

【大数据分析平台】应用

总体设计说明书

**2017年2月**

**文档修订记录**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本号 | 日期 | 简要说明 | 修订人 | 审核人 |
| V1.0 | 2017-02-22 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

目 录

[一、 引言 3](#_Toc475543445)

[1. 概述 3](#_Toc475543446)

[2. 术语定义 3](#_Toc475543447)

[3. 参考文档 4](#_Toc475543448)

[二、 平台设计 5](#_Toc475543449)

[1. 数据交互 5](#_Toc475543450)

[2. 逻辑架构 6](#_Toc475543451)

[3. 物理架构 7](#_Toc475543452)

[三、 数据库设计 9](#_Toc475543453)

[1. 内部软件数据结构 9](#_Toc475543454)

[2. 全局数据结构 9](#_Toc475543455)

[3. 临时数据结构 9](#_Toc475543456)

[4. 数据库说明 9](#_Toc475543457)

[四、 用户接口设计 10](#_Toc475543458)

[1. 登陆画面 10](#_Toc475543459)

[2. 客户查询 12](#_Toc475543460)

[3. 资金链 13](#_Toc475543461)

[4. 担保圈 13](#_Toc475543462)

[5. 客户图谱 13](#_Toc475543463)

[五、 运行设计 15](#_Toc475543464)

[5. 总体要求 15](#_Toc475543465)

[6. 数据库备份及清理设计 15](#_Toc475543466)

[7. 文件系统备份及清理设计 15](#_Toc475543467)

[8. 批量设计 15](#_Toc475543468)

[9. 监控设计 16](#_Toc475543469)

[六、 接口设计 17](#_Toc475543470)

[七、 安全设计 18](#_Toc475543471)

[功能性安全 18](#_Toc475543472)

[数据安全 18](#_Toc475543473)

[八、 模块设计 20](#_Toc475543474)

[九、 附件 21](#_Toc475543475)

# 引言

上海银行大数据分析平台是基于Teradata数据探索平台Aster Data 以及分析和可视化平台AppCenter建立的大数据分析应用，通过App的形式实现快速客户查询、资金链查询、担保查询以及客户关系图谱查询功能。

## 概述

在“大数据时代”，银行所面临的竞争不仅仅来自于同行业内部，外部的挑战也日益严峻。[互联网](http://tech.hexun.com)、电子商务等新兴企业在产品创新能力、市场敏感度和“大数据”处理经验等方面都拥有明显的优势，一旦涉足金融领域，将对银行形成较大的威胁。这些挑战主要体现在如下两个方面：对公客户方面，聚焦中小企业贷款市场上，银行与互联网起家的小额贷款公司开展竞争；零售客服方面，在互联网支付中，网银支付所占比重越来越低，这使得银行越来越难以知道客户的消费行为；互联网融资模式的出现，在未来可能会超过以银行为中心的间接融资和以交易所为中心的直接融资模式，这会使得银行逐渐被边缘化。

大数据是从线上线下资源中收集到的大量商业信息，这些信息包括网站、社会网络、手机APPs、软件、文件、电脑记录、传感网络等信息，越来越多的企业从数据的分析中获利，对大数据的分析正在帮助各种企业快速成长。然而，海量数据的爆炸，最大意义上不在于其数量，而在于大数据能够帮助企业做什么。不少大公司目前仅能从他们所拥有的海量数据的初级数据分析中获利。

提到大数据，人们经常用海量化(Volume)、多样化(Variety)、快速化(Velocity)来描述大数据，常把它想像成高风险、耗资巨大、运行耗时多的IT投资。数据容量增长的速度大大超过了硬件技术的发展速度，引发了数据存储和数据处理的危机。结构化数据、非结构化数据以及多结构化数据的数据构成，对传统的数据分析处理算法和软件提出了挑战。

