ICS 国际标准分类号
CCS 中国标准文献分类号

# 团　　体　　标　　准

T/CES XXX-XXXX

# 基于联邦学习框架体系的电力数据隐私计算平台总体技术规范

Overall Technical Specification for Power Data Privacy Computing Platform Based on Federated Learning Framework System

XXXX-XX-XX 发布　　XXXX-XX-XX 实施

中国电工技术学会　发布

# 目　　次

# 前　言

本文件按照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》给出的规则起草。

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这些专利的责任。

本文件由中国电工技术学会提出。

本文件由中国电工技术学会标准工作委员会能源智慧化标准工作组归口。

本文件起草单位：国网信息通信产业集团有限公司，国网山西省电力公司电力科学研究院，国网湖北省电力有限公司信息通信公司，国网吉林省电力有限公司营销服务中心，国网浙江省电力有限公司信息通信分公司，国网河北省电力有限公司信息通信分公司，国网思极网安科技（北京）有限公司。

本文件主要起草人：文爱军，付昀夕，刘泽三，张敏，高紫婷，刘泽辉，余明阳，鞠默欣，周鹏，陈连栋，王振亚，张文娟，闫廷廷，李瑞，刘昕，宋昊燃，杨帆，赵林丛，闫晨阳，张攀，李兆隆，王健，马东娟，戴俊峰，龚小刚，高丽芳，倪鹏翔，王凌，周晨轶，辛晓鹏，赵莉，张崇超，李燕超，李亚红。

