

从边缘到洞察力

面向数据和 AI 的存储 化繁为简

吴磊

IBM大中华区存储总经理

融合

分析

组织

收集

AI

解锁数据价值
以全新的方式
推动业务转型

- 预测和调整未来结果
- 让人们以更好的方式从事更高价值的工作
- 自动化决策、流程和体验
- 重构新的业务模式

发挥数据的力量，为您提供可圈可点的**竞争优势**



不过，仅有 15%
的数据价值被挖掘

数据是 AI 的第一个挑战

FORRESTER

AI 的三大瓶颈

- 技术资源
- 数据管理
- 缺乏数据

AI 之旅

融合 实现流程自动化并对其进行扩展

分析 按需扩展洞察力

组织 建立可信的分析基础

收集 让数据变得简单、可访问

无IA，不AI

企业对 AI 的需求给如今的基础架构带来了压力

收集

摄入数据

随增长扩展的能力
现代数据访问

边缘



过时

组织

对数据进行编目和索引处理

集成多个来源，
实时数据视图



过多

分析

高性能数据

快速访问 GPU/CPU
不断演变的数据科学
工具



瓶颈

融合

扩展基础架构

最大程度降低复
杂性
易于实施



复杂

洞察力

有助于简化从边缘到洞察力的整个 AI 基础架构的存储

收集

摄入高吞吐量的数据

扩展容量和性能
集成到核心和云端

组织

大数据量的目录和索引

从边缘到核心到云端的
360 度视图
缩短数据分析所需时间

分析

高性能数据访问

无瓶颈的更快速分析
通过全球访问简化基础架构

融合

面向整个组织扩展

在整个组织中更轻松
地集成

边缘

洞察力



IBM
Spectrum
Discover



IBM
Spectrum
Scale



IBM Cloud
Object
Storage

IBM Storage
Suite for IBM
Cloud Paks

从边缘到洞察力简化支持 AI 的解决方案

收集

摄入高吞吐量的数据



现代化

组织

大数据量的目录和索引



集成式

分析

高性能数据访问



高性能

融合

面向整个组织扩展



简化

边缘

洞察力



IBM
Spectrum
Discover



IBM
Spectrum
Scale



IBM Cloud
Object
Storage

IBM Storage
Suite for IBM
Cloud Paks

一家大型电信公司借助 IBM Storage 的产品整合了边缘数据并节省了大量成本

电信 - 来自边缘设备的云存储

节省 25%

相比之前的解决方案

减少 15%-35%

相比之前的解决方案，物理资产减少了 15%-35%

300 PB+

装机容量超过 300 PB，并以每月 PB 级容量的速度持续增长

业务挑战：

- 存储现代化
- 降低成本
- 减少物理资产
- 维持高吞吐量
- 快速、灵活的扩展

成效：

- 实现了新的云原生访问
- 减少了物理资产
- 节省了 25% 以上的成本
- 不影响性能
- 多个在线扩展



“客户支持服务非常有用，产品非常稳定，让我们的运营变得更轻松。” - 某位 IT 经理

某家医疗中心希望更好地组织和管理研究数据与关键实验数据

医疗保健 | 数据治理与优化

业务挑战：

- 对大型的基因数据集进行分类
- 监控并报告数据
- 查找 PHI/PII 数据
- 确立数据使用模式

成效：

- 数据报告速度更快，也更轻松
- 现在可以分析新的数据模式，以实现进一步分类
- 可以更轻松地识别 PHI/PII 数据
- 将用例扩展到超过 100 PB 的额外数据

30PB

的初始数据由 Spectrum Discover 进行管理

识别个人数据

从医疗成像数据集中查找 PHI/PII 数据

识别数据使用量

确立模式，以供监控和报告使用



从边缘到洞察力融合 IBM Storage for Data and AI

- AI 是未来的发展方向，而未来的发展要在于现在
- 如果没有稳健的信息架构 (IA)，AI 项目就会失败
- IBM Storage for Data and AI 可提供这种 IA 基础架构

边缘



IBM
Spectrum
Discover



IBM
Spectrum
Scale



IBM Cloud
Object
Storage

IBM Storage
Suite for IBM
Cloud Paks

洞察力

面向数据和 AI 的存储化繁为简

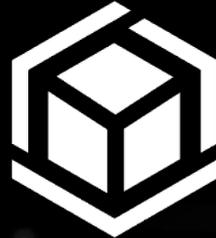
- 从收集、组织到分析和融合，简化数据和 AI 基础架构
- 从边缘到洞察力简化支持数据和 AI 的解决方案



IBM
Spectrum
Discover



IBM
Spectrum
Scale



IBM Cloud
Object
Storage

简单的IBM存储，让数据不简单

- 2020年三季度IBM 存储新品发布

周立旻

数据与AI存储及数据保护现代化

IBM系统部

zhouly@cn.ibm.com

以数据为基础，打造 AI 成功的阶梯



AI

融合

分析

组织

收集

现代化的
数据存储架构

无 IA (information architecture) 不 AI

IBM 提供完整的数据与 AI 存储平台

- ↓ 复杂性
- ↓ 成本
- ↑ 整合度



收集



发布

文件数据湖
IBM ESS 5000

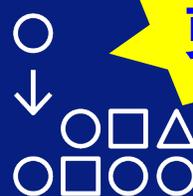


更新

对象数据湖
IBM COS



组织



AI 策略引擎与数据目录



分析



NVMe 高速闪存
IBM ESS 3000



文件数据湖

- 基于磁盘的系统，提高总体容量并降低成本
- SL 型号可安装在标准机柜中，单个模块最大可扩展到6个磁盘扩展柜（92块磁盘）及8.8PB容量
- SC 型号可安装在加深机柜中，单个模块最大可扩展到8个磁盘扩展柜（106块磁盘）及13.5PB容量
- 基于容器的管理堆栈，实现分钟级安装
- 支持实时向IBM Spectrum Discover 推送更新的元数据，无需扫描所有文件
- 可以无缝地和 ESS 3000 NVMe 模块或前几代 ESS 产品整合，实现投资保护