所以，本次选取资金链，担保圈作为主要分析对象，另外还有客户的图谱分析。借助ASTER大数据分析的特性，整合我行现有的数据，开展了相关专题的大数据分析。

## 术语定义

* 担保圈是指多家企业通过多种担保业务而形成的以担保关系为链条的特殊利益体。

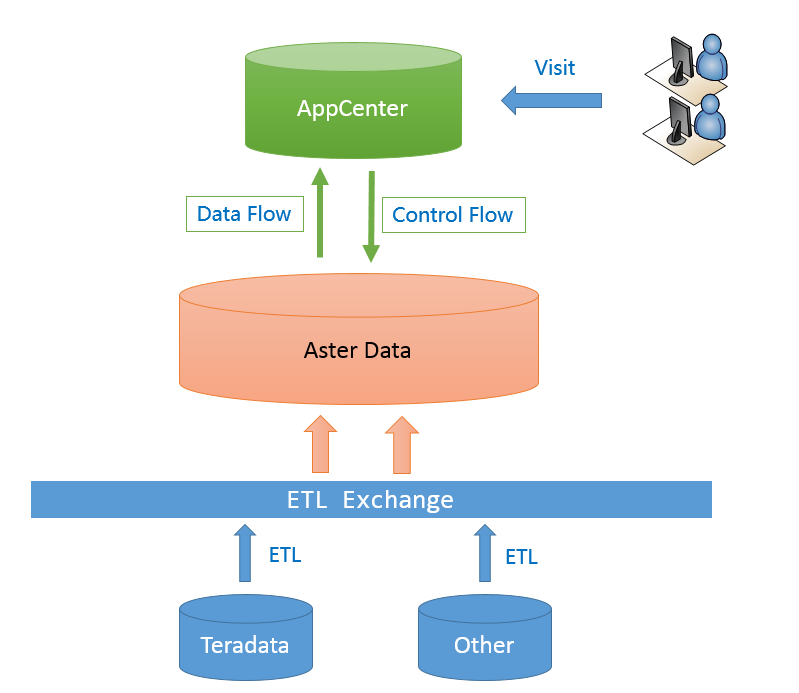
## 参考文档

* 《大数据分析平台应用项目业务需求书》

# 应用架构设计

## 数据交互

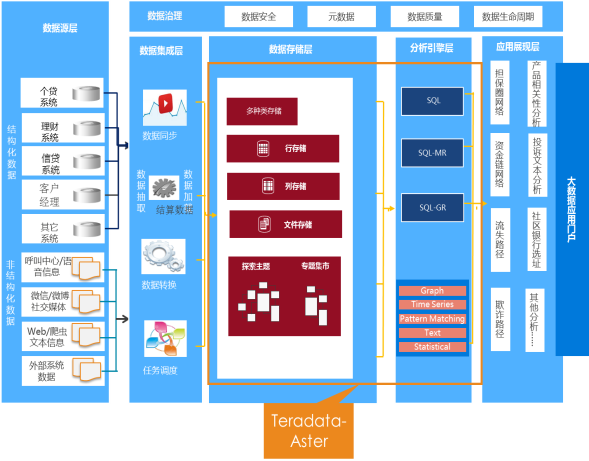
大数据分析平台数据交互如下：



ASTER大数据分析平台的数据来源主要来自于Teradata数据仓库，数据通过ASTER内置的分析引擎，将计算的结果通过AppCenter反馈给用户。

## 逻辑架构

大数据分析平台逻辑架构如下：



**数据源层:**该层支持传统业务系统的结构化数据，也涵盖客户文本、音频、视频等非结构化数据，还包括对外部数据（如移动互联网、社交网络）的支持。

**数据集成层：**该层包含对数据源层的事件同步、数据抽取、数据传输、数据转换、以及数据加载到大数据平台等各种处理。数据集成层通过各种处理，将异构的、结构化和非结构化的、实时和批量的数据，集中到统一数据中心，整合成为集中、一致、360度的数据视图。

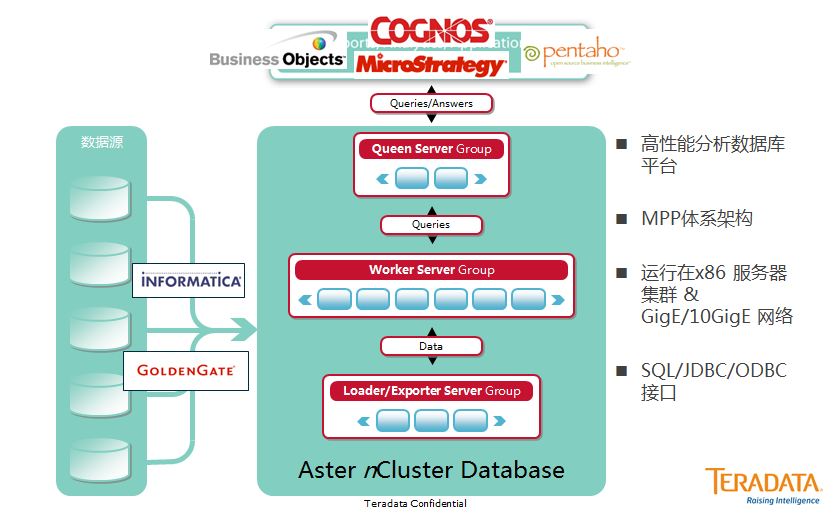
**数据存储层：**该层包含了对结构化数据、半/非结构化数据、实时数据等的贴源存储和预处理，以及根据探索不同的业务专题进行数据的分类统一存储。数据存储层可分数据汇总模型（DW）和ST数据应用模型（ST）。

