.246852445,-1.761531756,-2.840375975,-0.562750693,-2.23499737
6b86b273ff34fce19d6b804eff5a3f5747ada4eaa22f1d49c01e52ddb7875b4b,-1.216062821,-1.093614452,-1.632396806,0.887601314,-4.40930101
4e07408562bedb8b60ce05c1decfe3ad16b72230967de01f640b7e4729b49fce,-0.150047899,-1.323266508,3.01679156,1.728583156,0.656158732
4b227777d4dd1fc61c6f884f48641d02b4d121d3fd328cb08b5531fcacdabf8a,-0.333871414,-1.21968931,-0.082894791,0.020390259,-0.076884947
ef2d127de37b942baad06145e54b0c619a1f22327b2ebbcfbec78f5564afe39d,-2.438861166,0.111880807,-3.51428545,1.123004835,0.228893969
e7f6c011776e8db7cd330b54174fd76f7d0216b612387a5ffcfb81e6f0919683,-2.759963795,0.405262468,1.264947591,1.027350049,1.293868423
7902699be42c8a8e46fbbb4501726517e86b22c56a189f7625a6da49081b2451,0.189352371,-0.607297495,-0.808339321,2.048455567,1.303872778
2c624232cdd221771294dfbb310aca000a0df6ac8b66b696d90ef06fdefb64a3,0.390064223,0.664175034,3.20228741,0.380574513,0.017733811
19581e27de7ced00ff1ce50b2047e7a567c76b1cbaebabe5ef03f7c3017bb5b7,0.379250902,1.962293246,0.066277661,3.083228267,1.952626328
4a44dc15364204a80fe80e9039455cc1608281820fe2b24f1e5233ade6af1dd5,-0.070919538,-2.219653517,1.461645551,1.66185096,0.778770954
4fc82b26aecb47d2868c4efbe3581732a3e7cbcc6c2efb32062c08170a05eeb8,-0.771151327,-1.184821181,-0.674077615,-0.379858223,0.158957184
```

6b51d431df5d7f141cbececcf79edf3dd861c3b4069f0b11661a3eefacbba918,-0.738091802,-1.474822882,2.934
75295,-3.763763721,-1.817301398
3fdba35f04dc8c462986c992bcf875546257113072a909c162f7e470e581e278,-0.483250226,0.616586578,3.001
851708,2.407914633,0.856369412
8527a891e224136950ff32ca212b45bc93f69fbb801c3b1ebedac52775f99e61,-0.789268594,1.071733834,3.763
254446,-3.760298263,0.49776472
e629fa6598d732768f7c726b4b621285f9c3b85303900aa912017db7617d8bdb,-0.372531118,1.559382514,2.4
03559204,-0.041093457,0.169341125
b17ef6d19c7a5b1ee83b907c595526dcb1eb06db8227d650d5dda0a9f4ce8cd9,-2.773477116,-1.137653133,-1.
50133841,0.82842642,-1.25476711
4523540f1504cd17100c4835e85b7eefd49911580f8efff0599a8f283be6b9e3,-1.542814756,1.019110477,1.395
515599,0.539956076,0.100325065
4ec9599fc203d176a301536c2e091a19bc852759b255bd6818810a42c5fed14a,0.024227451,-1.087235302,3.67
470964,-2.420729037,-3.132456573

其中为了保证数据安全，企业A和大数据厂商B通过讨论决定使用hash过后的手机号作为已有数据的唯一标识id字段，并将唯一标识作为数据对齐的依据。

3.2.3.3 发布数据集

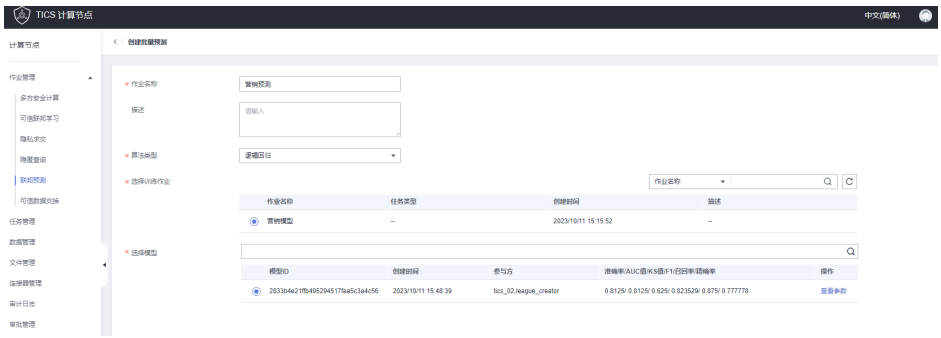
企业A将自己的需要预测的csv数据文件上传到自己的计算节点上，通过“数据管理”模块创建用于预测的数据集。

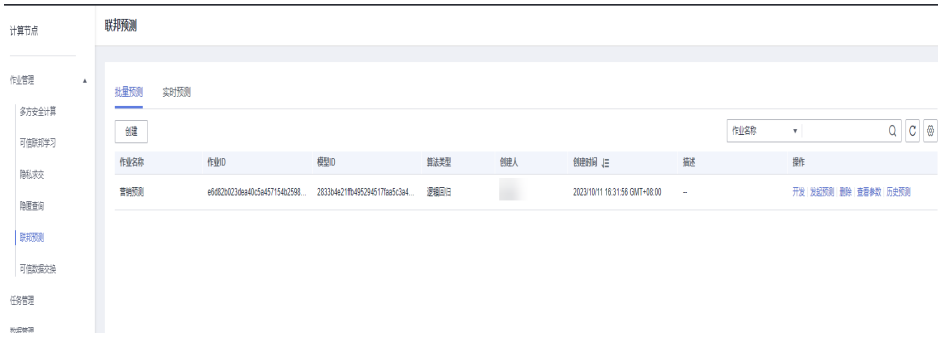
企业A预测数据集如下：

大数据厂商B仍使用训练时的提供的全量数据作为预测数据集，没有发布新的数据集。

3.2.3.4 创建联邦预测作业

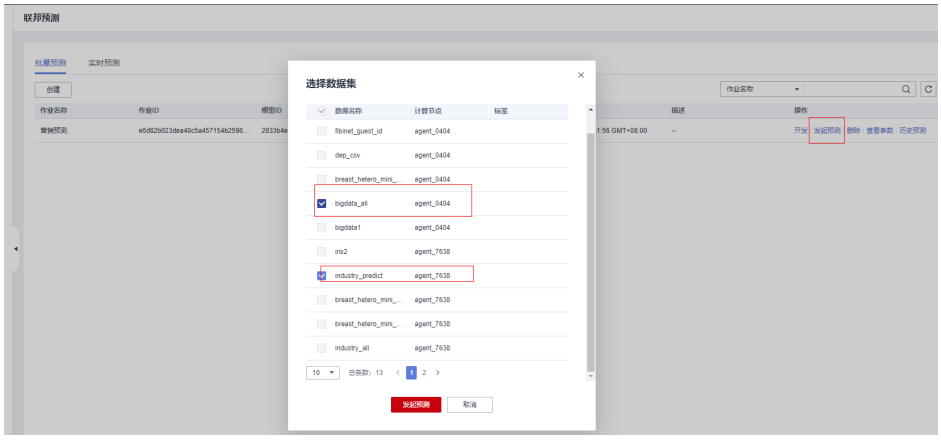
企业A单击“联邦预测 > 批量预测 > 创建”按钮，进入联邦预测作业的创建页面。企业A需要通过“算法类型”、“训练作业”等筛选条件可以找到用于预测的模型，点选使用的模型后单击“确定”按钮即完成联邦预测作业的创建。

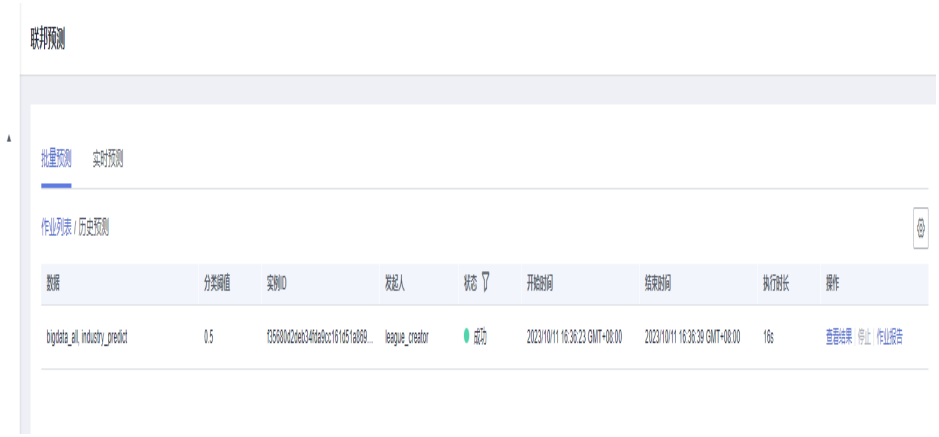




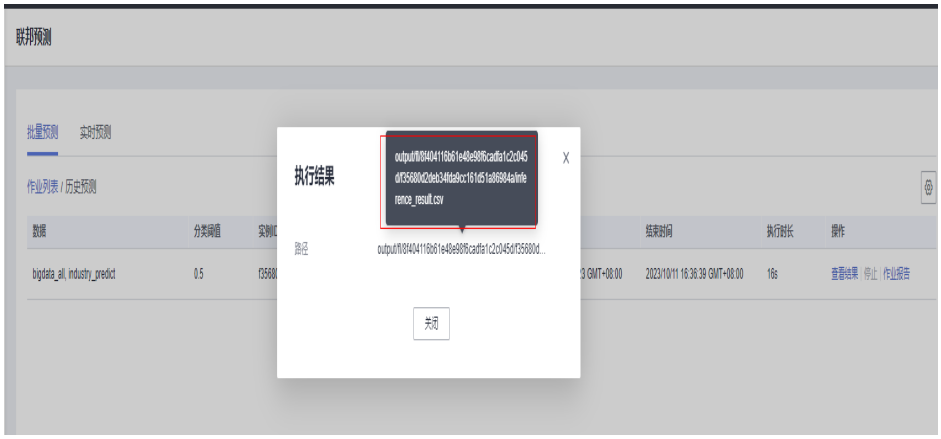
3.2.3.5 发起联邦预测

企业A单击“发起预测”按钮，选择己方和大数据厂商B的预测数据集，单击确定即可发起预测。





TICS服务会对两方的数据先进行样本对齐，并对双方共有的数据进行联邦预测，预测的结果会保存在企业A（作业发起方）的计算节点上。企业A可以通过obs服务或者登录到计算节点后台获取到对应路径的文件。



当只有一方提供特征时，预测的结果如下，第一列是用户的id，第二列是用户是否是高价值用户的标签，第三列、第四列是对应的概率：

