解决方案

基于InfiniBand网络的Hadoop集群

在运营商领域的实践案例

数据爆发给运营商带来挑战

在电信运营商大力推进3G网络建设，移动智能终端迅速普及的背景下，网络应用已经成为比肩语音通信的重要应用之一，其发展在很大程度上影响着运营商的未来业务前景。但同时，由于数据量的迅速增长，许多运营商的流量经营分析系统面临着巨大的性能压力，一旦处理不好，很容易成为消费者积累、爆发矛盾的环节之一。在近些年，某些运营商由于流量经营分析系统不完善或是处理能力有限，导致部分用户的账单出现疏漏、错误，这不但会增加应对消费者不满情绪的额外时间成本，也会给运营商的品牌形象带来损失，造成不良的社会影响。

因此，移动互联网和大数据汹涌来临的今天，流量经营分析系统的升级换代已经成为了运营商的大势所趋。一方面，传统的流量经营分析系统的性能即将达到满负荷，不能为流量提供长期的支撑。同时，大数据时代也要求运营商对用户终端产生的流量进行更精细的分析，对来源于微信、微博等互联网应用的数据进行精确的统计、管理，从而准确把握消费者的流量使用习惯，实现精准营销。这就迫切的要求运营商搭建一个处理能力更强，扩展性更高的平台来替代现有系统。

传统的以太网网络已经不能完全满足大数据环境所需要的性能。运行TCP/IP的典型数据中心集群，通过一个或多个千兆以太网网卡连接到千兆以太网主网，也只能达到每端口125MB/s的带宽。多CPU多核服务器对网络的需求早已超过了千兆以太网的网络容量，随着处理器技术的进步，主流市场上很快就会出现具有上百个内核的计算服务器。想要针对这样的服务器达到最高效率，就必须为每个服务器提供足够的带宽，以及提供避免数据通信大量占用CPU资源的CPU　卸载能力。

InfiniBand帮助运营商突破性能瓶颈

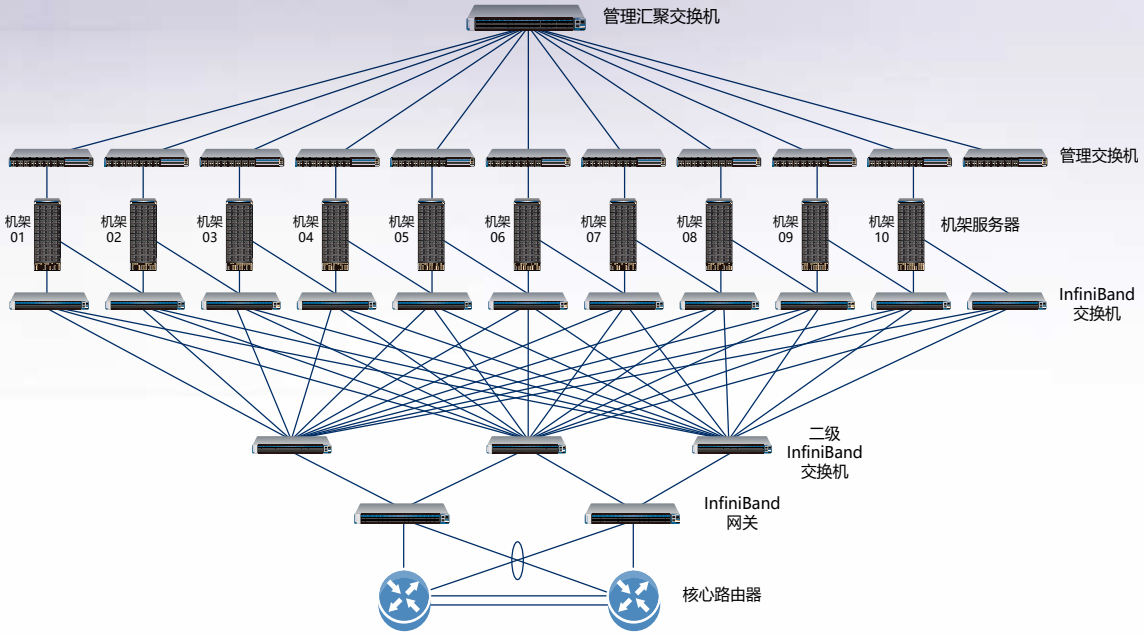
InfiniBand技术是一种开放标准的高带宽、高速网络互联技术。借助InfiniBand架构组建的网络，从处理器到系统I/O，到存储网络，甚至整个数据中心都达到处理器级的带宽。InfiniBand网络可为用户提供云计算和下一代数据中心最为全面的解决方案，实现高带宽速率、高扩展性、高性能、低延迟，同时可以提供高密度、低功耗、低成本、构架简单的数据中心解决方案。

