

联想凌拓

小芯片 大动能

专为半导体集成电路EDA打造的可靠高效存储方案



存储角度看半导体行业

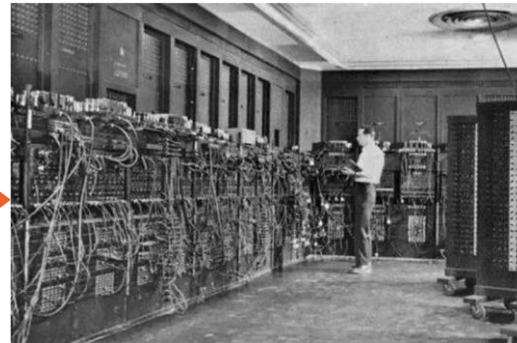
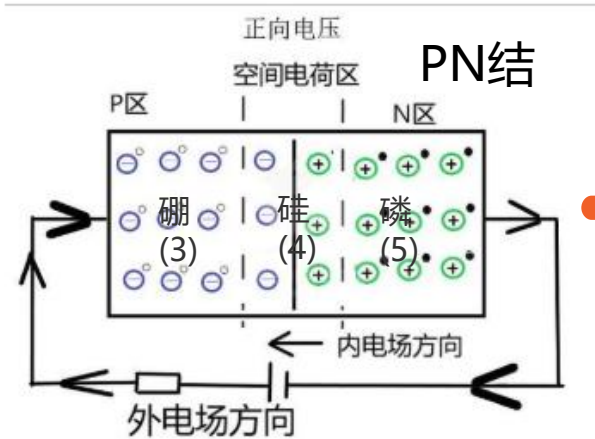
联想凌拓EDA解决方案

成功案例分享

半导体行业已在政策支持和终端市场需求强劲的双重动力推动下，**半导体产业链实现了持续快速增长和重要支柱行业之一。**

在芯片设计的模拟过程，**通过网络共享存储(NAS)结构，生成巨量的混合IO、写入和读取海量文件。**

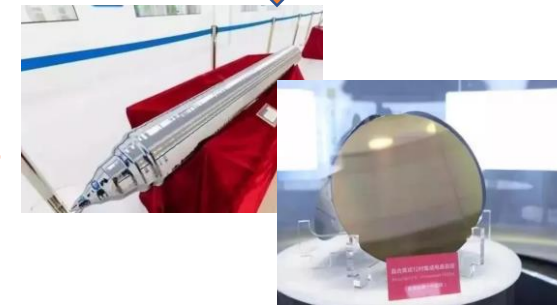
PN结， 半导体芯片的原理



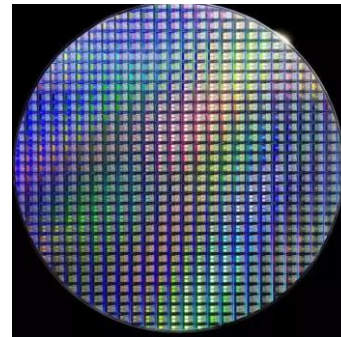
18000只电子管，50万条线，重达30吨，运算能力5000次/秒，人类第一台计算机，



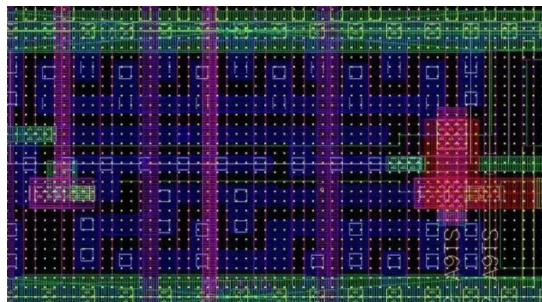
砂子，电子级高纯硅要求99.99999999%



硅提纯旋转，切片后的硅片也是圆的，因此就叫“晶圆”，晶圆上把成千上万的电路装起来的，干这活的就叫“晶圆厂”



在晶圆上涂一层感光材料，非常精准的光线刻出很多沟槽，在沟槽里掺入磷/硼元素，就得到了一堆N型半导体
晶圆上的小方块就是芯片。



芯片放大了看就是成堆成堆的非门电路，更多的器件，组成了更庞大的电路，运算性能自然就提高了

芯片设计与制造的难点及关注点

用数以亿计的器件组成如此庞大的电路，稍微有个PN结出问题，电子在整个电路就会出现大量的故障。
这种精巧的线路设计，只有一种办法可以检验，那就是：**验证！验证！大量大量的验证！**

生产投入成本极高，一款14nm芯片约2亿元~3亿元，研发周期约1~2年。一条14nm工艺生产线高达100亿美元。

风险点：集成电路设计存在技术和市场两方面的不确定性

芯片产业结构向高度专业化转化,形成了三大产业独立成行的局面:
- 芯片设计 - 芯片制造 - 封装测试
芯片行业目标: - **缩减上市时间(Faster time to market)**, 加速设计、模拟、验证流程,降低工程耗时

芯片与硅周期，目标实现行业跨越式发展



- 接触最多的半导体是来自于CPU、DRAM内存、NAND闪存、以及各种电子控制芯片，这些芯片都是基于单晶硅加工而成。

1.功能芯片

- 桌面级CPU主要是Intel和AMD两家瓜分了全部市场
- 移动处理器，ARM移动处理器主要厂家：高通、联发科、三星、海思、展锐等

2.存储芯片


- 闪存芯片按照类型分为两种，**DRAM颗粒**和**Nand闪存**。
- 3D XPoint的工作原理与NAND，DRAM存在着根本性的不同。

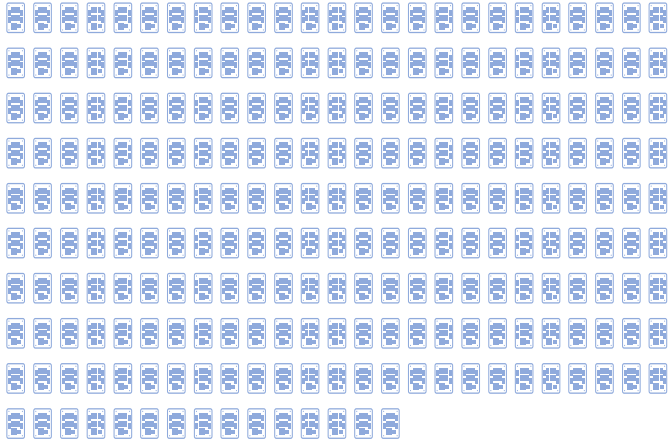
- 硅芯片周期的发展

- 如经济周期的发展，一般**10年为一次硅周期**
- 2030年中国集成电路产业链主要环节达到国际先进水平
- 下游产业（手机应用，物联网、区块链、汽车电子、5G、AR / VR及AI）

关注SSD固态硬盘存在半导体行业的使用

- Metrics through September 2019

 **5,300**
total petabytes
of flash shipped



49,000+
all-flash controllers


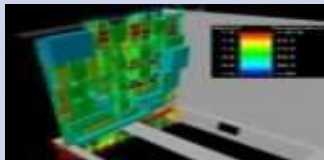

Figure 1. Magic Quadrant for Primary Storage



As of August 2019 © Gartner, Inc
Source: Gartner (September 2019)

