团体标准

发 布

中国电机工程学会

20XX—XX—XX实施

20XX—XX—XX发布

基于隐私计算的电力数据共享业务互联互通接口规范

（草案）

Interface Specification for Interconnection of Power Data Sharing Services Based on Privacy-Preserving Computation (Draft)

（初稿）

T/CSEE XXXX—YYYY

代替 T/XXXX

ICS 19.020

CCS K85

目 次

前 言 III

1 范围 4

2 规范性引用文件 4

3 术语和定义 4

4 缩略语 5

5 隐私计算基本要求 6

5.1 隐私计算应用场景 6

5.2 隐私计算参与方 6

5.3 隐私计算典型流程 7

6 互联互通概设 8

6.1 互联互通意义 8

6.2 互联互通框架 8

6.3 互联互通总体架构 8

7 互联互通对象 9

7.1 对象实体 9

7.2 对象关系 9

8 互联互通协议 10

8.1 节点互联协议 10

8.1.1 节点发现 10

8.1.2 节点管理 10

8.1.3 节点更新 11

8.2 项目互联协议 11

8.2.1 项目创建 11

8.2.2 项目查询 11

8.2.3 项目更新 11

8.2.4 项目删除 11

8.3 流程互联协议 12

8.3.1 流程创建 12

8.3.2 流程查询 12

8.3.3 流程审批 12

8.3.4 流程更新 12

8.3.5 流程删除 12

8.4 作业互联协议 12

8.4.1 作业创建 12

8.4.2 作业查询 13

8.4.3 作业启动 13

8.4.4 作业停止 13

8.4.5 作业重跑 13

8.5 任务互联协议 13

8.5.1 任务启动 13

8.5.2 任务查询 13

8.5.3 任务停止 14

8.5.4 任务回调 14

8.5.5 任务重跑 14

8.5.6 任务结果上报 14

8.6 组件互联协议 14

8.6.1 组件注册 14

8.6.2 组件查询 14

8.6.3 组件检查 15

8.6.4 组件注销 15

8.7 统一资源互联协议 15

8.7.1 数据集互联协议 15

8.7.2 模型互联协议 16

9 互联互通要求 16

9.1 通信要求 16

9.2 数据安全合规要求 16

9.3 性能要求 18

9.4 准确性要求 18

附 录 A 19

前 言

本文件按照《中国电机工程学会标准化管理办法》、《中国电机工程学会标准化管理办法实施细则》的要求，依据GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国电机工程学会提出。

本文件由中国电机工程学会电力信息化标准专业委员会技术归口和解释。

本文件起草单位：国家电网有限公司大数据中心、国网智能电网研究院有限公司、云南电网有限责任公司、中国大唐集团科学技术研究总院、中国电力建设集团有限公司、国网山东省电力公司、国网河北省电力有限公司、国网北京市电力公司、国网上海市电力公司、国网浙江省电力有限公司、国网安徽省电力有限公司。

本文件主要起草人：朱洪斌、刘圣龙、江伊雯、石聪聪、于鹏飞、张春林、杨国玉、朱晨、吕艳丽、姜嘉伟、安冰、高原、翟雨佳、周鑫、吴裔、严莉、康之增、王超、胡聪、毛冬、陈剑飞、于航、彭潇、李云昭、赵涛。

本文件首次发布。

本文件在执行过程中的意见或建议反馈至中国电机工程学会标准执行办公室（地址：北京市西城区白广路二条1 号，100761，网址：http：//www.csee.org.cn，邮箱：cseebz@csee.org.cn）。

基于隐私计算的电力数据共享业务互联互通接口规范

1. 范围

为了更好利用隐私计算技术实现电力数据的跨业务融合应用，破解“数据孤岛”困境，本文件规定了电力数据多方协同建模和分析预测时，与合作方隐私计算平台之间的接口规范，包括数据联邦学习建模和多方安全计算时的信息交互技术框架和接口要求，以满足与合作方不同技术架构下的隐私计算互联互通需求。

本文件适用于隐私计算电力数据共享业务中的互联互通开发、测试，以及电力数据多方融合应用业务建设时的隐私计算技术选型。

1. 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 25069-2010 ｜ 信息安全技术 术语

GB/T 35273-2020 ｜ 信息安全技术 个人信息安全规范

JR/T 0196—2020 ｜ 多方安全计算金融应用技术规范

BDC 79-2021 ｜ 隐私计算 跨平台互联互通 第1部分：总体框架

1. 术语和定义

下列术语和定义适用于本文件。

* 1.

隐私计算 privacy-preserving computation

在保证数据提供方不泄露原始数据的前提下，对数据进行分析计算的一类信息技术，保障数据在产生、存储、计算、应用、销毁等数据流转全过程的各个环节中“可用不可见”。

1. 术语条目注1：隐私计算常用技术方案有多方安全计算（Secure Multi-Party Computation）、联邦学习（Federated Learning）、可信执行环境（ Trusted Execution Environment）等。
2. 术语条目注2：隐私计算常用的底层技术有混淆电路（Garbled Circuit）、 不经意传输（Oblivious Transfer）、秘密分享（Secret Sharing）、同态加密（Homomorphic Encryption）等。

[来源：BDC 79-2021，3.1]

* 1.

联邦学习 federal learning

拥有本地数据的两方或多方开展机器学习联合建模，各方在建模过程中不交换本地数据，而是交换本地训练的局部参数或中间结果，从而协同构建基于各方数据的全局模型。

* 1.

隐私计算节点 privacy-preserving computation node

各隐私计算技术平台部署的互联互通网络基本组成单元，对外提供交互接口。

[来源：BDC 79-2021，3.5]

* 1.

隐私计算平台 privacy-preserving computation platform

一套完整的、严密的用于实现隐私计算功能的技术平台。

[来源：BDC 79-2021，3.3]

* 1.

