Gfarm Workshop 2020
2020年2月7日＠宮崎

Gfarmファイルシステムの概要と最新機能

建部修見
筑波大学

Gfarmファイルシステム

GRID DATA FARM

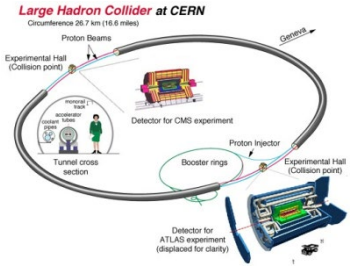
- オープンソース広域分散ファイルシステム
  - http://oss-tsukuba.org/software/gfarm/
- サポート
  - NPO法人つくばOSS技術支援センター（日本他）
  - Libre Solutions Pty Ltd（オーストラリア）
- 特徴
  - 性能・容量がスケールアウト
    - データアクセス局所性、ファイル複製
    - 無停止で拡張、更新可能
  - 単一障害点なし
    - 複製数維持機能、ホットスタンバイMDSサーバ
  - データ完全性を保証しサイレントデータ損傷も対応可

ossTsukuba

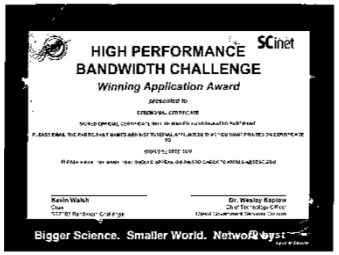
oss-tsukuba.org

Congratulation

20th anniversary!



- 2000年6月　電総研と高エネ研のGRIDミーティング
  - LHC実験の要求に応えるため日本独自研究開発開始
- 2002年11月　日米間ファイル複製実験で世界最高の741Mbpsを達成！
- 2003年11月　SC03において「分散インフラストラクチャ賞」を受賞！Gfarm 1.0を公開
- 2005年9月　Gfarm Workshop 2005開催！Gfarm 1.2公開
- 2005年11月　SC05において「Most Innovative Use of Storage In Support of Science賞」を受賞！
- 2006年11月　SC06におけるHPC Storage Challengeで優勝！Gfarm 1.4を公開
- 2007年11月　Gfarm 2.0.0公開！
- 2019年11月30日　Gfarm 2.7.16公開！！！