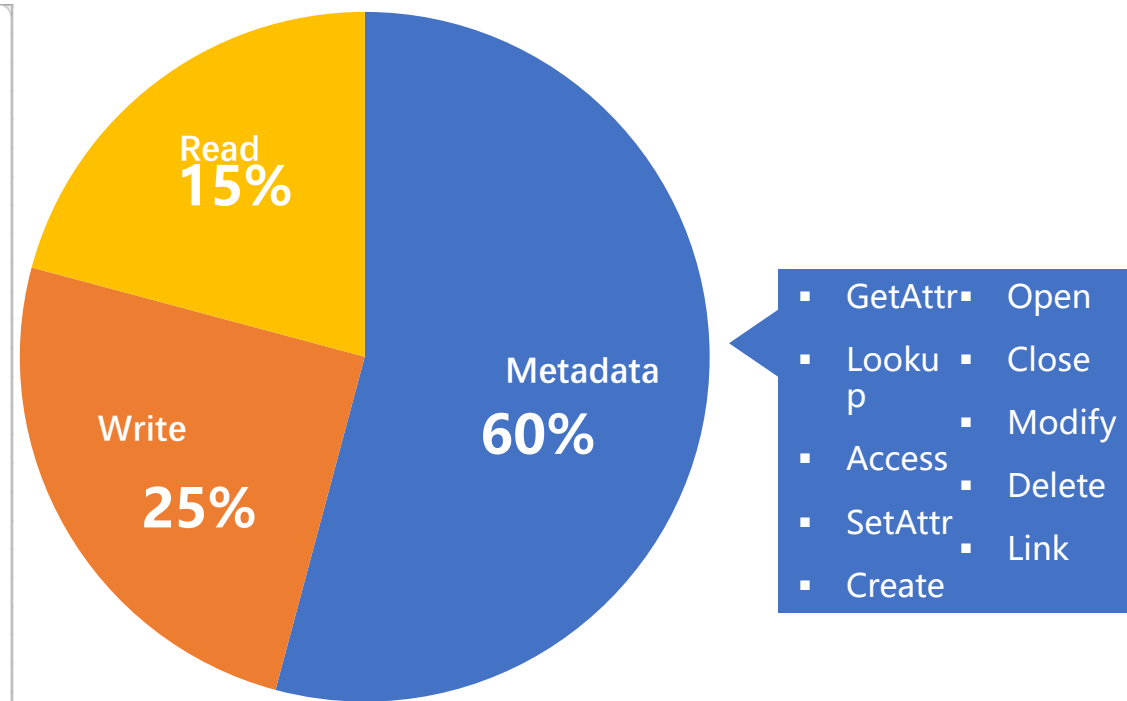
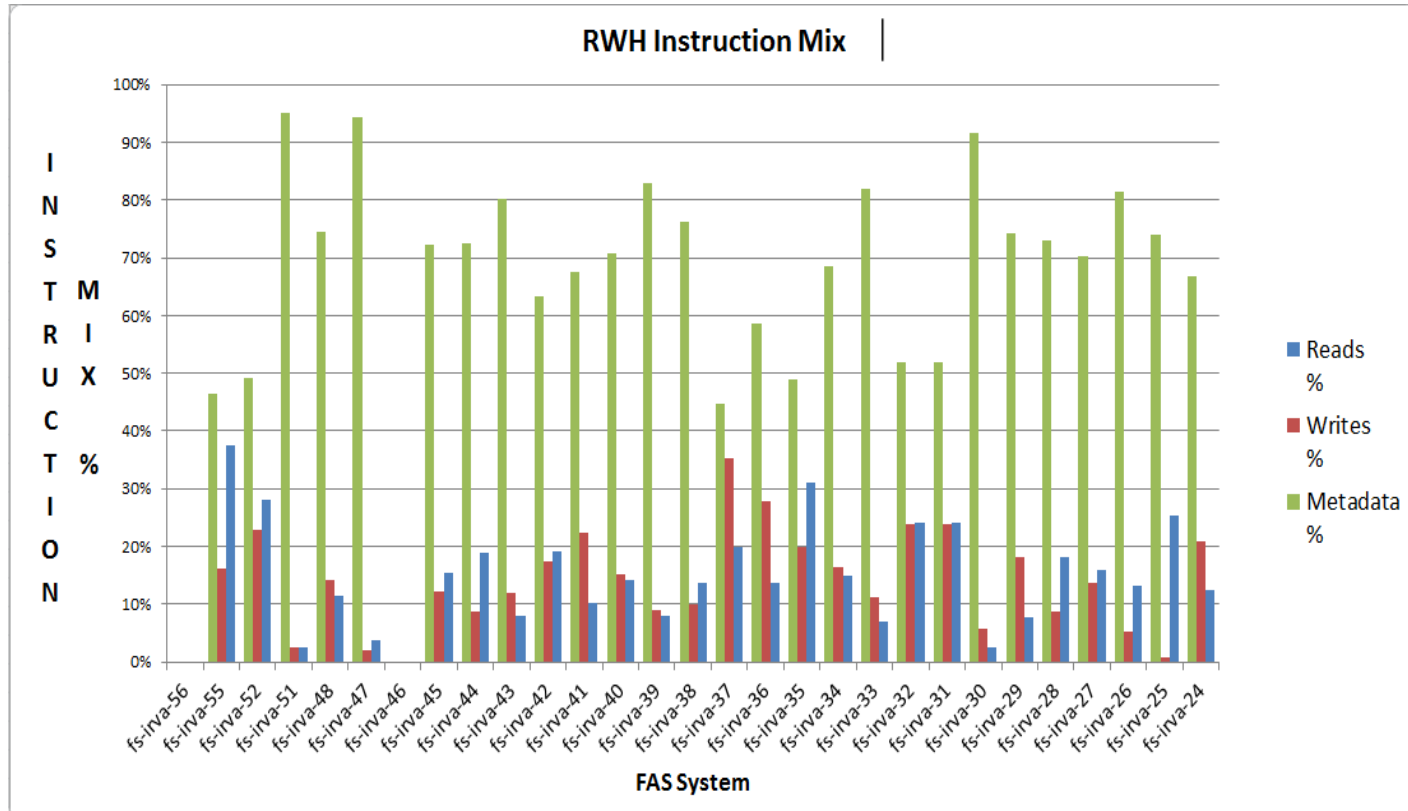
从存储的角度了解EDA不同工作阶段的应用特点

