

# 基于云平台的气象数据共享服务平台架构建设

■ 陈京华 何小明 张志强

气象数据共享服务平台正在从传统自建服务系统向更具弹性的云平台迁移。结合中国气象数据网（阿里云版）的建设，探索了基于云平台的气象共享服务平台架构，并针对该架构提出了文件向对象存储迁移方案、对象存储省流量方案和基于阿里云的中国气象数据网v2.0数据传输策略。

2015年9月中国气象数据网（简称“数据网”）v1.0上线对外服务，取代了中国气象科学共享服务网（CDC）。中国气象数据网是一个完整的气象数据服务体系，包括来源于中国及全球的、高质量的观探测气象资料及数据加工产品（含卫星遥感反演产品），以实况数据与历史气象数据集为主，并提供数据统计分析与可视化服务。同时，网站提供界面优化、方便快捷的数据服务方式，用户可以快速定位、检索、下载气象数据。中国气象数据网涵盖了数据目录与元数据服务分系统、数据检索与服务分系统、数据分析与可视化分系统、服务产品加工与管理分系统、数据服务移动应用分系统等八个分系统。在对外服务情况下，截止到2016年3月底，中国气象数据网累计注册用户数113124人，其中从CDC迁入用户101502人，数据网新注册用户11622人，服务订单数192271个。在数据存储方面，截止到2016年3月，数据网的数据库存储总量约47.5 GB，平均每日新增约301 MB，文件存储总量约32 TB，每日新增约23.1 GB（图1）。



图1 2015年中国气象数据网v1.0的订单量和文件下载量

中国气象数据网v1.0版本存在的问题是骨干网与Internet服务区之间存在多处系统耦合，这其中包括气象数据推送流程、元数据更新流程、临时工作区的加工处理以及订单处理流程。数据流跨越骨干网和Internet服务区，因为系统的耦合性，不利于数据网的升级。最重要的是因为用户量的快速增长，及气象数

据服务量的逐日递增，未来中国气象数据网将面临存储和计算资源问题，以及用户高并发、定制化等特殊访问需求，这将会给Internet服务区带来巨大压力。因此有必要充分利用公有云相关技术及网络服务资源，研究开发中国气象数据网v2.0（公有云版），建设基于公有云的权威和统一的气象数据共享服务平台。

## 1 中国气象数据网特点

### 1.1 资料种类丰富

中国气象数据网对社会公众服务的气象数据包括2015年共享清单资料和中国气象科学共享服务网存储的资料，这其中涵盖了地面、高空、海洋、辐射、农气等14大类气象行业标准所列资料。在数据集产品方面包括了地面气象观测站2170个、高空观测站88个的基本气象要素实时观测资料和气候标准值数据集、风云气象卫星的实时云图产品、定量产品和历史数据产品，天气雷达实时图像产品，全球数值天气预报（T639）产品、区域数值天气预报产品等。在数据存储方式上又分为要素型资料 and 文件型资料，文件型资料包括雷达图像产品（每日新增14 GB）、雷达基数据（每日新增15 GB）、T639产品（每日新增4.3 GB）、Grapes-MESO产品（每日新增2.5 GB）、CDC资料文件（每日新增2.1 GB）等，要素型资料包括服务数据（配置信息、用户信息等，每日新增3 MB）、CDC数据（含文件索引信息，每日新增42 MB）、CIMISS数据（含文件索引信息，每日新增256 MB）。

### 1.2 稳定与可靠

中国气象数据网是一个重要的业务系统，在软硬件的设计实现上考虑系统长期运行的稳定性和可靠性。软件在运行期间，针对任何一个重要操作，都具有判断错误的的能力，必要时可以进行恢复性操作。在支撑平台可靠性的保证下，关键高可用业务系统在3

分钟内恢复运行状态。中国气象数据网除了在系统上考虑安全性外（如内外网隔离、防火墙等），同时提供了基于角色（例如：系统管理员、操作员、高级用户、一般用户等）的安全控制策略，从访问级别上严格控制不同用户的权限，避免用户越权使用或非法使用系统资源。

