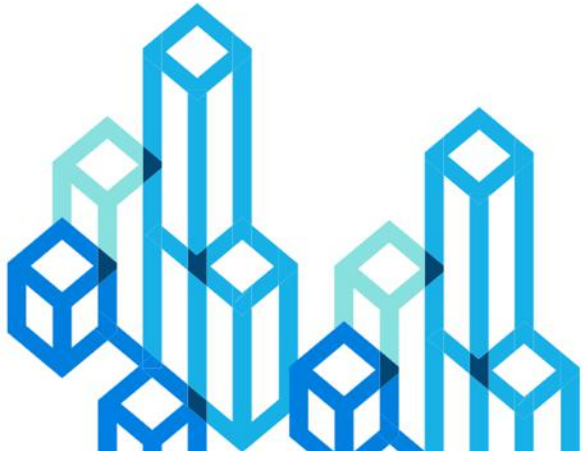
应用场景驱动的

隐私计算产品测试评估

2021.07.31



目录
CONTENT

《数据安全法》正式发布

全 国 人 民 代 表 大 会
The National People's Congress of the People's Republic of China

首页 | 宪法 | 人大机构 | 栗战书委员长 | 代表大会会议 | 常委会会议 | 委员长会议 | 权威发布 | 立法 | 监督 | 代表
对外交往 | 选举任免 | 法律研究 | 理论 | 机关工作 | 地方人大 | 图片 | 视频 | 直播 | 专题 | 资料库 | 国旗 | 国歌 | 国徽

当前位置： 首页

中华人民共和国主席令

来源： 中国人大网　浏览字号： 大 中 小　2021年06月10日 17:42:04

中华人民共和国主席令

第八十四号

《中华人民共和国数据安全法》已由中华人民共和国第十三届全国人民代表大会常务委员会第二十九次会议于2021年6月10日通过，现予公布，自2021年9月1日起施行。

中华人民共和国主席　习近平

2021年6月10日

■ 数据安全治理关键要点

- 应当坚持总体国家安全观，**建立健全数据安全治理体系**，提高数据安全保障能力；
- **国家统筹发展和安全，坚持以数据开发利用和产业发展促进数据安全，以数据安全保障数据开发利用和产业发展**；
- 明确提出了“**国家建立数据分类分级保护制度**”，并在数据分类分级保护制度基础上，又专门规定了对重要数据的保护；
- 明确支持“**数据安全检测评估、认证等服务的发展**”；
- 国家大力推进电子政务建设，提高政务数据的科学性、准确性、时效性，**提升运用数据服务经济社会发展的能力。**

