|  |
| --- |
|  |
| **平台组（Platform\_V1.11）** |
| **数据平台解决方案说明书** |
|  |
|  |
|  |

|  |  |  |
| --- | --- | --- |
| 文件状态  [√]草稿  [ ]正式  [ ]修订 | 文件编号 |  |
| 文档名称 | 数据平台解决方案说明书 |
| 当前版本 | 0.1 |
| 文档类型 | 项目文档 |
| 发布部门 | 开发部 |
| 创建日期 | 2017/02/07 |
| 最后修订日期 | 2017/02/24 |
| 文档作者 | 吴雄波 |

文档修订记录

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 版本 | 文档状态 | 变化状态 | 变更内容 | 日期 | 变更人 | 备注 |
| 0.1 | 草稿 | A | 创建文档 | 2017/02/07 | 吴雄波 | 创建文档 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

\* 文档版本： <主版本号.子版本号>

\* 文档状态： 草稿， 正式，修订

\* 状态变化： A – 增加内容， M – 修改内容，D – 删除内容。可以组合使用，如AMD。

\* 日期格式： yyyy/mm/dd

文档评审记录

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本 | 日期 | 评审人 | 评审结果 | 备注 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

\* 文档版本： <主版本号.子版本号>

\* 日期格式： yyyy/mm/dd

\* 评审结果： 通过，未通过，修改后通过

目录

[1 文档介绍 5](#_Toc476554860)

[1.1 文档背景 5](#_Toc476554861)

[1.2 文档目的 5](#_Toc476554862)

[1.3 读者对象 5](#_Toc476554863)

[1.4 术语与缩写解释 5](#_Toc476554864)

[2 方案设计 6](#_Toc476554865)

[2.1 方案描述 6](#_Toc476554866)

[2.2 总体构架 7](#_Toc476554867)

[2.3 运行流程 7](#_Toc476554868)

[2.3.1 数据查询 7](#_Toc476554869)

[2.3.2 数据同步 8](#_Toc476554870)

[2.3.3 流程时序图 9](#_Toc476554871)

[2.4 数据服务站点 10](#_Toc476554872)

[2.4.1 数据查询接口 10](#_Toc476554873)

[2.4.2 数据同步接口 11](#_Toc476554874)

[2.5 作业 12](#_Toc476554875)

[2.5.1 作业配置 12](#_Toc476554876)

[2.5.2 查询作业 13](#_Toc476554877)

[2.5.3 同步作业 13](#_Toc476554878)

[2.6 数据平台 13](#_Toc476554879)

[后台管理 13](#_Toc476554880)

[作业产生 13](#_Toc476554881)

[作业队列 14](#_Toc476554882)

[Cache 14](#_Toc476554883)

[作业调度 14](#_Toc476554884)

[数据存取 14](#_Toc476554885)

[数据转换 14](#_Toc476554886)

[Worker 14](#_Toc476554887)

[监控 14](#_Toc476554888)

[日志 15](#_Toc476554889)

[容错处理 15](#_Toc476554890)

[更新 15](#_Toc476554891)

[3 实施部署 15](#_Toc476554892)

[3.1 部署 15](#_Toc476554893)

# 文档介绍

## 文档背景

公司产品及第三方产品之间产生了错综复杂的业务联系，有很多数据交互的需求。现各产品之间使用直连方式进行数据交互，导致各产品之间对接时接口不统一、开发工作量大、维护困难等诸多问题。现由平台组提供数据平台解决方案，解决上述问题。

