高安全

重点：

* 1. 数据不丢失，即高可靠的数据存储。即使发生自然灾害、人为损坏、系统错误等极端情况，企业应该也能保证核心数据可恢复，保证企业可持续运营。
  2. 数据不泄露。保证数据只有授权人员能访问，它包括“访问控制”、“身份认证”、“解密脱敏”、“安全审计”等，牵涉到业务前端、网络、后台系统等多个方面。
  3. 数据准确完整。核心金融数据的丢失、泄露造成的社会损失通常是企业所不能承受的。因此，金融企业的数据安全不仅一直都有严格的法律法规要求，相关的行业标准也在随技术、业务的发展而更新。

痛点：

* 1. 数据访问风险。这又包含缺乏统一账号管理、缺失身份认证管理、数据授权能力弱这三个问题。
  2. 数据流动风险。这包含缺乏审计溯源能力和数据保护能力弱两个问题。
  3. 数据运维风险。包括数据管理成本大、运维行为缺乏监督、高危操作缺乏管控等。

难点：

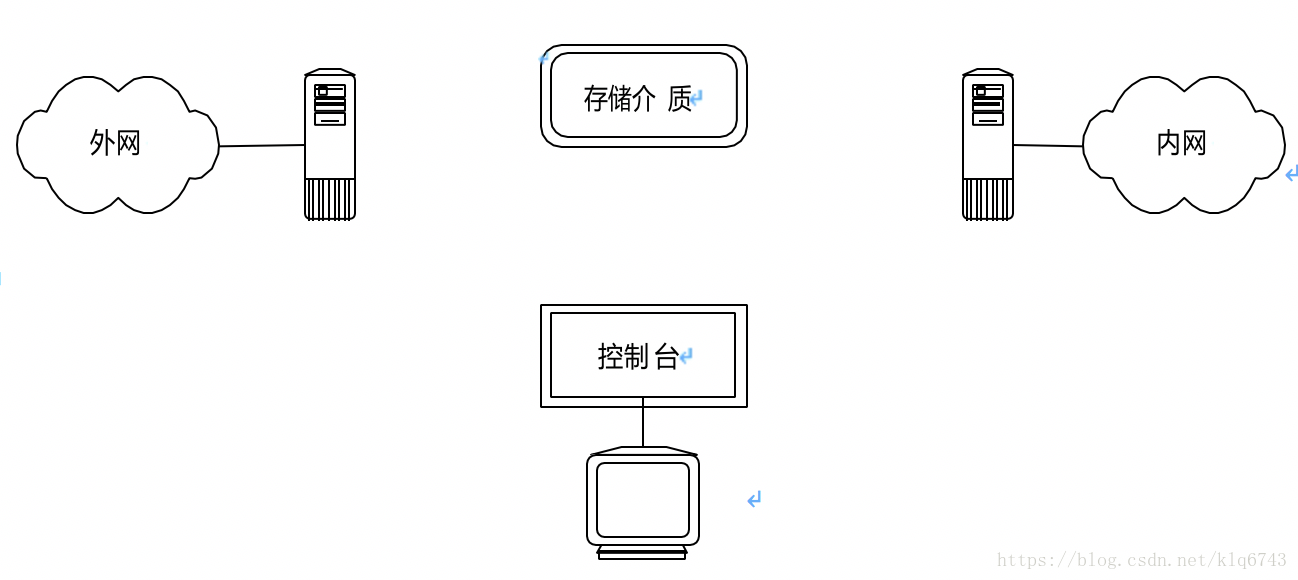
* 1. 法规多、标准多、更新快、检查多，怎样确保合规要求。
  2. 业务系统分散，涉及数据太广，难以归纳整理。
  3. 用户体验和安全性的平衡。

解决方案：

* 1. 规范大数据访问人员，统一规范访问控制。需要建立大数据统一用户管理系统，打通原有企业账号体系。还要建立大数据统一认证管理体系，多因子控制访问入口，防止数据裸奔。
  2. 建立大数据的资源管理能力和规范数据授权流程和手段。需要建立数据资产统一管理查询平台，开展数据分级分类管理。
  3. 对敏感数据访问进行控制保护。这包括：设置细粒度权限，控制敏感库表、字段、文件被低权限用户获取。提供精细化运营管控手段，基于数据等级、数据标签、数据分类进行保护。提供透明化的动态脱敏能力。控制用户访问数据频率和数据体量。控制用户的高危操作。
  4. 控制数据导出风险。基于敏感数据级别和权限，对导出场景进行审批。并限制开发导出数据落地，审批后通过统一的平台进行数据导出。
  5. 对内部人员内审、异常行为分析和事件溯源。包括以下三个方面：对大数据管理员的操作行为进行审计、UBA 分析和溯源。对大数据开发、分析的操作行为进行审计、UBA 分析和溯源。对数据导出的操作行为进行审计、UBA 分析和溯源。
  6. 借助外部的技术检测也是金融企业重要的手段之一，尤其是金融企业内很多安全产品采购自专业安全公司，第三方的安全检测报告和背书就显得很有必要。实际上，这还涉及到出现风险后责任划分的问题。

