

第二组集成场景调研：bilibili

张洪胤 庄子元 程荣鑫 师域飞

一、 应用概述	2
二、 bilibili 集成调研.....	2
(一) 表示集成	2
1. 概述	2
2. 细节描述	3
3. 总结	6
(二) 数据集成	6
1. 概述	6
2. 细节	6
3. 总结	10
(三) 功能集成	10
1. 概述	10
2. 细节	11
3. 总结	12
(四) 业务集成	12
1. 概述	12
2. 细节	13
3. 概述	18

一、应用概述

哔哩哔哩（英文名：bilibili，简称：B 站）是中国青年世代高度聚集的文化社区和视频平台，目前拥有动画、番剧、国创、音乐、舞蹈、游戏、知识、生活、娱乐、鬼畜、时尚、放映厅等 15 个内容分区，生活、娱乐、游戏、动漫、科技是 B 站主要的内容品类并开设直播、游戏中心、周边等业务板块。

近年来，bilibili 的用户数量快速增长，业务快速推进，平台快速扩张，已经逐渐成长成为集诸多功能和业务与一体的巨型平台，本次调研我们针对 bilibili 的 PC 端和手机 APP 端（以下简称 bilibili 客户端），从展示集成、数据集成、功能集成和业务集成四个方面展开，聚焦用户推荐、视频播放等功能。

二、bilibili 集成调研

对于 bilibili 客户端而言，我们在调研过程中发现其不仅仅是指存在某一集成方式，而存在着表示集成、数据集成、功能集成和业务集成这四种集成，分别在不同的方面和层次上完成了集成工作，共同提供了一致且有价值的用户服务。

（一）表示集成

1. 概述

bilibili 客户端为 bilibili 视频弹幕网站、bilibili 直播系统、bilibili 电子商城系统等诸多子系统提供了统一的用户界面，让用户可以通过简单的跳转即可获取到对应的服务。

2. 细节描述

(1) bilibili 视频弹幕网站

bilibili 视频弹幕网站主要功能包括了视频播放、视频上传、弹幕发送、点赞投币等功能。其本身具有自己的 Web 用户界面，页面内容基于其特有的有用户推荐算法，基本上能够做到“千人千面”（比如下图中右侧两张图为不同账号的推荐界面）。



bilibili 不得不说的特色便是“刷屏的弹幕”（如下图所示），处于好奇，我们对弹幕本身进行了进一步调研。



我们任意选择一个视频进行播放，关注弹幕，发现播放过程中，如果我们悬停一个弹幕，点击“点赞按钮”，在 F12 后的 Network 界面我们可以发现一个名为 add 的请求，这个请求的请求信息和请求体信息如下图所示。

Request URL: <https://api.bilibili.com/x/v2/dm/thumbup/add> ▼ Form Data view source view URL encoded

Request Method: POST

Status Code: 200

Remote Address: 112.25.54.211:443

Referrer Policy: no-referrer-when-downgrade

dmid: 46673886542561283

oid: 305845418

op: 1

platform: web_player

csrf: 3ffb35bcceea011b771610302c6cc365

在上图请求中我们可以发现其中有一个参数为 **platform**，由此我们结合 **bilibili** 客户端的多样性可以推测出 **bilibili** 视频弹幕网在这部分完成了表示集成。

(2) bilibili 直播系统

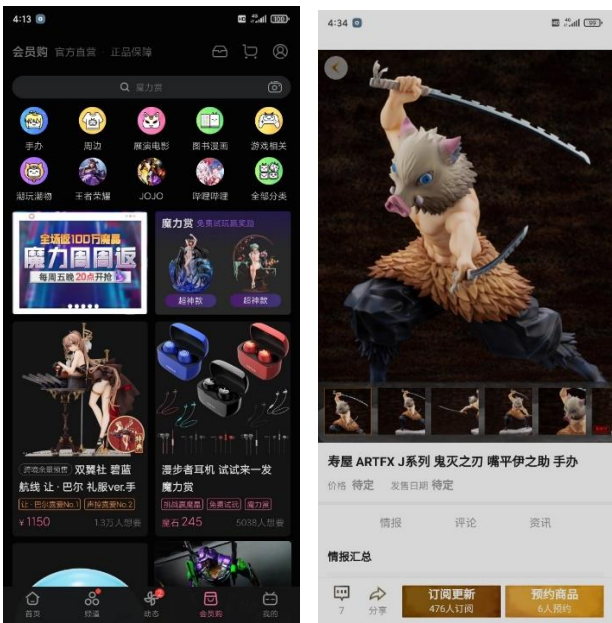
bilibili 直播系统主要功能包括了观看直播、发起直播、送礼物、开舰长等。直播系统有 APP 端界面，如下左图所示，将打赏等部分进行了折叠，也有 PC 端界面，如下右图所示，相对于 APP 端折叠情况有所减少。