## 文档目的

本文将阐述多个业务系统之间如何通过数据平台进行数据交互，用户如何配置与使用数据平台，方案部署等。

## 读者对象

开发人员，产品人员，运维人员，管理人员

## 术语与缩写解释

|  |  |
| --- | --- |
| **缩写、术语** | **解释** |
|  |  |
|  |  |
|  |  |
|  |  |

# 方案设计

## 方案描述

数据平台作为各业务系统（包括公司内部系统及第三方系统）之间数据交互中心节点，与各业务系统进行直连交互。一个业务系统与多个业务系统进行数据交互时，不再需要开发和维护多套交互接口，只需与数据平台交互即可做到。将原各业务系统之间的数据交互的那种复杂网状图（包含数据同步服务）（如图1），改成星形交互图（如图2）。

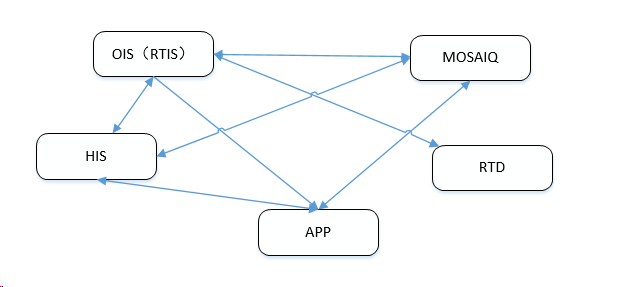


图1 网状式交互图

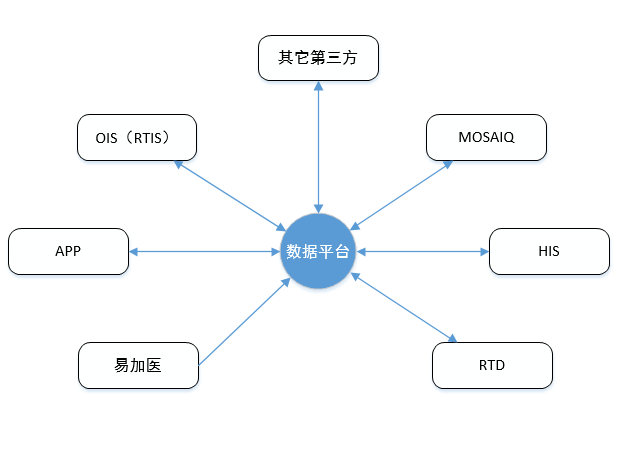


图2星形交互图

## 总体构架



## 运行流程

数据平台解决方案主要为解决两类问题：

1. 业务系统数据查询：业务系统发出请求，另一个业务系统响应，通过数据平台来完成请求与响应的对接。
2. 业务系统数据同步：业务系统中部分数据需要与另一个业务系统保持一致，通过数据平台来保证前者与后者的数据变化一致。

### 数据查询

数据平台提供两种查询业务数据方式：

直接从业务系统查询数据；

从数据平台中心库查询数据（中心库的数据来源于业务系统数据同步）。

两种来源通过接口参数控制，查询速度方面，优于，而数据实时性方面，优于，调用方可按需选择。

数据查询基本运行流程如下：

1. 业务系统通过http向服务接口发起查询请求，传入业务名称及相关参数(业务名称关联作业配置，依据配置可以知道从何处取数据，取哪些数据) ；
2. 数据平台根据业务名称获取详细配置信息，并根据请求参数生成相关作业；
3. 如果是同步请求，则等作业完成后再响应请求；
4. 如果是异步，则临时保存回调接口并立即响应请求，待作业完成后执行回调接口推送数据。

图4数据查询作业流程图

### 数据同步

数据同步的目的是保证两个业务系统的部分或全部数据一致，数据平台触发数据同步有三种机制：

1. 接口触发：业务系统调用数据同步接口后触发，数据提供者主动调用接口通知数据平台数据有更新。
2. 定时触发：定时启动同步数据作业，数据平台并不知道业务系统数据是否有更新，只能轮询检查。
3. 依赖触发：作业链形式，前置作业完成后启动后置作业。一般用于有一定前后关联的作业上。

