**时代新能源TDP**

**测试大数据平台项目**

**详细设计方案**

**目录**

1. 建设意义 3

2. 建设目标 4

2.1业务目标 4

3. IT目标 6

4. 系统整体架构设计 8

4.1 拓扑架构设计 9

4.2 数据扭转设计 13

4.3 应用层设计 16

4.4 数据层设计 17

4.5 系统接口设计 17

5. TDP平台详细建设方案 33

6.1 平台概述 33

6.2 平台体系建设 33

6.3 功能性设计方案 34

6.4 性能要求 49

名词解释

|  |  |
| --- | --- |
| 术语名词 | 描述 |
| DB | Database数据库 |
| TDP | Testing bigdata platform |
| TES | Testing E system |
|  |  |

# 建设意义

TDP可以将宁德CATL多年积累的数据唤醒，不仅为企业管理好这些海量数据，而且挖掘数据潜在的价值，从而成为企业运营维护系统的亮点之一。TDP的底层数据建设，为支持大数据平台的分析、报表和[其它](http://www.vsharing.com/industry/1625.html)商业智能（BI）功能所需要的各种数据源创建一个中央的位置和永久储存的空间。TDP底层MongoDB数据库的建设意义，有如下几点：

* 使时代新能源企业高层领导能够从全局角度出发，推动企业数据的统一规划，便于业务人员对企业数据的分析与理解。
* 可以形成时代新能源企业的概念模型，帮助时代新能源企业人员更好地理解业务的核心概念和业务之间的关系。
* 帮助设计人员制定出更加合理的数据架构和统一的数据分布图。
* 可以明确各个业务部门之间的关系和在分析应用工作中的主要职责，有利于实现统一的报表体系规范，便于实现时代新能源企业的运营指标分析和统一的口径。
* 形成有效的数据管理体系，保证企业在业务部门众多，内部数据和外部数据复杂的情况下，数据只有唯一事实的特点。
* 为业务人员提供各种报表查询功能，为应用系统提供强大的数据分析功能。
* TDP底层数据库提供加强的商业智能（BI）

利用从各种数据源提供的数据，管理人员和高管们将不再需要凭着有限的数据或他们的直觉做出商业决策。此外,“数据仓库及相关商业智能（BI）可直接用于包括电池测试、库存管理、[财务管理](http://finance.vsharing.com/)、销售这样的业务流程中。”

* TDP底层数据库可节省时间

因为业务用户可以在一个地方快速访问许多数据源，他们就在关键方案上迅速做出知情的决策，而不会用浪费宝贵的时间从多种数据源中检索数据。

不仅如此，业务主管们可以在很少或者根本没有IT的支持下自己查询数据 节约了更多的时间和资金。这意味着商业用户不需要等待IT的出现就能生成报表，而那些在IT努力工作的人员可以做他们最好该做事情维。

* TDP底层数据库能提高数据的质量和一致性

MongoDB数据库的实施包括将数据从众多的数据源系统中转换成共同的格式。由于每个来自各个部门的数据被标准化了，每个部门将会产生与所有其它部门符合的结果。而准确的数据是强大的商业决策的基础。

* TDP底层数据库能提供历史的智慧

MongoDB数据库储存了大量的历史数据，所以你可以通过分析不同的时期和趋势来做出对未来的预测。这些数据通常不能被存储在一个交易型的数据库里或用来从一个交易系统中生成报表。

* TDP底层数据库能创建高的投资回报率

已经安装了数据仓库和完善了商业智能（BI）系统的企业比没有在商业智能（BI）系统和数据仓库投资的企业能产生更多的利润和节约更多的资金。而这应该成为高级管理层快速加入到大数据平台建设这个潮流中的足够理由。

# 建设目标

## 2.1业务目标

MongoDB数据库是在数据库已经大量存在的情况下，为了进一步挖掘数据资源、为了决策需要而产生的，它并不是所谓的“大型数据库”。数据仓库的方案建设的目的，是为TDP上层查询和分析作为基础，由于有较大的冗余，所以需要的存储也较大。为了更好地为上层如电池寿命预测算法等应用服务，MongoDB数据库必须满足如下几点特点：

### 2.1.1高效率

MongoDB数据库的分析数据一般分为日、周、月、季、年等，日为周期的数据要求的效率最高，要求平台能显示到MongoDB数据库的12小时内实时更新数据信息。

### 2.1.2高数据质量

MongoDB数据库所提供的各种信息，保证数据的准确性，但由于数据仓库流程通常分为多个步骤，包括[数据清洗](https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E6%B8%85%E6%B4%97)，装载，查询，展现等等，复杂的架构会更多层次，那么由于[数据源](https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E6%BA%90)有[脏数据](https://baike.baidu.com/item/%E8%84%8F%E6%95%B0%E6%8D%AE)或者代码不严谨，都可以导致数据失真，客户看到错误的信息就可能导致分析出错误的决策，造成损失，而不是效益。所以MongoDB数据库需要避免数据失真、错误的数据信息。

### 2.1.3扩展性

MongoDB数据库需要考虑到了未来3-5年的扩展性，满足能正常存储未来3-5年时代新能源企业增加的数据量。

### 2.1.4面向主题

操作型数据库的数据组织面向[事务处理](https://baike.baidu.com/item/%E4%BA%8B%E5%8A%A1%E5%A4%84%E7%90%86" \t "_blank)任务，各个业务系统之间各自分离，而MongoDB数据库中的数据是按照一定的主题域进行组织的。主题是与[传统数据库](https://baike.baidu.com/item/%E4%BC%A0%E7%BB%9F%E6%95%B0%E6%8D%AE%E5%BA%93)的面向应用相对应的，是一个抽象概念，是在较高层次上将企业信息系统中的数据综合、归类并进行分析利用的抽象。每一个主题对应一个宏观的分析领域。数据仓库排除对于决策无用的数据，提供特定主题的简明视图。

### 2.1.5实现数据的自动化协同

* 消除分散的数据，打通数据之间的通道
* 统一整合，统一访问入口，逐渐消除“信息孤岛”，消除多个访问途径，为用户提供1个访问入口，最终实现业务流程端到端的管理

### 2.1.6灵活高效的数据运营体系

* 打通业务系统之间的数据流程壁垒，具备通过将业务数据系统内部流程外化的能力
* 形成数据定制和数据移动化改造+标准规范的后台业务数据流程平台 (如SAP)的架构体系
* 为商业化业务系统的快速标准化实施和平台的轻量化、高性能提供基础

### 2.1.7提供上层支撑

* 通过对数据执行的多维度分析，提供有参考价值的直观数据；
* 通过TDP平台为整体降低现有测试的数据运行时间提供技术支撑；

## 2.2 IT目标

满足时代新能源企业大数据信息化的企业级应用，MongoDB数据库需要满足如下目标：

### 2.2.1 面向主题

* + MongoDB数据库中的数据是按照一定的主题域进行组织。主题是指用户使用MongoDB数据库进行决策时所关心的重点方面，一个主题通常与多个操作型信息系统相关；
  + 不同主题域相互独立，互不干扰；

### 2.2.2 集成性

* + 数据有来自于分散的操作型数据，将所需数据从原来的数据中抽取出来，进行加工和集成，统一与综合之后才能进入MongoDB数据库；
  + MongoDB数据库中的数据是在对原有分散的数据库[数据抽取](https://baike.baidu.com/item/%E6%95%B0%E6%8D%AE%E6%8A%BD%E5%8F%96)、清理的基础上经过系统加工、汇总和整理得到的，必须消除源数据中的不一致性，以保证MongoDB数据库内的信息是关于整个时代新能源一致的全局信息。
  + MongoDB数据库的数据主要供时代新能源决策分析之用，所涉及的数据操作主要是数据查询，一旦某个数据进入MongoDB数据库以后，一般情况下将被长期保留，也就是MongoDB数据库中一般有大量的查询操作，但修改和删除操作很少，通常只需要定期的加载、刷新。
  + MongoDB数据库中的数据通常包含历史信息，系统记录了时代新能源企业从过去某一时点(如开始应用数据仓库的时点)到当前的各个阶段的信息，通过这些信息，可以对时代新能源企业的发展历程和未来趋势做出定量分析和预测。

### 2.2.3 数据同步性

* + MongoDB数据库主要是为[决策分析](https://baike.baidu.com/item/%E5%86%B3%E7%AD%96%E5%88%86%E6%9E%90)提供数据，所涉及的操作主要是数据的查询；
  + MongoDB数据库中的数据须与TES中的HANA数据库保持一致；

### 2.2.4 随时而变性

* + MongoDB数据库是不可更新，而又随时间不断变化的
  + MongoDB数据库是历史性的，过往的已存在的业务数据一般不会UPDATE
  + 随着时间积累，会逐步增加数据，可能由于数据量的膨胀，而优化体系或者改变[存储结构](https://www.baidu.com/s?wd=%E5%AD%98%E5%82%A8%E7%BB%93%E6%9E%84&tn=44039180_cpr&fenlei=mv6quAkxTZn0IZRqIHckPjm4nH00T1YLm1T4m1Tvn1IWuycLrjc10ZwV5Hcvrjm3rH6sPfKWUMw85HfYnjn4nH6sgvPsT6KdThsqpZwYTjCEQLGCpyw9Uz4Bmy-bIi4WUvYETgN-TLwGUv3EnWTsnjn1rjfd" \t "_blank)，甚至增加层次，这种因为时间引起的变化称之为随时间变化的。MongoDB数据库是随时间不断变化的。

### 2.2.5 汇总性

* + 操作性数据映射成决策可用的格式。

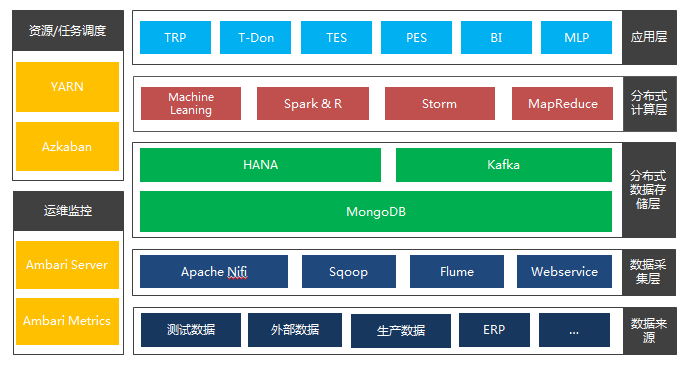
### 2.2.6大容量性

* + [时间序列数据](https://baike.baidu.com/item/%E6%97%B6%E9%97%B4%E5%BA%8F%E5%88%97%E6%95%B0%E6%8D%AE)集合满足大容量性需求

### 2.2.7元数据

* + MongoDB数据库将描述数据的数据保存起来。

# 系统整体架构设计



TDP平台的主要特点就是要打通不同数据之间的协作壁垒，所以，在设计TDP平台时，必须充分研究现有业务数据之间的指令传递及信息共享方式，充分研究当前不同业务数据之间存在的协同规则，明确当前各种需要克服的管理和业务传递过程中的消极因数，并针对性地提出相应的执行规则和连通逻辑，这里强调数据存储规则和数据交互逻辑，是因为数据存储规则和数据交互逻辑是我们固化大数据平台的第一步，任何数据之间的连通问题，归根结底都和缺乏明确的、可执行的规则和逻辑有关。其次，我们应该明确指令传递过程中的技术障碍，也就是说，即使有规则，但仍可能存在人为执行上的不确定性，仍可能存在和作业经验有关的作业效果的差异性，这是因为“规则是死的，但人是活的”，作业人员可以不执行规则、也可以有条件地执行，在对象众多，任务繁杂的大型企业，管理人员根本就无法进行有效的过程监管。所以，我们必须研究各种技术手段，把作业人员的“活的”部分，尽可能地消除掉。这也是流程管理平台设计中关注协同作业效果的重要思路之一。

## 拓扑架构设计

**方案特点**

* 高性价比：能够让企业低投资拥有高性能、经济的网络；
* 简易性：结构简单、安装快速、简单，维护无需配置专职人员；
* 高性能：最低投资做到千兆骨干、百兆接入；
* 可扩展性：灵活的网络架构，能根据用户需要随时扩展，并保护已有投资。

