北京郵電大學

网络存储技术课程设计

题目：大型云存储中心分析 - 12306 存储架构分析

姓　　名　史文翰

学　　院　计算机学院

专　　业　计算机科学与技术

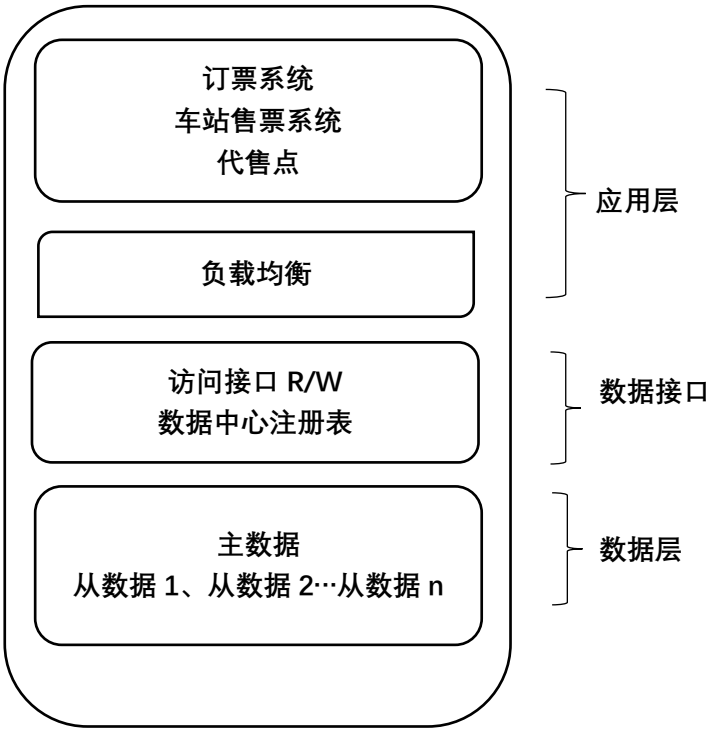
班　　级　2014211304

学　　号　2014211218

2016 年　12 月

12306 存储架构分析

一、 系统总体架构



1、 应用层

包括各种售票和订票系统，如车站售票系统、代售点系统和 12306 网络订票系统。对于车站售票、代售点售票均采用 C/S 架构。如车站售票系统中，每一个车站都是一个客户端节点而服务端节点是铁路中心的票池和数据服务器。C/S 架构认为，客户端和服务端的程序应该不同，即作为数据中心的服务端提供了数据管理、数据共享、数据及系统维护和并发控制等机制，而客户端应着眼于服务性功能如提供售票服务、查询服务、退票服务等。具体到每一个售票中心，任何一个窗口作为一个客户端与数据中心相连接进行数据的交互，人机接口需通过专业的售票员即售票系统来实现。

而对于 12306 售票系统，可以采用 B/S 架构。由于互联网售票具有 B2B 的电子商务特点，即可以由专用网络负责，完成节点和节点之间的信息交互，从而完成交易、信息交换等数据通信。而由此衍生的电子商务模式恰好符合了用户对于火车票订票的需求，即将火车票看作一种电子商务产品，通过 WWW 实现的用户界面来完成对火车票的一系列操作。此时客户端不需要任何专门的应用程序，将大部分的应用程序转移至服务器端。