## 2 公有云的特点——以阿里云为例

### 2.1 数据安全，稳定可靠

阿里云服务能够保证数据高可靠性，在不发生海啸、地震等特大自然灾害的情况下，阿里云服务器基本实现99.9999%的数据安全。在硬件方面，阿里云服务器在物理架构设计上，采用了分布式高可靠存储系统，数据在不同的硬盘上按照随机原理实现多重备份，即便其中一份数据因为某种原因遭到破坏，按照预定的数据修复算法，阿里云服务器可以实现在最短时间基于备份的数据将被破坏的部分实现修复。利用快照快速回滚技术，即便系统遭到计算机病毒的入侵而导致破坏，数据依然可以快速还原。阿里云服务的高可靠性还体现在对服务器硬件自动检测，一旦故障发生，阿里云服务器可以实现数据的自动迁移，并保证迁移后磁盘数据、MAC地址、IP地址等信息不发生变化。在软件方面，阿里云服务采用安全隔离和自定义防火墙技术，能够避免地址解析协议攻击和网络地址欺骗，同时云盾还提供木马扫描、端口入侵扫描和系统漏洞扫描，网络安全保障业务的稳定运行。在法律层面，阿里云完全遵守《中华人民共和国网络安全法》，2016年10月14日，阿里云宣布完成公安部组织的等级保护标准和云计算等级保护新标准试点工作，成为全国首家通过国家级权威测评的云计算服务商，其中公有云平台通过等级保护三级备案。

### 2.2 独享带宽，高速访问

阿里云服务提供高性价比的千兆光纤接入，阿里云服务系统内部数据流量不存在计费问题，宽带扩容按天计算，传统带宽无法与其相比。同时阿里云服务对于中国移动、中国联通、中国电信以及教育网等主干线路进行高速连接，实现了南北网络互通互联，保障国内访问顺畅。阿里云在国外共有八个节点，分别部署在新加坡、悉尼、弗吉尼亚、法兰克福和迪拜等地，充分保障国外用户的访问效率。阿里云服务保证带宽资源绝对独享，消除传统带宽资源在不同应用系统间抢夺的问题，彻底消除带宽带来的系统访问瓶颈。

### 2.3 简单易用、弹性伸缩

阿里云服务器的创建在几分钟内便可完成，而且拥有超级管理权限，对于资源的数量和配置完全按需

分配，无论增加或者减少资源，均在几分钟内就能实现。对于一次开启服务器的数量没有限制，方便实现互联网应用的部署。购买资源、系统配置变更、续费等通过Web方式实时完成。阿里云服务器相比于传统硬件采购，成本降低30%~80%，阿里云服务支持任意时刻升级CPU、新增数据盘和系统内存。

因为阿里云所具有的数据可靠、高速访问、弹性伸缩、节约成本等特点，可以有效的解决中国气象数据网面临的存储和计算资源问题，所以中国气象数据网v2.0采用阿里云进行部署，通过Web或应用程序接口（API）方式中国气象数据网v2.0实现便捷的自助管理，而且根据业务的具体需要动态的增减虚拟硬件的数量、调整系统配置。

## 3 基于阿里云的中国气象数据网v2.0物理架构

### 3.1 基于云的总体架构

基于阿里云的中国气象数据网v2.0服务应用全部部署于阿里云上，包括前端数据网Web门户、移动应用与API应用，以及后端订单处理及数据下载服务等。网站应用主要系统软件采用阿里云服务替代（表1）。

表1 软件替代汇总

	DMZ版本	阿里云版本
(分布式)数据库	Oracle/MySQL 虚谷数据库	云数据库 (RDS) 分布式云数据库 (DRDS)
文件存储	文件系统	云存储 (OSS)
缓存服务	Memcache	云缓存 (OCS)
虚拟主机	虚拟主机 (Linux)	云主机 (Linux)
负载均衡	Nginx	云负载均衡 (SLB)
下载服务	FTP (文件系统)	HTTP (Web代理云存储)

具体软件功能如下：

服务器（ECS）：分别用于Web服务器、可视化服务器、应用服务器、订单处理服务器。

缓存服务（OCS）：用于缓冲Web数据缓存。

数据库服务（RDS）：用于部署数据库结构与内容，存储气象要素内容、气象文件元数据和相关配置与服务信息。

存储服务（OSS）：用于存储气象文件，同时提供影像的对外Web访问服务。

根据服务需要，数据网使用云服务器（ECS）14台、一个分布式云数据库（DRDS，6个节点）、负载均衡服务（SLB）公网带宽150 MB、开放缓存服务（OCS）30 GB、开放存储服务（OSS）50 TB。

在国家气象信息中心的骨干网，部署两台数据和元数据同步服务器，负责将气象数据定时推送到信息中心的Internet服务区，两台服务器的关系为一台在线

运行，一台进行热备。在信息中心的Internet服务区同样部署两台数据和元数据同步服务器，部署方式一主一备，负责将骨干网推送过来的气象数据定时推送到阿里云。与此同时阿里云中两台（ECS）服务器，部署方式一主一备，负责接收并处理从信息中心Internet服务区推送过来的气象数据，将数据分类别存储到阿里云的一台存储服务器和一个（6个节点的DRDS）云数据库，存储服务器和数据库服务器共同构成了气象共享服务平台的底层数据环境。两台（LBS）负载均衡服务器，部署方式一主一备，负责接收外部用户的前端请求。在负载均衡服务器的后端，分别是一台GIS图像处理应用服务器，负责服务门户Web部件组