HPCI共用ストレージ

- 大学情報基盤センターをはじめ全国からマウント可能な共有ファイルシステム（～100PB）
- スパコン間のデータ共有、共有データ格納

西拠点（R-CCS）

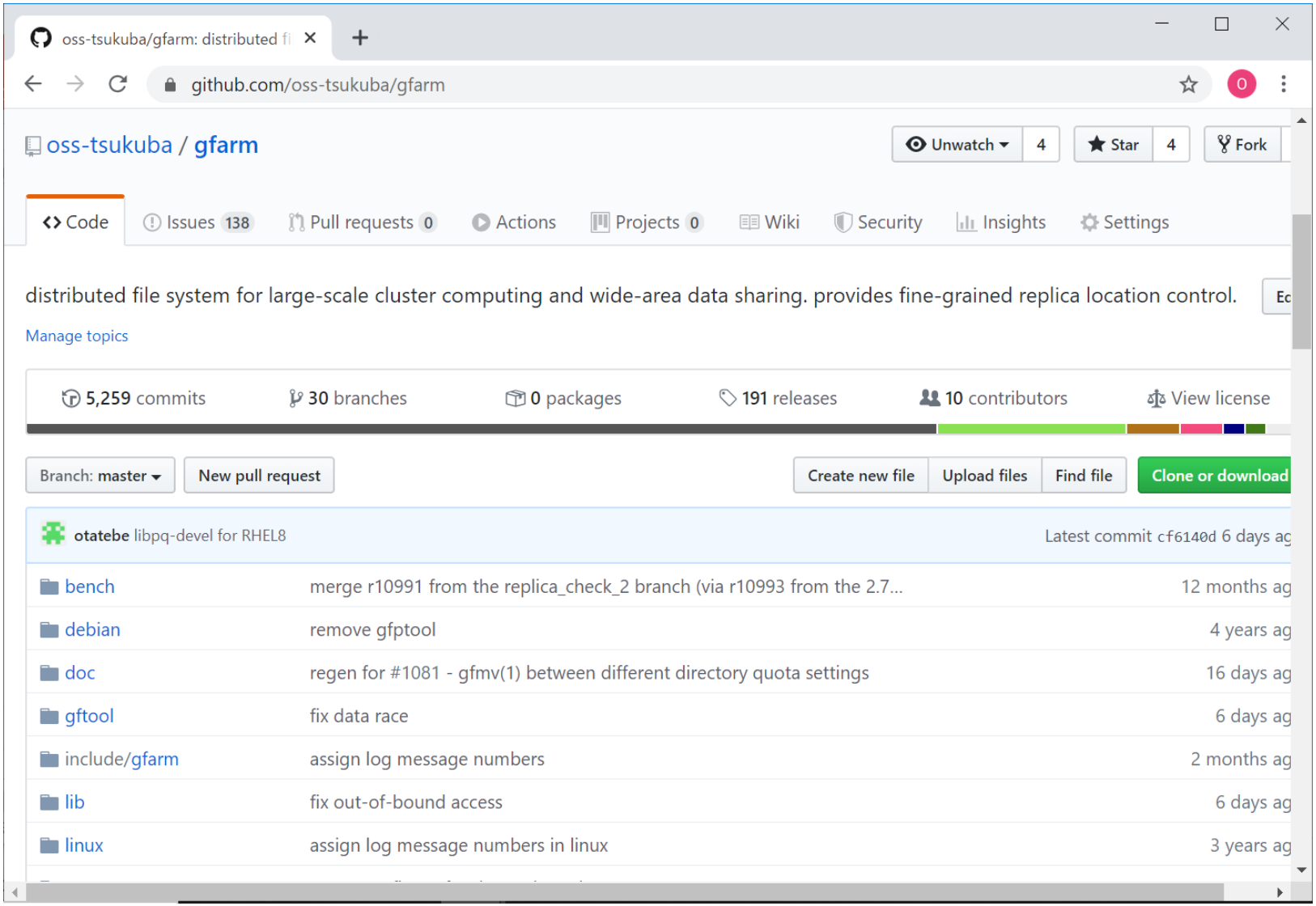
東拠点（東京大）

最新機能・状況紹介

主なリリース

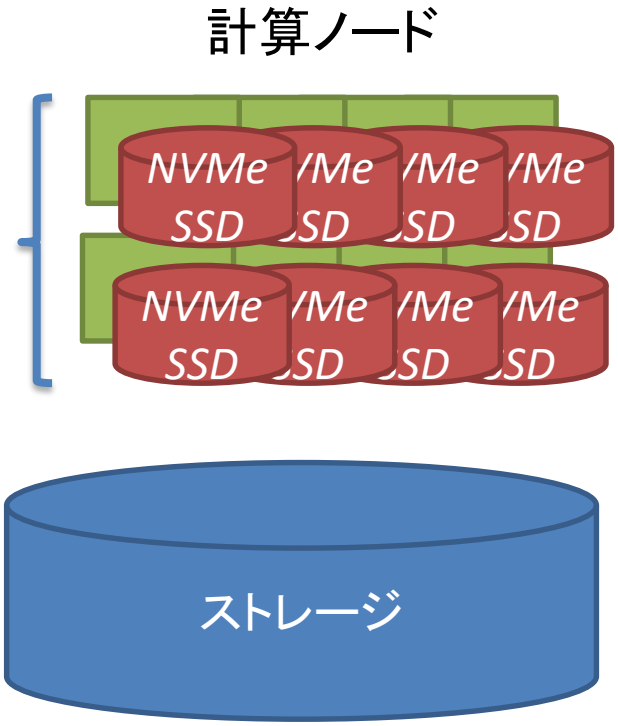
日付	version	新機能、更新機能
2019/11/30	2.7.16	• Githubへの移行！
2019/10/24	2.7.15	• Gfarmbb status, IB GRH対応
2019/9/10	2.7.14	• Gfarm/BBバーストバッファ
2016/12/8	2.7.0	• InfiniBand RDMAサポート • ディレクトリクオータ
2016/1/16	2.6.8	• 書込後ベリファイ