本文件为首次发布。

# 基于联邦学习框架体系的电力数据隐私计算平台总体技术规范

## 1 范围

本文件规定了基于联邦学习框架体系的电力数据隐私计算平台开发的参考框架、平台能力要求、平台建设技术要求等内容。

本文件适用于开展基于联邦学习的电力数据隐私安全共享应用的产品设计、软件开发。

## 2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是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

JR/T 0171-2020 个人金融信息保护技术规范

JR/T 0184-2020 金融分布式账本技术安全规范

JR/T 0218-2021 金融业数据能力建设指引

Q/CMB 006-2022 隐私计算总体技术规范

Q/BCTC 0002-2022 联邦学习金融应用技术要求

BDC 79-2021 隐私计算 跨平台互联互通 第一部分：总体框架

T/ISC 0015-2022 金融场景隐私保护计算平台技术要求与测试方法

## 3 术语和定义

下列术语和定义适用于本文件。

3.1

**隐私保护 privacy protection**

为保护隐私而采取的措施。例如：对个人数据的收集、处理和使用加以限制。

3.2

**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，保障数据在生产、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”。

3.3

**隐私计算技术 privacy-preserving computation techniques**

隐私计算技术主要分为三大方向，多方安全计算、联邦学习和可信执行环境。

3.4

**隐私数据 private data**

数据提供方输入的数据、结果使用方获得的数据，以及算法参数和模型参数种需要被保护的数据。

3.5

**联邦学习 federated learning**

联邦学习是一种多个参与方在不交互原始数据的情况下，通过安全机制交互模型参数，从而达到协同训练效果的分布式机器学习方法。

3.6

**节点 node**

隐私计算技术平台中提供所有功能或部分功能的部署实例，是联通网络的基本组成单元，对外提供交互接口：

a） 中心节点:提供管理和协调作用的节点；

b） 分布式节点:完成联邦学习任务的节点。

3.7

**特征 feature**

数据提供者提供用于联邦模型训练的指标。

3.8

**标签 label**

联邦学习发起方提供的用于模型训练的目标数据。

3.9

**联邦学习算法 federated learning algorithm**

完成联邦模型训练所使用的机器学习算法。

3.10

**联邦学习模型 federated learning model**

联邦参与方通过与其它联邦参与方共同训练生成的模型。

3.11

**联邦学习参与方 federated learning participant**

提供联邦学习数据或作为模型发起者的组织或机构，联邦参与角色可分为发起方、参与方、协调方：

a） 发起方：一般是任务的发起方，在纵向联邦学习场景中，一般为带有标签的一方；

b） 参与方：是数据提供方之一，在纵向联邦学习场景中，一般是没有标签的一方，仅提供数据和协同模型训练、模型预测；

c） 协调方：为联邦学习参与方的模型训练、模型预测等活动提供协调、辅助等支撑功能的组织或机构。

3.12

**多方安全计算协议 multi-party security computing protocol**

多方安全计算协议是一种保护数据隐私的计算协议，可以在多个参与方之间进行计算，同时保证数据的安全性和隐私性。

3.13

**电力数据业务场景 state grid data scenario**

一种电力数据流通共享和协同应用的业务场景，该业务场景中使用了电力数据。

## 4 符号、代号和缩略语

下列符号、代号和缩略语适用于本文件。

FL：联邦学习（Federated Learning）

AUC：ROC 曲线下的面积（Area Under Curve）

KS：洛伦兹曲线（Kolmogorov-Smirnov Curve）

## 5 概述

### 5.1 基于联邦学习的隐私计算平台

基于联邦学习的隐私计算平台逐渐成为主流，联邦学习隐私计算平台基于机器学习、深度学习算法和多方安全计算协议开发，支持一系列的联邦学习架构和安全计算算法，数据无需离开本地，将模型下发到本地服务器进行训练，以很小的数据交换量对模型进行迭代和更新，最终输出联合训练后的算法模型，高效实现和完成多方的联合建模与分析。隐私计算平台在数据驱动的领域中具有广泛的应用，促进了跨组织、跨边界的数据共享与合作，推动了各个领域的创新与发展，打造了数据和价值之间最安全可靠的桥梁，以支持联邦智能生态的发展和运作。

### 5.2 联邦学习邀约及生态构建

有意参与联邦学习的机构（数据提供方）将自身样本数据信息注册到数据隐私计算平台中心管理平台，供其他机构（数据需求方）查看。如果其他机构（数据需求方）需要该数据共同参与联邦学习，则对数据提供方发出联邦学习邀约，以申请数据使用权限。通过注册及邀约机制，形成联邦学习生态。

a） 邀约标的:
- 样本数据
- 计算资源

b） 邀约内容:
- 许可使用（无限期）
- 在许可的时间内使用
- 在许可的次数内使用
- 以数据及资源的使用量计费

c） 邀约方:
- 数据供需双方
- 供（数据提供方）、需（数据需求方）、中心管理平台（起到代理人的作用）三方

## 6 平台参考架构

### 6.1 联邦学习参考架构

在联邦学习架构中，分为中心节点与分布式节点。中心节点主要承担有协调方角色，协调方协调各参与节点协作构建联邦模型，协调方也承担计算方的角色，协调方宜取得其它参与方的信任或由具备公信力的第三方机构担任；分布式节点分为发起方和参与方，主要承担承担数据方、算法方、计算方、结果方的