2、 数据接口

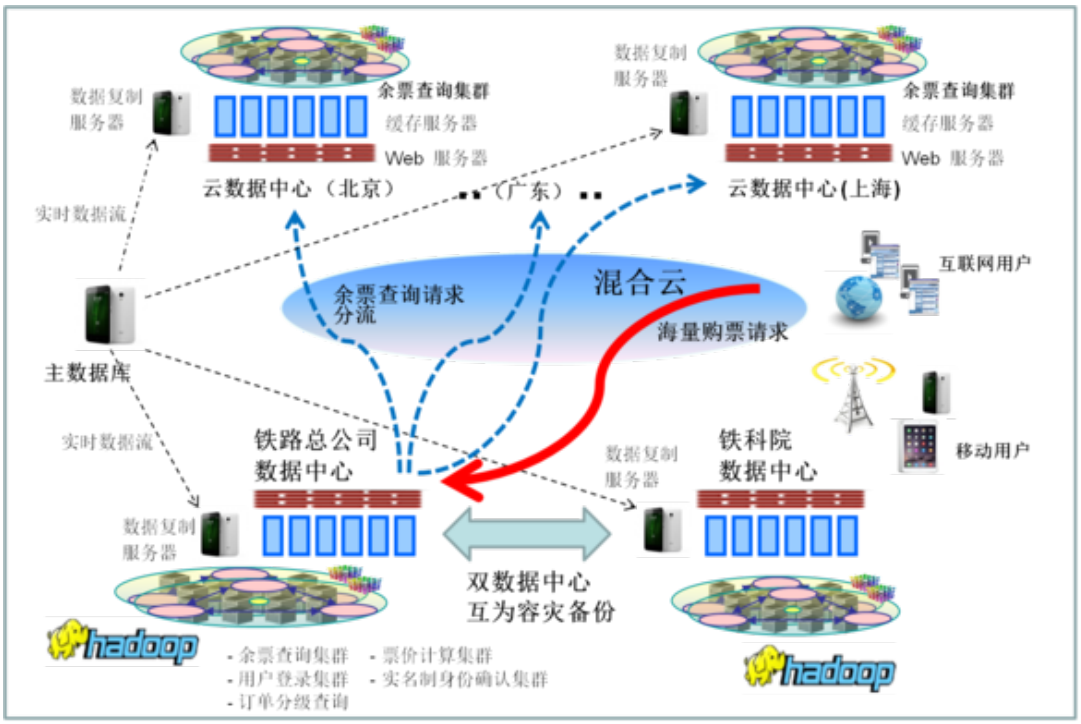
出于对高安全性、便于维护、高并发等特点的考虑，用户不能直接通过前端来访问后端的数据，这也就需要在应用和数据之间增加中间层。由接口服务器提供一系列的操作接口来实现对数据的访问。

3、 数据层

每一个数据中心都应该承载着一定量的、具有高聚合度的数据，如可以按照车次对所有的车票信息进行分组，也可以在此基础上再根据车次类别（Z,T,K 等）进行再分组以减少数据中心的数量。而数据中心的分组是为了分散客户订票、查票时对数据中心的服务器造成的压力。

对于每一个数据中心的数据，12306 采用主从数据的存储模式（分别对应一台主数据服务器和多台从数据服务器），分离订票和查询操作。主数据用于给用户出票，对应于“订票”服务。完成一次对“票”的购买操作后，主数据应该及时与从数据同步。并且，每一次“订票”的操作是不可以并行的，因此对于“订票”这一过程必须增加互斥操作，以避免多个用户同时对同一票务资源进行操作。这对应于操作系统的“读者-写者问题”，“订票”对应于数据“写”，广义上讲是对票池数据库的更改。而查票则是“读”，读操作不具有互斥性，因此可以多用户同时访问从数据。实现订票和查询的分离，可以根据数据流的不同特点进行针对性管理（查询频率远大于订票频率）。对于查询操作只需要访问其中一个从数据服务器即可，在各个从数据服务器中可以使用动态平衡算法实现负载均衡。也就是说，从数据服务器提供了可供查询的数据副本，把大量的操作分散到各个从数据服务器上，提高了查询操作的可靠性和安全性（各个从服务器之间形成冗余，具有一定的容错性），每个从数据服务器之间需要同步。