https://github.com/oss-tsukuba/gfarm



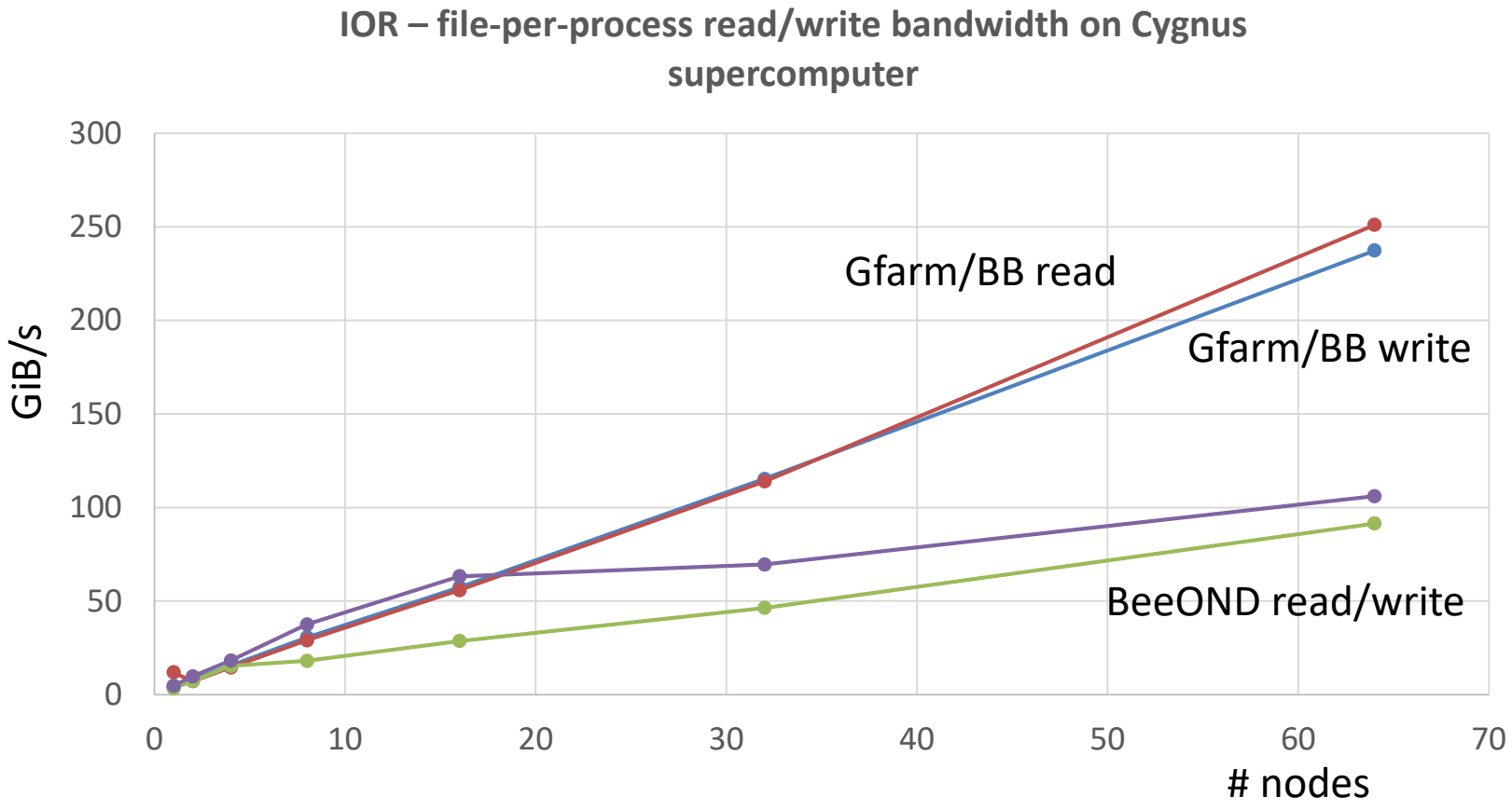
Gfarm/BBバーストバッファ [建部 2020]

- ノードローカルNVMe SSD等高速ストレージによる一時的な分散ファイルシステム
- アクセス性能の向上
  - ファイルディスクリプタパッシングによるgfsdを経由しない直接アクセス
  - RDMAアクセス
- メタデータ性能の向上
  - メタデータの永続性、冗長性なし
  - ジャーナル書込み、バックエンドDB、スレーブgfmdのオーバヘッドの削減
- 冗長性オーバヘッド削減
  - ファイル複製によるデータの冗長性なし
- ファイルシステム構築、撤去の高速化



Gfarm/BBバーストバッファ（2）

```
gfarmbb –h hostfile –m mount_point start
…
gfarmbb –h hostfile stop
```

