

INFINIDAT

大容量ストレージ InfiniBox のご紹介

INFINIDAT Japan 合同会社
橋口 慎一
shashiguchi@infinidat.com

Feb 10th, 2022



SCALE TO WIN

Gfarm Workshop

INFINIDAT & InfiniBoxご紹介

INFINIDATのマイルストーン



創業者（現CTE）
モシェ・ヤナイ

2011
2014

シリーズA 投資 製品開発

創業資金調達

EMERGING FROM STEALTH

初号契約はフォーチュン
500の金融、ヘルスケア、
小売企業



シリーズB 投資



TPGによる1億5千万ドルの
資金調達

企業価値は12億ドルに上昇

グローバル展開

USに新本社、EMEAおよび APAC
に新たな拠点を開設
特許申請件数は130件まで増加

2015
2016

ポートフォリの拡張



2017
2018

シリーズC 投資

ゴールドマンサックスおよびTPG
の追加投資による9,500万ドルの
資金調達

企業価値は
16億ドルに上昇



ELASTIC INFRASTRUCTURE ADVANTAGES

- ITインフラのリスク軽減
- 100%可用性保証
- 無停止データ移行
- Elastic Pricing



2019
2020

Scale to Win

Infinidat Elastic DataFabric

更なる製品価値の向上



INFINIDAT 10周年

オールフラッシュ版InfiniBoxを
リリース



2021



CEO
フィル・ブリンガー

INFINIDAT 創業の経緯

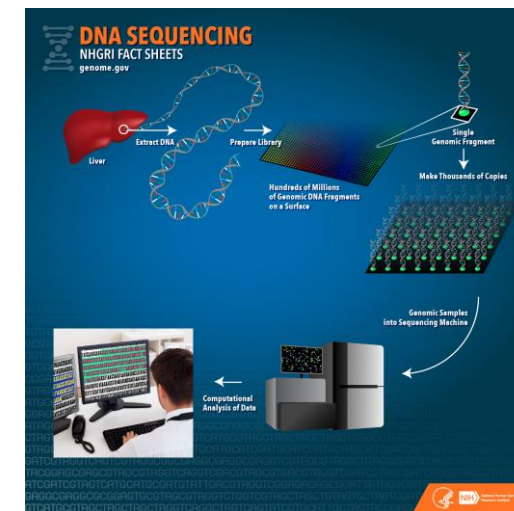
大量データの保管に悩みを抱えている
息子や米国大手企業のCIOからの要望に応える