```
id,label,proba_0,proba_1
4e07408562bedb8b60ce05c1decfe3ad16b72230967de01f640b7e4729b49fce, 1,0.268941,0.731059
2c624232cdd221771294dfbb310aca000a0df6ac8b66b696d90ef06fdefb64a3, 0,0.731059,0.268941
8527a891e224136950ff32ca212b45bc93f69fbb801c3b1ebedac52775f99e61, 0,0.731059,0.268941
```

可以看出企业A提供的预测数据集中有部分用户被模型预测成了高价值的客户，后续企业A可以对这一部分用户进行定向精准营销，缩小营销广告的投放范围，减少了营销的成本。

当两方都提供特征时，预测结果分为对齐id文件（只有一列id）和预测结果文件（包括预测结果标签、0的概率、1的概率），两个文件的行数相等且每一行相互对应。

至此，企业A完成了整个TICS联邦建模的流程，并将模型应用到了营销业务当中。这个预测作业可以作为后续持续预测的依据，企业A可以定期地使用模型预测自己的新业务数据。同时企业A也可以根据新积累的数据训练出新的模型，进一步优化模型预测的精确率，再创建新的联邦预测作业，产出更精准的预测结果供业务使用。

3.3 隐私求交黑名单共享场景

3.3.1 场景描述

有效的风险控制能够消灭或减少风险事件发生的各种可能性，或减少风险事件发生时造成的损失，对于企业具有重要意义。现阶段，企业级的单方风控体系已逐步建立，在机构内数据统一共享的基础上实现了覆盖业务前、中、后各环节的智能风控。然而，单方数据风控面临存在数据不全面、风控不及时的问题。随着隐私计算等技术为数据要素的有效流通提供了必要手段，多方数据联合风控成为新趋势。其中，黑名单共享查询是风控中的一个重要环节，企业间的黑名单共享能有效发挥风险联防联控效用。

在信息核验过程中，通过隐私计算技术实现多方黑名单数据共享，对电诈、洗钱、骗贷等行为的黑名单用户进行安全求交、匿踪查询，能够有效提升客户背景调查的安全可信程度。

现有两家企业A、B，双方决定通过TICS平台实现黑名单数据共享，通过隐私求交作业计算两方黑名单ID交集。本文以企业A为计算作业的发起方为例。

3.3.2 准备数据

A方提供了待查询的用户ID数据，样例如下：

blacklist_query.csv

```
id
1914fd1aef9346e7a1b0a63c95aa918e
6b86b273ff34fce19d6b804eff5a3f57
66985617b4f74d14b4eceeaa25d61f5e
935098fe075343a5864db257bc40dfa5
d4735e3a265e16eee03f59718b9b5d03
4ec9599fc203d176a301536c2e091a19
79f2c17703f0442ebb4d57800e2666a2
ea9257156eca491cada3e38b6d1000e9
acdef43b89014fab8d5916b7ce756f75
3855ad3d30c446ed94a0f34dbd7be63e
d9465361c73549739fd0ae252705c190
```

B方提供了共享的黑名单用户ID名单，样例如下：

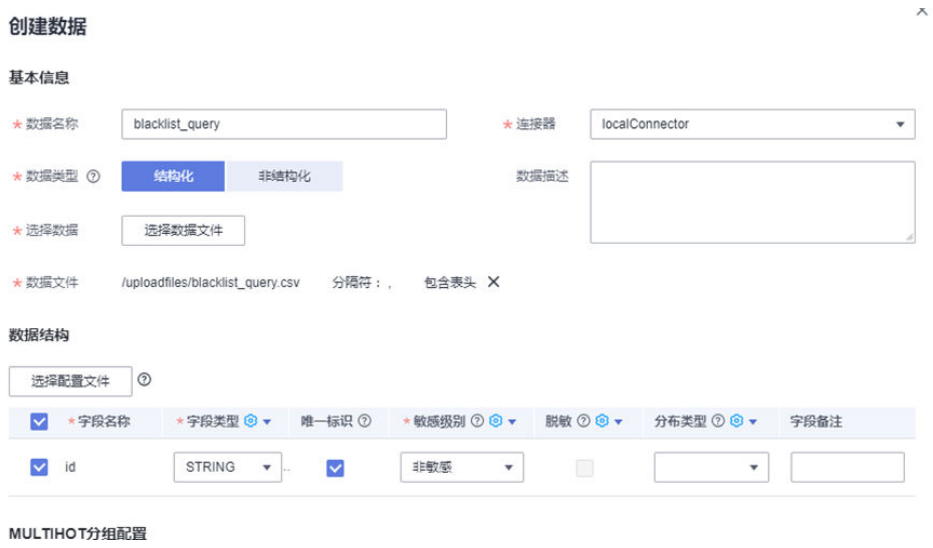
blacklist_all.csv

