

Gfarm Performance on Azure

*Gfarm*シンポジウム2016

2016年12月9日
株式会社H P Cソリューションズ
河野 証



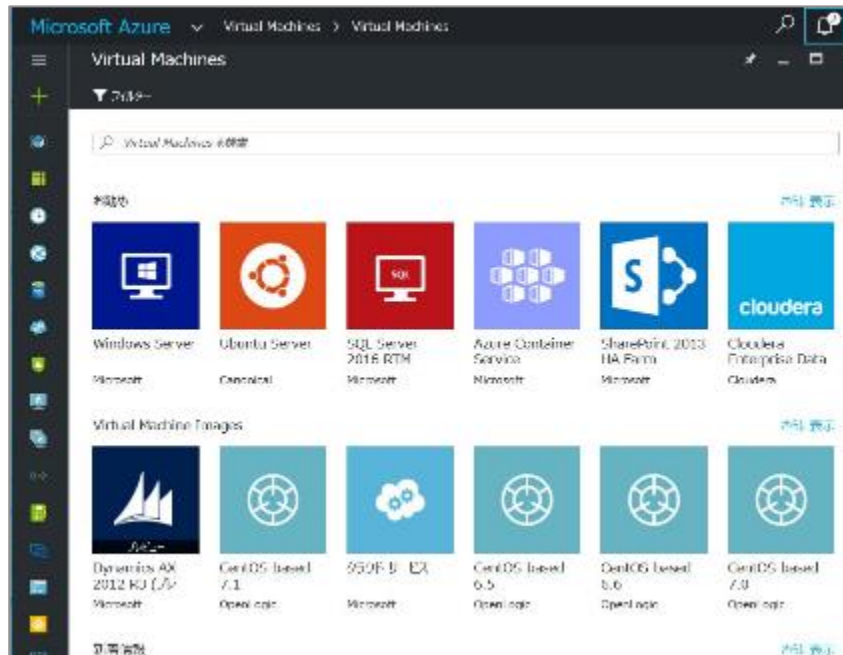
Gfarm Performance on Azure

内容

1. Microsoft Azure概要
2. クラウド環境におけるGfarmの利用イメージ
3. 性能テスト

Microsoft Azure（アジュール）概要

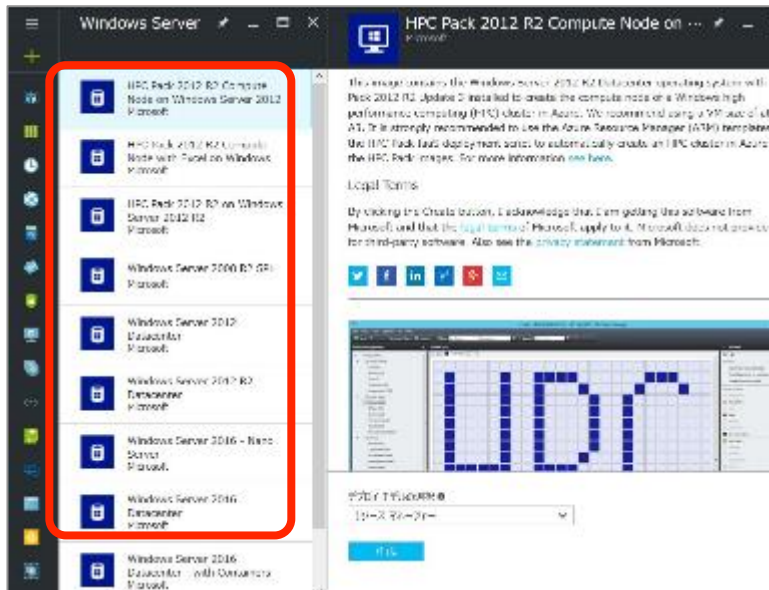
Microsoft Azure『仮想マシン』サービス



- Windows Server あるいは 各種 Linux の「OS 導入済み仮想マシンイメージ」が多数用意されており、それらを選択することでサーバー環境を容易に構築することが可能です。
- また、必要なアプリケーションを導入するなどカスタマイズを施した仮想マシンを「自分用のカスタムイメージ」として登録できます。
- HPC Pack を導入済みの仮想マシンイメージも用意してあります。