保障数据安全与数据开发利用协调发展

数据在带来巨大价值的同时，也引入了大量的安全风险与挑战。**保障数据安全**和**促进数据开发利用**是相辅相成的车之两轮、鸟之双翼。

《数据安全法》提出了发展与安全并重的具体要求

——国家**统筹发展和安全**，坚持以数据开发利用和产业发展促进数据安全，以数据安全保障数据开发利用和产业发展。（第十三条）

在数据成为关键要素的数字经济时代，数据开发利用为数据安全提供了新的应用场景和发展机遇，数据安全为数据开发利用提供了基础的保障和稳固的底盘，两者相互促进。

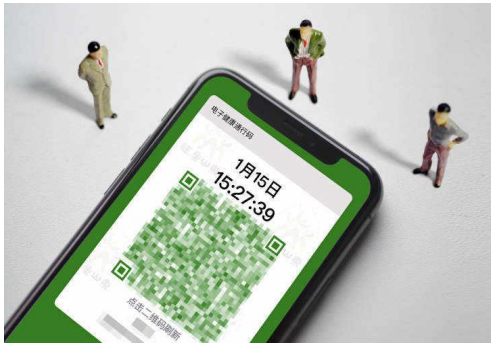
数字经济不是信息化时代的思路

IT时代，先建应用再有数据

DT时代，先有数据再建应用

数字应用创新需要多方数据的共享与协同

打通有价值的数据资源可以形成有价值的基于数据的数字创新应用



京津冀通关

实现海关、商务、税务、园区、货站等国际贸易各环节数据的共享交换，促进京津冀三地海关通关便利化。

防疫健康码

充分的数据互联互通实现一码通行，简化流程。

开放数据应用创新大赛

通过开放共享公共数据，广泛邀请各界人士参与数据资源的开发利用，寻找创新思路。

汇聚、流动、共享和协同成为数据的新特征

数据跨组织、跨平台、甚至跨境的汇聚、流动和共享，协同计算等，越来越多，数据流通产生价值。

新应用场景下数据安全保障问题重重

数据上云，数据泄露、数据滥用、隐私侵犯、网络暴力，现有的数据安全解决思路和方案已难以应对。

数据安全威胁无处不在

数据被攻击窃取&泄露

- 2020年全球公开报道了3932起泄露事件，泄露数据370亿条

内鬼泄露贩卖数据

- 中国电信“内鬼”出售超2亿客户信息
- 圆通“内鬼”泄露40多万条个人信息

数据污染&投毒

往数据集中掺入错误数据，破坏数据完整性、真实性，改变数据分析结果（污染AI训练数据集、污染基因数据集）

《全球数据安全观察》第50期

原创 国家工程实验室

大数据协同安全国家工程实验室 今天

收录于话题
#全球数据安全观察 20个 >

全球数据安全观察

总第50期 2021年第26期

2021.07.05-2021.07.11

目录

政策形势

1、滴滴旗下25款APP再遭下架：严重违法违规收集使用个人信息

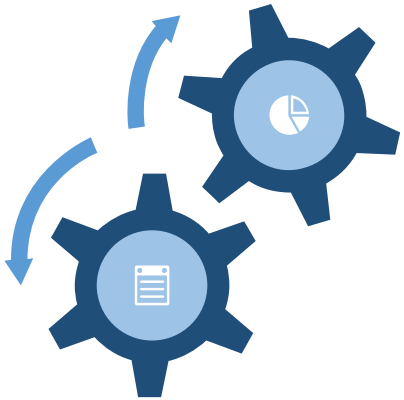
数据被加密勒索

- 美国最大燃油管道运营商遭勒索攻击停止运营
- 计算机巨头宏碁（Acer）受到勒索软件REvil攻击，赎金高达5000万美元，创下勒索软件赎金的新记录

合规挑战

抖音海外版，因非法收集儿童数据，分别在2020年7月被韩国罚款1.86亿韩元；2019年2月被美国罚款570万美元。

隐私计算进入了用户的视线



数据汇聚融合+兼顾发展与安全 双重需求驱动

隐私计算技术应运而生，成为解决数据流通过程中隐私保护问题的方案，有助于破解数据安全与开发利用的矛盾

隐私计算（Privacy Computing）是一类由两方或多方参与，在保护各自数据本身不对外泄露的前提下实现数据协同分析计算的技术。

Gartner发布的2021年重要战略科技趋势中 ，将隐私增强计算（Privacy Preserving Computing）列为未来几年内科技发展的九大趋势之一，指出其是专门为满足不断增长的共享数据需求同时保持隐私或安全性而设计的。

隐私计算成为了当前数据保护领域各界关注的热点。

隐私计算三大关键技术路线

随着密码技术、硬件技术的发展加速商业化， 隐私计算的技术路径也处于高速的演进状态，其中**联邦学习**、**安全多方计算**和**可信执行环境**是当前主流技术路径，也是当下产品化的主要方向。

在隐私合规、数据安全条件下的多方数据协同安全计算，解决数据孤岛问题

安全多方计算

一组相互独立的数据所有方在互不信任、且不信任任何公开第三方的条件下，应用各自的私有输入联合完成某个函数的计算。

可信执行环境/安全屋

提供可信的环境来执行处理或分析，该环境与其他应用以及操作系统隔离，保护其保密性和完整性。

安全屋：基于可信第三方，多方数据安全传输到安全屋，进行明文高效计算。

联邦学习

数据提供方的数据在不出本地的情况下还能通过联邦学习技术进行协同计算、建模，并辅助加密通信、同态加密、差分隐私、hash函数等技术保证数据隐私。

$|V\_FED-V\_SUM|<\delta$

隐私计算产品蓬勃发展

公司	核心产品	主要技术路线	是否开源	主要应用行业
微众银行	FATE、WeDPR-PPC	联邦学习	是	金融
蚂蚁	蚂蚁摩斯	MPC，TEE	否	金融
百度	PaddleFL框架，MesaTEE，点石	联邦学习，MPC，TEE	底层框架开源	金融、营销
腾讯	Angel PowerFL框架，腾讯安全联邦学习	联邦学习	底层框架开源	金融、广告
字节跳动	Fedlearner	联邦学习	是	金融、电商
360数科	FastFL	联邦学习	否	金融
富数科技	Avatar，FMPC安全计算产品	联邦学习，MPC	否	金融、医疗
华控清交	PrivPy多方计算平台	MPC	否	金融、政务
翼方健数	翼数坊	MPC	否	医疗
同盾科技	智邦iBond平台	联邦学习，MPC	否	金融
矩阵元	Rosetta框架	联邦学习，MPC，TEE	是	金融
锘崴科技	隐私机密计算平台	联邦学习，MPC，TEE	否	医疗
光之树	天机可信计算框架，云间联邦学习平台	联邦学习，MPC，TEE	否	金融

隐私计算应用场景

从金融、医疗等领域逐渐向其他行业延伸

隐私计算助力银行联合建模，提升反欺诈模型水平

银行应用隐私计算产品，融合多方的黑灰产行为等特征，反欺诈模型的KS提升30%以上，每年阻止数亿资金的风险贷款申请。

隐私计算有效助力医学影像识别、疾病筛查、AI辅助诊疗等

多家医疗机构可以通过横向联邦学习联合构建目标检测模型，用于辅助通过医疗图像的疾病检查（如肺部X光片检查等）。

隐私计算助力政府数据开放，实现精准施策

通过联邦学习平台，实现政务、银行、企业的三方的协作建模

联邦学习助力广告程序化交易联合建模，提升广告主投放效果和用户体验

通过广告主和流量主的联邦建模，融合双方的数据优势，在游戏、金融、教育、电商行业的广告应用案例中能够取得显著效果提升。

隐私计算的应用还面临哪些困难？

隐私计算虽然已经开始在不同行业初步应用，但是受限于**计算复杂度**、**多方交互效率**、**模型性能**等问题，大部分的应用场景均聚焦于少量数据的支持，**对海量数据场景的支持能力还有待提升**。

隐私计算普遍面临**计算效率低**和**应用场景有限**两大主要瓶颈，例如：

- MPC和联邦学习技术都受制于网络传输的带宽、通信速率和网络稳定性，计算和建模效率尚不够令人满意。
- 同态加密的计算有严重的性能瓶颈。与明文的计算相比，密文计算所需的时间与空间资源都较大；在运算类型上，只支持加法和乘法运算，目前没有较好的解决除法计算的方案。
- 针对特定问题和场景，需要设计专用协议，缺少针对特定场景动态数据集的算法框架，更缺少适应多场景动态数据集的普适性算法框架。