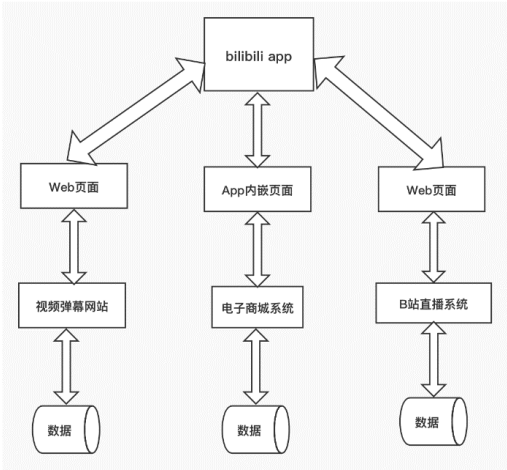
(3) bilibili 电子商城

bilibili 电子商城的主要功能包括了购买商品、支付账单、浏览商城等，但是 **bilibili** 电子商城本身没有独特的用户界面，其本身就伴随着 **bilibili** 客户端而出现，如下左图所示。而电子商城的手机端商品详情界面如下右图所示，虽然电商系统与视频播放系统在功能上、接

口上、实现上并不相似，但是因为 App 集成了通用的用户界面，因此界面风格还是与前两者保持一致。



(4) bilibili 客户端展示集成架构图



bilibili 视频弹幕网站、bilibili 直播系统、bilibili 电子商城系统三者有不同的接口、功能，使用不同的技术和开发方法实现，属于三个不同的系统，但是他们之间存在数据的交换和互通（例如 bilibili 账户信息），而 bilibili APP 为他们三者提供了通用的用户界面，如上图所示，实现了表示集成。

3. 总结

bilibili APP 的表示集成设计结构简单，易于实现，使得用户不必分别使用多个软件系统的界面，而是通过任意一个 bilibili 客户端便可以快捷方便地查看 bilibili 各个子系统的界面，为用户带来了极大的方便。

（二）数据集成

1. 概述

近年来，bilibili 的用户持续增长，覆盖的领域逐渐扩张，比如视频播放、直播、购物等等，在数据集成部分我们聚焦两部分，一部分是为了满足 UP 主流量完成变现的需求而推出的花火平台，另一部分我们聚焦 bilibili 的数据一致性问题，从这两个部分来调研 bilibili 已经完成的数据集成工作。

2. 细节

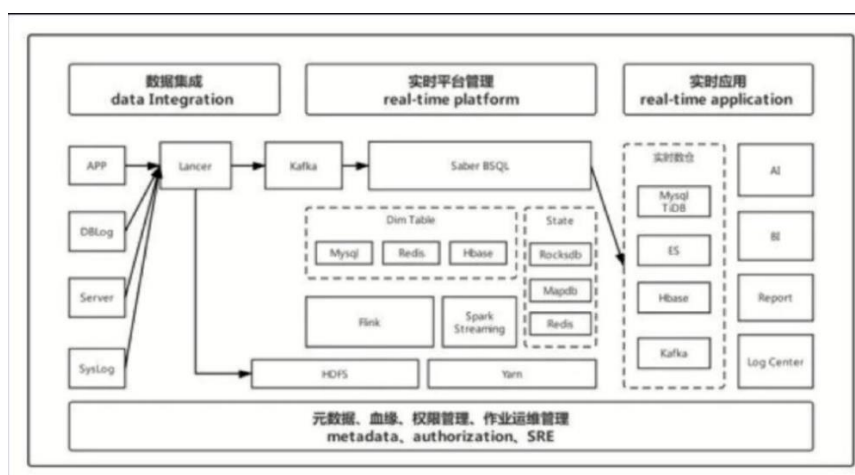
a) 花火平台描述

bilibili 一直有着关于用户增长分析、渠道投放分析等 BI 分析需求，因此 bilibili 选择使用了基于 Apache Flink 的流式计算平台（提供包含 SQL 化编程、DAG 拖拽编程、作业一体化托管运维等功能），即花火平台。花火平台主要涵盖了以下多个方向：

