

XXXXXX平台

<总体设计>

|  |  |
| --- | --- |
| **版本** | V1.0 |
| **组织** | 京东科技 |
| **发布日期** | YYYY-MM-DD |

京东科技信息技术有限公司

**目录**

[1. 法律声明 3](#_Toc87898980)

[2. 引言 5](#_Toc87898981)

[2.1. 目的 5](#_Toc87898982)

[2.2. 建设目标 5](#_Toc87898983)

[2.3. 建设原则 5](#_Toc87898984)

[2.4. 参考资料 5](#_Toc87898985)

[3. 项目建设计划 5](#_Toc87898986)

[4. 总体架构方案 6](#_Toc87898987)

[4.1. 应用架构说明 6](#_Toc87898988)

[4.2. 系统技术架构图 6](#_Toc87898989)

[4.3. 系统数据架构及支持业务 7](#_Toc87898990)

[4.4. 部署架构 8](#_Toc87898991)

[4.5. 功能模块清单 10](#_Toc87898992)

[4.6. 系统硬件 10](#_Toc87898993)

[4.7. 系统软件 12](#_Toc87898994)

[4.8. 接口及兼容性说明 13](#_Toc87898995)

[4.9. 安全设计 13](#_Toc87898996)

# 法律声明

京东科技提醒您在阅读或使⽤本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使⽤本⽂档，您的阅读或使⽤⾏为将被视为对本声明全部内容的认可。

1. 您应当通过京东科技⽹站或京东科技授权通道下载、获取本⽂档，且仅能⽤于自身的合法合规的业务活动。本⽂档的内容视为京东科技的保密信息，您应当严格遵守保密义务；未经京东科技事先书⾯同意，您不得向任何第三⽅披露本手册内容或提供给任何第三方使⽤。

2. 未经京东科技事先书⾯许可，任何单位、公司或个人不得擅⾃摘抄、翻译、复制本文档内容的部分或全部，不得以任何⽅式或途径进⾏传播和宣传。

3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。京东科技保留在没有任何通知或者提⽰下对本⽂档的内容进⾏修改的权利，并在京东科技网站或京东科技授权通道中不时发布更新后的用户⽂档。您应当实时关注用户⽂档的版本变更并通过京东科技网站或京东科技授权渠道下载、获取最新版的用户⽂档。

4. 本文档仅作为用户使⽤京东科技产品及服务的参考性指引，京东科技以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本⽂档。京东科技在现有技术的基础上尽最⼤努⼒提供相应的介绍及操作指引，但京东科技在此明确声明对本⽂档内容的准确性、完整性、适⽤性、可靠性等不作任何明⽰或暗⽰的保证。任何单位、公司或个⼈因为下载、使⽤或信赖本⽂档而发⽣任何差错或经济损失的，京东科技不承担任何法律责任。在任何情况下，京东科技均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使⽤或信赖本⽂档而遭受的利润损失，承担责任（即使京东科技已被告知该等损失的可能性）。

5. 京东科技⽂档中所有内容，包括但不限于图⽚、架构设计、⻚⾯布局、⽂字描述，均由京东科技和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。⾮经京东科技和/或其关联公司书⾯同意，任何⼈不得擅⾃使⽤、修改、复制、公开传播、改变、散布、发⾏或公开发表京东科技网站、产品程序或内容。此外，未经京东科技事先书⾯同意，任何⼈不得为了任何营销、⼴告、促销或其他⽬的使⽤、公布或复制京东科技的名称（包括但不限于单独为或以组合形式包含“京东科技”、“JDT”等京东科技和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案表示、标志、标识或通过特定描述使第三⽅能够识别京东科技和/或其关联公司）。