隐私计算应用受制于对技术的信任度

由于隐私计算技术复杂且常常呈现“黑盒化”现象，且处理对象常涉及敏感数据资产，隐私计算众多技术提供方须首先建立信任，提升需求方接受程度。

但是，国家**缺乏明确的监管文件和技术标准**认可隐私计算的安全可靠性，**缺乏明确的拉动性政策和标杆性示范项目。**

因此，很多客户都对如何证明技术的安全合理性有疑虑，只能在小范围内测试和监管沙箱内应用。

需要

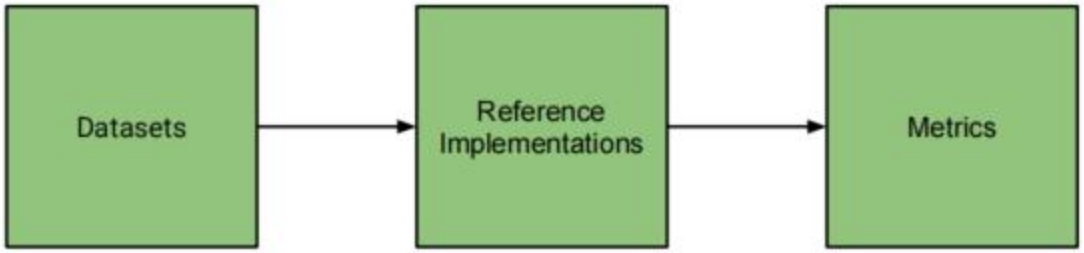
- ✓ 制定面向隐私计算技术应用的**测试评估标准**，在市场端对技术落地应用**加强监管，保证技术得到正确实现和安全**；
- ✓ 通过具有资质的**国家级、行业级第三方评估机构**对隐私计算技术**应用效果及安全性进行检测**，提升隐私计算需求企业及公众对技术的信任度。

国家工程实验室的使命担当

- 我们是大数据领域安全方向唯一的国家工程实验室
  - 大数据的安全开发利用是数字经济时代的主旋律；
  - 我们需要站出来。
- 隐私计算的应用还不成熟，特别是在大数据的安全利用方面
  - 受限于产品成熟度；
  - 受限于性能。
- 缺乏应用引领，需要在应用单位和数据之间搭起桥梁
  - 跨行业、跨领域的数据开发，数字创新应用还严重不足；
  - 将应用单位的“idea”转化为多方数据之上的应用。

目录
CONTENT

隐私计算评估系统的有关研究



现有的一类开源联邦基准测试系统包含了3个核心组件：

- “Datasets”模块负责数据的预处理，将其转换为标准化格式。
- “Reference Implementation”模块是一个联邦学习常用方法库，每种实现方式都会生成各种不同的统计以及系统指标日志。
- “Metrics”模块用来汇总和分析这些指标日志。

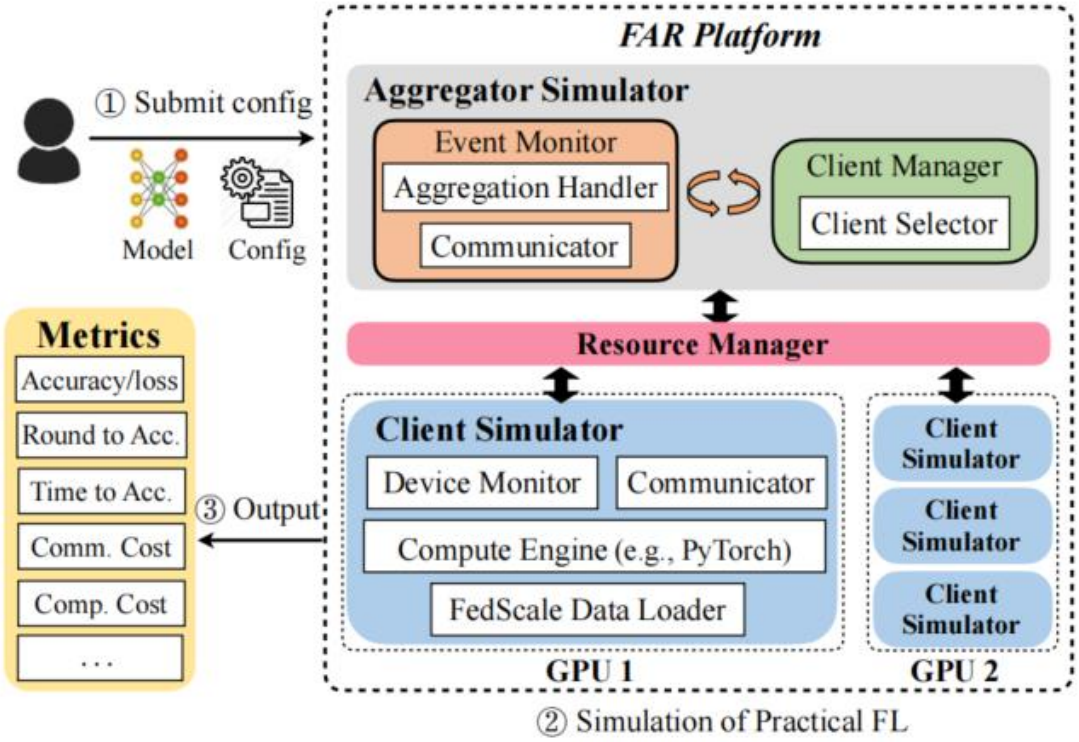
在数据集方面，其提供了一套**来自真实环境的联合数据集**，数据集关注以下几个特点：（1）数据具有自然键生成过程（可标识特定的用户等）；(2) 数据产生于数千至数百万台设备的网络等等。

Caldas, S., Duddu, S. M. K., Wu, P., Li, T., Konečný, J., McMahan, H. B., ... & Talwalkar, A. (2018). Leaf: A benchmark for federated settings. arXiv preprint arXiv:1812.01097.