様々な仮想マシンイメージをご提供

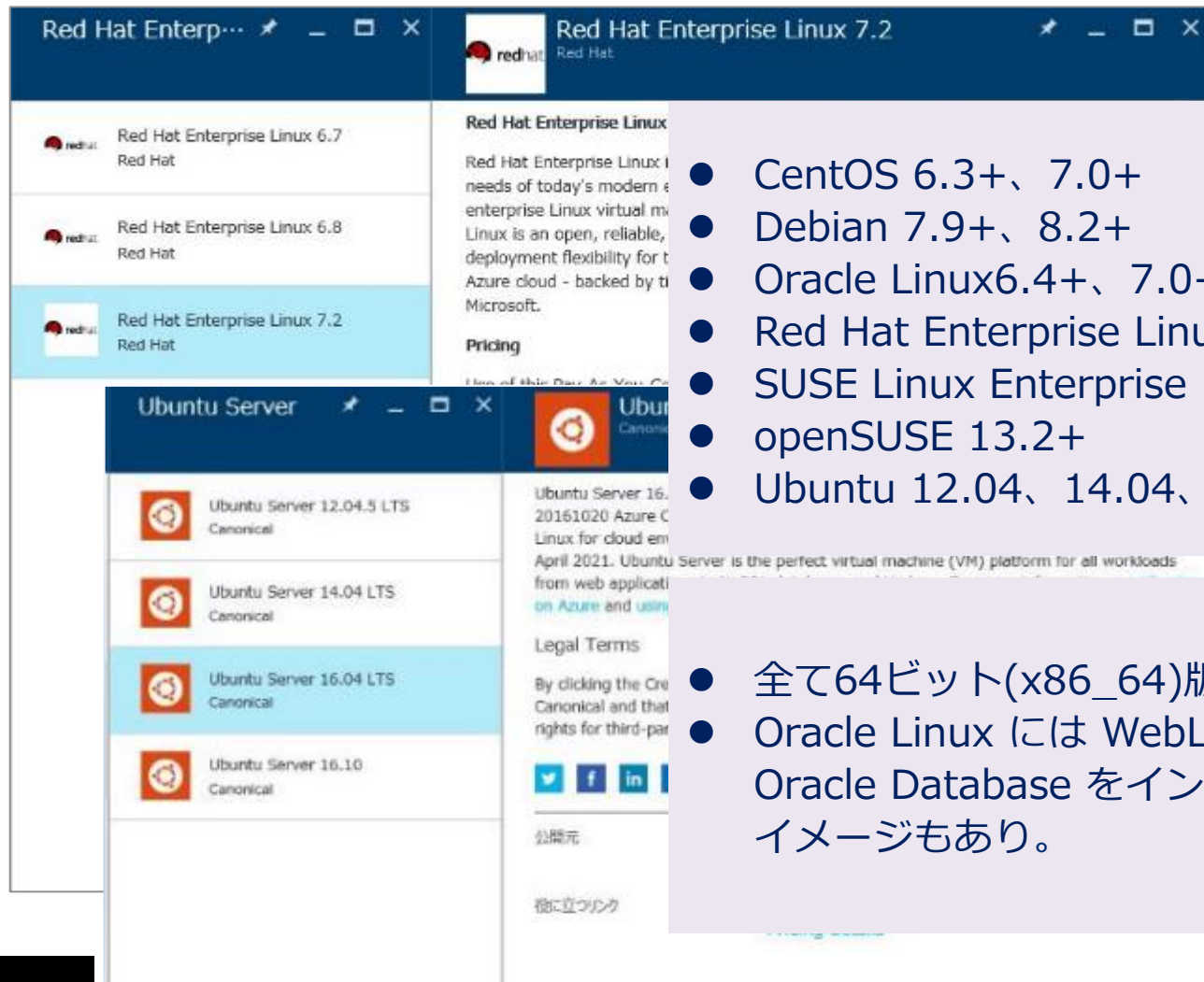
基本的な Windows Server



- Windows Server 2016
- Windows Server 2012 R2
- Windows Server 2012
- Windows Server 2008 R2 SP1
- Windows Server Essentials Experience
- HPC Pack

- OS インストール済み。すぐに使い始められます。
- Windows Server のライセンスは仮想マシンの料金に含まれており、別途調達は不要です。
(Windows Server CALも含まれています)
- OS は英語版ですが、日本語化可能です。
- 日本語化などのカスタマイズを施した上で「マイ イメージ」として登録しておけば、2台目以降の展開が楽になります。

様々な仮想マシンイメージをご提供 複数の Linux ディストリビューションを用意



- CentOS 6.3+、7.0+
- Debian 7.9+、8.2+
- Oracle Linux 6.4+、7.0+
- Red Hat Enterprise Linux 6.7+、7.1+
- SUSE Linux Enterprise 11 SP4、SLES 12+
- openSUSE 13.2+
- Ubuntu 12.04、14.04、および 16.04

- 全て64ビット(x86_64)版
- Oracle Linux には WebLogic Server や Oracle Database をインストール済みのイメージもあり。

Azure VMの仮想CPUは物理コアを占有



- オンプレミスの Hyper-V と異なり、Azure 上の仮想マシンは、LP と VP の比が 1:1 です。
- つまり、物理的なコアが一つ丸ごと、仮想マシン上の CPU コアになります。
- 「他の仮想マシン」とコアを共有することがないため、物理コアの性能をすべて享受できます。

※ ただし、A0 (XS)サイズだけは例外で、複数の仮想マシンでコアを共有します。
(LP:VP が 1:Nになります)

Azureリージョン



場所

ワールドワイドに30リージョン
(リージョンは増設進行中)

日本国内に2つのリージョン

- ・ 東日本
- ・ 西日本

アメリカ

リージョン	場所
米国東部	バージニア州
米国東部 2	バージニア州
米国中部	アイオワ州
米国中部	イリノイ州
米国中部	テキサス州
米国中部	米田中部
米国西部	カリフォルニア州
米国西部 2	米国西部 2
米国政府バージニア	バージニア州
米国政府アイオワ	アイオワ州
カナダ東部	ケベックシティ
カナダ中部	トロント
ブラジル東部	サンパウロ州

新規発表

米国防衛東部	後日発表予定
米国防衛西部	後日発表予定

ヨーロッパ

リージョン	場所
北ヨーロッパ	アイルランド
西ヨーロッパ	オランダ
ドイツ中部	フランクフルト
ドイツ北東部	マクデブルク
英国西部	ハーディフ
英国南部	ロンドン

アジア

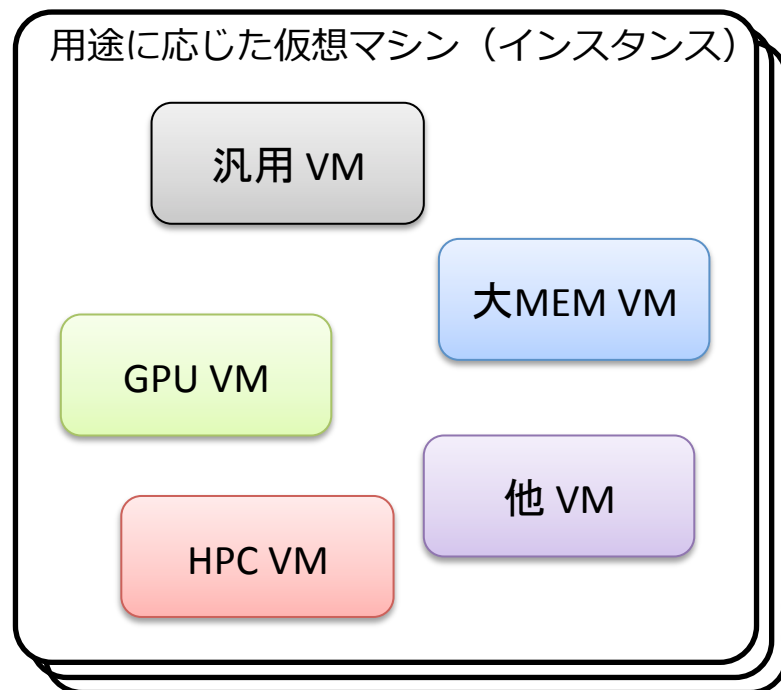
リージョン	場所
東南アジア	シンガポール
東アジア	香港
オーストラリア東部	ニュー・サウス・ウェールズ州
オーストラリア東部	ビクトリア州
インド中部	ブネー
インド西部	ムンバイ
インド南部	チェンナイ
東日本	東京、埼玉
西日本	大阪
中国東部	上海
中国北部	北京

新規発表

韓国中部	ソウル
韓国南部	後日発表予定

仮想マシンと仮想ストレージ(1)

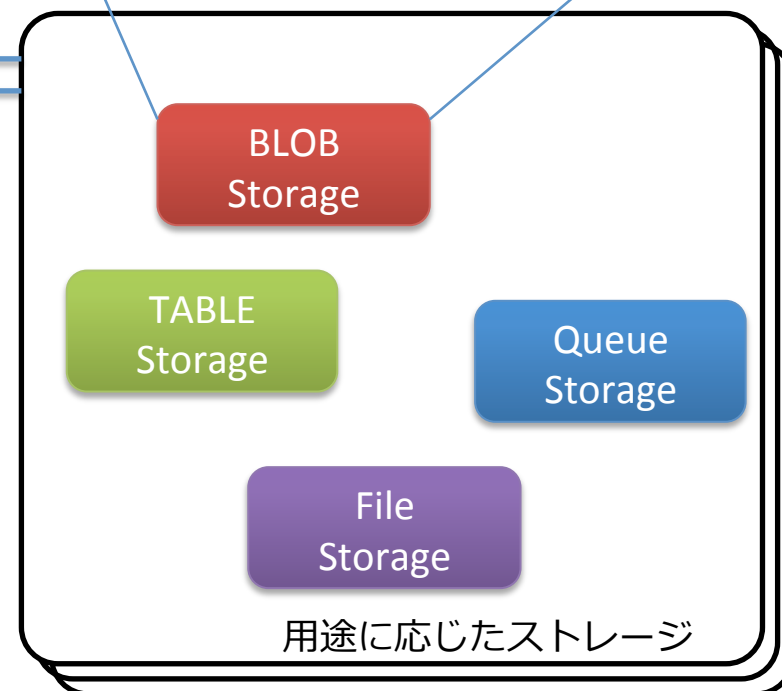
概要



- 汎用 VM
E5プロセッサシリーズ
- 大MEM VM
～ 448GB RAM
- GPU VM
GPGPU搭載
- HPC VM
高クロックCPU
Infiniband対応