IBM Elastic Storage System 5000

SL 型号
552TB - 8.8PB
260TB/u

SC 型号
1PB - 13.5PB
375TB/u

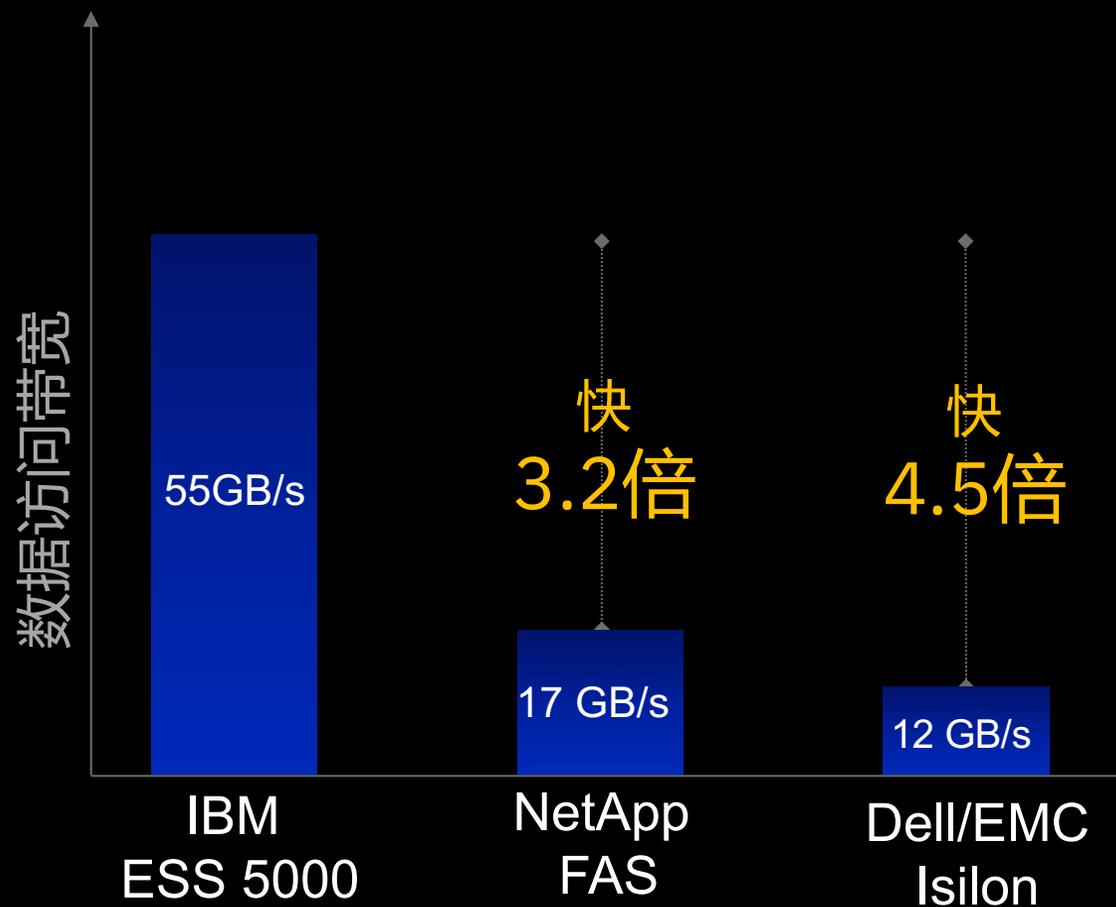


标准机柜

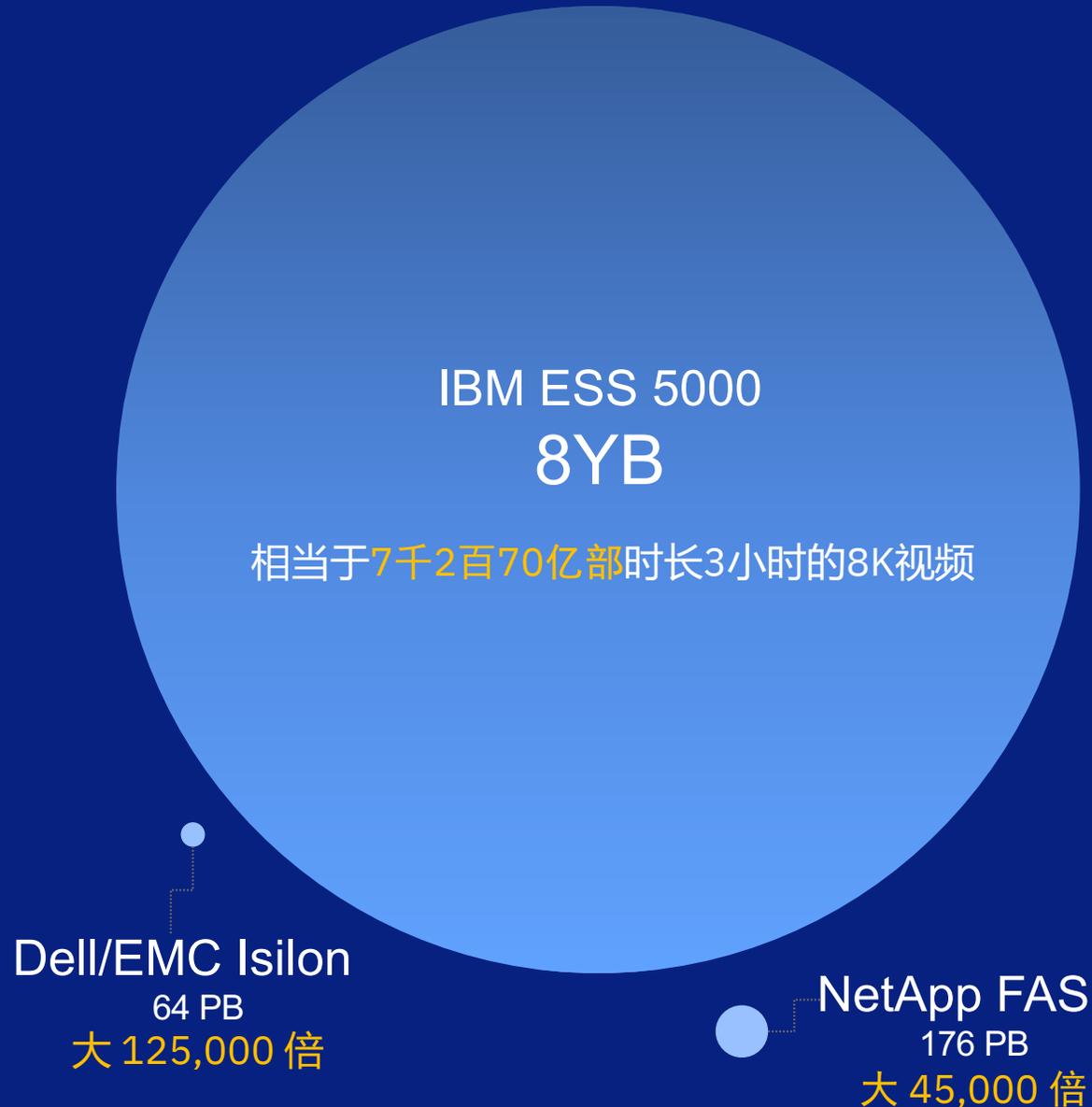


加深机柜

IBM ESS 5000 的性能



IBM ESS 5000 的容量



IBM Elastic Storage System

打造具有市场领先的性能、密度和可扩展性的企业级数据湖

只有 IBM 能提供：

- 简单化
 - IBM ESS: 3 种型号
 - Dell-EMC Isilon: 10种
 - NetApp FAS: 9种
- 分钟级的安装，轻松管理数据中心内的多个节点
- 达到 370TB/u 的密度