**分析引擎层：**建立各类业务探索模型，通过各种大数据分析探索分析算法等，分析客户、产品、渠道、营销等多个域的数据。分析发现业务价值，以支撑上层数据应用的开发。

**数据应用展现层：**在分析引擎层的基础上，建设各种创新业务应用，包括对内、对外数据分析专题或应用，从而实现大数据的商业价值。这些应用包含了渠道管理分析、理赔管理分析、运营管理分析、风险管理分析、客户管理分析、财务管理分析、营销管理分析等。这些应用及分析的结果可以通过报表、网页等形式展现，满足决策分析支持、内外部客户服务及大数据营销的需要

## 物理架构

Aster Data的数据库体系结构如下图所示：



在Aster Data的系统中包含Queen节点、Worker节点、Loader节点和Exporter节点，其中Queen节点和Worker节点是必须包含的，Loader节点和Exporter节点是可选的。

Queen节点：

管理系统、配置、语法和错误处理；

提供查询结构，包括SQL/ODBC/JDBC/OleDB；

提供全局的查询优化器，具体步骤：

解析SQL声明，并传递给Planer；

Planer生成一整套子查询供处理；

Executor将处理处理子查询，并完成结果的聚合。

Worker节点：

本地化查询优化。

存储数据，与Queen节点和其他Worker节点进行通讯；

处理Queen节点分配的任务（执行查询，复制，存储均衡，过程均衡）。

Loader节点：

从客户接收新数据（如nCluster\_loader，Informatica等）；

将这些数据进行合适的分段；

将这些分段数据分配给相应的Worker节点。

Exporter节点：

负责Aster数据库数据的导出至Exporter节点；

负责将导出数据分发至其它平台。

在我行，系统建设在初级阶段，为了减少设备投入，当前仅采用必须的Queen节点和Worker节点。ETL服务器和AppCenter服务器，在物理上使用Queen节点。

# 数据库设计

## 内部软件数据结构

|  |  |  |  |
| --- | --- | --- | --- |
| 组件或框架 | 提供者 | 收益 | 存储位置 |
| 名称和版本 | 开发商、其他项目、开放代码等 | 项目的收益 | 可以找到可重用组件或框架的URL |
| Aster | Teradata | 1）高扩展性的MPP关系数据库，支持传统关系型数据分析，支持高性能高可扩展性  2）基于MapReduce引擎，很好的支持非结构化数据的存储及分析  3）200多个算法函数，支持多种高级数据挖掘分析 | https://queen\_node:8080 |
| Aster AppCenter | Teradata | 担保圈、资金链可视化能力 | https://appcenter\_node/ |

## 全局数据结构

本系统是基于数据库的应用，该部分不适用。

## 临时数据结构

本系统是基于数据库的应用，该部分不适用。

## 数据库说明

Aster 是基于1Queen多Worker和多个loader节点组成的数据库集群

# 应用设计

## 登陆画面

在浏览器中输入：IP地址(由科技部门分配。如：生产环境10.232.33.68),

网站自动跳转进入安全模式的连接:

[https://10.232.33.68/appserver/portal/login](https://10.20.158.39/appserver/portal/login)