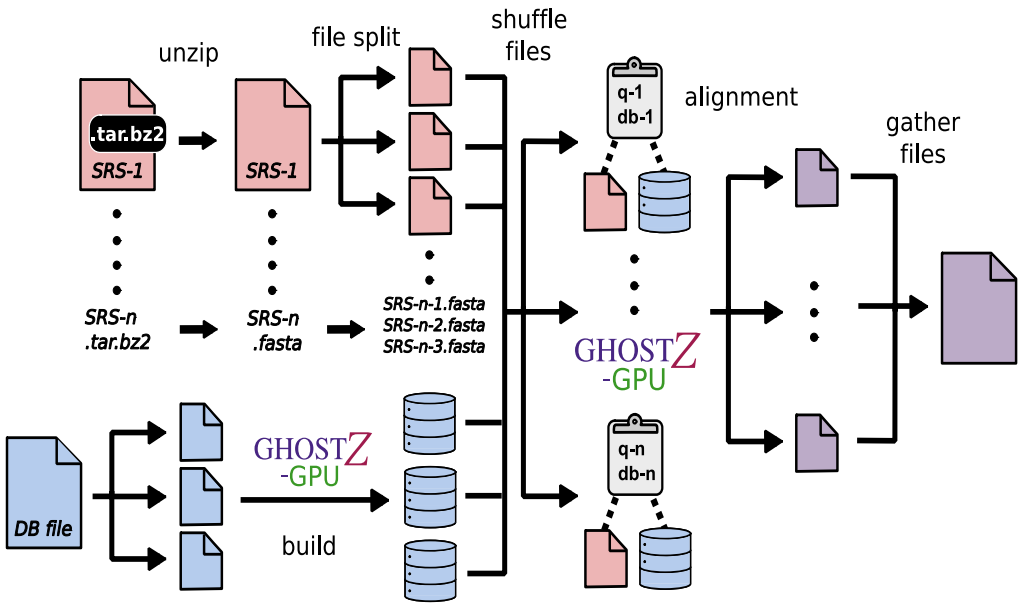


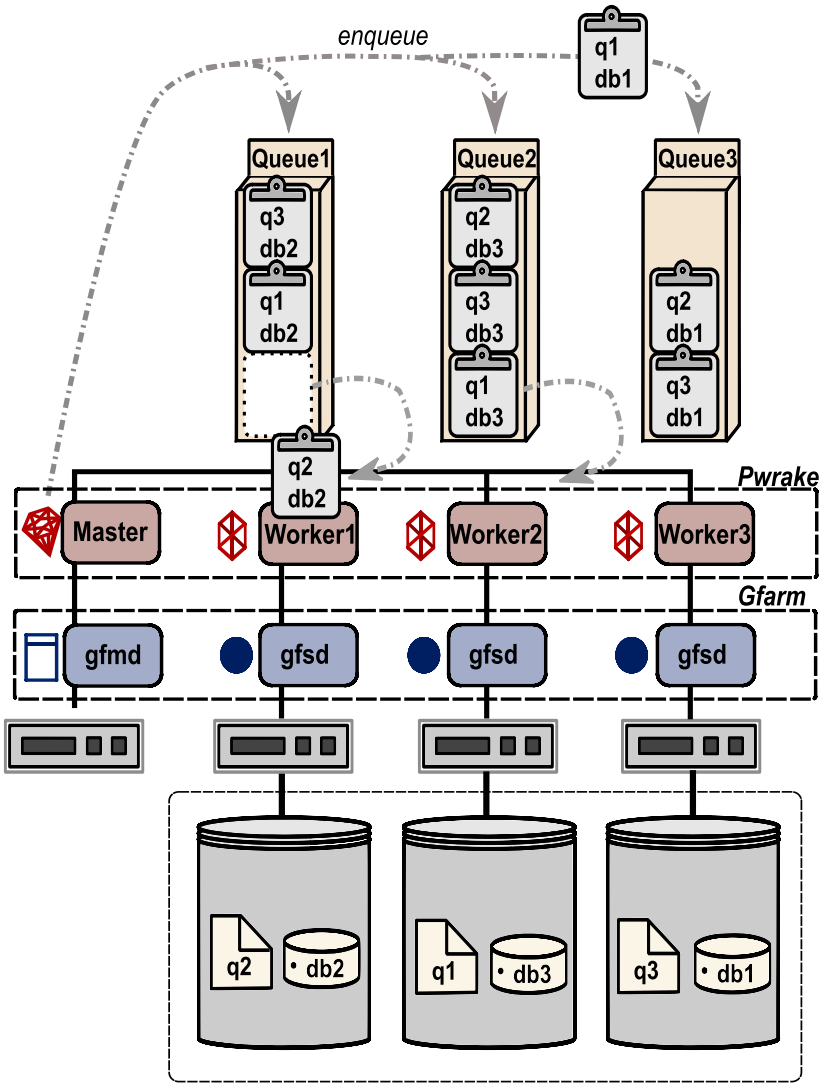
大規模メタゲノム解析 [町田 2019]