参与方 parties

一个或一组自然人或法人，指参与隐私计算跨平台互联互通任务的主体，包含发起方、数据方、算法方、计算方、结果方、协调方等多种角色。

[来源：BDC 79-2021，3.2]

* 1.

互联互通 interconnection

不同系统架构、功能实现方案或算法原理的异构隐私计算平台之间，通过约定最小必要范围的互通协议、接口和实体等定义，能够跨平台地进行数据多方协同交互，共同完成同一隐私计算任务。

[来源：BDC 79-2021，3.4]

1. 缩略语

下列缩略语适用于本文件。

HTTP：超文本传输协议（Hyper Text Transfer Protocol）

HTTPS：超文本安全传输协议（Hyper Text Transfer Protocol over Secure socket layer）

RPC：远程过程调用（Remote Procedure Call）

TCP：传输控制协议（Transmission Control Protocol）

TLS：传输层安全协议（Transport Layer Security）

Web：万维网（World Wide Web）

XML：可扩展标记语言（Extensible Markup Language）

WOE：证据权重（Weight of Evidence）

IV：信息价值（Information Value）

* 1.

加密 encipherment/encryption

对数据进行密码变换以产生密文的过程。一般包含一个变换集合，该变换使用一套算法和一套输入参量。输入参量通常被称为密钥。

* 1.

算法 algorithm

用于完成隐私计算任务的一系列指令，能够对一定规范的输入，获得所要求的输出。

* 1.

算法组件 algorithm component

用于执行隐私计算任务的一种可代替、可组合的部件，封装了算法功能的实现并提供一系列可用的接口。

* 1.

互联互通网络 interconnection network

指不同隐私计算技术平台部署后相互连接，通过交互与协同连接形成的提供跨平台联合隐私计算服务的网络。

[来源：BDC 79-2021，3.4]

1. 隐私计算基本要求
	1. 隐私计算应用场景

隐私计算的应用可根据实际场景需求灵活进行扩展、组合或新增，主要分类如下：

* 1. 隐私集合求交：支持对两方或多方不同数据集进行安全求交，在不泄露交集外其他数据的前提下输出正确交集。
	2. 隐私信息检索：支持对两方或多方不同数据集进行依据某条件的数据检索，在保证查询方不会泄露查询目标，且不能从被查询方获得查询结果之外的其他信息的前提下，输出正确的检索结果。
	3. 联合基础运算：
		+ 支持对两方或多方不同数据集进行安全数值运算，如加、减、乘、除、数乘等；
		+ 支持对两方或多方不同数据集进行安全单元运算，如标量运算、矢量运算、矩阵运算、多维数组运算等；
		+ 支持对两方或多方不同数据集进行安全逻辑运算，如与运算、或运算、非运算、与非运算、或非运算、与或非、异或运算、同或运算等。
	4. 联合统计分析：支持对两方或多方不同数据集进行各种统计分析，如求和、平方和、均值、方差、分位数等计算。
	5. 联邦学习：
		+ 联邦特征工程：支持对两方或多方不同数据集进行特征工程，如在保护标签信息不被泄露的前提下，进行WOE和IV值计算等；
		+ 联邦建模：支持对两方或多方不同数据集的特征和标签进行模型训练，如回归算法、决策树算法、聚类算法等；
		+ 联邦预测：支持通过两方或多方不同数据集，使用已有模型对具有指定特征的数据进行预测，并输出正确的预测结果；
		+ 联邦模型评估：支持对已有隐私计算模型进行模型评估，并输出常见的模型评估指标。
	6. 隐私计算参与方

在一个完整的隐私计算中，包含的逻辑角色包括：

* 1. 任务发起方：任务发起方是发起隐私计算任务的角色，并在任务执行前核实各方资源。任务发起方可兼任计算方、任务调度方、结果使用方等角色；
	2. 任务调度方：任务调度方是分发隐私计算任务，分配控制各方资源及协调各方行为的角色，通过协调数据提供方、计算方、算法提供方等角色来完成隐私计算任务，宜获得可信第三方的认证。任务调度方可由任务发起方兼任；
	3. 算法提供方：算法提供方是根据隐私计算的任务需求，提供隐私计算算法组件和算法描述的角色。算法组件包括计算逻辑、算法实现和算法参数，算法描述包括算法的版本号、接口说明、安全性说明、兼容性说明等；
	4. 数据提供方：数据提供方是提供执行隐私计算任务所需的数据的角色，为联合建模、分析预测任务提供模型训练数据和分析预测数据；
	5. 计算方：计算方是提供执行隐私计算任务所需算力的角色。计算方自身算力不足时，可能会设立辅助计算方的角色进行辅助计算。计算方可与其他计算方执行协同计算，并确保任务执行过程不会造成隐私数据泄露；
	6. 结果使用方：结果使用方是最终获得隐私计算任务计算结果的角色。只有结果使用方才能获取计算的最终结果，其它非结果使用方不能获得结果，也不能通过中间计算过程，计算或推断出最终结果。
	7. 隐私计算典型流程

隐私计算典型工作流程如图1所示，各阶段工作内容为：

* 1. 任务创建：任务发起方利用调度方发起隐私计算任务；
	2. 任务分配：调度方将任务下发到每个数据提供方的计算节点，当某个节点有辅助计算节点时，在任务运行过程中会将某个计算逻辑下发到辅助计算节点；
	3. 数据输入：数据提供方的计算节点获取本地数据并进行本地计算；
	4. 任务计算：数据提供方的计算节点进行交互，完成隐私计算任务；如有辅助计算方，辅助计算节点同时参与隐私计算计算任务；
	5. 结果分析：结果使用方从各个计算节点获取结果数据，合并成最终结果。