随着尺寸越来越小，更多时间花费在逻辑和物理验证阶段

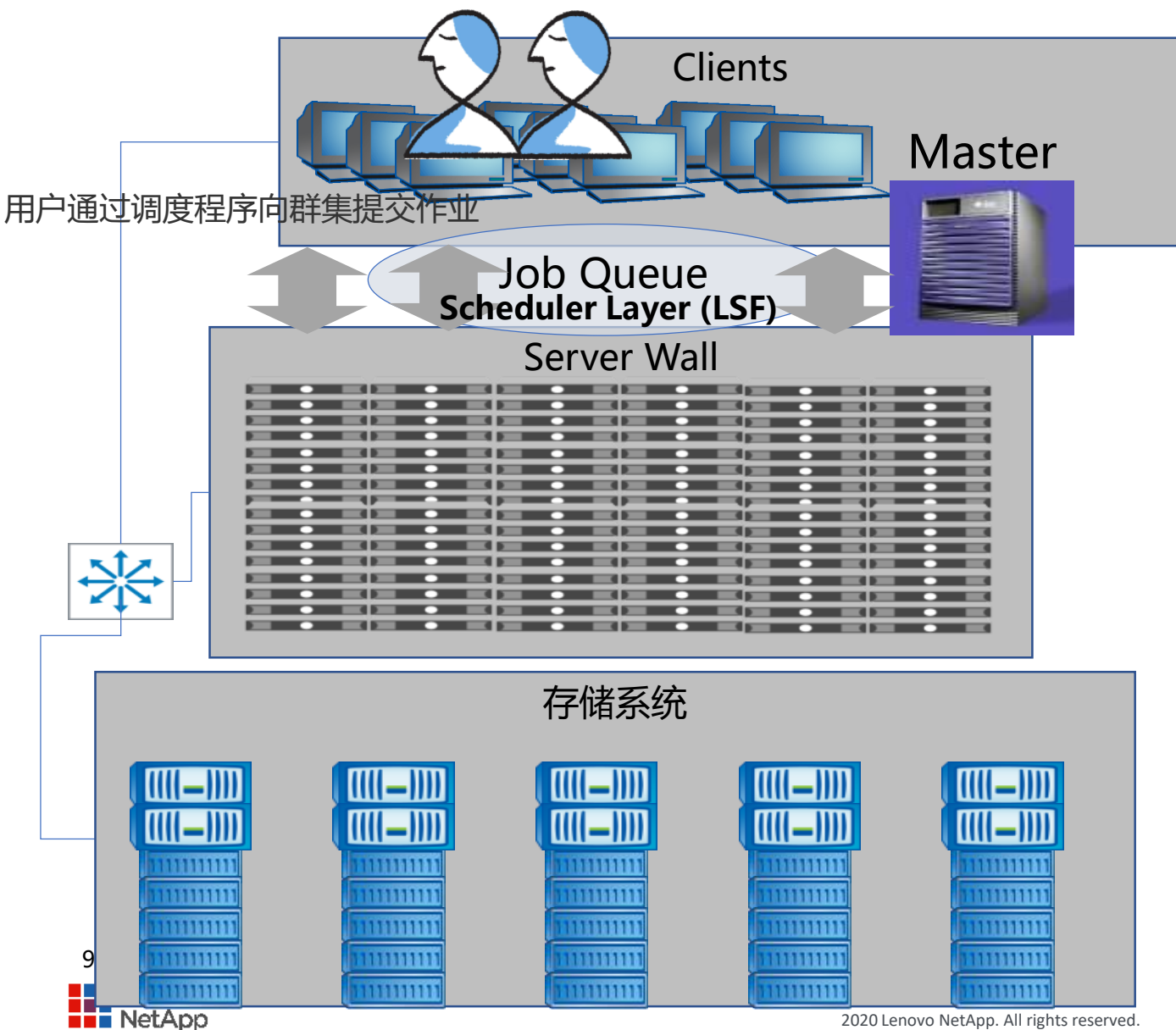
	Logic Design	Physical Design	TapeOut
			
工作应用	Logic design将源文件被读取和编译成一个可执行的芯片模型，验证逻辑设计行为，并检查是否设计相匹配	包括逻辑合成，布局 and 路由，以及各种性能和可制造性检查	在所有检查都通过后，该设计被记录到指定的铸造和工艺节点用于制造
服务器数量	500-5000	300-800	6000-20000
协助度	高度（数造工程师）	中等	低
应用类型	串行	串行	并行
IO带宽需求	600MB/s-6GB/s	3-15GB/s	25GB/s+
作业运行时间	Weeks – Months (7nm, 6~12months)	Weeks – Months (7nm, 1~2 months)	Weeks+
文件大小	MB~GB, Output file=10~50GB	300GB per layer	TB+
产生文件数量	Hundred Thousands	20~50 model, many rule files	10~30 files
数据访问方式	随机	混合 (sequential/random)	混合 (sequential/random)
数据访问负载类型	8% Read; 2% Write 90% Metadata(getattr, Lookup)	15% Read; 5% Write 80% Metadata(getattr, Lookup)	Many Large Read Many Large Write
缓存读取	高	高	中高

7

EDA工业应用不同阶段混合I/O的负载



EDA应用典型存储架构



- 芯片密度快速增长22nm,14nm,10nm,7nm
- 存储承担的负载快速增加,容量每两年翻倍
- Clustered servers, storage farms主流
- NFS客户端(少量CIFS)
- 存储更可能成为系统性能瓶颈
- 重要资产重复使用和数据备份恢复/容灾
- 大量计算单元并发/协同工作
- I/O通路Trunking
- 计算/存储负载不可预知 (QOS)
- 作业时序规划
- 混合工作负载 -设计 /验证/个人主目录
- 巨量元数据访问 (SSD或性能加速卡)
- Tape-Out到另外一套存储设备

存储角度看半导体行业

联想凌拓EDA解决方案

成功案例分享

国内外芯片行业巨头（无论是设计，制造和封装企业）使用NetApp（联想凌拓）存储存取构建和模拟测试设计所需的文件，以及在此阶段的生命周期内文件管理的**解决方案**。