规划方案：

**网络安全规划**

网络安全是整个系统安全运行的基础，是保证系统安全运行的关键。网络系统的安全需求包括以下几个方面：

Ø 网络边界安全需求

Ø 入侵监测与实时监控需求

Ø 安全事件的响应和处理需求分析

这些需求在各个应用系统上的不同组合就要求把网络分成不同的安全层次。

我们针对企业网络层的安全策略采用硬件保护与软件保护，静态防护与动态防护相结合，由外向内多级防护的总体策略。

根据安全需求和应用系统的目的，整个网络可划分为六个不同的安全层次。具体是：

Ø **核心层：**核心数据库；

Ø **安全层：**应用信息系统中间件服务器等应用；

Ø **基本安全层：**内部局域网用户；

Ø **可信任层：**公司本部与营业部网络访问接口；

Ø **危险层：**Internet。

信息系统各安全域中的安全需求和安全级别不同，网络层的安全主要是在各安全区域间建立有效的安全控制措施，使网间的访问具有可控性。具体的安全策略如下：

**核心数据库采用物理隔离策略**

应用系统采用分层架构方式，客户端只需要访问中间件服务器即可进行日常业务处理，从物理上不能直接访问数据库服务器，保障了核心层数据的高度安全。

**应用系统中间件服务器采取综合安全策略：**

应用系统中间件的安全隐患主要来自局域网内部，为了保障应用系统中间件服务的安全，在局域网中可通过划分虚拟子网对各安全区域、用户和安全域间实施安全隔离，提供子网间的访问控制能力。同时，中间件服务器本身可以通过配置相应的安全策略，限定经过授权的工作站、用户方能访问系统服务，保障了中间件服务器的安全性；

**内部局域网采取信息安全策略：**

公司本部及营业部内部局域网处于基本安全层的网络，主要是对于安全防护能力较弱的终端用户在使用，因此考虑的重点在于两个方面，一个是客户端的病毒防护，另一个是防止内部敏感信息的对外泄露。因此，通过选用网络杀毒软件达到内部局域网的病毒防护，同时，使用专用网络安全设备（如硬件防火墙）建立起有效的安全防护，通过访问控制ACL等安全策略的配置，有效地控制内部终端用户和外部网络的信息交换，实现内部局域网的信息安全。

**公司本部与下属机构之间网络接口采取通讯安全策略：**

处于可信任层的网络，其安全主要考虑各下属单位上传的业务数据的保密安全，因此，可采用数据层加密方式，通过硬件防火墙提供的VPN隧道进行加密，实现关键敏感性信息在广域网通信信道上的安全传输。

**Internet采取通讯加密策略：**

Internet属于非安全层和危险层，由于Internet存在着大量的恶意攻击，因此考虑的重点是要避免涉密信息在该层次中的流动。通过硬件防火墙提供专业的网络防护能力，并对所有访问请求进行严格控制，对所有的数据通讯进行加密后传输。

同时，建议设置严格的机房管理制度，严禁非授权的人员进入机房，也能够进一步提升整个网络系统的安全。

1) 广域网安全规划

企业广域网安全，主要是通过防火墙和VPN等设备或技术来保障。

防火墙对流经它的网络通信进行扫描，能够过滤掉一些攻击，防火墙还可以关闭不使用的端口，防火墙具有很好的保护作用，入侵者必须首先穿越防火墙的安全防线，才能接触目标计算机，所以出于安全考虑，企业必须购置防火墙以保证其服务器安全，将应用系统服务器放置在防火墙内部专门区域。一般硬件防火墙比软件防火墙的性能更好，建议选择企业级的硬件防火墙，硬件防火墙市场知名度高的品牌有CISCO、Check Point、Juniper、H3C、天融信、华为赛门铁克、联想网御等，用户应根据应用情况选择合适的防火墙。

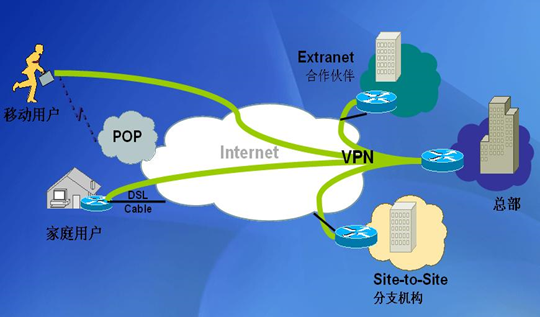
VPN 即虚拟专用网(Virtual Private Networks) 提供了一种通过公共非安全介质(如Internet)建立安全专用连接的技术。使用VPN技术，甚至机密信息都可以通过公共非安全的介质进行安全传送。VPN技术的发展与成熟，可为企业的商业运作提供一个无处不在的、可靠的、安全的数据传输网络。VPN通过安全隧道建立一个安全的连接通道，将分支机构、远程用户、合作伙伴等和企业网络互联，形成一个扩展的企业网络。

VPN基本特征：

Ø 使企业享受到在专用网中可获得的相同安全性、可靠性和可管理性。

Ø 网络架构弹性大——无缝地将Intranet延伸到远端办事处、移动用户和远程工作者。

Ø 可以通过Extranet连接企业合作伙伴、供应商和主要客户（建立绿色信息通道），以提高客户满意度、降低经营成本。

[](http://images.cnitblog.com/blog/15172/201503/211053579379751.png)

VPN实现方式：

Ø 硬件设备：带VPN功能模块的路由器、防火墙、专用VPN硬件设备等，如Cisco、H3C、深信服、天融信等。

Ø 软件实现：Windows 自带PPTP或L2TP、第三方软件（如CheckPoint、深信服等）。

Ø 服务提供商（ISP）：中国电信、联通、网通等。目前一些ISP推出了MPLS VPN，线路质量更有保证，推荐使用。

2）内网安全规划

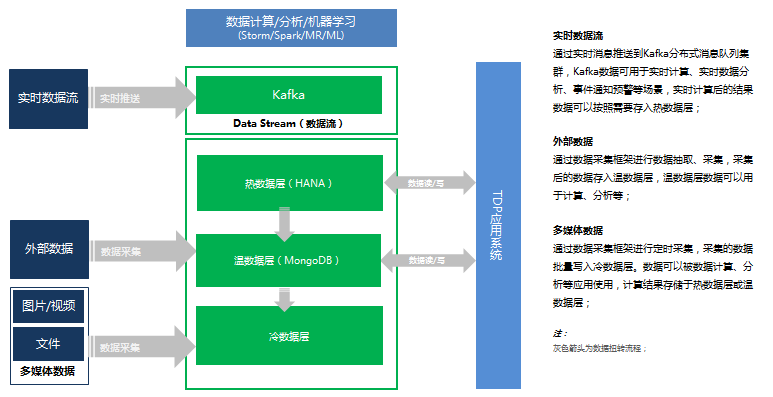
企业内网安全系统包括防病毒系统、内网安全管理系统，上网行为管理系统等。

防病毒系统可以采用网络版防病毒系统或防毒墙等产品（比如金山、瑞星、卡巴斯基等解决方案）。

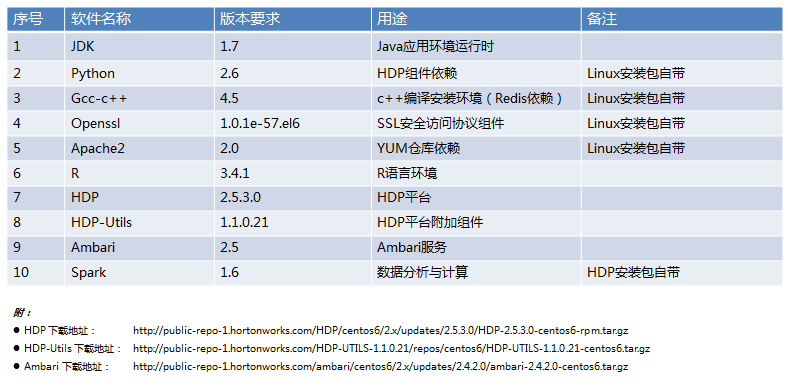
内网安全系统和上网行为管理系统可以选择深信服、任子行、IP-guard等解决方案，企业可以通过部署内网安全系统实现研发网和商业信息的信息安全防范工作。对于研发网的信息安全、数据集中存储和计算资源的统一协调配置，可以通过部署企业桌面虚拟化解决方案来实现。



## 数据扭转设计



### 开发平台



### 3.2.2 **面向SOA集成架构**

面向SOA，为支持跨系统间信息共享以及工作协同而建设的信息服务体系。该体系围绕各类应用主题，满足系统间的在线实时信息的横向交换和业务协同等需求，为各部门的资源共享、工作协同等提供信息交换服务。

针对新能源集团目前的业务需求，本期信息资源交换将重点建设三个方面的内容：目录管理系统、通用交换接口和专业交换接口。

1. 目录管理系统

目录管理系统主要实现各种数据、服务的编目和访问管理，其运行模型如下：



从上图可知，提供者从中分别提取公共资源和交换服务的特征信息，编目形成公共资源核心元数据，注册到公共资源核心元数据库和交换服务核心元数据库中，分别生成公共资源目录和交换服务目录。使用者通过信息资源目录一站式服务对信息资源进行目录查询，在通过访问权限认证后即可访问需要的资源。

1. 通用交换接口

作为一个建立在大数据平台基础上的、可升级的通用程序和解决方案，TDP管理系统能够同企业已有的软件产品和技术完全无缝的集成在一起，并为企业提供了一个高质量的、可靠的、安全的以随着外界环境的变化不断地发展的解决方案构架，并可与客户现有的其他用系统实现集成，从直接利用企业现有的技术资产中发掘和创造更大的效益。

1. 专业交换接口

通过MongoDB提供的数据仓库，上层各种应用通过专业的交互接口，进行数据的互通，做到不同业务数据的不同数据交互，保证数据的独立可用。

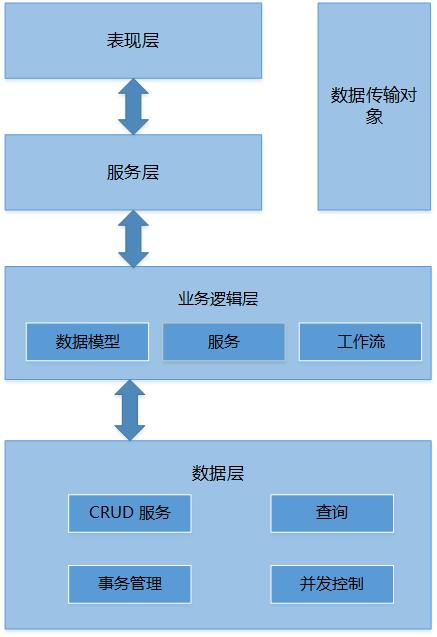
## 应用层设计

应用层是指在快速开发平台之上开发而成的最终供用户使用的业务模块，系统的快速开发平台中集成了规则引擎、调度引擎等众多接口，所以可以非常快速的并具使用最少的代码量来开发流程应用，这些应用可以集成其他业务系统也可以开发独立于其他业务系统的业务模块。应用层的应用开发可以由企业的IT运维人员或者企业的业务人员进行开发和维护。企业的业务人员可以开发基于纯业务流程的应用，而IT运维人员则可以开发复杂的集成其他业务系统的端到端的业务流程。

### 调用业务系统集成

API接口是提供给快速开发平台集成其他业务系统而提供的一系列标准接口，这些接口包括Oracle、Sql Server等数据库的调用接口以及调用其他业务系统发布的WebService、Http等接口。通过这些接口功能TDP系统可以快速的抽取其他业务系统的数据进行流转并把算法验证完成的业务数据重新写回到业务系统中。从而完成主动集成其他业务系统的功能。并且可以把一个业务系统中的数据通过流程审批后转入到另一个业务系统中从而完成不同业务系统的数据交换功能。

## 数据层设计



## 系统接口设计

### 安全认证和鉴权

在本系统中，一般的安全认证是与时代新能源认证相关，但是考虑到新能源希望对保密性极高的要求，可以使用相关的加密技术来确保相关的业务数据不被流失。如采用如下业界较为流行的USB KEY安全认证的方式。