YB级高性能数据湖

针对“**收集**”优化



ESS 5000

针对“**分析**”优化



ESS 3000

Powered by



IBM Spectrum Scale

针对“**组织**”优化
实时更新元数据



IBM Spectrum Discover

只有 IBM 能提供：

- 通过在不同应用阶段将数据放在合适的存储上，加速 AI应用
- 单一方案可扩展到8YB，消除数据孤岛
- 单模块 55GB/s 的速度

对象数据湖

增强性能并降低成本

- 采用专利技术对存储进行直接读写，避免本地文件系统开销，大幅提升**性能**
- 在线软件升级和数据迁移，提升**效率**
- **业界首家能够**提供主机管理（Host Managed）SMR 磁盘的对象存储

高达

300%

数据读性能提升

高达

30%

地理分布节点时延缩短

高达

55GB/s

12节点配置的数据访问带宽

提高密度

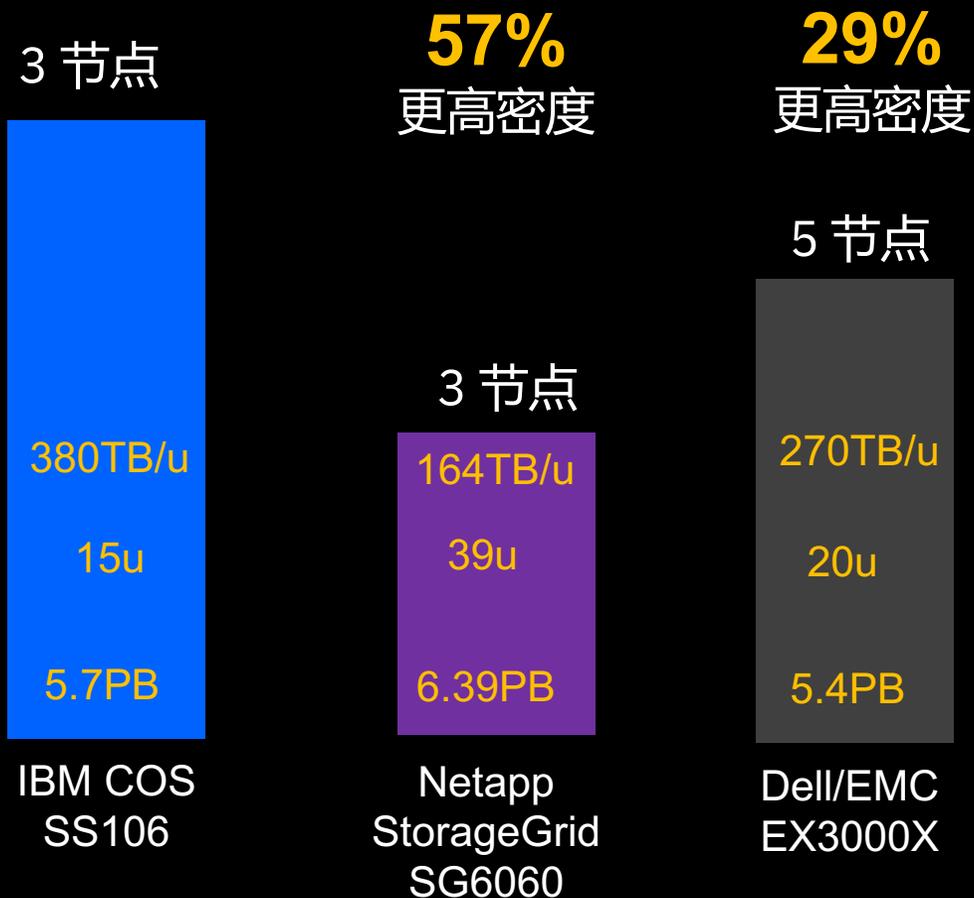
12%

采用 18TB SMR 磁盘

IBM Cloud Object Storage 保持业界技术领先地位

消除云对象存储数据访问 瓶颈

对象存储集群配置对比



图像与视频文件备份
或归档到云存储

从 20 分钟
缩短到 5 分钟



读



写

从 20 分钟
缩短到 8 分钟



IBM Cloud Object Storage

IBM Cloud Object Storage

提高对象数据湖的效率，12节点集群速度可达 **55GB/s**

只有 IBM 能提供：

- 为现有安装的系统带来 300% 的性能提升
- 提供市场领先的容量扩展能力和数据保护能力，消除数据孤岛并降低成本

EB级多站点多活的数据湖



针对“**组织**”优化
实时更新元数据



只有 IBM 能提供：

- 针对 AI 优化的对象存储，提升 AI 和 HPC 应用性能
- 支持 18TB SMR 磁盘，将存储密度提高12%

组织



AI

融合

分析

组织

收集

现代化

AI 基础架构的**组织**阶段需要：

灵活

- 可支持多种混合云部署

异构

- 支持不同数据源

整合

- 持续的、实时的数据更新来支持数据收集和分析
- 支持不同厂商

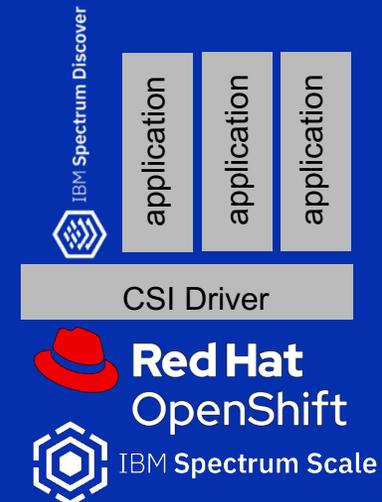
易于使用

- 支持不同应用场景

Red Hat OpenShift 容器应用可以方便地对数据进行索引，进一步优化 AI workflow，实现快速数据分类

- 部署在 Red Hat OpenShift 上
- 易于在多云环境部署
- 减少 50% 的内存资源消耗，和现有虚拟机部署方式相比可进一步降低成本

