DDN 持续支持 Lustre 社区发展

DDN中国 陈道碧 2023-11-3





报告提纲

- 关于DDN公司
 - 。概述
 - o DDN在中国
 - 本地技术服务能力
 - **合作伙伴**

- DDN 与 Lustre 社区
 - DDN 对 Lustre 的持续支持
 - 。 商业版 EXAScaler 与社区版 Lustre 的区别
- EXAScaler应用案例
- · 结语 DDN的优势



DDN 创建和提供数据存储解决方案 助力企业、政府和学术界实现数字化转型

11,000 + 客户 企业, 政府, 学术机构 市场覆盖广 科研, AI, 医疗健康, 云, 金融服务, 政府 全球性企业 客户,实验室和员工遍布 美国,欧洲中东非洲,亚太

一<mark>流团队</mark> 1,000 多名团队成员 600 多名工程师 稳健、可持续 24+年踏实经营 工程师文化 大规模存储领导者 经过实践检验的 创新与交付





EXAScaler 先进的商业版 Lustre 并行文件系统

- —— 面向AI 和 HPC的智能优化环境
- 性能 -性能最佳的并行文件系统,具有高级的应用程序、客户端、网络和集群级别的优化
- 易用 新的 EXAScaler 管理框架简化并自动化管理, 使用CLI 和 API进行在线升级和扩展,实现更简单的管理
- 数据管理 客户端缓存,透明的闪存/HDD混合,先进的自动条带化和平衡, DDN Dataflow 软件可在系统间和站点间移动数据
- 安全 加密,审计,多租户
- · 监测和分析 -DDN Insight 4.0 软件可实时监测分析





集成 EXAScaler 并行文件系统的一体化存储平台

型号	ES200NVX2	AI/ES400NVX2	ES7990X	ES18KX
	ddh	ddn		Section 1
性能	顺序读带宽 48 GB/s IOPS: 150万	顺序读带宽 90GB/s IOPS: 300万	顺序读带宽 24GB/s IOPS:80万	顺序读带宽 76GB/s IOPS: 300万
最大 支持 硬盘 数量	24颗 NVMe 盘	24颗 NVMe 盘 900颗SAS盘(HDD/SSD)	450颗 SAS 盘	48颗 NVMe盘, 1872颗 SAS盘
主机 接口	4 x HDR/HDR100 或 100/200 GbE	8 x EDR/HDR200 IB 或 200GbE	4 x EDR/HDR100 或 40/100 GbE	16 x EDR/HDR100 或 40/100 GbE

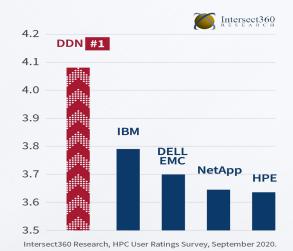


AI/ES400NVX2 - 适用于 AI 和 HPC 的全新 一级存储

交钥匙设备 - 全面优化以实现最大应用性能,已经过大规模验证







最受HPC用户认可的存储厂商

- 市场研究公司 Intersect360 对全球HPC 用户的调查: **DDN获得用户最高评分**
 - 。 最高总分
 - 。 最高技术分
 - 。 最高实际运行分
 - 。 未来最希望采购的产品

DDN所获荣誉





连续12年最佳 HPC 存储产品奖

- HPC领域专业媒体 HPCwire 组织全球 HPC 专家读者和编辑投票产生年度"读者之选" 和"编辑之选"
- 连续12年最佳 HPC 存储产品和技术

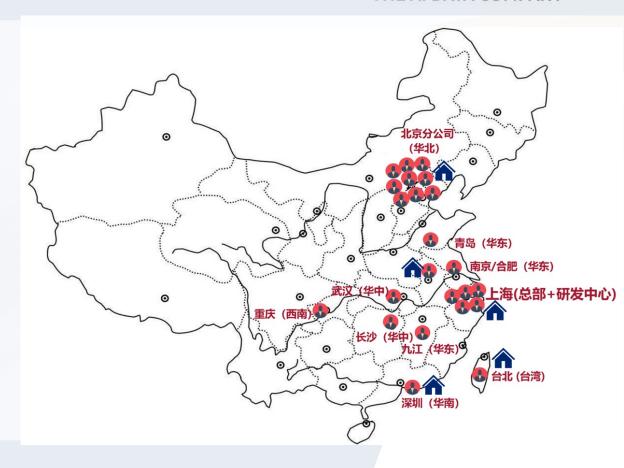
连续3年 AI 突破奖

- 第三方市场研究机构年度评选
- DDN A³I 解决方案在"领导者" 类别中荣获 AI硬件创新奖



DDN在中国

- 2000年进入中国市场, 2011年起 致力于HPC/AI市场推广
- 产品已广泛用于知名高校、科研院所、超算中心、能源、制造、 生物医药和金融等行业
- 2022年在中国投资设立全资子公司:上海数直通信息科技有限公司以及上海数直通信息科技有限公司北京分公司
- DDN中国团队包括 Lustre文件系统 研发、售前/售后技术支持、销售、 市场
- 中文网站: <u>ddnstorage.com.cn</u>
- DDN中国微信公众号: DDN_China