```
id
5feceb66ffc86f38d952786c6d696c79
6b86b273ff34fce19d6b804eff5a3f57
d4735e3a265e16eee03f59718b9b5d03
4b227777d4dd1fc61c6f884f48641d02
ef2d127de37b942baad06145e54b0c61
e7f6c011776e8db7cd330b54174fd76f
7902699be42c8a8e46fbbb4501726517
2c624232cdd221771294dfbb310aca00
19581e27de7ced00ff1ce50b2047e7a5
4a44dc15364204a80fe80e9039455cc1
6b51d431df5d7f141cbececcf79edf3d
3fdba35f04dc8c462986c992bcf87554
8527a891e224136950ff32ca212b45bc
e629fa6598d732768f7c726b4b621285
b17ef6d19c7a5b1ee83b907c595526dc
4523540f1504cd17100c4835e85b7eef
4ec9599fc203d176a301536c2e091a19
9400f1b21cb527d7fa3d3eabba93557a
```

3.3.3 发布数据集

企业A和企业B分别将自己的csv数据文件上传到自己的计算节点上，通过“数据管理”模块创建各自的数据集，并单击“发布”。

以企业A为例，数据集信息如下：

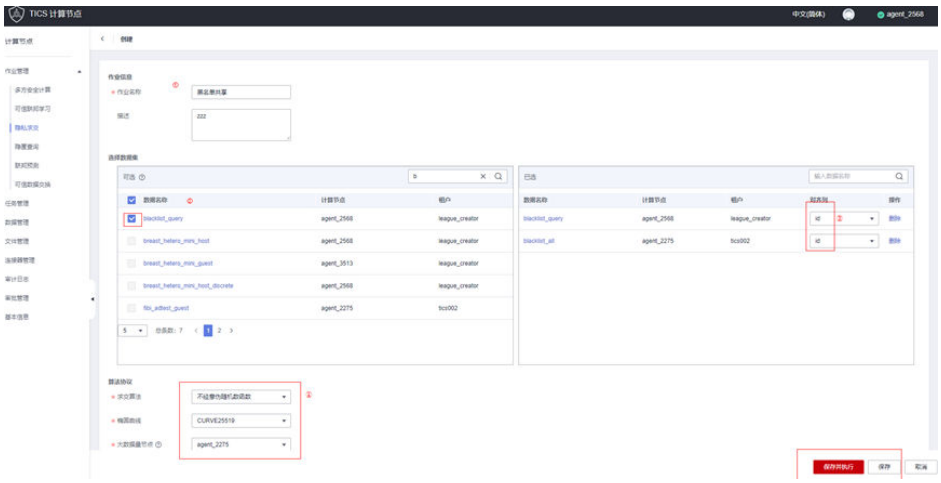


说明

隐私求交场景需要将求交的字段设置为“非敏感”的唯一标识。

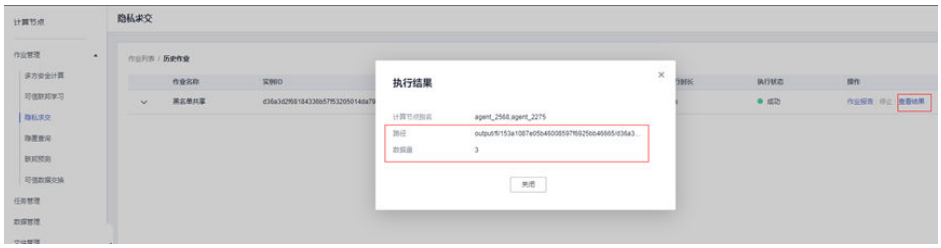
3.3.4 创建并运行隐私求交作业

企业A单击“作业管理 > 隐私求交 > 创建”，依次填写作业名称、选择需要求交的数据集和对应的求交列、选择算法协议及各种参数，再单击“保存并执行”即可发起一次隐私求交查询。



3.3.5 查看求交结果

隐私求交作业执行完成后，企业A可以通过单击“历史作业 > 查看结果”看到隐私求交作业的运行结果，包括交集的大小和交集文件的路径。



打开obs到指定目录下查看，可以看到有两个结果文件，其中一个是交集记录的序号alignedIds.csv，另一个是交集记录的id alignedOriginalIds.csv。

alignedIds.csv的内容如下：

```
1
4
5
```

即指企业A提供的数据中，第1条、第4条、第5条（从0开始）记录在企业B的黑名单当中。

blacklist_query.csv

```
id
1914fd1aef9346e7a1b0a63c95aa918e
6b86b273ff34fce19d6b804eff5a3f57
66985617b4f74d14b4eceeaa25d61f5e
935098fe075343a5864db257bc40dfa5
d4735e3a265e16eee03f59718b9b5d03
4ec9599fc203d176a301536c2e091a19
79f2c17703f0442ebb4d57800e2666a2
ea9257156eca491cada3e38b6d1000e9
acdef43b89014fab8d5916b7ce756f75
3855ad3d30c446ed94a0f34dbd7be63e
d9465361c73549739fd0ae252705c190
```

alignedOriginalIds.csv中即为这些在黑名单交集中的用户ID：

```
6b86b273ff34fce19d6b804eff5a3f57
d4735e3a265e16eee03f59718b9b5d03
4ec9599fc203d176a301536c2e091a19
```

至此，企业A可以得知这些用户存在于企业B的黑名单当中，这些用户的业务存在较高的风险。企业A可以提前做好风险预案，控制风险的发生或者减少风险造成的损失。

3.4 实时隐匿查询场景

3.4.1 外部数据共享

3.4.1.1 场景描述

现有企业A和企业B达成了一项数据共享合作协议，企业B允许企业A根据用户id查询企业B的数据，辅助企业A的实时分析业务。而企业A不想暴露给企业B自己查询的用户id，因为查询该用户的信息隐含着“该用户是企业A的客户”的信息，存在用户隐私泄露的风险。

企业A和企业B可以使用TICS服务的实时隐匿查询功能，既能满足实时业务高效低延迟的业务需求，又能避免暴露企业A想要查询哪个用户的隐私安全风险。

3.4.1.2 准备数据

企业A的实时业务不需要准备数据，在发起查询时通过参数传递需要查询的用户id。

表 3-10 企业 B 用户画像数据

字段名称	字段类型	描述
id	string	hash过后的手机号字符串
f0-f4	float	用户数据画像特征

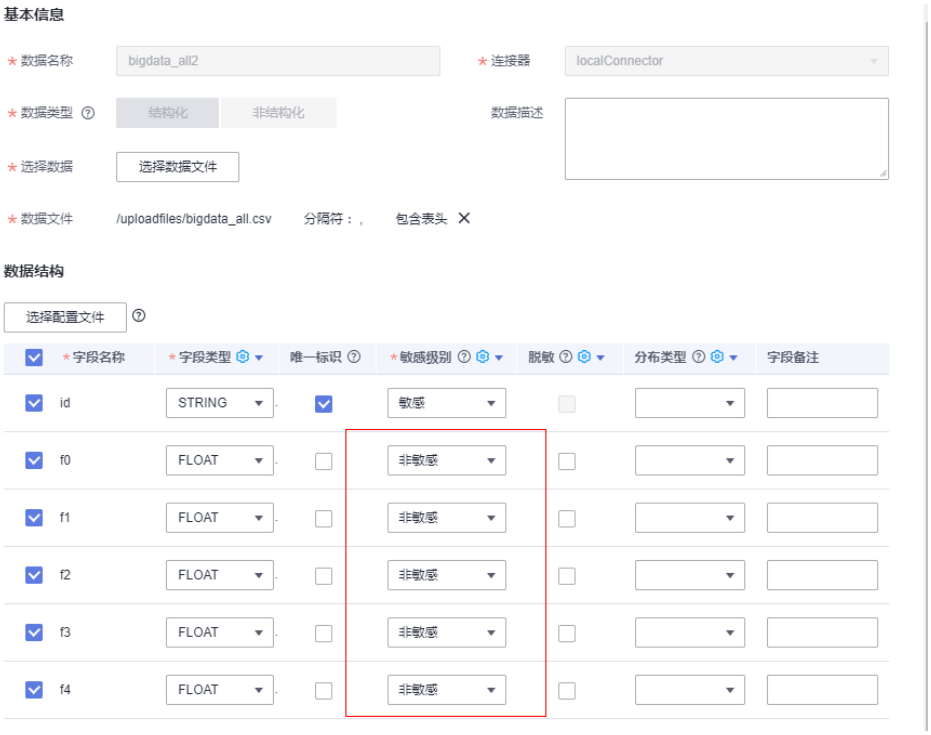
bigdata_all.csv