IBM Spectrum Scale + IBM Spectrum Discover + Red Hat OpenShift all integrated



Planned GA Q4

IBM Spectrum Discover 策略引擎升级

更好地利用大容量、低成本的存储

提供 REST API 可和第三方数据移动工具或 IBM Spectrum Scale 整合

- 将数据归档到低成本的存储上
- 利用低成本存储优化数据成本
- 迁移数据以实现数据整合并降低整体成本

IBM Spectrum Discover 能够支持更多的应用场景

数据归档



AI 加速



数据迁移



IBM Spectrum Discover

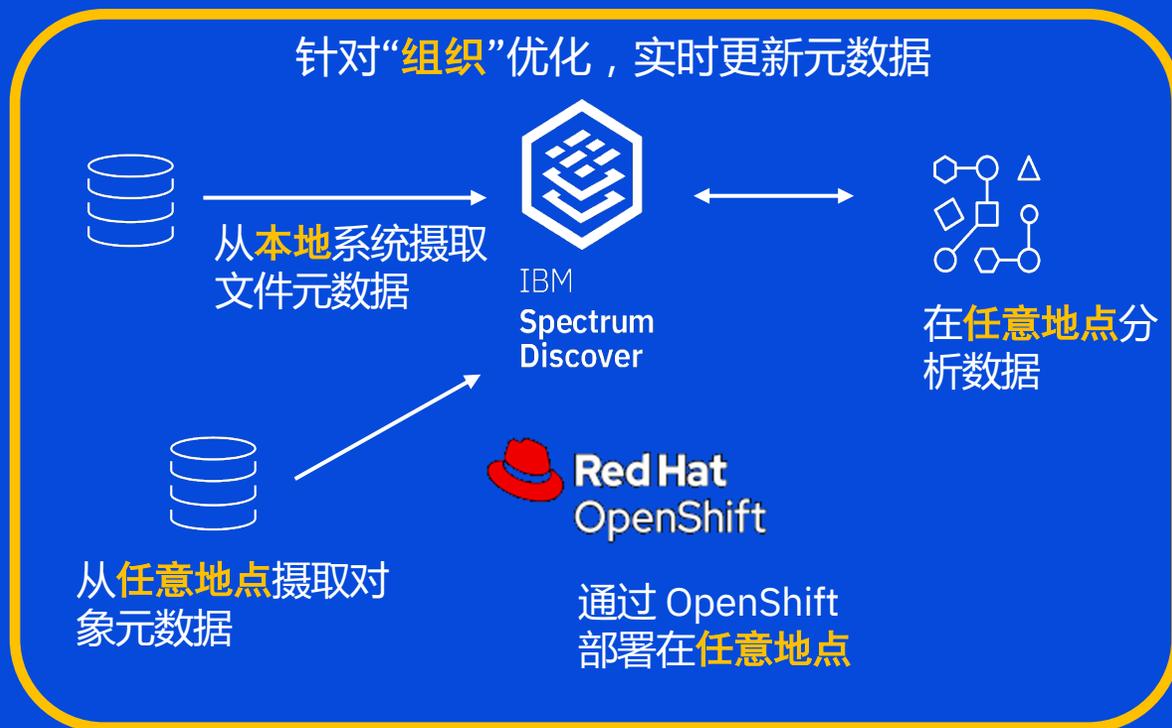
秒级搜索数十亿个文件/对象，可部署在 Red Hat OpenShift 上

只有 IBM 能提供：

- 持续的、实时的数据更新来更好地实现数据洞察，减少低效的数据重复扫描
- 通过 Red Hat OpenShift 和 IBM Spectrum Scale 实现数据目录的多云存储

异构多云存储的数据目录

针对“组织”优化，实时更新元数据



只有 IBM 能提供：

- 支持 IBM 或第三方数据移动工具优化存储
- 可和 IBM Cloud Pak for Data 实现一键集成的数据目录

引入 IBM Spectrum Scale Data Acceleration for AI

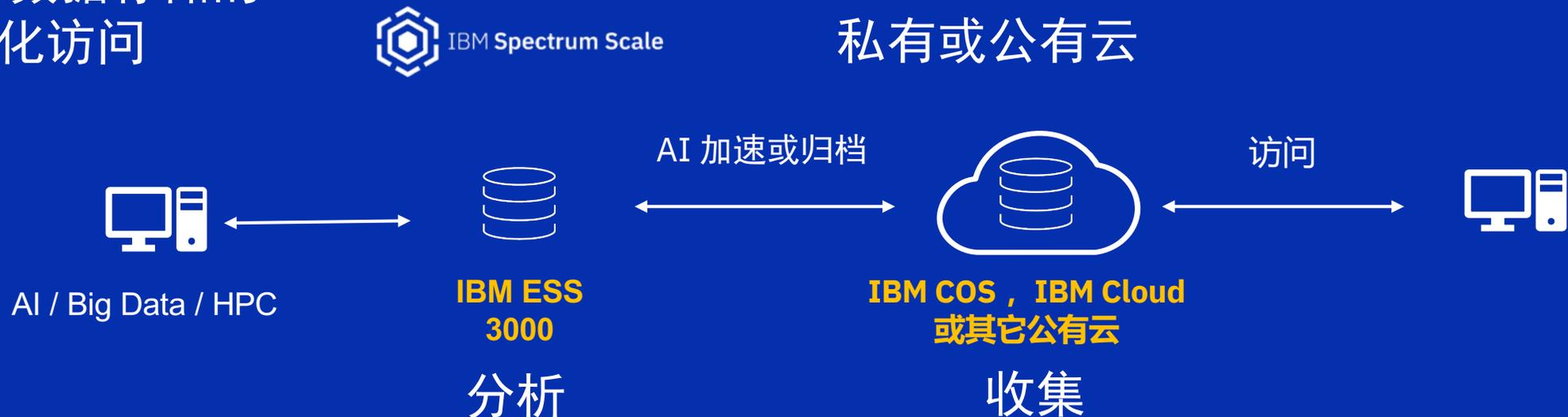
支持数据在 IBM Spectrum Scale 和对象存储（无论是数据中心内还是云端）之间的灵活访问和移动

Planned GA Q4

降低 50% 的成本, 减少不必要的副本

简化数据移动, 自动或手动对存放在大容量、低成本对象存储上的数据进行加速

AI 和数据存储的
一体化访问



IBM Storage Suites 简化采购并降低成本

IBM Storage Suite for IBM Cloud Paks

开源 Red Hat 存储软件 + IBM 企业级 Spectrum Storage 软件

- 包括: Red Hat Ceph, OpenShift Container Platform, IBM Spectrum Virtualize, **IBM Spectrum Scale, IBM Cloud Object Storage and IBM Spectrum Discover**
- 为容器应用/ IBM Cloud Paks 提供必需的软件定义存储环境
- 降低存储软件开销