的图像展示与WebGIS服务，一台Web应用服务器，负责网站门户的运行，一台对外接口应用服务器，负责支持数据网用户的接口调用，以及一台下载服务器。当负载均衡服务器接收到用户的前端请求后，按照请求的类别，请求将被转发到上述四台应用服务器。数据网在阿里云还部署了一台配置与监视服务器，主要负责监测负载均衡服务器、GIS图像处理应用服务器、Web应用服务器和对外接口应用服务器的运行状况，并进行必要的日志记录。另外在阿里云中数据网还部署了订单处理服务器一台，数据检索服务器一台，以及数据缓存服务器一台。系统物理架构见图2。

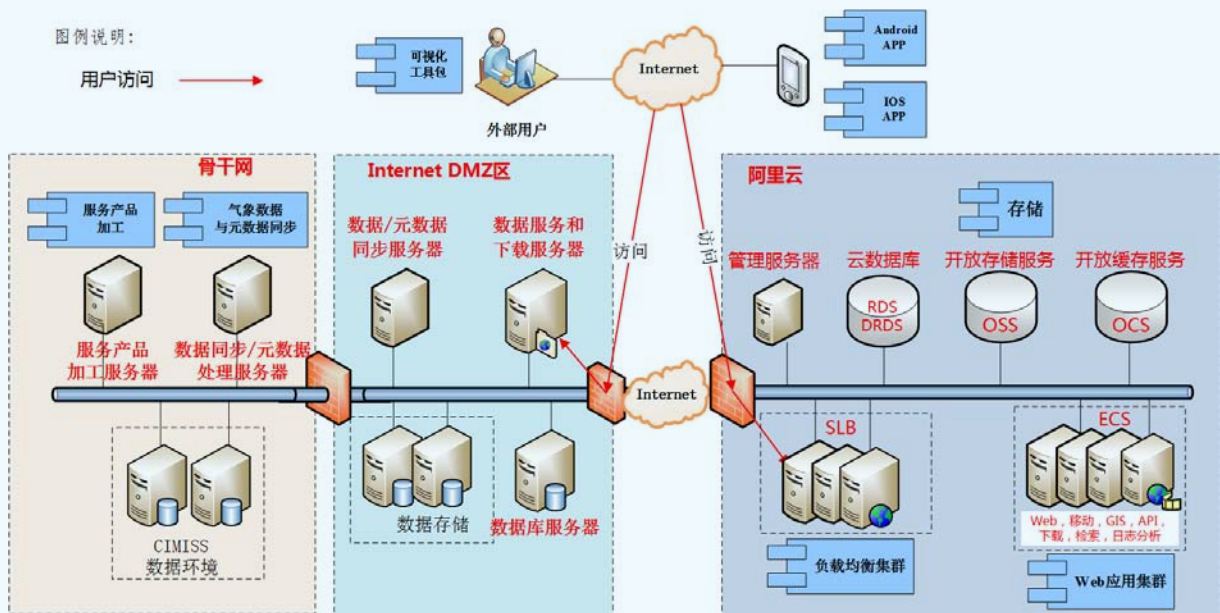


图2 系统物理架构部署图

方案优点是利用阿里云服务资源进行气象数据的存储和服务，可以省去相关的硬件采购，简化服务运维、保证高可靠。同时，数据网应用可以根据服务需要，面向服务优化存储设计与存储管理，重新规划建设适用的应用以此来支持阿里云数据环境。

该方案直接面临的两个问题，一是目前阿里云面向对象的数据存储服务与现有业务的文件共享存储模式之间存在较大差异，二是阿里云计费问题，存储服务按外出流量计费，直接利用存储服务进行数据服务费用较高。

### 3.2 文件向对象存储迁移方案

阿里云与现有业务的文件共享存储不同之处首先体现在文件夹的概念上，现有业务的文件夹是一个实体，文件存储在文件夹里面，方便进行文件夹下信息的汇总。而OSS中文件夹的概念仅是一个逻辑概念，