角色。在联邦学习架构实际部署中，一个联邦学习发起方/参与方可承担多个角色。基于联邦学习框架体系的隐私计算平台参考架构如图 1。

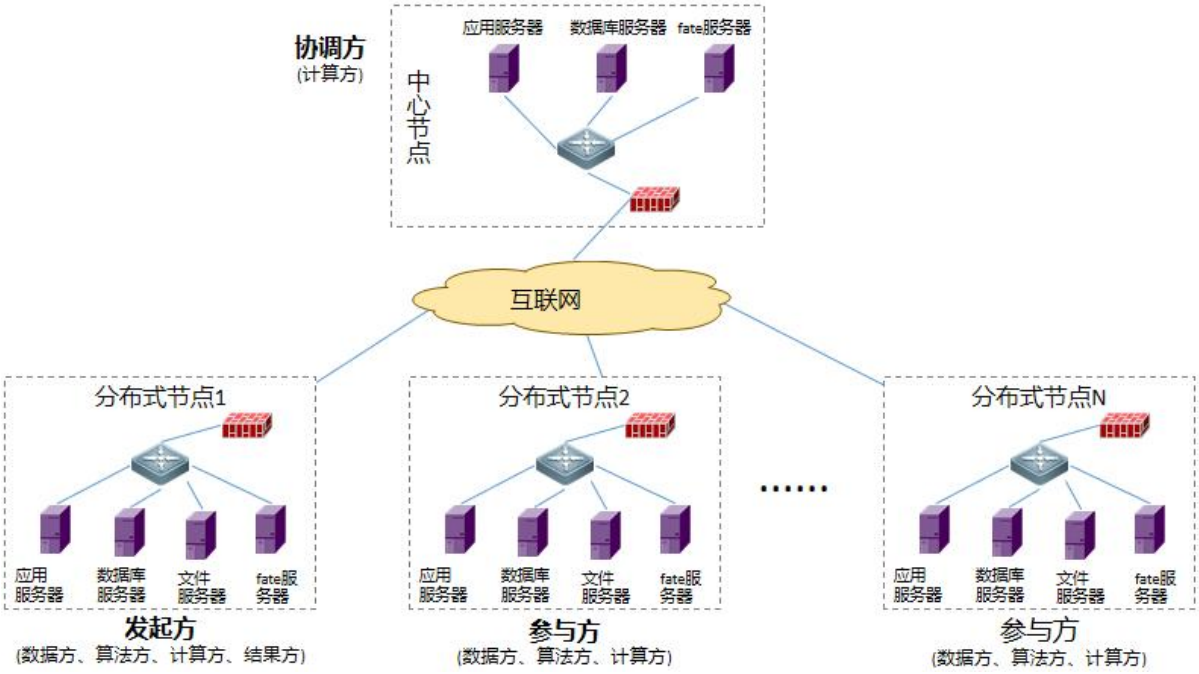


图 1 基于联邦学习框架体系的隐私计算平台参考架构

## 6.2 联邦学习节点角色

### 6.2.1 发起方

发起方是发起联邦学习任务的参与方，负责发起联邦学习邀约、模型训练、模型预测。一个联邦学习任务只有一个任务发起方。

### 6.2.2 参与方

参与方是接受联邦学习任务邀约申请并参与联邦学习任务的参与方，通过与发起方建立联邦学习合作，协助发起方完成联邦学习。一个联邦学习任务有一个或多个参与方。

### 6.2.3 协调方

协调方是根据联邦学习任务流程，通过协调各参与方完成联邦学习并同步联邦学习过程中各方任务状态的参与方。

### 6.2.4 数据方

数据方是负责提供联邦学习任务所需数据的参与方，包括：

a) 提供模型训练所需的训练样本数据；

b) 提供预测样本数据。

### 6.2.5 算法方

算法方是根据联邦学习任务需求，提供联邦学习算法的参与方。算法方提供的算法宜以算法组件形式输出。一个联邦学习任务有一个或多个算法方。

### 6.2.6 计算方

计算方是为执行联邦学习任务提供算力的参与方。一个联邦学习任务有一个或多个计算方。计算方与数据方通常由同一隐私计算节点承担。

#### 6.2.7 结果方

结果方是最终获得联邦学习任务预测结果的参与方。联邦学习流程应约定结果方，仅有结果方可获取模型预测的最终结果，其它非结果方不能获得结果，也不能通过中间计算过程推断出最终结果。一个联邦学习任务有一个或多个结果使用方。