技术实现：

1. 网络隔离
   1. 目的：训练中心系统内网和外网之间，不相连，不相通，相互断开，同时在需要的时候能够做到两者间的信息交换。全面断开TCP/IP和OSI数据模型的所有七层，以消除TCP/IP网络存在的攻击。比防火墙的安全系数更高。
   2. 分类：
      1. 物理隔离，两个网络物理上不互相链接，一般分为内网和外网，客户端一般安装隔离卡，隔离卡分为两种
         1. 数据隔离：硬盘电源接口接主板电源，数据接口接隔离卡。
         2. 电源隔离：硬盘电源接口接隔离卡，数据接口接主板电源。
      2. 逻辑隔离，一般两套或者几套网络共用一套网络设备，在网络设备上做配置，各个网络不能互相访问。这种隔离技术非常不安全，容易泄漏数据
         1. 虚拟局域网VLAN
         2. 虚拟路由和转发
         3. 多协议标签转换（MPLS）
         4. 虚拟交换机
   3. 实现技术



* + 1. 网闸是实现网络隔离又安全的交换数据的最为成熟的产品，包括两个主机和一个固态存储介质，网闸技术实现包括动态断开技术和固态断开技术
  1. 隔离方案
     1. 双机双网，双机双网隔离技术方案是指通过配置两台计算机来分别联接内网和外网环境，再利用移动存储设备来完成数据交互操作，然而这种技术方案会给后期系统维护带来诸多不便，同时还存在成本上升、占用资源等缺点，而且通常效率也无法达到用户的要求。
     2. [双硬盘](https://baike.baidu.com/item/%E5%8F%8C%E7%A1%AC%E7%9B%98/5345210)隔离。双硬盘隔离技术方案的基本思想是通过在原有客户机上添加一块硬盘和隔离卡来实现内网和外网的物理隔离，并通过选择启动内网硬盘或外网硬盘来连接内网或外网网络。由于这种隔离技术方案需要多添加一块硬盘，所以对那些配置要求高的网络而言，就造成了成本浪费，同时频繁的关闭、启动硬盘容易造成硬盘的损坏。
     3. 单[硬盘](https://baike.baidu.com/item/%E7%A1%AC%E7%9B%98)隔离。单硬盘隔离技术方案的实现原理是从物理层上将客户端的单个硬盘分割为公共和安全分区，并分别安装两套系统来实现内网和外网的隔离，这样就可具有较好的可扩展性，但是也存在数据是否安全界定困难、不能同时访问内外两个网络等缺陷。
     4. 集线器级隔离。集线器级隔离技术方案的一个主要特征在客户端只需使用一条网络线就可以部署内网和外网，然后通过远端切换器来选择连接内外双网，避免了客户端要用两条网络线来连接内外网络。
     5. 服务器端隔离。服务器端隔离技术方案的关键内容是在物理上没有数据连通的内外网络下，如何快速分时地处理和传递数据信息，该方案主要是通过采用复杂的软硬件技术手段来在服务器端实现数据信息过滤和传输任务，以达到隔离内外网的目的。
  2. 技术选型

项目采用网闸方式实现训练中心的内部和外部网络隔离，保证网络的安全性。

1. 鉴权
   1. 技术方案
      1. HTTP Basic Authentication 通过Authorization字段，用户密码登录的方式进行验证，缺点：仅仅使用了base64加密，安全系数低
      2. session/cookie，通过服务器端的session和浏览器端的cookie实现，用户登录完成后，将登录信息存储在session中，session信息会存储在cookie中，每次调用时候会通过cookie里的信息鉴权，缺点：
      3. token，用户登录后，会生成一个token发送给客户端，token可以使用jwt生成，客户端每次请求的时候带着token，服务器根据token验证用户。
      4. OAuth2.0，目前主流的用户身份验证和授权方式，
         1. 授权码模式，安全性高，流程复杂，要求第三方必须有服务器
         2. 密码模式，不安全，需要高度信任第三方
         3. 隐式授权模式，流程简单，不安全，token有效期端
         4. 客户端凭证模式，授权维度是应用维度，使用于应用维度的共享资源
   2. 技术选型

