# 虚拟化世界中的大数据：

# 云端数据中心的容量，速率和多样性

***Robert Birke+, Mathias Bjorkqvist ¨ +, Lydia Y. Chen+, Evgenia Smirni∗, and Ton Engbersen+***

+IBM Research Zurich Lab, ∗College of William and Mary

+{bir, mbj,yic,apj}@zurich.ibm.com, [∗esmirni@cs.wm.ed](mailto:∗esmirni@cs.wm.ed)

**摘要：**虚拟化是为数据中心终端用户提供计算和存储服务的无处不在的方式。支持足够的数据存储和有效的数据访问是所有数据中心操作的核心，但我们队虚拟化对存储工作负载的影响知之甚少。在本次研究中，我们收集和分析了三年内在私有云模式下运行的生产数据中心的现场数据。我们研究的数据中心由8,000个物理盒组成，托管超过9万个虚拟机，而这些虚拟机又使用了超过22 PB的存储空间。存储数据是从容量，速度和各种虚拟机的存储需求以及对其他资源的依赖性的角度分析的。跟踪数据除了分析和使用存储量的增长率和流失率之外，还讨论了虚拟化和整合对IO读写速度的影响，包括IO去重比率。我们分析了各种应用，这些应用大致分为应用程序，Web，数据库，文件，邮件和打印的各种应用程序，并将其存储、IO需求与CPU、内存和网络使用相关联。本研究通过显示使用趋势以及应用程序类型如何在大型数据中心中创建存储流量来提供关键的存储工作负载特性描述。

1 引言

数据中心提供了广泛的数据相关服务。它们具有强大的计算能力，可靠的数据存储，快速的数据检索以及更重要的资源可扩展性。虚拟化是以无缝和安全的方式增加资源共享的关键技术，同时降低运营成本，同时不影响数据相关操作的性能。为了优化虚拟化数据中心的数据存储和IO访问，现有研究已经提出了存储和文件系统缓存技术[13,18,28]以及数据重复和重复数据删除技术[22]。中心的主题是将正确的数据移动到正确的存储层，特别是在共存虚拟机（VM）的高峰负载期间。因此，对于各个虚拟机的IO工作负载的特点，以及托管服务器的负载情况，这一点至关重要。有几个以存储为中心的研究已经阐明了文件系统卷[14,20,31]和IO速度，即读/写数据访问速度[15,17,28]。尽管如此，研究人员并不清楚虚拟化如何在数据中心范围内满足存储和IO需求，以及它们与CPU、内存和网络工作需求之间的关系。

本文的目的是从以下几个角度提供对数据中心存储工作负载更好的理解：存储量，读写本地性和应用程序多样性。使用来自在私有云数据中心内运行的生产数据中心的现场数据，我们分析了对应于8,000个物理盒子上托管的90,000个虚拟机的踪迹，并且包含了超过22 PB的活跃使用的存储，从2011年1月至2013年12月的三年时间范围内，由于可用数据的规模，我们采用黑盒方法对各种性能指标进行统计表征。由于缺乏关于系统拓扑和所采用的文件系统体系结构的信息，这项研究不足以分析存储设备的延迟，文件内容和数据访问模式。我们的分析提供了有代表性的虚拟存储工作负载的多层面视图，并阐明了高度虚拟化数据中心的存储管理。

通过收集的数据，我们可以查看每个虚拟机中虚拟磁盘的分配，使用和可用空间的数量，特别关注每年的增长率和每周的流失率。我们通过统计表征读写操作的负载（以GB / h为单位）以及每秒IO操作（IOPS）（即每小时，每天和每月）来测量速度，重点在于时间变异性和峰值负荷分析。我们通过分析盒子内共同定位的虚拟机的IO工作负载来推断虚拟环境中存储重复数据消除的效率。要了解存储和IO工作负载如何由不同的应用程序驱动，我们执行每应用程序分析，使我们能够专注于几个典型的应用程序，如Web，应用程序，邮件，文件，数据库和打印应用程序，突出其差异和IO使用的相似之处。最后，我们提出了一个以数据存储/访问为中心的详细的多资源依赖性研究，并提供了当前数据中心数据管理实践的见解。

我们的研究发现可以总结如下：虚拟机容量和使用空间的年增长率分别为40％和95％。每台虚拟机的充满度都有19％的增长率，尽管三年来虚拟机的存储容量分布保持不变。虚拟机存储空间流失率下限为17％，略低于备份系统报告的21％的流失率[31]。

对于IO速度，尽管虚拟化的开销不可忽略，但是盒子的IO访问速率几乎与整合的虚拟机数量呈线性关系。虚拟机和虚拟机都由写入工作负载占主导地位，有11％的虚拟机的虚拟IO速率高于物理虚拟机。重复数据删除率随虚拟化程度呈线性增长。峰值负载在下班时间发生，并且只贡献给极少数的虚拟机。高流速的VM往往具有更高的存储容量和更高的流失率。

对于IO类型，不同的应用程序使用不同的存储方式，文件服务器应用程序具有最高的容量，满量程和流失率。数据库具有相似的特征，但是不够丰满。总的来说，我们观察到高IO需求与CPU和网络活动强烈和积极相关。

本文工作的大纲如下。第2节介绍相关工作。第3部分概述了数据集。体积，速度和品种分析分别在第4,5和6节详细介绍。第7节将讨论一个以数据为中心的多资源依赖性研究，接下来是第8节的结论。

2 相关工作

管理存储是一项昂贵的业务[19]。鉴于存储硬件的成本几乎是服务器硬件的成本，数据中心的有效使用存储变得至关重要[29]。存储/ IO的工作负载特性研究对于开发更好地使用系统的新技术是至关重要的，但是由于工作负载的种类繁多，很难确定真正具有代表性的系统。一般来说，从文件系统工作负载的各种研究中，除了大量较低级别的存储研究外，还有那些基于学术原型和基于个人计算机的研究。虚拟化为我的任何存储增加了额外的复杂层次[10,16]。由于虚拟化确实是数据中心使用的标准，因此虚拟化IO的工作负载研究是重要和相关的。尽管如此，分析所有相关虚拟化IO工作负载的所有相关特性不在本文的讨论范围之内。在这里，为了更好地理解在一个非常大规模的虚拟化环境中IO如何发生，给定收集的跟踪数据，我们进行统计分析。