```
id,f0,f1,f2,f3,f4
5feceb66ffc86f38d952786c6d696c79c2dbc239dd4e91b46729d73a27fb57e9,-0.246852445,-1.761531756,-2.840375975,-0.562750693,-2.23499737
6b86b273ff34fce19d6b804eff5a3f5747ada4eaa22f1d49c01e52ddb7875b4b,-1.216062821,-1.093614452,-1.632396806,0.887601314,-4.40930101
4e07408562bedb8b60ce05c1decfe3ad16b72230967de01f640b7e4729b49fce,-0.150047899,-1.323266508,3.01679156,1.728583156,0.656158732
4b227777d4dd1fc61c6f884f48641d02b4d121d3fd328cb08b5531fcacdabf8a,-0.333871414,-1.21968931,-0.082894791,0.020390259,-0.076884947
ef2d127de37b942baad06145e54b0c619a1f22327b2ebbcfbec78f5564afe39d,-2.438861166,0.111880807,-3.51428545,1.123004835,0.228893969
e7f6c011776e8db7cd330b54174fd76f7d0216b612387a5ffcfb81e6f0919683,-2.759963795,0.405262468,1.264947591,1.027350049,1.293868423
7902699be42c8a8e46fbbb4501726517e86b22c56a189f7625a6da49081b2451,0.189352371,-0.607297495,-0.808339321,2.048455567,1.303872778
2c624232cdd221771294dfbb310aca000a0df6ac8b66b696d90ef06fdefb64a3,0.390064223,0.664175034,3.20228741,0.380574513,0.017733811
19581e27de7ced00ff1ce50b2047e7a567c76b1cbaebabe5ef03f7c3017bb5b7,0.379250902,1.962293246,0.066277661,3.083228267,1.952626328
4a44dc15364204a80fe80e9039455cc1608281820fe2b24f1e5233ade6af1dd5,-0.070919538,-2.219653517,1.461645551,1.66185096,0.778770954
4fc82b26aecb47d2868c4efbe3581732a3e7cbcc6c2efb32062c08170a05eeb8,-0.771151327,-1.184821181,-0.674077615,-0.379858223,0.158957184
6b51d431df5d7f141cbececcf79edf3dd861c3b4069f0b11661a3eefacbba918,-0.738091802,-1.474822882,2.93475295,-3.763763721,-1.817301398
3fdba35f04dc8c462986c992bcf875546257113072a909c162f7e470e581e278,-0.483250226,0.616586578,3.001851708,2.407914633,0.856369412
8527a891e224136950ff32ca212b45bc93f69fbb801c3b1ebedac52775f99e61,-0.789268594,1.071733834,3.763254446,-3.760298263,0.49776472
e629fa6598d732768f7c726b4b621285f9c3b85303900aa912017db7617d8bdb,-0.372531118,1.559382514,2.403559204,-0.041093457,0.169341125
b17ef6d19c7a5b1ee83b907c595526dcb1eb06db8227d650d5dda0a9f4ce8cd9,-2.773477116,-1.137653133,-1.50133841,0.82842642,-1.25476711