在阿里云上创建文件夹有多种方式，其中最方便的就是通过新建文件夹按钮来创建文件夹和通过API/SDK的方式设置文件夹，其底层实现的逻辑方式是一样的。比如在OSS存储中创建一个“surf”文件夹，在这个文件夹下存储一个“V0001.txt”的文件，系统会在后台定义一个“对象”，其Key值就是surf/V0001.txt，但“surf”文件夹只是在逻辑概念上存在，其实质是一个大小为0KB的空文件。因为OSS采用的是分布式存储方式，对象并不是根据文件夹的物理位置进行存储的，虽然给用户的感觉是在“surf”文件夹存储了N个文件，但这N个文件并不一定存储在同一个物理位置上，在后端存储的过程中不同文件夹下的文件仅仅是key值的前缀不一样。而想要遍历某个文件夹下的所有的文件也需要首先通过ListObject接口获取文件夹下的所有文件的key值（这里需要通过prefix指

定文件夹)，然后再进行操作。由于OSS的数据存储对象化、服务化，不支持标准文件协议，不支持文件目录，其虚拟目录功能对目录名与文件名也有着严格的关联限制，数据网现有业务数据存储管理模式已不适用，必须基于阿里云面向服务重新设计建立符合服务要求的文件存储规则来支持数据环境。在阿里云内部开发数据文件转换模块，将原有的数据文件在OSS中进行地址定义并存储。对于中国气象数据网v1.0中涉及的系统内部文件访问模块，对其进行文件访问方式的改造，将原来的绝对物理地址，修改为内部OSS服务地址。

### 3.3 对象存储省流量方案

阿里云OSS主要计费项包括按存储容量时间计费 and 按外出流量计费。存储容量时间计费标准大约1万元/5T·年，外出流量计费标准为0.05万元/T。2014年中国气象科学共享服务网存储数据量不超过10T，服务量（含离线）合计为206 T，服务量约为存储量的20倍。存储容量计费比较简单，不含卫星数据，2016年中国气象数据网v1.0数据量拥有量约为33 TB，OSS存储收费7万元/年。按中国气象科学共享服务网服务量与存储量的比例计算，数据网年外出服务流量约为660 TB，OSS流量收费33万元/年。两项合计，仅OSS服务收费可达40万元/年。若考虑费用问题，一个可行的解决方式是不直接使用OSS存储提供对外服务，而是对数据网应用软件根据OSS存储进行改造，将OSS数据服务转向阿里云应用内部（OSS数据服务在阿里云内部环境免费），再通过数据网的固定带宽

（20 MB）应用提供对外服务，尽量减少或避免OSS外出流量，数据网应用软件主要服务和管理功能模块的设计需要在此基础上进行调整并进行代码开发。

## 4 基于阿里云的中国气象数据网v2.0数据传输策略

利用阿里云服务资源，考虑到网络安全问题，中国气象数据网v2.0数据传输应用涉及三个网络，分别是国家气象信息中心内部骨干网、国家气象信息中心Internet服务区（DMZ区）和阿里云。数据网提供服务的气象数据全部存储于阿里云。

从国家气象信息中心骨干网气象数据环境中同步数据和文件，对于要素型资料，对需要同步更新的数据存储成数据文件，然后将数据文件从骨干网传输到DMZ区，然后从DMZ区通过FTP推送到阿里云ECS中，在ECS通过程序解析数据文件中的数据，将数据同步更新到阿里云数据库中；对于文件型资料，需要将相应的资料文件从骨干网推送到DMZ区，然后通过FTP推送到阿里云ECS，通过调用API的方式，将资料文件从ECS推送OSS进行存储。

在数据传输策略（图3）中充分考虑了数据同步的稳定性和数据传输的时效性。针对数据同步的稳定性，不论在国家气象信息中心内部骨干网，还是在国家气象信息中心的DMZ区，都设置了相应的数据传输监控管理，一旦有数据同步失败情况发生，系统会自动触发补录机制，如果补录连续失败三次，监控系统会发送消息给系统管理员，管理员进行问题排查，然后进行人工补录。针对数据传输