通常情况下，相关工作涵盖了数量方面[2,14,20,30]，速度[17]和多样性，重点是文件系统。关于文件系统的数量，有几个研究集中在台式电脑上[2,14,20]。在四周[20]和五年[2]的时间段内使用文件系统元数据，提供了充分显示的性能趋势和统计数据，文件/目录计数，文件大小和文件扩展名。认识到需要更好地理解备份工作负载的行为，Wallace等人[31]提出了广泛的特征分析，并指出数据流失率大致为每周21％。他们的研究表明，物理驱动器的容量每年大约增加一倍，而其利用率仅略有下降。该研究将备份存储系统与主存储系统进行了比较，发现其完整性分别为60-70％和30-40％。传统上，备份系统的特征被用来推动重复数据删除技术的发展[20,24]。

大多数关于IO特性分析的工作都集中在非虚拟化环境中的特定文件系统，例如NFS [7]，CIFS [17]，Sprite / VxFs [9]，NTFS [25]以及EMC Data Domain备份系统[31]。共同的特征包括大的和顺序的读取访问，增加读/写比率，突发性的IO，以及占据大部分文件活动的一小部分工作。自相似行为[9]被识别出来并被提议用来合成文件系统流量。据观察，备份系统[31]的写入次数比读取次数多得多，而主应用程序的文件系统的读取次数是写入次数的两倍[17]。

随着虚拟化技术的进步，近期的一些研究着重于优化虚拟环境中的数据存储和访问性能，重点在于新型共享存储设计[11,13]和数据管理[15,18,28]。为了减少共享存储系统的负载，提出了分布式虚拟机存储系统，如Lithium [13]和Capo [28]。古拉蒂等人通过考虑多个虚拟机的IO需求，设计和实现在大量虚拟机之间共享的存储资源池的概念[11]。已经提出了旨在改进使用性能模型的虚拟化数据中心的IO负载平衡的系统[10,23]。结合智能缓存，IO重复数据删除可以通过减少不同存储层（如虚拟机，托管框[18]和磁盘[15]）之间的重复数据来实现。珠峰[21]通过允许将写入过载卷的数据暂时卸载到短期虚拟存储中来解决数据中心中IO负载峰值的挑战。 Nectar [12]提出通过自动生成数据的过程来交换管理数据中心的计算和数据存储，从而释放不经常使用的数据的空间。针对评估能源使用的目的，已经完成了专注于特定服务器工作负载（即，应用程序多样性）的工作量特征，例如web，数据库，邮件和文件服务器。到目前为止，只有一个相当小规模的虚拟存储工作负载特征[28]，指出虚拟桌面工作负载是由峰值定义的。

这里介绍的工作负载研究展示了生产数据中心的虚拟机存储需求的广泛概述，包括IO容量，速度和变量，以及这些与虚拟化程度以及其他资源使用情况的关系。此处的分析通过在高度虚拟化的环境中使用来自生产数据中心的非常大的数据集，恭维了许多现有的IO和文件系统研究。

3 统计

收集我们调查了90,000个虚拟机，托管在分布在全球各地的不同数据中心的8000台物理服务器上，在三年内为来自各行业的300多家企业客户提供服务，并计算出22 PB的存储容量。服务器使用多个操作系统，包括Windows和不同的UNIX版本。 VMware是流行的虚拟化技术。对于当前虚拟化实践的工作负载研究，我们引导感兴趣的读者[5]。

收集到的跟踪数据通过vmstat，iostat和监控程序特定的监控工具进行检索，并收集到VM和物理服务器（称为托管箱）。每个物理盒子可以承载多个（虚拟）文件系统，这是本研究中考虑的存储介质的最小单位。为了表征虚拟化数据中心中的数据工作负载，我们专注于三种类型的与IO相关的虚拟机统计信息。

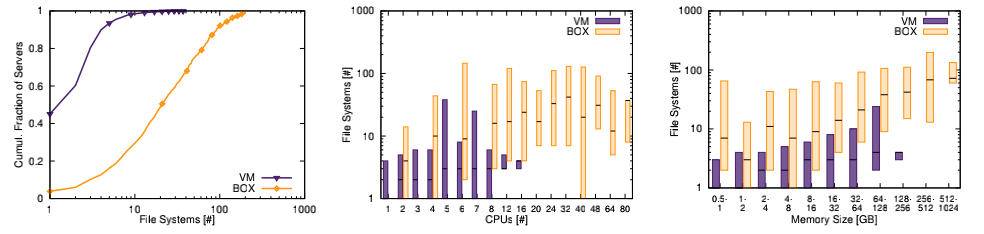
容量是指虚拟机在聚合所有文件系统后，分配的空间，可用空间和完整度，定义为虚拟机使用总空间量和总分配空间量之比。在这里，我们关注长期趋势，即增长率和短期变化，即流失率。

速度是指以操作次数和每时间单位传输的字节数来衡量的读写速度，分别为IOPS和GB / h。我们比较了在虚拟机上测量的虚拟IO速度和在底层机器上测得的物理IO速度。

各种指的是特定应用程序的数量和速度，即在特定的虚拟机上的应用程序，网络，数据库，文件，邮件和打印。为了进行以存储为中心的多资源分析，我们还收集虚拟机和盒子的CPU利用率，内存使用情况和网络流量。

跟踪数据有两种粒度可供选择：（1）2013年4月15分钟/小时平均;（2）2011年1月至2013年12月粗粒月平均。一天中，我们从04/17和04/21使用15分钟/小时粒度的详细跟踪。每月平均值用于推导长期趋势。

我们注意到，感兴趣的统计数字有很长的尾巴，因此我们着重介绍CDF以及一定的百分位数，即第10，第50和第90百分位数。由于盒子上的虚拟化程度（即整合）是相当动态的，因此我们会报告每个物理盒子的每日平均值。为便于将每个虚拟机存储需求与每个文件系统存储需求进行连接分析，我们提供跨越虚拟机和盒子的文件系统数量的CDF（见图1（a））以及文件系统分布在不同的系统之间是不同的，我们根据每个盒子和内存的CPU数量来区分（见图1（b）和（c）中的盒子图）。图1（a）显示，虚拟文件系统的虚拟文件系统通常比虚拟文件系统（虚拟文件系统平均具有2个以上）多得多。这些值与桌面研究[2]非常不同，并强调了我们数据集的独特性，尤其是在虚拟化数据中心的情况下。而且，从图1（b）和图1（c）的中位数趋势来看，文件系统的数量随着装有更多CPU的服务器而增加，特别是更大的内存。