- 性能オプション
Standard: HDD
Premire: SSD
- 冗長オプション
LRS : 3xコピー(同一施設内) デフォルト
~~ZRS : 3xコピー(1-2リージョン、2-3施設内)~~
GRS : 6xコピー(プライマリリージョンに3
セカンダリリージョンに3)
RA-GRS : GRS+セカンダリリージョンへの
Read-only Access

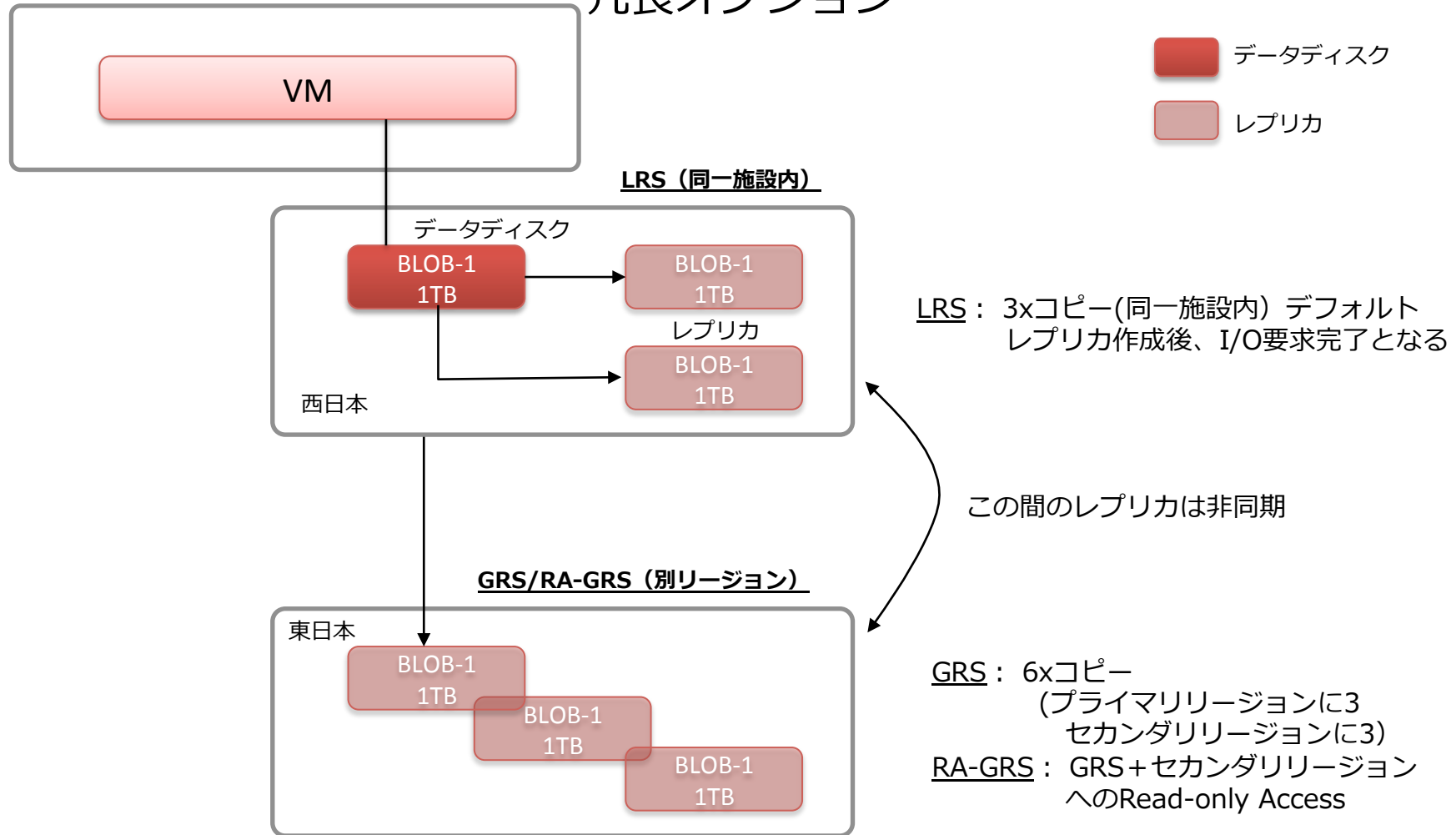
...



用途に応じたストレージ

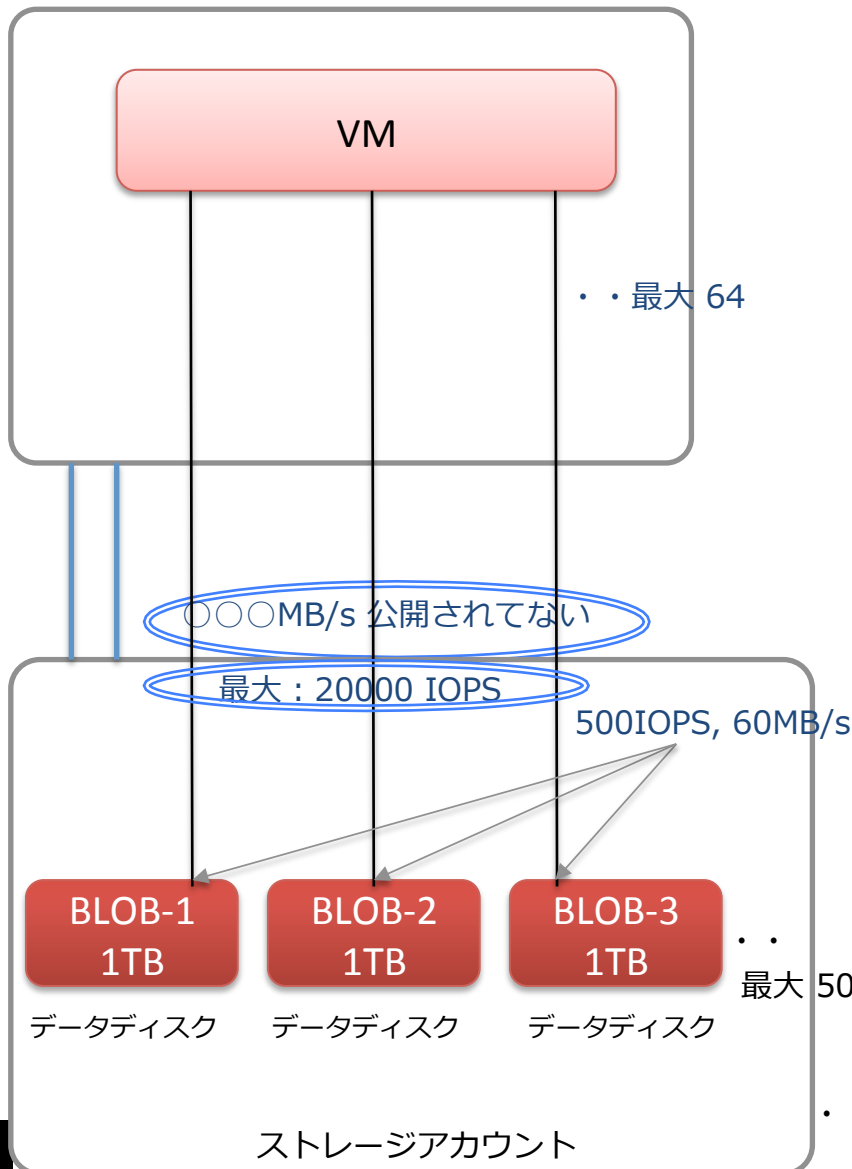
...