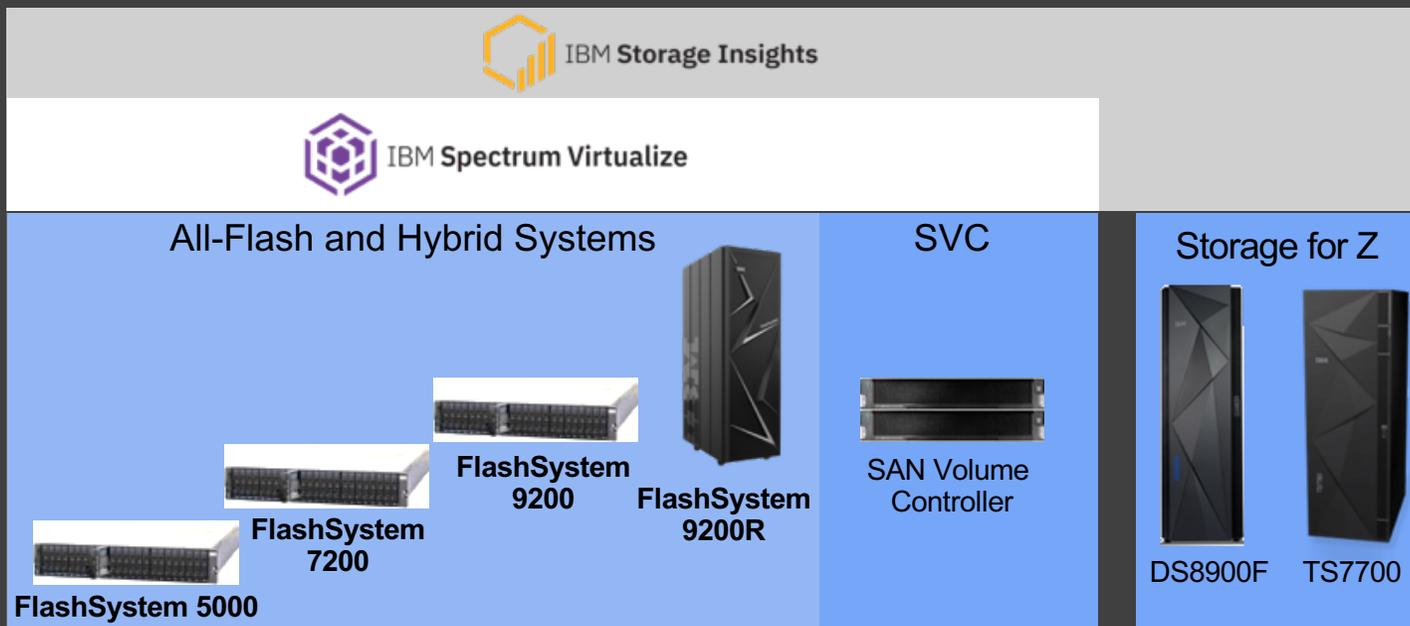
IBM Spectrum Storage Suite

打包销售的 IBM Spectrum Storage 套件

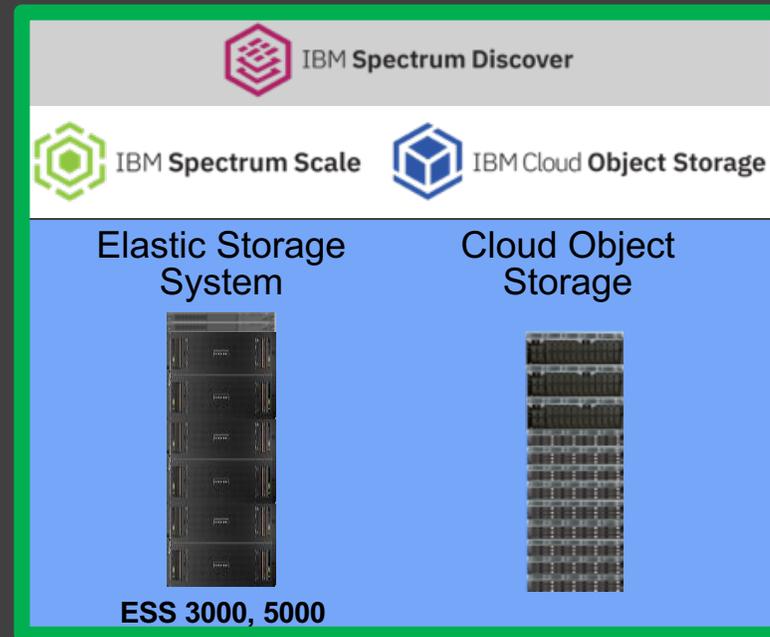
- 包括: IBM Spectrum Storage Software only: IBM Spectrum Control, IBM Spectrum Protect, IBM Protect Plus, IBM Spectrum Virtualize, IBM Spectrum Archive, IBM Spectrum CDM, **IBM Spectrum Discover, IBM Spectrum Scale, IBM Cloud Object Storage**
- 为各种应用提供软件定义存储环境
- 降低多达 40% 的存储软件开销

IBM 屡获殊荣的存储产品家族

混合多云存储



数据与 AI 存储



网络安全弹性与数据保护现代化

实现开放的、端到端的 AI 数据管道

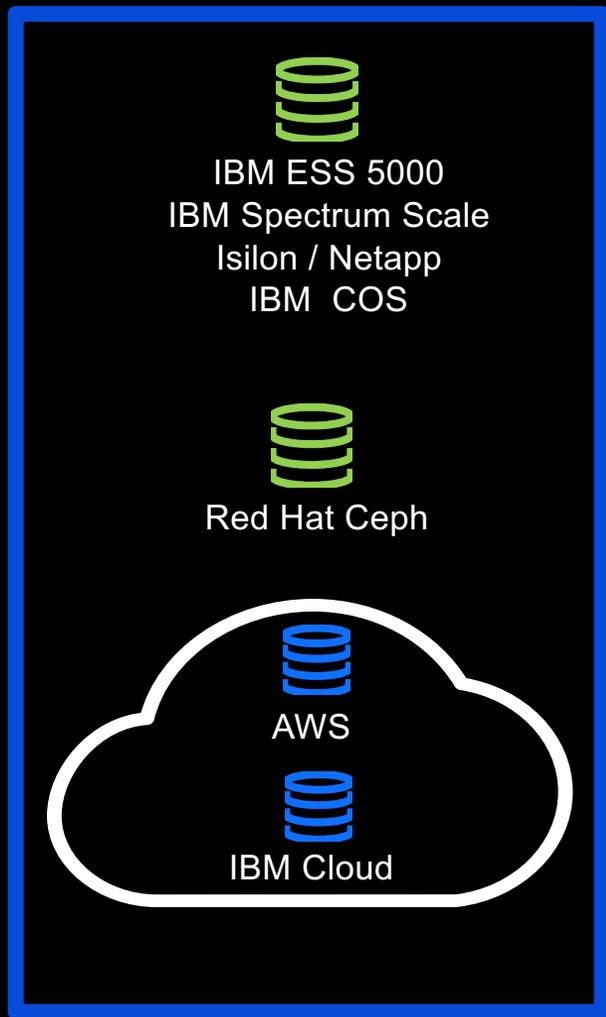
收集

组织

分析

数据
摄取

数据
洞察



AI 加速
或归档

IBM Watson
& IBM Cloud
Pak for Data

搜索与整理



简单的 IBM 存储，让数据不简单

IBM Elastic Storage System 5000

新一代 ESS

- 更简单: 易于安装, 易于配置
- 和其它产品相比数据访问速度快4.5倍
- 单模块容量可达 13.5 PB (降低 CAPEX/OPEX)

IBM Spectrum Scale

AI 数据加速

- 通过整合数据孤岛, 降低多达 50% 的总体成本
- 实现自动的 AI 数据流, 降低运维成本, 提升效率
- 无缝对接对象存储, 支持AI和大数据应用

IBM Cloud Object Storage

- 读性能提升 300%, 写性能提升 150%
- 12个节点对象存储集群性能可达 55GB/s
- 通过支持18TB SMR 磁盘提升 12% 的存储密度

IBM Spectrum Discover

- 实时更新元数据, 更好地实现数据洞察, 减少低效的数据重复扫描
- 通过Red Hat OpenShift 和 IBM Spectrum Scale 实现数据目录的多云存储
- 支持 IBM 或第三方数据移动工具优化存储