**(a)CDF (b)Number of CPUs (c)Memory size[GB]**

**图1 Number of file systems associated with a VM and a box**

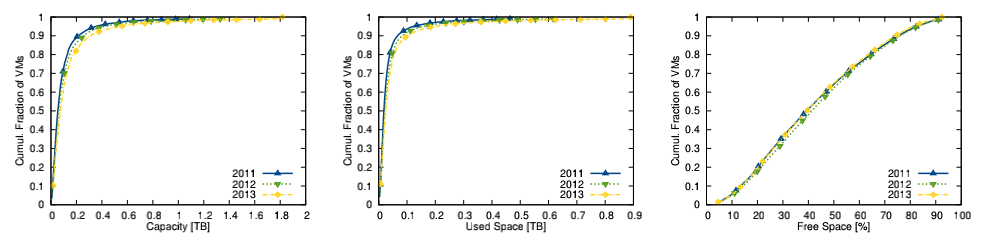
由于我们的数据是由操作系统级别的标准实用程序获得的，因此我们缺少关于文件系统的具体信息，例如类型，文件数量，深度和扩展。另外，由于追踪数据的细粒度为15分钟/小时，所以不能观察到这样的间隔内的IO峰值。例如，本研究中确定的一天内最大GB / h是基于小时平均值，并且远低于瞬时最大GB / h。收集到的信息的粗糙程度与本研究的庞大数据集形成鲜明对比：在三年的时间内观察到8,000箱高平均固结水平，即每箱10 VM。

4 容量

数据中心管理的核心业务之一是规划存储容量以处理短期和长期的存储需求波动。重复数据删除和备份活动进一步强化了这些波动[6,20]。数据需求和数据保存周期的不断增长推动了存储决策;现有的预测研究或者采用用户或者每个文件系统的视角，而不一定针对整个数据中心。在这里，我们的目标是采取不同的观点，并提供虚拟机级别的存储需求的年增长率和每周流失率。在下面的小节中，我们分析了90,000个虚拟机的存储需求，使用/分配的存储空间和容量，然后对其年增长率和每周流失率进行统计分析。

4.1跨越虚拟机的数据存储需求

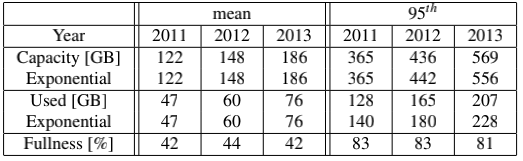
以2011年，2012年和2013年的受监控虚拟机的年平均值为例，介绍存储需求随着时间的推移如何演变以及如何在虚拟机之间分布。图2（a）和2（b）给出了属于每个虚拟机的所有文件系统上每个虚拟机分配和使用的存储量总和的CDF。图2（c）总结了由此产生的丰满度。目视检查显示，每个VM的总体容量和使用空间同时增长，并导致随着时间的推移而变得丰满。这个观察结果说明了与文件系统级别[20]相似的行为，并提供了如何为以数据中心为基本计算单位的数据中心维护存储系统的情况。

**（a）Storage Capacity (b)Used Volume (c)% of fullness**

**图2 CDF of storage volume per VM over three years**

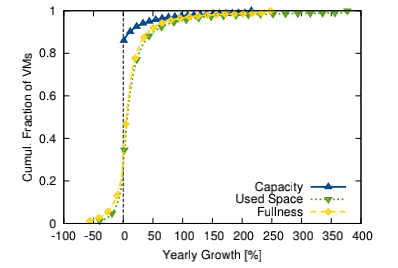
通过简单的统计拟合，我们发现指数分布可以很好地捕获虚拟机的存储需求，分配的存储容量和使用的存储容量。表1总结了测量值和拟合值，平均值和第95百分位的容量和使用量。由于平均有10个虚拟机共享同一个物理盒子[5]，因此系统需要配备450GB的存储空间用于非常积极的存储复用方案，即只考虑已用空间（45× 10）或1120 GB，根据分配的容量（112×10）进行更保守的整合方案。均匀分布可以近似地建立充满度。由于两个独立的指数随机变量的相对比较均匀[26]，这进一步证实了指数分布是一个很好的拟合。总的来说，上面的分析给出了整个虚拟机群体的趋势，这些虚拟机群体随着年份的增长而增加，但是没有提供关于单个虚拟机的存储容量如何变化的任何信息。在下面的章节中，我们通过给整个虚拟机群体提供CDF来关注每个虚拟机的每年增长率和每周虚拟机的流失率的上限。

**表1 Three year storage volume**

平均而言，虚拟机有2.55个文件系统，总容量为185 GB，其中大约占42％，这意味着每个虚拟机平均存储77 GB的数据。一般来说，多年来分配的容量和空闲的存储空间增加，而存储容量保持不变。

4.2 年增长率

数据增长率预计每两年翻一番[1]。但是，在每个虚拟机数据量水平上，这个价值如何转化为增长，或者更重要的是，现有的存储资源是否能够支撑未来的数据需求，目前还不清楚。在这里，我们从两个角度来分析长期的成交量增长率：供给，即从存储能力和需求角度，即从使用的存储容量的角度来看。

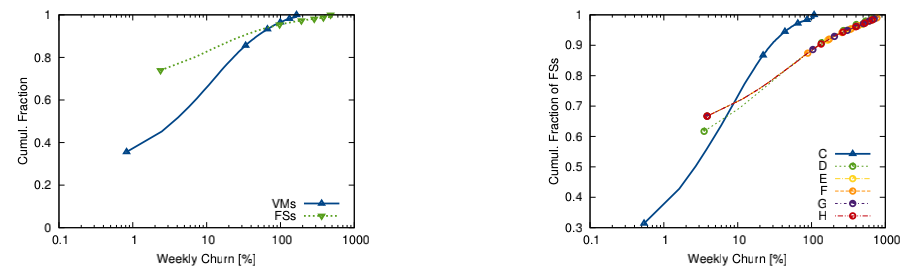


**图3 CDF of yearly growth rate of VM storage volume**

在图3中，我们显示所有虚拟机的分配容量，已用空间和全部年度相对增长率的CDF。我们计算2012年6月至2013年5月的相对年增长率作为使用容量的差异，并将其除以起始值。正（负）增长意味着增长（减少）趋势。总的来说，已用空间的CDF非常接近完整性，这意味着存储空间的利用受到数据需求的高度影响，而不是由容量的供应。

人们可以看到，大多数虚拟机（大约86％）没有提高存储级别，而剩下的14％虚拟机将存储容量显着提高，即高达200％。由于这个长尾巴，平均增长是40.8％。对于空间的需求，几乎所有的虚拟机都增加了他们的存储空间。只有少量（低于25％）的虚拟机减少了使用空间，并有负增长。另一方面，一些虚拟机在使用空间上有三倍的折叠。最终结果是，使用空间的平均增长率为95.1％。最小的增长属于丰满：平均比率为19.1％。图3：VM存储容量年增长率的CDF：容量，使用空间和丰满度。而不是图2（c）中整个虚拟机群体的饱满度趋势。每个单独虚拟机的存储容量和使用空间都会随着时间的推移而逐渐增加，分别年均增长率分别为40％和95％。由此产生的丰满度也每年增加19％。