二、 存储架构



12306 基于云的存储架构可用上图来表述，主要基于混合云来分流余票查询请求，并将海量的购票请求引流到数据中心。铁路总公司的数据中心和铁科院数据中心形成虚拟化双活架构，并接收由主数据库发送而来的实时数据流进行数据复制和数据备份。同时，它还具备两个云数据中心，部署了余票查询服务的缓存服务器和 Web 服务器集群，接收余票查询请求的流量并将查询结果返回至数据中心。

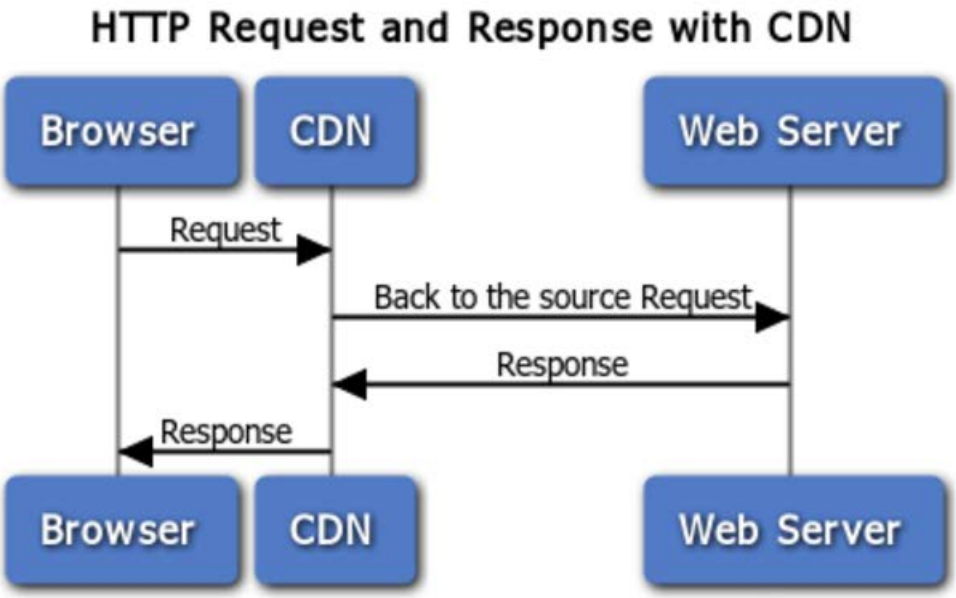
为了便于关键技术的分析，可总结为如下特点：

1) 铁路总公司与铁路院的数据中心形成虚拟双活，1:1 分担流量，且互为灾备备份（见三、5）
2) 余票查询、购票请求是 12306 业务流量的主要来源，核心业务是相互分离的，其中余票查询的服务被分散部署在各地云数据中心（见三、4）
3) 最外围部署 CDN 网络，用户可就近取得所需信息（见三、1）
4) 利用公有云（或混合云），将车票查询服务分流，以缓解处理资源的压力，一系列的核心子系统高度“云化”，按需部署在不同的数据中心，建立可弹性扩展的平台。
5) 采用数据库复制服务器，将主数据库服务器汇总的实时数据复制到各个数据库服务器之间，完成数据中心与数据中心、数据中心与公有云的同步。