仮想マシンと仮想ストレージ(2) 冗長オプション



※ 日本国内では、東日本と西日本が互いにプライマリ/セカンダリ リージョンになります。

仮想マシンと仮想ストレージ(3) データディスク

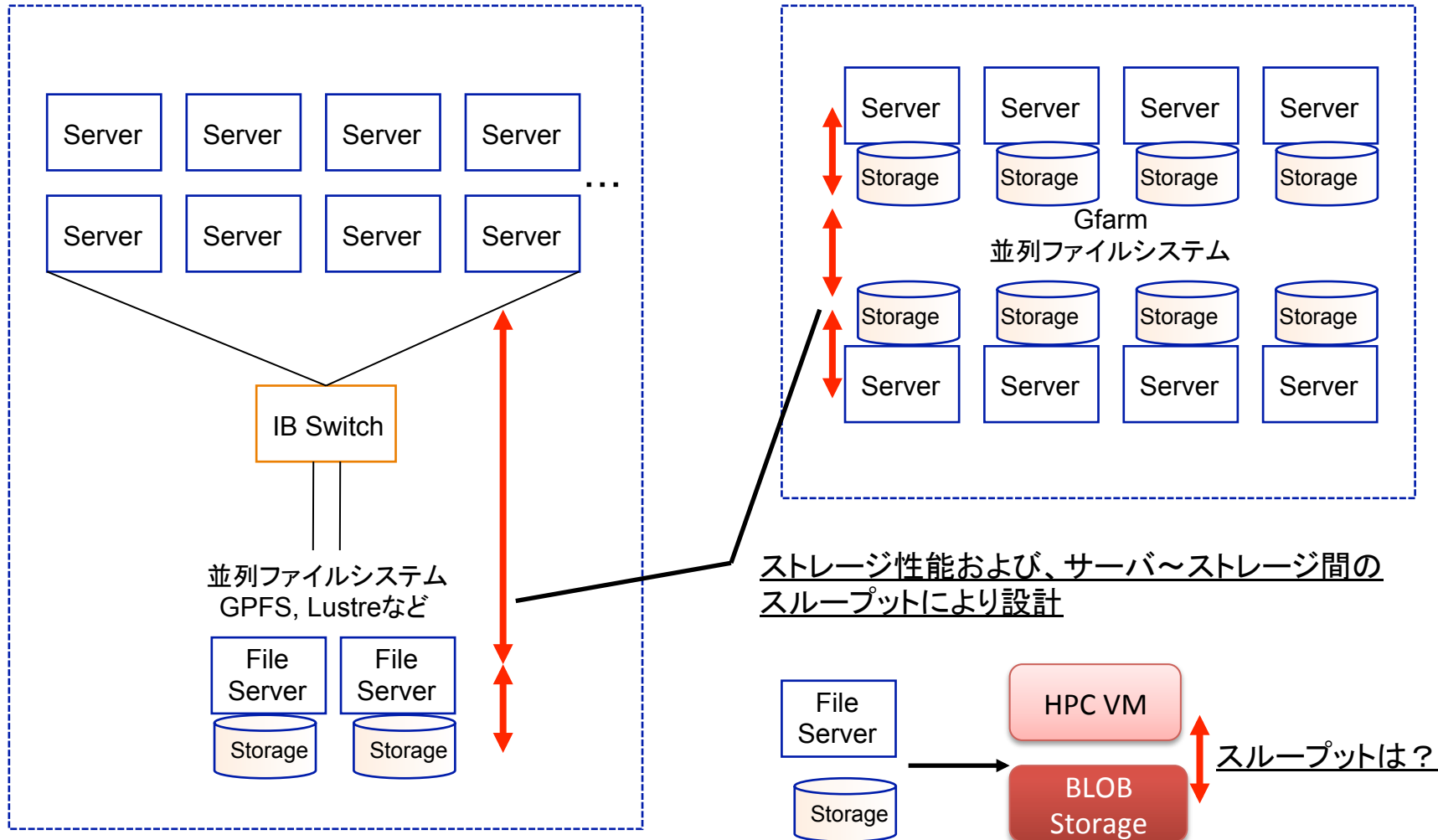


- マウント数
VM（インスタンス）毎に異なる（1～64ディスク）
- ストレージアカウント
HDD：データディスクあたり
最大 500 IOPS または 60MB/s
SSD：データディスクあたり
最大 5000 IOPS または 200MB/s
 - ・1ストレージアカウントで4つのBLOB領域を作成した場合、4つの合計が性能割り当て値となる
 - ・1ストレージアカウントは、最大 20000 IOPS
 - ・サブスクリプション（課金単位）あたりの最大ストレージアカウントは、200個まで
- データディスクサイズ
HDD, SSDいずれも最大約1TB
 - ・・・最大200個まで