4523540f1504cd17100c4835e85b7eefd49911580f8efff0599a8f283be6b9e3,-1.542814756,1.019110477,1.395515599,0.539956076,0.100325065
4ec9599fc203d176a301536c2e091a19bc852759b255bd6818810a42c5fed14a,0.024227451,-1.087235302,3.67470964,-2.420729037,-3.132456573
```

3.4.1.3 发布数据集

企业B分别自己的csv数据文件上传到自己的计算节点上，通过“数据管理”模块创建各自的数据集，并单击“发布”。

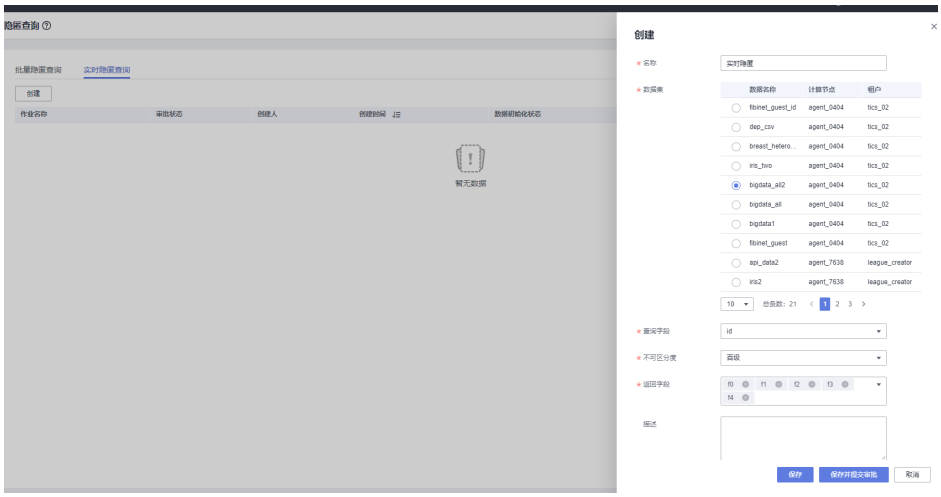
企业B的数据集如下：



创建数据集后单击“发布”按钮即可将数据的元数据信息发布到tics空间侧，供其他合作方参考。

3.4.1.4 创建实时隐匿查询作业

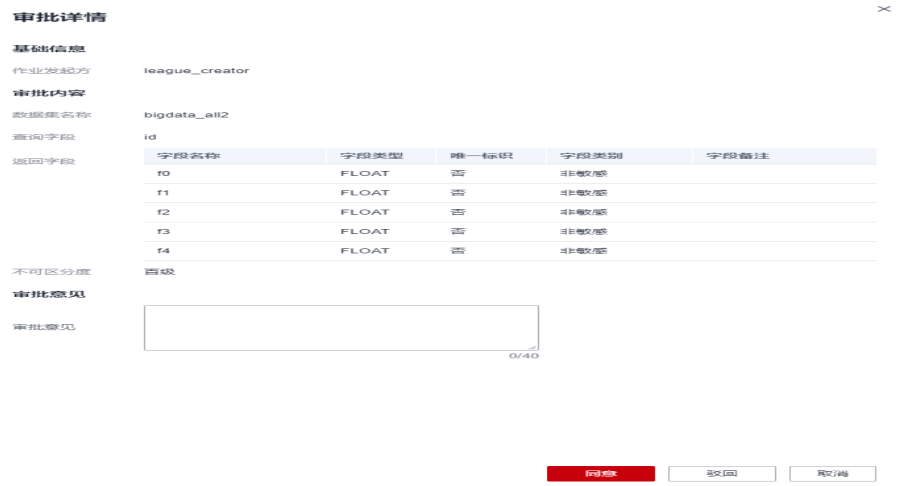
实时隐匿查询作业需要由数据查询方创建作业，企业A单击“作业管理 > 隐匿查询 > 实时隐匿查询”页面的创建按钮，填写相关信息，例如：



其中“不可区分度”即为实时隐匿查询的安全级别，不可区分度越高，则安全级别越高，但查询的速度会变慢，传输的数据量也会变大。

企业A完成信息选择后，单击“保存并提交审批”即可向数据提供方企业B发送一条审批信息。

企业B在自己的计算节点上可以单击“审批管理”，选择“待处理”的实时隐匿查询作业审批，可以看到自己的数据被如何使用。待企业B同意审批之后，企业A可以开始执行实时隐匿查询作业。

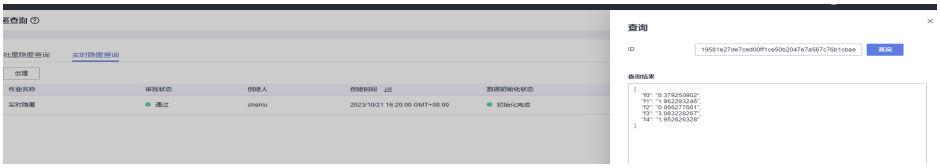


3.4.1.5 执行实时隐匿查询作业

企业A在发起实时隐匿查询前需要先执行数据初始化。

待实时预测作业初始化完成后，企业A可以通过页面单击“执行”试用发起查询。

例如查询id为“19581e27de7ced00ff1ce50b2047e7a567c76b1cbaebabe5ef03f7c3017bb5b7”这样的一条数据，查询结果中即会返回企业A所选择的企业B的数据字段。



同时企业A的业务系统后台也可以通过API调用的方式调用企业A计算节点的接口发起实时隐匿查询，更好地服务生产业务。

3.5 可信数据交换场景

3.5.1 场景描述

数据商业空间中公司B针对公司A的某些数据资产存在业务需求，由于安全性和数据主权的考虑，公司A与公司B基于TICS完成数据资产的交换。基于TICS进行数据资产交换，保证公司A的数据主权、公司B的数据可获得，同时保证交换过程安全可信。

以下是数据拥有方公司A和数据需求方公司B基于TICS平台的操作。

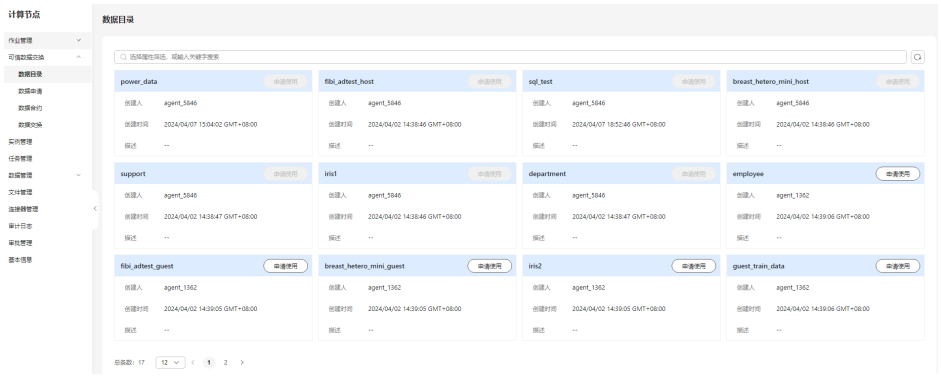
3.5.2 创建数据

数据拥有方公司A创建和发布数据集。可供选择有两种数据资产类型：结构化数据集、非结构化数据集。创建数据集后，发布数据集，此时对空间内的所有代理可见。

3.5.3 申请使用数据

步骤1 数据需求方公司B在自己的计算节点页面上可以查看数据目录，找到数据拥有方公司A创建并发布的数据。

图 3-24 创建数据申请

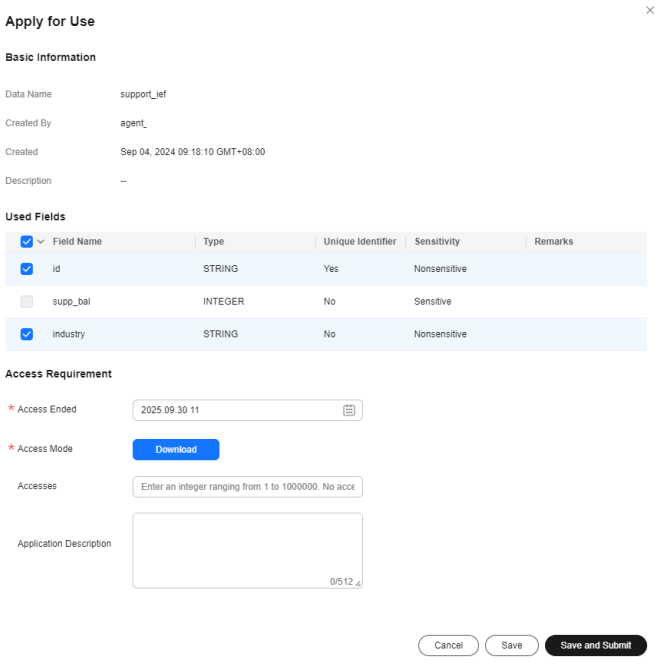


步骤2 对数据集单击“申请使用”，在弹窗中填写需要使用的字段和访问需求，保存后可以提交审批，由公司A审核。

访问需求包括：

- 访问截止时间：设置访问的时间限制，超过访问时间后，对方的访问权限将被收回，交换至对方的加密文件将被删除。
- 访问方式：基于TICS平台进行下载。
- 访问次数：用户可以访问次数的最大限制；超过访问次数，用户将无法访问作业文件。如果不填写，用户在访问截止时间前无限次访问。

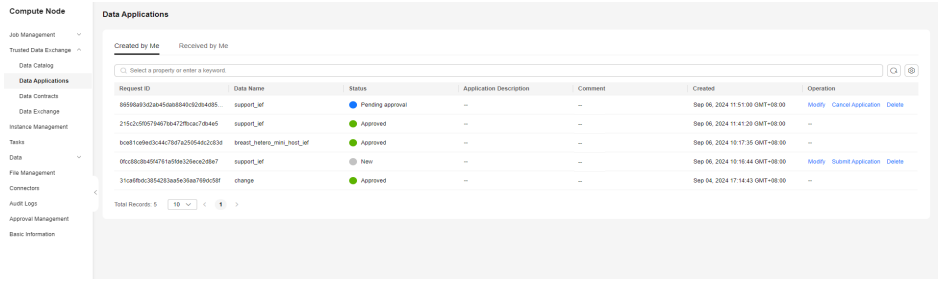
图 3-25 设置使用的字段及访问的需求



步骤3 单击保存或者保存并提交审批。

在“可信数据交换 > 数据申请 > 我创建的”的页签下可以查看、编辑、删除已创建的申请及对应的状态。

图 3-26 已创建的申请



----结束

3.5.4 审批数据申请

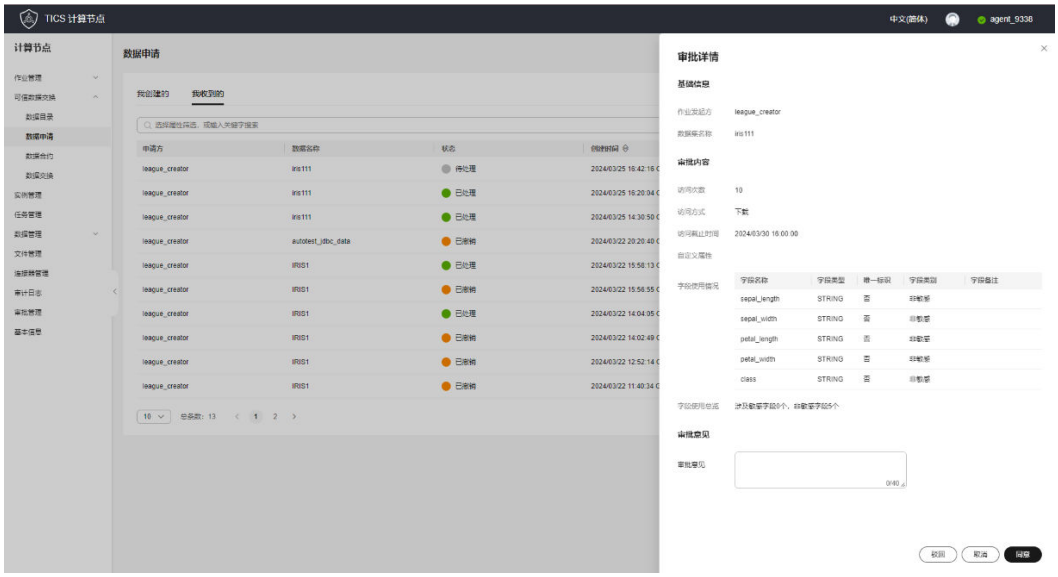
步骤1 数据拥有方公司A登录进入计算节点页面。

步骤2 在左侧导航树上选择“可信数据交换 > 数据申请”，打开数据申请页面。