NetApp存储解决方案在芯片设计行业的地位

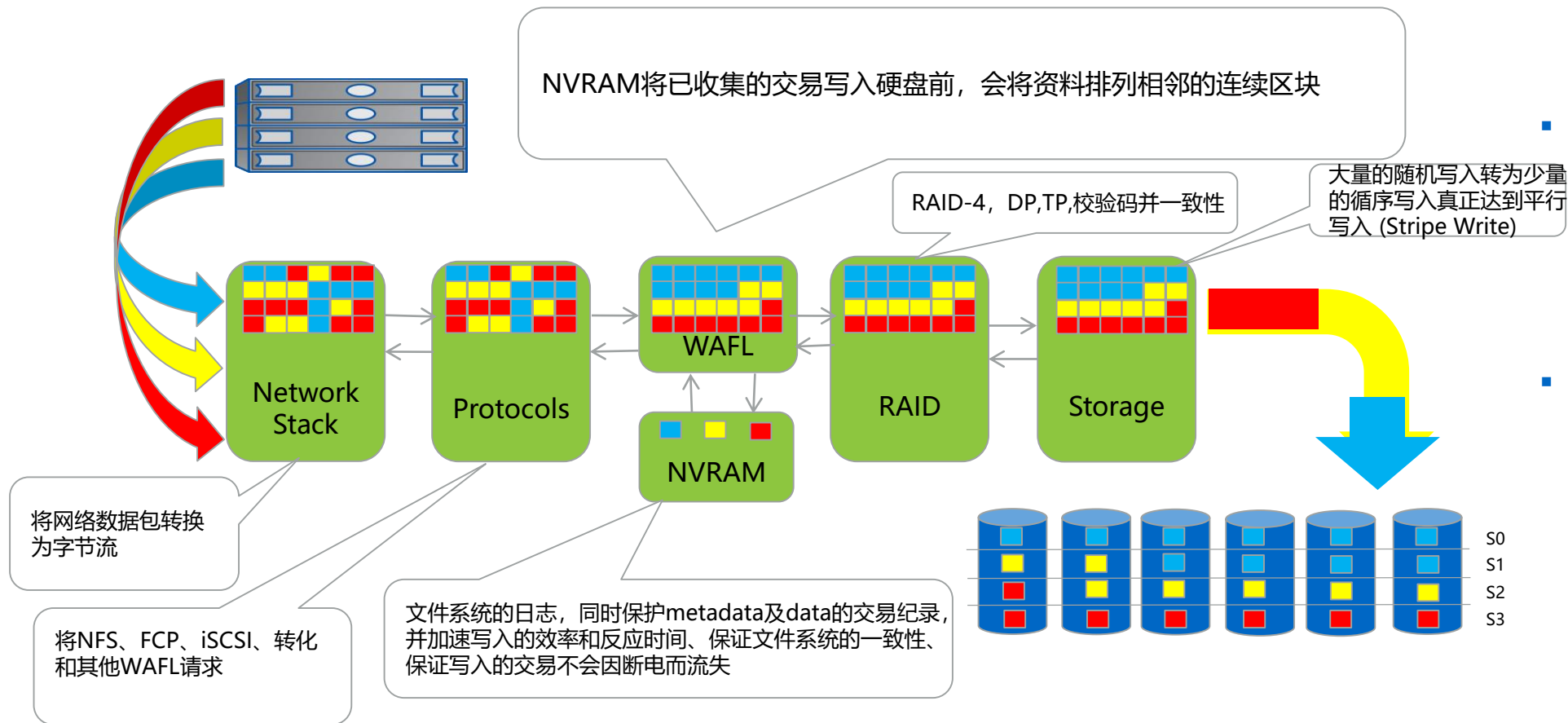


- Top 300 EDA设计企业中 65+%是NetApp用户
- 绝大多数EDA应用运行于 x86/Linux (基于NFS协议)
- 所有EDA应用软件可运行于NetApp存储
- 所有EDA软件厂商都使用NetApp产品
- 专业的NetApp团队/社区专注于EDA解决方案



NetApp WAFL原理

优化了写IO性能，文件系统一致性，SSD优化



- 结合有电池保护的NVRAM来担任WAFL的日志，并藉由一致点提供文件系统一致性
- NetApp ONTAP独特的配合NVRAM日志功能，可将大量的随机写入转为少量的循序写入相邻可用的数据块，这与SSD减少写放大原理一致，优化SSD的寿命。
- WAFL结合NVRAM、RaidDP、Snapshot的设计难度极高，故从1992年至今仍未有其它厂商可以做到。

1998年10月获得专利 (专利号码: 5,819,292) , 题目为「Method For Maintaining Consistent States of A File System and For Creating User-Accessible Read-Only Copies of A File System」 ,

NetApp WAFL天然地优化SSD功能

White Paper

Enterprise Storage: The Foundation for Application and Data Availability

Sponsored by: NetApp
Eric Burgener
October 2018

Gartner魔力象限第一的全闪存存储

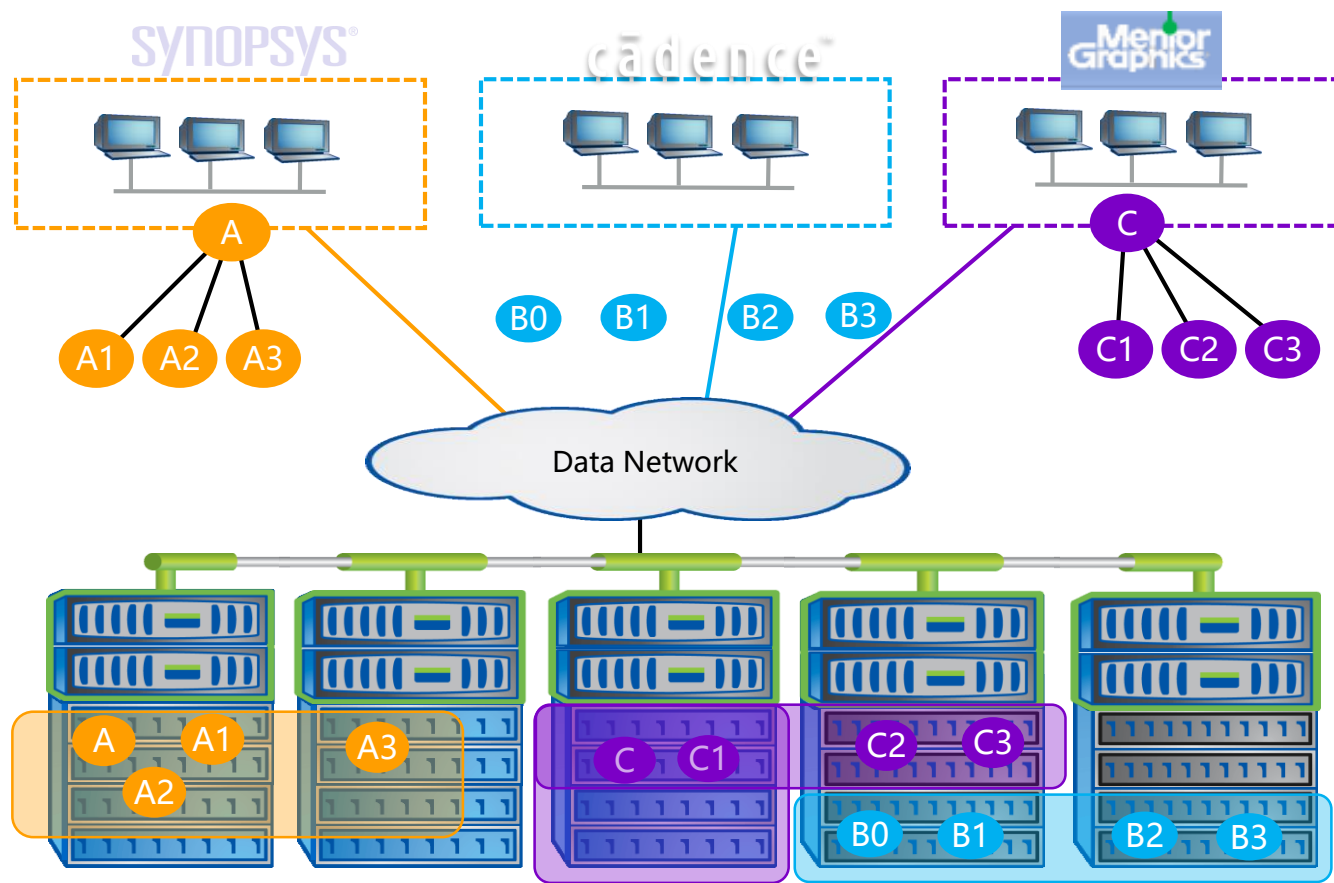
IDC报告：WAFL管理磁盘上数据的布局，检测和纠正存储错误，优化性能，并启用ONTAP的许多独特功能，包括快照拷贝、克隆和存储效率。WAFL的“随时随地写入”设计特别针对闪存的高性能读写进行了优化。第8页

一块普通SSD的组成：NAND，SSD主控制器，辅助芯片



SSD功能特点:	ONTAP是否支持?
内存中合并写	支持 (since 1992)
条带写以磨损均衡	支持 (since 1992)
避免覆盖写以减少写放大	支持 (since 1992)
无损快照	支持 (since 1992)
在线压缩	支持 (since 2015)
在线重复数据删除	支持 (since 2015)
读写性能优化	支持 (since 2014)
写单元匹配SSD页	支持 (since 2015)
高效克隆	支持 (since 2006)