- クエリ、DBの増大

→ クエリ、DB双方を分割して分散計算

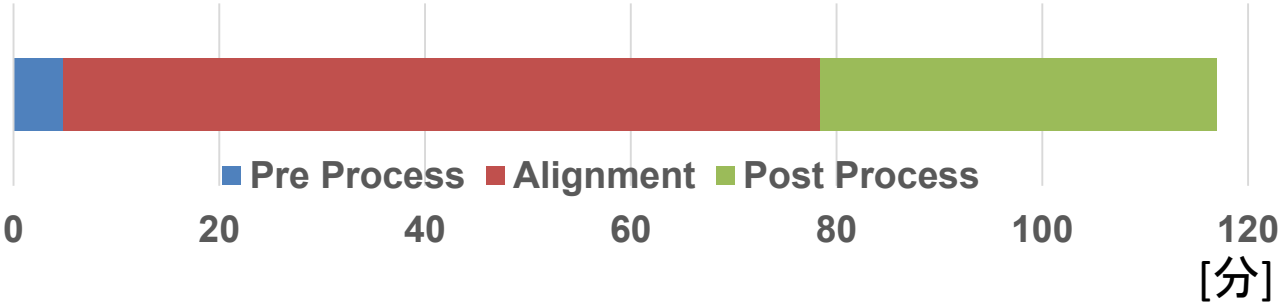
- Gfarm/BB, Pwrake, GhostZ-GPU