隐私计算评估系统的有关研究

第二类自动化评估平台同样提供了**真实数据集**，包含小、中、大体量的数据。这些数据集的原始数据收集来源不一，同时以不同格式存储，适用于各种任务类别，例如：图像分类、对象检测、语言建模、语音识别、推荐和强化学习等，能够在更真实的环境中做到简化、标准化评估。

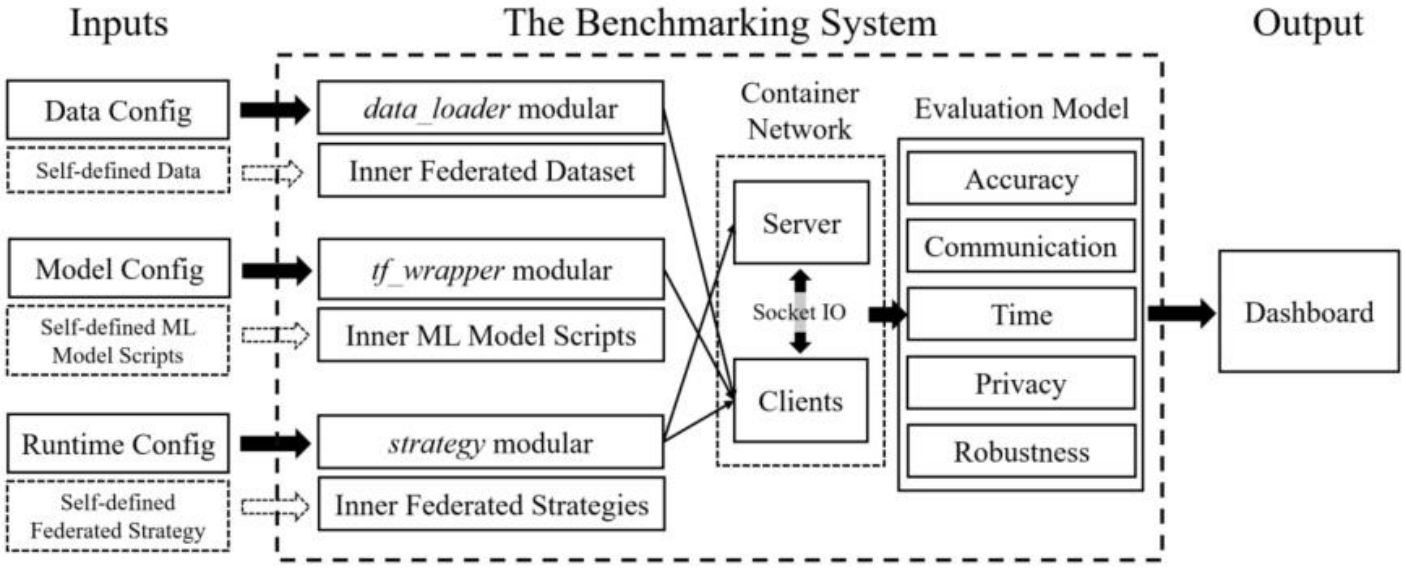
通过在 GPU/CPU 上模拟真实的联邦学习行为，同时利用精确度、计算/通信成本、延迟等指标，来评估隐私计算的效能。



- **Aggregator Simulator：**充当实际联邦学习中的聚合器，负责选择参与方、分配执行配置文件（例如模型权重）并处理结果（例如模型更新）。
- **Client Simulator:**充当联邦学习中的各个客户端，加载该客户端的数据集并将此数据提供给计算引擎以运行训练/测试。
- **Resource Manager:**协调可用的资源。

Lai, F., Dai, Y., Zhu, X., & Chowdhury, M. (2021). FedScale: Benchmarking model and system performance of federated learning. arXiv preprint arXiv:2105.11367.

隐私计算评估系统的有关研究



提供若干组**真实的数据集**，以精确度, 通信, 时间效率, 隐私保护, 鲁棒性五大指标作为评估模型，使评估更加标准化。

- Step1：确定基准数据集、模型和联邦学习策略，然后修改数据配置、模型配置和运行时配置。
- Step2：使用内置工具为客户端生成数据，启动客户端与服务端的通信。
- Step3：通过内置评估模型来全面评估不同联邦学习模型，并通过仪表板监控查看评估状态和结果。

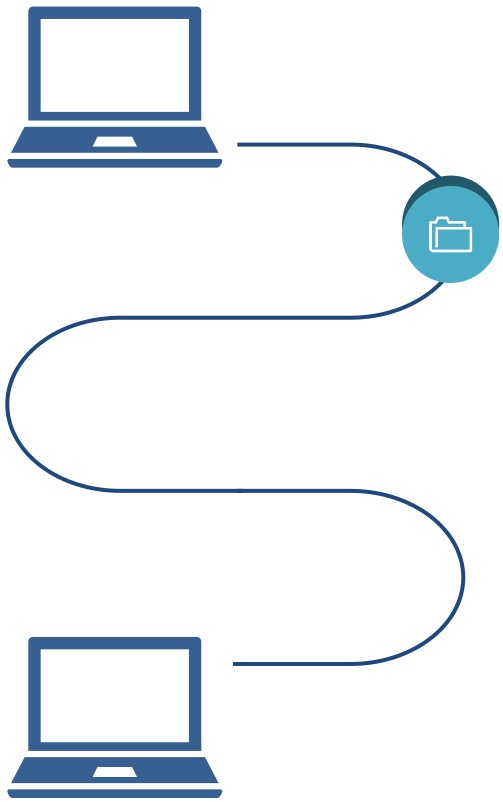
Chai, D., Wang, L., Chen, K., & Yang, Q. (2020). FedEval: A Benchmark System with a Comprehensive Evaluation Model for Federated Learning. arXiv preprint arXiv:2011.09655.