4.3每周流失率：下限

 在这里，我们研究存储容量利用率的短期波动，通过在一周的时间段内相对于已用空间删除的字节的百分比来定义。请注意，此值表示流失率的下限，因为跟踪中可用的是15分钟间隔内的总量，即如果VM在15分钟内写入和删除相同数量的数据，那么没有办法知道在这段时间内真正删除了多少。因此，我们在这里报告了流失率的下限;真正的价值可能比这里报道的更大。流失率下限的倒数显示了数据保留期的上限。例如，这里每周20％的流失率意味着数据在被删除之前保留5周。我们以2013年4月22日至2013年4月28日期间收集的15分钟数据为基础计算虚拟机的每周流失率。流失率被计算为已用空间中所有相对下降的总和，即两个相邻的15分钟样本之间的所有负差异。我们注意到，由于数据也在这一个星期的时间范围内增加，我们考虑所有删除的数据的总和，这个值可以超过100％。

**(a)VM and file system (b)Specific file stystems**

**图4 CDF of the weekly churn rate computed based on single VM and a single file system**

我们在图4（a）中给出了两类每周流失率的CDF：虚拟机和文件系统（FS）。前者给出虚拟机删除的数据量，后者侧重于从单个文件系统中删除的数据量。从文件系统CDF的起点和长尾看，很大一部分文件系统的流失率为零，而一小部分文件系统的流失率非常高。因此，在文件系统中观察到较高的流失率变化率。进一步验证这个观察，我们计算了Windows系统中最常见的文件系统卷标（即C，D，E，F，G和H）的流失CDF，这大概占了整个虚拟机的流行。如图4（b）所示，可以清楚地看到，与其他标签相比，卷标C的流失率非常低。这样的观察与Windows系统上的C驱动器存储很少更新的程序文件和其他驱动器用于存储用户数据的公共惯例相匹配。

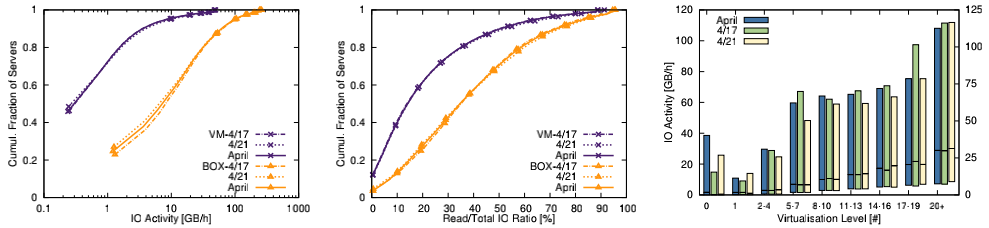
总的来说，虚拟机的流失率平均在17.9％左右，而文件系统的流失率平均在20.8％左右。这个值是一个下限，与文献中以前的结果是一致的，在这个结果中，从详细的文件系统轨迹计算出来的真正的流失率是21％[31]。大多数虚拟机具有相当低的流失率下限;从图4（a）可以看出，75％的虚拟机的流失率低于15％。但是，有10％的虚拟机的流失率高于50％。具有高流失率的虚拟机对存储系统提出了挑战，因为需要回收和写入大量的空间。

5 速率

存储系统最直接的性能衡量标准就是IO速度，在虚拟机访问数据中心的大数据的情况下，我们称之为速度。高峰期的性能[21]一直是优化的目标。为了加速IO操作，高速缓存[28]和IO重复数据删除[15]算法是至关重要的。在用于IO活动的系统栈（例如额外的管理程序层）变得更深和更复杂的虚拟化数据中心的情况下尤其如此。对虚拟化数据中心的缓存和IO重复数据删除方案的评估通常是在小规模或类似实验室环境下进行[15,28]。我们通过数据中心存储和检索数据的速度来量化虚拟机速度，并从IO角度进一步查明“热”或“冷”虚拟机。下面的小节中的统计数据是根据2013年4月17日起的小时平均值来表示的，它代表了5.1节中的IO速度。重点在于了解它们随时间的变化以及它们对虚拟化级别（即同时执行的exe虚拟机的数量）以及IO IO负载分析的依赖程度。

5.1 概述

1. 包括读写操作在内的每小时传输数据（GB / h），我们从虚拟机（及其对应的盒子）的日常速度概述开始本节。 （2）与读取操作相关的传输数据的百分比。图5描述了三种统计数据的上述信息：2013年4月17日（平日），2013年4月21日（周末）的平均小时数以及2013年4月整个月的平均日数。看看随机选择日期的IO速度是否足够代表性。总体而言，2013年4月17日的日均速度统计数据与周末日均速度的统计数据非常接近，而整个4月份的日均数据汇总数据也是非常接近的， 5.因此，在本文的其余部分，我们将重点放在2013年4月17日的特定日子，我们认为这是代表性的。

**(a)IO in GB/h (b)Percentage of reads (c)IO by virtualization level**

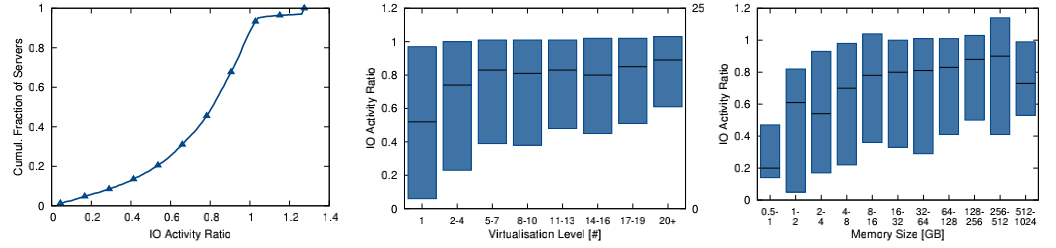
**图5 Daily velocity**

图5（a）中由较低的CDF显示，盒子具有比VM更高的IO速度。盒子和虚拟机的平均IO速度分别为26.7 GB / h和2.9 GB / h，即盒子的速度大概大了9倍。这个系数与平均固化水平一致[5]，即每盒10个虚拟机，并提示IO活动的线性缩放。关于读取操作的百分比，盒子的读取工作负载比虚拟机要大，如图5中的CDF曲线所示（对应盒子，大约有12％的虚拟机只有写入工作负载，最左边的点虚拟机CDF，同时，只有不到1％的虚拟机具有读取工作负载，实际上，虚拟机和虚拟机的平均读取率分别为38％和21％，虚拟机和盒子的速度主要由写工作负载。