三、 关键技术

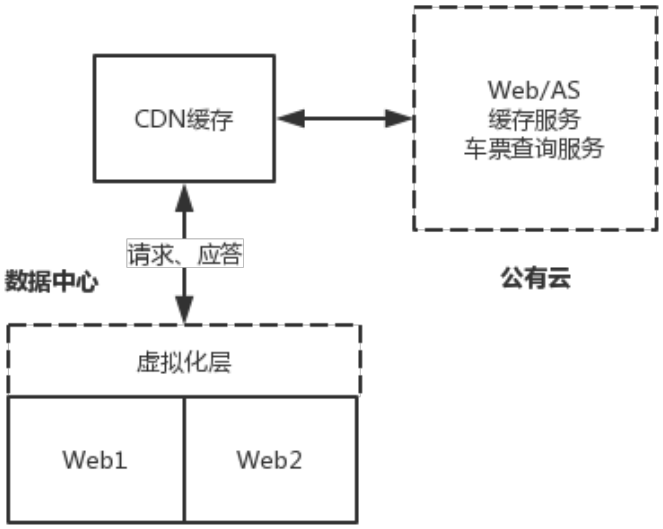
1、CDN 缓存

12306 采用 CDN 缓存技术。



CDN 为浏览器和 Web 服务器之间提供了缓存机制，也就是在二者之间假设了桥梁，使得流量可以通过 CDN 本身的特点和动态算法完成内容分发、流量均衡等任务。在 CDN 缓存下，浏览器不再直接从 Web 服务器请求数据，而是就近从缓存服务器请求，这个特点也打破了集中在数据中心的 Web 服务器的拓扑，即使客户端距离 Web 服务器十分遥远，也可以通过 CDN 缓存服务器来间接通信。

12306 在互联网公用云上部署了车票查询集群和查询服务，有 CDN 完成车票查询流量的分配，这种分配涉及到虚拟化（见二.3）架构。从单个 Web 服务器来看，它们的关系如下：



可见 CDN 为查询缓存提供了便利性，通过不同地点的 CDN 缓存可以大大降低访问延时，大量的请求在 CDN 边缘节点完成，起到了相应的分流作用，减轻了源站的负载。通过负载平衡等技术，CDN 可以将用户请求定位到最近的缓存服务器上获取内容，提供用户访问网站的响应速度，这也为 12306 缓存服务器的部署提供了方便。CDN 通过用户的就近性和服务器的负载进行判断自行选择缓存服务器来请求缓存资源，而不是直接向 Web 服务器请求资源。

2、 NoSQL 数据库

用 NoSQL 数据库取代传统数据库以大幅提升并发查询能力，使得 12306 的 TPS（transaction per second）提升了二十倍。RT（response time）由原来的 1s 缩减至 10ms，使得用户可以快速获取车次和余票情况（加速查询）。

NoSQL 数据库适用的对象和场合 :数据模型比较简单（车票的数据结构并不复杂，只是由一系列的属性值组合成的一个结构，各个属性之间具有无关性）、需要灵活性更强的 IT 系统（显然，12306 的订票和查询系统是其业务能力的瓶颈）、对数据库性能要求很高（对于查询系统，要求能承载相当大的输入流量）。而 NoSQL 在数据一致性方面没有传统的 SQL 那样的即时、准确，然而因此而换得的性能上的提升对于铁路售票系统来说还是十分有利的。

12306 采用以 Key-Value 为主的查询服务，此类数据库主要用于处理大量数据的高访问负载。对于每一个<Key,Value>对通常用哈希表来实现（定位相应的 key 的复杂度只是 O（1）的），其最大优势在于查找速度很快，有助于解决 12306 查询功能的瓶颈。

3、 内存数据库

12306 将用户登录及常用联系人查询等业务迁移至内存数据库中，提高了相关业务的处理性能和可靠性。

内存数据库是将数据放在内存中直接操作的数据库，相对于磁盘来说，内存的 IO 速度要高出几个数量级。就如 Disk 和 RAM 在个人电脑中的关系一样，内存与外存的出现实际上是存储器分级的结果。有些不必要的数据、非核心的业务等可以将其存放至普