应用场景驱动的隐私计算测试验证

从用户的实际业务需求出发，平台能否充分满足业务需求，是应用场景驱动的测试验证最关切的问题。**从功能、性能、安全性和可用性四个方面，对隐私计算平台及其应用效能进行测试评估**，目的促进隐私计算相关产品的计算性能和安全性水平的提高，推动相关产品和解决方案在实际应用场景的落地，为数据的安全利用找到最佳方案。

- 利用真实应用场景和真实业务数据，支撑隐私计算相关产品或模块的测试评估；
- 从安全漏洞层面，对隐私计算平台的渗透测试；
- 从密文计算、可信执行环境、联邦学习等层面，对隐私计算本身进行安全性测试，测试是否存在信息泄露、数据污染等安全风险；
- 从性能层面，计算耗时是评估的主要参考维度，和明文本地计算结果、明文基准模型的相对误差和评价指标进行核验；
- 从可用性、易用性层面，邀请真实用户，通过用户的使用来评估产品是否满足用户需求。

利用真实的应用场景和真实的数据



- 用户在选择产品时往往会选择更适合自己的，而不仅仅是追求“高性能”，使用场景的区别，使得用户的侧重点不一。
- 真实的业务应用环境往往是多变且复杂的。
- 模拟仿真测试数据通常体量较小且关系简单，不足以体现“大数据”的协同计算。

采用真实生产环境数据以及结合具体行业应用和需要打通数据的场景，**验证在各自核心应用场景下，隐私计算相关产品是否能够输出最佳效能**。此类在实际应用的测试验证将有更高的参考价值。

面向典型的应用场景

面向智慧城市和行业领域的典型应用场景，支持政务数据的安全开放利用，支持各行业领域数据的安全打通，根据他们的数字创新应用对多方数据的需求，来测试隐私计算产品。

- 测评机构（离线仿真测试）
  - 数据：仿真数据；
  - 目的：在特定硬件资源、特定数据集、特定算法要求和特定目标要求条件下，模拟实际需求场景，测试不同产品各维度指标，如功能、安全性、性能等指标，通常是标准符合性测试，例如各大测评机构的测试。
- 应用场景驱动的测试验证
  - 数据：不是生成的数据，而是来自于真实业务场景的“活数据”；
  - 目的：在真实的应用场景下，基于实际场景中的数据，开展功能、性能、安全性和可用性的测试验证，一方面指导数据治理，另一方面看哪个产品更能解决实际场景中的问题。

应用场景驱动测试的目标

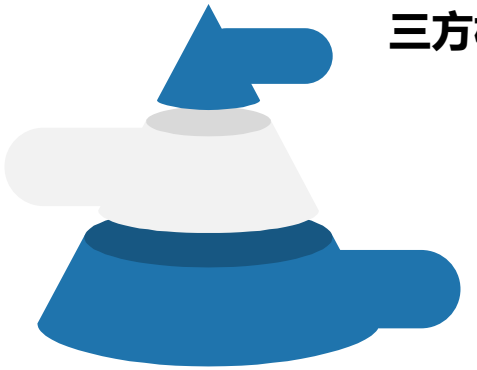
- 推动隐私计算产品的发展
  - 隐私计算产品更加成熟。
- 推动隐私计算的实践进程
  - 应用得更好、更充分；
  - 探索新的应用场景，更好地促进数据的安全开发利用。
- 推动数据安全治理工作
  - 来自于真实业务场景的“活数据”，需要经过治理才能更好的使用；
  - 为各领域数据安全治理、数据利用提供可落地的具体建议，使数据“荒山”成为可用好用的“良田”。

目录
CONTENT

利用国家工程实验室的第三方属性

大数据协同安全技术国家工程实验室是由国家发改委批复，我国大数据安全领域唯一的一个国家工程实验室，由360牵头承建，十家单位联合建设的具有第三方属性的机构，验收后将成为法人实体独立运营。

打造具有第三方公信力的、权威的开放大数据协同安全计算中心，选择最适合场景需求的隐私计算产品



结合具体场景和问题，建立第三方机制，持续运营、开放创新

为隐私计算的测试验证设计场景并提供解决方案

作为参与计算的一方进行测试评价

在无可信第三方的应用场景下，作为参与计算的一方，例如，作为某个委办局单位、某个医院；通过本方固定数据集和相同的应用，可以对模型算法效能、参与方的数据质量进行评价。