クラウド環境におけるGfarmの利用イメージ

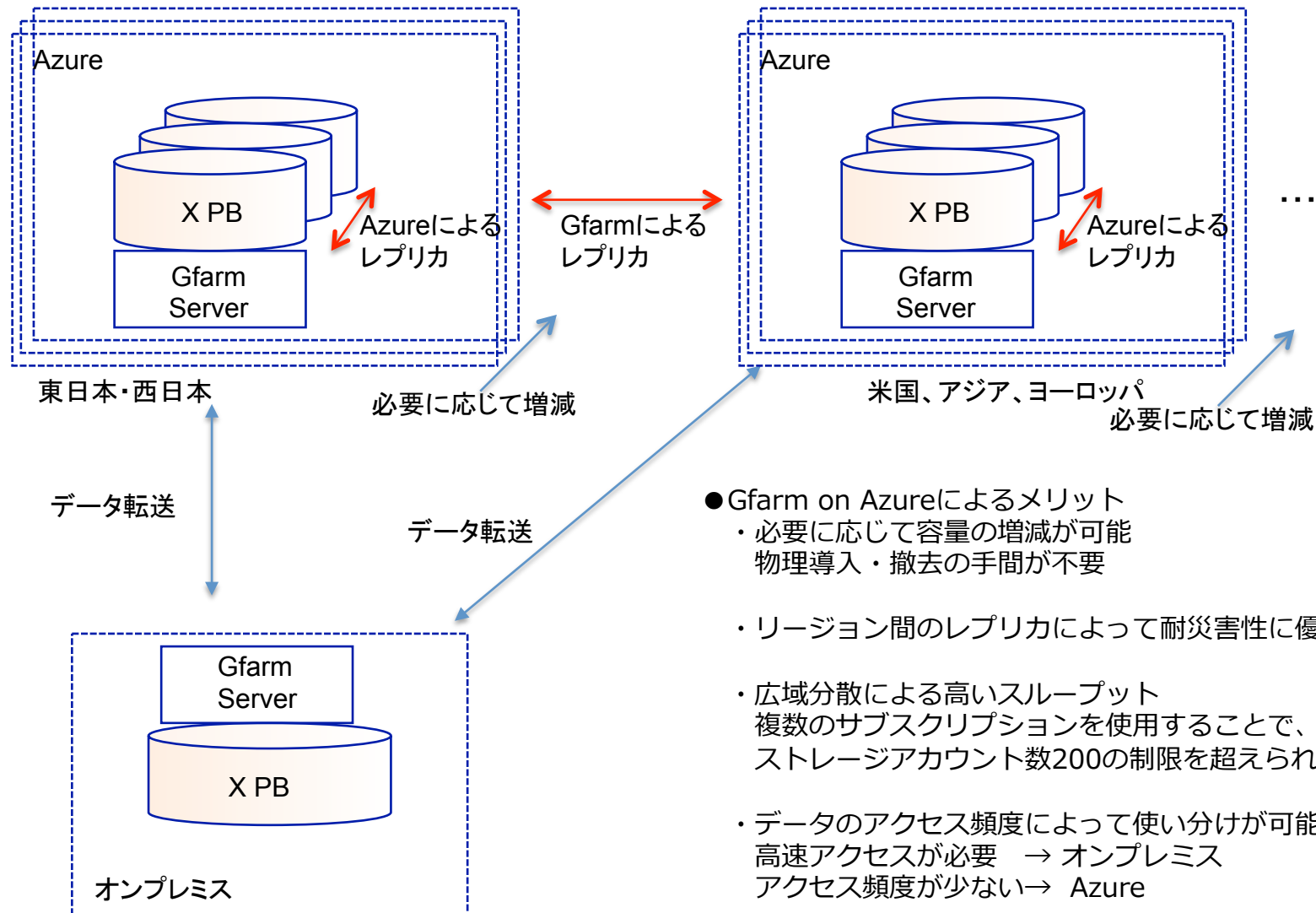
HPCクラスタシステムの並列ファイルシステム

HPCクラスタシステムの並列ファイルシステム



広域ストレージ領域

広域ストレージ領域



● Gfarm on Azureによるメリット

- ・ 必要に応じて容量の増減が可能
物理導入・撤去の手間が不要
- ・ リージョン間のレプリカによって耐災害性に優れる
- ・ 広域分散による高いスループット
複数のサブスクリプションを使用することで、
ストレージアカウント数200の制限を超えられる
- ・ データのアクセス頻度によって使い分けが可能
高速アクセスが必要 → オンプレミス
アクセス頻度が少ない → Azure
- ・ オンプレミスシステム入れ替え時のデータ移行の軽減
- ・ 1 オンプレミスとの連携以外に、公共データの公開にも向く

前回確認したこと

1. Gfarm on Azure

Azure上および、オンプレ + Azure環境で Gfarmが容易に構築できることが確認できた。
Gfarmの基本機能も問題無く動作確認できた。

2. HPCクラスタシステム用並列ファイルシステムとしてのGfarm

問題無く動作確認できた。

データディスクの刻みが1TB, HDD:60MB/s、SSD:200MB/sであるため、これを束ねた時のピーク性能の計測は今後の課題

3. 性能

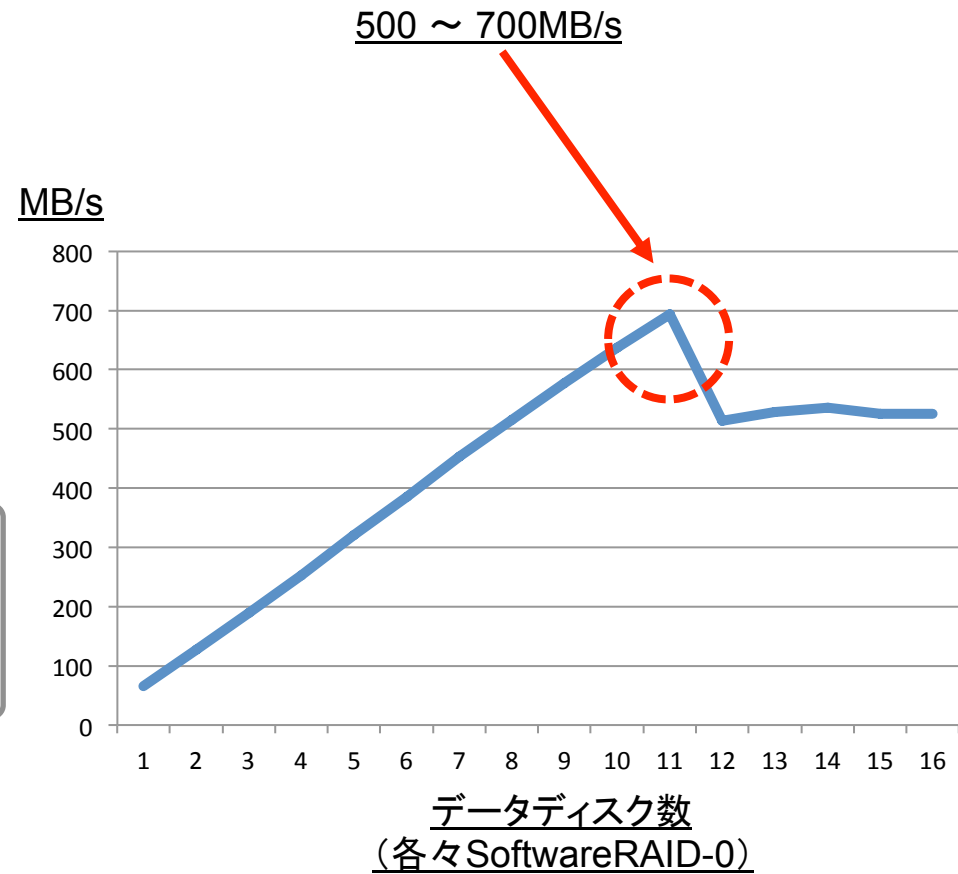
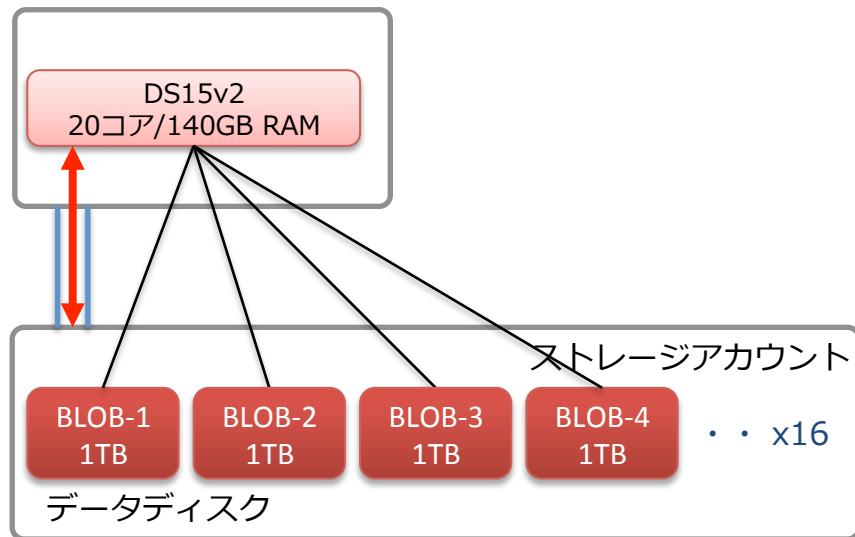
基本動作確認程度