* 方案概述

数字身份认证系统能够完成数字证书的申请、登记、审核、签发、吊销和查询，实现CA认证中心的全部功能。签发的数字证书符合ITUT X.509及相关的PKCS标准，可在支持SSL、S/MIME、SET协议的标准应用系统（如Outlook Express，Internet Explore，Microsoft Authenticode等）和自主开发的应用系统中使用。

数字身份认证系统是政府、企事业单位信息安全的基础平台，通过使用数字身份认证系统签发的数字证书可以为用户提供信息安全的全面服务：

• 保密性 — 保证信息是秘密的

• 完整性 — 能检验信息未被篡改

• 身份鉴别 — 检验个人或机构的身份

• 不可否定性 — 确保信息不能被否认

数字身份认证系统应用国际先进技术，拥有高度的加密算法，高可靠性的安全机制及完善的管理及配置策略。提供自动的密钥和证书管理服务。

* 系统架构

数字身份认证系统由签发系统、CA中心系统、WEB发布系统、证书业务终端系统四个部分组成。数字身份认证系统（ECA）的系统结构如下图所示：



• 签发系统（SIGNER）

由签名服务单元（SIGNServer）和密码服务单元构成。密码服务单元生成并保存CA系统的根密钥，签名服务单元通过连接加密机在线响应CA服务器的证书和CRL签发请求。

• CA中心系统

由CA服务单元（CAServer）、CA系统管理终端构成。CA服务单元在线响应证书业务终端和WEB发布系统的业务请求，提供证书申请、废除、证书查询等服务；CA服务器上安装数据库，保存CA系统的所有证书用户信息、证书信息、操作记录等。

CA系统管理终端实现ECA系统的全面的配置管理和操作审计功能。

• WEB发布系统

由WEB服务单元和ECA CGI构成，提供根证书下载，用户证书下载，CRL下载，证书查询功能。

• 证书业务终端

以界面的方式由CA系统操作员操作，直接面对证书用户办理证书申请和证书作废业务。证书业务终端可以按照证书业务的要求灵活部署。

* 权限控制

系统供了多层的权限控制机制，确保授权的人员访问授权的功能及信息。系统首先提供功能级的权限控制，可以授权使用人员可以使用系统的那些功能，没有被授权的功能不能被使用者访问和使用。

系统提供文件级的权限控制，首先文件在发布时可以设置发送给哪些人和部门，这样只有选定的人和部门才能访问该文件，其他人无法访问和查看，其次，文件本身设定有密级的文件，无法越级查看文件。

### 与HUE门户集成

**HUE是什么**

HUE=Hadoop User Experience

Hue是一个开源的Apache Hadoop UI系统，由Cloudera Desktop演化而来，最后Cloudera公司将其贡献给Apache基金会的Hadoop社区，它是基于Python Web框架Django实现的。

通过使用Hue我们可以在浏览器端的Web控制台上与Hadoop集群进行交互来分析处理数据，例如操作HDFS上的数据，运行MapReduce Job，执行Hive的SQL语句，浏览HBase数据库等等。

HUE链接

* Site: <http://gethue.com/>
* Github: <https://github.com/cloudera/hue>
* Reviews: [https://review.cloudera.org](https://review.cloudera.org/)

核心功能

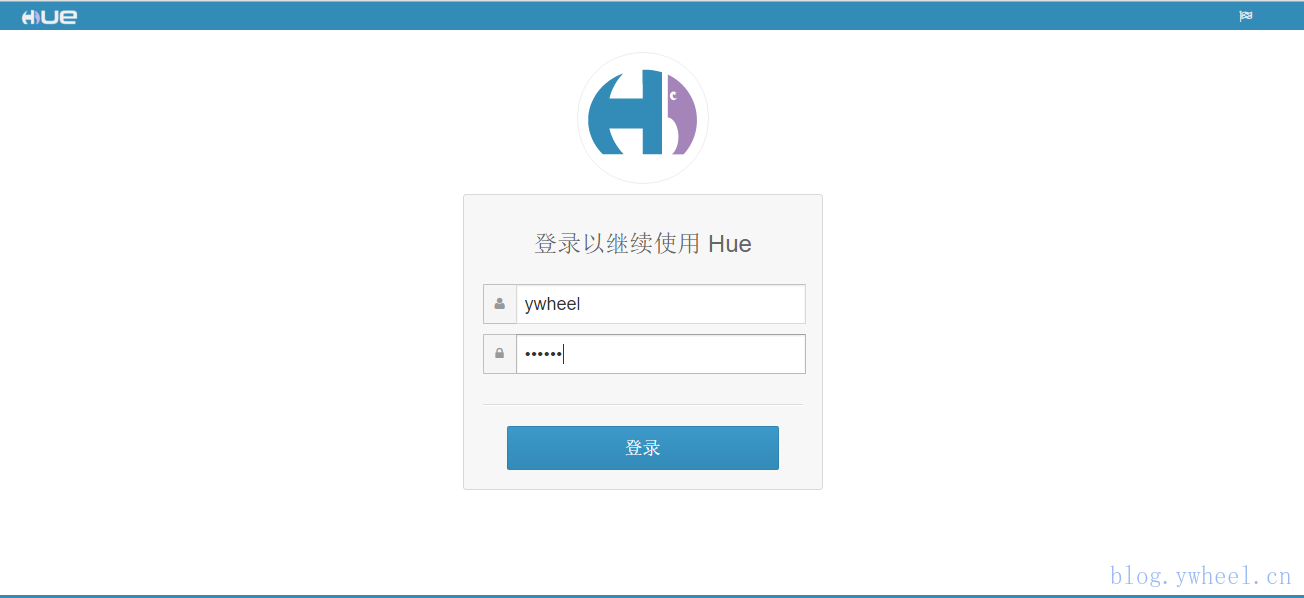
* SQL编辑器，支持Hive, Impala, MySQL, Oracle, PostgreSQL, SparkSQL, Solr SQL, Phoenix…
* 搜索引擎Solr的各种图表
* Spark和Hadoop的友好界面支持
* 支持调度系统Apache Oozie，可进行workflow的编辑、查看

HUE提供的这些功能相比Hadoop生态各组件提供的界面更加友好，但是一些需要debug的场景可能还是需要使用原生系统才能更加深入的找到错误的原因。

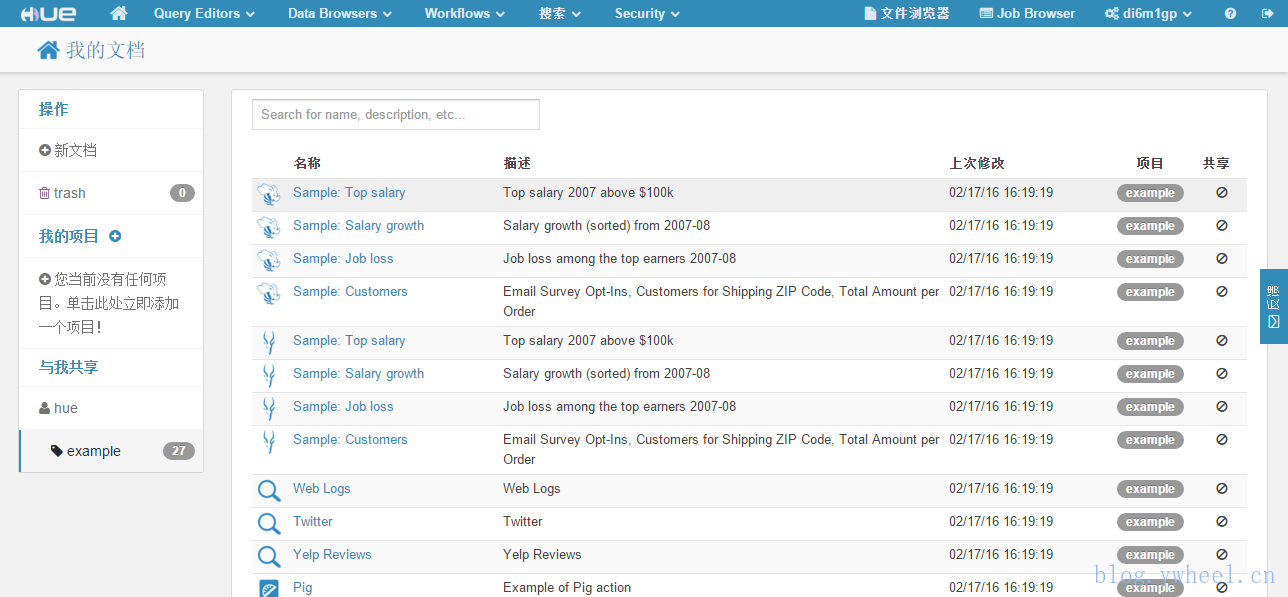
HUE中查看Oozie workflow时，也可以很方便的看到整个workflow的DAG图，不过在最新版本中已经将DAG图去掉了，只能看到workflow中的action列表和他们之间的跳转关系，想要看DAG图的仍然可以使用oozie原生的界面系统查看。

**HUE登录**

如果自己搭建了HUE，则可以使用管理员账户创建一个新的用户，然后使用新的用户进行登录，见下图：



使用[HUE官网](http://gethue.com/)上的live demo可以尝鲜。如果大家自己没有搭建大数据平台，没有安装HUE的话，可以先在该demo上尝试。点击[Play with the live Demo now!](http://demo.gethue.com/),将会进入HUE的”我的文档”：

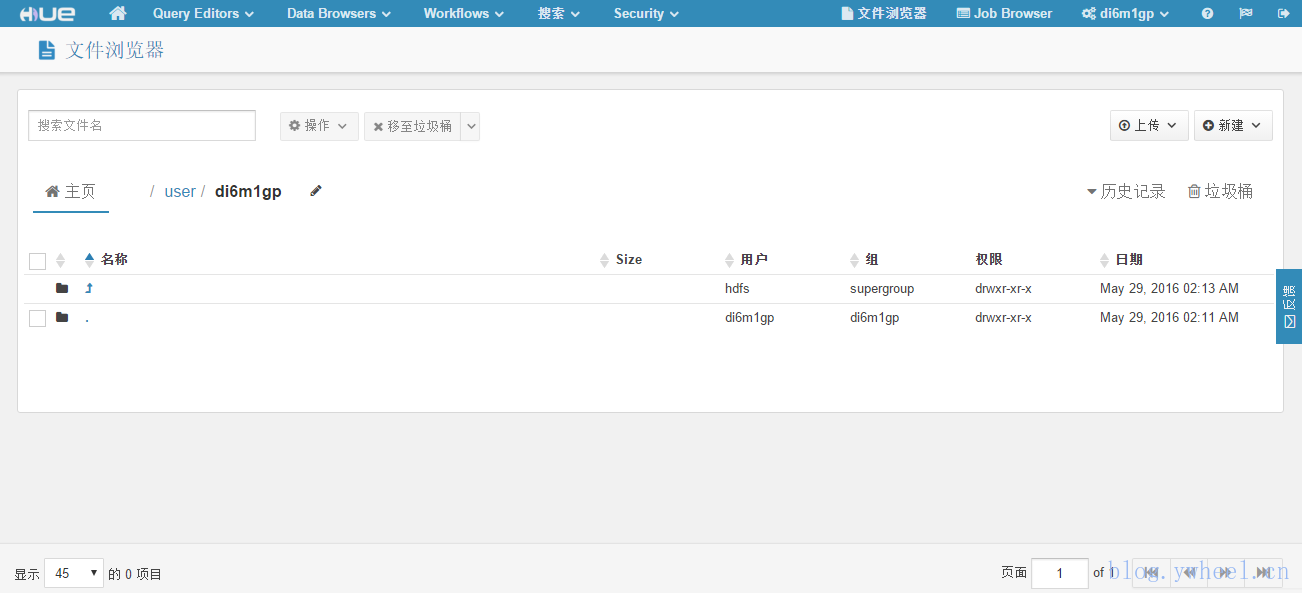


HDFS文件浏览

HUE可以很方便的浏览HDFS中的目录和文件，并且进行文件和目录的创建、复制、删除、下载以及修改权限等操作。

HDFS实现了一个和POSIX系统类似的文件和目录的权限模型。每个文件和目录有一个所有者（owner）和一个组（group）。文件或目录对其所有者、同组的其他用户以及所有其他用户分别有着不同的权限。**但，用户身份机制对HDFS本身来说只是外部特性。HDFS并不提供创建用户身份、创建组或处理用户凭证等功能。** 使用HUE访问HDFS时，HDFS简单的将HUE上的用户名和组的名称进行权限的校验。