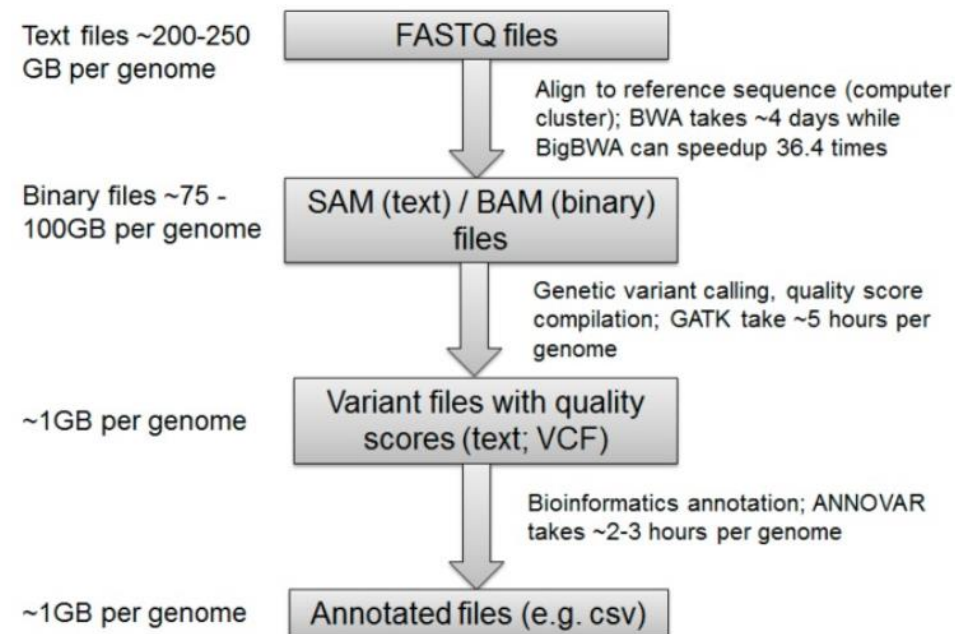
モシェ・ヤナイの息子
イタイ・ヤナイ

ニューヨーク大学 ランゴーン医療センター
計算医学研究所所長
生化学・分子薬理学科教授

ゲノム解析で生成される大量のデータを高速に処理ができコストを抑えて安心して保管できるストレージがなく困っている



ゲノムシーケンサーから収集したDNA情報をデジタル化



INFINIDATのマイルストーン



創業者（現CTE）
モシェ・ヤナイ

2011
2014

シリーズA 投資 製品開発

創業資金調達

EMERGING FROM STEALTH

初号契約はフォーチュン
500の金融、ヘルスケア、
小売企業



シリーズB 投資



TPGによる1億5千万ドルの
資金調達

企業価値は12億ドルに上昇

グローバル展開

USに新本社、EMEAおよび APAC
に新たな拠点を開設
特許申請件数は130件まで増加

2015
2016

ポートフォリの拡張



2017
2018

シリーズC 投資

ゴールドマンサックスおよびTPG
の追加投資による9,500万ドルの
資金調達

企業価値は
16億ドルに上昇



ELASTIC INFRASTRUCTURE ADVANTAGES

- ITインフラのリスク軽減
- 100%可用性保証
- 無停止データ移行
- Elastic Pricing



2019
2020

Scale to Win

Infinidat Elastic DataFabric

更なる製品価値の向上



INFINIDAT 10周年

オールフラッシュ版InfiniBoxを
リリース



2021



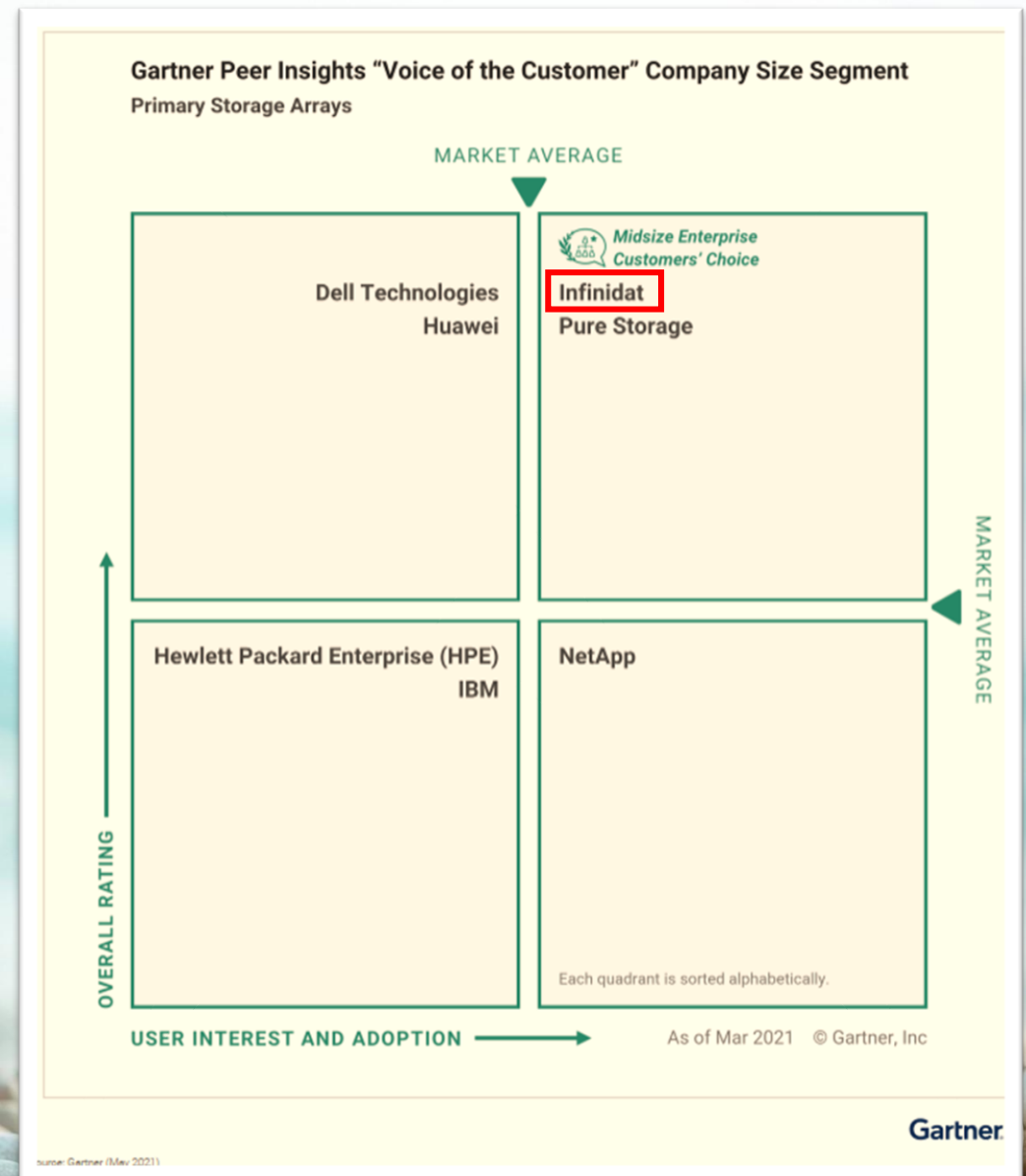
CEO
フィル・ブリンガー



年商50M\$以上の企業で利用されるプライマリストレージの評価
選出されたのはINFINIDATとPureStorageの2社のみ

Recognized as at May 2021 Gartner Peer Insights
Customers' Choice for Primary Storage

<https://www.gartner.com/reviews/market/primary-storage-arrays>



Gartner Peer Insights Customers' Choice badge is a trademark and service mark of Gartner, Inc. and/or its affiliates, and is used herein with permission. All rights reserved. Gartner Peer Insights Customers' Choice constitute the subjective opinions of individual end user reviews, ratings, and data applied against a documented methodology; they neither represent the views of, nor constitute an endorsement by, Gartner or its affiliates.