实现EDA平台性能及容量的横向扩展



- 从入门级起步，横向扩展节点
- 统一集群命名空间建立和验证/模拟
- 各应用单元间性能和容量无缝扩展，并保证安全隔离
- 对关键应用实现性能和容量的负荷调整 (Vol Move)
- 支持不同型号
- 支持全闪存阵列

对所有EDA用户和应用均透明实现

FlexGroup 高性能的NAS空间

- 由FlexGroup將资料分配到各个volume中
- 性能及容量可线上横向扩展

460* Over 193PB

Customers

Early Validation Program

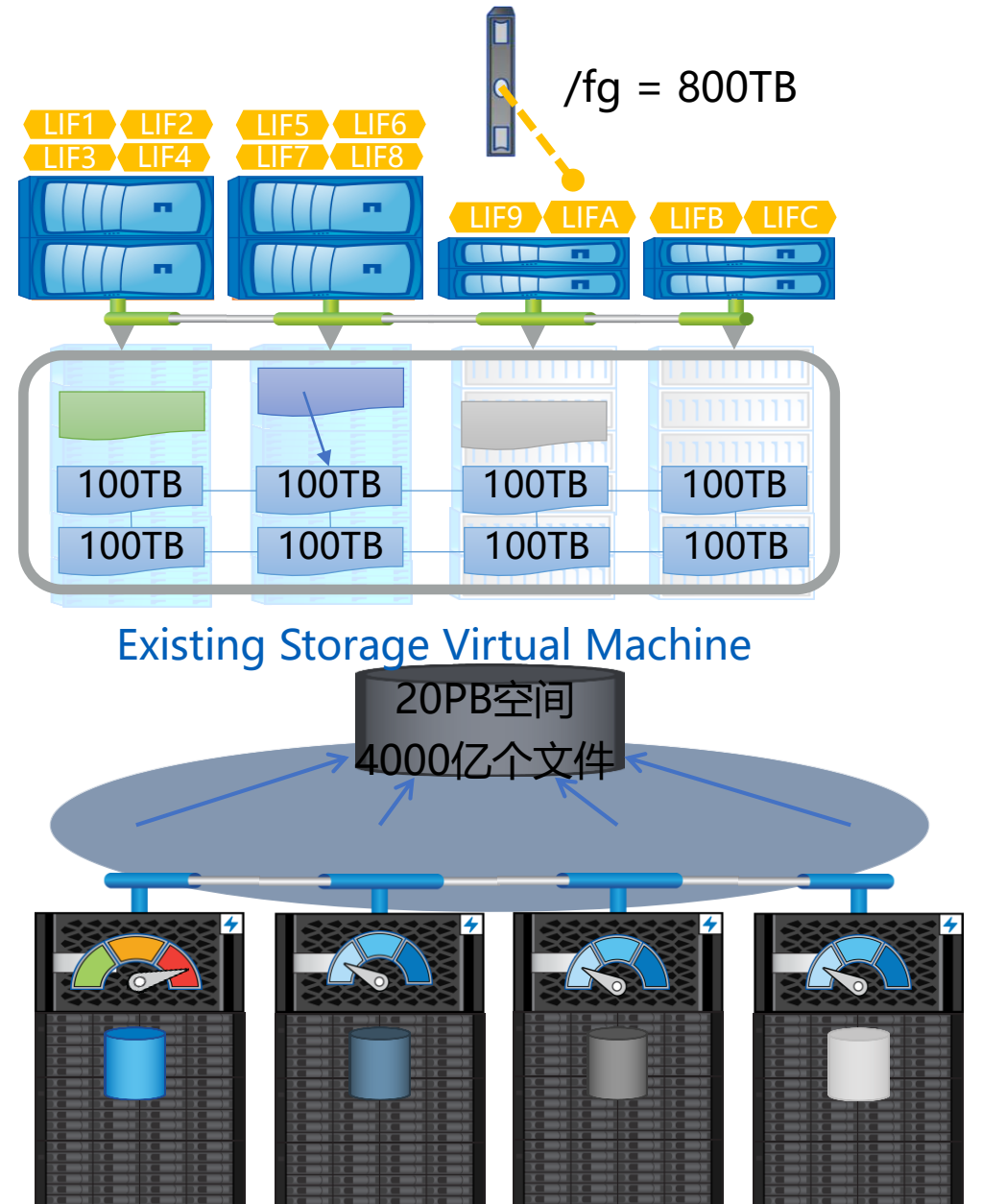
TSMC
Globalfoundries
MediaTek
Thomson-Reuters (*)
NVIDIA
Cisco (Build) (*)
Synopsys
Intel

15

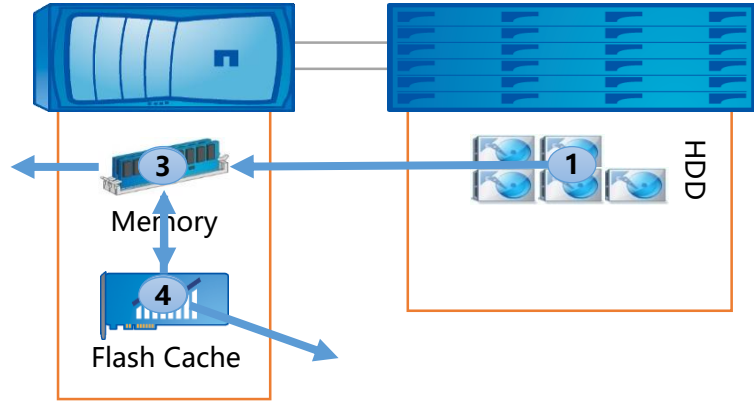


Deployed on FlexGroup

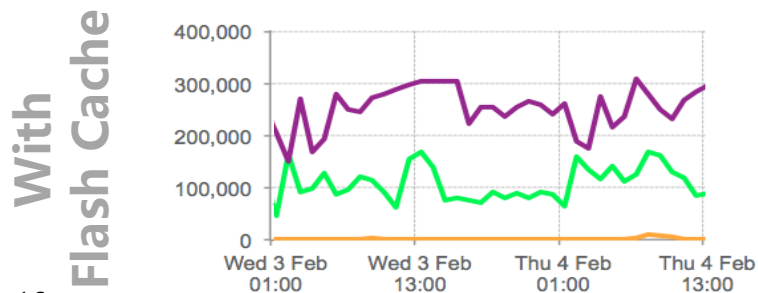
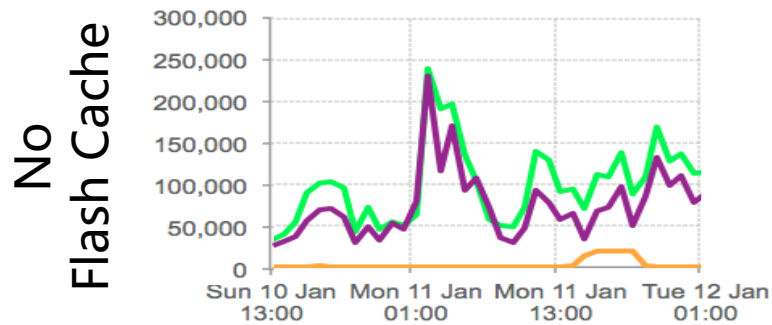
2020 Lenovo NetApp. All rights reserved.