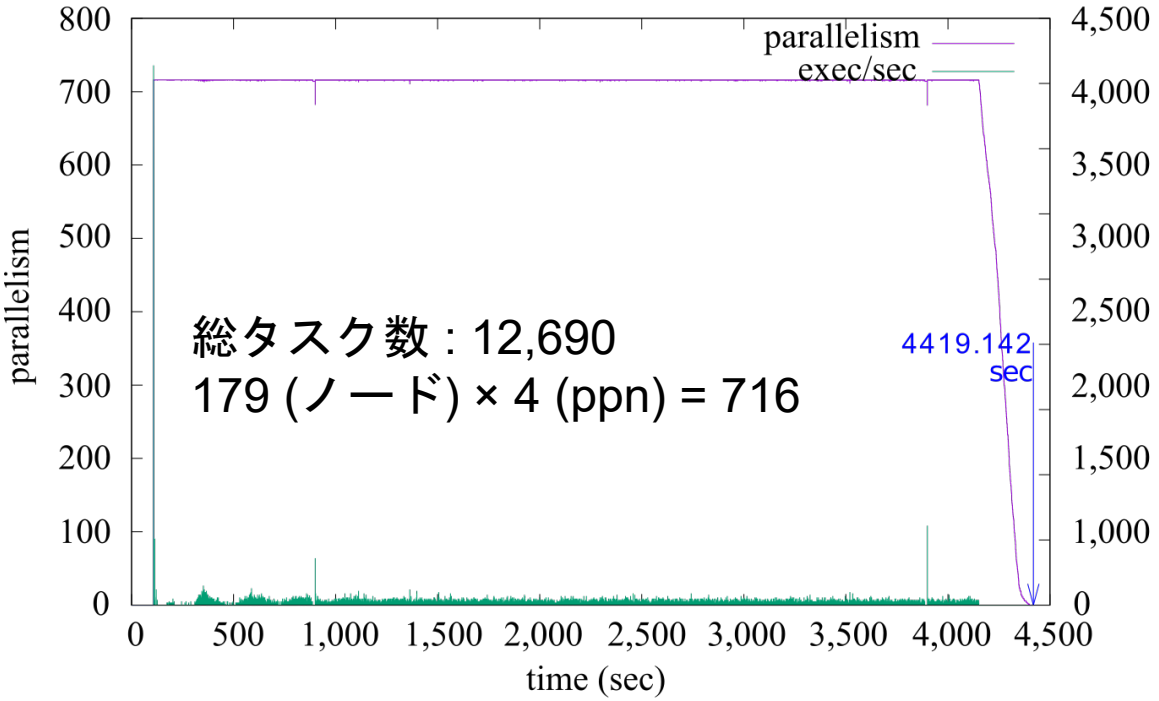


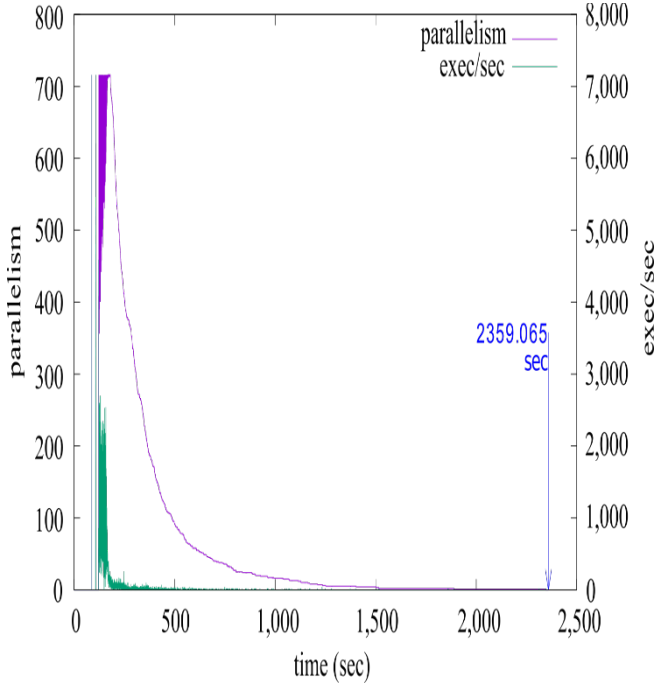


TSUBAMEグランドチャレンジ