**图 1** 隐私计算工作流程图

1. 互联互通概设
	1. 互联互通意义

不同隐私计算平台在系统架构、实现方案、算法原理等方面存在差异，无法直接建立连接并共同执行隐私计算任务，因此必须提出异构隐私计算平台之间的互联互通方案。通过约定最小必要范围的互通协议、标准化的接口定义以及明确的实体交互规则，异构隐私计算平台能够实现跨平台的数据多方协同交互，共同完成同一隐私计算任务，进而有效支持多方数据融合共享。

实现与合作方异构隐私计算平台间的互联互通协作，才能彻底的打破“数据孤岛”，保障电力数据安全多方共享协同，促进电力行业数据要素流通和价值释放。

* 1. 互联互通框架

通过建立协议接口规范，实现异构隐私计算平台之间的兼容与互联互通。如图2所示，电力应用隐私计算互联互通包含三个层次。



**图2** 电力隐私计算互联互通框架

1. 节点的互联互通，参与隐私计算任务的不同节点可以相互发现、合作授权、建立连接，是跨平台互联互通的最基础环节；
2. 资源的互联互通，已建立连接的节点间可以相互发现、申请、使用对方的数据资源、算法资源、模型资源等，在节点互联的基础上实现资源互通；
3. 算法组件的互联互通，在节点、资源互通的基础上，通过调用相同的算法组件或按照已约定的交互协议执行算法，实现具体计算任务的协同。
	1. 互联互通总体架构

电力隐私计算互联互通总体架构包含以下部分：

1. 对外接口：对节点互联、统一资源和算法组件互联协议中的核心流程进行接口设计和功能实现，供上层业务场景调用。
2. 互联协议：互联协议覆盖节点、项目、流程、作业、任务、资源和算法组件维度的互联协议栈。
3. 通信协议：隐私计算互联互通以统一的平台间通信要求为基础，包括通信接口、通信框架、数据格式、加密传输机制、网络环境等。



**图3** 隐私计算互联互通总体架构

1. 互联互通对象
	1. 对象实体

电力隐私计算互联互通最小必要的实体元素如下，参与互联互通的各方需要对齐这些实体元素的定义和数据结构，对与节点互通无关的个性化属性不做限制。

1. 节点（Node）：隐私计算生态中的抽象功能单元，用来指代由机构或组织部署的隐私计算平台；
2. 数据集（Dataset）：可供参与隐私计算的数据资源；
3. 项目（Project）：面向特定目标的，提供一项独特产品、服务或成果的隐私计算方案，该实体为可选项；
4. 组件（Component）：独立执行隐私计算任务的模块单元，其经过封装、符合开放接口规范、可以完成某个特定计算或算法，可独立部署；
5. 流程（Flow）：采用DAG结构定义的、可编排的隐私计算作业运行模板；
6. 作业（Job）：一个隐私计算流程通过配置运行参数后的运行实例；
7. 任务（Task）：组件运行实例的载体，每个运行实例通过任务来管理；
8. 模型（Model）：指通过隐私计算技术训练完成，可用于进一步推理的模型，或是已有的现成模型。
	1. 对象关系

参与互联互通合作的异构平台作为节点（Node）加入到互联互通网络，一个节点可以创建多个项目（Project），每个项目中可能包含多个参与该互联互通项目的多个节点。通过项目实体来统一管理项目中相关的节点、数据集（Dataset）和模型（Model）等资源，项目中多个合作的节点可以创建多个流程（Flow）流程运行多次生成多个作业（Job），不同作业可以按照“Conf通用化设计”配置不同的运行时参数，一个作业可能由多个任务（Task）组成，任务承接作业中对应的运行时参数配置，每个任务通过调度底层组件（Component）对应的算法资源进而执行任务，生成模型、报告等隐私计算结果。每个组件都是独立的算法模块，例如：PSI、LR等，组件可以按照“DAG通用化设计”提前配置调度顺序、输入输出关系等，从而编排形成一个具有完整计算逻辑的流程，作业和任务按照DAG配置和Conf配置调度和运行相应资源，生成对应结果。结合7.1实体定义，实体间的关系图如图4所示。



**图4** 隐私计算互联互通组最小必要对象

1. 互联互通协议
	1. 节点互联协议

节点互联协议是对异构隐私计算平台间的节点发现、节点管理和节点更新等互联操作内容与流程进行规范。节点互联协议是隐私计算跨平台互联互通的基础。

* + 1. 节点发现

节点发现约定节点基本信息查询、合作节点查询和节点合约查询的技术要求。

1. 节点基本信息查询：根据节点ID，查询该节点基本信息；
2. 合作节点查询：发起方查询合作方节点信息，无需鉴权；
3. 节点合约查询：发起方向合作方查询合约状态，确认回调失败的兜底操作。
	* 1. 节点管理

节点管理约定了节点签约、节点合约确认、节点签约取消、节点健康检查、节点解约的技术要求。

1. 节点签约：根据节点ID、节点名称、节点描述、节点认证方式、签约有效期、对外服务地址等有效信息，发起方向合作方申请签约，且签约结果需等待对方通知，节点双方签约成功会生成有效的合约ID；
2. 节点合约确认：在合约有效期内，合作方通过认证即可向发起方确认是否接受本次签约；
3. 节点签约取消：发起方取消指定合约ID的签约申请，取消合约为待处理状态才有效；
4. 节点健康检查：检查节点的状态，如发起方取消签约申请时，待处理状态才有效；
5. 节点解约：合作节点撤销合约关系，双方都可发起。
	* 1. 节点更新