FlashCache 加速磁盘性能，特别加速EDA应用metadata性能



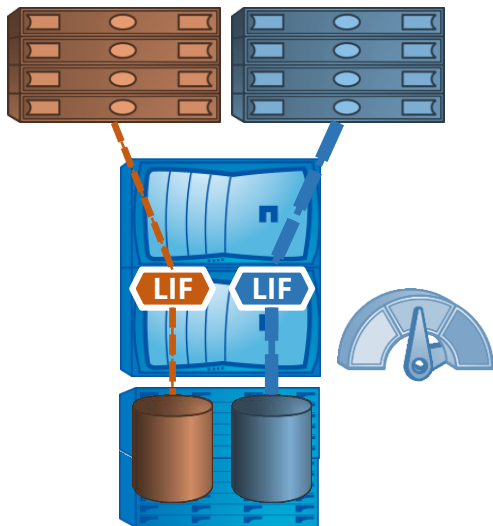
- Flash Cache 可作为主内存或基本内存的扩展缓冲区。
- Flash Cache 可以非常有效地为元数据和读取工作负载实现低延迟。
- 由于 EDA 验证工作负载包含数百万个小文件、大量的元数据并需要进行大量读取，FlashCache 可以为其缓存元数据和数据，从而加快 I/O 请求处理速度，并减少后端磁盘访问次数。
- FlashCache已经获得我们大多数 EDA 客户的一致认可
- FAS(DM)存储默认自带NVMe卡做FlashCache加速



16

服务质量QoS控制—大型EDA用户必须的功能

- 控QoS workload 包括I/O操作和数据吞吐量，分别以IOPS和MBps为单位来度量
 - Storage virtual machines (SVMs)
 - FlexVol® volumes
 - LUNs
 - Files
- 控制抢占资源的工作负载
- 管理不同租户不同应用要求
- 实时调整



Client

```
a700s::statistics top*> client show -max 100
```

```
a700s : 4/17/2019 14:28:25
```

```
*Estimated
```

Total IOPS	Protocol	Node	vserver	Client
10858	nfs	a700s-02	svm_nfs	192.168.200.204
8841	nfs	a700s-02	svm_nfs	192.168.100.204
7644	nfs	a700s-01	svm_nfs	192.168.200.202
7600	nfs	a700s-01	svm_nfs	192.168.200.200
7541	nfs	a700s-02	svm_nfs	192.168.100.206
7541	nfs	a700s-01	svm_nfs	192.168.100.200

File

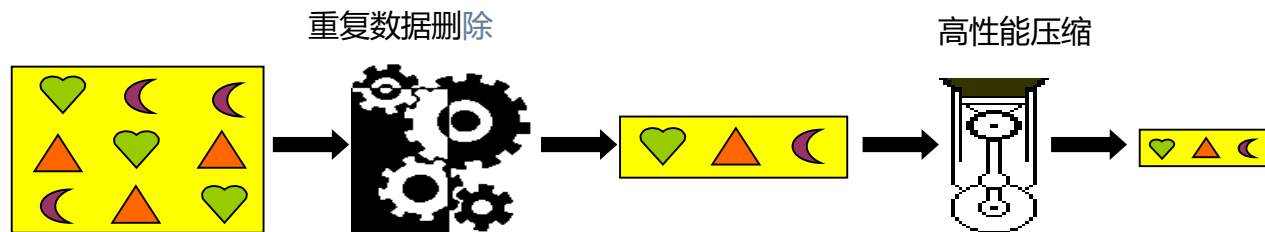
```
a700s::statistics top*> file show -max 100
```

```
a700s : 4/17/2019 14:30:09
```

```
*Estimated
```

Total IOPS	Node	vserver	Volume	File
924	a700s-01	svm_nfs	fg1	/mp2-1m8/vdb.1_1.dir/vdb_f8451.file
924	a700s-01	svm_nfs	fg1	/mp2-1m8/vdb.1_1.dir/vdb_f8403.file
924	a700s-01	svm_nfs	fg1	/mp2-1m8/vdb.1_1.dir/vdb_f8341.file
924	a700s-01	svm_nfs	fg1	/mp2-1m8/vdb.1_1.dir/vdb_f8072.file
924	a700s-01	svm_nfs	fg1	/mp2-1m8/vdb.1_1.dir/vdb_f8036.file

EDA应用主存储上的存储效率数据及容灾备份



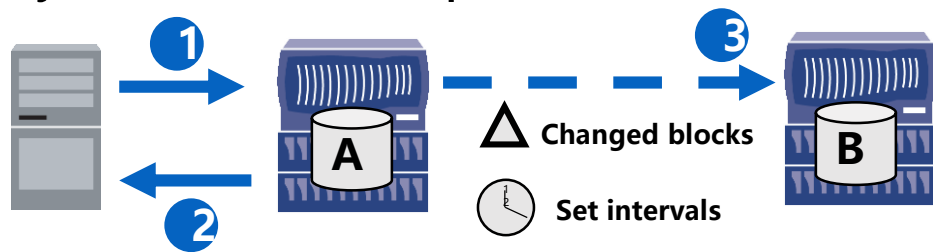
TR-4476 NetApp Data Compression and Deduplication, Page6

全闪存默认重删除/压缩都是打开的，按文档所讲，只有**1%到2%**的性能下降，但获得的全闪存效率收益巨大。

TR-4617, Electronic Design Automation Best Practices, Page20

由于EDA工作负载和文件的性质，ONTAP通过使用重复数据消除可以提供高达20%的效率，通过使用压缩可以提供高达30%的效率，**总共节省了高达50%的空间。**