INFINIDATの設計思想



Flashを超える パフォーマンス

2.0M+ IOPs
95%+のキャッシュヒット率
130 μ s以下のレスポンス



TCO削減

TB単価の削減
低消費電力
CoD



最高レベルの可用性

セブンナイン（99.99999%）の可用性
InfiniRAID, InfiniSnap, Replication



数PBの スケーラビリティ

42Uの1ラックで
最大4.1PBを提供



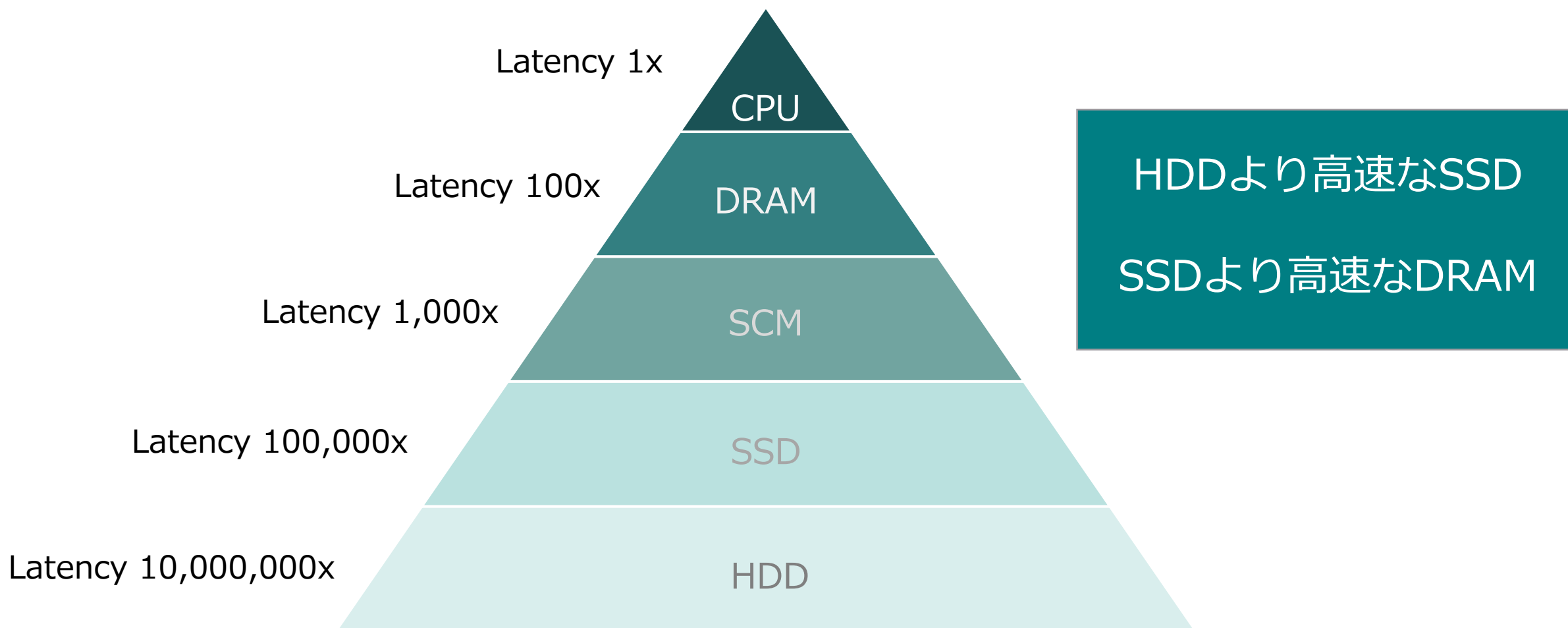
管理

自動化に向けたAPI
サーバー管理者視点のツール

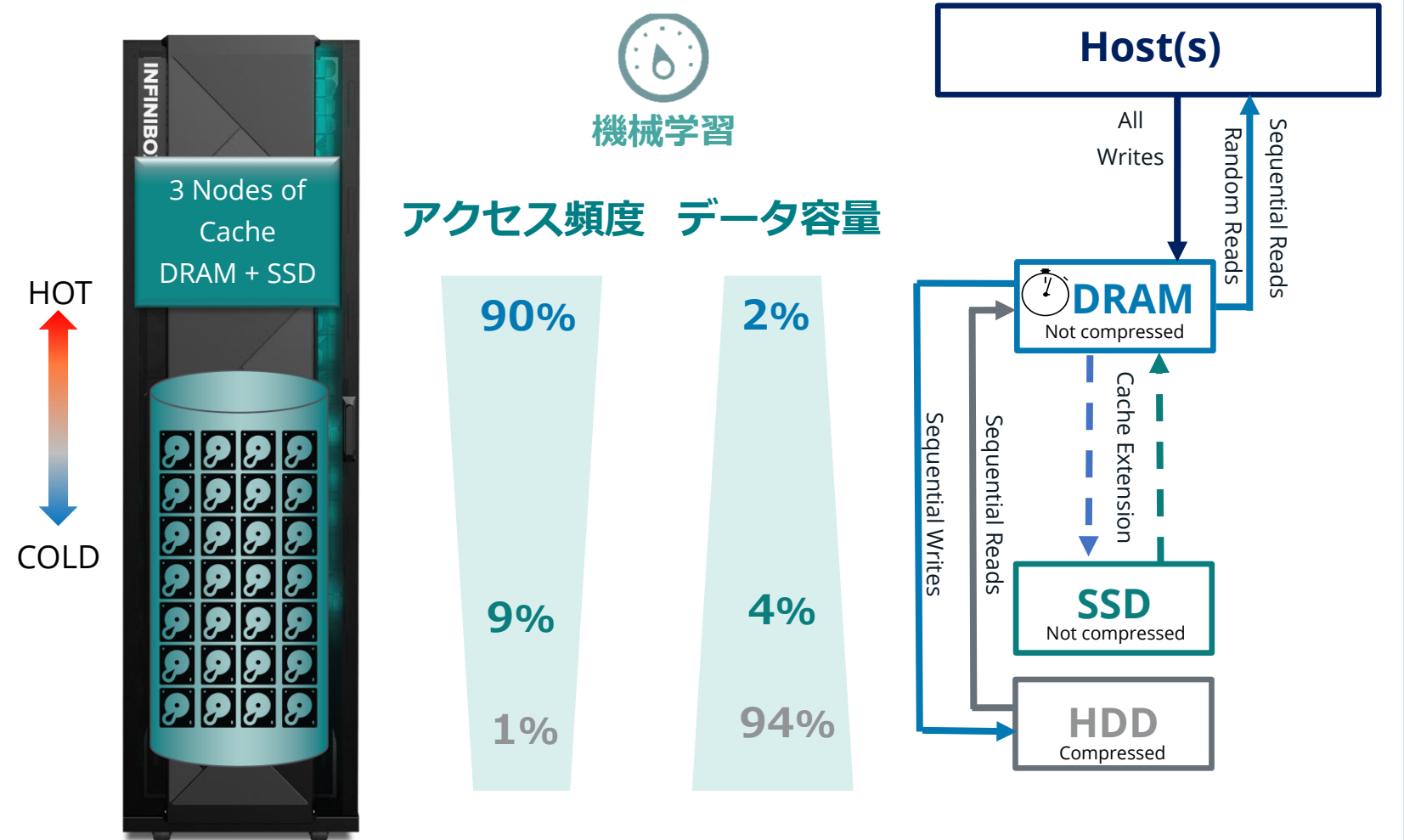


フラッシュを超えるパフォーマンスの実現

各デバイスのパフォーマンス比較



最適化されたIOフロー



Writeキャッシュの属性

- DRAM へ100% 書込み
- 5分以内にデステージ
- ノード間で2重に保護

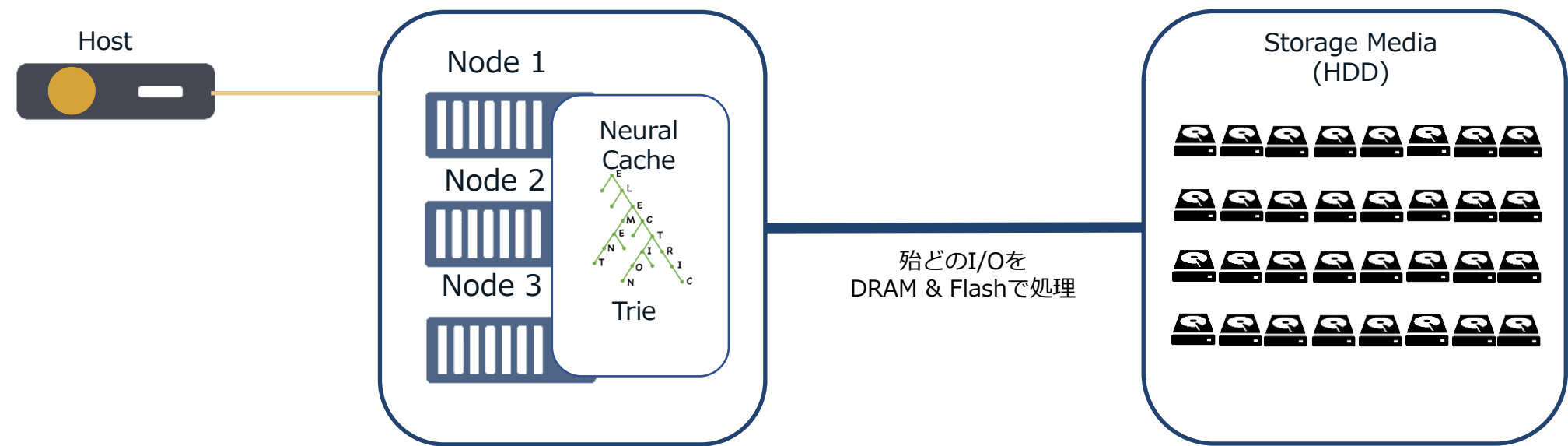
Readキャッシュの属性

- 積極的に先読み
- 1次層と2次層
- 低遅延を実現するために非圧縮

データ保持層の属性

- 最適なデータ配置
- 常時Sequentialアクセス
- 高性能ダブルパリティによるデータ保護

Neuralキャッシュ



InfiniBox特許取得済 Trie