节点更新约定了节点信息更新、节点信息同步的技术要求。

1. 节点信息更新：更新当前节点信息；
2. 节点信息同步：在合约有效期内，通过合作节点对外提供服务的地址，将节点更新内容同步到签约方。
	1. 项目互联协议

项目互联协议是对异构隐私计算平台间的项目创建、项目查询、项目更新以及项目删除等互联操作内容与流程进行规范。项目互联协议是隐私计算跨平台互联互通的应用基础协议。

* + 1. 项目创建

项目创建约定了自有项目创建和合作项目创建的技术要求。

1. 自有项目创建：发起方创建一个项目，支持设置项目名称、项目描述、和项目类型等基本初始参数；
2. 合作项目创建：发起方向合作方发起项目创建请求，用于数据同步、审核等场景。
	* 1. 项目查询

项目查询约定了自有项目查询、合作项目查询、自有项目信息查询、合作项目信息查询的技术要求。

1. 自有项目查询：查询发起方的自有项目，查询方式支持模糊查询和精确查询，查询结果支持列表展示和统计说明；
2. 合作项目查询：查询发起方可以查询指定合作方的全部合作项目，查询结果支持列表展示和统计说明；
3. 自有项目信息查询：根据项目ID，查询单个自有项目详情；
4. 合作项目信息查询：根据项目ID，查询单个合作项目详情。
	* 1. 项目更新

项目更新约定了自有项目更新、合作项目更新的技术要求。

1. 自有项目更新：根据项目ID，发起方更新单个自有项目；
2. 合作项目更新：根据项目ID，发起方向合作方发起项目更新请求，用于数据同步、审核等场景。
	* 1. 项目删除

项目删除约定了自有项目删除的技术要求。

1. 自有项目删除：根据项目ID，发起方删除单个自有项目。
	1. 流程互联协议

流程互联协议是对异构隐私计算平台间合作项目的流程创建、流程查询、流程审批、流程更新以及流程删除等互联操作内容与流程进行规范。流程互联协议是隐私计算跨平台互联互通的应用基础协议。

* + 1. 流程创建

发起方在指定项目ID下创建流程，支持设置流程名称、添加流程描述、DAG和运行时任务参数和配置信息，流程创建成功后可以根据推荐雪花算法生成或以节点ID为前缀生成对应的流程ID，流程ID支持跨节点共享。

其中，DAG是用来描述流程中组件依赖关系的语言，包括描述流程使用的组件列表的components字段、输入数据列表的input字段和输出数据列表的output字段，一旦流程确定则DAG内容不变，Config是运行时任务参数和配置信息。

* + 1. 流程查询

流程查询约定了流程列表查询、流程信息查询的技术要求。

1. 流程列表查询：根据项目ID，查询该项目下的所有流程，结果应支持列表显示和查询总数统计说明；
2. 流程信息查询：根据流程ID，查询流程信息，结果应支持列表显示。
	* 1. 流程审批

流程审批约定了流程审批发起、流程审批确认、流程审批查询的技术要求。

1. 流程审批发起：通过项目ID、流程ID、流程描述、DAG等参数，触发合作节点对指定ID的流程内容进行审批，并生成特定的审批ID；
2. 流程审批确认：合作方向发起方确认审批结果，并返回该审批ID及其审批状态；
3. 流程审批查询：发起方通过审批ID向合作方查询流程审批状态，流程审批查询操作是确认回调失败的兜底操作。
	* 1. 流程更新

发起方对指定流程进行更新，支持对流程名称、流程描述、DAG以及作业运行配置进行更新，且流程更新操作不影响已经在运行的Job。

* + 1. 流程删除

发起方根据指定流程ID，删除对应流程。

* 1. 作业互联协议

作业互联协议是对异构隐私计算平台间合作项目流程的作业创建、作业查询、作业启动、作业停止以及作业重跑等互联操作内容与流程进行规范。作业互联协议是隐私计算跨平台互联互通的应用基础协议。

* + 1. 作业创建

作业创建约定了自有作业创建、合作作业创建的技术要求。

1. 自有作业创建：基于流程创建新的作业，支持设置DAG和作业运行时配置信息，如作业优先级、作业超时秒数等，作业创建成功后生成作业ID，且该作业ID在整个流程运行期间不变；
2. 合作作业创建：根据作业ID，创建合作方作业。
	* 1. 作业查询

作业查询约定了作业列表查询、作业信息查询的技术要求。

1. 作业列表查询：根据指定流程ID，查询所属该流程下的全部作业，结果支持列表显示和记录总数统计说明；
2. 作业信息查询：根据指定作业ID，查询作业详情信息。
	* 1. 作业启动

作业启动约定了自有作业创建、合作作业启动的技术要求。

1. 自有作业创建：根据作业ID，启动作业，并开始调度其相关任务；
2. 合作作业启动：根据作业ID，启动合作方作业。
	* 1. 作业停止

作业停止约定了自有作业停止、合作作业停止的技术要求。

1. 自有作业停止：根据作业ID，停止作业；
2. 合作作业停止：根据作业ID，停止合作方作业。
	* 1. 作业重跑

重跑指定ID作业，重跑作业过程中支持对DAG中的组件执行新的运行配置。

* 1. 任务互联协议

任务互联协议是对异构隐私计算平台间合作项目的任务启动、任务查询、任务停止、任务回调、任务重跑以及任务结果上报等互联操作内容与流程进行规范。任务互联协议是隐私计算跨平台互联互通的应用基础协议。

* + 1. 任务启动

任务启动约定了自有任务启动、合作任务启动的技术要求。

1. 自有任务启动：根据指定作业ID，启动新的组件任务，支持对DAG中的组件名称、新的任务参数、组件任务运行参数、任务优先级进行配置，任务启动成功后生成对应的任务ID；
2. 合作任务启动：根据指定作业ID，启动新的组件任务，支持对DAG中的组件名称、新的任务参数进行配置，任务ID支持由发起方指定，合作方直接使用，若任务ID由发起方指定，则必须和入参 task\_id 相同。
	* 1. 任务查询

