

# 一种高效高可靠的数据中心网络架构设计

田征 何恒宏 周琰

在对目前主流数据中心网络架构设计方法进行对比分析的基础上，通过对气象业务的数据交互特征进行分析和归类，提出一种少跳转、低延时、高可靠的适用于气象业务系统的数据中心网络架构设计。该网络架构将业务和管理分别组网，并采用异构的网络结构，在满足业务高速高效交互需求的基础上，兼顾数据中心网络的安全性和可靠性。

数据中心以集约共享的基础设施资源池作为资源载体和补充，承载各类气象应用端系统等业务网区气象业务和科研应用系统，提供统一的气象数据环境，以服务接口的方式支撑气象应用，实现天气、气候、探测等加工处理算法的整合，融入一体化加工流水线。

网络系统作为数据中心的基础设施，为各类气象业务系统提供稳定、高效、可靠的网络传输平台，数据中心的网络架构关系到数据中心的配置、性能、管理维护等多个方面，数据中心网络系统的结构和性能至关重要。

随着几年来云计算技术的逐渐成熟和推广应用，数据中心内部一般采用分布式架构处理海量数据存储、挖掘、查询等相关业务，服务器之间需要进行大量的协同工作，因此产生了大量的东西向流量。其次，数据中心中越来越多的采用虚拟化技术，物理服务器的吞吐量是虚拟化前的数倍之多。同时，为了更大幅度地增大数据中心业务的可靠性、提高业务部署的灵活性，虚拟机需要实现在数据中心内进行动态迁移。作为核心基础设施，数据中心网络是连接数据中心大规模服务器进行大型分布式计算的桥梁，适应云计算和大数据应用的数据中心网络已经成为了近年来引人瞩目的研究热点。国际国内学术界、标准组织、网络设备厂商、云计算提供商等都对数据中心网络研究给予了非常大的关注，PortLand、VL2等多种新型数据中心网络结构被提出广泛部署和应用用于当前的商用数据中心。

本文在对目前数据中心分析研究的基础上，结合目前主流数据中心网络架构设计方法，提出一种满足少跳转、低延时、高可靠的数据中心网络架构设计。

## 1 网络架构分析

### 1.1 数据中心网络架构

传统数据中心网络架构一般采用以交换机为核心的三层树形网络架构，即采用核心层、汇聚层、接入层三层结构，如图1所示。这种传统树形网络结构的根节点带宽会成为瓶颈，存在以下缺点。

- 带宽受限，在汇聚层，其连接边缘层的带宽远远大于连接核心层的带宽；
- 网络时延开销大，部署在不通根节点交换机的业务系统进行数据交换，需要经过多级多台交换机的转发，而普通根节点接入交换机的转发性能远低于核心交换机；
- 流量工程难度大，数据中心流量是高动态和突发的，约80%的流量都是内部流量，且流量持续变化，难以预测，从而使得传统流量工程方法无法有效工作。

近年来，随着数据中心规模的不断膨胀，以及虚拟化、云计算等新技术的不断发展，仅仅使用传统的

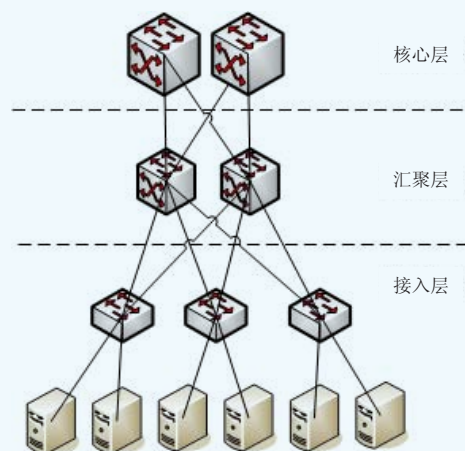


图1 传统数据中心网络架构

网络技术越来越无法适应业务发展的需要。传统的数据中心内，服务器主要用于对外提供业务访问，不同的业务通过安全分区及VLAN隔离。一个区域通常集中了该业务所需的计算、网络及存储资源，不同的分区之间经由核心交换通过三层网络交互，数据中心的网络流量大部分集中于南北向。在这种设计下，不同分区间计算资源无法共享，资源利用率低下的问题越来越突出。通过虚拟化技术，将各个分区的资源进行池化，实现数据中心内资源的有效利用。而随着虚拟化、云计算技术的兴起和应用，新的业务需求如虚拟机迁移、数据同步、数据备份、协同计算等在数据中心内开始实现部署，数据中心内部东西向流量开始大幅度增加。