6. 如若发现本⽂档存在任何错误，请与京东科技取得直接联系。

# 引言

*【内容待填入】*

## 目的

*【内容待填入】*

## 建设目标

*【样例】大数据分析平台建设将基于现有场景，对应用场景进行合理评估，从功能需求、运行环境、安全需求等维度，确定本项目的平台需求规划和实现方法。坚持以统一标准，互联互通；经济安全，保驾护航；灵活扩展，长远规划；逐步治理，持续优化为原则，采取安全性高的访问认证机制。同时在平台建设中要充分重视系统自身的安全性以及其他环节的安全性，成为大数据分析平台建设和应用的基础保障。*

## 建设原则

*【内容待填入】*

## 术语定义

*【内容待填入】*

## 参考资料

*【内容待填入】*

# 项目建设计划

*【内容待填入】*

# 总体架构方案

## 应用架构

*【样例】从应用角度，整个平台分为数据源、数据计算与存储、数据仓库、数据治理、数据开发、数据服务、数据应用这个7个模块。*

*附加应用架构图*

## 技术架构

*【样例】从技术架构层次分为数据源、数据采集、数据存储、数据处理、技术框架、数据分发、数据服务、数据应用、数据开发支撑。主要采用Hadoop、Hive生态相关的分布式技术，满足计算能力线性扩展、数据汇总能力。*