丰富数据资源

实验室在开展数据安全业务的过程中积累了大量的数据和应用场景，拥有众多大数据资源

结果验证

实现数据参与方之间的互联合作，对计算结果进行验证

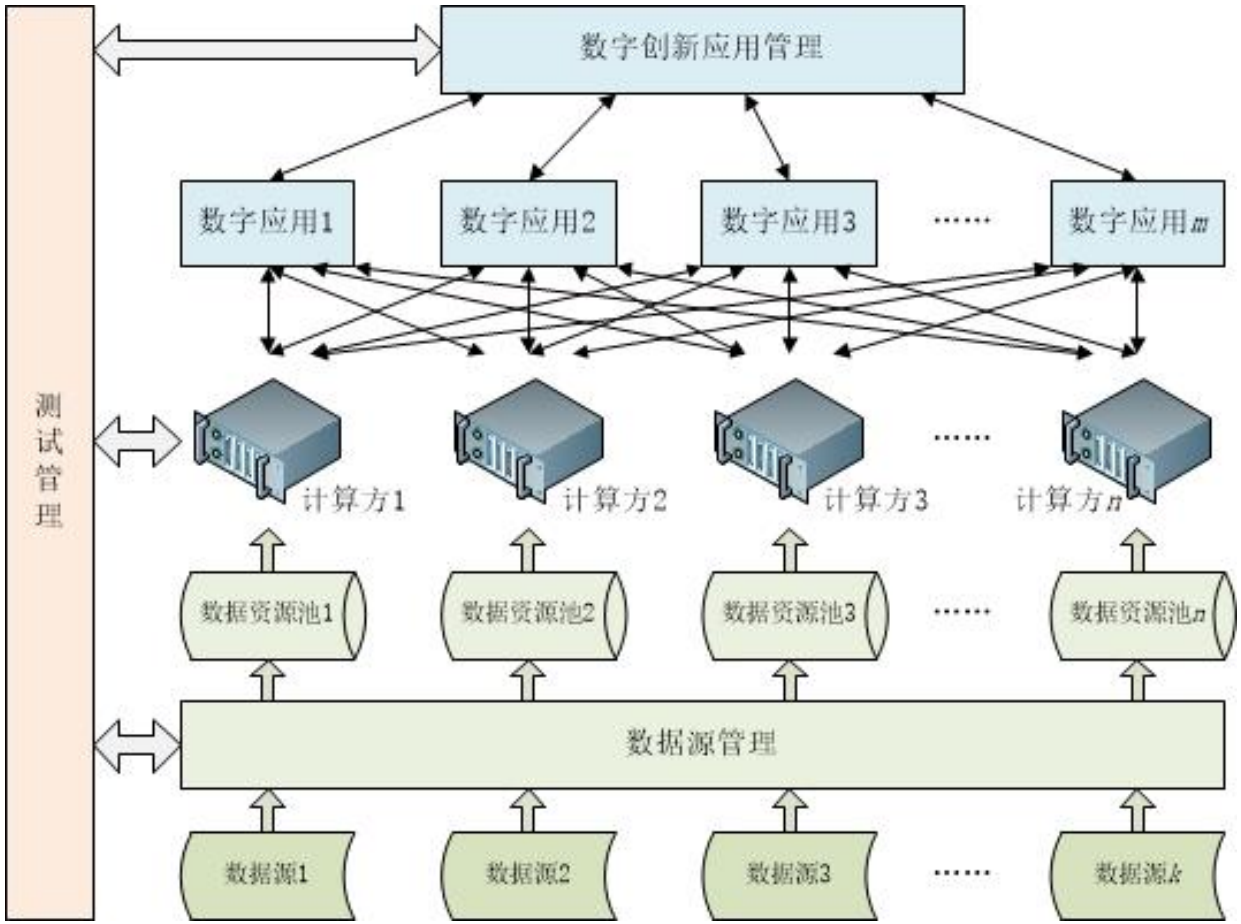
监管审计

对多方安全计算节点中的日志存证进行安全审计

隐私计算的测试验证平台架构

- 测试管理：对测试验证平台本身各系统或组件进行管理，包括数字创新应用管理、各计算方、数据源管理等，管理测试项目或任务，生成测试报表和报告。

- 数字创新应用管理：根据智慧城市和行业应用需求，开发、吸纳并管理需要多方数据协同计算的数字创新应用，根据测试项目或任务，启动具体的数字应用；
- 计算方：部署隐私计算产品，根据应用和测试需求确定需要多少个计算方；
- 数据源管理：对数据进行管理，根据数字应用的计算需求，从各智慧城市或行业领域的数据源中，梳理出用于隐私计算的数据资源，形成每个计算方的数据资源池。