1. AI 工程方向，解决了广告、搜索、推荐的流式 Joiner 和维表 Joiner 问题。

2. 特征支持方向：提供了实时计算的特征支持，支持 Player 以及 CDN 的质量监控，包括了直播、PCU、卡顿率、CDN 质量等多质量指标。
3. 用户增长方向：即如何借助实时计算进行渠道分析、调整渠道投放效果。
4. 实时 ETL 方向：实现包括 Boss 实时播报、实时大屏、看板等功能。

b) 花火平台架构细节



bilibili 的数据集成部分负责实时传输的工作，完成了 APP 日志、数据库 Binlog、服务端日志或系统日志的传输。接下来 bilibili 内部的 Lancer 系统就会解决数据落地到 Kafka 或 HDFS 的问题。

上层核心基于 Flink 构建运行池。再向上一层满足多种维表场景，包括 MySQL、Redis、HBase。状态 (State) 部分在 RocksDB 基础上，还扩展了 MapDB、Redis。Flink 需要 IO 密集是很麻烦的问题，因为 Flink 的资源调度体系内有内存和 CPU，但 IO 单位未做统一管理。

当某一个作业对 IO 有强烈的需求时，需要分配很多以 CPU 或内存为单位的资源，且未必能够很好的满足 IO 的扩展。所以本质上 bilibili 现阶段是将 IO 密集的资源的状态转移到 Redis 上做缓解。数据经过 BSQL 计算完成之后传输到实时数仓，如 Kafka、HBase、ES 或 MySQL、TiDB。最终到 AI 或 BI、报表以及日志中心。

c) bilibili 推荐机制与数据源

bilibili 十分注重首页推荐机制，其推荐算法主要是先将内容进行分类，然后筛选出优秀的作品，推荐主要权重落在了用户的喜好上。

之前暴露出的视频权重的计算方式是：硬币*0.4+收藏*0.3+弹幕*0.4+评论*0.4+播放*0.25+点赞*0.4+分享*0.6，最新视频（一天内发布）提权[总值*1.5]，如下图所示。

```
//countArchHot: 视频-硬币*0.4+收藏*0.3+弹幕*0.4+评论*0.4+播放*0.25+点赞*0.4+分享*0.6 最新视频（一天内发布）提权[总值*1.5]
func countArchHot(t *api.Stat, ptime int64) int64 {
    if t == nil {
        return 0
    }
    hot := float64(t.Coin)*0.4 +
        float64(t.Fav)*0.3 +
        float64(t.Danmaku)*0.4 +
        float64(t.Reply)*0.4 +
        float64(t.View)*0.25 +
        float64(t.Like)*0.4 +
        float64(t.Share)*0.6
    if ptime >= time.Now().AddDate(0, 0, -1).Unix() && ptime <= time.Now().Unix() {
        hot *= 1.5
    }
    return int64(math.Floor(hot))
}

// countArchHot: 专栏-硬币*0.4+收藏*0.3+评论*0.4+阅读*0.25+点赞*0.4+分享*0.6 最新专栏（一天内发布）提权[总值*1.5]
```

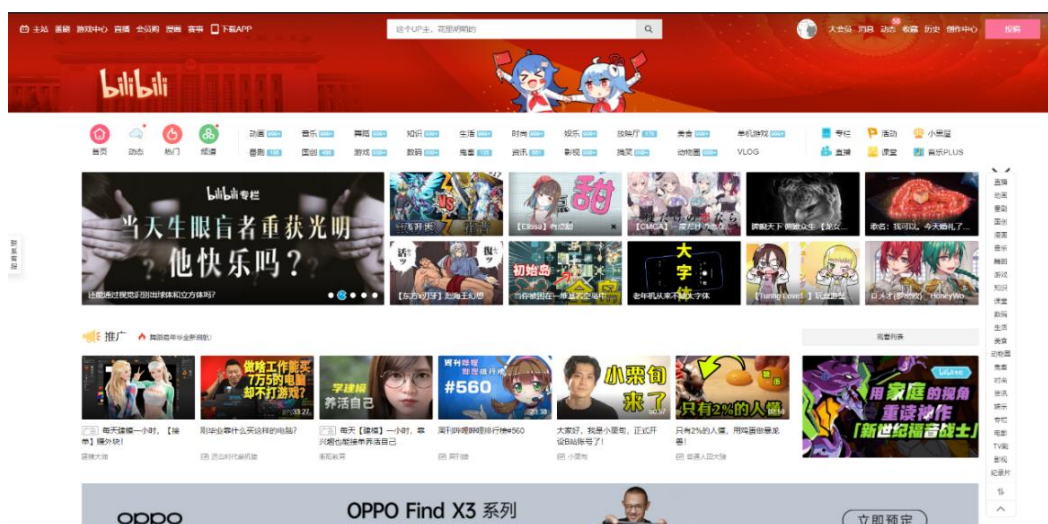
可见 bilibili 的推荐主要是基于视频本身属性的推荐。但对于用户本身来说，可能是基于关注的 up 与常看视频的标签进行推荐。