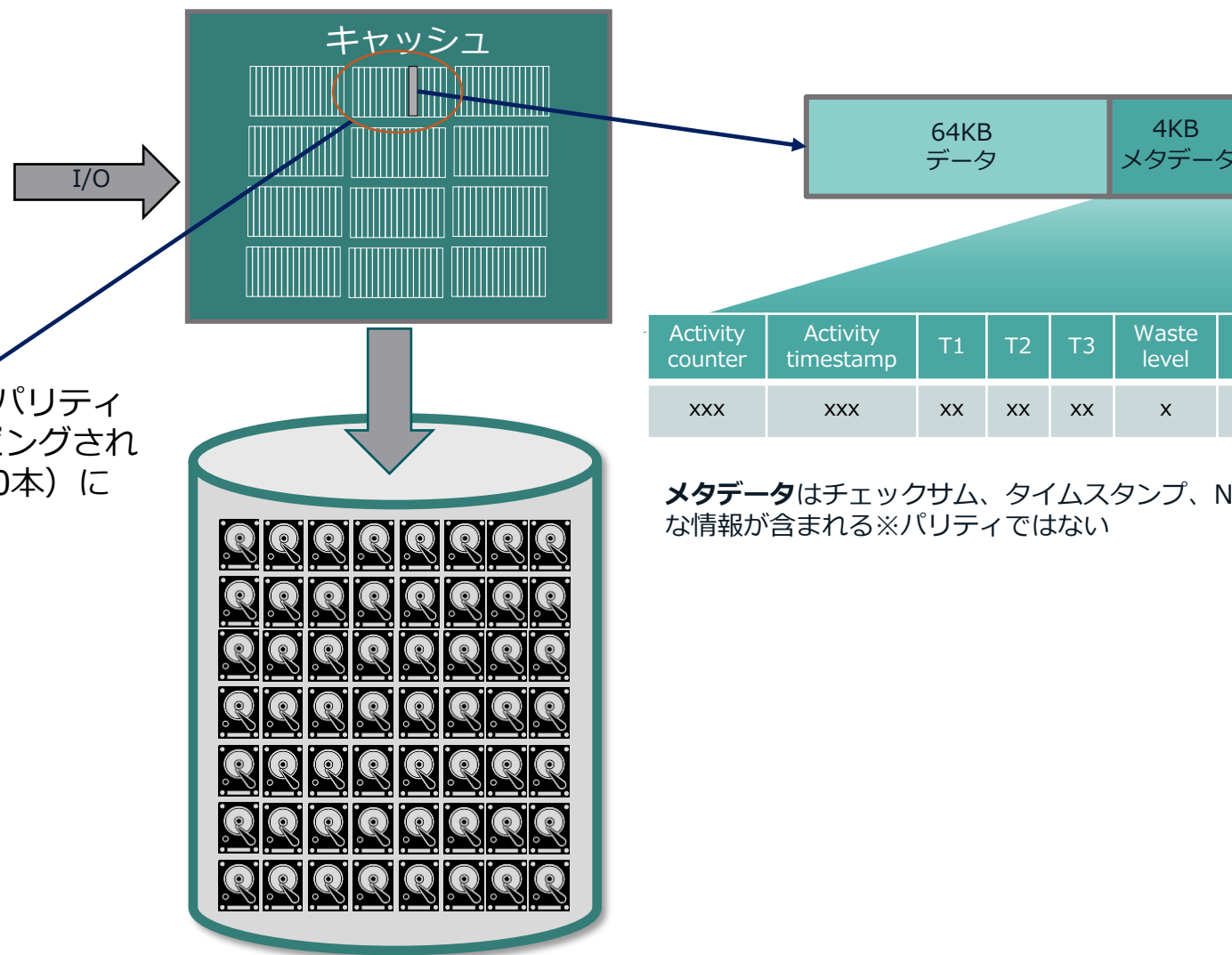
- ・ データ量に依存しないデータ検索時間

マシンラーニング&AIを活用したキャッシュ

- ・ ワークロード特性に合わせたキャッシュメカニズム

非常に高いキャッシュヒット率を実現

データの書き込みイメージ（独自のメタデータ管理）



- 「データセクション×14」と「パリティセクション×2」単位にグルーピングされ全ドライブ（F6000の場合は480本）に分散して書込み
- 書出しはSequential I/Oに変換

メタデータはチェックサム、タイムスタンプ、Neural Cache等で利用するための様々な情報が含まれる※パリティではない

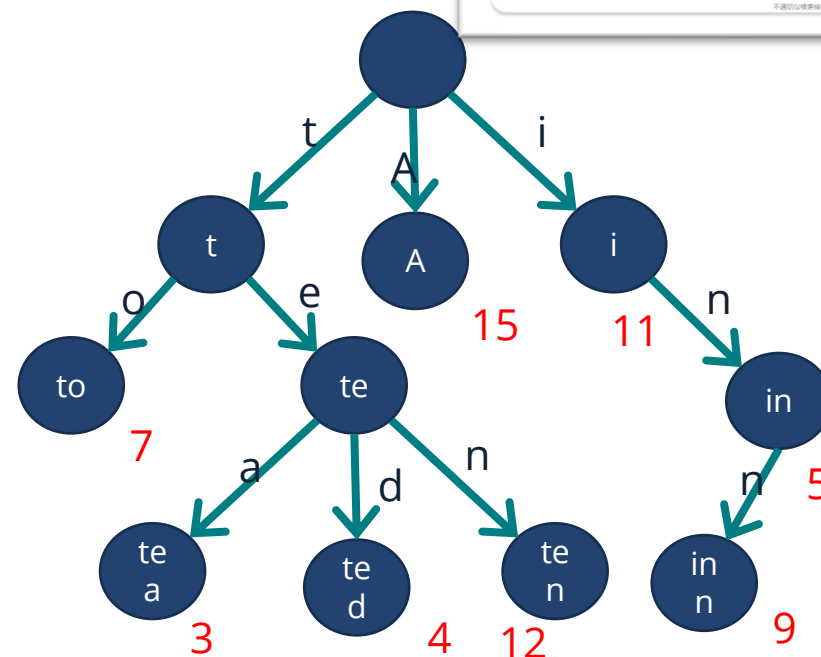
Trie技術をストレージに活用（特許取得済）

■ Trieとは

- 高速にデータを検索するためのデータ構造
- 検索する単語の長さがMの時、検索するための計算量は $O(M)$
- 格納されているデータ量によらず、検索する単語の長さに比例した時間で検索

■ Trie技術をストレージに活用

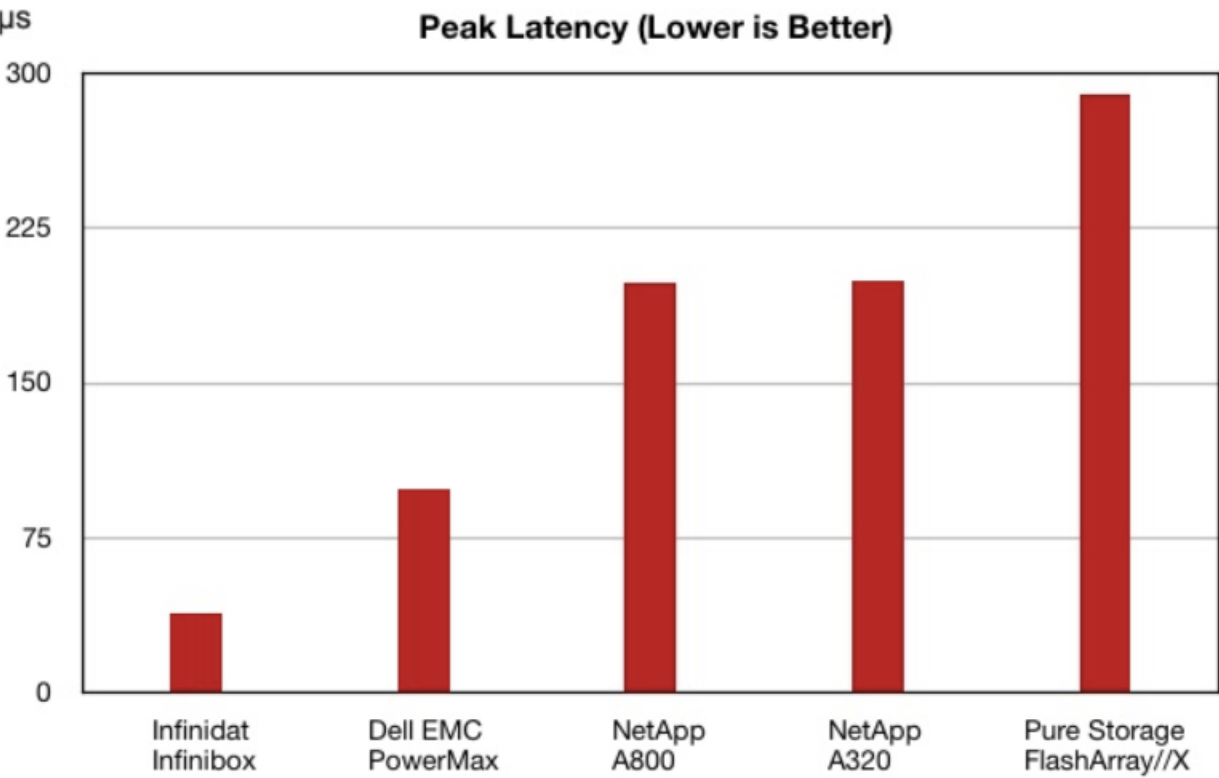
- 大量データの追跡を高速に実行
- メモリ内の処理だけでなく、システム内の全データを追跡するためにTrieを利用
- ホットデータ&コールドデータを正確に判断しより高いキャッシュヒット率を実現