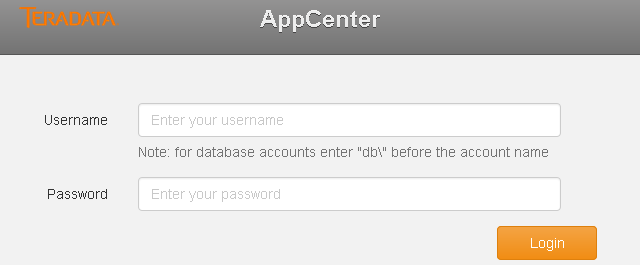
首次登录需要将其加入信任。

|  |
| --- |
| 提示安全风险时，请将AppCenter网址加入信任列表 |

以Firefox为例，永久添加AppCenter网址到信任区域，其它浏览器操作类似



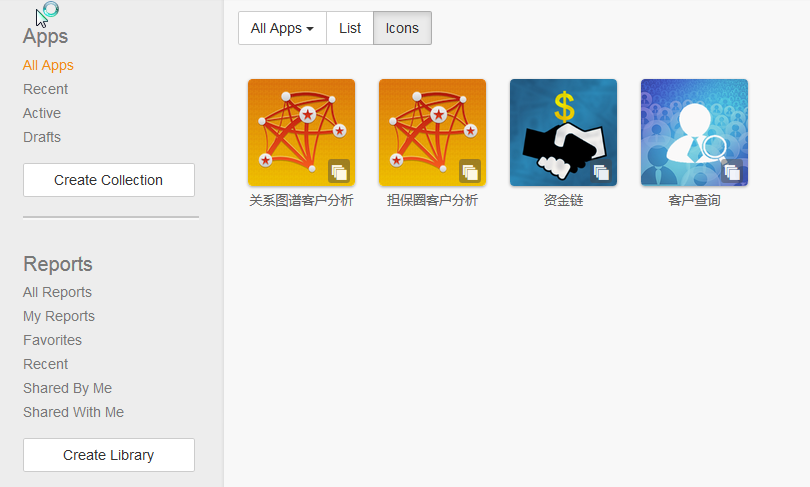
成功进入登录页面后，键入用户、密码等信息。请留意用户中的斜线，其中用户名前辍db指从数据库建立的用户，还可用域用户登录，具体可通过管理员分配。



|  |  |
| --- | --- |
| Username用户名 | Password密码 |
| db\用户名 | 密码 |

大数据分析平台的功能在画面中以图标方式显示，不同的角色进入后，仅能看到和该角色相关的功能点。

整个画面风格如下:



点击进入相应的功能。点击后，将进入功能录入画面。

## 客户查询

该APP的设计主要是作为一个公用的APP，对所有用户开放，根据模糊信息精确查找需要分析的客户

输入：

|  |  |  |
| --- | --- | --- |
| 选项 | 说明 | 示例输入 |
| **\***Title | App执行任务的名称 | 建议使用“查询的信息” |
| Tags | App执行任务的标签，作为任务关键字以便搜索 | 建议描述分析的客户名 |
| \*客户信息 | 下拉选择查询对公或是对私客户 | 仅可以选择一个 |
| \*名称 | 客户名称 | 对公对私均要输入名称：万尚地产发展有限公司 |
| 证件类型 | 若选择的是对私客户，建议输入证件类型 |  |
| 证件号码 | 若选择的是对私客户，建议输入证件号码 |  |

输出：客户信息表

## 资金链

该款App提供资金链查询，即通过输入信贷客户号，最小交易金额以及最大交易金额，搜索满足条件的交易。

输入：

|  |  |  |
| --- | --- | --- |
| **选项** | **说明** | **示例输入** |
| **\***Title | App执行任务的名称 | 349985 |
| Tags | App执行任务的标签，用户可以自定义标签来标识本次任务的意义 | 349985客户2层资金网络 |
| **\***信贷客户号 | 信贷客户号 | 349985 |
| **\***最小交易金额 | 交易金额范围，最小值 | 1000000 |
| **\***最大交易金额 | 交易金额范围，最大值 | 10000000 |

输出：查询客户的资金网络图

## 担保圈

该功能可以查询客户的担保情况。

输入：

|  |  |  |  |
| --- | --- | --- | --- |
| **栏位** | **必填** | **说明** | **示例输入** |
| Title | **Y** | 报告名称。有了名称后，可以用该名字识别这份报告，以后可以在报告集中搜索查找。 | 建议用名单名称+查询条件描述 |
| Tags |  | 标签。作为关键字搜索报告 |  |
| ASTER客户键 | **Y** | Aster客户键可以通过客户查询功能获取 | 0\_上海博应信息技术有限公司\_8624437 |

输出：担保网络图

## 客户图谱

该功能是客户的关系图，描述客户与客户的上下三层关系。该主题分为两个应用模块

* 关系图谱客户分析：主要是查询指定客户的关系
* 分行关系图谱：展示中心客户是该分行的分行所有客户关系

关系图谱客户分析输入：

|  |  |  |
| --- | --- | --- |
| **栏位** | **必填** | **说明** |
| Title | **Y** | 报告名称。有了名称后，可以用该名字识别这份报告，以后可以在报告集中搜索查找。 |
| Tags |  | 标签。作为关键字搜索报告 |
| 信贷客户号 | **Y** | 10292662 |