步骤3 在数据申请页面单击“我收到的”，查看供数方节点收到的申请列表。

数据来源为数据需求方公司B发送来的使用申请：申请交换的数据集、数据集字段（结构化数据才有该字段）。

步骤4 在申请列表中选择申请状态为“待处理”，单击“查看详情”了解用数方需求。根据实际情况同意或者驳回申请。



----结束

3.5.5 创建合约

数据拥有方公司A同意数据需求方公司B的数据使用申请后，可以由公司A创建合约，合约是需要双方同意的数据使用证明。

合约内容包括：合约名称、合约描述、数据信息、公司B的访问需求、访问限制和自定义限制。其中数据信息、公司B的访问需求来自于公司B的数据使用申请，合约名称、合约描述、访问限制和自定义限制由公司A在创建合约时定义。

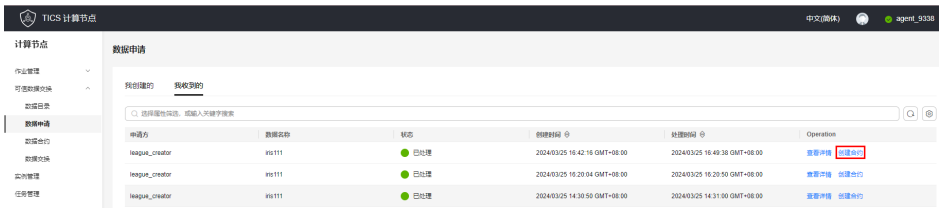
约束限制

- 访问限制：目前不限制数据的使用环境和应用。
- 自定义限制：自定义其他属性，比如设置文件访问者的名称、工号等，使用“=”相连。比如：name=huaxiaowei，code=996181。

操作步骤

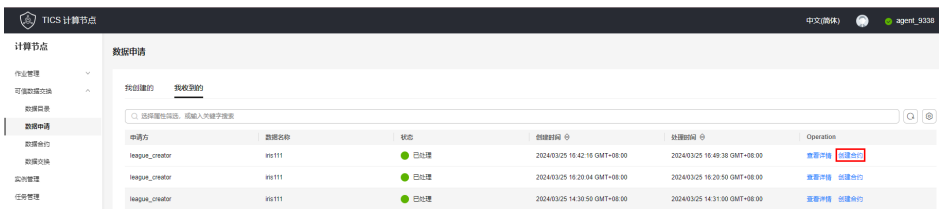
1. 创建合约。
   a. 用户登录进入计算节点页面。
   b. 在左侧导航树上选择“可信数据交换 > 数据申请”，打开数据申请页面。
   c. 在数据申请页面单击“我收到的”。
   d. 在“我收到的”数据申请页签中，选择已经确认的申请，单击“创建合约”。

图 3-27 创建合约



   e. 在创建合约对话框填写合约信息。
   数据合约的内容有五个部分，包括：
   1、合约内容：合约名称、合约描述。
   2、数据信息：主要描述结构化数据的列信息，包含数据名称、创建人、创建时间描述等信息。
   3、访问需求：主要描述数据用方的需求，包含访问截止时间、访问方式、访问次数。
   4、访问限制：暂不支持。
   5、自定义限制：自定义策略支持“<”、“>”和“=”。
   供数方可以设置自定义属性来进一步强化数据访问控制。

图 3-28 创建合约



f. 单击“保存并提交审批”。

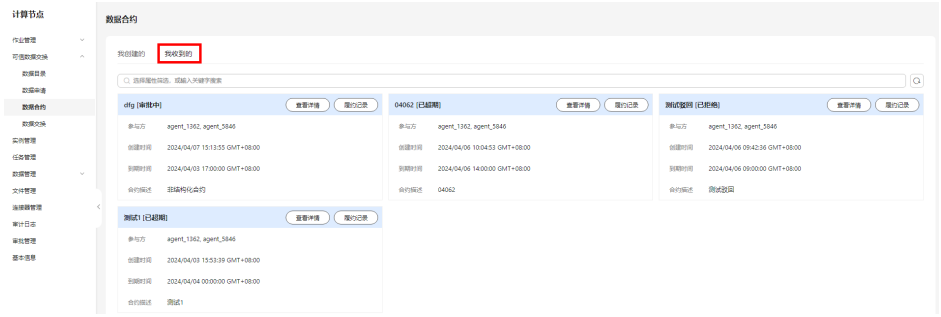
公司A编辑合约完成后，可以提交由公司B审批。在“可信数据交换 > 数据合约 > 我创建的”的页签可以看到合约及状态，并允许撤回再次编辑。

2. 审批合约。

a. 数据拥有方公司B在左侧导航树上选择“可信数据交换 > 数据合约”，打开数据合约页面。

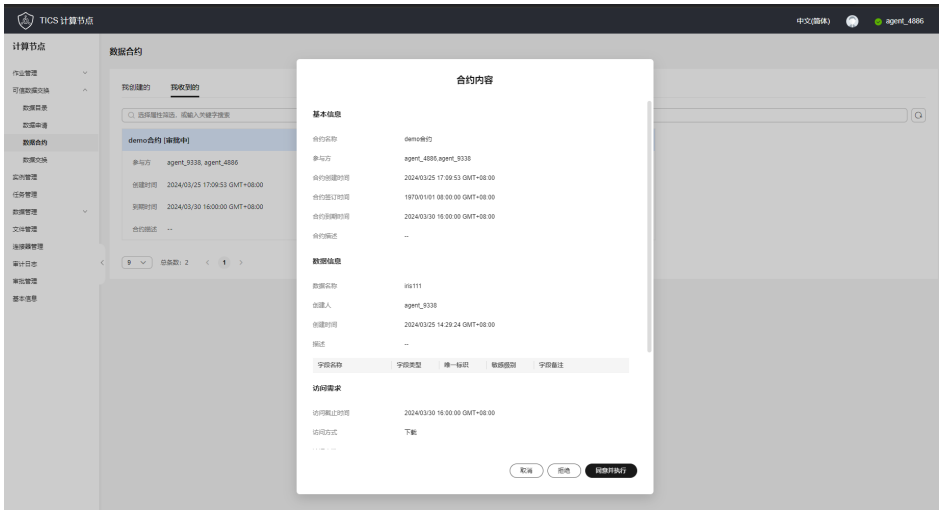
b. 在数据合约页面单击“我收到的”。可以看到数据需求方公司A发送来的数据合约。

图 3-29 我收到的数据合约



c. 单击“查看详情”可以查看合约的具体内容，检查无误后单击“同意并执行”可以执行合约进行数据交换。

图 3-30 查看合约内容



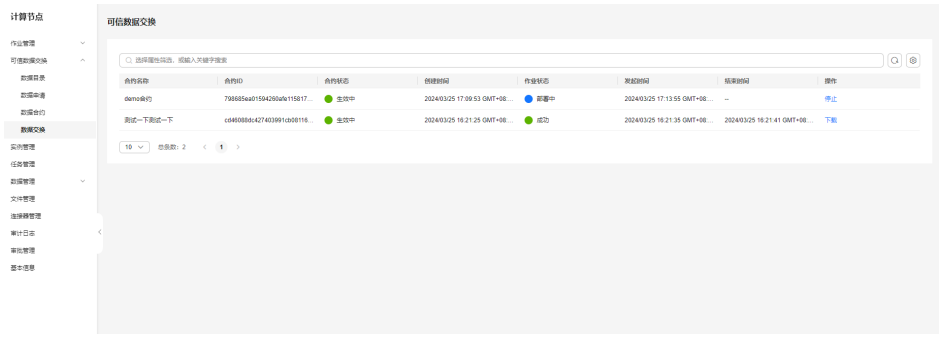
3. 数据交换。

a. B公司单击合约执行后，在左侧导航树上选择“可信数据交换 > 数据交换”，在可信数据交换页面，查看交换作业的执行情况。

b. 单击“下载”，即可使用数据。

数据交换作业执行完成后可以进行下载，将文件保存到本地。如有必要，下载时还需要填写公司A制定的自定义限制。

图 3-31 下载合约



3.6 横向联邦学习场景

TICS从UCI网站上获取了乳腺癌数据集Breast，进行横向联邦学习实验场景的功能介绍。

乳腺癌数据集：基于医学图像中提取的若干特征，判断癌症是良性还是恶性，数据来源于公开数据**Breast Cancer Wisconsin (Diagnostic)**。

3.6.1 场景描述

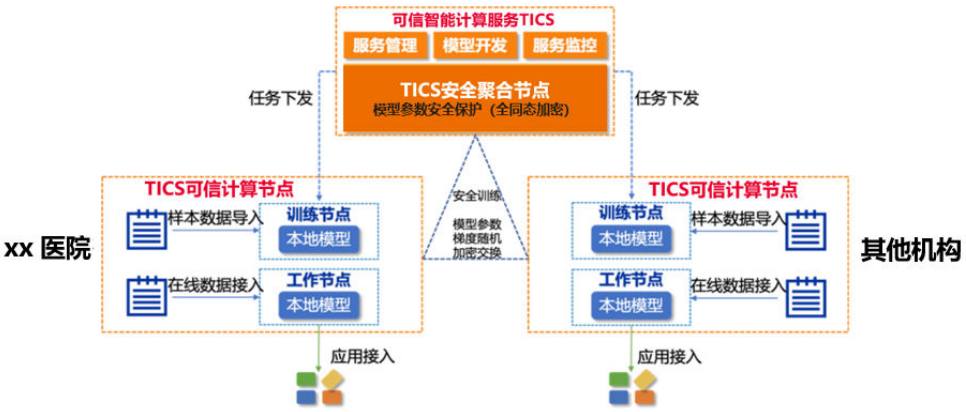
背景信息

本案例以“预测乳腺癌是良性/恶性”的场景为例。假设一部分的乳腺癌患者数据存储在xx医院，另一部分数据存储在某个其他机构，不同机构数据所包含的特征相同。

这种情况下，xx医院想申请使用其他机构的乳腺癌患者数据进行乳腺癌预测模型建模会非常困难。因此可以通过华为TICS可信智能计算平台的横向联邦功能，实现在患者隐私不泄露的前提下，利用其他机构的医疗数据提升乳腺癌预测模型的准确率。

进一步地，可根据该模型案例发散，构建老年人健康预测、高血压预测、失能早期预警模型等。

图 3-32 乳腺癌预测研究应用场景示意



1. 作业发起方通过计算节点上传数据、待训练模型的定义文件；
2. 作业发起方配置TICS的横向联邦学习作业，启动训练；