（bilibili 是直接基于视频进行推荐，不可能检索所有的视频，故采用推荐算法每次运行时，其检索的视频数量一直保持相对稳定。）

因此，bilibili 获取的数据源有很多种，在视频方面，首先有视频本身的信息与标签，视频制作者/上传者的信息，在用户方面，有用户

本身个人信息，关注的 up 主信息，常看的视频标签，常搜索的关键词，购买记录等等。在数据源的获取，在不同的终端有不同的方式。

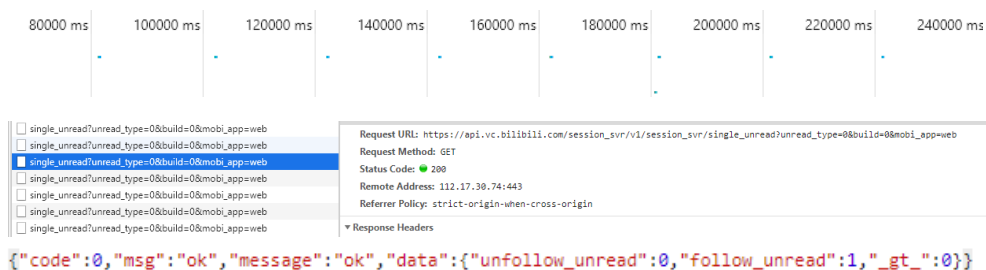
bilibili 根据这些丰富的数据源，通过实时的流计算，为 AI 推荐、搜索提供支持，从而达成了上文中提到的“千人千面”，得到如下图所示的首页界面。



d) bilibili 的数据一致性问题：以未读信息为例

经过了之前的表示集成的分析，我们已经充分意识到 bilibili 多客户端之间容易出现信息不一致的问题，我们选择未读信息为例，从网页端入手尝试找到不同客户端之间的未读信息的同步问题。

进入我的消息界面，我们发现了如下的 single_unread 请求，其参数 mobi_app 指出这个是 web 界面，同时我们可以看到这个请求大约是 20s 发送一次，来获取最新的未读信息，而返回体也如下图所示。



由此我们猜测，未读信息的部分使用了数据集成的方法，使用 `mobi_app` 参数来区分前台，统一从数据集成后的后台拿到数据，从而保证一致性。

3. 总结

bilibili 作为一个内容导向的视频网站，不仅需要处理用户画像，还需要满足 **AI** 推荐场景下的实时计算，也需要保证其内部的数据一致性。其数据集成方面做到了以下的三个方面：

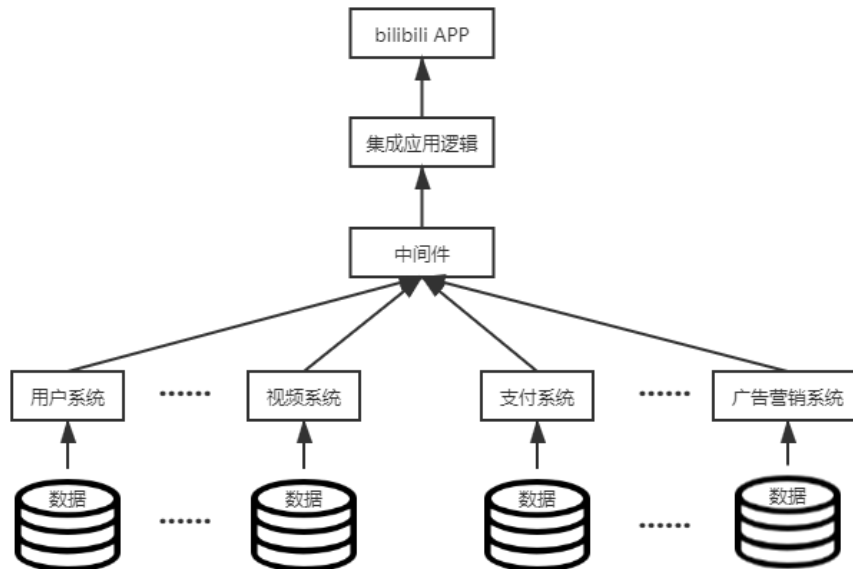
1. 更广泛的数据访问，整个的数据输入来自于多场景，但输出统一于 **Lancer**，方便新数据源的集成。并且采用多种不同的数据库来辅助实时平台的管理。
2. 解决了运维困难，方便系统逻辑演变的维护工作。
3. 充分挖掘大量数据中蕴藏的价值，提供出更加有价值的服务。