输出：该客户的网络图

分行关系图谱输入：

|  |  |  |  |
| --- | --- | --- | --- |
| **栏位** | **必填** | **说明** | **示例输入** |
| Title | **Y** | 报告名称。有了名称后，可以用该名字识别这份报告，以后可以在报告集中搜索查找。 | 建议用客户名称+日期 |
| Tags |  | 标签。作为关键字搜索报告 |  |
| 图谱类型 | **Y** | 全图谱：中心信贷客户在该支行的所有关系  违约客户图谱：选出包含违约客户的关系图谱（全图谱子集） |  |

输出：该支行的所有客户的网络图

# 运行设计

## 总体要求

|  |  |  |
| --- | --- | --- |
| 指标名称 | 指标值 | 备注 |
| 应用的服务时间 | 7\*24 |  |
| 故障恢复时间（RTO） | <=1.5 |  |
| 数据恢复时间(RPO) | 30 |  |
| 最大并发数 | 100 |  |
| 每秒处理事务数（TPS） | 无 |  |
| 交易响应时间 | 无 | 交易在本系统内处理时间 |
| 在线数据保留时间 | 三年 | 根据数据仓库保留时间 |
|  |  |  |

## 数据库备份及清理设计

【本章节主要阐述应用系统上线后如何进行数据库的备份、数据清理】

数据仓库备份：分析平台每天根据仓库数据实时计算结果，对数据不做保留。

数据清理：每日批量结束自动清理SDATA，PTEMP库下相应的表

其中：SDATA保留30天以内表

PTEMP则按照规则，删除计算过程中间表（AppCenter运行产生）

## 文件系统备份及清理设计

文件备份：暂时文件没有做备份，后续按照双活设计，即有2个QUEEN节点互相备份。

## 批量设计

【本章节主要阐述本应用系统有哪几类批量任务，批量任务的设计思路、批量任务的调度，执行频率，执行方式，批量任务之间的关系（并行/串行）等。批量的详细内容可参考《批量说明文档》。