相比较以太网，InfiniBand网络技术优势如下：

- 高带宽，单宽口可提供高达40/56Gbps的速率，适应各种应用环境
- 超强硬件卸载能力。采用RDMA（远程直接内存访问）技术，实现硬件卸载功能，充分释放CPU性能，CPU处理通信的资源消耗仅有3%；而以太网的CPU处理消耗大于55%
- 超低延时，InfiniBand网络端到端延时小于1us，万兆以太网延时一般大于20us
- 低成本数据中心架构，相同性能的系统其投资仅为万兆以太网价格的60%
- 为存储和服务器提供最大性能、无拥塞网络互联，集群能获得更多的应用
- 集群采用IB网络，性能是万兆以太网的2倍，功耗是以太网的1/2

InfiniBand技术可以显著提高Hadoop数据库集群的系统性能。在Hadoop集群中，InfiniBand网络将集群数据处理的吞吐量提高了超过一倍，并为每个计算节点减少了一半执行时间。它的设计为更大的数据集（dataset）提供了相同或更好的性能。通过降低全部任务执行时间，增加每个节点的CPU　利用率，集群能够为特定系统配置增加相应的功耗效率从而直接增加数据中心的功耗节约，并与数据中心的绿色倡议目标一致。更高的带宽与基于RDMA技术的可扩展架构，使得单网线能够传输更大的数据量。

基于InfiniBand架构的Hadoop集群解决方案



图：某运营商的流量运营Hadoop平台的IB网络规划

在该网络设计中，所有的网络设备的连接均采用冗余设计。每台服务器配置两个IB端口，IB端口通过冗余绑定方式分别与两台不同的InfiniBand交换机相连，此外通过高度冗余的三台InfiniBand二级交换机，最大限度的提高系统的冗余度和降低节点间通信带宽限制。两个InfiniBand网关设备底层与InfiniBand二级交换机相连，上层分别与两台核心路由器连接。上层的两台路由器通过VRRP技术实现IB网关的冗余连接，保证当任意一台网关或路由设备出现故障时，底层服务器均有链路连接到上层网络。

上述网络架构采用来自Mellanox（迈络思）公司的InfiniBand端到端网络互连产品，包括基于Mellanox SwitchX-2芯片的InfiniBand交换机和网关，以及基于Mellanox ConnectX-3芯片的QDR或FDR网络适配器，借助其自身的设计优势，使整个方案具备了行业领先的高效能，高密度，高性价比，以及超低延迟。

平台优势

基于InfiniBand技术的流量运营Hadoop平台可以为运营商带来诸多优势，体现在以下几个方面：

- 高性能：借助Mellanox交换机严谨设计的交换结构，连接到Mellanox交换机上的节点之间端口与端口之间的通信带宽可以达到40/56Gbps，端口与端口之间的延迟小于0.7us，实现PB级大数据的高性能多维分析
- 高扩展：支持在线线性扩展，满足业务发展需要
- 高可靠性：InfiniBand交换机具备系统运行的可靠性，用户更高的可靠性还可以通过构建冗余路径的InfiniBand Fabric来实现，同时系统多副本机制保证数据高可用，避免单点故障
- 高性价比：X86架构，节省硬件和存储设备投资
- 易用性：实现数据库集群的无缝集成

实现Hadoop性能提升的更多创新

Hadoop集群的Master主节点和Slave从节点以分布式方式部署在机架服务器上，运行HDFS，MapReduce等服务。为了充分发挥InfiniBand高速网络的性能优势，Mellanox进一步开发了针对Hadoop的RDMA加速插件，包括UDA（Unstructured Data Accelerator）和R4H（RDMA for Hadoop Distributed File-System），分别应用于Hadoop MapReduce和HDFS，实现TCP/IP到RDMA传输协议的转换。UDA和R4H基于Verbs和Accelio协议开发，在部分应用场景下可以为Hadoop带来2倍的性能提升。Mellanox对Hadoop插件采用开源策略，且持续进行版本演进和开发。

北京迈络思科技有限公司

咨询电话：+86-10-57892000

销售咨询：china_sales@mellanox.com

市场合作：marketing_cn@mellanox.com

*欲了解更多欢迎登陆www.mellanox.com

微信

CN-SB1512 Rev 1.0