任务查询约定了任务列表查询、任务信息查询、合作任务信息查询、任务日志查询的技术要求。

1. 任务列表查询：根据指定作业ID，查询所属该作业下所有的组件任务列表，结果支持列表显示和记录总数统计说明；
2. 任务信息查询：根据指定作业ID和任务ID，查询对应任务详情；
3. 合作任务信息查询：根据指定作业ID和任务ID，查询任务详情；
4. 任务日志查询：根据指定作业ID和任务ID，查询对应任务日志，结果显示应支持按时间排序，且日志查询结果应包括日志级别、日志内容等，显示的内容不应超过1M。
	* 1. 任务停止

任务停止约定了自有任务停止、合作任务停止的技术要求。

1. 自有任务停止：根据指定作业ID和任务ID，停止对应组件任务；
2. 合作任务停止：根据指定作业ID和任务ID，停止对应的组件任务。
	* 1. 任务回调

任务回调约定了自有任务回调、合作任务回调的技术要求。

1. 自有任务回调：根据指定作业ID和任务ID，回调当前节点组件任务；
2. 合作任务回调：根据指定作业ID和任务ID，回调相关组件任务。
	* 1. 任务重跑

任务重跑约定了自有任务重跑、合作任务重跑的技术要求。

1. 自有任务重跑：根据指定作业ID和任务ID，重跑组件任务，重跑时运行新的任务参数且不带起下游；
2. 合作任务重跑：根据指定作业ID和任务ID，重跑组件任务，重跑时运行新的任务参数且不带起下游。
	* 1. 任务结果上报

根据指定作业ID和任务ID，上报组件任务运行结果，组件任务运行完成上报运算结果，上报内容应包括任务输出类型、任务输出内容。

* 1. 组件互联协议

组件互联协议是对异构隐私计算平台间的节点发现、节点管理和节点更新等互联操作内容与流程进行规范。组件互联协议是隐私计算跨平台互联互通的应用基础协议。

* + 1. 组件注册

注册组件相关信息。

* + 1. 组件查询

组件查询约定了对外组件列表查询、对外组件信息查询、组件列表查询、组件信息查询的技术要求。

1. 对外组件列表查询：查询对外可见的相关组件，结果应包括组件编码、组件版本、组件引擎等组件概要说明，且结果支持列表显示和统计说明；
2. 对外组件信息查询：查询对外可见的相关组件信息，根据查询的组件编码及其版本描述返回对应的组件名称、组件地址、组件地址类型、组件入参、组件出参以及组件开发者等相关组件信息；
3. 组件列表查询：组件列表查询，结果应包括组件编码、组件版本、组件引擎等组件概要说明，且结果支持列表显示和统计说明；
4. 组件信息查询：组件信息查询，根据查询的组件编码及其版本描述返回对应的组件名称、组件地址、组件地址类型、组件入参、组件出参以及组件开发者等相关组件信息。
	* 1. 组件检查

组件检查约定了组件参数验证、组件健康检查的技术要求。

1. 组件参数验证：根据组件编码和组件版本信息，验证组件参数是否合法；
2. 组件健康检查：根据组件编码和组件版本信息，检查组件服务健康状态。
	* 1. 组件注销

根据组件编码及其版本描述，注销对应组件相关信息。

* 1. 统一资源互联协议
		1. 数据集互联协议
			1. 自有数据集创建

在当前节点创建自有数据集，并对该数据集添加数据集来源、数据集权限等信息进行设置，自有数据集创建完成后生成对应的数据集ID，且该数据集ID在节点内唯一。

* + - 1. 数据集查询

数据集查询约定了公开数据集列表查询、自有数据集列表查询、授权数据集列表查询、合作数据集信息查询、自有数据集信息查询、已授权数据集列表查询的技术要求。

1. 公开数据集列表查询：发起方向合作方查询公开的数据集列表，结果支持列表显示和记录总数统计说明；
2. 自有数据集列表查询：根据项目ID，查询当前节点自有数据集，结果支持列表显示和记录总数统计说明；
3. 授权数据集列表查询：发起方向合作方查询已授权本方的（不含公开的）数据集列表，结果支持列表显示和记录总数统计说明；
4. 合作数据集信息查询：根据数据集ID，发起方向合作方查询公开的或已授权本方的数据集信息；
5. 自有数据集信息查询：根据数据集ID，查询当前节点单个数据集详情；
6. 已授权数据集列表查询：遵循授权模式的规定，查询当前节点授权外部或被授权的数据集列表，结果支持列表显示和记录总数统计说明。
	* + 1. 数据集授权

数据集授权约定了数据集授权申请、数据集授权申请取消、数据集主动授权、数据集授权确认、数据集授权查询的技术要求。

1. 数据集授权申请：发起方向合作方发起数据集授权申请，授权申请包括申请授权的数据集ID、授权作用域类型以及权限有效期等，申请发起后会生成对应的申请ID，结果需等待对方通知；
2. 数据集授权申请取消：根据数据集授权申请ID，发起方向合作方取消数据集授权申请，待处理状态才有效；
3. 数据集主动授权：主动授予合作节点数据集权限，授权信息包括申请授权的数据集ID、授权作用域类型以及权限有效期等；
4. 数据集授权确认：根据数据集授权申请ID，合作方向发起方确认是否接受本次授权；
5. 数据集授权查询：根据数据集授权申请ID，发起方向合作方查询申请状态，确认回调失败的兜底操作。
	* + 1. 数据集权限撤销

合作方可以主动撤销授权通过的数据集权限。

* + 1. 模型互联协议
			1. 模型查询

模型查询约定了模型列表查询、模型信息查询的技术要求。

1. 模型列表查询：根据项目ID，查询当前节点模型，结果支持列表显示和统计说明；
2. 模型信息查询：根据模型ID，查询当前节点单个模型详情，包括模型名称、模型所属任务、模型文件位置、模型创建时间、模型算法等。
	* + 1. 模型审批