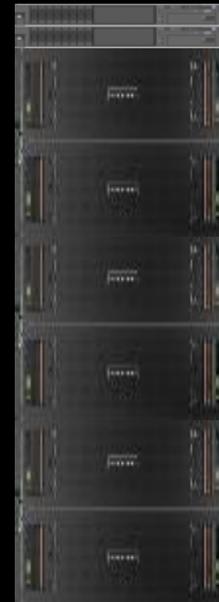
Reference Information



IBM Elastic Storage System 5000

	SL	SC
Number of Enclosures	Up to 6	Up to 8
Enclosures	Standard size (92)	Expanded Size (106)
SSDs	2 SSDs	
HDDs Raw	6, 10, 14 or 16TB	10, 14 or 16TB
capacity	552TB - 8.8PB	1PB - 13.5PB
Servers	2 x IBM Power9 Servers	

SL model



SC model



Leadership for capacity scale-out (HDD based) file storage

	IBM ESS 5000 SL SC	Dell EMC Isilon H400 H500 H5600 H600 A200 A2000	NetApp FAS	Pure Storage FlashBlade
Max throughput (capacity storage)	 55 GB/s	 12 GB/s	 17 GB/s	No capacity solution
Max total throughput (capacity storage)	 2.5 TB/s +	 756 GB/s	 408 GB/s	
Max capacity (per cluster)	 8 YB	 64 PB	 176 PB	
Integrate file/object data	 Both (announced)	 None	 File only	
Smallest Configuration	 552 TB	 12 TB	 10 TB	
Data lifecycle management NVMe/Flash/HDD/Cloud/Tape	 ALL	 No Tape/NVMe cache only	 No Tape/ NVMe cache only	
Raw density per rack unit	 1.7 PB	 TB	 1.8 PB	
Max capacity per enclosure	 13.5 PB	 960 TB	 14.7 PB	
Integrate with Public Cloud deployment	 Yes	 Limited as a service	 Yes	

Complex and Costly Portfolio

Collect

Organize

Analyze

Scalable file/object capacity

Metadata Catalog

Scalable high performance

Dell-EMC



ClarityNow

Data IQ



HPE

Object – Partners only

File – Partner only

NetApp



Cloud Insights



Pure



Hitachi Vantara



Hitachi Content Intelligence

IBM Cloud Object Storage vs. Competition

	IBM Cloud Object Storage	Dell/EMC ECS	Netapp StorageGrid	Clouidian
Largest capacity node	 1.7PB	 1.4PB	 2.1PB	 1.5PB
Largest deployed customers	 EB+	 unknown	 unknown	 u nknown
Accelerate object data to flash file data	 YES	 No	 No	 No
Increase performance with software upgrade	 YES	 No	 No	 No
Smallest configuration	 3 nodes (1 site or 3)	 5 nodes (1 site)	 3 nodes (1 site)	 3 nodes (3 copies) 5 nodes (1 site)
Geo-dispersed protection	 Yes / EC	 Yes / XOR	 No (replication)	 Yes / EC
File Access	 3 rd Party	 NFS only	 NFS/SMB	 Add -on
Tier to Cloud	 No	 No	 Yes	 Yes
Software only Suite option	 Yes	 No	 No	 No