## 7 平台能力要求

### 7.1 节点管理能力

节点管理模块应具备对参与联邦学习的隐私计算节点进行管理的能力，包括：

a） 节点注册与发现、节点上下线、删除功能；

b） 节点实体信息的信息维护，如保存、更新和同步；

c） 邀约基本操作，如节点发现、节点识别、节点通信验证、样本发现与邀约申请等。源节点应该能够根据目标节点的地址，交换双方节点基本信息，并且和目标节点建立邀约申请；

d） 节点间合作意向变更操作，如邀约拒绝、解约和合作有效期变更等；

e） 节点状态同步；

f） 合作有效性校验，应保证仅能与通过认证的合作节点开展正常工作；

g） 节点支持算法查询；

h） 应对节点运行状态以及节点与其他节点的连接进行监控，并收集节点状态数据写入日志供审查；

i） 应支持针对各节点的任务管理相关能力，包括但不限于任务的创建、任务调度、 状态监控和多任务并行等。

### 7.2 邀约管理能力

邀约管理模块应具备对联邦学习发起任务邀约进行管控的能力，包括邀约任务新增、邀约任务查询、邀约任务处理、邀约发起、邀约修改、邀约删除、邀约发布等，有联邦学习意愿的机构（数据提供方）将自身数据信息注册到数据隐私计算平台中心管理平台，供其他机构（数据需求方）查看。其他机构（数据需求方）若需要该数据共同参与联邦学习，则对数据提供方发出联邦学习邀约，申请数据使用权限。

### 7.3 样本管理能力

样本管理模块应具备存储和管理各类型样本资源的能力，为训练环境提供标注样本，对样本数据进行统一存储与管理，实现样本数据的更新、维护与共享，支撑模型训练。

a) 中心节点样本管理。建设注册样本列表查询、注册样本信息预览等功能；

b) 分布式节点样本管理。建设样本注册、样本上传、样本查询、样本编辑、样本解析、样本失效/激活、样本删除、样本预览、样本下载等功能，对样本的操作包括：

1） 数据预处理。联邦数据预处理可以分为线下数据预处理和线上数据预处理：

- 线下数据预处理。指联邦训练的参与方需要在数据加载之前将己方数据格式规范化，各方需要事先约定好输入数据的表头格式及内容，id 格式，特征格式；

- 线上数据预处理。包括联邦隐私求交，特征相关性计算，特征分箱，特征选择与过滤，特征缺失值处理，特征编码，特征降维，联邦特征稳定性计算等；

2） 用户配置本地样本数据信息；

3） 同一节点下不同用户的样本数据共享和同步、权限控制；

4） 数据字典的修改和管理；
5） 审批样本数据授权申请和更新授权意向更新；
6） 设置样本数据公开层级；
7） 查看合作方样本数据公开信息；
8） 根据合作方样本数据公开信息，申请样本数据授权。

### 7.4 算法管理能力

算法管理模块应具备对联邦学习算法进行管理的能力，包括：算法的基本操作，如添加、删除、更新等。

### 7.5 模型训练管理能力

模型训练管理模块应具备实现联邦学习模型训练的能力，支撑各类应用模型训练的需求，包括：联邦学习管理、提交模型训练、模型训练查询、模型训练复制等。

a） 联邦模型训练分为两种类型，横向联邦训练和纵向联邦训练；
b） 如果是有监督学习，则至少有一方提供标签数据（Y）；
c） 联邦模型训练是利用参与方提供的数据计算模型参数，使模型函数结果逼近标签（Y）；
d） 联邦模型训练过程对数据安全的要求是保证在数据交换过程中不泄露参与方的训练数据、不泄露参与方的模型参数、不泄露标签数据；
e） 常用模型有线性回归、逻辑回归、树模型、神经网络模型。

### 7.6 模型管理能力

模型管理模块应具备管理模型的能力，保存模型训练生成的模型，并为模型预测提供模型，包括：模型保存、模型查询、模型发布等。

### 7.7 模型预测管理能力

模型预测管理模块应具备实现模型预测的能力，支撑业务应用，包括：发起模型预测任务、模型预测查询、模型预测任务处理、提交模型预测、模型预测结果查询等。

a） 联邦模型预测分为两种情况，横向联邦预测和纵向联邦预测；
b） 横向联邦预测场景下，发起方以及各参与方都用于完整联邦模型、完整特征向量；
c） 横向联邦预测场景下，发起方、各参与方都可独自在本地完成联邦模型预测，输入本地预测样本数据，得出预测结果。