图5 数据同步作业流程图

外部通知消息一般来源于提供数据的业务系统主动调用，即在业务系统本身数据更新后，调用数据同步接口。

### 流程时序图

基本流程时序图示意如下：

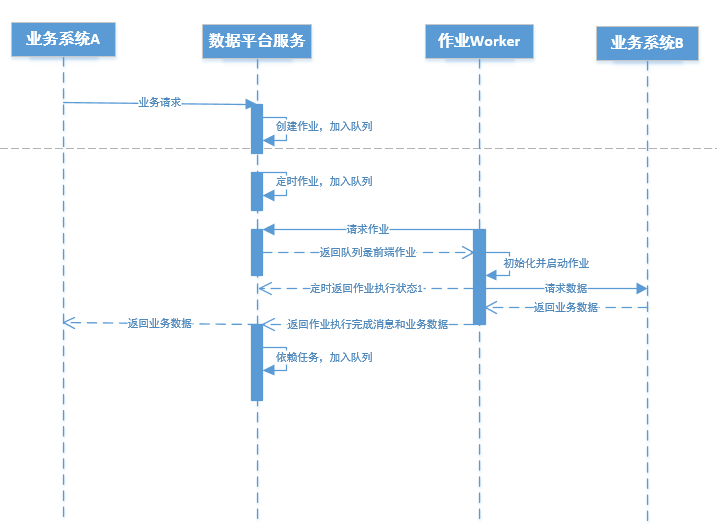


图6 基本流程时序图

## 数据服务站点

接收数据查询及数据同步请求。

### 数据查询接口

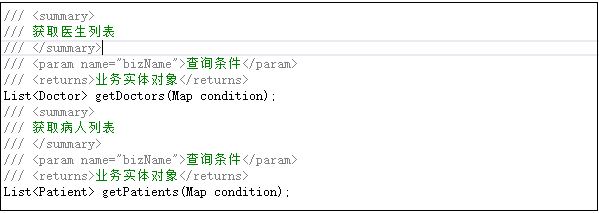
同时提供同步与异步接口，同步接口与异步接口区别仅在于异步接口比同步接口多一个回调参数，即回调地址。

调用方可以使用数据平台提供的Jar包调用接口获取数据，也可通过Http方式发起请求调用获取数据。两者均提供通用接口和专用数据访问接口，通用接口是对专用业务访问接口的补充。通用访问接口中的参数bizName的作用，类似于Jar包中的方法调用地址，Http发起请求的请求地址。

通用接口定义：

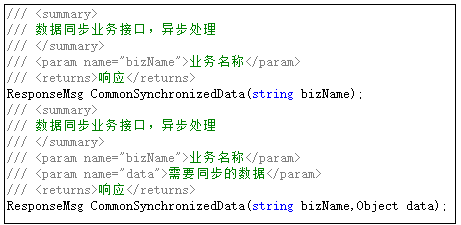


数据方问专用接口定义示例：



### 数据同步接口

业务系统主动发起数据同步请求。



## 作业

### 作业配置

作业配置为数据查询作业和数据同步作业运行所需要部分参数。接口调用时，提供业务名称就可关联出这部分参数，再结合接口中传递的参数，来完成作业所需做的工作。在使用时，每一个作业的参数都会提供模板供参考。

基本项配置大致示意如下：

|  |  |
| --- | --- |
| 配置id | 唯一标识 |
| 名称 | 唯一 |
| 作业类型 | 查询或同步 |
| 作业组别 | worker可领取指定组别的作业 |
| 是否可并发执行 | 同一id的作业能否并发执行 |
| 是否启用 |  |
| 触发类型 | 定时触发，依赖触发或接口触发 |
| 触发参数 | 定时触发参数、依赖触发参数、接口触发参数 |

作业参数配置项，大致示意如下：

|  |  |
| --- | --- |
| 参数id |  |
| 配置id | 与作业基本配置id关联 |
| 参数类型 | 参数类型，参数名，参数值等由各类作业识别并使用。 |
| 参数名 |
| 参数值 |