同时，随着虚拟化技术的进步，服务器间的低时延通信和更高的双向带宽需求变得更加迫切。然而传统的三层网络结构，造成数据中心计算节点间通信延时较高，这就驱动网络架构向扁平化方向发展，尽量减少网络层次的数目。

采用VSU技术（Virtual Switch Unit，即虚拟交换单元）的网络结构（图2）把传统网络中的两台核心层交换机虚拟化为一台逻辑虚拟交换机，和汇聚层交换机通过聚合链路连接，提高网络连接的可靠性和传输带宽。采用VSU技术，在核心层和汇聚层之间避免出现二层环路，同时由于VSU减少了路由实体的数目，简化了网络拓扑。在数据中心内部，通过采用VSU技术，可以构建大型的二层交换网络，具有网络结构简单、高可靠、高性能、避免环路的好处。

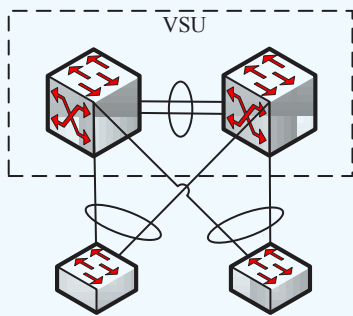


图2 VSU技术网络结构

## 1.2 数据中心网络安全架构

传统的数据中心安全架构通常采用串行部署方式，存在如下问题：

- 动网断网：设备升级、更换都需要进行网络割接，造成断网；
- 性能瓶颈：串接链路上，品牌不统一，性能参

差不齐，整个链路的性能较差，转发时延随着设备的增加而增大；

- 利用率低：无论是否有该方面安全需求，数据流量都要经过全部串接安全设备，无法按需定义分区流量。

SDN（Software Defined Network）软件定义网络，将传统网络设备的数据平面和控制平面两个功能模块相分离，通过集中式的控制器(controller)以标准化的接口对各种网络设备进行管理和配置。这种架构为网络带来的可编程的特性，使得网络资源的设计、管理和使用提供更多的可能性，使得SDN日趋成为解决数据中心网络安全问题的主要手段之一。

利用SDN Service Chain技术，可以改变传统数据中心中安全设备串行部署的方式，通过引流方式实现各区域及业务应用间的按需安全控制，为数据中心提供弹性、简单、高效的安全保障。

## 2 设计原则

- 稳定性和可靠性

采用主流网络技术和市场覆盖率高、技术成熟的产品，保障网络及其所承载的业务系统的稳定运行。

- 业务效率优先

在对业务需求进行充分分析的基础上，本着提升业务网络通信效率，降低跨网段、跨设备交换的通信消耗，进行网络结构设计和选型。

- 整体性

以整体效益为立足点，综合考虑业务、平台、网络、场地环境等因素，提升网络性能、可靠性、扩展性、可维护性，降低建设及维护成本。

- 可扩展性

在网络设计中充分考虑未来业务发展及其网络需求的变化，需满足业务需求。

- 集约适度

根据不同业务需求，进行网络分级分类，按照业务所需不同性能、可靠性要求匹配适当的网络类别及设备。

- 开放性和兼容性

所有技术和设备都遵循相关的国际、国家或行业规范。符合相关标准的不同厂家设备能够相互兼容。

## 3 网络架构设计

### 3.1 总体架构

当前流行的数据中心网络架构，通常按照业务和安全等级划分少量几个分区，各个分区连接到业务核心。为了实现气象业务的高速高效转发，减少网络性

能瓶颈，在总体架构上，通过采用VSU和SDN技术来实现高效、可靠、高安全的扁平化业务网络环境。

业务网主要承担各个业务系统的接入和数据交换功能，具有东西向流量大、交互频繁的特点，对网络转发效率、传输速率、网络可靠性的要求较高；管理网主要用于业务系统及设备的管理流量转发，主要包括系统管理流量和带外管理流量。管理流量具有流量较小，交互少，对网络时延的要求远低于业务流量，但对安全性要求较高。根据业务和管理流量的不同需求，本文采用异构的业务网和管理网架构，业务网采用核心层加接入层的大二层扁平化网络结构，管理网采用传统的分布式网络架构。

### 3.2 业务网架构

数据中心业务网采用核心层加接入层的大二层扁平化网络结构，与局域网进行互连，实现各业务系统的网络接入和数据交换。为满足数据中心各类资源系统对业务通信高速、高效、高可靠、低延时、低成本、少跳转、少瓶颈的要求，业务网将采用集中式接入的网络结构，交换机均采用高性能的数据中心核心级交换机，通过采用VSU技术，提高核心交换机转发性能和可靠性，实现各业务系统的网络连接高度集中，减少跨网络设备的数据交互及上联瓶颈，降低网络转发时延。