基本测试流程

① 创建测试项目或任务，一个项目可包括多个任务，例如1个针对三个隐私计算产品的比对测试项目，包含3个测试任务；

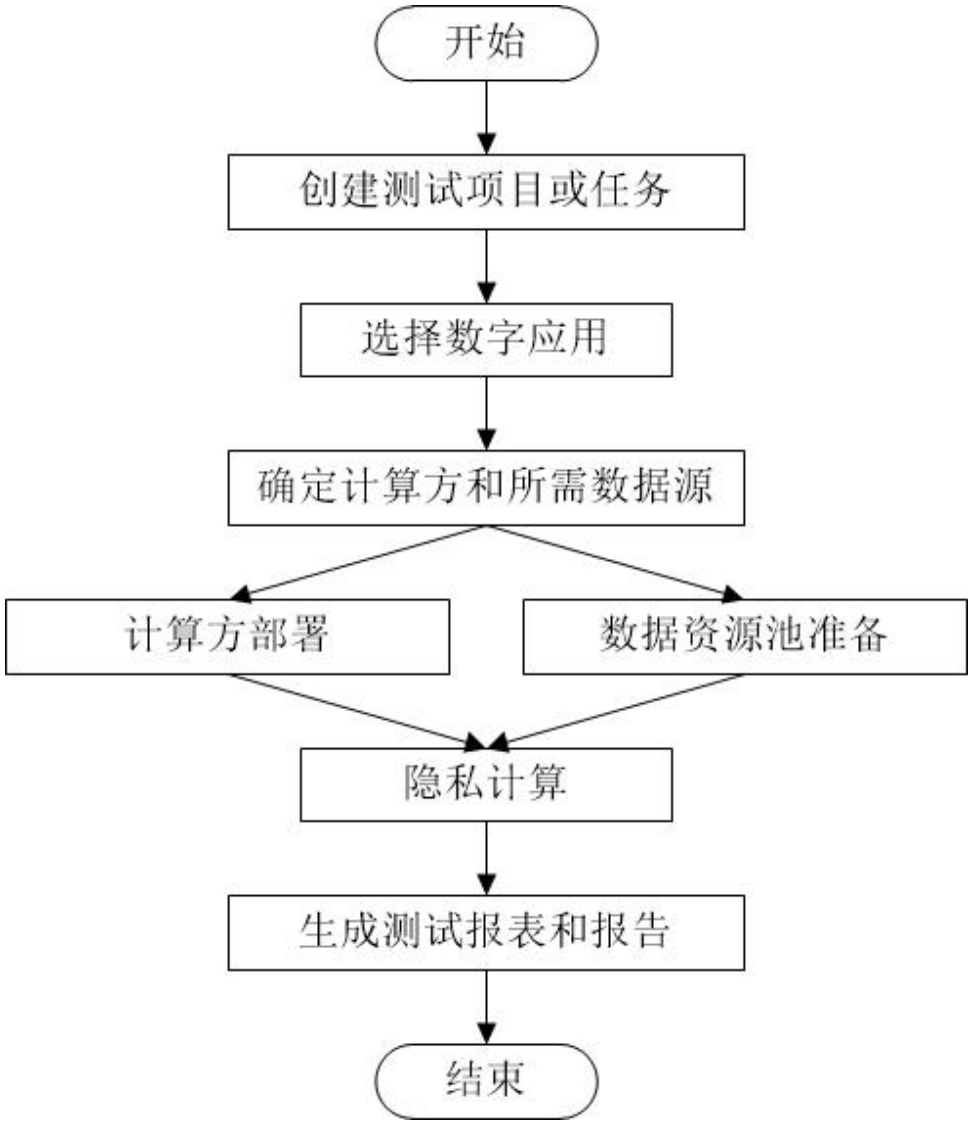
② 选择一种应用场景中的应用，根据该应用确定计算方数量和所需的数据资源；

③ 利用待测隐私计算产品去部署相迎数量的计算方；

④ 根据数据资源需求，从真实数据资源中为每个计算方调配相应的数据资源池；

⑤ 多个计算方进行隐私计算，满足设定条件后结束；

⑥ 根据测试结果生成测试报表和测试报告。



智慧城市的应用场景和数据

一个城市的应用场景，利用城市内各委办局的数据，以及医疗、交通、金融、公安等行业领域的数据，开展隐私计算的测试验证。

在智慧城市中，具体的应用场景例如：政务大数据打通

- 城市居民的多维度信用评级；
- 个人联合风控，其中需要横向打通的数据包括交通出行数据、水电燃气数据、公安数据、征信数据等；
- 机动车行车数据和保险公司数据的打通。

多个城市之间的应用场景和数据

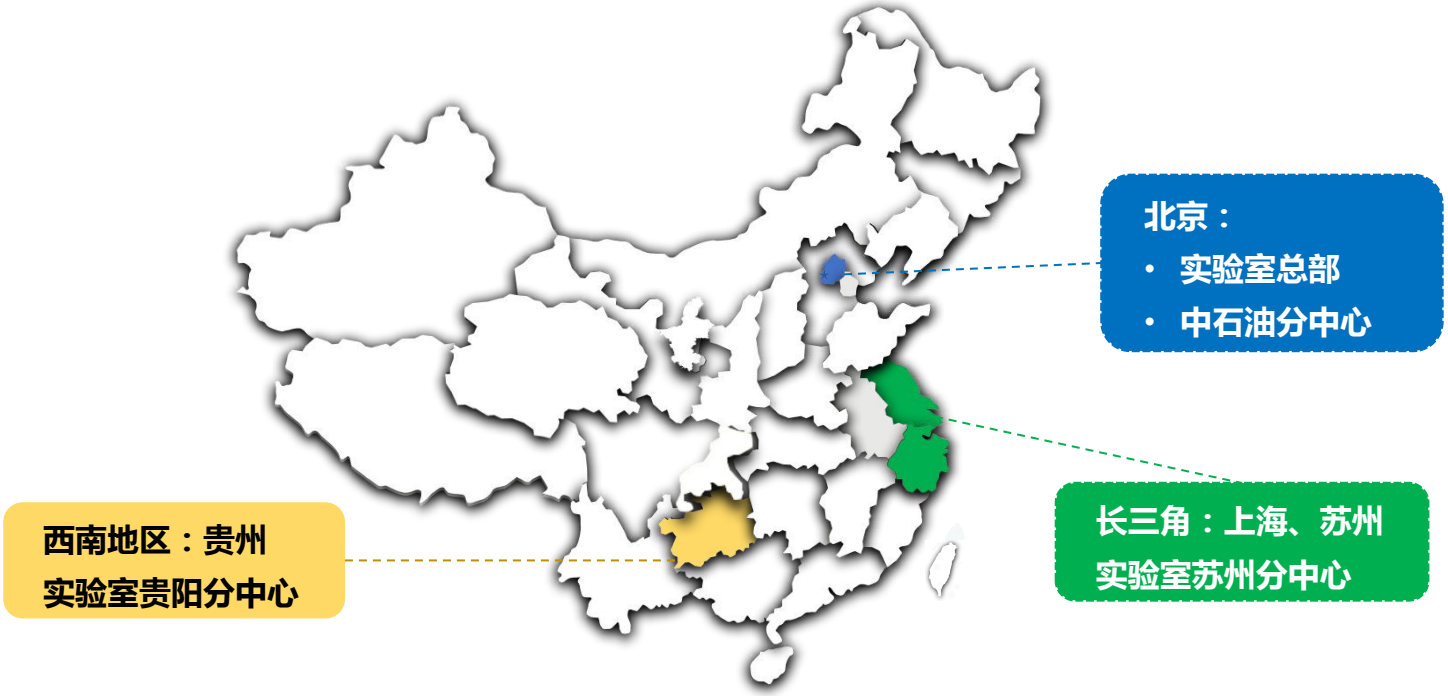
基于实验室在多个城市设立的分中心，进行城市间数据的融合打通，开展隐私计算的测试验证。