（三）功能集成

1. 概述

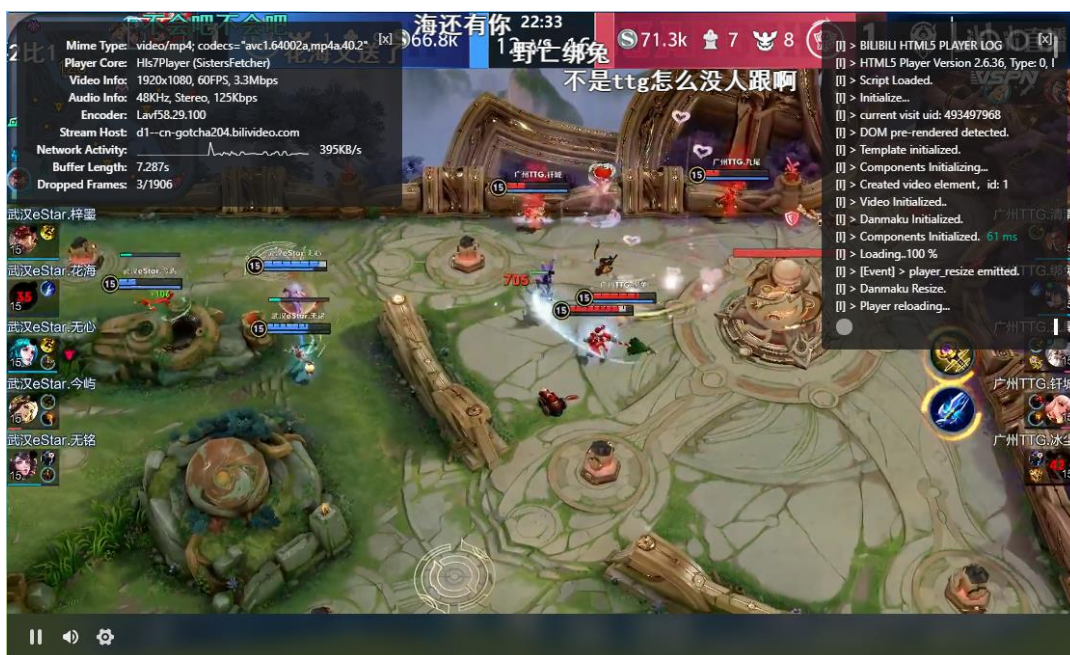
随着 **bilibili** 用户增长、业务增值和平台的发展，**bilibili** 已经成为一个集观看视频、购买商品、社交等诸多功能为一体的巨型平台。结合资料查询、**bilibili** 历次融资情况，我们猜测 **bilibili** 出于业务需求的需要，完成了平台内外的功能集成来增强功能间的交互、并提供出一致且有价值的功能给用户。

2. 细节



bilibili 平台上的功能集成如上图所示，可以划分为以下两部分：

其一，bilibili 平台内部的功能集成涵盖了包括用户个人信息、用户行为信息等诸多功能，结合上文中提到的花火平台，我们猜测 bilibili 的部分用户功能同时会集成到花火平台上，从而满足花火平台 BI 分析所需要的数据等。



其二，bilibili 平台外部的功能集成了广告营销等诸多的功能，从网络渠道我们了解到 bilibili 的播放器是完全自实现的，而在播放的过程中又会出现广告、链接、在线抽奖等部分，结合 PC 端分析，我们认为这部分存在功能集成的工作。