3. 模型参数、梯度数据在TICS提供的安全聚合节点中进行加密交换；
4. 训练过程中，各参与方计算节点会在本地生成子模型，由TICS负责安全聚合各子模型的参数，得到最终的模型；
5. 空间的整体配置通过空间管理员进行统一管理。

3.6.2 测试步骤

3.6.2.1 数据准备

乳腺癌数据集从UCI获取，该数据集只包含连续类型特征，因此对所有特征使用Scikit-Learn的StandardScaler进行了归一化。为了模拟横向联邦学习场景，将数据集随机划分为三个大小类似的部分：（1）xx医院的训练集；（2）其他机构的训练集；（3）独立的测试集，用于准确评估横向联邦学习得到的模型准确率。此外由于原始的数据集较小，采用了Imbalanced-Learn中的SMOTE算法，进行了数据集的扩充。下表为扩充过后的数据集统计信息。

乳腺癌数据集统计信息。

统计量	取值
特征数目	30
xx医院的训练样本数目	7366
其他机构的训练样本数目	7366
测试集样本数目	7257

操作步骤

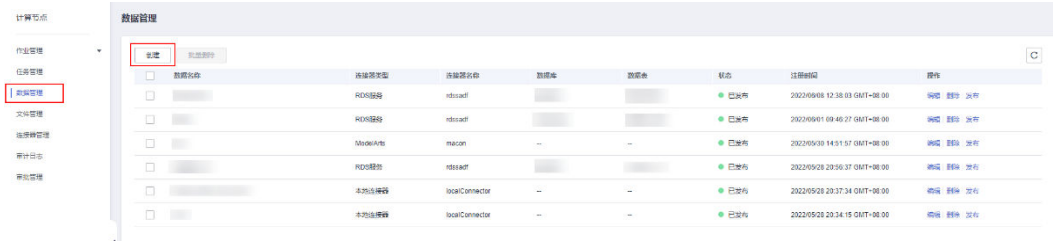
步骤1 进入TICS服务控制台。

步骤2 在计算节点管理中，找到购买的计算节点，通过登录地址，进入计算节点控制台。

图 3-33 前往计算节点

步骤3 登录到计算节点后，进入数据管理并进行数据集发布。

图 3-34 数据管理中新建数据集



步骤4 参考下图填写参数信息。（1）指定连接器为localConnector，选择数据文件的路径，填写数据名称；（2）字段配置中特征字段（x_{特征序号}）均配置为字段类型：FLOAT，字段类别：特征，特征类型：连续；标签字段（label）配置为字段类型：INTEGER，字段类别：标签。

图 3-35 配置数据集参数

步骤5 发布数据集。

图 3-36 发布数据集

说明

数据集发布的过程并不会直接从数据源中导出用户数据，仅从数据源处获取了数据集相关的元数据信息，用于任务的解析、验证等。

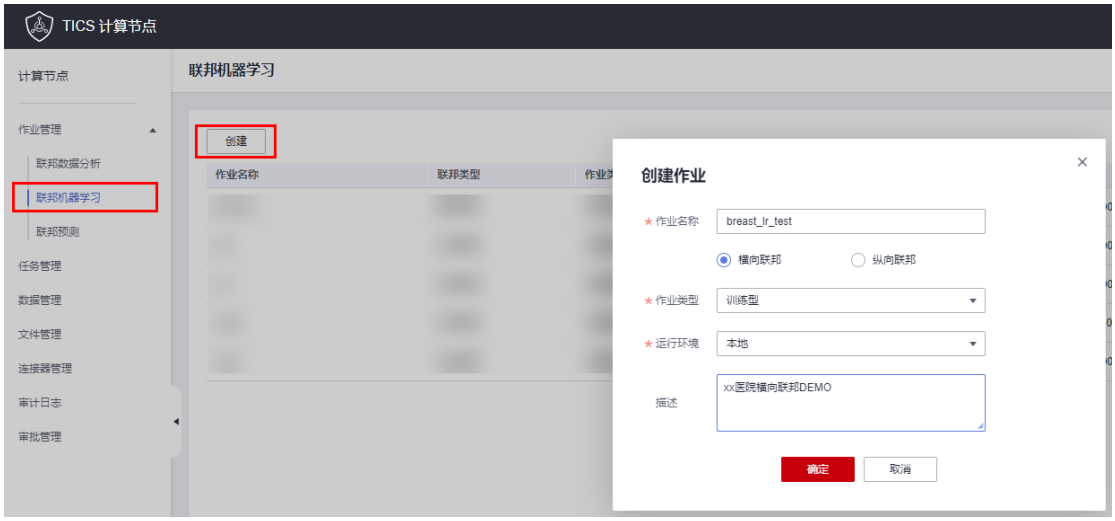
----结束

3.6.2.2 训练型横向联邦作业流程

联邦学习分为横向联邦及纵向联邦。相同行业间，特征一致，数据主体不同，采用横向联邦。不同行业间，数据主体一致，特征不同，采用纵向联邦。xx医院的应用场景为不同主体的相同特征建模，因此选用横向联邦。

步骤1 创建训练型横向联邦学习作业。

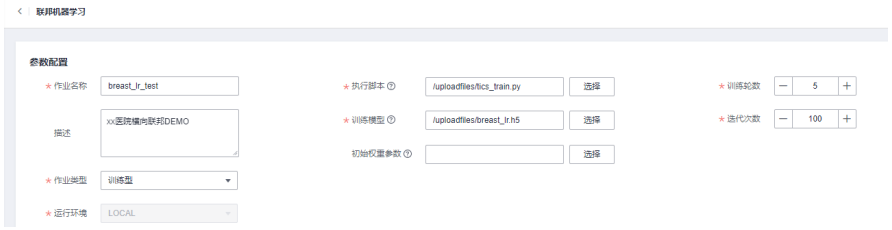
图 3-37 创建训练型横向联邦学习作业



步骤2 配置作业的执行脚本，训练模型文件。

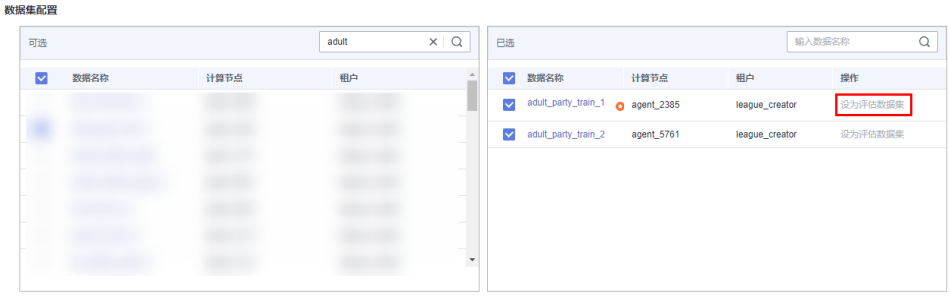
步骤3 执行脚本是每个参与方的计算节点在本地会执行的模型训练、评估程序，用于基于本地的数据集训练子模型。 训练模型文件则定义了模型的结构，会用于每个参与方在本地初始化模型。

图 3-38 配置执行脚本、训练模型文件



步骤4 配置已方、对方数据集。在作业的数据集配置中，选择己方、对方的本地数据集，此外需将已方的数据集设为评估数据集。横向联邦中，需要确保不同参与方的数据集结构完全一致。

图 3-39 配置数据集



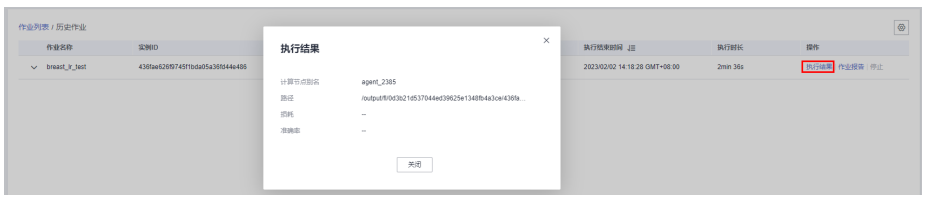
步骤5 保存并执行作业。单击下方的“保存并执行”按钮，即可发起执行横向联邦学习作业。

- 单击“历史作业”按钮，查看当前作业的执行情况。
- 单击“计算过程”按钮可以查看作业的具体执行计划。
- 单击“执行结果”按钮可以查看作业保存的模型文件路径，用于后续的评估型作业。

图 3-40 查看作业的执行情况

图 3-41 查看作业的具体执行计划

图 3-42 查看作业的执行结果



----结束

3.6.2.3 评估型横向联邦作业流程

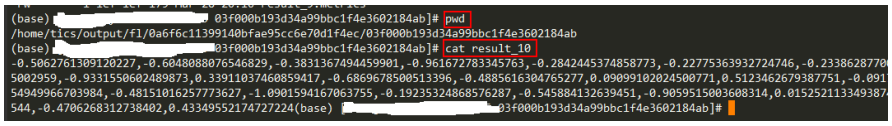
基于横向联邦作业的训练结果，可以进一步评估横向联邦模型，将训练好的模型用于预测。

步骤1 选择**对应训练型作业**的“历史作业”按钮，获取最新作业的模型结果文件路径。

图 3-43 查看模型结果文件的保存位置

步骤2 前往工作节点上步骤1展示的路径，下载模型文件。由于Logistic Regression模型本质上还是线性模型，因此模型文件result_10为该线性模型的系数加上偏置项。

图 3-44 查看模型结果文件



步骤3 本地利用测试集评估模型。可以采用如下脚本，会打印出模型在测试集上的准确率和AUC两个指标。

图 3-45 本地评估模型的 Python 脚本