根据各业务系统重要性、时效要求和系统耦合性不同，将交互耦合度和时延要求高的在线实时业务定义为高IO系统，如MICAPS、CIMISS等；交互耦合度和时延要求低的非在线实时业务或实验系统定义为中低IO系统。

高IO系统采用双链路捆绑直接上连至2台核心交换机，各业务间数据交换在核心交换机内部采用背板交换的方式进行转发，不存在跨交换机的数据转发，减少跨设备的转发时延。同时由于数据中心级交换机的性能更优，经测试得到转发时延小于4 μs，可将数据交换的网络时延降至最低，实现了实时业务系统数据的高速高效转发。

中低IO系统采用双链路捆绑直接上连至2台中低IO汇聚交换机，2台中低IO汇聚交换机与核心交换机间采用多40G链路捆绑互连，在降低转发时延的同时，保障了高IO系统和低IO系统间数据交换具备足够的网络带宽。

同时2台核心交换机和中低IO汇聚交换机分别采用VSU技术虚拟化为1台逻辑交换机，在增加业务系统接入带宽的同时，增加了系统可靠性。业务网网络架构如图3所示。

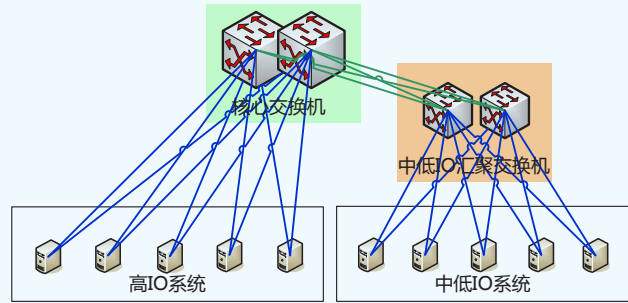


图3 业务网网络架构

### 3.3 管理网架构

管理网的数据流量一般为系统监控、远程管理及虚拟机迁移，网络流量较小，时效要求不高，目前局域网络时延不会影响管理通信；同时，其对网络可靠性要求不高，短时中断一般不会影响到业务运行；另外，统筹考虑综合布线和建设管理成本，管理网采用分布式接入的网络结构，即采用汇聚+接入的方式进行组网。同时考虑业务系统管理需求，将系统管理网和带外管理网分别单独组网，充分保障管理网络的稳定可靠。管理网架构如图4所示。

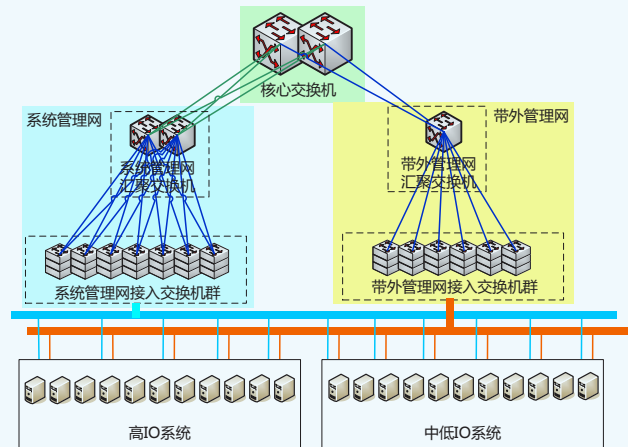


图4 管理网网络架构

### 3.4 转发时延测试

为了确认本文中网络架构的网络转发性能，采用如图5所示的测试拓扑，采用专用测试仪对核心交换机、汇聚交换机和接入交换机的单接口转发时延进行测试，据此可估算数据中心业务网络的转发时延。测试结果如表1所示。

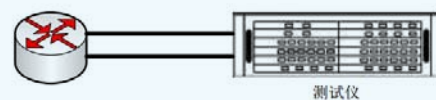


图5 转发时延测试拓扑

通过对核心交换机和接入交换机接口转发时延进行测试，以64 kbytes的数据包为例，可以估算出传统

表1 核心交换机、汇聚交换机、接入交换机转发时延测试结果

数据包/bytes	测试次数	最小时延/ $\mu$ s		最大时延/ $\mu$ s		平均时延/ $\mu$ s	
		核心/汇聚交换机	接入交换机	核心/汇聚交换机	接入交换机	核心/汇聚交换机	接入交换机
64	100	1.96	104.54	2.42	152.24	1.986	150.96
128	100	1.85	105.09	1.97	127.43	1.93	126.505
256	100	1.86	106.11	2.1	129.91	2.047	129.063
512	100	1.85	109.86	2.2	136.44	2.132	135.736
1024	100	1.88	115.52	2.34	146.55	2.295	145.794
1280	100	2.05	123.58	2.33	148.3	2.193	147.454
1518	100	1.89	124.18	2.3	150.45	2.253	149.663
9216	100	2.02	190.7	3.74	270.55	3.681	268.883