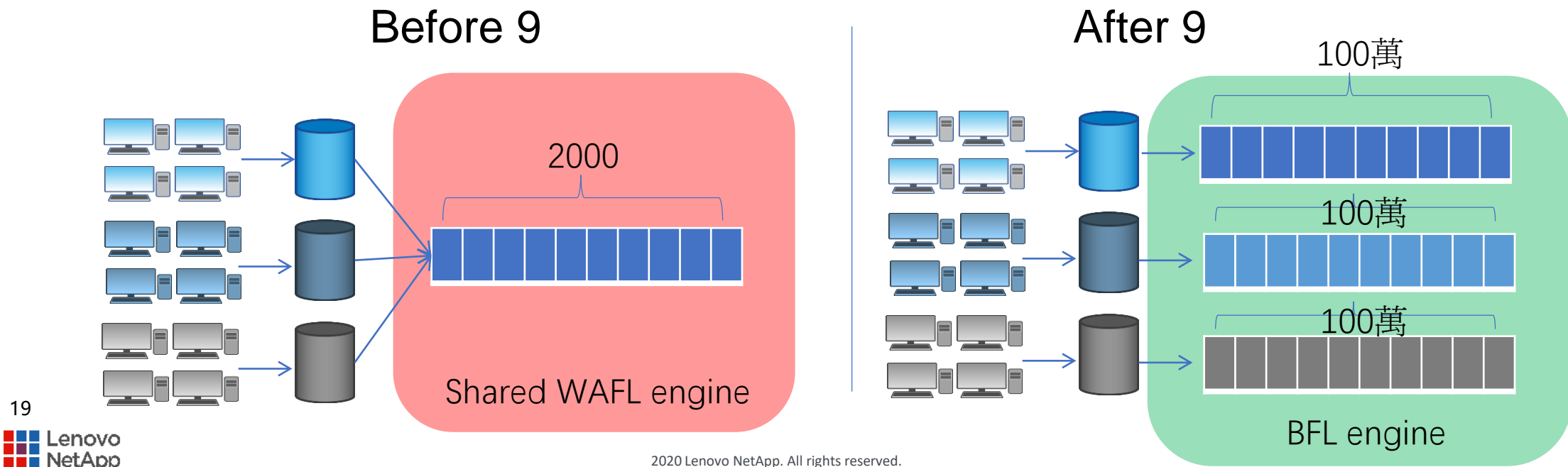
Asynchronous SnapMirror



- 对主存储进行重复数据删除/数据压缩
- 单一数据保护流和存储库节省带宽和存储，变化的4K数据块输送
- 实现磁盘到磁盘备份/还原和灾难恢复
- 在主站点损坏时进行时间点恢复

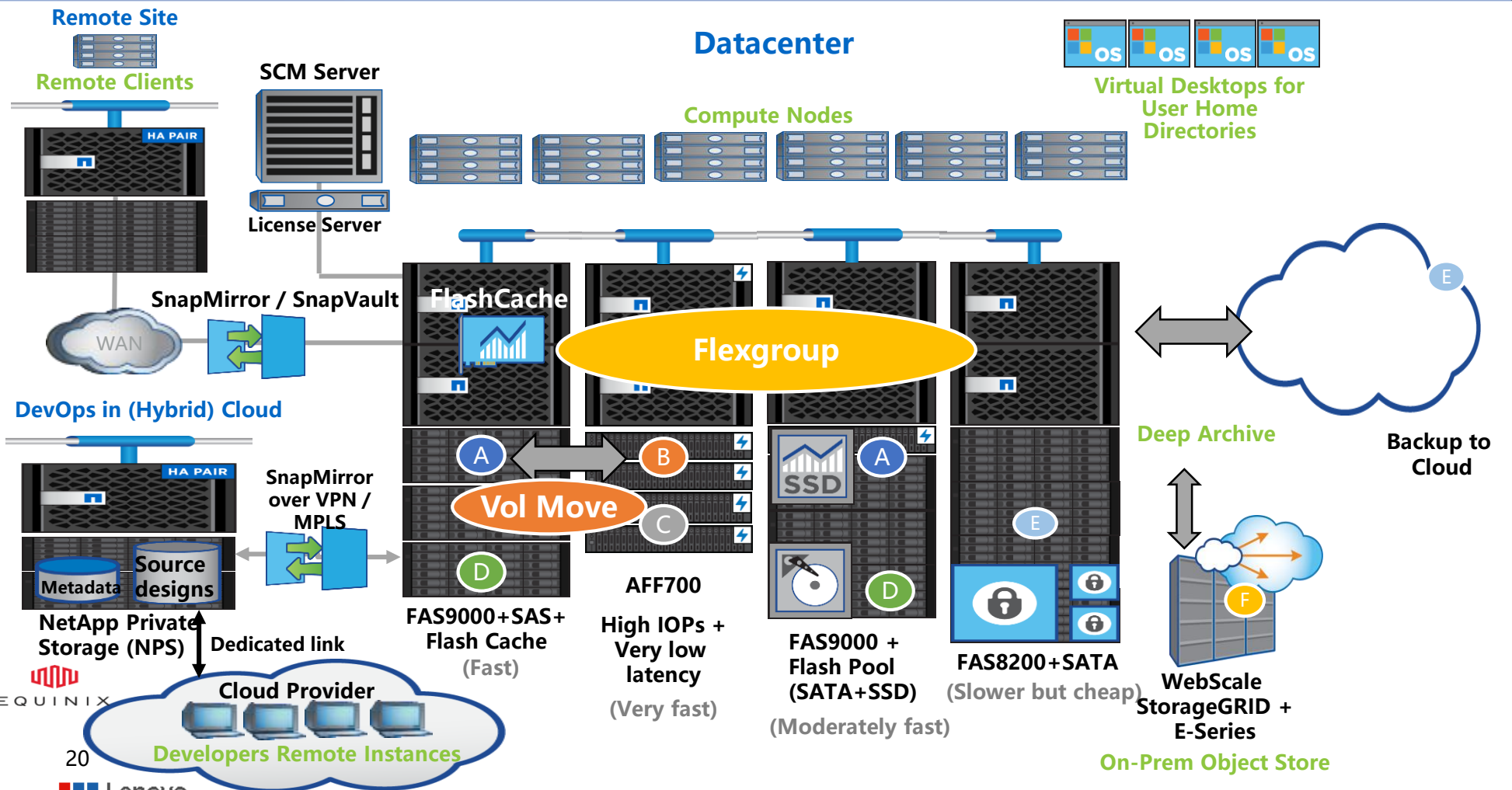
Delete速度是目前EDA环境的限制

- EDA业务特性
 - 开始会有大量文件读取，中间会有小文件读写与删除，最后会有大量读写与删除
 - 因为由computing farm同时存取，所以大量读写与删除会同时发生，若delete响应时间不够及时，可能导致timeout
- Delete queue的限制改为volume，而非node，Delete queue高达100萬 (size >1TB volume)



NetApp 集群存储针对EDA行业解决方案

Collaborations with the ECO system partners



- User Workspace Configuration
- VCS Simulation Acceleration
- Silicon SMART Cell Characterization
- Storage-Aware Job Scheduling
- Liberate Cell Characterization
- Incisive Verification
- Questa ModelSIM
- Calibre Tape-Out



