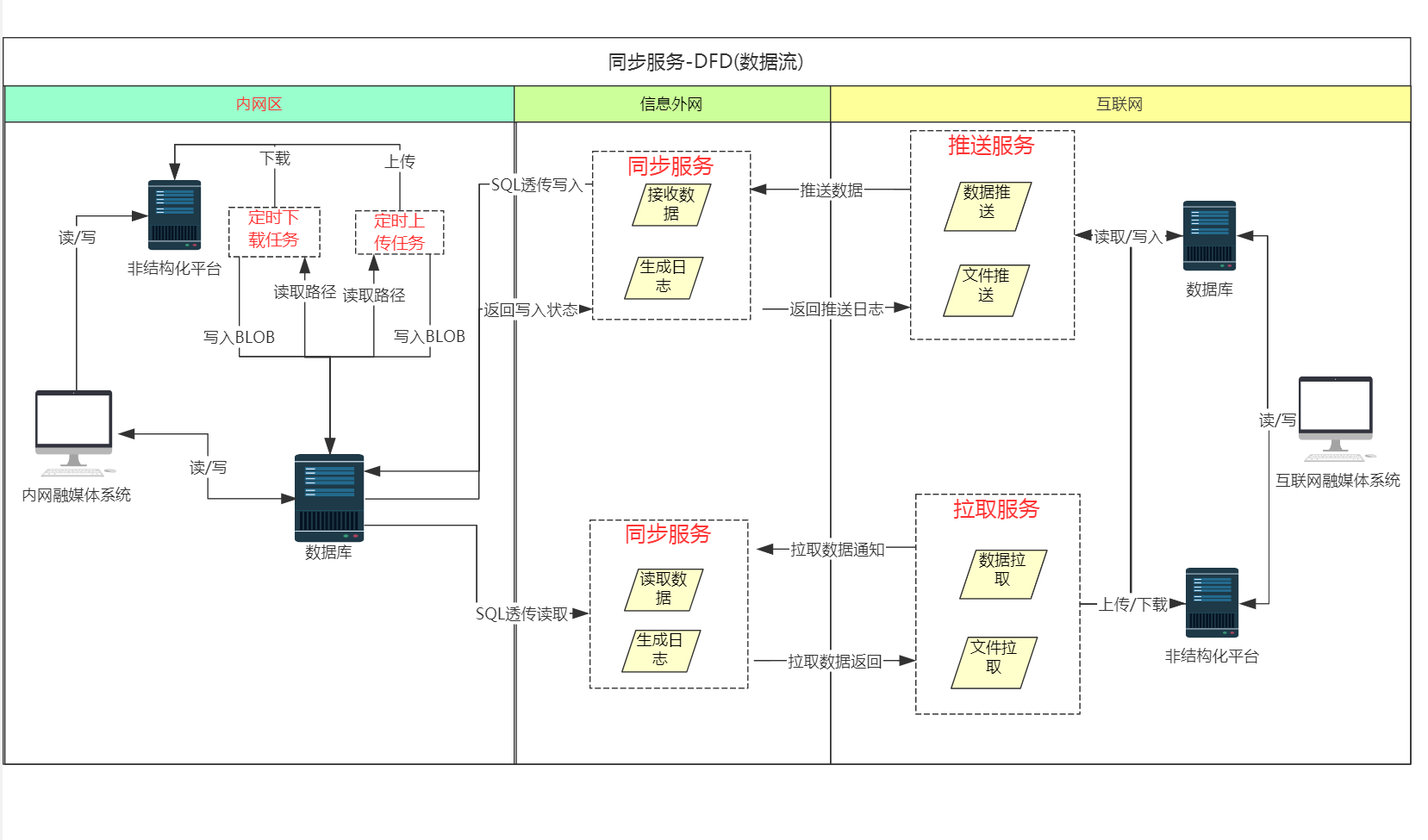
内外网同步实现方案和思路

# 数据流图：



# 实现流程和思路

## 互联网—内网数据同步

### 获取上一次互联网-内网的同步时间、当前数据库时间：

从日志表获取上一次同步成功的时间和获取数据库的当前时间形成待同步数据的时间段。

### 获取待同步的表名字：

从同步配置表获取需要同步的表名字。

### 获取待同步的数据：

通过表名字、上一次同步时间、当前数据库时间查询数据库中待同步的数据，并进行JSON封装。

### 文件路径提取与拆分：

使用异步线程对读取到的数据进行文件路径的提取（如果该地址存在数据中则不在进行文件路径提取，避免多次同步相同的文件），生成文件同步的记录并通过文件路径把文件流下载下来并保存到数据库以供下次文件同步使用，大文件拆分：获取文件流byte大小，通过byte大小和单文件拆分的大小计算该文件需要拆分的小文件个数，使用多线程对文件进行指定大小的读取然后另存为小文件，切分后的小文件名按顺序进行命名，以免在合并时造成小文件的合并顺序错误，最后保存到数据库。