提升城市间各种产业之间的联系紧密度，具体应用场景例如：

- 电子证照共享互认等跨省数据共享应用；
- 跨城市共享空气自动监测数据，为大气污染联防联控提供了信息共享渠道与数据支撑。

本地企业在其他城市的业务开展和收入，例如：

- 工业互联网数据资源跨城市合作共享，提升企业各地方分部产线的智能化水平。

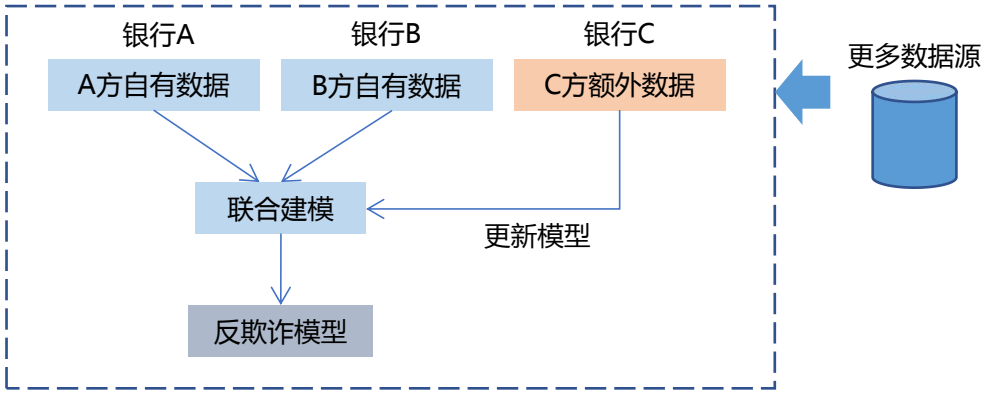


行业领域的应用场景和数据

面向行业领域，探索大数据利用和安全分析新方法，支持各行业领域数据安全打通。

在金融领域中，例如：

- 多个银行机构共同合作，实现多方联合风控，在保证数据安全的同时，有效降低业务风险和不良资产率；
- 反欺诈模型的训练。

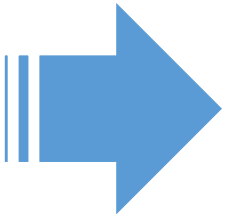


在网络安全领域的应用场景

在安全领域，以360安全大数据为基础，在保障数据安全条件下充分利用各城市或行业用户的大数据，进行联合建模与分析，提升网络攻击行为分析特别是APT监测分析的智能化水平。

- 360覆盖全球的网络神经元收集超250 亿恶意样本、22 万亿安全日志、80 亿域名信息、2EB 以上的安全大数据
- 360一线对抗和APT狩猎形成的安全知识及安全技术、安全专家

- 各城市或行业领域的安全大数据打通，提升安全能力



攻击监测模型训练

- 构建APT攻击模型
- 构建威胁画像
- 威胁分析
- 溯源分析

从当前应用现状来看，由于不同厂商的隐私计算技术平台基于自有知识产权算法和系统设计实现，技术路径与功能实现差异明显，导致数据的应用机构在面对与不同的数据提供机构合作时，需要部署不同隐私计算平台，大大增加了系统建设和平台运营的成本。**隐私计算跨平台的互联互通成为了新的挑战。**

通过**开展隐私计算产品的互联互通测试**，推动互联互通相关标准规范的制定，对隐私计算产品相关标准的符合性以及互联互通实际应用效果进行综合测试与评价，促进协议互通、系统互通，防止“新”的数据孤岛出现，推进隐私计算产业生态的建设。

国家工程实验室招贤纳士，欢迎加盟

大数据协同安全技术国家工程实验室在今年验收后将变更为工程研究中心，总部设在北京，成为法人实体独立运营，开展数据安全咨询、测评和培训服务，研发数据安全产品，搭建若干数据安全公共服务平台。

隐私计算的测试验证是其中的重要模块！

招聘动手实践能力强的数据安全工程师！

招聘数据安全研究人员！

共同打造隐私计算的测试验证平台，并结合智慧城市和行业领域的具体应用场景，推动隐私计算技术和产品的发展。

目录
CONTENT

大数据协同安全技术国家工程实验室获批准建设

国家发改委批复文件

国家发展和改革委员会办公厅文件

发改办高技〔2017〕147号

国家发展改革委办公厅关于开展大数据协同安全国家工程实验室组建工作的通知

北京市发展改革委、中国电子信息产业集团有限公司：

你们报来《关于报送大数据领域创新能力建设专项的请示》（京发改文〔2016〕473号）和《关于报送大数据领域创新能力建设专项的报告》（中电计〔2016〕522号）均悉。经研究，现函复如下：

一、原则同意由北京奇虎科技有限公司作为承担单位，中电长城网际系统应用有限公司为参与单位，联合相关单位筹建大数据协同安全技术国家工程实验室。

二、该工程实验室的主要任务：针对我国大数据环境下数据安全和系统安全监测、预警和控制处置能力不足等问题，围绕提升大

大数据协同安全技术国家工程实验室主要任务：针对我国大数据环境下数据安全和系统安全监测、预警和控制处置能力不足等问题，**围绕提升大数据安全分析能力和保障大数据自身安全的需求，**建设大数据协同安全技术与应用研究支撑平台，聚焦态势感知、漏洞挖掘、威胁情报、数据安全和隐私保护等重点方向，形成国内一流的科研环境，培养和汇聚大数据安全领域的高端技术人才，承担国家和行业的重大科研项目，孵化创新技术和产品，形成可持续的产学研用协同创新机制，为推动大数据安全的技术进步和产业发展提供技术支撑。

麦思博(msup)有限公司是一家面向技术型企业的培训咨询机构，携手2000余位中外客座导师，服务于技术团队的能力提升、软件工程效能和产品创新迭代，超过3000余家企业续约学习，是科技领域占有率第1的客座导师品牌，msup以整合全球领先经验实践为己任，为中国产业快速发展提供智库。

高可用架构公众号主要关注互联网架构及高可用、可扩展及高性能领域的知识传播。订阅用户覆盖主流互联网及软件领域系统架构技术从业人员。

高可用架构系列社群是一个社区组织，其精神是“分享+交流”，提倡社区的人人参与，同时从社区获得高质量的内容。