模型审批约定了合作模型审批、合作模型审批确认、合作模型审批查询的技术要求。

1. 合作模型审批：根据模型文件的摘要值和模型所属任务、所属作业、所属项目，由合作节点对模型（碎片）进行审批，并生成对应的审批ID；
2. 合作模型审批确认：根据审批ID和审批状态，由合作节向发起方确认审批结果；
3. 合作模型审批查询：根据审批ID，发起方向合作方查询审批结果，确定该模型的审批状态。
4. 互联互通要求
	1. 通信要求

电力隐私计算互联互通应满足通信要求包括：

1. 互联互通场景中，各隐私计算节点应通过协商机制确定统一的通信规范，包括通信框架、数据格式、安全机制、网络环境等；
2. 各隐私计算节点通信接口应采用国家密码管理主管部门认证核准的密码技术保证通信过程中数据的机密性、完整性和不可否认性，以确保传送或接收的通信数据不发生泄漏、篡改、抵赖等情况。
3. 应提供多种标准通讯协议支持节点间进行数据交互，如：HTTP(s)、GRPC(s)；
4. 各隐私计算节点间应通过协商确定网络传输环境，运行于公网、内网、运营商专线等物理网络环境之中，不宜采用无线网络连接，如确需采用应确保无线网络的安全性；
5. 建议使用一致的IPV4、IPV6网络，如果IP格式不一致，应具有格式相互转换功能；
6. 对于网络环境整体安全性和稳定性有更高要求的，建议使用专线网络。
	1. 数据安全合规要求

电力隐私计算互联互通应满足数据安全合规要求包括：

1. 节点参与互联互通前应经过双向身份认证，以确保连接节点身份有效性和合法性：
2. 各隐私计算节点的身份标识应具有唯一性，通过标识符可唯一确定连接节点身份；
3. 身份认证采用的密码技术应由国家密码管理主管部门认证核准；
4. 节点身份认证的内容应包括IP地址、端口、角色等；
5. 认证证书应满足国家密码管理部门的相关要求，同时符合GB/T 20518-2006、GB/T 27928.1等标准文件要求；
6. 宜具备身份认证失败的处理能力，当身份认证应答超过规定时限时，平台宜终止与待接入节点之间的当前会话；当经过一定次数的认证失败后，平台宜终止由该节点发起的建立会话的尝试，并在一定的安全时间间隔后恢复。
7. 节点间应协商建立统一的资源分层授权访问控制机制，防止资源越权访问：
8. 应保证权限最小化原则，满足资源使用的最低要求即可；
9. 各方应约定资源使用期限、使用范围、使用数量等细节，超出授权范围则资源无法被使用；
10. 资源所有者应能够对已授权的资源取消授权，取消授权后资源无法被外部使用。
11. 访问控制范围应覆盖各平台涉及的节点、资源和操作等，并能够根据业务需要对访问权限进行配置。
12. 节点提供的数据集、模型、算法等资源应提供安全性证明，资源被加载进入异构的隐私计算平台前应进行安全性认证和检测：
13. 节点提供的算法资源应确保输入数据明文不出可信域、电力明细数据原则上不出数据中台，且各方明文中间结果被安全保护；
14. 节点提供的算法资源应满足在半诚实安全模型下不会泄漏敏感信息，支持计算安全强度不低于112比特，统计安全强度不低于30比特；
15. 各节点可根据实际需求协商采取不同的资源安全性认证措施，例如：摘要值比对、漏洞扫描、数字签名等。
16. 各平台应具备对关键数据和关键行为进行记录的能力，以满足内外部监管审计以及任务责任归属追溯的需要：
17. 存证内容应支持依据监管方及业务的具体需求进行定制化，存证方式包括但不限于日志、区块链等；
18. 应对存证内容进行访问控制，防止非授权用户对存证内容进行查看、删除，存证内容应满足不可篡改性；
19. 应在各参与方约定的期限内对存证内容进行销毁，根据法律法规要求约定的存储时限应不少于6个月；应采用加密技术确保存证的数据不能暴露或推断出非授权的个人隐私数据；
20. 应对存证内容进行完整性保护和定时备份，当存证内容存储已满、失败和受到攻击时，应确保存证内容不被篡改、破坏；
21. 为保证后续对任务执行正确性和计算结果准确性进行深入审计，宜具备通过授权后恢复或部分恢复存证内容的能力。
22. 各隐私计算节点及平台应具备存储数据的完整性、机密性和可用性保护机制：
23. 各平台应采用国家密码管理主管部门认证核准的密码技术保障平台中存储数据的完整性，存储的重要数据应不能被篡改。
24. 各平台应采用国家密码管理主管部门认证核准的密码技术对平台中存储的数据进行加密存储，重要数据应以密文方式存储。
25. 各平台应采取实时备份方式保证重要数据的可用性，当数据受到破坏时应能及时恢复。
26. 各平台应采取冗余部署模式，保证单个节点发生故障时系统业务维持稳定。
27. 各平台发生逻辑错误时，应能够安全的从错误中恢复，并保证数据的一致性。
28. 隐私计算互联互通全过程中使用的密码技术包括加密、签名、哈希、密码交换技术等应在国密或国际标准密码管理相关规定范围内协商一致，优选使用由国家密码管理主管部门认证核准的国密算法加强自主可控，且至少满足安全强度的要求，包括：
29. 对称密码：SM4（128bits）、AES（128bits）；
30. 非对称密码：SM2（256bits）、ECDSA（256bits）、RSA-2048（112bits）；
31. 哈希函数：SM3（128bits）、SHA3-256（128bits）、SHA3-512（256bits）、SHA256（128bits）；
32. 当对密码技术有更高要求时，各方可以共同协商使用安全性比以上更高的加密算法
33. 互联互通应采用安全的通信信道，确保通信过程中的数据完整性、真实性，保证通信过程中敏感信息字段或报文的机密性，保证参与通信的各实体行为的不可否认性，建议采用SSL3.2/TLS1.2以上的安全通信机制。
34. 电力隐私计算平台应支持内外网穿透建模能力，确保外网隐私计算平台和内网隐私计算平台通过模型中间参数跨隔离交互，共同完成隐私计算建模。
	1. 性能要求