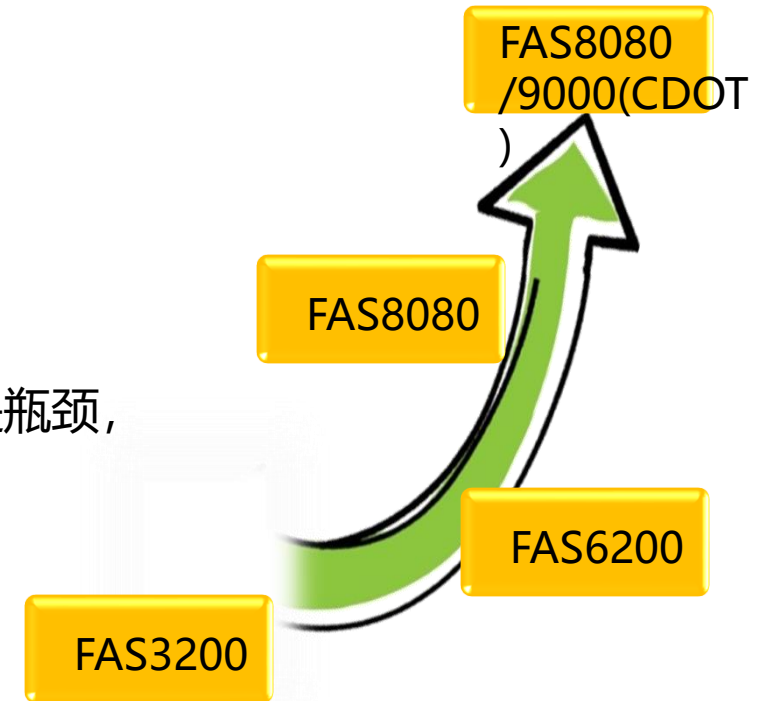
存储角度看半导体行业 联想凌拓EDA解决方案 成功案例分享

国内芯片设计行业巨头,从2006年底开始使用
NetApp(联想凌拓) 解决方案

国内案例：设备制造商-存储控制器CPU分析

High CPU utilization

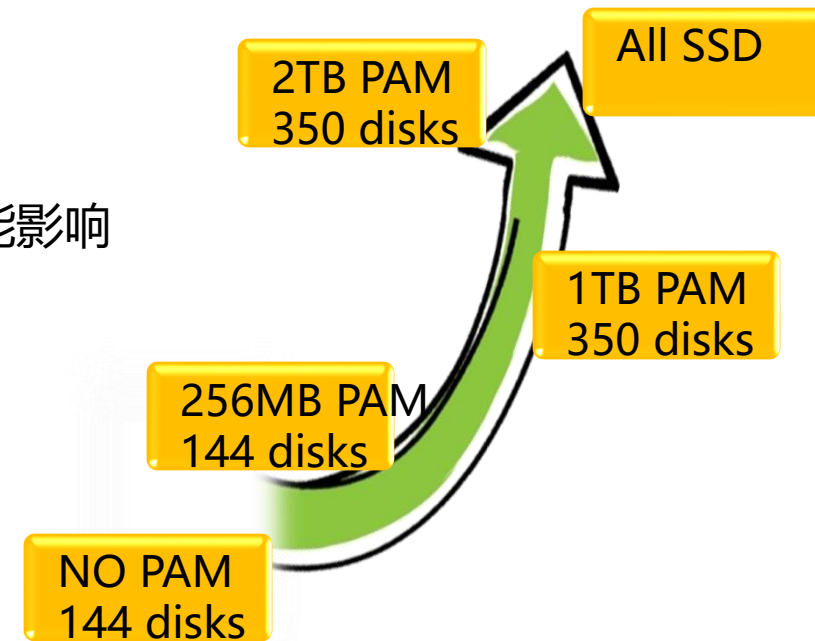
- 从20nm后，控制器是瓶颈，所以FAS3200→FAS6200最高端→到只买FAS8080最高端的阶段
- 多核平台使用较少的内核，多模式工作方式
 - * 大量 Meta-data读书
 - * 单线程读写 -High Kahuna
 - * 大量文件删除占据性能
- 20nm->14nm后，发现EDA性能爆炸性增长，单个控制器（7-mode）是瓶颈，CDOT多个控制器来做性能分担
- C-mode的QOS是性能控制的极好工具，XX用户一定需要此功能
- 现阶段，FAS9000（A700）是大型EDA企业的唯一选择。



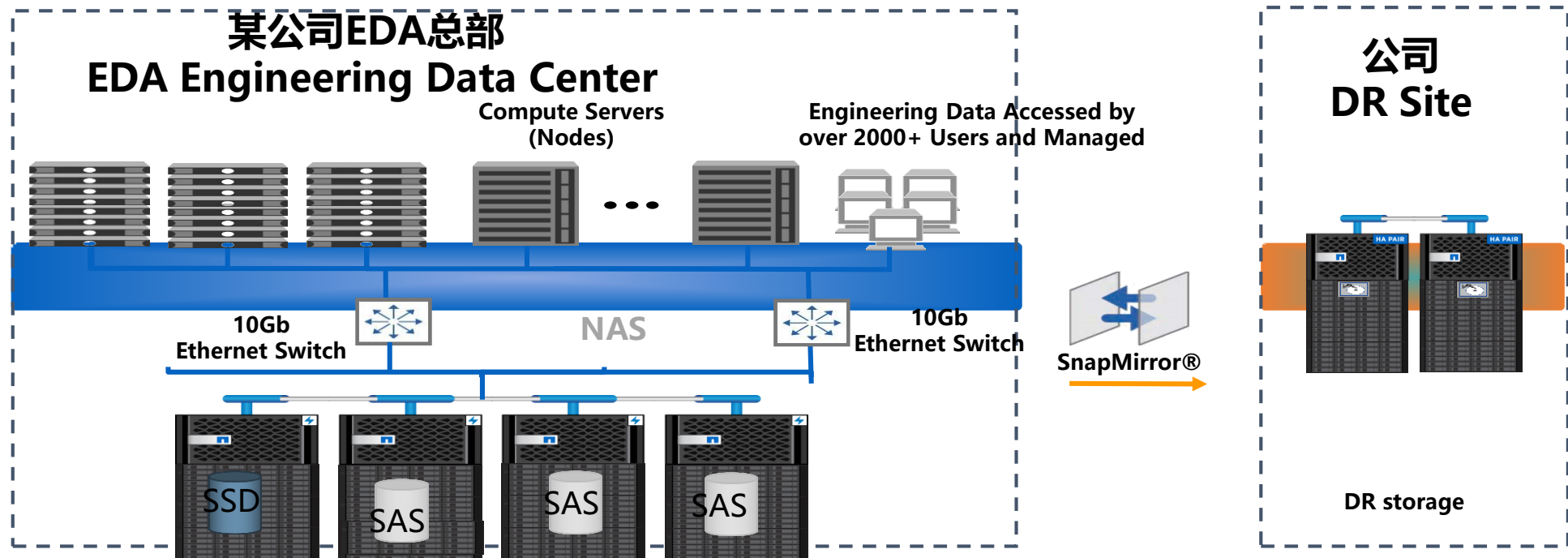
国内案例：设备制造商-存储磁盘和性能加速卡分析

High Disk Utilization

- 升级到最高端存储后 (7-mode) ,发现大量的metadata运算, 磁盘是瓶颈,购买 PAM和组建多磁盘Aggregate来解决问题
 - * 升级PAM卡: 256MB→1TB→2TB
- 增加写入磁盘的磁盘主轴数
 - * 每个控制器Aggregate 的数量为 (1 - 2 个) , Aggregate的数量对性能影响不大, 每个Aggregate的磁盘尽量多。
 - * 至少两个Raid Group, 每个Raid Group Size约22 (20 + 2)
 - * 启用CSC功能, 文件系统碎片整理
- **AFF全闪存存储是必然的选择**



国内案例：设备制造商EDA存储架构及容灾方案



注:这仅是一个研究所中配置:

1. 16个节点CDOT FAS8080/FAS90000(集群1), 2个节点 CDOT FAS8020A用于数据本地备份
2. 重要的成品数据, 通过2节点MCC (双活存储) 存放及容灾
3. QOS是设置为最大值, 在新上的项目2周观察并决定QOS值, 但QOS值会随业务变化。

WHY

EDA User Choose us?

- 强大的存储集群功能及可预知性能
- QOS功能，确保存储集群内用户性能获得保证
- 重要数据采用NAS双活解决方案
- 快照及容灾保护数据
- 能够帮助跨国、跨区域的研发团队，能够在24小时在线的系统平台上持续工作
- 长期的客户关系和优质售后服务

谢谢

智慧数据构建智能世界