为企业数据湖而生的 ESS 5000

- 技术路线详解

冯硕
IBM中国系统实验室技术经理



数据的能量

提供显著的竞争力

在2019年

40%

的数字化转型使用了AI

\$4.79B

与AI相关的存储开销

到 2021年

75%

的商业应用
会使用到AI

>90%

的消费者
会与客服机器人互动

>50%

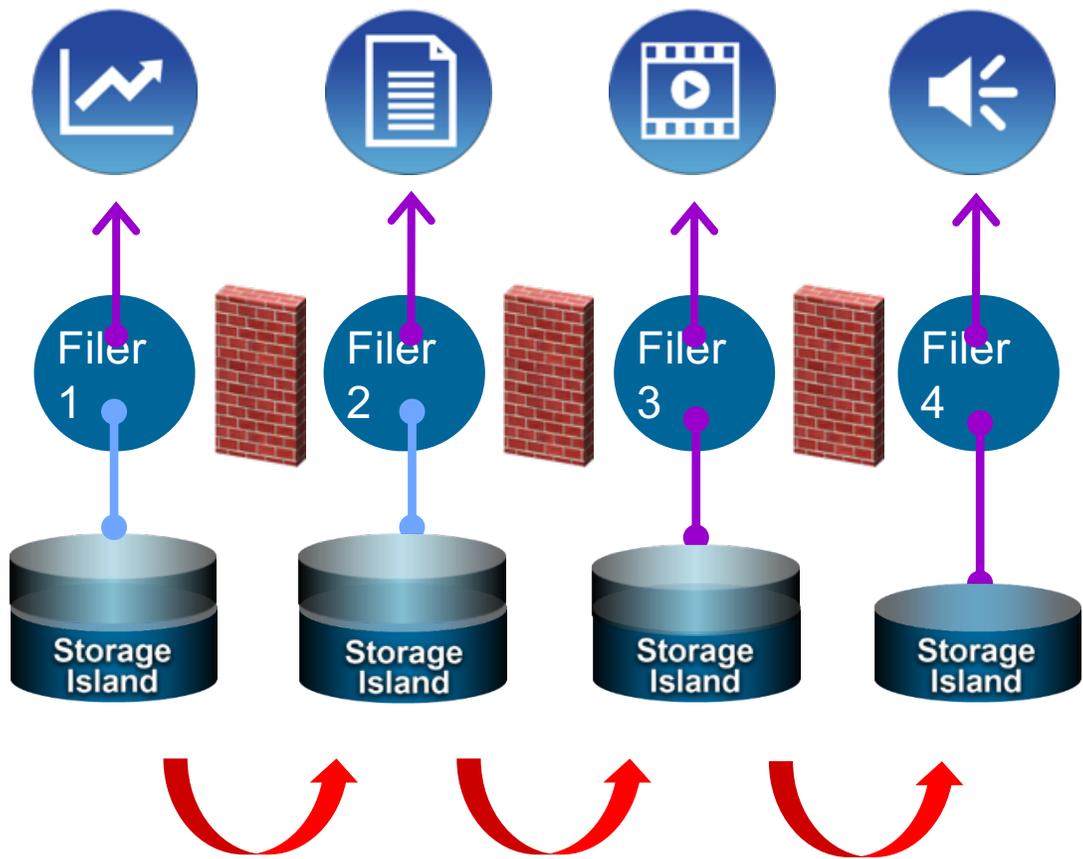
of 的新型工作机器人
会使用到 AI

企业数据难题： 复杂度高，数据量大，数据孤岛，成本昂贵

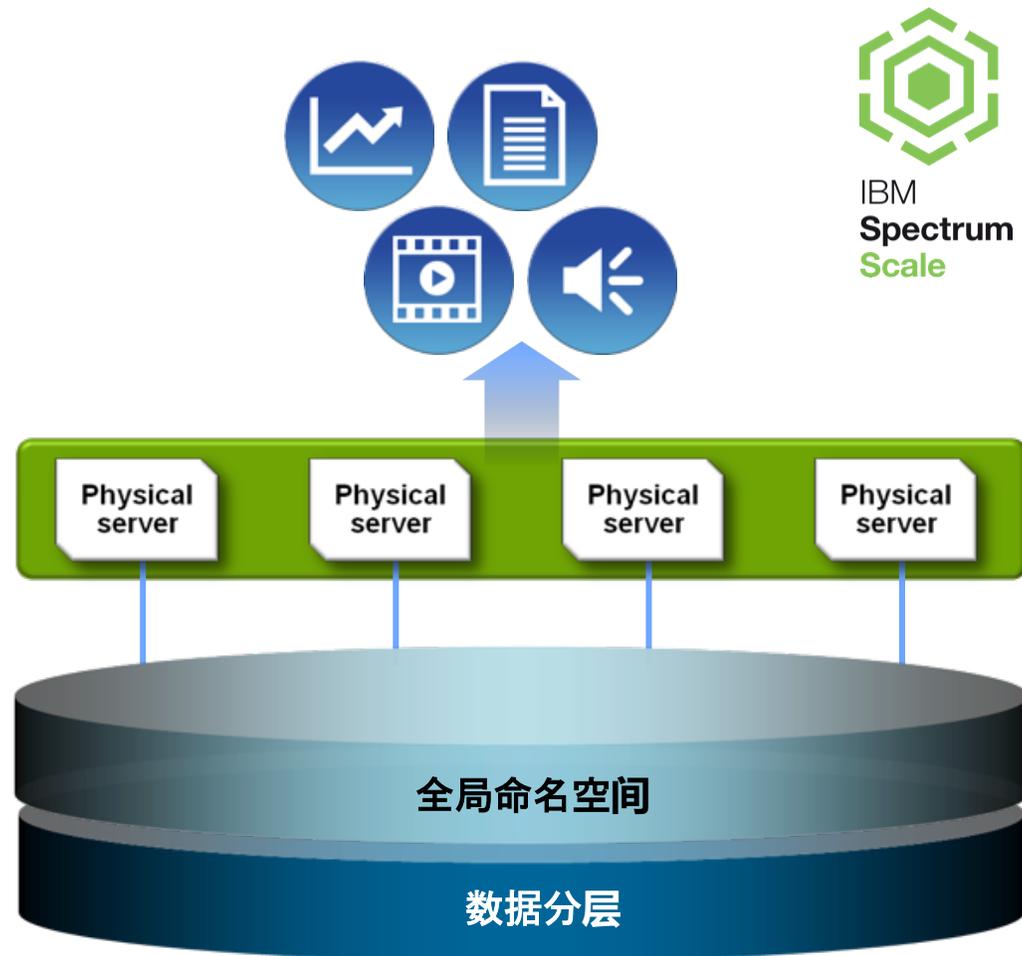


高性能数据结构

使用 Spectrum Scale之前



使用 Spectrum Scale之后



Spectrum Scale 的部署方式



IBM Elastic Storage Server 的多种型号与配置



性能配置

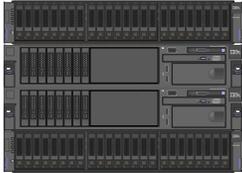
容量配置

Model GS1S
24 SSD



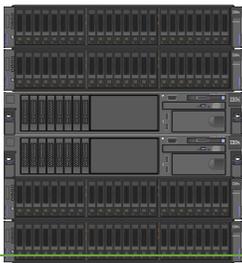
14 GB/s
92 TB raw/56(8+2),51(8+3)
360 TB/224/205

Model GS2S
48 SSD



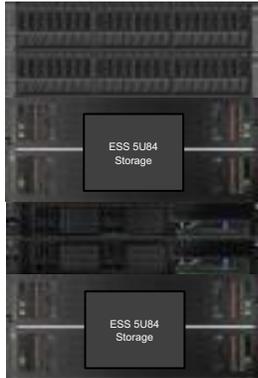
26 GB/s

Model GS4S
96 SSD

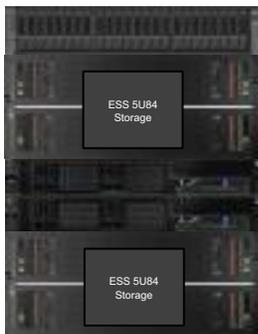


40 GB/s
368 TB raw/51(8+2),46(8+3)
1.14 PB/1/0.94

Model GH22:
2 2U24 Enclosure SSD
2 5U84 Enclosure HDD
166 NL-SAS, 48 SSD



Model GH12:
1 2U24 Enclosure SSD
2 5U84 Enclosure HDD
166 NL-SAS, 24 SSD



18 GB/s*

Model GH14:
1 2U24 Enclosure SSD
4 5U84 Enclosure HDD
334 NL-SAS, 24 SSD



38 GB/s*

Model GH24:
2 2U24 Enclosure SSD
4 5U84 Enclosure HDD
334 NL-SAS, 48 SSD



40 GB/s*

Model GL1S:
1 Enclosures, 9U
82 NL-SAS, 2 SSD



6 GB/s
328TB raw/227(8+2),208(8+3)
1.14PB/794TB/728TB

Model GL2S:
2 Enclosures, 14U
166 NL-SAS, 2 SSD



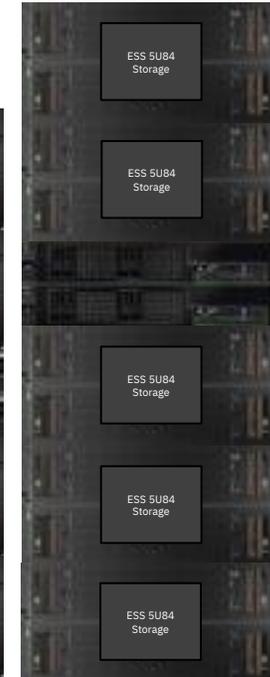
12 GB/s

Model GL4S:
4 Enclosures, 24U
334 NL-SAS, 2 SSD



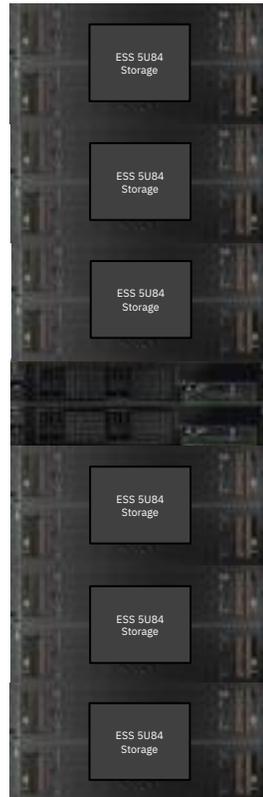
24 GB/s

Model GL5S:
5 Enclosures, 29U
418 NL-SAS, 2 SSD



30 GB/s

Model GL6S:
6 Enclosures, 34U
502 NL-SAS, 2 SSD



36 GB/s
2PB raw/1.45(8+2),1.32(8+3)
7PB/4.8/4.4

* Estimate of performance aggregated across SSD and HDD. NOTE: All estimates assume EDR Infiniband connections and are read performance