今後はI/Oスループットの設計ができるレベルの計測結果と構成情報の整理が必要

性能テスト

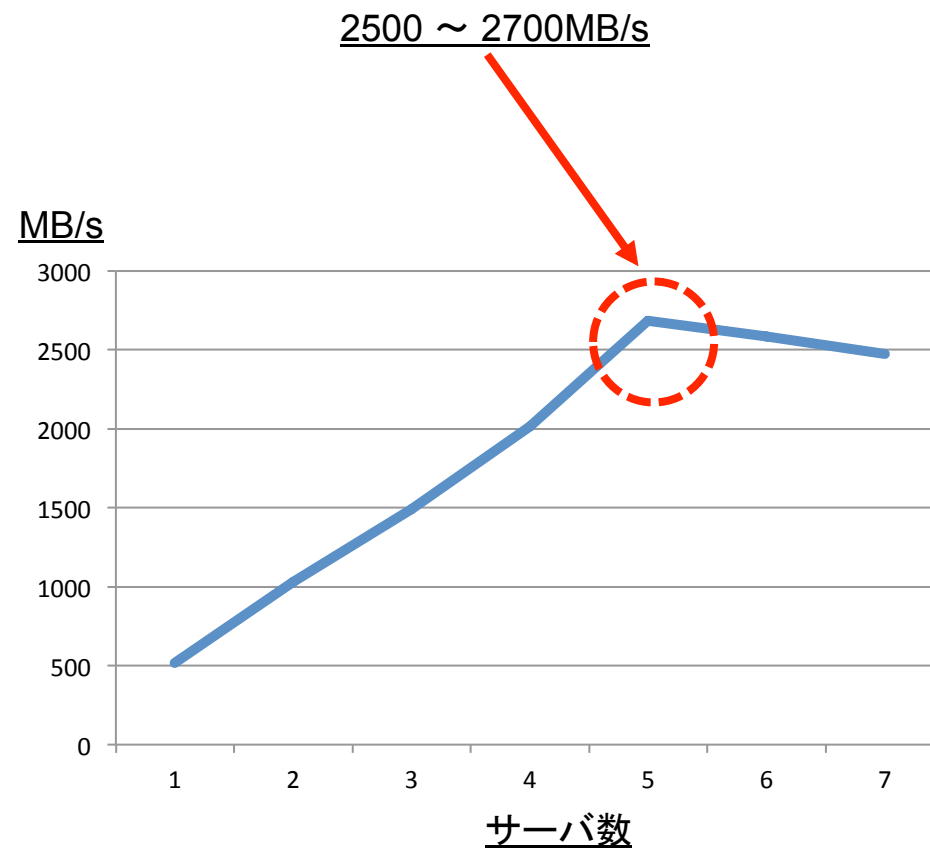
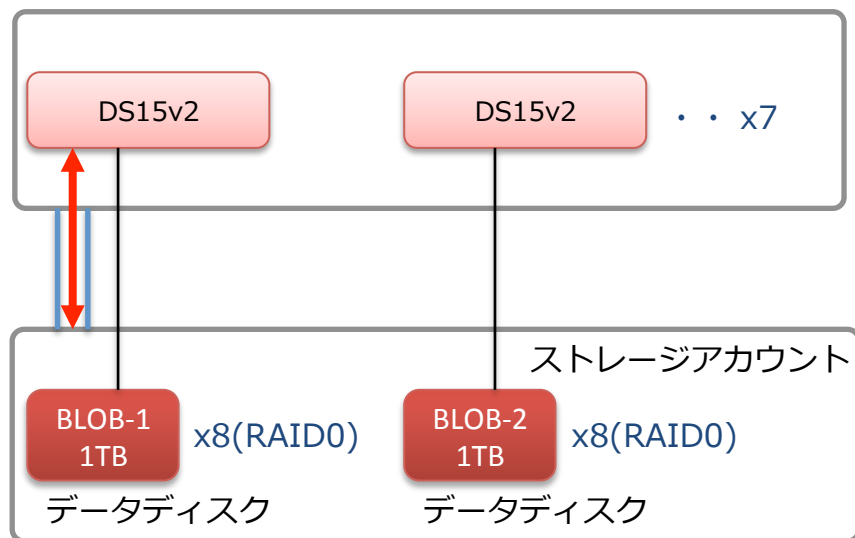
性能テスト 1

- 測定内容
ストレージアカウントの最大スループット
- 環境
サーバ : 1 x DS15v2インスタンス (20コア/140GB RAM)
CentOS 7.2
データディスク : 1~16 x 1TB (SoftwareRAID-0)
EXT4
- 測定方法
dd transfer size 1MBによる524GB WRITE