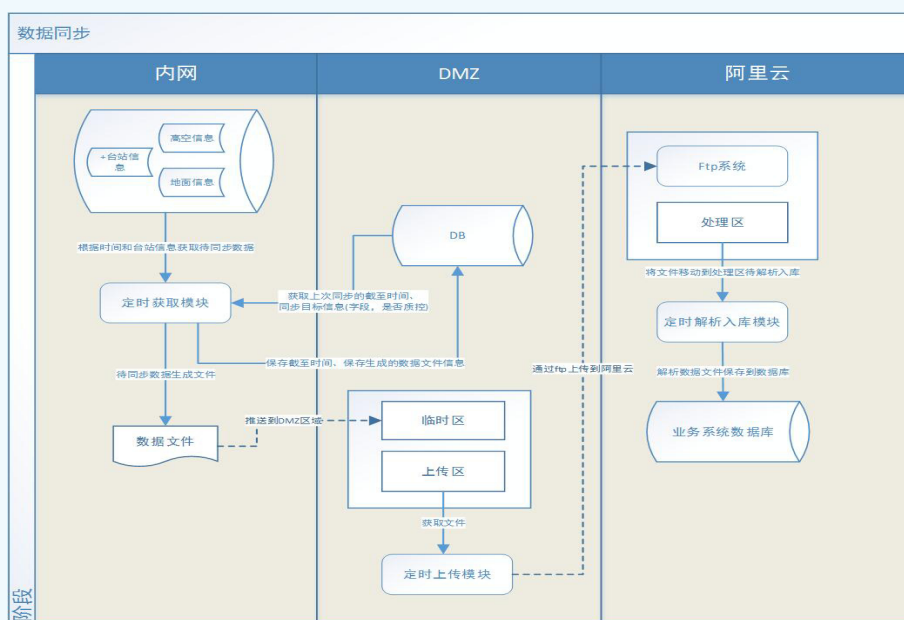


图3 数据传输策略

的时效性，数据网的数据更新频次包括逐小时、逐日、延迟三个月等，我们采用多进程和多线程等手段对资料分类进行数据传输，其中数据传输过程中时效性要求最高的是“中国地面气象站逐小时观测资料”，更新频次是逐小时，针对该资料在数据传输过程中时效性完全能够满足。

同步要求：共享目标数据表中存在创建时间和更新时间（对于有更新的数据），程序以时间为依据查找待更新数据。

## 5 效率和效益分析

2016年3月8日，中国气象数据网v2.0 (<http://data.cma.cn>) 正式迁址阿里云，通过利用阿里云的计算、存储与网络服务资源，进一步提高了中国气象数据网的高可靠、高并发服务能力，2016年新增用户数42929人次，新增人数比2015年增长173%（图4）。



图4 中国气象数据网用户统计

在性能对比上，可以看出阿里云版的中国气象数据网服务能力有了显著提升（表2）。

表2 性能对比分析

测试用例	(阿里云) 局域网内部IP		(阿里云) 互联网域名		DMZ版本	
	请求/s	响应时间	请求/s	响应时间	请求/s	响应时间
数据网首页显示	638	0.36 s	80.2	2.03 s	211	0.46 s
资料详细页面显示	570	0.55 s	98	1.98 s	180.6	0.67 s
中国地面逐小时检索	690.2	0.32 s	327	0.57 s	259	0.48 s
中国地面累年日值检索	1063	0.06 s	760	0.18 s	503	0.2 s
全球地面定时值检索	792.5	0.35 s	263	0.62 s	65.2	1.5 s
中国高空定时值检索	125.4	1.02 s	84	1.3 s	70.8	1.41 s
基本反射率打包检索	1768	0.02 s	1004	0.22 s	1739	0.02 s
T639模式产品检索	430.2	0.25 s	270.7	0.41 s	91.1	1.1 s
中国地面逐小时混合检索	820.2	0.14 s	359.6	0.23 s	230	0.43 s
组合检索测试	650.1	0.17 s	462	0.26 s	329	0.31 s

中国气象数据网v2.0（阿里云版）应用建设与上线服务，实现了基础气象数据服务业务与阿里云平台的有机结合。

由于公有云与生俱来的稳定可靠、高效访问和优越性价比等特点，未来的各类气象数据服务都将面临向公有云迁移的问题，本文对其他数据服务系统向云平台的迁移具有借鉴意义。

### 深入阅读

- 洪伟, 任芬, 2016. 云服务器技术的发展与安全远程访问研究. 网络安全技术与应用, (9): 70-70.
- 李国杰, 2015. 对大数据的再认识. 大数据, 1(1): 1-9.
- 林楠, 魏涛, 2015. 基于云服务的统计测试非监督故障诊断预测. 计算机测量与控制, (5): 1457-1459.
- 武小年, 邓梦琴, 张明玲, 等, 2013. 云计算中基于优先级和费用约束的任务调度算法. 计算机应用, 33(8): 2147-2150.

（作者单位：国家气象信息中心）

## 气象信息化标准体系框架下的大数据标准体系

气象大数据标准体系由七个类别标准组成，分别为：总体标准、大数据信息平台、数据分类与格式标准、数据元及元数据标准、大数据文件命名标准、安全管理、评估评价标准（图中蓝色阴影部分为重点制定的标准）。

——摘自《气象大数据行动计划（2017-2020年）》，2017年9月