为了验证虚拟化水平如何影响盒子IO活动，我们将虚拟盒子IO活动按虚拟化级别进行分组，并呈现第10个，第50个和第90个百分点，请参见图5（c）中的箱形图。盒子IO活动与虚拟化水平几乎呈线性关系，这可以从第50百分位看出来。当一个盒子的IO速度进一步正常化时，按虚拟机的数量来计算，每盒的平均值随着虚拟化水平而略有下降。这意味着与虚拟化相关的固定开销是不可忽略的。由于缺乏空间，我们省略了这个图表。

5.2 虚拟Io的重复数据删除

 IO重复数据删除技术[15]被广泛用于减少IO的数量。本节中的讨论仅限于虚拟IO，因为从跟踪中无法区分重复数据和/或缓存数据的方式和位置。我们比较一个框中所有整合虚拟机上聚合的所有虚拟IO活动的总和（称为虚拟IO）除以在底层物理框（称为框IO）中测量的IO活动，并将此比率称为虚拟重复数据删除率。与本文其余部分相比，我们在这里使用IOPS作为速度的度量，而不是GB / h。当重复数据删除率大于（或小于）一个时，虚拟IO分别高于（或低于）物理盒IO。以重复数据删除率1作为重复数据删除和放大的阈值。

**(a)CDF (b)By virtulization level (c)By memory size**

**图6 Virtual IO deduplication/amplification per box**

我们总结了图6（a）中重复数据删除率的CDF。大约50％的盒子具有从0.8到1.2的重复数据删除率，即接近于1，表明在物理和虚拟级别上具有类似的IO活动。另一个观察是大多数盒子经历放大，如通过重复数据删除率小于1（包括接近1）所指示的，即虚拟IO负载低于物理IO。这可以通过以下事实来解释：虚拟机管理程序由于虚拟机管理（例如，虚拟机迁移）而引发IO活动。

重复数据删除比率大于1的数据框表明，只有极少数的数据包（大约11％）出现IO重复数据消除。为了支持这种重复数据删除的原因，我们计算读写活动的单独重复数据删除率。我们发现观察到的重复数据删除比读取操作更多地取决于重复数据删除读取比率大于1的较高分组数（大约18％）。人们可以将这种观察与这样一个事实联系起来，即读取缓存技术比写入缓存技术更直接有效。

要了解虚拟化如何影响重复数据删除率，我们将重复数据删除率按其虚拟化级别进行分组，并使用箱形图表示，如图6（b）所示。查看每个箱型图的中下部，即第10和第50百分位数，我们看到，重复率随着虚拟化水平而增加。这样的观察可以通过以下事实来解释：位于同一位置的VM的IO活动具有一定的依赖性，这进一步为减少管理程序的IO操作提供了机会。更高的虚拟化级别可以带来更好的IO重复数据删除。我们注意到，使用IO以GB / h为单位可以得出类似的观察结果和结论，重复数据删除率大致在0到3之间。

除了虚拟化之外，IO重复数据消除的有效性可能高度依赖于缓存大小。不幸的是，我们的数据集不包含关于缓存大小的信息，只有内存大小，而这些大小通常与缓存大小正相关。因此，为了推断高速缓存大小与IO重复数据删除率之间的依赖关系，我们采用内存大小和猫箱内存大小的重复数据删除率，如图6（c）所示。趋势是，IO重复数据量随着内存容量的增加而增加，但是对于内存大于512 GB的系统来说则是下降的。

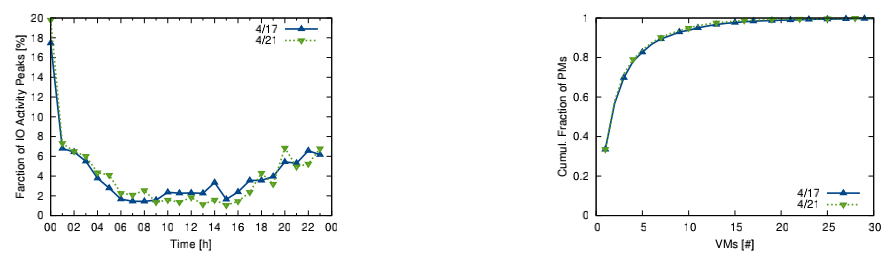
5.3 虚拟IO的峰值速度

虚拟化增加了访问模式的随机性，这是由于虚拟机和访问的数据量之间缺乏同步活动，这又给IO管理带来了一些挑战[8]。第一个问题是IO工作负载随着时间的推移如何波动。为此，我们使用小时数据计算每个虚拟机和盒子中IO活动的变化系数（CV），单位为GB / h。 CV值越高，白天IO工作负载的可变性越高。我们的研究结果表明，盒子的IO速度相当稳定，平均CV约为0.8，而VM的平均CV约为1.3。

虚拟机具有更高的时间可变性的确认导致我们将重点放在同一个盒子上托管的所有虚拟机上的虚拟IOAG的特性，特别是它们的峰值负载。我们试图捕捉什么时候发生聚合速度的峰值，以及每个虚拟机如何贡献最高峰。我们根据小时IO活动数据在星期三（2013年4月17日）和星期日（2013年4月21日）进行此操作。

5.3.1 峰值时间

图7显示了经验频率，显示了聚集的虚拟高峰IO负荷发生在一天中的哪个小时。很明显，大部分虚拟机在两个工作日内都会有高峰，也就是在下午六点到六点之间。这一观察结果与峰值CPU [4]和峰值网络[3]活动的时间非常吻合，但与IO工作负荷由工作时间表驱动的观点不符[18]。事实上，在之前的工作中[5]，我们观察到大多数虚拟机迁移发生在午夜/凌晨，这与图7中看到的活动一致。显然，虚拟IO工作负载的强度受备份等后台活动的影响并更新通常在下班后运行的操作。

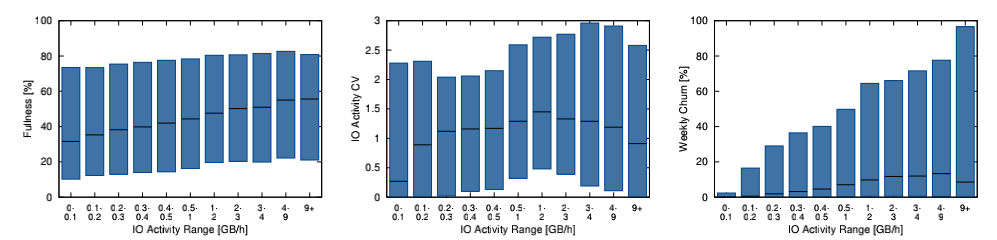
**图7 PDF of virtual loads peak times in a day overall consolidated VMs 图8 Number of VMs to reach 80% of peak load over all consolidated VMs**