3. 总结

bilibili 平台的功能集成一方面为用户提供了更多的功能、更好的体验，使得用户可以更流畅、更高效、更便捷的使用 APP；另一方面 bilibili 平台方带来了更好的重用性和更多收益的可能，降低了后续研发过程中的成本、提高了从用户身上获取到潜在收益的可能。

（四）业务集成

1. 概述

近年来，bilibili 快速发展，bilibili 的数据平台集成视频、直播、游戏、电商等四大 to C 核心业务和电子交易中心、商品供应链管理服务两大实体、虚拟盈利服务。

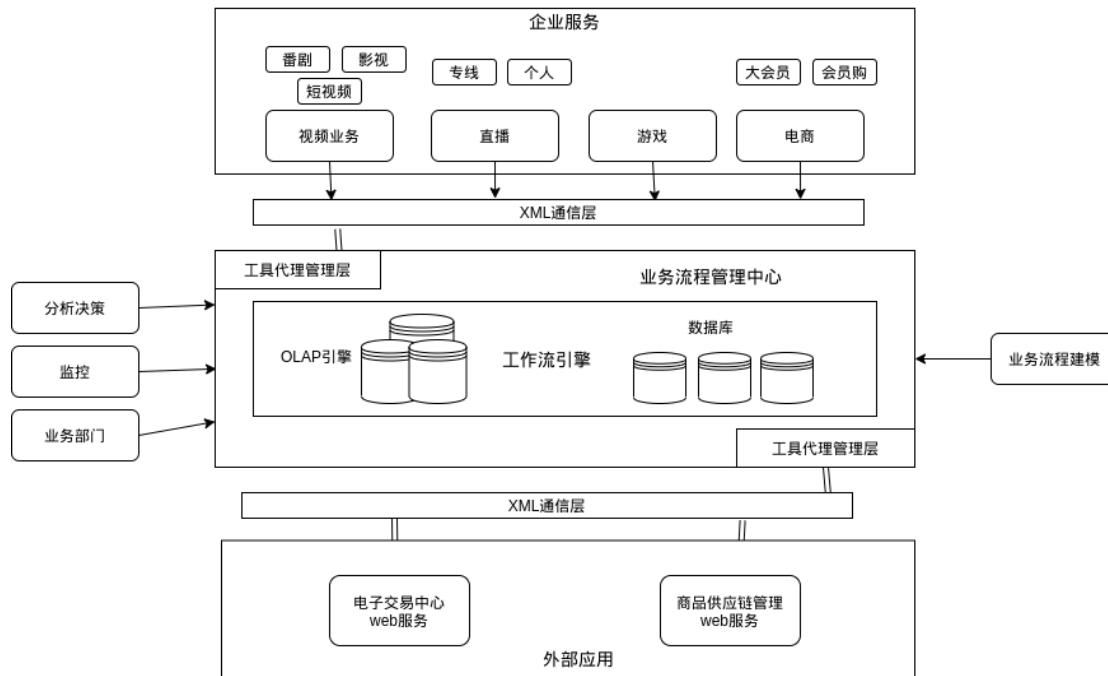
bilibili 对其复杂的各类业务实施了业务集成的工作，也就是为其复杂业务系统提供了平台化的底层支持，使得业务开发能以扩展、隔离的方式推进，驱动业务快速支持。这些业务都具有如下的设计共性：

1. 业务和平台分离：平台不关心具体的业务，通过抽象统一的模型去完成业务逻辑。而平台提供定制扩展机制，方便业务方通过定制扩展开发实现业务需求。

2. 业务隔离：业务方通过平台服务来设计开发自己所需要的业务，降低耦合，从调用者的角度来看，彼此之间相互透明。
3. 变化和实现分离：将业务变化通过规则引擎管理，实现变化实现分离，通过规则引擎快速响应需求而非硬编码实现，提升业务变动需求频繁情况下的业务服务能力和系统稳定性。
4. 职责功能分离：通过微服务架构，减少系统间耦合，使系统简明规范，提升性能和稳定性。

2. 细节

a) 整体业务架构



bilibili 的主要 To C 企业服务有：视频业务（包括番剧、影视和用户创作短视频）、直播（包括官方大型直播和个人直播）、游戏（bilibili 自营游戏和代理游戏）、电商（包括大会员等虚拟产品和会