参与电力隐私计算互联互通的隐私计算平台应满足性能要求包括：

1. 互联互通对隐私计算平台通信效率和计算效率产生的影响应在业务需求可接受的范围内；
2. 支持完成海量数据的处理、计算，如:十亿级数据的安全融合，百亿级数据的安全查询，亿级数据的安全计算，万级数据、百级特征的联合建模，亿级数据的模型在线推理等；
3. 宜支持多个维度的效率优化，包括软硬件升级、专用计算芯片、算法优化等。
	1. 准确性要求

参与电力隐私计算互联互通的隐私计算平台应满足准确性要求包括：

1. 在联合基础运算、联合统计分析场景中，明文本地计算结果与隐私计算结果之间的相对误差应保持在合理范围内，计算精度≥32。(该项指标取值代表计算结果与相同明文数据计算结果连续相同的有效位数（以二进制表示）；
2. 在隐私集合求交、隐私信息检索场景中，得到的结果应与明文本地计算结果保持一致；
3. 在联邦机器学习场景中，使用相同的数据集样本、输入特征和训练参数进行训练，隐私计算得到的模型应与在明文数据集上训练得到的模型在新数据上的预测结果保持一致。

附 录 A

（资料性）

隐私计算电力应用实践用例

* 1. 基于联邦学习的重点污染企业排放监测
		+ 1. 业务背景

环境问题关乎国计民生。为响应生态环境部门关于重点污染企业排放监测中对电力数据的需求，将电力数据赋能重点污染企业排放监测工作，同时为保证生态环境部门重点污染企业负面数据不外泄，以及企业用电数据不外泄。通过隐私计算技术实现“重点污染企业排放监测”场景。利用电力数据和生态环境数据，能够对重点污染企业的生产排污情况进行监测，实现企业污染排放全场景在线监测，及时发现企业生产过程中的违规生产的现象，提升污染防治精准管控水平。

* + - 1. 数据准备

电网企业和生态环境部门各自拥有重点污染企业的关键企业信息、用电数据、企业限停产政策数据，以及重污染天气气象等级等数据，从联合建模分析的角度出发，首先构建融合电力和环境的多维数据特征。

**表1** 数据特征维度

|  |  |
| --- | --- |
| **数据种类** | **数据特征** |
| 电力数据 | 企业名称 |
| 统一社会信用代码 |
| 表号 |
| 企业日用电量 |
| 合同容量 |
| 企业用电量级别 |
| 企业历史典型用电量 |
| 是否为高耗电企业（布尔） |
| 政府生态环境部门数据 | 重点污染企业名称 |
| 统一社会信用代码 |
| 停限产政策 |
| 停限产时间 |
| 重污染天气气象等级 |
| 有无响应重污染天气停限产政策 |

* + - 1. 模型设计

（1）数据虚拟融合

建模中涉及的相关方在数据上不但由参与方各自持有，还呈现出多尺度，多模态等特性。结合上述特点，模型中采用了数据源的虚拟融合和特征工程两个环节实现了安全联邦建模前的数据准备，如图5所示。前者解决了任务合作方数据安全，实现了任务发起方对数据持有方的数据申请授权以及对齐；后者解决了复杂数据的特征统计学习分析。



**图5** 联合建模数据准备工作流程

（2）数据特征工程

特征工程环节中，采用了数据挖掘和统计学习方法。重点污染企业违规排放建模中使用了9项处理流程：特征筛选、样本抽样、缺失值处理、异常值处理、标准化归一化处理、数据集划分、变量分项、WOE转化、特征统计。下表中阐述了特征工程环节中的9项操作。

**表2** 数据特征工程流程

|  |  |  |
| --- | --- | --- |
| **步骤** | **实施流程** | **描述** |
| 1 | 特征筛选 | 根据数据特点选择缺失值比例筛选、方差筛选、变异系数筛选、iv值筛选、逐步回归筛选、手动筛选等，保证特征有效性。 |
| 2 | 样本抽样 | 对样本中的类别不均衡问题实现样本过采样或欠采样，按Y值取值样本量的比例设置抽样或按Y值取值样本量的数量设置抽样。 |
| 3 | 缺失值处理 | 联合建模前对缺失值进行处理，解决多数算法无法直接处理缺失数据的问题 |
| 4 | 异常值处理 | 数据中的异常值会影响模型计算的准确性，建模前先识别是否存在异常值，如果存在则对异常值进行替换处理。 |
| 5 | 标准化归一化处理 | 归一化/标准化模块用于数据缩放，数据缩放可以消除数据量纲影响，提升模型的收敛速度。数据进行归一化后可缩放到0-1之间。数据进行标准化后服从标准正态分布。 |
| 6 | 数据集划分 | 在进行模型训练时，将建模样本划分为训练集和测试集两部分。训练集用于模型训练，测试集用于测试模型的效果。 |
| 7 | 变量分箱 | 根据等频分箱、等距分箱、卡方分箱等连续型变量分箱和一般离散型变量分箱、高基类变量分箱等离散型变量分箱对连续变量离散化，特征离散化后，保证模型稳定，降低模型过拟合风险。 |
| 8 | WOE转化 | 分箱之后得到了一系列的离散变量，对变量进行编码，保证字符型变量可在算法中进行入模。 |
| 9 | 特征统计 | 根据变量类型不同，按不同维度进行特性统计并生成统计分析报告。 |