All Flash Storage パフォーマンス比較

「Math > Media」を体現

数学がメディアの壁を打ち破る



過去記事抜粋

<https://blocksandfiles.com/2019/05/08/infinidat-claims-fastest-storage-array-crown/>
<https://www.infinidat.com/blog/dell-please-stop-calling-powermax-worlds-fastest-storage-array>

Storage Performance Latency

Product	Peak Performance Latency	Front-End Technology	Infinidat Advantage
InfiniBox	38 <u>uSec</u>	NVMe-oF RoCE v2	
Dell EMC PowerMax	290 <u>uSec</u> *	FC	7.6x
NetApp AFF A800	"Sub 200 <u>uSec</u> "**	FC-NVMe	5.3x
Pure Storage FlashArray//X	200 <u>uSec</u> ***	NVMe-oF RoCE v2	5.3x

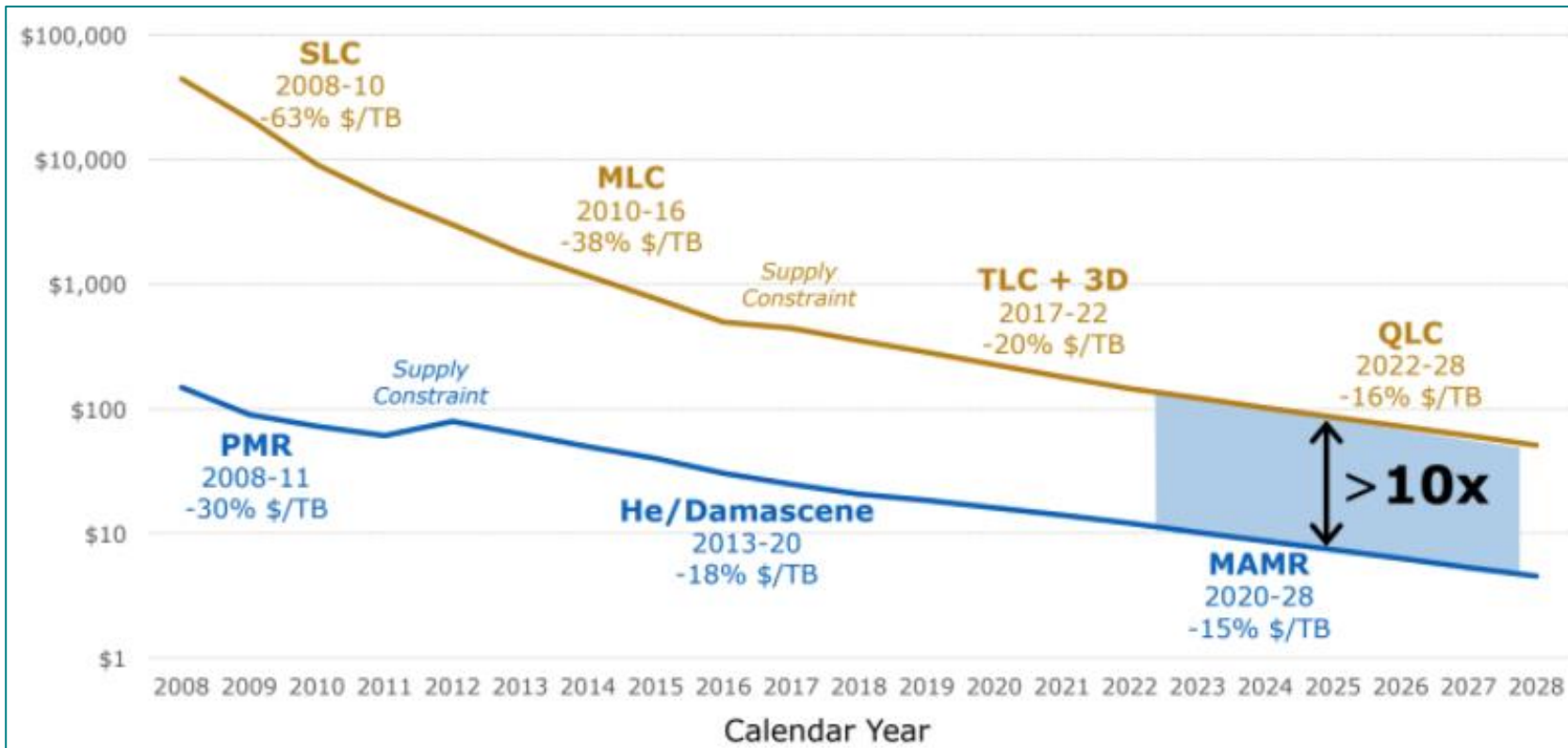


大容量データのコスト最適化

NL-SASとSSDのコスト比較

SSDとSAS HDDの価格差はなくなっているが、NL-SAS HDDの価格差は変わらない

NL-SASとSSDのTBコストの比較



Western Digital, Q4'2017

All Flashベンダーのメッセージ

容量効率化機能で物理容量を
10分の1に削減できます



実際は

VDI（しかもOS部分）なら
可能性はあるが
多くの環境では2分の1も
削減できない

HDDでも
容量効率化機能は利用可能

INFINIBOXのハードウェア

Software Definedで汎用品を利用した低コストの実現

1U	3x ATS: 3重化された電源機構
3U	3x BBU: 3重化されたバッテリーバックアップ
6U	3x Server Nodes: Active/Active/Active - 最大3TBのDRAMを搭載 - 最大368TBのSSDを搭載 (Fx3xx) - 最大207TBのSSDを搭載 (Fx2xx)
32U	最大8台のディスクシェルフ (1シェルフは60ドライブ搭載) 120, 240, or 480 High Capacity Drives



Software Defined
(汎用のIAサーバー、ディスク
バッテリーで構成)

大容量・低コストの
NL-SASを活用しながらも
高パフォーマンス

独自の容量効率化機能

InfiniBoxによるマルチPBシステムの統合効果



VS.