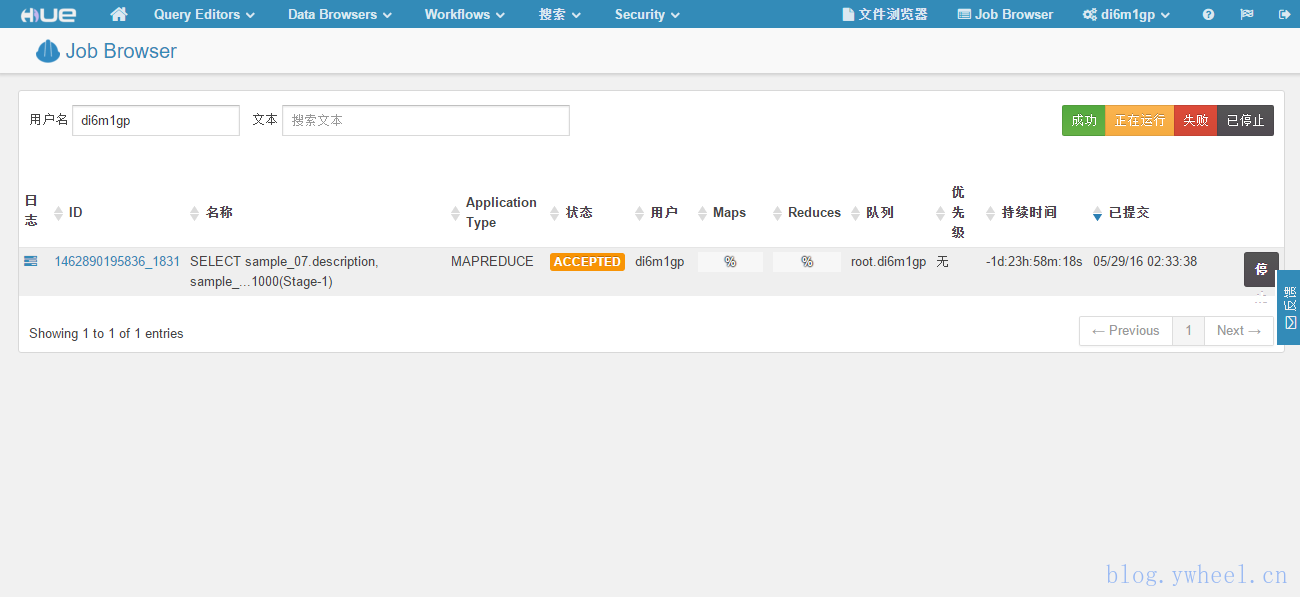
在Live Demo中,点击”文件浏览器”, 进入HDFS的家目录：



**PS:** Live Demo中禁了文件上传功能。

作业浏览

点击Job Browser，可以查看作业列表，并且可以通过点击右上角的”成功”,”正在运行”,”失败”,”停止”来筛选不同状态的作业：



Hive查询

HUE的beeswax app提供友好方便的Hive查询功能，能够选择不同的Hive数据库，编写HQL语句，提交查询任务，并且能够在界面下方看到查询作业运行的日志。在得到结果后，还提供进行简单的图表分析能力。

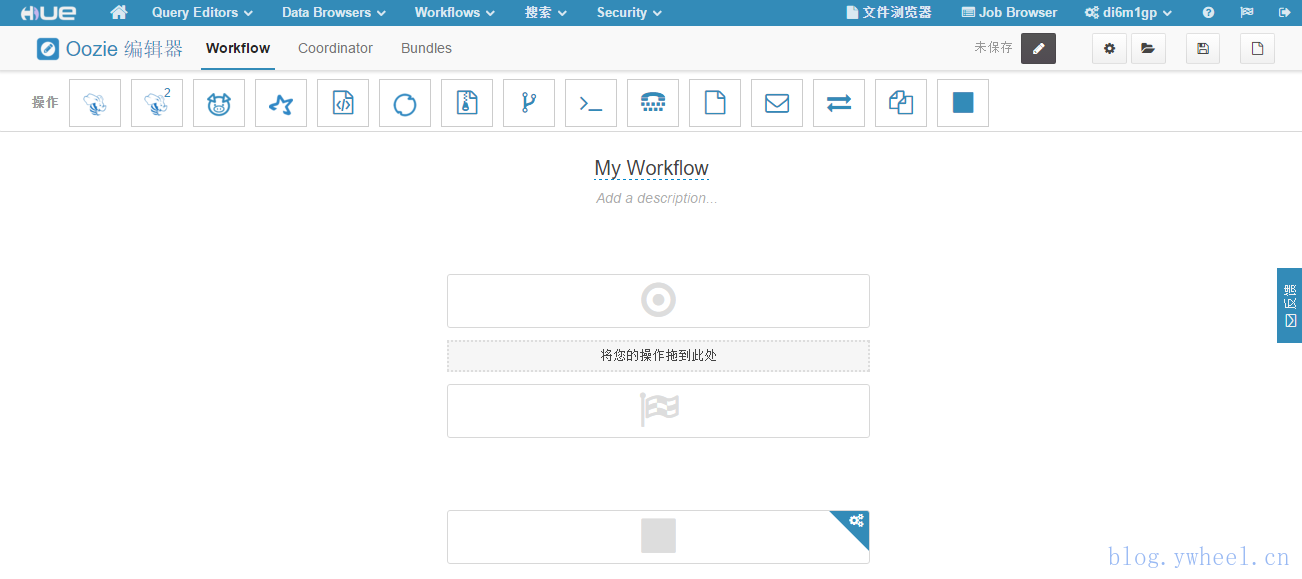


点击”Data Browsers”->”Metastore表”，还可以看到Hive中的数据库，数据库中的表以及各个表的元数据等信息。

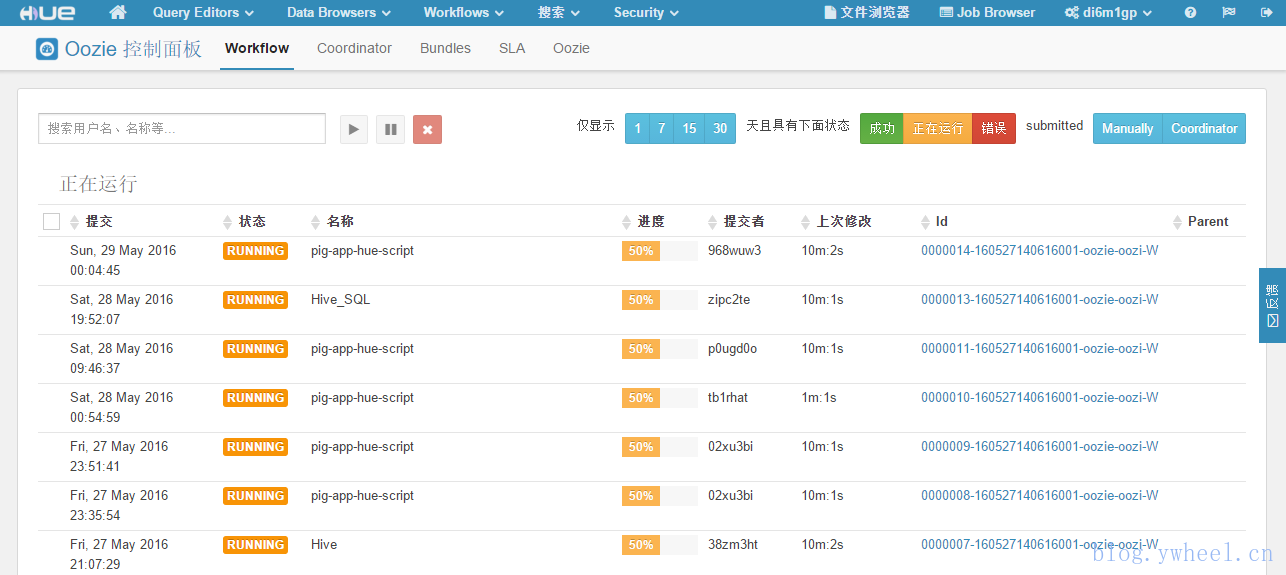


Oozie Workflow编辑

HUE也提供了很好的Oozie的集成，能够在HUE上创建和编辑Bundles, Coordinator, Workflow. Oozie的介绍可以去官网查看。下图为在HUE上创建一个新的workflow，在该界面上，可以直接拖动不同的组件，变成DAG中的节点，并且设置各个action的流转逻辑。



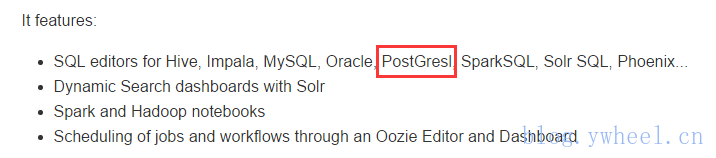
当然Oozie也可以通过命令行的方式提交B,C,W. 不过是使用HUE创建的workflow，或者是通过命令行提交的workflow，都可以在HUE上查看运行的状况：



只是通过命令行提交的workflow就不可以在HUE上进行编辑了。使用配置文件、命令行提交的方式能够保证在生产环境上运行的和在测试环境上运行的版本一致，而使用HUE界面编辑的方式虽然方便，但也可能会带来人工操作在生产环境中失误的风险，有利也有弊。

Contribution

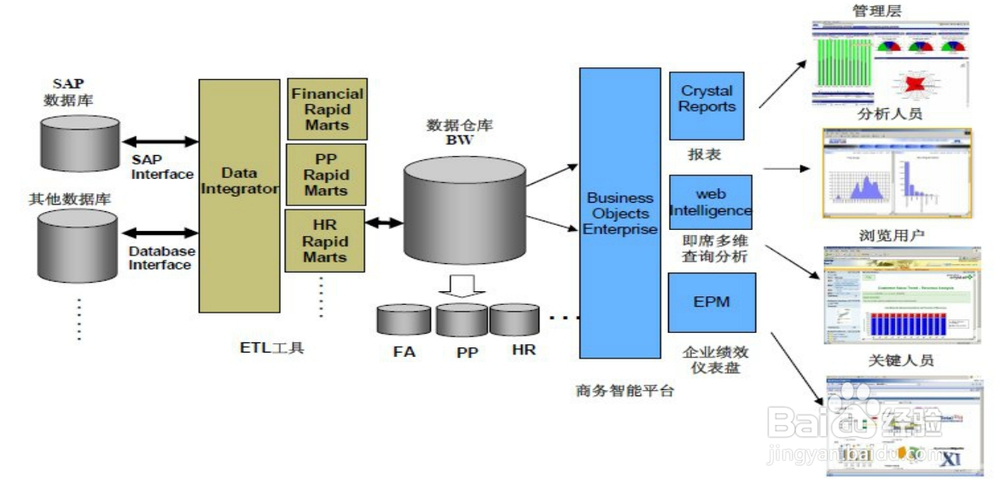
HUE的主要功能时，github上的原文是这样的：



**集成设计**

TDP通过底层MongoDB的数据交互到spark计算框架，随后再通过HUE的相关页面进行操作，更加方便高效的进行系统使用。

### 与BO协作集成



SAP BO架构

TDP通过与SAP BO的集成，把TDP的业务数据通过SAP BO的报表展现，使输出结果一目了然。SAP BO是报表展现类，本身带有数据抓取器，用起来很方便，也可直接用sQL编写取数逻辑。业务对象层包括表示应用程序领域内"事物"的所有实体类。这些实体类驻留在服务器上，并利用服务类来协助完成它们的职责。