### 发送数据到信息外网同步服务：

加密数据然后通过HTTP发送到信息外网同步服务器。

### 信息外网写入数据到内网数据库：

信息外网接收到来自互联网的数据后提取数据中的文件路径，根据文件路径查找内网数据库中文件路径映射表，如果存在映射则把文件路径替换成内网的地址，通过二代隔离设备写入数据到内网数据库，最后写入数据库完成后生成同步日志返回到互联网。

### 互联网写入同步日志

接收到信息外网返回的同步日志后写入到数据库中，标识该次同步任务的状态成功与否（同步日志统一在互联网记录）。

## 内网—互联网数据同步

### 获取上一次内网-互联网的同步时间、当前数据库时间：

从日志表获取上一次同步成功的时间和获取数据库的当前时间形成待同步数据的时间段。

### 获取待同步的表名字：

从同步配置表获取需要同步的表名字。

### 发送HTTP请求到信息外网：

封装上一次同步时间、当前数据库时间、待同步的表名集合到信息外网读取内网的数据。

### 信息外网读取数据：

信息外网同步服务接收到来自互联网读取内网数据的请求后解析查询参数，查询内网数据库中待同步时间段之内的数据。

### 文件路径提取

对2.2.4中读取到的数据进行文件路径提取，提取思路与2.1.4一致，唯一的区别是内网的文件地址下载只能通过内网的定时任务去下载文件流数据并保存到内网数据库。

### 互联网写入数据

互联网同步服务接收到信息外网返回的数据后进行解密然后写入到互联网数据库并生成同步日志。

## 互联网—内网文件同步

### 读取互联网未同步的文件记录：

读取互联网数据库中2.1.4中提取到的文件记录，和未同步的文件流数据，文件记录通过HTTP加密发送到信息外网同步服务器，文件流则通过异步线程分批次发送到信息外网同步服务。

### 信息外网写入文件到内网：

信息外网接收到互联网发送的文件记录和文件流数据后写入到内网数据库，并返回是否写入成功的标志。

### 互联网修改状态

互联网接收到信息外网同步服务返回的文件同步状态后修改数据库中该文件的同步状态是否成功。

### 内网文件上传

内网同步服务定时去扫描数据库中来自互联网且未上传到内网非结构化平台的文件记录和文件流数据，读取出来后对大文件进行文件合并，然后检验MD5值与记录表中存在的MD5值是否一致，如果一致则上传到内网非结构化平台，上传成功后建立内网与互联网的文件地址映射记录，并修改文件的上传状态为成功。

## 内网—互联网文件同步

内网到外网的

### 内网文件下载 对2.2.5中提取的文件记录进行查询是否已经下载文件流数据到内网数据库，如果未下载则通过非结构化平台提供的SDK接口进行文件下载，对大文件进行拆分后保存到数据库。

### 互联网发送文件记录同步请求：

互联网先发送同步内网的文件记录请求到信息外网，信息外网接收到请求后读取内网数据库中的未同步的文件记录，然后返回到互联网。

### 互联网写入文件记录到数据库：

接收到信息外网返回未同步成功的文件记录后写入数据库，并且再发送通知到信息外网去修改同步成功的记录ID，信息外网接收到同步成功的文件记录ID后修改内网数据库的文记录为同步成功。

### 互联网发送文件流数据同步请求：

1.查询内网未同步成功的文件流数据条数总数和ID号，

2.根据读取到的总数和批次大小计算出要读取文件流数据的线程数，

3.使用多线程进行分批次读取内网的文件数据，

4.读取成功后发送请求到信息外网修改内网的文件流同步状态。

### 上传文件

在这次整个同步任务完成后，再开一个线程统一对来自内网的文件进行上传到文件服务器，上传成功后建立文件路径映射记录，并替换数据库中对该文件路径引用的数据为互联网的文件路径。