稳定性极高的专业技术支持团队



- 中国技术团队占DDN中国员工总数 70% 50%的技术员工在DDN工作时间 >5年,80%的技术员工在DDN工作时间 >3年
- 中国团队拥有丰富的高性能存储行业经验 100%的技术团队成员在存储行业的工作经验 >10年; 70%的销售团队成员具有技术背景
- · 技术团队将为用户和合作伙伴提供商业版文件系统 EXAScaler 的培训
- 技术服务热 线: <u>400 825 0880</u>



本地实施能力- 大规模HPC/AI集群



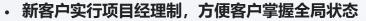












- 最佳实践 流程规范 文档齐备
 实施前出具工作说明书 (SOW) → 安装报告
 → 性能测试报告 → 用户培训
 - → 提供培训资料
- 实施人员经培训认证, 具备最新专业技能
 - 大规模并行存储 (Lustre) 安装部署
 - InfiniBand 网络部署
 - Kubernetes & Slurm & 监测系统部署
 - 节点 & 集群 性能基准测试和验证

HPC/	AI*	集群部署
------	-----	------

国家级超算 中心	20+套,	百PB
高校实验室	70+套,	百PB
科研院所	30+套,	百PB
能源/制造行 业数据中心	30+套,	百PB

* 支持DDN 存储与基于任意品牌的GPU或 AI 加速器的OEM系统进行集成优化







DDN 拥有广泛而专业的合作伙伴

- 合作伙伴生态系统: 全球解决方案合作伙伴 + 技术合作伙伴 + 高质量的经销商
- 国内外头部的芯片提供商
- 国内外系统集成商
- 行业解决方案提供商
- 在国内已建立了经销商体系,书面授权的经销商可在DDN中文官网查询

DDN将与国内合作伙伴一起拓展市场 合作共赢!



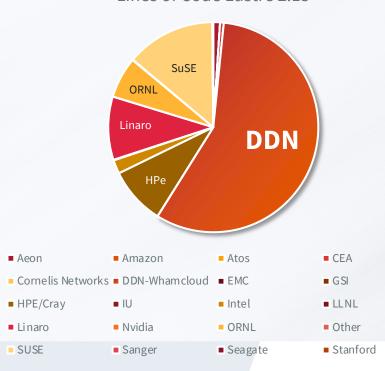




对 Lustre 2.15 (OpenSFS)的贡献统计

- DDN一直秉持厂商中立的态度,积 极参与和支持开源社区的各项工作 与活动
- DDN公司旗下的 Whamcloud工程 部门主要负责Lustre的持续研发工作, 管理所有开源流程并维护源代码树
- DDN把许多自研代码贡献到社区,是 Lustre 源代码的主要贡献者
- DDN已经连续十二届参加并支持 CLUG峰会

Lines of Code Lustre 2.15





商业版 EXAScaler 6 相对于社区版 Lustre 的优势

Hot Pools -在 SSD 和 HDD 池之间自动数据分层

Hot Nodes - 把数据自动缓存到计算系统本地的闪存上

End-to-End Data Integrity - 从客户端到介质的校验和

Enhanced Multi Rail - 简化客户端网络*

Data On Metadata – 小文件的最高效率

NFS/SMB/S3 Data Services - 可扩展的 MultiOS 协议访问

DDN A3I - 面向AI和DL的交钥匙、全面优化的解决方案

User Analytics - 节点作业监测, 文件大小分布报告

Container Optimized Client - 容器完全并行数据访问

Multi Tenancy - 面向共享环境的安全部署

DDN DataFlow Solutions - 备份, 归档, 移动, 同步大规模数据

Tight GPU Integration - GPUDirect, 面向 DGX 容器友好型部署

In the Cloud -在AWS, Google Cloud, Azure 都能使用 DDN 并行文件系统



Advanced Management	MultiProtocol Access	Al Integration	
Performance Optimizations	Added Security	Data Protection	
Open Industry Leading Parallel Filesystem, Lustre			
DDN Appliances and Public Cloud			



DDN 商业版 EXA6 vs 社区版 Lustre

	EXA6	Lustre
可扩展监测	DDN Insight	DIY with open source tools
详细的用户监测	Insight Workload Analyzer	DIY integration with Jobstats
系统管理	EXAScaler Management Framework	Linux based facilities and ad hoc tools
闪存管理	FSTRIM support	None
AI 集成	Reference Architectures for AI/GPUs	None
Declustered RAID	DDN DCR	None

	EXA6	Lustre
支持 GPU	Containerized Client	Nothing specific
多租户	Secure multi tenancy	DIY using open tools
端到端数据保护	End to end data protection	No
数据管理	DDN DataFlow	DIY with open tools
闪存介质集成	Hot Pools	No Flash Tiering
NAS 访问	DDN NSDS	No NAS Access or DIY