**集成设计**

### 与portal集成

  Portal以用户为中心，提供统一的用户登录，实现信息的集中访问，集成了办公商务一体的工作流环境。利用Portal技术，可以方便地将员工所需要的，来源于各种渠道的信息资料集成在一个统一的桌面视窗之内。根据Portal提供的定制功能，部门主管可以为本部门人员量身定制一套特有的信息门户，将部门共同所需信息有效地组织在统一的Web浏览器之中，并可根据人员级别和职能来设定相应的访问操作权限。

一.Portal主要功能

1、单点登录（SSO—Single Sign-On）：Portal提供对各种应用系统和数据的安全集成，用户只需从Portal服务器登录一次就可以访问其它应用系统和数据库。对于安全性要求较高的业务系统，如电子银行、电子交易系统等，通过传递用户身份信息，如数字证书信息、数字签名信息等进行二次身份认证，保证单点登录的安全性。单点登录既减少了用户在多个应用系统反复登录多次认证的麻烦，更是简化了各种应用系统对用户及其权限的一致性维护管理。

2、资源整合：能够把各种不同应用的内容聚合到一个统一的页面呈现给用户，实现同应用系统实时交换信息。能够从各种数据源如数据库、多种格式的文件档案、Web页面、电子邮件等集成用户所需的动态内容。

3、定制与个性化：能够为不同角色的用户制定不同功能权限的Portal页面。同时，用户自己也能够按照喜好在规定的权限下定制自己风格的页面和内容，如可以定制Portal页面，取舍不同功能和内容的Portlet窗口，自行布置Portlet窗口的摆放位置，可以对Portlet窗口外观，如标题，图标，颜色等进行个性化设置。

4、协作功能：为用户提供即时讨论，聊天，论坛，电子邮件以及语音或视频会议等功能。

5、工作流：支持根据业务处理规则建立起来的工作流任务处理，比如审批流程等待办事宜。

6、信息检索：从多种数据源检索动态信息资料。

7、客户端：除了Web浏览器外，可以为PDA和手机提供接口，实现移动接入服务。

二、Portal标准

建立一个以标准为依托的Portal才能很好地保护自己的投资，既便于同现有应用系统连接，也使得同第三方的相关产品更容易接口。在2003年先后发布的JSR-168和WSRP两大标准为Portal的发展奠定了基础，结束了战国纷争的局面，Portal的发展和应用将会更加广阔长远。

下面介绍几个与Portal紧密相关的技术标准：

1、WSDL——Web Service Description Language

Web服务描述语言。WSDL是用来描述Web服务和说明如何与Web服务通信的XML语言。WSDL语言使用XML格式来描述信息的接口、访问格式和处理形式。

WSDL描述信息内容。

2、SOAP——Simple Object Access Protocol

简单对象访问协议。SOAP是一种在无中心的分布式环境下，应用系统之间交换结构化信息和特定类型的信息所使用的基于XML的轻量级协议。SOAP允许任何信息对象在任何语言、任何平台上使用多种传输协议实现传输处理。

SOAP定义信息的传输处理。

在Web应用环境中，通常把SOAP同WSDL结合起来，利用HTTP协议实现应用系统之间交换各种类型的信息对象。

3、JSR -168——Java Specification Request - Portlet Specification

Java Portlet规范。JSR-168为业界明确了Portal的定义，制定了Portlet规范标准，从而解决了基于Java的Portal之间，以及同其他Web应用系统之间的互操作性。遵循JSR-168的Portlet将具有适用于所有Portal服务器和Web应用系统，支持多种类型的客户端，支持本地化和国际化，具备确定的安全性，允许Portal应用程序热部署和重新部署。

4、WSRP——Web Services for Remote Portlets

远程Portlet Web服务协议。WSRP定义了Portal和Portlet容器服务之间标准化接口的一个Web服务标准。WSRP允许在Portal之间或其他Web应用上即插即用，具有互操作性，提供可视化的、面向用户的远程Web服务。

远程Portlet在远程服务器上作为Web服务运行，其服务可以发布到公共的或单位自己的UDDI服务器上。Portal或其他支持WSRP的应用系统通过UDDI服务来查找并使用远程系统提供的WSRP服务内容。

WSRP的典型应用是把天气预报、即时新闻、股市行情等嵌入到自己的Portal中（在国外有专门的WSRP内容提供商提供这种服务）。

WSRP使用了WSDL定义应用程序的接口，并以SOAP作为通讯标准。

5、其他规范标准

此外还有一些与Portal有一定关联的技术标准，在开发建立Portal应用中将会使用到：

UDDI：Universal Description，Discovery and Integration

JSR-170/283：Java Specification Request - Content Repository for Java Technology API

JAAS：Java Authentication and Authorization Service

LDAP：Lightweight Directory Access Protocol

SAML：Security Assertion Markup Language

BPEL：Business Process Execution Language for Web Services

 三、Portal应用实现

Portal可以从本地或远端获得数据资源：数据资源可以来自于本地或异地的数据库，应用系统，公共信息内容供应商（RSS，提供新闻、财经信息、天气预报等），Web站点或其他Portal。此外，Portal还可以提供日历、工作流、电子邮件、论坛、博客、Wiki、即时交谈、电子会议等等协同工作的应用功能。可见，Portal是一个有别于传统桌面应用和Web网站的全新应用系统，是各种信息处理的集中展现平台，是用户日常工作的综合台面。

显然，建立一个良好的Portal应用需要充分考虑各种应用系统和数据资源的整合问题：

1、现有应用系统和数据资源的利用  对能够改造利用的，要开发相应的Portlet组件来重新实现；对不能改造的可以通过链接的方式跳转到这些系统，其中的数据库资源可以采取单纯读取的方式获得；还有些封闭的专业应用系统可能完全无法接入Portal，可以采取定期卸载的方式获得它的数据库资源。

2、新建应用系统的考虑  需要以Portal理念进行设计，按照相关标准来开发实现应用功能的Portlet组件，然后集成到Portal系统使用。

3、单点登录与权限管理  对于新建应用系统或能够改造的现有应用系统，通过Portlet组件比较容易实现单点登录，进行统一用户认证和用户权限的控制。当然，对那些安全性要求较高的应用系统还可以在这些系统内部进行二次认证和授权处理。对那些不能改造的应用系统显然也无法实现单点登录，用户需要重新登录到这类系统，用户管理和权限控制还依赖于这些系统自己处理。

  （一）单点登录和权限控制

单点登录是为了方便用户进入多个应用系统，减少用户多次登录，免除用户记忆多套用户名和密码的麻烦。

单点登录涉及到两个问题，一是身份认证，二是权限控制。

身份认证是Portal系统提供访问控制的第一步，即确认用户是谁，能否进入系统。通常要求用户提供用户名和口令，必要时要求提供用户的数字证书，也可以配合使用IC卡、指纹等验证手段。

权限控制或授权确定一个用户的角色和级别，从而控制用户的访问许可，即决定用户能查阅哪些资料，能进行哪些操作等等。Java EE架构采用了基于角色的访问控制策略（RBAC）。RBAC的基本思想是把对用户的授权划分成两个分配关系，即“用户—角色”和“角色—权限”。RBAC的好处是便于应用系统的开发，使得程序设计相对独立和透明化，只是在应用系统部署使用时才通过“角色”把“用户”和“权限”关联起来，而且对用户和权限的调整配置容易实施。

用户与角色之间是多对多的关系，即一个用户可以被分配给多个角色，多个用户也可以分配给同一个角色。

角色与权限之间也是多对多的关系，即一个权限可以与多个角色相关，一个角色也可以包含多重权限。

Sun公司建立了具有可堆叠和可插接功能的JAAS框架，为Java应用系统提供安全而灵活的身份认证和授权机制。如在JBossAS应用服务器环境下，利用JAAS技术，JBossAS提供了几种身份认证模块（使用者也可以自己编写新的身份认证模块），并且可以实现多级认证（堆叠）。对Web应用和EJB等则通过配置文件定义访问角色和访问方式以实现安全控制。对没有提供JAAS支持或支持不足的Web服务器或Java应用服务器，使用者需要在自己的应用程序中编写相关代码或模块来支持JAAS实现身份认证和权限管理。

此外还需要考虑不同领域内不同应用之间的信任关系，解决跨越应用系统的身份认证和访问控制问题。这一点需要用到SAML，Federation或Liberty等技术规范，通过传递用户认证资料取得应用系统之间的认证互信。

专业的Web服务软件公司能够提供比较全面的解决方案，其产品为Web环境下的各种应用提供可靠的身份认证和访问权限管理。比如Sun公司的Access Manager 7.1，Oracle公司的Oracle Access Manager 10g，IBM公司有Tivoli Access Manager（提供用户单点登录）和Tivoli Identity Manager（解决访问控制问题），还有BEA公司的BEA AquaLogic Enterprise Security等。

相比之下，开源软件在这方面比较瘦弱，基本上没有专门的软件产品，只能提供一定的支持方案。例如通过对JBossAS和Tomcat做一些配置，可以支持使用耶鲁大学的CAS实现单点登录功能，也能够实现JAAS来获得用户身份认证和对应用资源的访问控制。这样，建立在JBossAS和Tomcat之上的应用如JBoss Portal，Liferay Portal和Apache JetSpeed当然也就具备了这些功能了。

在用户管理、身份认证和权限控制方面，无论是商业的或开源的Portal产品多数喜欢采用LDAP，当然也有的支持使用数据库。LDAP的好处一是它可以方便的按类别存储任何类型的数据信息；其二，它的树形存储结构类似于一个企事业单位的组织架构，容易对应；三是它同应用系统接口容易，各个LDAP产品的接口都一致无需特别配置；四是它对数据信息的访问安全控制方便；五是它偏向于相对固定数据信息的查询使用，效率较高，维护也方便。

|  |
| --- |
| 四Portal作用 |

**消除信息孤岛，提高工作效率和企业宏观决策的准确性和及时性**   
Portal提供了为企业员工提供了各应用系统的统一内容聚集和展现，员工不必再到分散的应用系统中浏览、查找、分析相关的信息，而是能通过统一的Portal入口，迅速地找到自己关心的信息和情报。对于企业普通员工而言，这样将提高了工作效率，减少了反复查找信息而进行的重复劳动和由此引起的信息失真和缺漏; 对于企业的管理者和决策者而言，统一的应用系统内容聚集和展现，将企业各个环节和不同维度的信息切片综合地、有机地、即时地展现在使用者面前，方便进行日常管理和宏观决策。

**简化工作环节，快速适应新业务和新战略的需求**   
Portal框架下通过提供统一认证和单次登录的功能，改变了用户必需使用不同的登录方法，不同的用户和密码反复登入不同的系统而带来的操作复杂性和安全隐患，通过提供单次登录的功能，简化了用户访问各种应用系统的环节，提高了效率。  
Portal还提供了可高度定制的个性化内容展示框架，以一种统一的架构和模式，适应企业内部不同组织层次，不同业务部门对各自工作界面的特殊要求。在这个具备高度扩展性和灵活性的展示聚集框架的支撑下，企业才有可能随时根据业务发展的变化，即时调整相关部门和人员关注的焦点和内容，使底层的IT支撑能够赶上业务的变化和员工的要求。

**减轻企业管理负担，适应未来IT基础支撑体系的发展**  
Portal采用基于浏览器方式的、统一的访问和展示界面，提供单次登录功能，并配合远程桌面管理系统，对于普通用户而言，无需安装另外的客户端即可访问众多的后台应用系统和企业的内部信息共享平台，基本无需额外的培训，即可适应新的系统的基本操作，个人用户可以将精力集中在业务和工作本身上，而不是反复适应系统上; 对于系统管理人员而言，集中授权式的企业安全管控模式、瘦客户端的发布方式，基于桌面管理系统的远程部署、诊断模式，都大大减轻了管理人员的工作负担和复杂度; 对于企业而言，Portal是基于开放的J2EE架构的，高度灵活、可定制的展示框架系统，其采用的企业集中LDAP目录服务和统一管理、统一认证机制，为企业的后续IT建设和应用集成，奠定良好的构架和基础。

|  |
| --- |
| **五Portal架构** |

Portal采用三到N层次的应用架构，逻辑三层次架构具有展现层 (Presentation Tier-View), 应用逻辑层 (Application Tier- Control)， 及数据层 (Enterprise Tier-Data Model)。展现层负责输入及输出和展现客户所要看的数据信息，应用逻辑层 (Application Tier- Control) 是专门用来处理应用逻辑及访问和查询后台的企业业务应用与数据。