```
import keras
import numpy as np
import pandas as pd
from sklearn.metrics import accuracy_score, roc_auc_score


if __name__ == "__main__":

    weight =
    weight = weight.split(",")
    weight = np.array([float(elem) for elem in weight])

    coefficient = weight[:30].reshape(-1, 1)
    bias = weight[30].reshape(-1)

    model = keras.models.load_model("breast_lr.h5")
    model.set_weights([coefficient, bias])

    test = pd.read_csv("breast_test.csv")
    X = test.values[:, :30]
    y = test.values[:, 30]

    y_pred = model.predict(X)
    y_pred_binary = (y_pred > 0.5).astype(int)

    print("Test Acc: {:.3f} %".format(100. * accuracy_score(y, y_pred_binary)))
    print("AUC: {:.3f}".format(roc_auc_score(y, y_pred)))
```

----结束

3.6.3 实验结果

3.6.3.1 乳腺癌数据集作业结果

本节实验包含了如下三个部分：（1）训练轮数对联邦学习模型分类性能的影响；（2）迭代次数对联邦学习模型分类性能的影响；（3）参与方数据量不同时，本地独立训练对比横向联邦的模型性能。

- 不同训练参数对模型准确率、训练时长的影响

  训练轮数对模型准确率的影响（迭代次数固定为20）

训练轮数	1	10	20
测试集准确率 (%)	98.016	98.016	98.016
测试集AUC	0.996	0.996	0.996
训练时长 (秒)	19	173	372

迭代轮数对模型准确率、训练时长的影响（训练轮数固定为10）

迭代次数	10	25	50
测试集准确率 (%)	97.065	98.140	98.415
测试集AUC	0.995	0.996	0.997
训练时长 (秒)	166	167	216

从上面两张表可以看出：

（1）训练轮数对于联邦学习模型的性能影响不大，这主要是由于乳腺癌数据集的分类相对简单，且数据集经过了扩充导致的；

（2）增大每个参与方本地模型训练的迭代次数，可以显著提升最终联邦学习模型的性能。

- 参与方数据量不同时，独立训练对比横向联邦训练的准确率

  本节实验不再将训练集均匀划分到两个参与方，而是以不同的比例进行划分，从而探究当参与方数据量不同时，模型性能的变化情况。具体划分如下所示。实验中训练轮数固定为10，迭代次数固定为50。

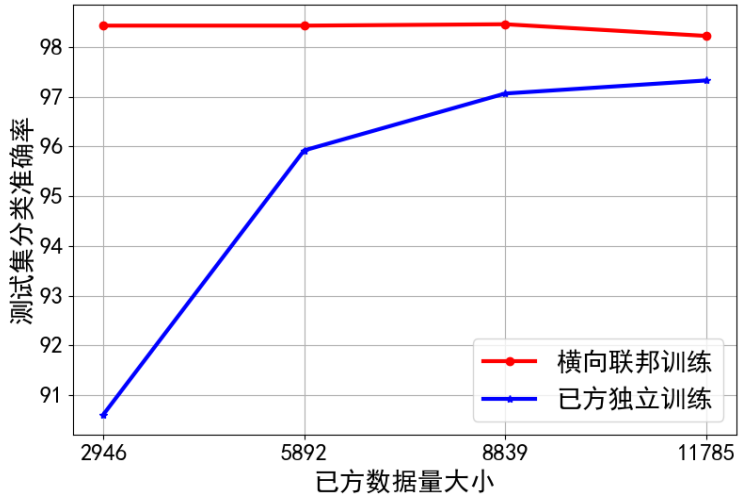
  参与方持有的样本数目信息

Host所持样本占比(%)	Host样本数	Guest样本数
0.2	2946	11786
0.4	5892	8840

Host所持样本占比(%)	Host样本数	Guest样本数
0.6	8839	5893
0.8	11785	2947

下图为当Host方拥有不同数据量时，使用横向联邦对比己方独立训练的性能对比。

图 3-46 Host 方拥有不同数据量时，横向联邦对比对立训练的模型性能



结论为：使用横向联邦学习，在己方拥有不同数据量的情况下都可以显著提升模型性能。