*附加技术架构图*

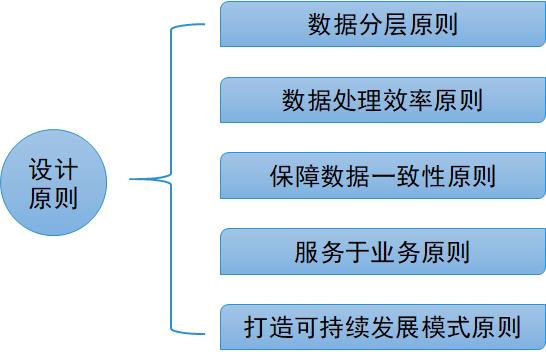
## 数据架构及支持业务

*【样例】在数据全生命周期过程中，数据架构立足于解决从数据产生，到中间的流转、整合、分析应用、归档和消亡等环节的数据的存储方式、模型策略、访问机制等问题。*

*附加数据架构图*

### 设计原则

数据架构设计，应遵循以下原则进行处理：



### 数据架构分层设计

*【样例】通过借鉴业内先进经验，基于业务及技术现状，以此为基础，遵守数据架构的设计原则，设计数据仓库的数据架构。*

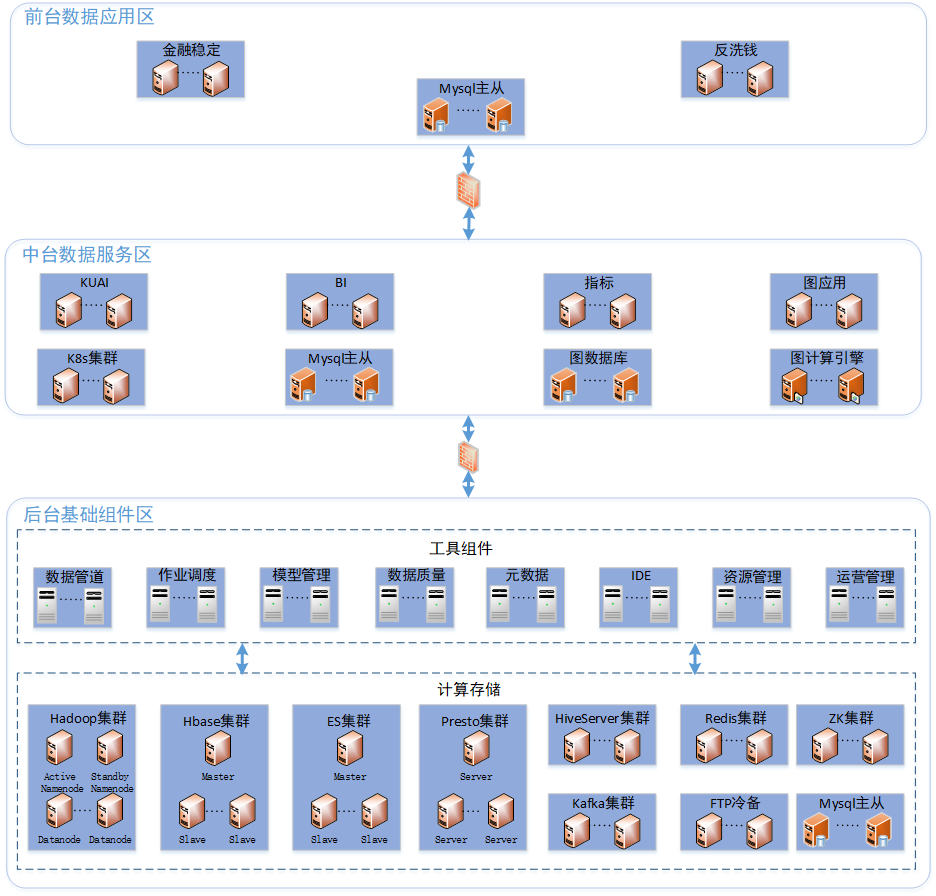
*主要包括如下几层：*

### 支撑业务

*【样例】基于上述设计原则和数据分层设计，实现大数据平台具备支持金融稳定，国际收支，国库，调查统计，反洗钱，货币信贷等业务系统的数据服务能力。*

## 部署架构

*【样例】大数据平台部署的网络拓扑示意图如下所示：*



*附加部署架构图*

*【样例】大数据平台部署在大数据区，主要分为大数据应用服务和大数据集群、消息中间件。大数据应用服务部署大数据管理平台、离线计算模块、数据开发IDE等web应用；大数据集群包括大数据的管理节点服务器、数据节点服务器；管理节点服务器主要部署NameNode、ResourceManager、HMaster等组件，离线计算节点部署DataNode、NodeManager、spark、mr、HiveMetastore、HiveServer2、Presto等组件。*

## 功能模块清单

|  |  |  |
| --- | --- | --- |
| **产品大类** | **产品名称** | **产品描述** |
| *数仓管理* | *工程管理* | *管理业务线，例如反洗钱，金稳等* |
| *数据管道* | *支持多种数据格式，定时调度，准确快速采集多源数据* |
| *数据模型* | *数据建模工具* |
| *作业中心* | *可靠、高效、安全的作业调度平台* |
| *资源平台* | *集群管理* | *进行集群定义* |
| *权限管理* | *管理用户库表权限* |

## 系统硬件

### 系统硬件要求

*【样例】大数据平台系统目标架构硬件选型，必须支持在以PC服务器集群为架构基础的运行环境下稳定运行。*

### 生产环境配置方案

*【样例】生产环境满足客户项目大数据基础平台、数据研发平台、数据应用能力、客服数据门户、数据仓库、数据治理功能和性能要求，推荐申请91台主机，包括：物理主机16台，虚拟主机75台，75台虚拟主机可以通过9台物理主机进行虚拟化隔离，所有的服务器推荐采购4个万兆网卡。具体硬件推荐配置如下：*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **资源类型** | **资源型号** | **数量** | | |
| **应用服务主机（做虚拟化）** | **大数据计算节点** | **数据库服务节点** |
| *1* | *CPU* | *Intel ® Xeon ® Silver 4110* | *464（总）核* | *928（总）核* | *200（总）核* |
| *Intel ® Xeon ® Gold 5120* |
| *Intel ® Xeon ® Gold 5117* |
| *Intel ® Xeon ® Gold 5118* |
| *2* | *内存* | *DDR4 2666* | *2560(总)G* | *4608(总)G* | *1024（总）* |
| *3* | *硬盘* | *SATA* | *8.7T* | *4.4T+12\*2.2T* | *2T* |
| *5* | *网卡* | *10GB* | *4* | *4* | *4* |
| *6* | *磁盘阵列* | *RAID 版本* | *支持 RAID 10* | *支持 1组RAID 0 + 1组RAID 1* | *支持 RAID 10* |
| *7* | *建议主机* | *物理机数量* | *7* | *9* | *4* |