### 7.8 跨站点网络通信能力

联邦学习跨站点网络通信模块应具备协调不同参与方之间数据交换，确保数据能够以安全、高效的方式在各个节点之间传输的能力。管理模型参数的同步，确保所有参与方的模型保持一致。

a） 通信模块必须支持数据加密、身份验证和授权，以保护数据和模型的安全性；
b） 具备容错机制，能够自动处理通信错误或节点故障，以确保系统的可靠性；
c） 记录通信活动的详细日志，以便跟踪问题并进行审计。

### 7.9 多方安全计算协议

多方安全计算协议需要保证数据的隐私性、保证计算结果的正确性、保证计算的效率，即在计算过程中不会泄露数据的任何信息、计算结果应该与单独计算的结果相同、计算过程应该尽可能地快速。

### 7.10 平台运营管理能力

运营管理模块应具备提供平台支撑功能的能力，包括机构管理、用户管理、角色管理、菜单管理、权限管理、日志管理、字典管理、用户中心等。

## 8 平台建设技术要求

### 8.1 平台建设安全要求

a） 平台应具备系统日志和审计相关能力，能够通过日志对用户操作、数据使用等关键信息进行存证记录，便于审计和追溯；

b） 各参与方对联邦学习模型训练和预测过程中的操作日志、相关结果进行存证，用于复查和审计；

c） 平台应具备访问控制机制；

d） 平台接口应做好权限管理，防止未授权的调用，应具备对外公开接口的安全限制和安全控制措施，包括但不限于接口访问和调用的身份鉴别、设备鉴权、访问控制和审计机制等 ；

e） 数据应做好权限控制，保障数据经授权查看或使用。应保证在隐私计算全流程中各参与方不泄露原始数据信息,参与方无法获取其它参与方原始数据，且无法反推出原始数据，只有规定的参与方(发起方)可以获得模型预测结果，其它参与方以及第三方无法获得模型预测结果。只有本次参与方可获得本次联邦学习日志，且参与方无法通过日志反推出模型结果和中间数据；

f） 多方联合建模安全性，应保证安全原理与系统代码、运行日志、通信数据的一致性；

g） 多方联合预测安全性，应保证模型在预测服务时数据和模型参数的安全，应保证模型在线服务不会泄露数据隐私，包含模型服务的输入数据信息、模型服务地址、模型服务参数等。多方联合预测安全性应保证安全原理与系统代码、运行日志、通信数据的一致性；

h） 平台应具备节点身份认证的功能，对联邦学习过程中的关键环节进行身份认证，保证操作行为的合法性、合规性；

i） 应能对不同节点进行相应的权限设置和控制，避免出现信息泄露和操作风险；

j） 应具备与区块链技术融合的能力，提供可追溯性和数据不可篡改性；

k） 数据传输安全性要求：

1） 数据传输完整性。使用密码技术来提供通信数据的完整性保护和校验，节点间通信协议应具备对通信延时、中断等的处理机制；

2） 数据传输保密性。通信节点应对通信过程中的报文或会话进行加密处理，建立数据传输安全通信通道，避免因传输协议受到攻击而出现的信息泄露或篡改等保密性破坏；

3） 数据传输可靠性。采用可靠的数据传输协议、设置数据传输超时时间、监控数据传输过程等，通信节点之间应保持稳定的连接并能够及时传递数据，建立容错机制，以应对硬件故障、网络中断等问题，确保通信持续进行；