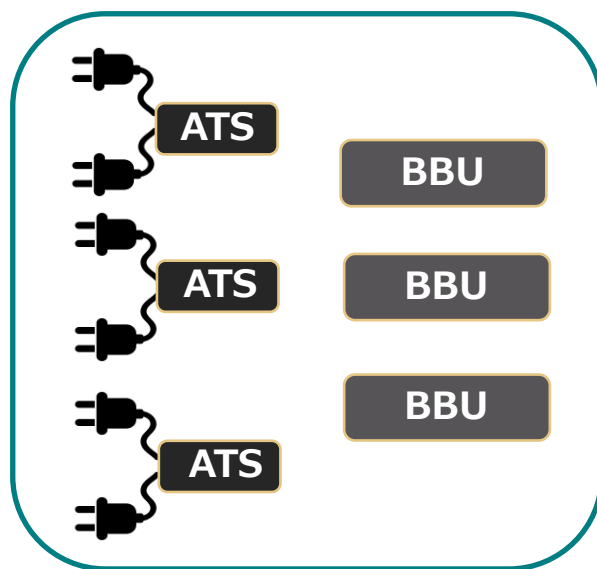
	2 x EMC VMAX 40K 6 x XIV Gen3 (14 racks)	InfiniBox (2 racks)
Power	50.6 KW	16 KW
Space	28 Floor Tiles	2 Floor Tiles
Availability	99.999% (Five Nines)	99.99999% (Seven Nines)
Data Protection	RAID-5 RAID-6	InfiniRAID
Usable Capacity	~1.8 PB	2.4 PB
Relative CapEx Cost	\$3k per TB	\$1k per TB



最高レベルの可用性と信頼性

セブンナイン（99.99999%）業界No.1の可用性

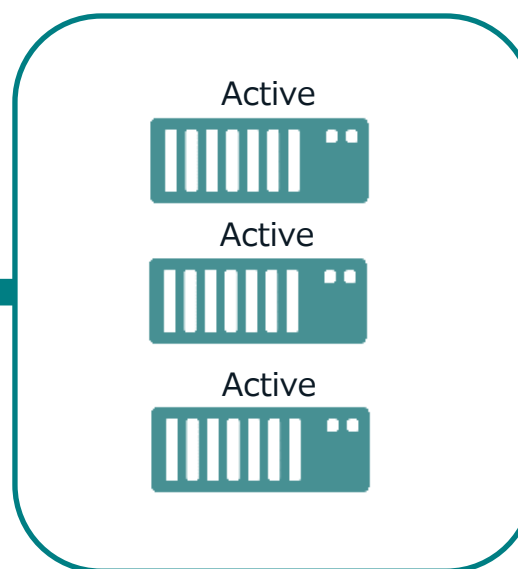
3重化された ATS/BBU/Nodes



N+2 の電源機構

- 2重障害のケースに対応

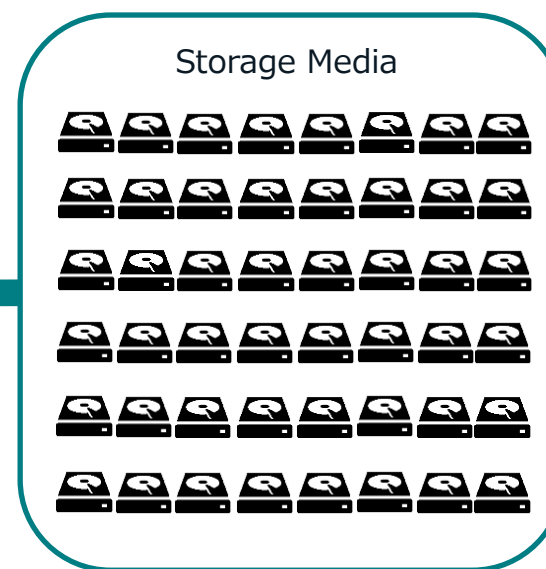
マルチレイヤキャッシュによりバランスが取れた稼働



高い可用性を実現するフロントエンドの接続性

- Active-Active-Active コントローラノード
- 全ノードが全ドライブに接続
- ソフトウェア・デファインドのFCターゲットとIP アドレス

Storage Media



InfiniRAID

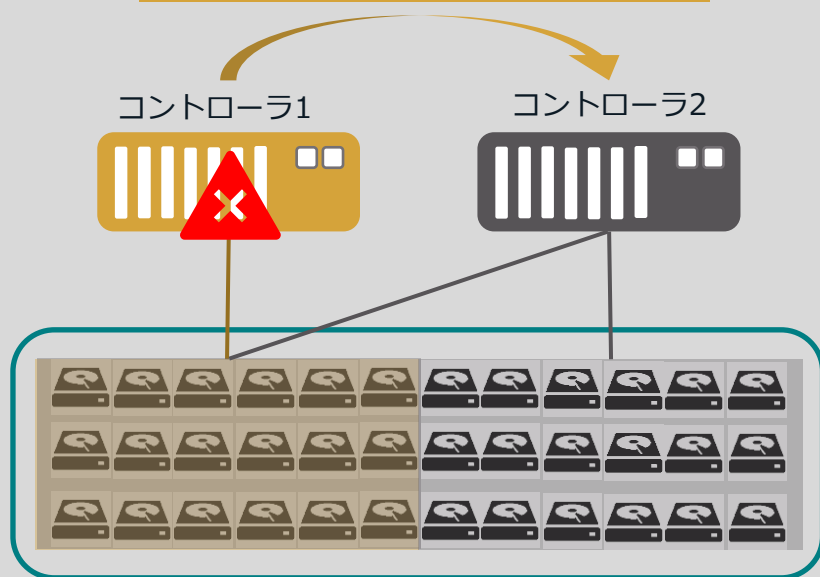
- 14Data + 2Parity単位で最大480本のディスクにデータを分散
- 複数ドライブの障害に対応
- リビルド時間は最大でも15分
- スペアディスク不要