5.3.2顶级VM贡献

另一个有趣的问题是如何整合虚拟机对峰值负载的贡献。关于顶级VM控制器的峰值负载信息对于通过缓存来提高峰值负载性能至关重要[21,28]。为了达到一定的阈值，即本研究的峰值负荷的80％，我们定义了对峰值负荷贡献最大的同位置VM作为最高负荷。我们总结了图8中两个最高虚拟机的数量的分布情况。有趣的是，可以看到一个明显的趋势，表明少数虚拟机在两天中都支配高峰负载是很常见的。这样的发现与文献[28]中报道的类似，只考虑了独立（即不共存）的虚拟机。这些结果进一步表明，优先考虑一些顶级虚拟机的IO的优化可能会对整体性能产生很大的影响。

5.4 冷热VM的特点

受到少数VM支持峰值负载这一事实的驱动，我们试图根据虚拟机的IO活动以GB / h来捕获虚拟机的特性，目的是将虚拟机分类为冷/热。数据的热度对维度和层级存储系统非常有用;例如，慢速存储介质中的冷数据和闪存驱动器中的热数据。为此，我们比较了不同级别的IO活动分组的虚拟机的使用量，时间变化率和流失率，分别参见图9（a），（b）和（c）。每个框表示一组具有落入x轴所示的IO活动范围的平均活动的VM。

**(a)Fullness (b)Time variabilty (c)Weekly churn rate**

**图9 Cold vs Hot VMs**

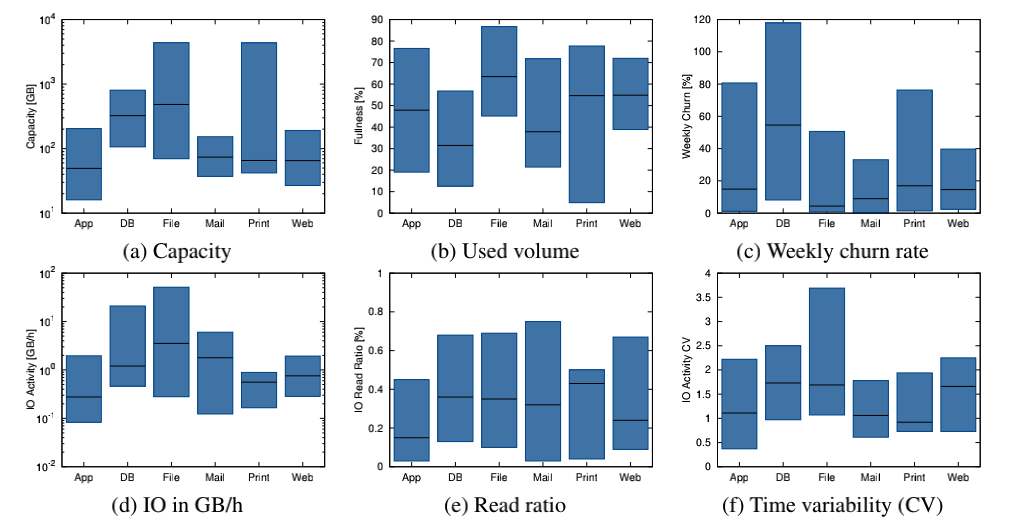
第50个百分点，即每个箱形图中的中间栏，随着IO活动水平的升高和流失率的增加而增加。总体而言，与IO活动较少的虚拟机相比，具有较高IO活动的虚拟机也更加完善，流失率更高。为了丰满，不仅是第50百分位，而且整个盒子也随着IO活动水平而变化。为了查看反向是否也是如此，我们按照GB和百分比的不同使用空间等级对IO活动等级进行分类。数据显示，高空间使用确实导致了高活动性，特别是当以GB测量时。但是，具有非常全面的存储系统（即90-100％占用率）的虚拟机的IO活动略低于80-90％的虚拟机。这源于大多数存储系统在没有完全满的情况下具有最佳性能的事实。通常的经验法则是当使用空间高达80％时，达到最佳性能。因此，只有冷数据放置在使用空间百分比较高的磁盘上。由于空间的限制，我们省略了这套结果的介绍。

时间变异性表现出不同的趋势，即CV随着IO速度增加而先增加，但后来减小，见图9（b）。最热的虚拟机，即IO大于9 GB / h的VM，具有第二低的CV，从第50百分位可以看出。因此我们得出这样的结论：热虚拟机具有相对恒定的IO负载。

就流失率而言，每百分位50和90分明显随着IO活动水平而增长，表明IO活动和流失之间有很强的相关性。这样的观察与常见的情况非常吻合，因为热虚拟机具有频繁的读/写，导致频繁的数据删除和短的数据保存期。数据显示，50％的热门虚拟机，即IO活动级别为9 GB / h或以上的虚拟机的数据保留时间在11.11（1 / 0.09）和1.02 1 / 0.98）周。总的来说，虚拟机虚拟机具有较高的容量消耗（55％）和流失率（9％）。

6 种类

跟踪数据允许区分VM的一个子集的应用程序类型。在这里，我们选择以下应用程序：应用程序，网络，数据库（DB），文件，邮件和打印，并表征其体积和速度。我们的目标是提供定量分析和定性分析，可用于存储系统的应用驱动优化研究。应用程序服务器托管客户端的关键应用程序，如业务分析。数据库服务器运行不同的数据库技术，比如DB2，Ora cle和MySQL。文件服务器用于远程存储文件。由于商业机密，不可能提供有关这些应用程序的详细信息。我们使用每个应用程序类型的盒形图来总结存储容量，已用空间，每周流失率，IO速度，读取操作的百分比和时间变化性，参见图10.我们将属于每个应用程序的虚拟机的第10，第50和第90百分位。除了基于2013年4月22日至2013年4月28日期间的数据的每周流失率外，大部分统计数据基于2013年4月17日收集的数据。

**图10 Application’s storage volume and IO velocity**

存储容量：文件虚拟机具有最高的容量，其次是数据库虚拟机 - 查看各自的第50百分位的相对值。邮件，打印，网页和应用程序具有相似的存储容量，但打印虚拟机具有最高的差异 - 请参阅箱线的高度。