（3）模型设计

重点污染企业违规排放模型中，生态环境部门作为企业排放监测任务发起方，需要电力有关数据支持和赋能，基于企业用电量等数据和重点污染企业数据，通过隐私计算平台的联邦学习功能模块构建重点污染企业违规排放模型，通过模型确定各重点污染企业有无按照政策在重污染天气时安规定减产或停产。重点污染企业违规排放模型选择了基于同态加密的纵向逻辑回归V-LR算法。

如下图所示，在样本对齐及模型训练过程中，参与方各自的数据均保留在本地，且训练过程中各节点之间只交互梯度值，不会导致数据隐私泄露，模型训练结果参与方只保存己方特征对应的模型参数，无法获取其他参与方特征的模型参数。因此，双方在联邦学习的帮助下得以实现合作训练模型。



**图6** 重点污染企业违规排放建模

重点污染企业违规排放模型构建完成后，通过AUC、KS等评价指对模型评价，根据模型报告了解各字段对模型训练的贡献率。同时使用LR模型产出的模型系数标结合企业日用电量、企业历史典型用电量、重污染天气气象等级等x值对重点污染企业进行预测，判断企业有无响应重污染天气停限产政策。

* + - 1. 应用成效

通过对电力数据及重点企业污染数据进行联邦学习模型构建，打造重点污染企业违规排放模型落地重点污染企业排放监测场景，可将企业有无响应重污染天气停限产政策结果数据反馈给生态环境部门，帮助生态环境部门及时了解企业排放现状及政策落实有效性，敦促在严控生产过程中的污染物排放的同时，达成特定天气环境状态下的节能要求。推进重点企业污染防治综合性工作，强化重点企业污染防治目标责任、调整优化产业结构、实施重点企业污染防治重点工程、加强重点企业污染防治管理。减轻了生态环境部门的工作量，提高了生态环境部门监管的效率和及时性。建立基于电力大数据的重点污染企业污染监测模型，能够统一企业电使用和污染排放监管于一体，创新重点企业污染防治管理体系，为建设生态文明提供有力支撑。

* 1. 基于多方安全计算的碳能耗监测
		+ 1. 业务背景

2021年9月中共中央、国务院印发《关于完整准确全面贯彻新发展理念做好碳达峰碳中和工作的意见》，提出要建立统一规范的碳核算体系，健全电力、钢铁等行业领域能耗统计监测和计量体系。

* + - 1. 数据准备

能耗监测数据包括但不限于煤、油、电、气、水等能源种类，并梳理其他相关的数据标签和维度，包括但不限于监测时间、企业、企业所属行业、企业所属区域、能耗种类、能耗数值等。

生产经营数据包括但不限于企业的工业总产值和工业增加值等财务统计数据。其中，工业总产值是指以货币表现的工业企业在一定时期内生产的已出售或可供出售的工业产品总量。工业增加值是指工业企业在报告期内以货币形式表现的工业生产活动的最终成果，是工业企业生产过程中新增加的价值。并梳理其他相关数据的标签维度，包括但不限于企业名称、工业总产值、工业增加值、时间、企业所属行业、企业所在区域等。

**表3** 企业能耗监测数据维度

|  |  |
| --- | --- |
| **数据种类** | **数据特征** |
| 能耗监测数据 | 煤-能耗数值(标准单位) |
| 油-能耗数值(标准单位) |
| 电-能耗数值(标准单位) |
| 气-能耗数值(标准单位) |
| 水-能耗数值(标准单位) |
| 热-能耗数值(标准单位) |
| 总能耗-能耗数值(标准单位) |
| 时间(年/季/月/日) |
| 生产经营数据 | 企业及ID |
| 企业所属行业及ID |
| 企业所属区域及ID |
| 工业总产值(标准单位) |
| 工业增加值(标准单位) |

* + - 1. 模型设计

开展重点用能企业碳排放及能源消耗量特征分析，提取企业碳能耗、重点行业碳能耗、重点区域碳能耗监测指标和预警指标。

（1）数据虚拟融合

基于多方安全计算构建联合统计分析解决方案，使同一重点用能企业的能耗监测数据和生产经营数据实现数据不离库的虚拟融合，然后基于虚拟数据集进行联合统计分析。

采用纵向虚拟融合的方法，重点企业能耗监测数据方作为参与方Guest与重点企业生产经营数据方Host共同对齐用户ID集，找到共同拥有的用户ID集合。



**图7** 企业能耗数据与生产经营数据虚拟融合

（2）耗能特征分析模型

针对重点用能企业，能源、交通、建筑、工业生产等重点行业以及重点关注区域，分别进行包括但不限于如下所示的分析：

1）汇总分析：对区域、行业及用能单位综合能耗计算汇总分析；

2）分时统计：统计企业/行业/区域某段时间内逐日、逐月、逐季、逐年用能情况数据，查看用能趋势，可根据已有的日期或者自定义时间进行查询，以图表形式显示；

3）同比环比分析：将企业/行业/区域的能耗与去年同期值和上月值进行同比环比分析，分析节能效果，以及节能目标的修正；

（3）能耗对标排序模型

基于多方安全计算，构建跨区域重点用能企业碳能耗对标、碳能耗排序分析模型，对所有参与方全量数据进行共同的比较和排序，针对不同区域、不同行业以及不同企业间的能耗进行对标分析，实现可视化的排序输出。

* + - 1. 应用成效

根据工业企业能源结构，匹配相应碳排放系数，计算各企业碳排放量。结合全社会对电力、燃气、煤、油耗用量的耗用量信息，对重点企业的碳排放量、碳强度进行预测，引导企业有序开展碳交易，为相关的单位、部门和领域的调度指挥和优化调整提供数据支撑。