ESS C系列 与 ESS 3000

性能配置

ESS 3000
40 GB/s in 2U
24 NVMe



Model GL8C
8 Enclosures, 36U
846 NL-SAS, 2 SSD

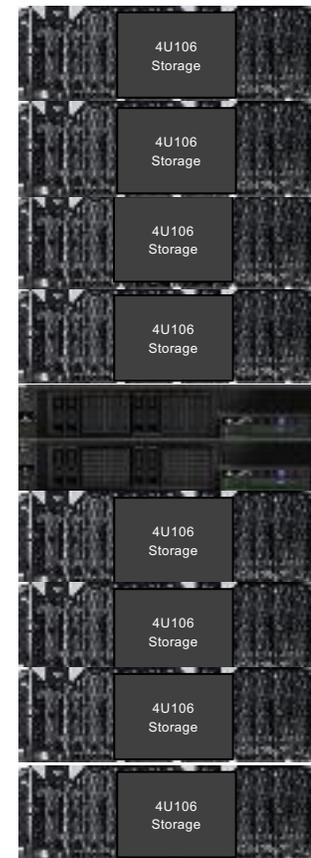
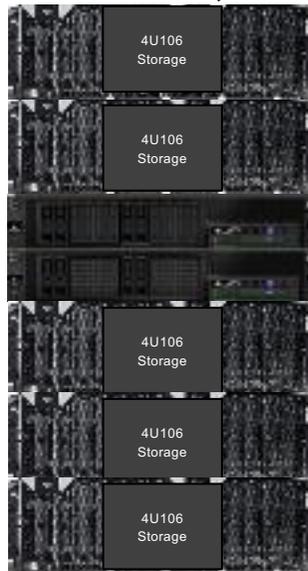
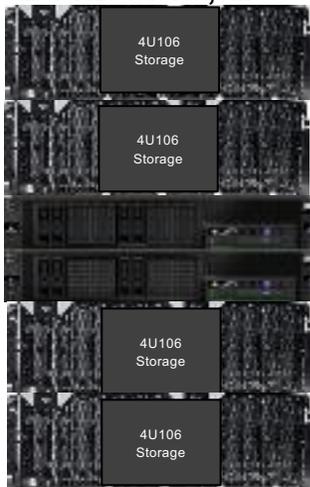
Model GL6C
6 Enclosures, 28U
634 NL-SAS, 2 SSD

Model GL5C
5 Enclosures, 28U
528 NL-SAS, 2 SSD

Model GL4C
4 Enclosures, 16U
432 NL-SAS, 2 SSD

Model GL2C:
2 Enclosures, 12U
210 NL-SAS, 2 SSD

Model GL1C:
1 Enclosure, 8U
104 NL-SAS, 2 SSD



1.46 PB raw
1 (8+2P)
0.93 (8+3P)

2.9 PB

5.9 PB

7.3 PB

8.8 PB

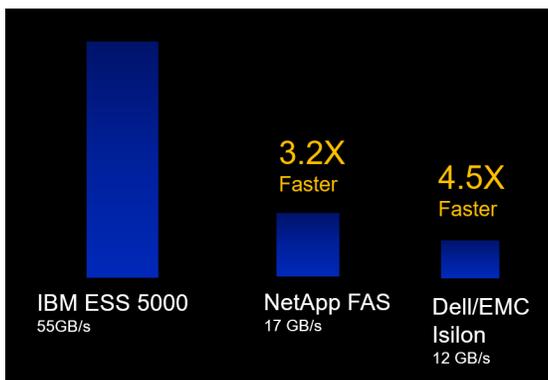
11.8 PB raw
8.5 (8+2P)
7.5 (8+3P)

容量配置

ESS 5000 极致的容量、性能与存储密度

硬件

- 基于IBM Power 9处理器，更强大的计算能力
- PCIe Gen4, 翻倍的总线带宽
- 日志使用NV-DIMM, 远低于PCIe总线的延迟和更高的带宽
- 支持最多9个Enclosure (Model SC9), 单机柜可达13PB
- 持续55GB/s的读写性能



软件

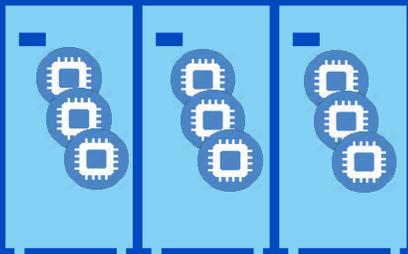
- 基于Spectrum Scale V5.0.5 EUS 扩展升级版
- Redhat Enterprise Linux 8.1
- 与ESS3000相同的容器化的部署工具, 便于安装与升级,
- 支持全系列各型号间升级
- 与已有ESS完全兼容, 可横向扩展组成混合集群



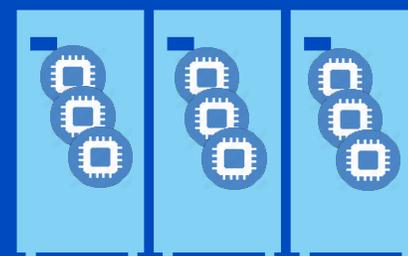
计算 / 应用层



客户
工作站



Compute / app
farm with GPUs



Metadata,
indexes,
databases,
backup targets

存储层

高性能存储层

AI Model
training and
inference



Metadata,
indexes, hi-perf
storage, DB,
Spectrum Protect
databases

容量存储层



Capacity storage
working
seamlessly with
High Performance
storage

Cloud Data
Sharing

Hybrid

Public

Private

Multi-Cloud

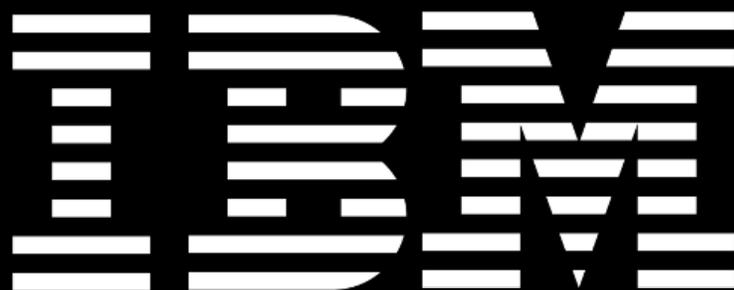
Automated
Synchronization



Tape

Spectrum Scale与ESS完整支持 Redhat OpenShift和容器接口





Social Distancing Save Lives
Thank You and Be Safe

欢迎致电 IBM 存储热线：400 – 810 – 1818 转 5123