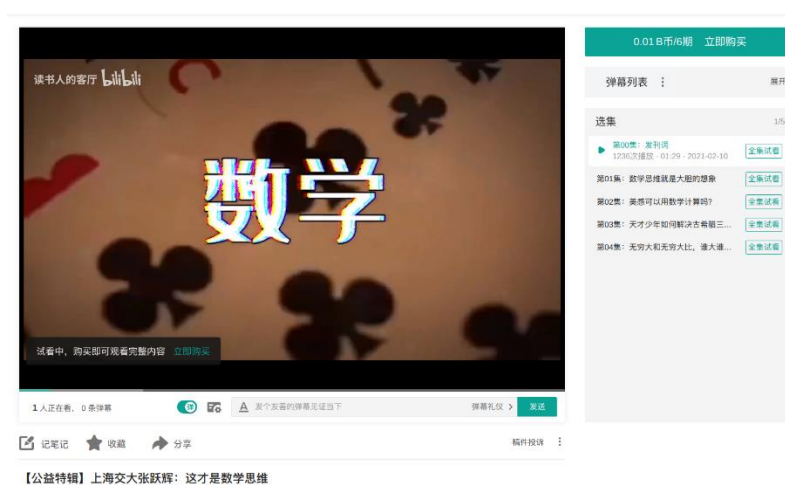
员购例如手办周边等实体业务) 这四个部分。

bilibili 的外部应用有电子交易中心(用于用户购买增值服务, 购买实体产品等), 商品供应链管理服务(管理产品供应)这两个部分。

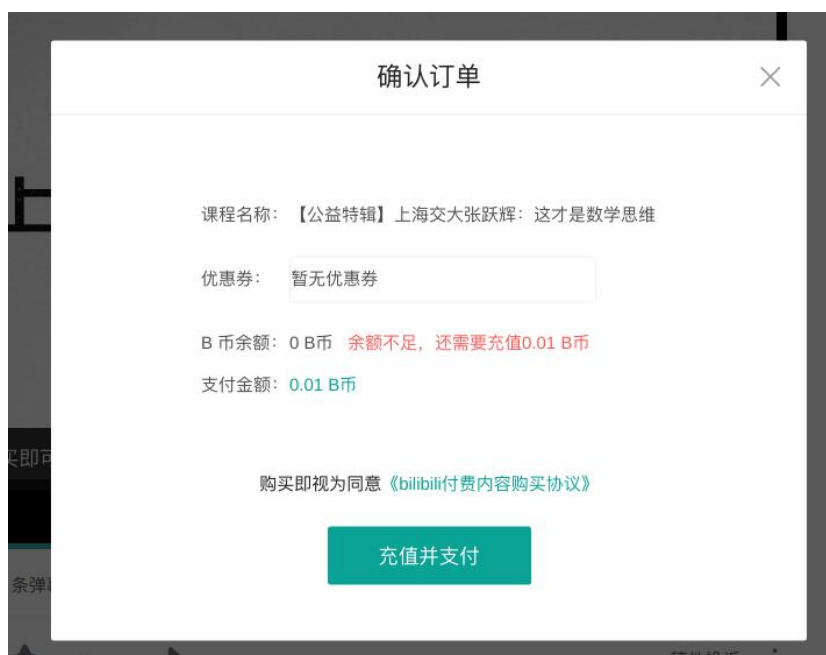
b) 从支付服务看业务集成

例如 B 站的课程在线平台中, 一门叫做《【公益特辑】上海交大张跃辉: 这才是数学思维》课程中, 我们需要花费 0.01B 币来购买这门课。这门课程的播放数据从 Bilibili 的视频接口中获取, 但是具体的购买业务需要从 Bilibili 的电商接口来进行。

如下图是 Bilibili 的课程页面。



当我们点击 **0.01B 币/6期 立即购买** 按钮时, Bilibili 会调用 `sentry-5.6.3.min.js` 文件中的 `sentryWrapped` 接口以调用付费服务的 SDK, 展示可以使用 QQ/微信/支付宝的支付二维码, 如下图所示



点击充值并支付按钮，我们会看到支付二维码，解析二维码便可以得到 bilibili 的付费服务网站如下：

<https://pay.bilibili.com/payplatform-5/qrpays.html?codeType=0&domain=pay.bilibili.com&payToken=1GR0KOIKj19LNZwQz8TPZNGeSJ1TloXS%2BorbARVyM5jm%2BWf1yZ%2FuUQ96aA%3D%3D>

