**BlueDBM: An Appliance for Big Data Analytics**

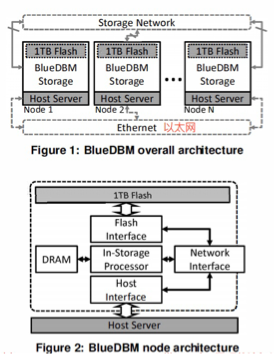
主机处理器➕FPGA模拟控制器

应用领域：一般性（计算功能可由用户编程）数据密集型应用

复杂的数据查询操作随机访问频繁，会带来很大的延迟和吞吐量开销，为减少开销并提高查询速度，可以将数据存储于内存DRAM中。然而，应用数据集往往是很庞大的，其相应所需的DRAM内存空间也将很庞大。

研究发现，架式结构（rack-sized，按文章可理解为分布式）闪存较易于存储规模庞大的数据库数据，且闪存相对于硬盘更优的随机访问性能也使得外存访问性能的问题迎刃而解。

据此，本文提出BlueDBM，一个基于闪存的可扩展的系统，该系统包括一个全局地址空间，一个灵活的inter-controller网络及一个in-store计算单元。

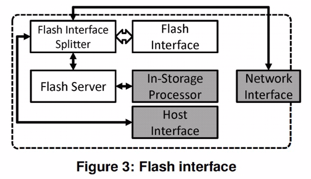


BlueDBM结构如图1所示，其结构是耦合有存储设备的主机服务器的集群。图2是其单个节点的结构，存储设备包括in-store处理器、多个接口及板上DRAM。

in-store处理器提供闪存接口、(存储)网络接口、主机接口和DRAM缓存功能以实现存储内计算。

**①闪存接口（移除闪存管理功能/高性能硬件接口/多硬件访问）**

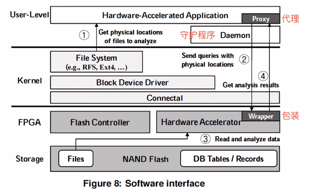
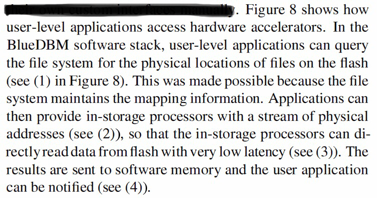
一般商用SSD控制器中包含有闪存管理功能，但其与FS部分功能重复，带来显著延迟和I/O开销。同时，BlueDBM的分布式结构使其独立存储单元缺少全局视角，难以有效管理全局闪存。为高效，闪存接口为高性能硬件接口，直接暴露给原始闪存芯片、总线、块和页面。并将闪存管理功能移交给软件层（即后续将提到的软件接口）。



对于某个节点的闪存控制器，可能会访问它的有1）本地（即该节点的）in-store处理器（通过flash interface）2）本地主机软件（通过host interface）3）远程（即其他节点的）in-store处理器（通过network interface），这会产生多重访问的问题。对于多硬件访问，flash Interface Splitter带tag对不同硬件访问进行区分，flash sever帮助梳理并简化存储内处理器对不同硬件访问的处理。（图3）

**②软件接口**

软件接口包括1) 文件系统接口，2) 块设备驱动接口，3) 加速器接口，其中文件系统提供有FTL的闪存管理功能，有助于开发者快速完成in-store处理操作。in-store处理器提供的支持用户自定义的软硬件协同设计环境的操作流程如图8所示。可以观察到，**查询和数据的传输均交由硬件处理，绕开了内核软件层，大大降低了能耗。**



③**（存储）网络接口**

集群中单个存储设备由高效连续链连接成整体的存储网络，具有全局地址空间，是一个灵活的inter-controller网络。各节点根据确定的路由通过网络接口传送数据。

Summary：

美国麻省理工学院于2015年针对大数据分析设计了一套分布式闪存存储系统BlueDBM。BlueDBM是一个基于闪存的可扩展的系统，该系统包括一个全局地址空间，一个灵活的Inter-Controller网络以及一个In-Store的计算单元，其中In-store处理引擎提供了一个支持用户自定义的软硬件协同设计环境。软件接口提供给用户文件系统接口、块设备接口、加速接口等三个可编程接口。最终实现一个复杂数据分析器，与同样功能的云系统相比，其能耗少一个数量级。