详细参考



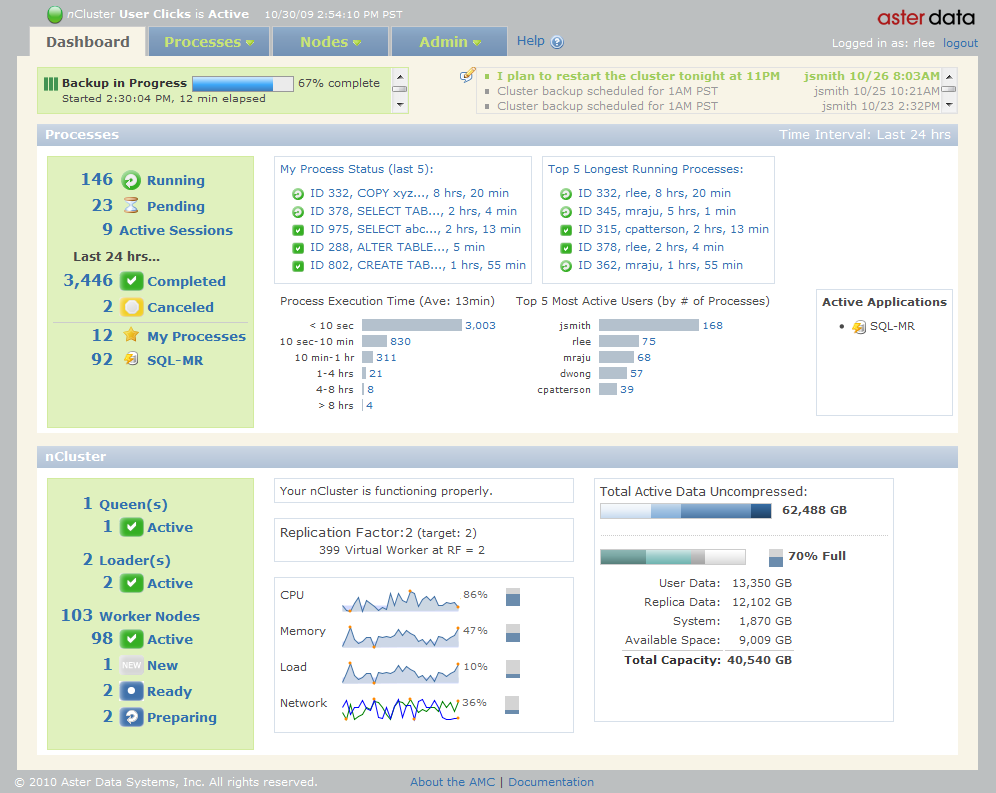
## 监控设计

【本章节主要阐述本应用进程、日志、交易的运行监控要求、监控方式、监控方法等】

Aster Data系统提供统一的实时管理监控平台，对于整个硬件、软件、操作系统、数据库、网络、存储等进行统一的管理、监控、优化等系统管理功能。

Aster Data系统具备良好的系统管理方式、方法与操作界面；提供人性化的图形界面，展示数据库的整体状态及SQL/分析查询的运行状态等；提供通过界面查看相应的系统管理功能。

Aster Data提供了在线监控和管理平台Aster Management Console (AMC)，使系统监控和管理简单方便。Aster Management Console (AMC)可实时监控节点、SQL、作业等元素的状态，并根据配置的规则进行实时预警，Aster Data的系统管理监控界面如下图所示：



# 接口设计

【简要描述本应用内部子应用之间，与外部应用之间的交互接口，包括对外提供接口和调用外部接口，说明通讯方向、通讯协议、交互信息、接口规范等，优先图示说明。

接口设计需要考虑服务请求的注册，否则不能使用该接口。

接口设计需要考虑限流控制，自动或手动方式。

接口设计需要考虑防重发机制。】

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 编号 | 源应用 | 目的应用 | 交互信息 | 通讯协议 | 接口规范 |
| 1 | 数据仓库 | 审计集市 | 视图 | 无 | 汇聚字典接口 |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |

# 安全设计

【本章节对可行性方案阶段的安全设计方案予以细化补充, 主要包括但不限于以下内容:

* 客户、用户密码生成、存储、验证安全设计方案；
* 用户登录、认证、权限的安全控管设计；
* 敏感信息传输、存放安全设计；
* 外联文件及报文传输安全设计方案】

审计二期系统的安全性主要是通过功能性安全、数据安全、审计追踪等多方面配合，从而达到预定的信息安全目的。

## 功能性安全

按照团队管理需要配置角色和角色可用资源；系统采用标准RBAC权限模型控制功能访问权限。用户没有权限的内容，将不会显示在界面上。

## 数据安全

数据权限控制

系统提供了多种方式进行数据权限的控制：机构数据过滤、目录进入的权限控制、共享和验证、在线授权机制；

机构数据过滤（记录级控制）

* 在有“机构”字段的数据里，用户只能看到有权限的机构数据；
* 通过给用户分配归属部门的操作，可以让用户获得归属部门管辖机构的权限。

目录权限控制（目录级控制）

* 有些数据仅提供给部分用户查看，你可以设置一个目录存放这些数据，再通过标准目录与部门和角色绑定，没有权限的用户无法进入这个目录；

共享和验证（文件级控制）

* 为了尊重文件拥有者的权利，只有拥有者才能将DAP模型、工作区里的文件、等成果文件共享给其他人。系统管理员可以强行更改成果文件的拥有者。

在线授权机制（事件级控制）

* 有些数据需要有授权才能查看或者导出，用户可以在线向负责人提出申请，负责人会在自己的消息窗中对你的申请进行批复。
* 支持的在线授权事项有数据导出、敏感字段访问、给用户分配角色、用户归属到部门等；
* 授权人是上级单位或者同级单位的用户。
* 不需要的在线授权事项可以在console中“关闭”；

数据存储安全

* 对系统数据实施每日备份，一旦发生数据损坏，可进行数据恢复。
* 系统内敏感数据如登录密码等均采用密文存放，防止泄密。
* 系统配置文件中的敏感数据均采用密文存放，数据库连接等均使用应用服务器管理，应用本身不保存这方面的信息。

数据传输安全

信息在网络上进行传输时都可通过加密处理，以防止信息在网络上被窃听使用。

审计追踪

系统按不同层次生成业务操作日志，系统运行日志，系统错误日志，开发调试日志，按要求保留系统运行操作相关痕迹，供审计追踪用。

# 模块设计

【简要描述本应用实现的功能，优先图示说明，可采用列表，如功能过多可以在八、附件中说明】

参见附件

# 附件

【是对上述模块设计的一个补充】