如果使用支付宝扫一扫该二维码，我们便会进入 <https://mclient.alipay.com> 开头的支付宝付款链接，与会员购业务进行比较，我们看到 B 站的购买链接都可以通过调用 B 站的支付 SDK 或是直接访问，以通过支付宝等第三方交易平台进行在线交易。

c) 从视频服务看业务集成

B 站的视频服务统一通过位于 <https://s1.hdslb.com> 的 SDK 来进行资源调度获取，其中视频服务统一集中在/main/video 路径下。而视频资源都是由分布在全国各地的 MCDN 服务(Media Content Distribution Network 媒体内容分发技术)进行资源分发，通过区域负载均衡来降低服务器和贷款资源的无谓消耗，为用户提供稳定的视频观看体验。

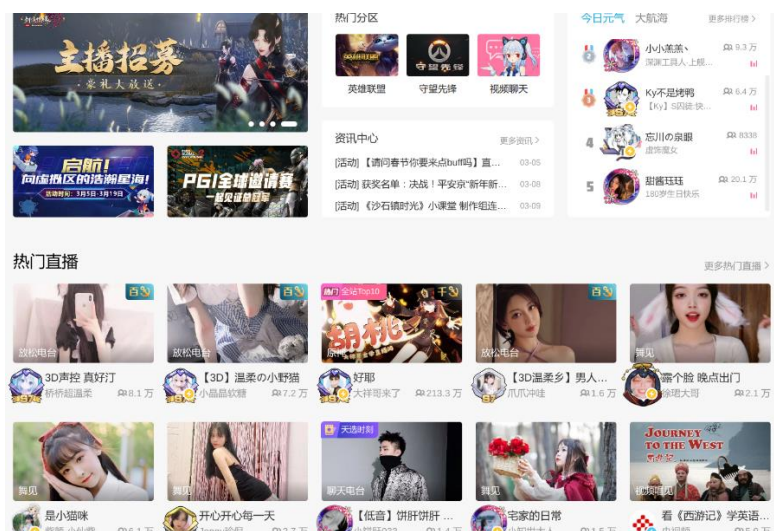
我们通过分析视频的来源，可以看到视频都是由 B 站的 mcdn 进行分发。例如近期的热门番剧《堀与宫村》就是位于 <https://xy120x195x25x95xy.mcdn.bilibili.com:4483/> 的 cdn 进行资源分发的。其中视频资源切分成了约 2M 每片来优化传输。



普通的视频片段，例如首页有 10W 播放的《【司藤·岳绮罗】大姑奶奶 + 小姑奶奶=杀疯了!!! //疯狂踩点·疯批向|景甜·陈瑶|》，是位于 <https://xy183x211x222x80xy.mcdn.bilibili.com:4483/> 的 cdn 进行资源分发。

d) 从直播服务看业务集成

我们打开 B 站的直播服务，可以看到 B 站提供了很多的直播服务。从类型上看，包括网游、手游、单机游戏、娱乐、唱见、电台、Vtuber 等；从用户数量来看，包括英雄联盟赛事、PUBG 赛事等大用户数量的官方直播、一米八的坤儿、梦醒三生狗等头部主播和众多小主播。但总体来看，B 站的直播服务都是通过 UDP 协议向 <https://api.live.bilibili.com/xlive> 的直播源获取直播数据。



调研了其中的一个 PGLS 的 PUBG 比赛直播间，通过分析 ajax 请求，可以看到他们的数据都是向类似于 https://d1--cn-gotcha204.bilivideo.com/live-bvc/137006/live_19738891_5360819_4000/h81126840.m4s?trid=07fac740f36143bc98953dcf5ce7884a 的一个直播源发出 GET 请求，以获取一个 m4s 文件用于直播数据缓存和播放。

此外，类似甜酱钰钰等个人主播的直播数据也是通过 bilivideo.com/live-bvc 的直播源开展直播业务，只是他们提供的直播源和服务流量都会小于官方主办的直播活动。



3. 概述

bilibili 通过业务集成，使用 XML 通信层将企业服务（视频、直播、游戏、电商）和外部应用（电子交易、商品供应链管理）集成到核心构建业务功能集，将数据存储到数据库和 OLAP 引擎（目前是 ClickHouse）中。

bilibili 的工作人员（包括业务人员、IT、决策层等）藉此利用业务流程引擎和资源管理工具集成业务，进行业务流程建模。并通过调度工具、审计管理工具、商务智能工具等进行业务数据监控、分析决策等。由此使得业务贡献，获得对业务的全面透视掌握能力。