

**网络存储期末论文**



**题目: 阿里云的存储架构、关键技术和适用应用类型**

**姓 名 李楠轩**

**学 院 计算机学院**

**专 业 计算机科学与技术**

**班 级 2015211308**

**学 号 2015211362**

**班内序号 22**

**指导教师 姚文斌**

**2017 年 12 月**

**阿里云的存储架构、关键技术和适用应用类型**

阿里云作为国内数一数二的云平台，不仅在国内口碑极佳，在国外也有着广阔的市场，目前在阿里云不断创新和在经营者的良好运营的情况下，阿里云逐渐缩短与国际知名厂家AWS以及AZURE的距离。作为学习和研究存储技术的我们，很有必要去了解一下阿里云的架构以及与其相关的一些技术。本文通过分别介绍阿里云的关键技术：OSS,NAS以及盘古来了解国内先进存储技术。

1. 存储架构

OSS云存储：

*阿里云对象存储服务（Object Storage Service，简称 OSS），是阿里云提供的海量、安全、低成本、高可靠的云存储服务。它具有与平台无关的RESTful API接口提供99.99999999%的服务持久性。您可以在任何应用、任何时间、任何地点存储和访问任意类型的数据。OSS适合各种网站、开发企业及开发者使用。您可以使用阿里云提供的API/SDK接口或者OSS迁移工具轻松地将海量数据移入或移出阿里云OSS。数据存储到阿里云OSS以后，您可以选择标准类型（Standard）的阿里云OSS服务作为移动应用、大型网站、图片分享或热点音视频的主要存储方式，也可以选择成本更低、存储期限更长的低频访问类型（Infrequent Access）和归档类型（Archive）的阿里云OSS服务作为不经常访问数据的备份和归档。*

*----摘自《阿里云文档》*

OSS云存储架构:

OSS分为客户端和阿里云开放云存储系统，客户端是用户直接操作。

阿里云开放云存储系统：

首先是由一个监控系统，这个监控系统课以监督整个阿里云存储系统。接着是负载均衡器，作用是把网络请求分散到一个服务器集群中的可用服务器上去。接着则是HTTP服务器，功能是提供Web信息浏览服务。它只需支持HTTP协议、HTML文档格式及URL。接下来是分布式机构化数据存储系统和分布式文件系统，这两个负责oss的存储。

对象存储是将数据保存到一个对象当中，并且这个对象不能是别的对象的子对象，这样的做法使得都想存储不再具有层次结构，如果想要查询对象则不知道数据的物理地址，这种方法对于云计算环境自动化和简化数据存储起到帮助

分布式机构化数据存储系统：

这个分布式机构化数据存储系统的作用是保存结构化的数据，包括具有固定字段格式和类型定义并可用二维表结构逻辑表达的规整数据，这些数据的特点是价值密度高，使用频繁，对于阿里云的数据存储系统，有必要将频繁的结构化数据和非机构话数据相分离。

nosql的非结构化数据存储，由于数据不是一个标准的格式，所以数据保存不能按照结构化数据进行存储，这是使用nosql可以对非机构话数据的存储，在这里中常用的非机构话存储有Hbase。对于对象存储由于兼容nosql，也可以基于Hbase，所以对象存储可以很好的保存非结构化数据。

二、盘古-飞天同意底层存储平台

盘古关键技术：



阿里云的盘古技术是目前阿里云其他云技术的基础，主要负责的是存储，盘古分为Client,Mater，ChunkServer。需要发起一次写入的时候，Client向Master创建一个文件，并且打开这个文件，此时Master会选好三个副本的位置反馈给Client。Client根据三个副本的位置找到ChunkServer，把数据写进去。也就是说，Client做整体的控制，Master提供源数据的存储，ChunkServer提供数据的存储。

盘古的主要优点：

1. 数据高可靠。

由于三个副本位于不同的位置，所以再算法RACK-3的基础上会保证一致性和数据的可靠性。

1. 数据完整性好。

数据校验采用CRC，采用端到端的数据校验，静默错误检查，这样的好处是对于数据的完整性，可以极大的确保安全。

主要结构的信息：

Master：这个需要解决三类问题：容量，效率和稳定性。在盘古系统中，为了防止单节点的问题，阿里云采用了自研发的Paxos来保护数据的一致性，使Master防止单点，以及多用户的打死问题。

## Chunkserver：采用面临的问题是IO与存储器件价格之间的问题。对于快速IO的器件很贵，对于慢速的便宜，但是导致速度下降。解决方案是使用少量的缓存搭配大量的机械硬盘，数据前台先写入缓存，后台将其转储到机械式硬盘。

Client：需要解决用户的极致性能，多核系统上线程较多时，切换代价非常高，高性能的程序无法容忍这一点，所以在Client的设计中采用了协成技术不过对于协成来说，一是很多的编程语言没有做到协成，而是协成具有数据不安全的风险，所以在阿里云的盘古系统中，主要采取两种不同的方式，一种支持多线程，另一种不支持，这样就可以把不开多线程和开了多线程分开。

1. 阿里云的适用应用类型：

**在线考试系统**

首先适合在阿里云上构建。

1. 对于在线考试系统，需要存储结构化和非结构化的数据，在阿里云中可以使用多结构数据存储，即采用OSS和RDS来混和使用，其中RDS可以存储数据的有关填空题的方面和选择题方面，在OSS中则可以保存问答题。
2. 对于试题的分发，可以采用阿里云所带的内容分发网络CDN
3. 对于试卷的分发安全问题可以采用HTTPS加密，海量DDoS清洗。
4. 对于试卷存储的安全问题采用加密服务，然后使用云防火墙来保护、
5. 对于数据上传可以使用专有网络 VPC。对数据进行上传。

阿里云的平台拥有完善的数据保护，存储和分发功能，适合构建在线的考试系统。

**在线购物系统**

适合在阿里云上构建。

1. 大规模的数据访问，阿里云CDN支撑日常的数亿用户对网页、图片等的访问，保证了极大访问量时的系统访问不掉线。
2. 对于大规模的数据保存，采用CDN+SLB +ECS +OSS + 云数据库memcache +RDS。

可以有效的防止数据库瓶颈。

1. 采用云服务器ecs，保证数据的处理速度，降低 IT 成本，提升运维效率。
2. 对于数据信息安全性，采用Web攻击防护，保证网站不被黑客攻击。
3. 对于网站安全性，采用DDoS攻击，防止网站被DDoS攻击，页面打不开，防止攻击最高峰值上百G，业务面临全面中断的风险。