DDN 商业版 EXA6 vs 社区版 Lustre

	EXA6	Lustre
高可用	Optimized, Integrated HA	Deploy Open Tools
S3 访问	DDN S3 Data Services	NONE
云集成	DDN EXA in Azure, GCP, AWS	AWS or Azure managed
简单性	Integrated NVMe and Hybrid Appliances	External servers and/or systems with loose integrations between Lustre and appliance
能效	90GB/s in ~1KW	Typically at least 2x more power









EXAScaler 用于高性能计算中心

典型应用:传统高性能计算,机器学习、数据分析

客户挑战

• 节省空间: 高效利用机柜空间

• 高性能:大量CPU和GPU核心;高速低延迟网络需求,GPU需要RDMA和大规模高性能并行文件系统

• 简单运维:对海量数据存储设备进行简单高效的运维服务

- 规模 168U空间容纳 36PB容量,单一文件系统
- •性能 实测读写带宽 400GB/s,全年保持99.99%不停机稳定性能输出
- EXAScaler 商用版Lustre并行文件系统
- 简洁的融合存储架构,无需独立服务器做OSS,简单方便运维
- •全方位售后服务,研发级支持



EXAScaler 用于AI 智算中心

典型的人工智能应用包括:大规模交通图像视频数据分析,城市级交通态势感知、建模与预测;多模态人工智能辅助诊疗;赋能智慧金融

客户挑战

- 计算数据迅速增长, 要求存储提供高吞吐量;
- •自动化的安装部署工具,操作简单,维护管理;
- ·满足AI、云原生等多种应用要求
- •智算中心的算力需不断演进

- •规模:15PB容量 性能:聚合带宽 140GB/s
- EXAScaler商用版Lustre并行文件系统:支持约50亿个数据文件的存储和高速并发访问,满足城市交通图像视频数据的分析需求
- •面向未来可实现<mark>无缝横向纵向扩展</mark>,性能容量线性增长,最大容量可达EB 级





EXAScaler 用于海洋及气候研究

典型应用:海量、多源、异构的海洋调查大数据、观测大数据、计算大数据和网络大数据的快速及 时分析处理



客户挑战

- 数据离散分布:科研结果数据分散在不同机构和 人员手中,需通过多种渠道稳定、持续地获取大 数据
- 业务数据以PB级增加,累积数据量大,需大容量的数据存储空间;存储系统在不停机的情况下扩展,不影响业务运行
- 高通量、高性能计算分析需要高性能的存储能力, 保障数据支持, 保证科研效率

- 规模 总容量超过14PB;
- 性能 总实测带宽~100GB/s;
- 同时支持两台768口的OPA网络,940台双路 CPU计算节点
- EXAScaler 商业版Lustre并行文件系统 统一命名空间,支持以太网、InfiniBand、OPA混 合组网,一套存储系统可同时接入前端数据采集和 高性能计算平台



EXAScaler 用于冷冻电镜平台

某高校的冷冻电镜平台配备了高端的冷冻透射电镜及数据存储和处理的超算中心,旨在从最根本的蛋白质分子和生物大分子复合体的结构入手,结合亚细胞、细胞以及组织层面的三维结构,揭示蛋白质机器和生命过程调控的机理,为生命科学、基础医学、临床医学、农学、环境科学等学科提供关键性技术服务。



客户挑战

- 冷冻电镜三维重构软件例如 RELION2 (GPU) EMAN (MPI) 需要存储设备提供不同 I/O 模式
- 业务需求高性能、大容量、高可靠存储系统
- 对海量数据存储设备进行简单高效的运维服务

- 规模 8U空间内 2PB容量
- 性能 聚合带宽 20GB/s
- EXAScaler 商用版Lustre并行文件系统 统一命名空间,扩容简单方便,性能和容量按需线性扩展
- 存储系统统一管理, 灵活易用, 减少运维管理难度



EXAScaler 用于自动驾驶解决方案研发



DDN提供不同类型的存储产品,帮助客户 打通数据采集、海量数据处理、模型训练 等环节,形成数据闭环



采集存储 IntelliFlash IF6200



海量数据存储 SFA200NVXE+ ES7990



训练存储 Al400NVX2



EXAScaler 用于大模型训练

某企业研究院建立了中国首个基于 NVIDIA SuperPOD 架构的超大规模计算集群,它已支持该企业完成参数量近五十亿的 大语言模型的训练。

客户挑战

- 大模型训练需要多节点协同
- 在训练模型时,需要从存储中多次读取训练数据,而读取操作的耗时也 会在一定程度上影响训练的时效
- 训练过程中需要经常地保存检查点数据,数据写入性能对此至关重要

- DDN AI400X 匹配 DGX SuperPOD ,以超过1TB/s的速率把数据从存储系统传输到 AI 环境中
- 支持 GPU Direct Storage(GDS) 技术,可以绕过 CPU 直接连通 GPU 和存储,从 而进行高速低时延的数据传输
- DDN出色的写入性能让用户快速保存检查点到DDN的环境中,从而保证数据安全,并且方便重新启动模型以采用一些不同的参数。





结语



- 1. 设计、实施、服务
- 金方位本地技术能力

2. 满足客户的

家黑定弧

3. 风格稳健