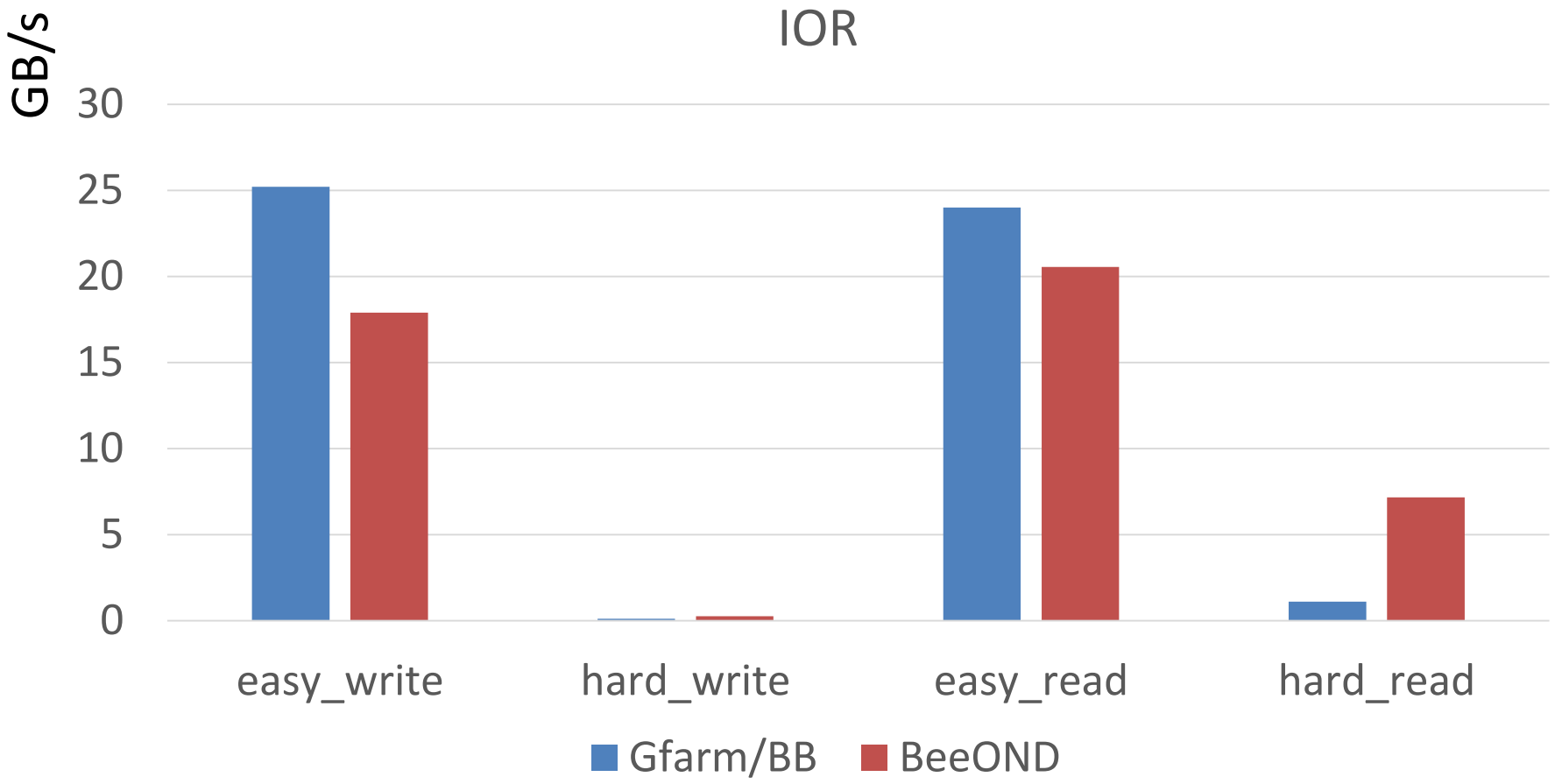
- 180ノード
- 62GB nr DB, 71GB クエリ（口内歯垢）



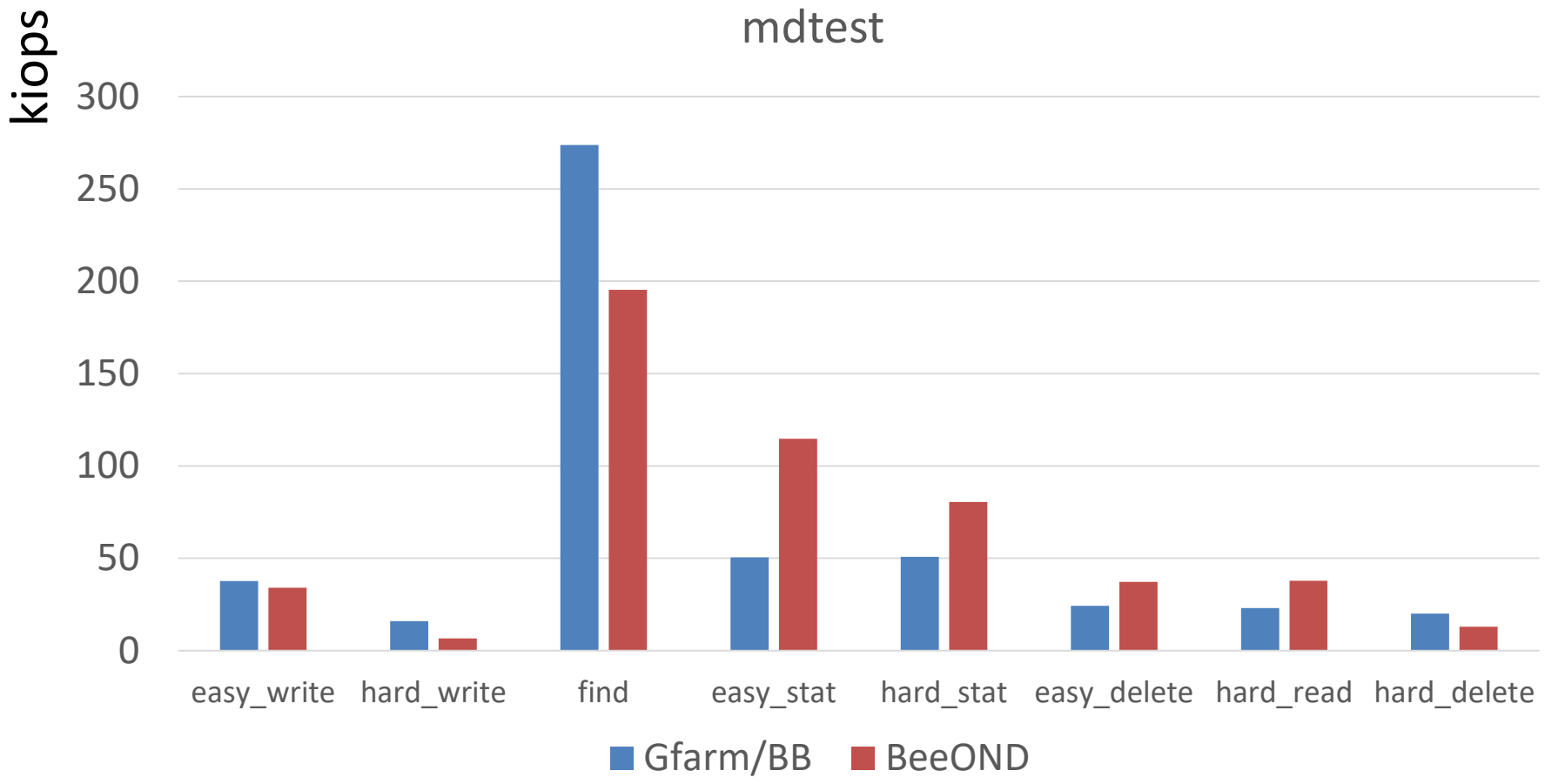