「3-Active Shared Everything」方式採用

障害時のフェイルオーバー時間の短縮およびフェイルオーバー障害を回避

Shared Nothing 方式

コントローラ1障害時はサービス or
コントローラのフェイルオーバーが発生



通常時

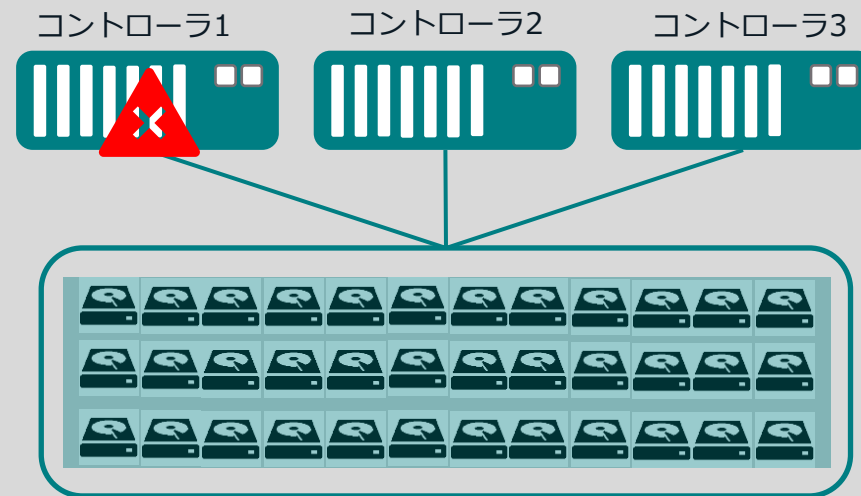
- ・各コントローラはオーナーディスク（ボリューム）のみアクセス可能

障害時

- ・障害コントローラのサービスを正常コントローラが引継ぎ、障害コントローラのオーナーディスク（ボリューム）もアクセス

3-Active Shared Everything 方式

コントローラ1障害時もコントローラ2
とコントローラ3はそのままActive
（IPアドレスのFailoverはあり）



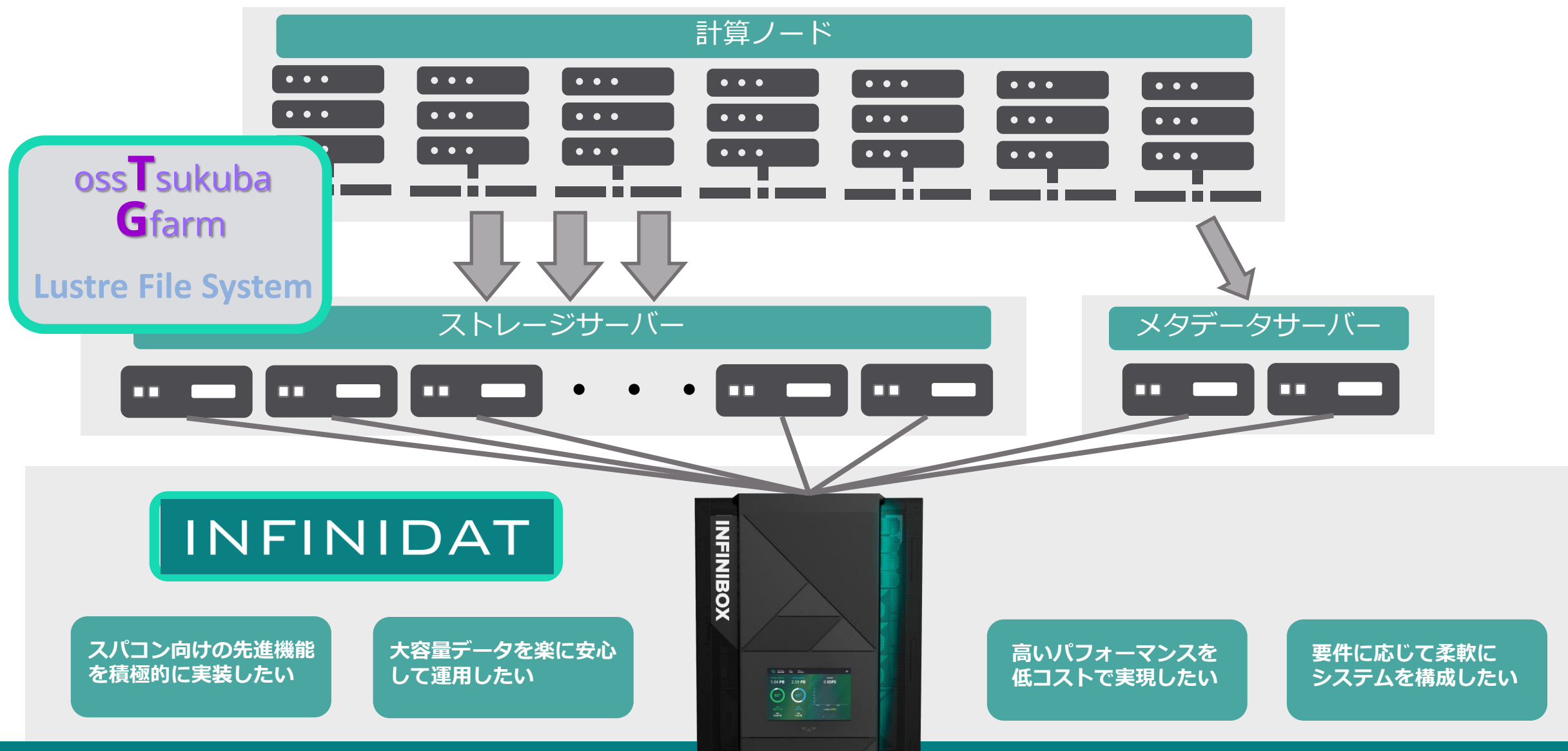
通常時

- ・各コントローラはどのディスク（ボリューム）にアクセス可能

障害時

- ・障害コントローラから正常コントローラにマルチパスの仕組みで引き継ぎ（サービス引継ぎはなし）
- ・IPアドレスのFailoverはあり

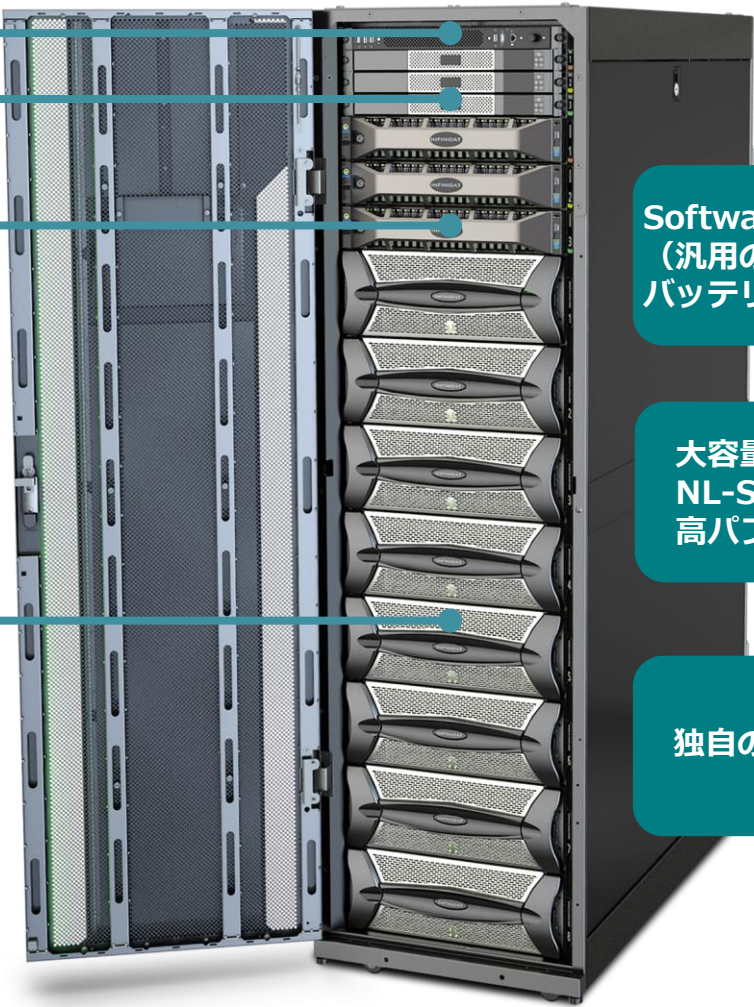
次世代HPCプラットフォームを支えるブロックストレージ



補足資料

InfiniBoxのハードウェア

1U	3x ATS : 3重化された電源機構
3U	3x BBU : 3重化されたバッテリーバックアップ
6U	3x Server Nodes : Active/Active/Active - 最大3TBのDRAMを搭載 - 最大368TBのSSDを搭載 (Fx3xx) - 最大207TBのSSDを搭載 (Fx2xx)
32U	最大8台のディスクシェルフ (1シェルフは60ドライブ搭載) 120, 240, or 480 High Capacity Drives