系统接入层

客户端是最终用户，他们处于非安全区域，可以采用多种接入方式，一般以浏览器为主，可是也可以通过常用的移动终端设备如PDA，WAP手机等接入。

**业务展示层**

Portal主要处理用户访问统一信息平台的访问安全控制管理(后台认证功能由认证子系统和目录服务数据库共同完成)、策略管理及内容、应用聚集的功能。另外，Portal将支撑用户使用不同访问设备的内容格式提交，通过门户的渠道功能将企业内部的信息资源呈现给用户。

**业务汇聚层**

业务逻辑层汇集了Portal的具体业务应用逻辑，主要包括: 协同工作(工作流)、搜索引擎、文档管理、网上教育、统计查询、身份认证、个性化服务等。对系统数据层的数据进行操作访问，实现各种系统服务功能，是Portal框架下应用资源的有机载体。

**核心数据层**

核心数据层就是Portal的数据采集与交互平台，包括: 目录服务数据库、群件数据库、多媒体课件服务器、访问控制数据库、包括BOSS、网管等核心业务系统在内的核心数据源等。

**六 集成设计**

### 与其他系统集成

系统集成有以下几个显著特点：

* 系统集成要以满足用户的需求为根本出发点。
* 系统集成不是选择最好的产品的简单行为，而是要选择最适合用户的需求和投资规模的产品和技术。
* 系统集成不是简单的设备供货，它体现更多的是设计、调试与开发的技术和能力。
* 系统集成包含技术、管理和商务等方面，是一项综合性的系统工程。技术是系统集成工作的核心，管理和商务活动是系统集成项目成功实施的可靠保障。
* 性能性价比的高低是评价一个系统集成项目设计是否合理和实施是否成功的重要参考因素。

总而言之，系统集成是一种商业行为，也是一种管理行为，其本质是一种技术行为。

# TDP平台详细建设方案

## 4.1 平台概述

大数据问题的分析和解决通常很复杂。大数据的量、速度和种类使得提取信息和获得业务洞察变得很困难。以下操作是一个良好的开端：依据必须处理的数据的格式、要应用的分析类型、使用的处理技术，以及目标系统需要获取、加载、处理、分析和存储数据的数据源，对大数据问题进行分类。

大数据可通过许多方式来存储、获取、处理和分析。每个大数据来源都有不同的特征，包括数据的频率、量、速度、类型和真实性。处理并存储大数据时，会涉及到更多维度，比如治理、安全性和策略。选择一种架构并构建合适的大数据解决方案极具挑战，因为需要考虑非常多的因素。

这个 “大数据架构和模式” 系列提供了一种结构化和基于模式的方法来简化定义完整的大数据架构的任务。因为评估一个业务场景是否存在大数据问题很重要，所以我们建设TDP测试大数据平台就是为了解决上述问题应运而生，TDP平台，作为宁德新能源平台级产品，为新能源各业务方提供平台级数据支撑，为算法应用、科学预测、智能制造打造良好的通用平台。

## 4.2 平台体系建设

现在来分享TDP业务流程管理体系建设过程：

* 第一步：搭建企业统一的数据底层架构。从战略和业务出发，在企业内部建立统一涵盖所有业务活动的数据架构（目录），打破数据的部门化（孤岛化），建立数据的衔接体系，为构建端到端的数据平台体系建立基础。业务数据架构搭建的重点是建立业务数据的分类和分层体系，以清晰展现企业的业务分类、层次和逻辑。
* 第二步：建立业务数据管理平台系统，作为企业大数据管理的统一平台。基于TDP平台，建立统一的业务数据描述方法和规范，根据已经建立的统一的业务数据架构，对企业内部所有的业务数据进行梳理、补充、完善和优化，搭建端到端的流程体系。流程设计完成之后，统一发布实施，作为企业内部业务数据执行的唯一依据，实现“一个企业，一套流程”的管理目标。
* 第三步：建立统一的平台数据管理制度。成立专门的部门或者指定相应部门承担业务数据管理的职责，负责组织进行业务数据的持续优化和改进。建立业务数据设计、发布、监控、优化、变更方面的流程和制度。
* 第四步：多管理主题的集成。根据企业的管理需要，对与数据相关的各种管理主题，例如智能营销、机器学习、电池寿命预测、科学决策管理等在业务流程管理信息系统中基于统一的业务流程进行整合，实现“一套流程、多种应用”的管理目标。

## 4.3 功能性设计方案

### 4.3.1 底层MongoDB数据库要求

**技术要求**：

1.TES中的HANA表，全量导入MongdDB数据库中，表明和字段不发生变化；

TES中的HANA表清单：

* Table Name: MD\_PROCESS\_INFO（主数据表）
* Table Name: MD\_STEP\_INFO（主数据对应的工步信息表）
* Table Name: TX\_ORIGINAL\_PROCESS\_DATA（测试数据原始数据表\_存所有数）
* Table Name: TX\_RESULT\_PROCESS\_DATA分区（抽取结果数据表4个）
* Table Name: TX\_ZIP\_PROCESS\_DATA 分区（压缩结果数据表12个）