体积：丰满度与分配的存储容量略有不同。文件虚拟机也是最丰富的，因此它们存储最大的数据量。具有第二高分配容量的数据库虚拟机现在是最少的，这意味着大量的可用空间。就相同应用程序类型的虚拟机的完整性的可变性而言，打印虚拟机仍然具有非常不同的存储丰满度。

每周流失率：数据库虚拟机的每周流失率最高，一些虚拟机的流失率大于120％，这意味着频繁更新，大量存储卷被删除和回收。由于跟踪数据的粗糙性，我们不能确定这是否是由于用于大型查询的tmp空间造成的，尽管这是一个可能的解释。这样的观察与DB的低饱满度齐头并进。基于第50百分位的值，打印虚拟机具有第二高的客户流失率，因为打印虚拟机存储了许多临时文件，这些临时文件在打印作业完成后被删除。由于内容的动态性，应用程序和网络虚拟机的流失率也很高，即与4.3节所示的17.9％的平均流失率类似。

IO Velocity：应用第5节中概述的热/冷VM的特性，文件VM具有最高的IO速度（以GB / h衡量）并不奇怪。根据第50百分位，邮件和数据库虚拟机具有第二和第三高的IO速度。打印，网页和应用程序虚拟机的访问速度相似。

读写比例：所有的应用程序虚拟机的读取率都低于50％，即所有的应用程序类型的读写操作都比读操作更多。事实上，正如第5节所讨论的，虚拟机的写入密集程度更高。其中，应用VM具有最低的读取比例，即低于20％。相反，打印虚拟机具有最高的读取率接近50％，这是合理的，因为打印虚拟机具有相当对称的读/写操作，即将文件写入存储并读取它们以发送到打印机。

时间可变性：为了查看每个应用的IO时间变化性，我们使用他们的一天中的CV，从24小时平均值计算。数据库和文件显示，第50百分位的时间变化率大约为1.8。由于Web虚拟机经常与具有强大时间模式的用户进行交互，因此Web虚拟机显示的时间可变性与文件和数据库虚拟机一样高。邮件，打印和应用程序虚拟机的CV值略高于1，即IO活动在一天中分散开来。

总之，文件虚拟机具有最高的容量，速度和IO负载变化性，但每周流失率相当低，在10％左右。数据库虚拟机具有高容量，高可靠性，IO负载可变性和流失率，但具有非常低的丰满度。邮件虚拟机音量适中，一天中的速度均匀。所有应用程序虚拟机都是写入密集型

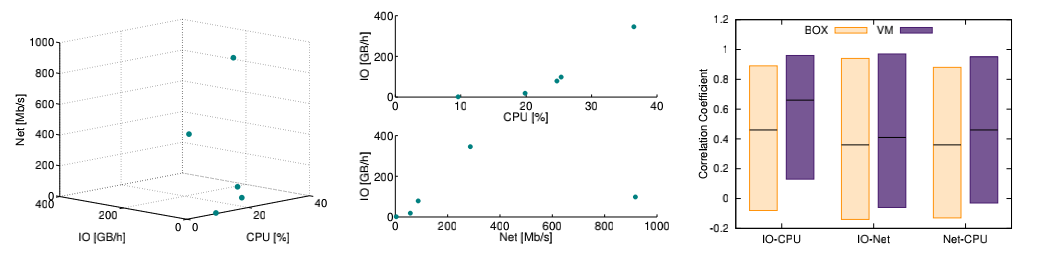
7 CPU和网络的相互依赖性

由于这里给出的统计分析是基于虚拟机和盒子的角度，所以可以将存储工作负载与其他资源（特别是CPU和网络）的工作负载相比较。使用2013年4月17日的小时平均值，我们捕获了VM IO活动对CPU利用率和网络流量的依赖关系，以兆比特每秒（Mb / s）为单位。我们关注以下两个问题：（1）IO，CPU和网络使用的最具代表性的模式是什么; （2）这三种资源之间的依赖程度如何？对于第一个问题，我们使用K-means聚类来查找具有代表性的VM工作负载。对于第二个问题，我们使用每个虚拟机的IO，CPU和网络的相关系数，然后求和它们的分布。

7.1 代表性的VM工作负载

当通过三维散点图呈现虚拟机的日常平均IO，CPU和网络时，大约有90,000个VM点。由于不可避免的过度绘图，没有明显的模式可以通过视觉识别来识别。为了识别代表性的虚拟机工作负载，我们采用K均值聚类。由于缺乏有关虚拟机群集数量的先验知识，我们首先将目标群集数量从3个变为20个，以观察群集数量增加的聚类趋势。我们的结果表明，集群质心的整体轨迹在不同数量的集群上是一致的。在图11（a）中，我们给出了5个簇的质心。当簇号增加超过5时，在前两个最低质心之间的线上出现更多的质心。

为了采取以IO为中心的观点，我们通过查看IO-CPU和IO-网络平面上的VM质心预测来分析代表性的VM工作负载，请参见图11（b）。在查看IO-CPU平面时，我们看到IO工作负载随着CPU利用率以指数方式增加。具有最高IO（大约342GB / h），即最右边的VM质心具有最高的CPU利用率（约36％）。在IO网络平面的趋势不太清楚。可以看到，前四个虚拟机质心大致位于网络流量与IO速度相同的速度上。但是，最高网络流量（大约917Mb / s）的最后一个VM质心的IO活动相对较低（约97GB / h）。总的来说，大多数具有代表性的虚拟机的IO工作负载随着CPU负载和网络流量而增加，而IO密集型虚拟机倾向于大量占用CPU而不是网络。

**(a)VM workload centroids (b)IO-CPU and IO-Net projections (c)Correlation coefficients**

**图11 Dependency among IO [GB/h], CPU[%], Network [Mb/s]**

7.1.1相关系数

在图11（c）中，我们给出IO-CPU，IO网络和CPU网络相关系数的每百分位10，50和90。为了计算上述三对的相关系数，对于每个VM /盒，我们使用三个时间序列的24小时平均值：IO GB / h，CPU利用率和网络流量。

在所有三对中，IO-CPU显示的是最高的相关系数，特别是对于虚拟机。虚拟机和盒子的IO-CPU相关系数的百分之五十分别为0.65和0.45。这表明IO活动紧跟CPU活动。这样的观察与聚类结果是一致的。箱子的相关系数略低于VM的相关系数。事实上，有一定比例的盒子和虚拟机表现出负面的依赖性，这在IO和网络之间更为显着。对于网络-CPU对来说，虚拟机和盒子对资源的需求大致相似，相关系数值大都在零以上。