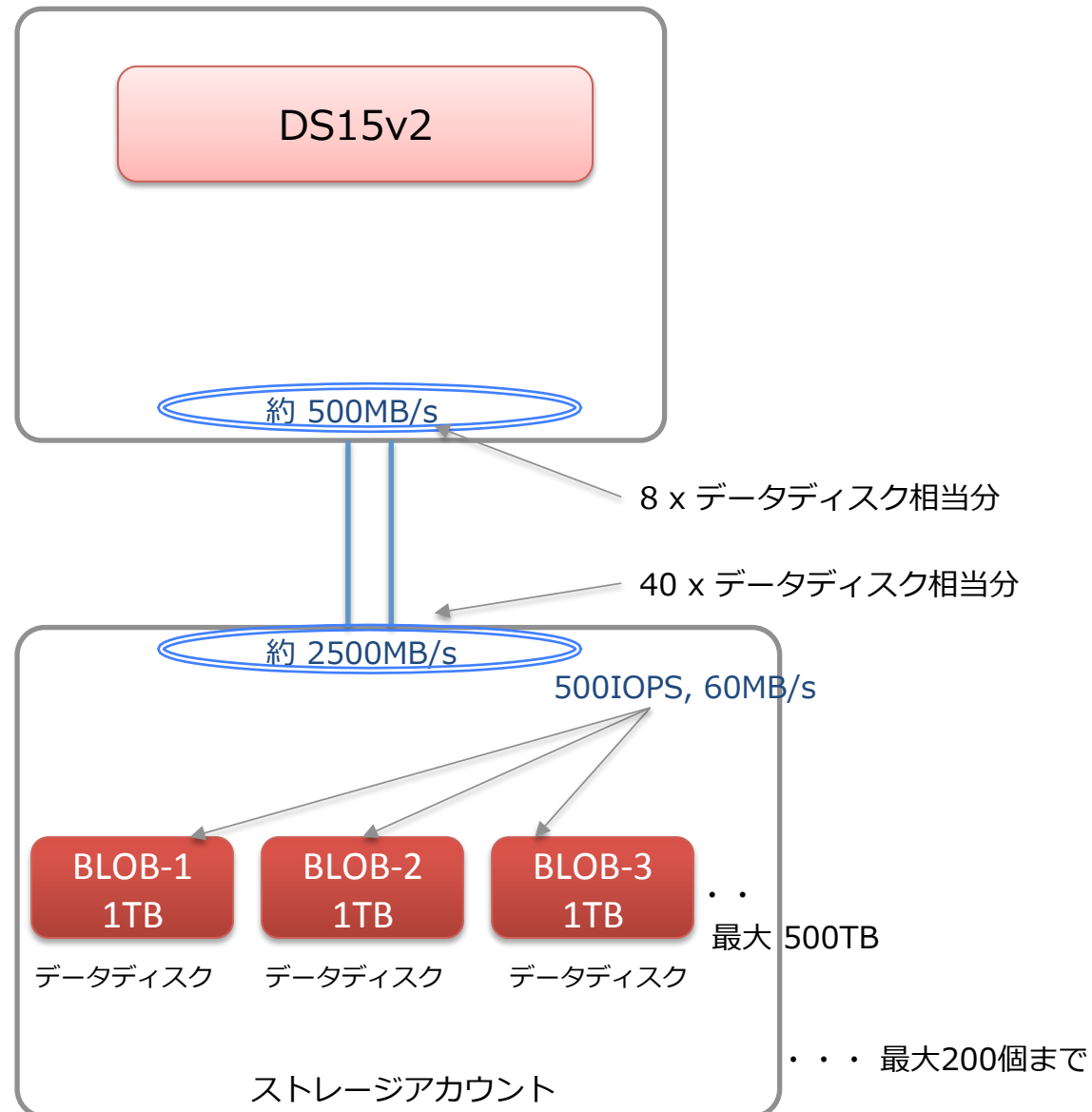


性能テスト2

- 測定内容
ストレージアカウントの最大スループット（つづき）
- 環境
サーバ：1～7 x DS15v2インスタンス（20コア/140GB RAM）
CentOS 7.2
データディスク：1～56 x 1TB
EXT4
- 測定方法
dd transfer size 1MBによる524GB WRITE



性能テスト1、2まとめ

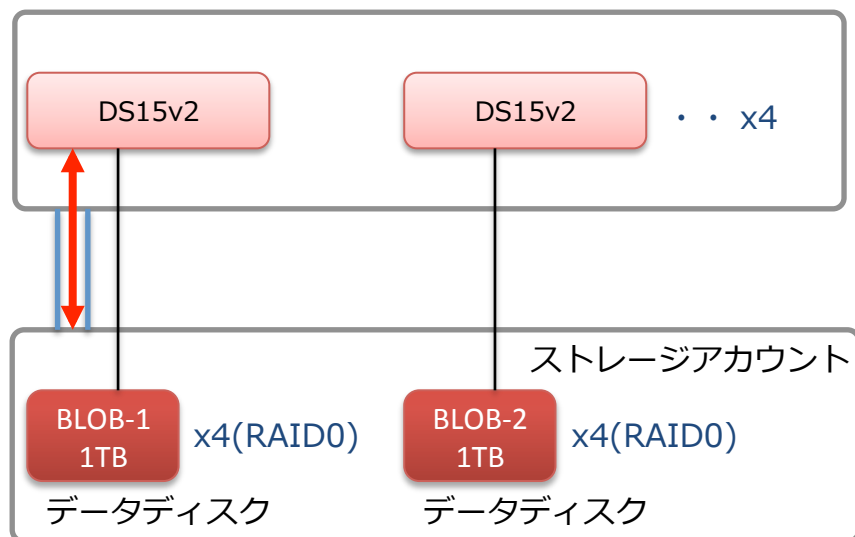


性能テスト3

- 測定内容
gfarm スループット

- 環境
サーバ : 1~4 x DS15v2インスタンス (20コア/140GB RAM)
CentOS 7.2
データディスク : 4~16 x 1TB (SoftwareRAID-0)
local-dd: EXT4
gfarm-dd, gfarm-IOR: Gfarm 2.6.14