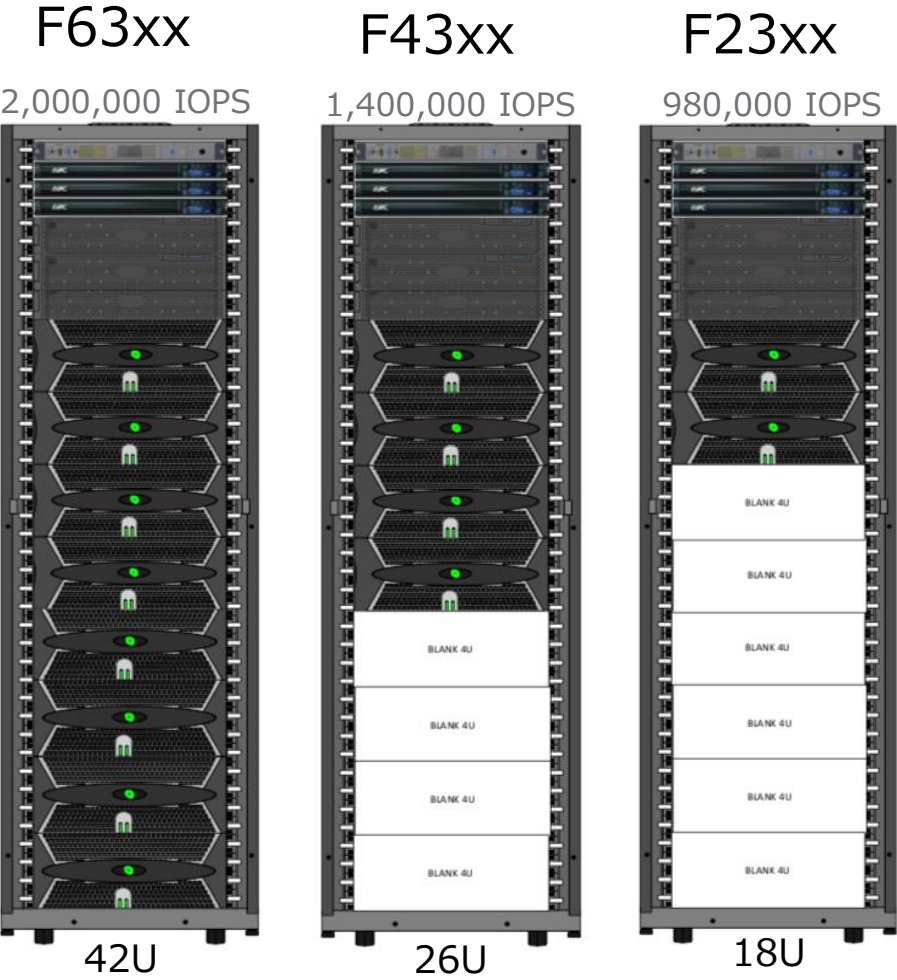


Software Defined
(汎用のIAサーバー、ディスク
バッテリーで構成)

大容量・低コストの
NL-SASを活用しながらも
高パフォーマンス

独自の容量効率化機能

Fx3xx モデル一覧



Series	Model	Usable Capacity	Effective Capacity	DRAM	Power & Cooling
F63xx	6312	4.149 PB	10.372 PB	3.072 TB	8 kW @ 27,500 BTU/H
	6308	2.766 PB	6.915 PB		
	6306	2.074 PB	5.185 PB	2.304 TB	
	6304	1.383 PB	3.457 PB	2.304 TB	
	6303	1.037 PB	2.592 PB	1.536 TB	
F43xx	4312	2.050 PB	5.125 PB	2.304 TB	5.5 kW @ 18,500 BTU/H
	4308	1.367 PB	3.417 PB		
	4306	1.025 PB	2.562 PB	1.536 TB	
	4304	0.683 PB	1.707 PB	1,152 GB	
	4303	0.512 PB	1.280 PB		
F23xx	2306	0.499 PB	1.247 PB	1,152 GB	3.5 kW @ 9,500 BTU/H
	2304	0.333 PB	0.832 PB	768 GB	
	2303	0.249 PB	0.622 PB	768 GB	

F63xx モデル一覧

IOPS	Throughput	Max DRAM	Power	Cooling	Form Factor	Max Weight
2M**	25 GB/s	3.072 TB	8 KW	27,300 BTU/Hour	42U in a Standard 19" Rack	1,176kg

各モデル仕様

Model	DRAM Capacity	Usable Capacity	Flash Cache	Ethernet Ports	FC Ports	Number of SSDs per node	Encryption @ Rest Capable
6312	3.072 TB	4.149 PB	0, 28, 46, or 92 TB 184 TB config is under Directed Availability and requires management approval	18 × 10GigE	24 × 16Gbps	0, 5, 8, or 16	Yes*
6308	3.072 TB	2.766 PB					
6306	2.304 TB	2.074 PB					
6304	2.304 TB	1.383 PB					
6303	1.536 TB	1.037 PB					

*Encryption At Rest feature (introduced in 2.2) - the system's ability to encrypt customer data on the enclosure drives and local drives (HDDs only). In order for the system support this feature, it must be selected in the Site Survey. The system does not support this by default.
**NFSv3 = 2M IOPS; FC/iSCSI = 1.4M IOPS

F43xx モデル一覧

IOPS	Throughput	Max DRAM	Power	Cooling	Form Factor	Max Weight
1.4M**	20 GB/s	2.304 TB	5.5 KW	19,110 BTU/Hour	26U in a Standard 19" Rack	795kg

各モデル仕様

Model	DRAM Capacity	Usable Capacity	Flash Cache	Ethernet Ports	FC Ports	Number of SSDs per node	Encryption @ Rest Capable
4312	2.304 TB	2.050 PB	0, 28, 46, or 92 TB 184 TB config is under Directed Availability and requires management approval	18 × 10GigE	24 × 16Gbps	0, 5, 8, or 16	Yes*
4308	2.304 TB	1.367 PB					
4306	1.536 TB	1.025 PB					
4304	1.152 TB	0.683 PB					
4303	1.152 TB	0.512 PB					

*Encryption At Rest feature (introduced in 2.2) - the system's ability to encrypt customer data on the enclosure drives and local drives (HDDs only). In order for the system support this feature, it must be selected in the Site Survey. The system does not support this by default.
**NFSv3 = 1.4M IOPS; FC/iSCSI = 1.0M IOPS

F23xx モデル一覧

IOPS	Throughput	Max DRAM	Power	Cooling	Form Factor	Max Weight
980K***	14 GB/s	1.152 TB	3.6 KW	12,300 BTU/Hour	18U in a Standard 19" Rack	583kg

各モデル仕様

Model	DRAM Capacity	Usable Capacity	Flash Cache	Ethernet Ports	FC Ports	Number of SSDs per node	Encryption @ Rest Capable
2306	1.152 TB	0.499 PB	0, 28, 46, or 92 TB	18 × 10GigE	24 × 16Gbps	0, 5, 8, or 16	Yes*
2304	768 GB	0.333 PB	184 TB config is under Directed Availability and requires management approval				
2303	768 GB	0.249 PB					

*Encryption At Rest feature (introduced in 2.2) - the system's ability to encrypt customer data on the enclosure drives and local drives (HDDs only). In order for the system support this feature, it must be selected in the Site Survey. The system does not support this by default.
***NFSv3 = 980K IOPS; FC/iSCSI = 700K IOPS

THANK YOU!



INFINIDAT

SCALE TO WIN