2.上述五张表，在MongoDB中作数据仓库使用；

3.TDP中的spark计算框架（MongoDB的数据输出），调用上述存储数据时

3.1能随时调用，确保数据的准确性

3.2能随时调用，确保数据的及时性

3.3能以结构化的数据调用方式调用

3.4调用时发生中断，能自动断线重连

4.五张表结构如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table Name** | **MD\_PROCESS\_INFO** | | | | |
| Column Name | SQL Data Type | Key | Not Null | Default | Comment |
| HANDLE | NVARCHAR(255) | X | X |  | MdProcessInfoBO:<SITE>,<REMARK> |
| SITE | NVARCHAR(6) |  |  |  |  |
| REMARK | NVARCHAR(80) |  |  |  | 流程文件标识号  T3-2016041501317-246454\_DCR\_35(测试单号+流程名称+电芯号)要求全局唯一号 |
| ROOT\_REMARK | NVARCHAR(80) |  |  |  | 接续的上一个REMARK |
| PROCESS\_ID | NVARCHAR(50) |  |  |  | 流程号 |
| TEST\_REQUEST\_ID | NVARCHAR(50) |  |  |  | 测试申请单号 |
| GROUP\_NAME | NVARCHAR(50) |  |  |  | 组别 |
| SFC | NVARCHAR(50) |  |  |  | 电芯Barcode |
| ENGINEER | NVARCHAR(50) |  |  |  | 测试工程师 |
| USER\_ID | NVARCHAR(20) |  |  |  | 测试工程师工号 |
| PACKAGE\_ID | NVARCHAR(50) |  |  |  | E-package ID |
| PROJECT | NVARCHAR(50) |  |  |  | 测试项目 |
| PROJECT\_CODE | NVARCHAR(50) |  |  |  | 项目代码 |
| PROJECT\_NAME | NVARCHAR(50) |  |  |  | 项目名称 |
| DESCRIPTION | NVARCHAR(50) |  |  |  | 改善信息 |
| SV\_MODEL | NVARCHAR(50) |  |  |  | Model |
| SV\_CAPACITY\_VALUE | DECIMAL(20,5) |  |  |  | 标称容量 |
| FIXTURE\_TYPE | NVARCHAR(50) |  |  |  | 夹具类型 |
| SV\_INIT\_FIXTURE\_FORCE | DECIMAL(20,5) |  |  |  | 初始夹具力 |
| SV\_INIT\_WEIGHT | DECIMAL(20,5) |  |  |  | 初始重量 |
| SV\_INIT\_VOLUME | DECIMAL(20,5) |  |  |  | 初始体积 |
| SV\_INIT\_IR | DECIMAL(20,5) |  |  |  | 初始电阻 |
| SV\_INIT\_OCV | DECIMAL(20,5) |  |  |  | 初始OCV |
| SV\_INIT\_THICKNESS | DECIMAL(20,5) |  |  |  | 初始厚度 |
| DAYS\_OR\_CYCLE | DECIMAL(20,5) |  |  |  | 天数/循环 |
| SV\_LOWER\_U | DECIMAL(20,5) |  |  |  | 电压下限 |
| SV\_UPPER\_U | DECIMAL(20,5) |  |  |  | 电压上限 |
| SV\_TEMPERATURE | DECIMAL(20,5) |  |  |  | 温度 |
| SV\_TIME\_DURATION | DECIMAL(20,0) |  |  |  | 测试时间 |
| SV\_CHAGE\_CURRENT | DECIMAL(20,5) |  |  |  | 充电电流 |
| SV\_DISCHARGE\_CURRENT | DECIMAL(20,5) |  |  |  | 放电电流 |
| SV\_CHARGE\_VOLTAGE | DECIMAL(20,5) |  |  |  | 充电电压 |
| SV\_DISCHARGE\_VOLTAGE | DECIMAL(20,5) |  |  |  | 放电电压 |
| SV\_CHARGE\_POWER | DECIMAL(20,5) |  |  |  | 充电功率 |
| SV\_DISCHARGE\_POWER | DECIMAL(20,5) |  |  |  | 放电功率 |
| CYCLE\_NUMBER | DECIMAL(20,0) |  |  |  | 循环数 |
| SOC | DECIMAL(20,5) |  |  |  | SOC |
| PULSE\_TIMES | DECIMAL(20,0) |  |  |  | 脉冲次数 |
| CHARGE\_MULTI | NVARCHAR(45) |  |  |  | 充电倍率 |
| DISCHARGE\_MULTI | NVARCHAR(45) |  |  |  | 放电倍率 |
| STORE\_END\_CONDITION | NVARCHAR(45) |  |  |  | 存储结束条件 |
| STORE\_CROSS\_CONDITION | NVARCHAR(45) |  |  |  | 存储交叉条件 |
| CYCLE\_END\_CONDITION | NVARCHAR(45) |  |  |  | 循环结束条件 |
| CYCLE\_CROSS\_CONDITION | NVARCHAR(45) |  |  |  | 循环交叉条件 |
| SIM\_END\_CONDITION | NVARCHAR(45) |  |  |  | 工况结束条件 |
| SIM\_CROSS\_CONDITION | NVARCHAR(45) |  |  |  | 工况交叉条件 |
| IS\_ISO\_DISCHARGE | TINYINT |  |  |  | 是否恒压放电，  0 – false  1 – true |
| ISO\_DISCHARGE\_DURATION | DECIMAL(20,0) |  |  |  | 恒压放电时长 |
| CONSTANT\_IR\_VALUE | DECIMAL(20,5) |  |  |  | 横阻值 |
| CYCLE\_TEMPERATURE | DECIMAL(20,5) |  |  |  | 循环温度 |
| STORE\_TEMPERATURE | DECIMAL(20,5) |  |  |  | 存储温度 |
| TX\_STATUS | NVARCHAR(30) |  |  |  | 流程状态，  new – 新建  inprocess – 进行中  protected – 保护  close – 完成 |
| TOTAL\_CYCLE\_NUM | INTEGER |  |  |  |  |
| CREATED\_DATE\_TIME | SECONDDATE |  |  |  | 创建日期 |
| CREATED\_USER | NVARCHAR(20) |  |  |  | 创建用户，USR.USER\_ID |
| MODIFIED\_DATE\_TIME | SECONDDATE |  |  |  | 最后修改日期 |
| MODIFIED\_USER | NVARCHAR(20) |  |  |  | 最后修改用户，USR.USER\_ID |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table Name** | **MD\_STEP\_INFO** | | | | |
| Column Name | SQL Data Type | Key | Not Null | Default | Comment |
| HANDLE | NVARCHAR(255) | X | X |  | MdStepInfoBO:<SITE>,<REMARK>,<STEP\_ID> |
| SITE | NVARCHAR(6) |  |  |  |  |
| REMARK | NVARCHAR(80) |  |  |  |  |
| STEP\_ID | INTEGER |  |  |  | 工步号 |
| STEP\_NAME | NVARCHAR(50) |  |  |  | 工步名称 |
| SCRIPT\_CURRENT | TEXT |  |  |  | 电流判异脚本 |
| SCRIPT\_VOLTAGE | TEXT |  |  |  | 电压判异脚本 |
| SCRIPT\_TEMPERATURE | TEXT |  |  |  | 温度判异脚本 |
| SCRIPT\_TIME | TEXT |  |  |  | 时间判异脚本 |
| SCRIPT\_CAPACITY | TEXT |  |  |  | 容量判异脚本 |
| SCRIPT\_ENERGY | TEXT |  |  |  | 能量判异脚本 |
| IS\_CYCLE\_SIGNAL\_STEP | TINYINT |  |  |  | 循环标识工步，是否是循环+1工步 |
| DELTA\_VOLTAGE | DECIMAL(20,5) |  |  |  | 电压差∆V |
| SV\_POWER | DECIMAL(20,5) |  |  |  | 功率 |
| SV\_IR | DECIMAL(20,5) |  |  |  | 内阻 |
| SV\_VOLTAGE | DECIMAL(20,5) |  |  |  | 电压 |
| SV\_CAPACITY | DECIMAL(20,5) |  |  |  | 容量 |
| SV\_CURRENT | DECIMAL(20,5) |  |  |  | 电流 |
| SV\_STEP\_END\_CAPACITY | DECIMAL(20,5) |  |  |  | 工步截止容量 |
| SV\_STEP\_END\_CURRENT | DECIMAL(20,5) |  |  |  | 工步截止电流 |
| SV\_STEP\_END\_VOLTAGE | DECIMAL(20,5) |  |  |  | 工步截止电压 |
| SV\_STEP\_END\_TEMPERATURE | DECIMAL(20,5) |  |  |  | 工步截止温度 |
| SV\_ENERGY | DECIMAL(20,5) |  |  |  |  |
| SV\_TEMPERATURE | DECIMAL(20,5) |  |  |  |  |
| SV\_TIME | DECIMAL(20,5) |  |  |  |  |
| CYCLE\_COUNT | INTEGER |  |  |  |  |
| CONDITION\_TYPE | NVARCHAR(45) |  |  |  |  |
| CONDITION\_OPERATIONAL\_CHARACTER | NVARCHAR(45) |  |  |  |  |
| CONDITION\_VALUE | DECIMAL(20,5) |  |  |  |  |
| GOTO\_STEP | NVARCHAR(45) |  |  |  |  |
| START\_STEP | NVARCHAR(45) |  |  |  | 起始工步 |
| START\_SOC | DECIMAL(20,5) |  |  |  |  |
| SOC\_INCREMENT | DECIMAL(20,5) |  |  |  |  |
| END\_SOC | DECIMAL(20,5) |  |  |  |  |
| CREATED\_DATE\_TIME | SECONDDATE |  |  |  | 创建日期 |
| CREATED\_USER | NVARCHAR(20) |  |  |  | 创建用户，USR.USER\_ID |
| MODIFIED\_DATE\_TIME | SECONDDATE |  |  |  | 最后修改日期 |
| MODIFIED\_USER | NVARCHAR(20) |  |  |  | 最后修改用户，USR.USER\_ID |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table Name** | **TX\_ORIGINAL\_PROCESS\_DATA** | | | | |
| Column Name | SQL Data Type | Key | Not Null | Default | Comment |
| HANDLE | NVARCHAR(255) | X | X |  | TxOriginalProcessDataBO:<SITE>,<REMARK>,<SFC>,<RESOURCE\_ID>,<CHANNEL\_ID>,<SEQUENCE\_ID>唯一ID |
| SITE | NVARCHAR(6) |  |  |  |  |
| REMARK | NVARCHAR(80) |  |  |  | 流程文件标识号  T3-20160415-1317-246454\_DCR\_35（测试单号+流程名称+电芯号）要求是全局唯一号 |
| SFC | NVARCHAR(50) |  |  |  | 电芯Barcode |
| RESOURCE\_ID | NVARCHAR(64) |  |  |  | 设备号 |
| CHANNEL\_ID | INTEGER |  |  |  | 通道号 |
| SEQUENCE\_ID | INTEGER |  |  |  | 顺序号 |
| CYCLE | INTEGER |  |  |  | 设备循环号 |
| STEP\_ID | INTEGER |  |  |  | 工步号 |
| STEP\_NAME | NVARCHAR(10) |  |  |  | 工步名称 |
| STEP\_LOGIC\_NUMBER | INTEGER |  |  |  | 逻辑工步号 |
| TEST\_TIME\_DURATION | DECIMAL(20,0) |  |  |  | 测试时间 |
| TIMESTAMP | SECONDDATE |  |  |  | 绝对时间 |
| SV\_IC\_RANGE | DECIMAL(20,5) |  |  |  | 电流量程 |
| SV\_IV\_RANGE | DECIMAL(20,5) |  |  |  | 电压量程 |
| PV\_VOLTAGE | DECIMAL(20,5) |  |  |  | 电压 |
| PV\_CURRENT | DECIMAL(20,5) |  |  |  | 电流 |
| PV\_IR | DECIMAL(20,5) |  |  |  | 内阻 |
| PV\_TEMPERATURE | DECIMAL(20,5) |  |  |  | 温度 |
| PV\_CHARGE\_CAPACITY | DECIMAL(20,5) |  |  |  | 充电容量 |
| PV\_DISCHARGE\_CAPACITY | DECIMAL(20,5) |  |  |  | 放电容量 |
| PV\_CHARGE\_ENERGY | DECIMAL(20,5) |  |  |  | 充电能量 |
| PV\_DISCHARGE\_ENERGY | DECIMAL(20,5) |  |  |  | 放电能量 |
| PV\_SUB\_CHANNEL\_1 | NVARCHAR(500) |  |  |  | 辅助通道1的值 |
| PV\_SUB\_CHANNEL\_2 | NVARCHAR(500) |  |  |  | 辅助通道2的值 |
| PV\_SUB\_CHANNEL\_3 | NVARCHAR(500) |  |  |  | 辅助通道3的值 |
| PV\_SUB\_CHANNEL\_4 | NVARCHAR(500) |  |  |  | 辅助通道4的值 |
| PV\_SUB\_CHANNEL\_5 | NVARCHAR(500) |  |  |  | 辅助通道5的值 |
| PV\_SUB\_CHANNEL\_6 | NVARCHAR(500) |  |  |  | 辅助通道6的值 |
| PV\_DATA\_FLAG | INTEGER |  |  |  | 数据类型 |
| PV\_WORK\_TYPE | INTEGER |  |  |  | 流程标识 |
| TX\_IS\_EXCEPTIONAL | NVARCHAR(5) |  |  |  | 是否为异常数据，true/false |
| TX\_ALERT\_CURRENT | NVARCHAR(6) |  |  |  | 是否为电流异常，true/false |
| TX\_ALERT\_VOLTAGE | NVARCHAR(6) |  |  |  | 是否为电压异常，true/false |
| TX\_ALERT\_TEMPERATURE | NVARCHAR(6) |  |  |  | 是否为温度异常，true/false |
| TX\_ALERT\_CAPACITY | NVARCHAR(6) |  |  |  | 是否为容量异常，true/false |
| TX\_ALERT\_DURATION | NVARCHAR(6) |  |  |  | 是否为时间异常，true/false |
| TX\_ALERT\_CATEGORY1 | NVARCHAR(6) |  |  |  | 预留字段 |
| TX\_ALERT\_CATEGORY2 | NVARCHAR(6) |  |  |  | 预留字段 |
| TX\_ROOT\_REMARK | NVARCHAR(80) |  |  |  |  |
| ST\_BUSINESS\_CYCLE | INTEGER |  |  |  | 业务循环号 |
| CREATED\_DATE\_TIME | SECONDDATE |  |  |  | 创建日期 |
| CREATED\_USER | NVARCHAR(20) |  |  |  | 创建用户，USR.USER\_ID |
| MODIFIED\_DATE\_TIME | SECONDDATE |  |  |  | 最后修改日期 |
| MODIFIED\_USER | NVARCHAR(20) |  |  |  | 最后修改用户，USR.USER\_ID |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table Name** | **TX\_RESULT\_PROCESS\_DATA** | | | | |
| Column Name | SQL Data Type | Key | Not Null | Default | Comment |
| HANDLE | NVARCHAR(255) | X | X |  | TxResultProcessDataBO:<SITE>,<REMARK>,<SFC>,<ST\_BUSINESS\_CYECLE>,<STEP\_ID>,<RESULT\_KEY> |
| SITE | NVARCHAR(6) |  |  |  |  |
| REMARK | NVARCHAR(80) |  |  |  |  |
| SFC | NVARCHAR(50) |  |  |  | 电芯Barcode |
| RESULT\_KEY=最后个放电容量，变容量保持率 | NVARCHAR(45) |  |  |  | 数据键 |
| RESLUT\_VALUE | NVARCHAR(45) |  |  |  | 数据值 |
| PROCESS\_ID | NVARCHAR(50) |  |  |  | 流程号 |
| ST\_BUSINESS\_CYCLE | INTEGER |  |  |  | 业务循环号 |
| STEP\_ID | INTEGER |  |  |  | 工步号 |
| SECOND\_CALC\_TAG | NVARCHAR(45) |  |  |  | 二次计算别名 |
| ALIAS | NVARCHAR(45) |  |  |  | 别名 |
| CREATED\_DATE\_TIME-更新日期 | SECONDDATE |  |  |  | 创建日期 |
| CREATED\_USER | NVARCHAR(20) |  |  |  | 创建用户，USR.USER\_ID |
| MODIFIED\_DATE\_TIME-更新日期，同样 | SECONDDATE |  |  |  | 最后修改日期 |
| MODIFIED\_USER | NVARCHAR(20) |  |  |  | 最后修改用户，USR.USER\_ID |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table Name** | **TX\_ZIP\_PROCESS\_DATA** | | | | |
| Column Name | SQL Data Type | Key | Not Null | Default | Comment |
| HANDLE | NVARCHAR(255) | X | X |  | TxZipProcessDataBO:<SITE>,<REMARK>,<SFC>,<RESOURCE\_ID>,<CHANNEL\_ID>,<SEQUENCE\_ID>,<ZIP\_CATEGORY> |
| SITE | NVARCHAR(6) |  |  |  |  |
| REMARK | NVARCHAR(80) |  |  |  | 流程文件标识号  T3-20160415-1317-246454\_DCR\_35（测试单号+流程名称+电芯号）要求是全局唯一号 |
| SFC | NVARCHAR(50) |  |  |  | 电芯Barcode |
| RESOURCE\_ID | NVARCHAR(64) |  |  |  | 设备号 |
| CHANNEL\_ID | INTEGER |  |  |  | 通道号 |
| SEQUENCE\_ID | INTEGER |  |  |  | 顺序号 |
| ZIP\_CATEGORY | NVARCHAR(50) |  |  |  | 压缩类型(I,V,T,C,E) |
| CYCLE | INTEGER |  |  |  | 设备循环号 |
| STEP\_ID | INTEGER |  |  |  | 工步号 |
| STEP\_NAME | NVARCHAR(10) |  |  |  | 工步名称 |
| TEST\_TIME\_DURATION | DECIMAL(20,5) |  |  |  | 测试时间 |
| TIMESTAMP | SECONDDATE |  |  | CURRENT\_TIMESTAMP | 绝对时间 |
| SV\_IC\_RANGE | DECIMAL(20,5) |  |  |  | 电流量程 |
| SV\_IV\_RANGE | DECIMAL(20,5) |  |  |  | 电压量程 |
| PV\_VOLTAGE | DECIMAL(20,5) |  |  |  | 电压 |
| PV\_CURRENT | DECIMAL(20,5) |  |  |  | 电流 |
| PV\_IR | DECIMAL(20,5) |  |  |  | 内阻 |
| PV\_TEMPERATURE | DECIMAL(20,5) |  |  |  | 温度 |
| PV\_CHARGE\_CAPACITY | DECIMAL(20,5) |  |  |  | 充电容量 |
| PV\_DISCHARGE\_CAPACITY | DECIMAL(20,5) |  |  |  | 放电容量 |
| PV\_CHARGE\_ENERGY | DECIMAL(20,5) |  |  |  | 充电能量 |
| PV\_DISCHARGE\_ENERGY | DECIMAL(20,5) |  |  |  | 放电能量 |
| PV\_SUB\_CHANNEL\_1 | NVARCHAR(500) |  |  |  | 辅助通道1的值 |
| PV\_SUB\_CHANNEL\_2 | NVARCHAR(500) |  |  |  | 辅助通道2的值 |
| PV\_SUB\_CHANNEL\_3 | NVARCHAR(500) |  |  |  | 辅助通道3的值 |
| PV\_SUB\_CHANNEL\_4 | NVARCHAR(500) |  |  |  | 辅助通道4的值 |
| PV\_SUB\_CHANNEL\_5 | NVARCHAR(500) |  |  |  | 辅助通道5的值 |
| PV\_SUB\_CHANNEL\_6 | NVARCHAR(500) |  |  |  | 辅助通道6的值 |
| PV\_DATA\_FLAG | INTEGER |  |  |  | 数据类型 |
| PV\_WORK\_TYPE | INTEGER |  |  |  | 流程标识 |
| TX\_IS\_EXCEPTIONAL | NVARCHAR(5) |  |  |  | 是否为异常数据，true/false |
| TX\_ALERT\_CURRENT | NVARCHAR(6) |  |  |  | 是否为电流异常，true/false |
| TX\_ALERT\_VOLTAGE | NVARCHAR(6) |  |  |  | 是否为电压异常，true/false |
| TX\_ALERT\_TEMPERATURE | NVARCHAR(6) |  |  |  | 是否为温度异常，true/false |
| TX\_ALERT\_CAPACITY | NVARCHAR(6) |  |  |  | 是否为容量异常，true/false |
| TX\_ALERT\_DURATION | NVARCHAR(6) |  |  |  | 是否为时间异常，true/false |
| TX\_ALERT\_CATEGORY1 | NVARCHAR(6) |  |  |  | 预留字段 |
| TX\_ALERT\_CATEGORY2 | NVARCHAR(6) |  |  |  | 预留字段 |
| TX\_ROOT\_REMARK | NVARCHAR(80) |  |  |  |  |
| ST\_BUSINESS\_CYCLE | INTEGER |  |  |  | 业务循环号 |
| CREATED\_DATE\_TIME | SECONDDATE |  |  | CURRENT\_TIMESTAMP | 创建日期 |
| CREATED\_USER | NVARCHAR(20) |  |  |  | 创建用户，USR.USER\_ID |
| MODIFIED\_DATE\_TIME | SECONDDATE |  |  | CURRENT\_TIMESTAMP | 最后修改日期 |
| MODIFIED\_USER | NVARCHAR(20) |  |  |  | 最后修改用户，USR.USER\_ID |