- 測定方法
dd transfer size 1MBによる524GB WRITE
IOR transfer size 1MBによる536GB WRITE



gfarm mount option: -o direct_io 無し MB/s

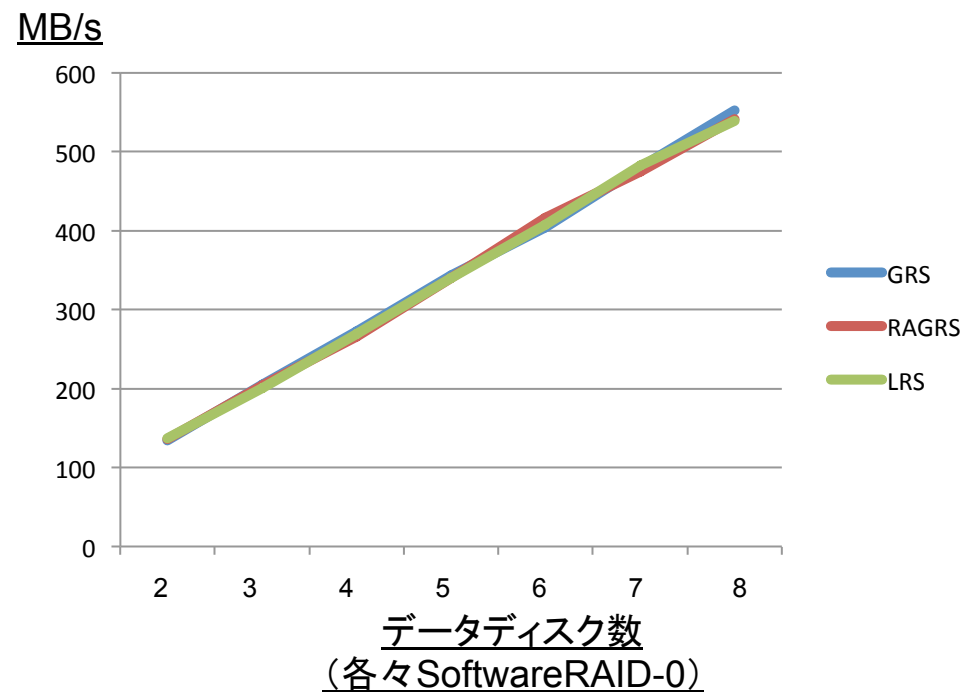
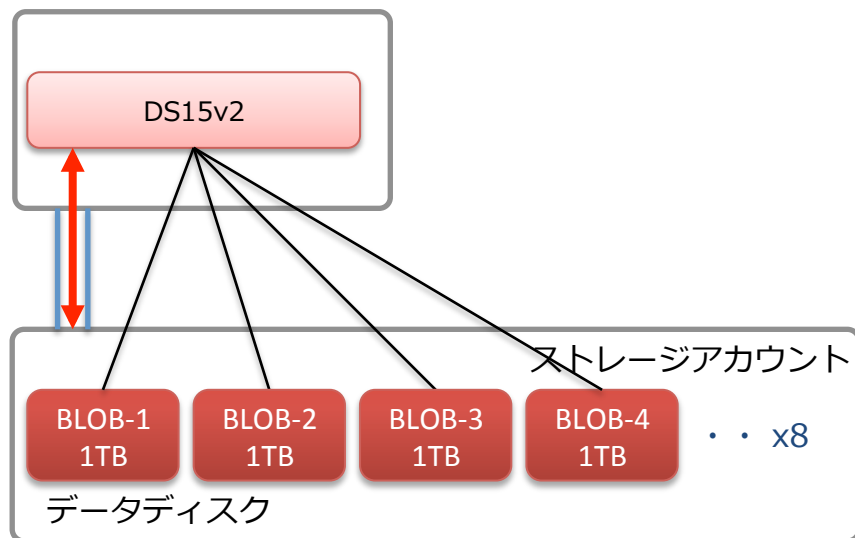
	local-dd	gfarm-dd	gfarm-IOR
1	250	272	140
2	548	390	280
4	1086	656	525

gfarm mount option: -o direct_io 有り MB/s

	local-dd	gfarm-dd	gfarm-IOR
1	250	278	260
2	548	543	519
4	1086	1083	1033

性能テスト4

- 測定内容
Azure 冗長オプション毎のスループット比較
- 環境
サーバ : 1 x DS15v2インスタンス (20コア/140GB RAM)
CentOS 7.2
データディスク : 2~8 x 1TB (SoftwareRAID-0)
EXT4
- 測定方法
dd transfer size 1MBによる524GB WRITE



性能テスト5

- 測定内容
Azure リージョン内／リージョン間でのGfarmレプリカ性能比較

- 環境
Azure : 2 x DS15v2 インスタンス (20コア/140GB RAM)
8 x 1TB BLOB(1 x ストレージアカウント)
CentOS 7.2
Gfarm 2.6.14

- gfrepによる7.8GBファイルのレプリカ作成

- 測定結果
[1] リージョン内 (秒)

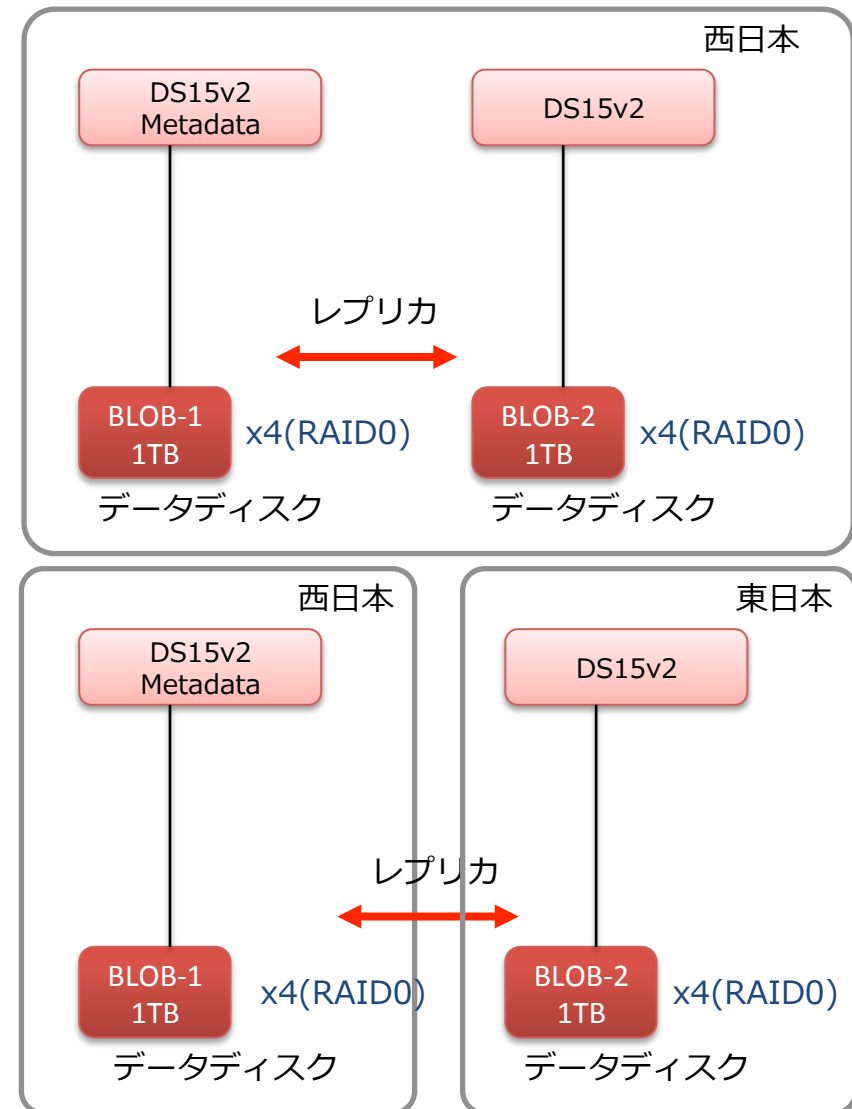
metadata	非metadata
239	274
226	603
379	418

- [2] リージョン間 (秒)

metadata	非metadata
55	53
82	76
76	53

要調査：

- ・リージョン間の方がリージョン内より高速
- ・ややバラツキ気味



まとめ

1. AzureのI/Oスループット

VM (DS15v2) あたり、約500MB/s (8xデータディスク)

ストレージアカウントあたり、約2500MB/s (40xデータディスク)

2. Gfarm on AzureのI/Oスループット

「1. AzureのI/Oスループット」相当

3. 今後

- ・リージョン内／リージョン間のgfrep性能については要調査
- ・その他の性能評価
- ・信頼性
- ・運用性コストなど