三层网络架构（图6）和本文提出的网络架构（图7）中业务通信的转发时延范围（表2）。采用二层扁平化网络架构的转发时延小于10  $\mu$ s，比传统三层树形网络架构降低了98.7%。同时由于将交互耦合度高的系统连接至同一对核心或汇聚交换机上，业务系统间通信主要为东西向流量，南北向流量大幅度减少，可降低上联接口成为带宽瓶颈的概率。

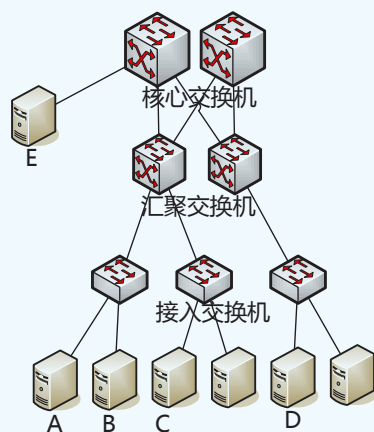


图6 传统三层树形网络架构拓扑

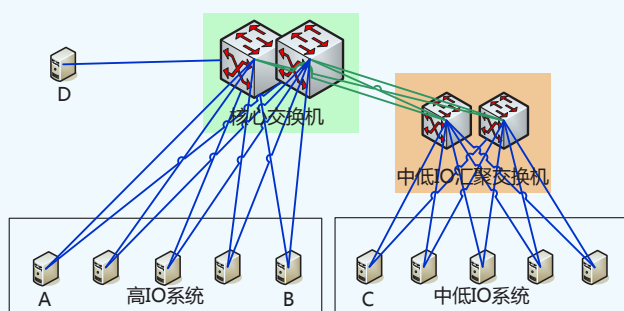


图7 本文业务网架构拓扑

表2 传统三层树形和本文网络架构的转发时延 (单位:  $\mu$ s)

传统三层树形网络架构转发时延		本文网络架构转发时延	
A-B (同一接入交换机)	301.92	A-B (同一核心交换机)	3.972
A-C (跨接入交换机)	607.812	A-C (跨汇聚交换机)	7.944
A-D (跨汇聚交换机)	615.756	A-D (跨核心交换机)	3.972
A-E (跨核心交换机)	309.864		

### 3.5 数据中心安全架构设计

为了提高数据中心的安全水平，采用如图8所示的SDN Service Chain架构，通过在数据中心核心交换机处部署SDN控制器，使用安全引流技术，使旁挂于核心交换机处的安全设备形成安全资源池，实现数据中心各业务系统安全控制的按需引流。在提高数据中心各区域安全防护能力，实现安全设备的高效集约利用，提高数据中心网络和安全架构的可靠性。

## 4 结语

数据中心网络系统是数据中心的基础，是气象业务系统、云平台、高性能计算机和存储等系统的基础运行环境和数据交换枢纽。为了适应云平台等不断发展的资源池网络需求，在网络架构方面，通过采用VSU技术，搭建扁平化数据中心网络架构，减少业务通信网络转发环节，可有效降低转发时延，避免网络瓶颈，满足气象数据中心内部大流量数据交换需求。通过建设异构的业务网和管理网，将业务和管理流量分开，提高业务系统的安全性和可靠性；在安全架构方面，采用SDN Service Chain 架构，实现数据中心各区域间安全控制的按需引流，提高安全设备的集约化水平，为各业务系统提供可定制化的安全方案。

#### 深入阅读

- Guo C, Lu G, Helen J, et al, 2010. SecondNet: A data center network virtualization architecture with bandwidth guarantees. ACM CoNEXT 2010.
- 程容斌, 王宏, 杜勇, 等, 2011. 数据中心网络结构综述. 电脑知识与技术, 7(35): 9029-9032.
- 王斌锋, 苏金树, 陈琳, 2016. 云计算数据中心网络设计综述. 计算机研究与进展, 53(9): 2085-2106.
- 魏祥麟, 陈鸣, 范建华, 等, 2013. 数据中心网络的体系结构. 软件学报, 24(2): 295-316.

(作者单位: 国家气象信息中心)