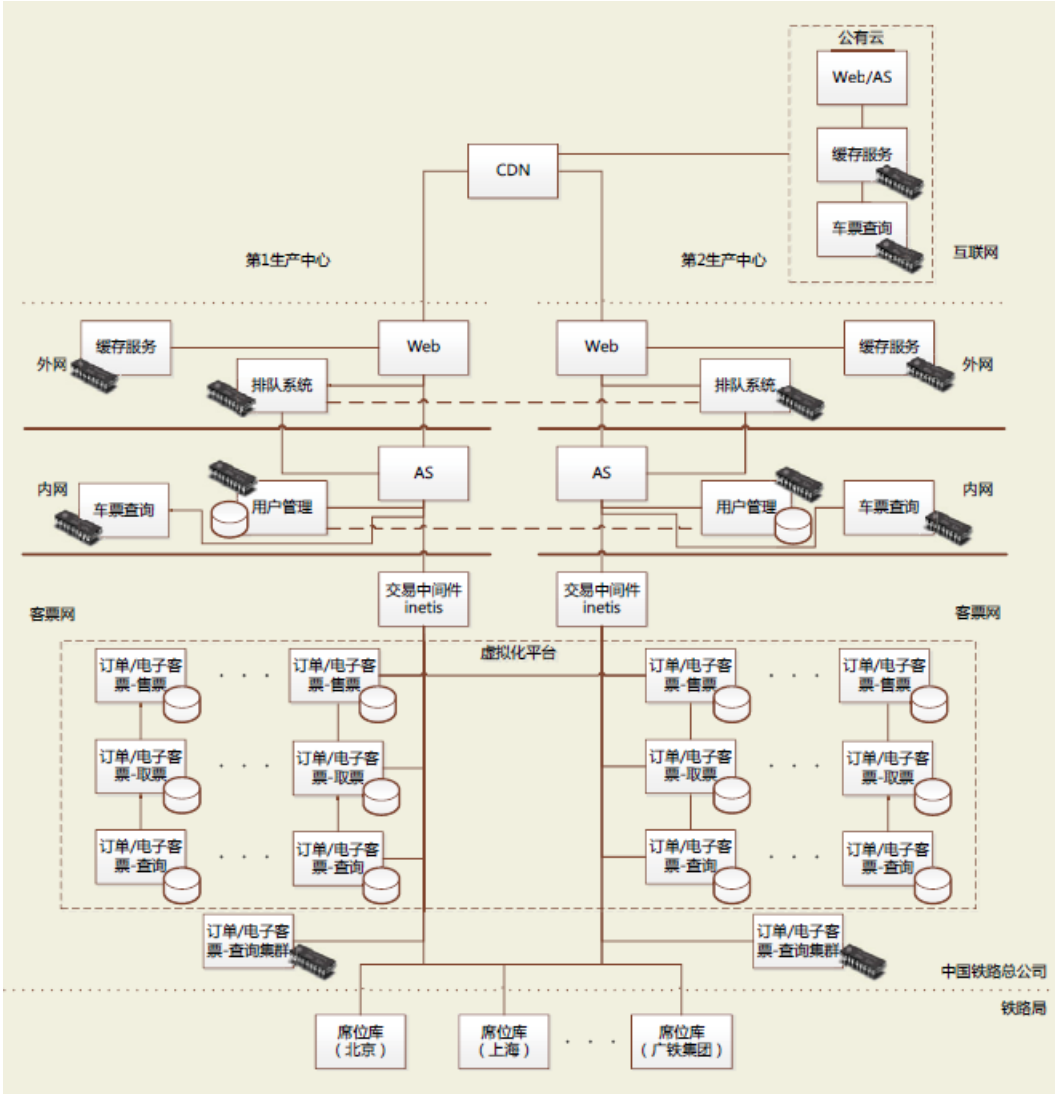
通数据库，而对于核心数据、IO 频率很高的数据而言，将其“主拷贝”或者“工作版本”常驻内存，即活动事物只与实使的内存数据打交道，会增加其 IO 速度。对于 12306 部分的查询业务，采用类似的存储技术是合适的。

4、 公有云服务

12306 将车票查询服务和集群部署在公有云上，通过策略配置可随时将车票查询流量分流至公用云，以缓解在售票高峰期网站的处理资源和带宽的压力。由 CDN 完成车票查询流量在若干个生产中心以及公有云之间的流量分配，流量分配可通过动态调整策略在线完成。