## 系统软件

*【样例】京东大数据平台产品，完全基于J2EE架构的应用，并可以独立部署，推荐Centos7.6以上版本操作系统和MySQL5.7以上版本数据库，Tomcat8.0和JDK1.8版本；支持与主流数据源集成，如ORACLE、DB2、MySQL等；自带容器用于部署和运行，也支持在主流web容器（Tomcat等）上部署运行。支持虚拟化硬件，软件需求，满足使用现有的硬件设备、数据库软件、应用服务器软件、客户端操作系统和浏览器等资源。*

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 产品 | 软件及版本号 | 备注 |
| *1* | *操作系统* | *centos7.5-1804* |  |
| *2* | *数据库* | *数据库:MySQL 5.7.22* |
| *3* | *中间件* | *Tomcat:7.0.61*  *JDK版本:1.8.0\_161*  *Nginx:1.16.0* |
| *4* | *开源程序* | *Hadoop2.7.3*  *Hbase1.2.6*  *Hive2.3.2*  *Spark2.4.3*  *Presto*  *Kafka1.0.0*  *Flink1.7.2*  *Elasticsearch5.5*  *Azkaban:3.43*  *SFTP:OpenSSH\_5.3p1,*  *OpenSSL 1.0.1e-fips*  *Zookeeper v3.4.11* |

## 接口及兼容性说明

*【样例】大数据平台提供调度系统的接口和presto查询接口供大数据平台内部业务系统使用。*

*调度系统接口采用dubbo接口形式提供使用，接口调用者需要引用exts-common-api.1.0-SNAPSHOT.jar；*

*Presto查询接口采用http接口形式提供使用，接口调用者配置好http接口的服务器地址即可实现远程调用。*

## 安全设计

*【样例】平台提供涉密数据保护功能，数据安全保护策略通过用户认证与传输通道数据脱敏, 保证用户使用过程中的数据安全, 全方位无死角的防篡改、防盗用、防截取, 为平台内部数据保驾护航。*

### 数据采集

*【样例】数据采集过程中，可对涉密信息部分的字段进行脱敏配置，以保证涉密信息的安全，在数据使用过程中如有需要，可通过对应的解密函数恢复脱敏信息。大数据平台的数据与办公网隔离，非授权无法访问。*

### 数据存储

*【样例】大数据存储与计算均基于Hadoop分布式系统实现。Hadoop系统在物理架构单独划分一个网络区域。其他所有产品和服务均没有与Hadoop系统共享存储，并从网络上也与Hadoop系统做了天然隔离操作。Hadoop系统作为大数据工业界的事实标准，对于存储的安全考虑，本身已实现了高容错、高可用的系统架构。物理上，单集群数据全部放在Hadoop系统的DataNode结点上；逻辑上，又划分为不同的数据集市做业务间的互斥和隔离。对于数据权限访问上，自研了一套非常完备的数据安全管理服务，经过多层访问控制，确保数据存储访问的安全。对于涉密数据单独加密存储在涉密数据区，和其他业务数据完全隔离，保证了涉密数据的安全。*

### 数据传输与共享

*【样例】对涉密数据的传输，平台通过AES加密技术在数据传输过程中保证数据的安全，防止数据被拦截篡改。通过AES加密技术，利用平台特定密钥，实现对涉密数据传输的加密。*

### 数据使用

1. *用户账号绑定*

*所有登录大数据分析平台的用户，必须使用发放的唯一账号，所有的数据查询都会对用户所授予的权限进行校验。*

1. *最小化授权策略*

*大数据平台权限系统提供了一个和HDFS原生权限相匹配适应的授权模型管理访问策略, 可以将用户的授权细化到文件级别。用户申请使用数据时，需要详细说明数据的使用场景与用途，并明确操作权限。权限需要经过上级领导、数据生产者、业务负责人、集市负责人等多层严格审批。所有授权记录需有据可查，保证数据最小化的访问与授权可追溯。*