项目采用OAuth2的授权码模式提供鉴权服务，安全性高，保证了项目数据访问的安全性。

1. 开放网关

外网网关，网关实现了系统验签、加密、鉴权等多种公共安全策略，保证外网访问系统时的安全。

1. 内管网关

内网网关实现了系统鉴权、防攻击、限流等多种公共安全策略，一定程度保证系统的安全性。

1. 数据分类分级
   1. 目的：是确定数据保护和利用平衡点的一个重要依据
   2. 标准：按照法律法规，对组织的价值、关键性和敏感性进行分类
      1. ISO/IEC27001:2013
      2. NIST Special Publication 1500-2[2]
      3. 贵州省，在主题、行业和服务三个维度进行分类，每个分类根据需要分为小类
      4. 银监会，从业务线触发，首先对业务细分，然后对数据细分
      5. 全国金融标准化委员会，数据安全遭到破坏后可能造成的影响是确定数据安全级别的重要判断依据，主要考虑影响对象和影响程度
      6. 工业和信息化部科技司
   3. 用途
      1. 满足合规要求
      2. 满足企业自身运营要去
   4. 数据分级
      1. 分级属于数据安全领域，表示了数据的敏感级
      2. 分级方式
         1. 梳理敏感数据域
         2. 元数据归属数据域
         3. 指定数据类别
         4. 指定敏感等级
         5. 给数据归类归级
      3. 系统实现
         1. 根据系统数据的敏感程度，对每个表字段进行敏感度分级，控制每个敏感等级的访问权限，保证数据的安全性。
   5. 数据分类
      1. 从业务角度分类，比如政府数据、企业数据、平台数据
      2. 从用途角度分类，比如训练数据集、生产数据集
2. 数据脱敏
   1. 目的：屏蔽敏感数据，如身份证、卡号、邮箱地址、手机号等，保护用户隐私安全
   2. 脱敏策略
      1. 脱敏函数
         1. 部分脱敏
         2. 全脱敏
      2. 脱敏列
      3. 生效条件
3. 流程：动态数据脱敏，是在查询语句执行过程中，根据生效条件是否满足，实现实时的脱敏处理。生效条件，通常是针对当前用户角色的判断。敏感数据的可见范围，即是针对不同用户预设的。系统管理员，具有最高权限，任何时刻对任何表的任何字段都可见。确定受限制用户角色，是创建脱敏策略的第一步
4. 系统实现：对系统内的手机号等敏感字段进行脱敏，保证用户隐私和数据安全。
5. 联邦学习
   1. 概念：联邦机器学习是一个机器学习框架，可以帮助多个机构在满足隐私保护、数据安全和政府法规下，进行数据使用和建模
   2. 背景：当前的人工智能，是数据驱动的，人工智能的发展主要基于机器学习技术，难以脱离数据独自前进
   3. 目的：为了解决数据孤岛和数据隐私问题。避免了非授权的数据扩散
   4. 分类
      1. 横向联邦学习：特征对齐的联邦学习，也就是两个数据集之间数据特征重叠部分比较多，用户重叠部分比较少
      2. 纵向联邦学习：样本对齐的联邦学习
      3. 联邦迁移学习：两个数据集数据特征和用户重叠都比较少
   5. 开源框架：
      1. FATE，是微众银行AI团队自主开发的开源联邦学习框架。为联邦生态提供了安全计算框架，安全底层支持同态加密、秘密共享，哈希散列等多种多方安全计算机制。
      2. TensorFlow FeDerated 是谷歌开源的联邦学习框架。是建立在TensorFlow上的，方便基于分布式数据做全局训练。
   6. 联邦学习和数据中台的区别：数据中台可以为联邦学习提供良好的数据环境，但是不是必备条件。数据中台是对数据的分层和解耦，联邦学习是分布式的机器学习框架。
   7. 技术选型：

系统选择FATE联邦学习框架，保证政府数据集的安全性，不被其他用户访问。其他用户的数据集之间如果需要共同生成模型，也可以保证每个用户数据集之间的安全性。

1. 防火墙
2. 入侵检测
3. 入侵防御
4. 安全审计
5. HTTPS
6. SSL加密网关
7. 金融密码机
8. 服务器密码机