8 结论

我们对在私有云范式下运行的虚拟化生产数据中心进行了大规模的研究。我们分析了三年内90,000个虚拟机的活动情况，托管在8000个物理盒子上，并且包含超过22 PB的活动存储。 IO和存储活动是从三个角度来报告的：容量，速度和多样性，即我们从整体上看整个系统，但也看单个应用程序。这个工作量表征研究与其他的不同之处在于从观察长度和追踪系统的数量两方面来看。然而，尽管我们的一些发现证实了那些小规模研究的报道，但另一些研究则提供了不同的规范。总体而言，虚拟化的程度被认为是性能感知的一个重要因素，对于每个应用程序的存储需求和需求也是如此，指向了关注虚拟化数据中心更好的资源管理的方向。

9 参考文献

1. Big Data Drives Big Demand for Storage http: // www. idc. com/ getdoc. jsp? containerId= prUS24069113 . 2013.
2. AGRAWAL, N., BOLOSKY, W., DOUCEUR, J., AND LORCH, J. A five-year study of file-system metadata. In FAST (2007), pp. 3–3.
3. BIRKE, R., CHEN, L. Y., AND MINKENBERG, C. A datacenter network tale from a server’s perspective. In IEEE IWQoS (2012), pp. 1–10.

[4] BIRKE, R., CHEN, L. Y., AND SMIRNI, E. Data centers in the cloud: A large scale performance study. In IEEE CLOUD (2012), pp. 336–343.

[5] BIRKE, R., PODZIMEK, A., CHEN, L. Y., AND SMIRNI, E. State-of-the-practice in data center virtualization: Toward a better understanding of vm usage. In IEEE/IFIP DSN (2013), pp. 1–12.

[6] DUBNICKI, C., GRYZ, L., HELDT, L., KACZMARCZYK, M., KILIAN, W., STRZELCZAK, P., SZCZEPKOWSKI, J., UNGUREANU, C., AND WELNICKI, M. Hydrastor: A scalable secondary storage. In FAST (2009), pp. 197–210.

[7] ELLARD, D., LEDLIE, J., MALKANI, P., AND SELTZER, M. Passive NFS tracing of email and research workloads. In FAST (2003). EVANS, C. Dedicated VM storage emerges to meet virtualisation demands. http: // www. computerweekly. com/ feature/ Dedicated-VM-storage-emerges-to-meetvirtualisation-demands. 2013.

[9] GRIBBLE, S. D., MANKU, G. S., ROSELLI, D. S., BREWER, E. A., GIBSON, T. J., AND MILLER, E. L. Self-similarity in file systems. In SIGMETRICS (1998), pp. 141–150.

[10] GULATI, A., SHANMUGANATHAN, G., AHMAD, I., WALDSPURGER, C. A., AND UYSAL, M. Pesto: online storage performance management in virtualized datacenters. In SoCC (2011), p. 19.

[11] GULATIAND, A., SHANMUGANATHAN, G., ZHANG, X., AND VARMAN, P. Demand based hierarchical qos using storage resource pools. In USENIX ATC (2012), pp. 1–14.

[12] GUNDA, P. K., RAVINDRANATH, L., THEKKATH, C. A., YU, Y., AND ZHUANG, L. Nectar: Automatic management of data and computation in datacenters. In OSDI (2010), pp. 75–88.

1. HANSEN, J. G., AND JUL, E. Lithium: virtual machine storage for the cloud. In SoCC (2010), pp. 15–26.
2. J.DOUCEUR, AND BOLOSKY, W. A large-scale study of file-system contents. In SIGMETRICS (1999), pp. 59–70.
3. KOLLER, R., AND RANGASWAMI, R. I/O deduplication: utilizing content similarity to improve I/O performance. In FAST 2010, pp. 16–16.
4. LE, D., HUANG, H., AND WANG, H. Characterizing datasets for data deduplication in backup applications. In FAST (2012), pp. 1–10.
5. LEUNG, A. W., PASUPATHY, S., GOODSON, G. R., AND MILLER, E. L. Measurement and analysis of large-scale network file system workloads. In USENIX ATC (2008), pp. 213–226.
6. LI, M., GAONKAR, S., BUTT, A. R., KENCHAMMANA, D., AND VORUGANTI, K. Cooperative storage-level de-duplication for I/O reduction in virtualized data centers. In MASCOTS (2012), pp. 209–218.
7. MERRILL, D. R. Storage economics: Four principles for reducing total cost of ownership. http: // www. hds. com/ assets/ pdf/ four\_ principles\_ for\_ reducing\_ total\_ cost\_ of\_ ownership. pdf . 2009.
8. MEYER, D., AND BOLOSKY, W. A study of practical deduplication. In FAST (2011), pp. 1–1. [21] NARAYANAN, D., DONNELLY, A., THERESKA, E., ELNIKETY, S., AND ROWSTRON, A. I. T. Everest: Scaling down peak loads through i/o offloading. In OSDI (2008), pp. 15–28.
9. NG, C.-H., MA, M., WONG, T.-Y., LEE, P. P. C., AND LUI, J. C. S. Live deduplication storage of virtual machine images in an open-source cloud. In Middleware (2011), pp. 81–100.
10. PARK, N., AHMAD, I., AND LILJA, D. J. Romano: autonomous storage management using performance prediction in multi-tenant datacenters. In SoCC (2012), p. 21.
11. PARK, N., AND LILJA, D. J. Characterizing datasets for data deduplication in backup applications. In IISWC (2010), pp. 1–10.
12. ROSELLI, D., LORCH, J., AND ANDERSON, T. A comparison of file system workloads. In USENIX ATC (2000), pp. 41–54.
13. ROSS, S. A First Course in Probability. 2004.
14. 27] SEHGAL, P., TARASOV, V., AND ZADOK, E. Evaluating performance and energy in file system server workloads. In FAST (2010), pp. 253–266.

[28] SHAMMA, M., MEYER, D., WIRES, J., IVANOVA, M., HUTCHINSON, N., AND WARFIELD, A. Capo: Recapitulating storage for virtual desktops. In FAST (2011), pp. 31–45.

[29] SIMPSON, N. Building a data center cost model. http: // www. burtongroup. com/ Research/ DocumentList. aspx? cid= 49 . 2009.

[30] VOGELS, W. File system usage in windows nt 4.0. SIGOPS Oper. Syst. Rev. 33, 5 (Dec. 1999), 93– 109.

[31] WALLACE, G., DOUGLIS, F., QIAN, H., SHILANE, P., SMALDONE, S., MARK, M. C., AND HSU, W. Characteristics of backup workloads in production systems. In FAST (2012), pp. 4–4.