

非结构化数据加速器 Unstructured Data Accelerator (UDA)



非结构化数据加速器(UDA)

随着当今数据量的爆炸性增长, 以及在多个领域, 包括科技, web2.0, 云计算, 银行, 政府, 娱乐, 以及在这些领域相关公司的数据分析工作正以指数倍的速度采用Hadoop来实现商业目标。Hadoop是对用于数据分析的MapReduce编程模型的开源实现。它是一个容易使用的编程接口, 众多组织采用它来处理爆炸性数量级的数据、执行大型计算、以及为商业情报提取关键信息。现在, 由Apache Foundation维护Hadoop, 同时它也得到一些领先的技术公司, 如Google, Yahoo!, Facebook 和 LinkedIn的支持。

传统的以太网网络已经不能再提供Hadoop集群所需要的性能。运行TCP/IP的典型多级Hadoop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, 以及提供避免数据移动大量占用服务器CPU资源的CPU 卸载的能力。高带宽的InfiniBand以及以太网技术提供高达40Gb/s的带宽, 并且都具备RDMA (远程直接内存访问) 能力, 以卸载数据移动。然而采用RDMA, Hadoop需要一个特殊的针对网卡驱动力的接口。

Mellanox UDA解决方案

Mellanox UDA作为一个软件插件加速Hadoop网络并提升了运行大量数据分析的Hadoop集群的扩展性。一项创新的数据移动协议采用了RDMA技术并结合了一套有效的合并检索算法, 使基于Mellanox 万兆以太网网卡 (支持RoCE) 的Hadoop集群可以有效地加速Hadoop服务器之间的数据移动。这是首次借助基于RoCE的卸载功能来处理Hadoop节点数据访问和运算问题的方案。UDA对Hadoop用户是透明的, 现有的应用依旧保持原有的运行方式。只有集群的管理员需要了解UDA, 通过配置集群来运用UDA的长处。终端用户则只需要关注Hadoop的好处。

UDA性能

将数据处理的吞吐量提高了超过一倍, 并为每个Hadoop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InfiniBand和以太网RDMA技术的可扩展架构, 使得单网线传输更大的数据量。

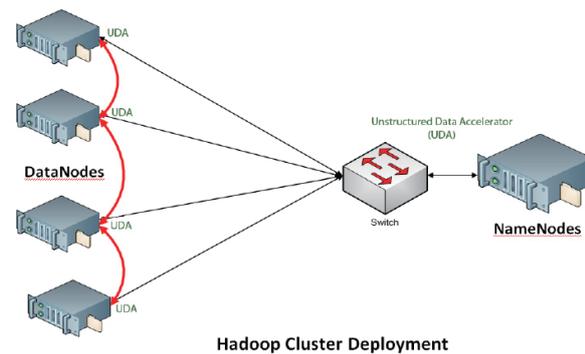


Figure1. Hadoop Cluster Deployment

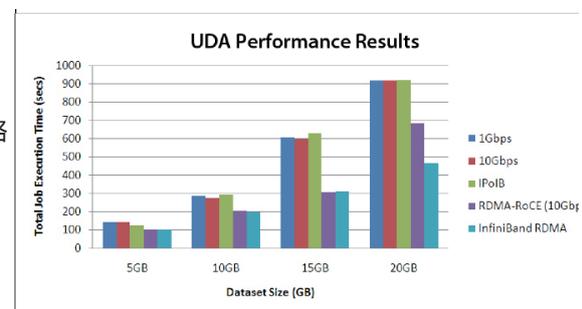


Figure2. UDA Performance Results

UDA的关键优势

- 采用世界最快的支持40/56Gb/s的InfiniBand或以太网底层结构
- 采用RDMA与高效Merge-Sort算法增加Hadoop效率
- 降低每个节点的全部任务执行时间
- 无损及可扩展的网络解决方案

如何获得UDA

有兴趣的用户可以获得UDA加速工具箱，在他们的数据中心上部署和评估UDA所带来的巨大好处。工具箱包含Mellanox UDA 加速器软件和组网硬件。有关如何订购工具箱的详细信息，请联系您的Mellanox销售代表或通过网站与我们联系 http://www.mellanox.com/content/pages.php?pg=buy_overview

UDA2.0是由Auburn大学Dr. Weikuan Yu领导的并行结构和系统实验室与Mellanox共同开发的。



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