### 4.3.2 MongoDB对接MLP数据需求

**技术要求**：

1.MongoDB新开辟一块存储空间，存放MLP应用需要的数据。

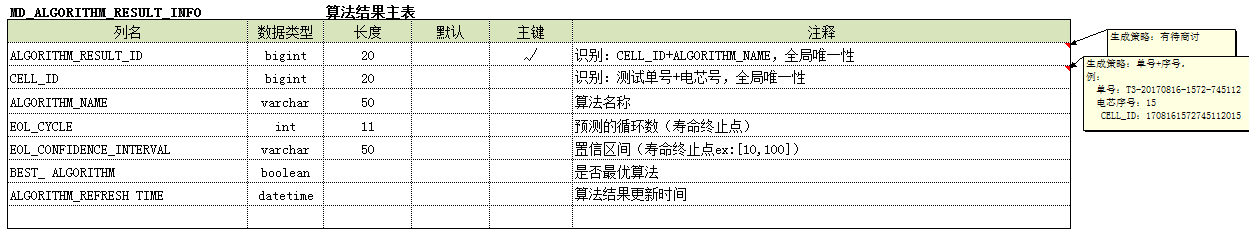
2.所有表结构生成参考MySQL

2.1 新开辟的存储空间存储如下四张新建表

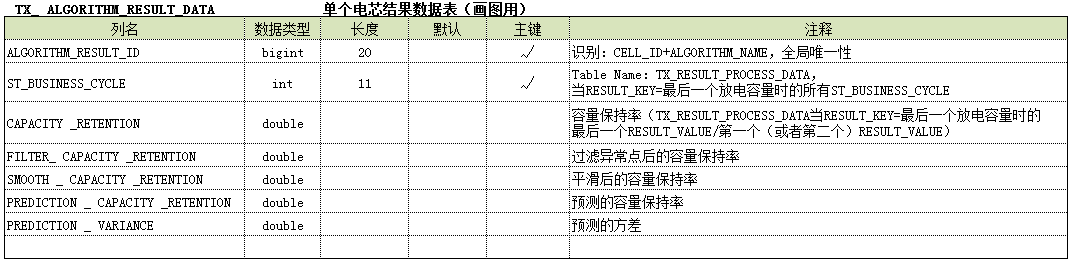
2.1.1 基础信息主表MD\_ALGORITHM\_BASE\_INFO



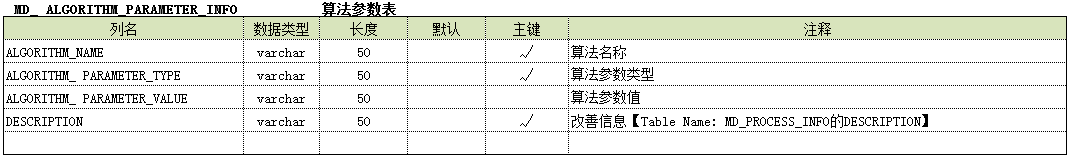
2.1.2算法结果主表MD\_ALGORITHM\_RESULT\_INFO



2.1.3单个电芯结果数据表（画图用） TX\_ ALGORITHM\_RESULT\_DATA



2.1.4算法参数表MD\_ ALGORITHM\_PARAMETER\_INFO



2.2上述四张新建表中的字段，部分来源6.3.1中的五张表，注视有标明的即是。通过调用五张表中对应的表名和字段，即可生成；

2.3注视中未标明来源字段，需在新建四张表中新建字段；

2.4新建四张表中的字段，随着MLP的操作而随之变化（增删改）；

### 4.3.2 spark计算要求

**技术要求：**

1. 支持底层MongoDB数据的输出；
2. 支持MLP算法的数据输入；
3. 支持MongoDB-spark-MLP的数据链路互通；

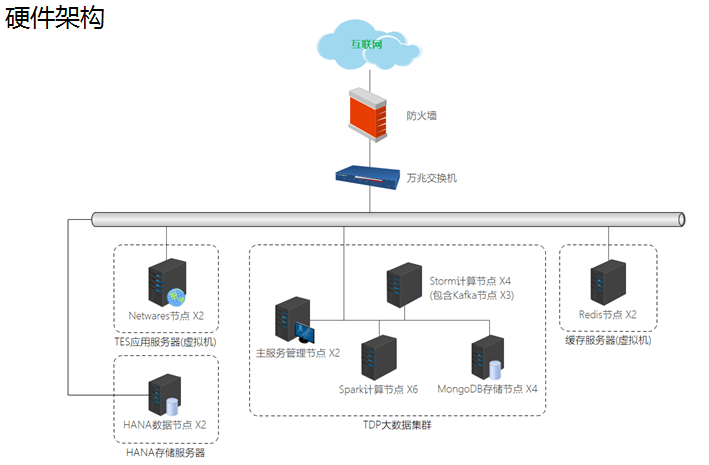
### 4.3.3 MLP应用要求

### 4.3.4 HUE应用要求

### 4.3.5 BO集成要求

### 4.3.6 portal集成要求

### 4.3.7 硬件配置要求



## 4.4 性能要求

### 4.4.1功能性要求

功能性指与一组功能及其指定的性质有关的一组属性，这里的功能是指满足明确或者隐含的需求的那些功能。具体包括：

• 适合性：与规定任务能否提供一组功能，以及这组功能的适合程度有关的软件属性，例如面向任务系统中由子功能构成的功能是否合适，表容量是否合适等等。

• 准确性：与能否得到正确或者相符的结果或者效果有关的软件属性。

• 互操作性：与其他指定系统进行交互的能力有关的软件属性。

• 依从性：使软件遵循有关的标准约定法规及类似规定的软件属性。

• 安全性：即与防止对程序技术局的非授权的故意或者意外访问的能力有关的软件属性。如用户权限、动态口令、数据库字段加密等。

对于这组非功能需求来说，绝大部分是满足功能需求的情况，他并不需要采用额外的措施，而安全性是一个例外，它会涉及具体的技术性功能需求。2

### 4.4.2可靠性

可靠性之与在规定的一段时间和条件下软件维持其性能水平的能力有关的一组属性。具体包括：

• 成熟性：与有软件故障引起失效的频度有关的软件属性。

• 容错性：与在软件故障或违反指定接口的情况下维持规定的性能水平的能力有关的软件属性。如离线录入支持等。

• 易恢复性：与在是小发生后重建其性能水平并恢复直接受影响数据的能力，以及为达到此目的所需时间和努力有关的软件属性。如表单数据自动保存等。

这类非功能需求通常是全局的，他除了与系统运行环境、平台选择、代码质量相关之外，还会涉及部分技术性功能需求，他别是容错性、易恢复性的实现都需要一些具体的功能来支持。

### 4.4.3易用性

易用性是与一组规定或者潜在的用户为使用其软件所需做的努力和对这样的使用所作的评价有关的一组属性。具体包括：

• 易理解性：与用户为人质逻辑概念即其应用范围所花的努力有关的软件属性。

• 易学习性：与用户为学习软件应用所花的努力有关的软件属性。

• 易操作性：与用户为操作和运行控制所花的努力有关的软件属性。如带首字母筛选功能的下拉列表等。

这类非功能需求是与UI设计、联机帮助系统有着直接关系的，易理解性和易学习性通常和界面导航、联机帮助有关，课归纳为界面友好性;易操作性则会和界面元素设计有关。也就是说这类属性会关联到具体的技术性功能需求。

### 4.4.4效率

效率是指与在规定的条件下软件的性能水平与所使用资源量有关的一组属性。具体如下：

• 时间特性：与软件执行器功能时响应和处理时间及吞吐量有关的软件属性。如数据缓存等。

• 资源特性：与在软件执行其功能时所使用的资源数量及其使用时间有关的软件属性。如数据压缩等。

这部分实际上就是通常所说的性能需求，他有一大部分是局部性的，在每个用力的描述中应该指出;另外它又会引申出一些相关的技术性功能需求，例如数据缓存等。

### 4.4.5维护性

维护性是指与进行指定的修改所需的努力有关的一组属性。具体包括：

• 易分析性：与为诊断缺陷或者失效原因及为判定待修改的部分所需努力有关的软件属性。如日志记录系统等。

• 易改变性：与进行修改排除错误或者适应环境变化所需努力有关的软件属性。

• 稳定性：与修改所造成的未预料结果的风险有关的软件属性。

• 易测试性：与确认已修改软件所需的努力有关的软件属性。

这部分通常是开发团队最容易投入时间和成本的地方，诸如动态属性支持、UI界面生成、流程引擎等都是为了提高系统的可维护性，因此它显然是会引申出相关的技术性功能需求的。

### 4.4.6可移植性

可移植性是指与软件可从某一环境转移到另一环境的能力有关的一组属性。具体包括：

• 适应性：与软件无需采用有别于为该软件准备的活动和手段就可能适应不同的规定环境有关的软件属性。如全球技术支持等。

• 易安装性：与在指定的环境下安装软件所需努力有关的软件属性。如在线更新、安装包自动生成等。

• 遵循性：使软件遵循与可移植性有关的标准或约定的软件属性。项目管理论坛

• 可替换性：与软件在该环境中用来替代指定的其他软件的机会和努力有关的软件属性。

这部分除了需要通过选择正确的开发工具、平台来支持外，也会涉及一些技巧性的功能需求，如全球语言支持等。