4） 监控和检测。应建立完善的监控和检测机制，及时发现并解决数据异常情况，以保证数据的安全性和可靠性。

### 8.2 平台建设可用性要求

a） 应保证系统的总体可用性，在业务要求的服务时间周期内具有快速恢复的高可用和备份恢复机制；

b） 应具备系统自动容灾恢复的能力，包括应用、数据库、中间件、网络和硬件平台等。

### 8.3 平台建设兼容性要求

a） 平台和组件应支持在不同环境进行部署，如云环境、虚拟机、物理机等；
b） 平台升级后新的应用协议格式应对涉及到对原来应用协议格式依然支持；
c） 应用系统升级后新的数据格式或文件格式应对涉及到对原来数据格式提供新旧数据转换的功能，原来用户的记录要能继承，在新的格式下依然可用；
d） 平台升级后新的应用接口格式宜对涉及到对原来接口格式提供支持。

### 8.4 平台建设易用性要求

a） 提供直观、用户友好的可视化操作界面，确保界面设计清晰、简洁；
b） 制定完善的说明文档，包括但不限于用户手册和技术文档，清晰地解释系统的功能和步骤。

### 8.5 平台建设容错性要求

a） 隐私计算平台应具备一定的容错能力，平台应具备在复杂网络环境中保持稳定运行的能力，应容忍网络震荡并快速恢复运行状态；
b） 引入容错冗余机制，保证在单个节点发生故障时，联邦学习任务不会受到致命性影响。容错实现方案：
1） 硬件冗余：通过提供多个硬件，以在硬件出现故障时保障系统可用性，如采用硬件镜像、多服务器均衡等技术；
2） 软件容错：通过应用软件技术、算法或系统设计来提高系统的容错性，如可采用数据备份、实时监控、多台设备负载均衡等技术；
c） 确保容错机制不会降低系统的安全性，在具有容错能力的同时，提供数据隐私安全保护。

### 8.6 平台建设性能要求

#### 8.6.1 计算性能

a） 应保证联邦学习模型评估指标，如 AUC（Area Under the Curve），Kolmogorov－Smirnov (KS) 统计量，准确率等：
1） AUC：ROC(receiver operating characteristic curve)曲线下与坐标轴围成的面积；
2） Kolmogorov－Smirnov (KS) 统计量: 两个概率分布函数之间的最大差异值；
3） 准确率: 模型正确分类的样本数量与总样本数量的比例；
b） 应保证预测效果相比明文计算，模型评估指标损失程度在合理范围。

#### 8.6.2 资源消耗

a） 与相同条件下的明文机器学习作业相比，联邦学习资源消耗合理；
b） 在线预测耗时、吞吐量等满足业务场景需要。

### 8.7 平台建设可扩展性要求

平台应遵循标准化、模块化的建设原则，应设置相对独立的功能模块，实现各项功能。

## 9 电力数据场景应用能力

电力数据隐私计算平台应具备电力数据业务场景应用能力，能够实现电网与能源网、交通网、信息网等的流通融合，充分发挥电力数据价值与应用能力，电力数据与政务、金融、通信、工业、农业、贸易等更多行业深入结合，从而形成多层级、多领域的密态数据智能流通生态网络。隐私计算组件需部署到相关方，完成协作完成联邦学习。

### 9.1 横向联邦学习电力数据业务场景应用能力

基于联邦学习框架体系的电力数据隐私计算平台应具备横向联邦学习业务场景应用能力，实现电力数据跨省区共享应用。具备基于电网能源大数据中心的横向联邦学习业务场景应用能力，跨省区分别部署联邦学习平台节点，各省各自接入本地重点用能企业能耗数据，监管部门可以通过任意节点进行跨省区的企业碳能耗对标。

### 9.2 纵向联邦学习电力数据业务场景应用能力

基于联邦学习框架体系的电力场景隐私保护计算平台应具备纵向联邦学习业务场景应用能力，实现电力数据与政务、金融、交通等行业数据相融合。具备应用于城市数字大脑、税收风险分析、小微信贷风控、精准电桩布网、电力负荷评估、电费回收预测等场景应用能力，为政务部门改善民生提供辅助决策、为电力行业业务赋能提供有效支撑。