5、 虚拟化存储与双活架构

为增强横向扩展能力，12306 引入了 vShpere 虚拟化平台技术。通过虚拟化技术将两个数据中心（生产中心）的资源整合到一个资源池中，两个中心同时对外提供服务，资源利用率更高。而基于虚拟化的双活数据中心可以帮助用户屏蔽底层细节，实施和管理都比非虚拟化方式更加简便。



对于计算环境的双活，由于虚拟机可以在两个数据中心自由迁移，两个数据中心应该保持相同的计算资源。即每个中心配置的物理服务器台数、数据库虚拟机台数都应该

相同。其中每一个业务群集配备 7 台虚拟应用机，采用 5 台运行 2 台备份的模式。

6、 Gemfire 集群

12306 曾经历过架构改造，即由传统的网络架构向云架构改造，手段是将子系统业务逻辑和数据都放在 Gemfire 集群上执行，利用 MapReduce 技术建立可弹性扩展的平台，提供高性能 CPU 计算能力。在经过这样的部署之后，12306 的每云化的子系统都具有特定的独立性，因为相关数据都放在 Gemfire 内存数据网格节点，这意味着其规模可随业务需求变大或变小，一分为多，以至于到不同的数据中心来协同合作，提高资源的利用率。

四、 适用应用类型

可从 12306 运用的关键技术分析其系统架构所适用的应用类型。

12306 核心的业务是余票查询和订票处理，两者的流量特性都具有高并发、对即时性要求较高的特点。且业务流量具有一定的时间依赖性，如遇到国庆高峰、春运等特定日期，会出现流量突发等情况，需要克服因流量溢出而产生的瓶颈问题。

12306 采用公有云（混合云）的模式，将“余票查询”的核心业务分散到各个公有云或子数据中心之中，分担了核心业务带来的高流量问题。可见此系统对于流量需求大、业务模式固定的情况较为适用。

12306 使用 NoSQL 的数据库承载数据，是由于其需要处理大量数据的高访问负载，必须拥有较快的查询速度。这对于一些实时性（real-time system）较强的系统来说是有效的改进方案之一。

12306 采用 CDN 的缓存机制，克服了因数据中心的地域性而带来的一系列问题，提高了交换速率和交换新能，也提供了流量分配、实时监控等机制。

12306 迁移至 Gemfire 云平台，而利用了 Gemfire 相应的功能特性，如可自定义数据结构、弹性扩展、数据切割、share noting 等。对于这些方面需求较强的应用可以使用 Gemfire 架构进行搭建或改造，如社保系统和金融单位 POC 测试系统等。

对于安全性，12306 采用了双活的数据中心模式，两个数据中心互相分摊流量并互为灾备，这对于灾备需求大、需要高可靠性和安全性的同时还承载大量的即时流量的系统是适用的。

五、 参考资料

1、 朱建生，王明哲，杨立鹏,等. 12306 互联网售票系统的架构优化及演进[J]. 铁路计算机应用, 2015(11):1-4.

2、 杨立鹏，梅巧玲，陈爱华,等. 铁路互联网售票系统的研究与实现[J]. 铁路技术创新, 2012(4):32-34.

3、 12306 上的分布式内存数据技术 GemFire
http://www.csdn.net/article/2013-12-30/2817959-look-at-12306

4、 浅谈 12306 核心模型设计思路和架构设计
http://www.cnblogs.com/netfocus/archive/2016/02/12/5187241.html

5、 透过 12306 五大焦点看高性能高并发系统
http://storage.it168.com/a2012/0217/1313/000001313424_3.shtml

6、 技术揭秘 12306 改造（一）：尖峰日 PV 值 297 亿下可每秒出票 1032 张
http://www.csdn.net/article/2015-02-10/2823900

7、 12306：分布式内存数据技术为查询提速 75 倍
http://www.ctocio.com.cn/cloud/120/12820120.shtml

8、 12306 票池
http://www.cnblogs.com/killmyday/archive/2012/11/12/2765714.html