### 查询作业

查询作业，由数据存取模块向目标业务系统取数据，然后经过数据转换模块处理后，将数据返回给数据平台即可，后续步骤由数据平台的数据查询接口完成。

每一个目标业务系统，数据存取模块都应有合适的接口来完成与该业务系统数据存取对接，如果没有，则需要数据存取模块进行扩展开发。数据存取模块是数据平台方案中需要持续开发扩展的模块。

数据转换模块同理，如果没有合适的转换接口，同样需要扩展开发。

### 同步作业

同步作业，由数据存取模块向业务系统A取数据，经数据转换模块处理后，再由数据存取模块将数据写入业务系统B。

数据模块需同时对接业务系统A和业务系统B，同样需要合适的接口来完成数据存取工作。

## 数据平台

提供数据查询服务，数据同步服务，后台管理，作业处理等核心功能。

### 后台管理

Web站点，不对外开放.提供作业配置增删改，查看监控数据，查看作业日志等功能。

### 作业产生

作业产生模块包含三个作业触发器：接口触发器，定时触发器，依赖触发器。各触发器产生作业优先级按高到低为：接口触发器，依赖触发器，定时触发器。

作业产生模块读取作业配置后，依据其触发类型，达到触发条件后立即产生作业，并设立作业优先级后，加入到作业队列中。注：由接口触发器产生的从数据中心查询数据的作业并不会加入到队列中，而是在产生后会立即启动执行。

数据服务接口的所有调用请求都将通过接口触发器处理产生作业。

### 作业队列

负责维护待执行作业队列与正执行作业队列。接收Worker返回的作业消息，更新正执行作业队列。

### Cache

缓存作业配置，缓存作业执行完成后所获取的业务数据。

### 作业调度

负责作业的分配，优先级调度，生命周期控制。

### 数据存取

与业务系统对接，进行数据存取。存取来源包括业务系统，中心库。

### 数据转换

作业执行时，此模块对数据进行必要的计算或转换，转换参数可配置。

### Worker

作业执行者，部署在终端。向数据平台轮询请求作业并启动，收集作业执行状态，进度，数据结果等信息，反馈回数据平台。可按需线性增加部署数量，可指定执行特定组别的作业。

### 监控

监控正在运行的作业，对执行耗时过高告警并记录日志。

监控待执行的作业队列，对待执行数过高告警并记录日志。

监控作业Worker运行状态。

### 日志

收集并记录作业相关的日志信息，如时间消耗，错误日志等信息，存储到数据库中或文件中，可在后台管理站点查看。

### 容错处理

各个环节的容错处理，建立一个完整的容错机制。

### 更新

与Worker通信，负责Worker的更新。

# 实施部署

## 部署

数据平台方案部署包含四部分：数据库，数据平台服务， Worker。

数据库，数据平台服务一般部署在同一网络环境中或可以相互访问的良好网络环境中，Worker部署在终端，业务系统所在网络与公网交互结点上。



图7 部署网络图（互联网）

对于在局域网络中的应用，四部分部署则简单些，基本可以部署在同一机器上，当然也可以分多个机器部署。

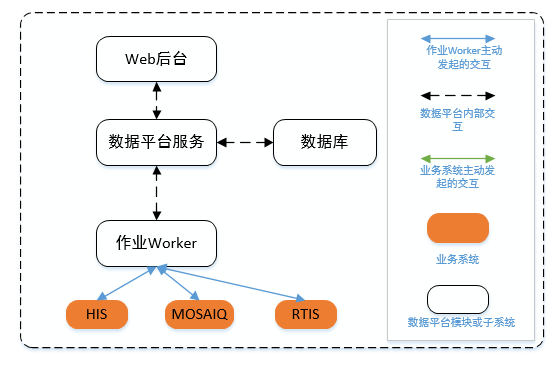


图8 部署网络图（局域网）