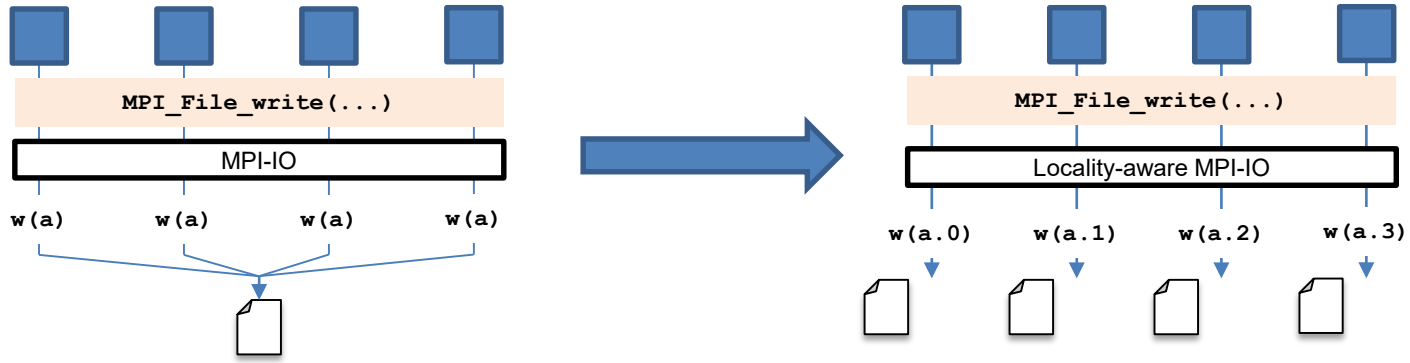
IO500 10 node challenge (1)



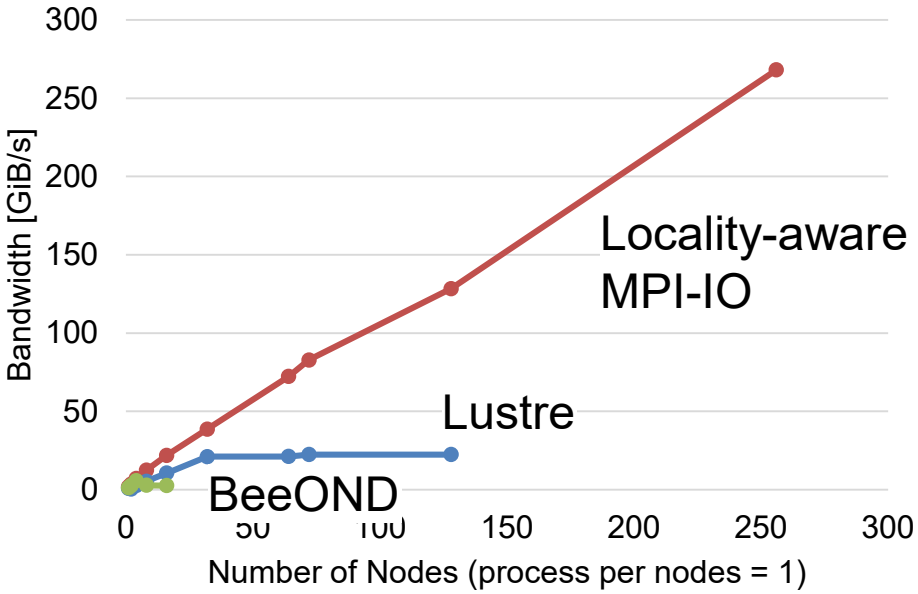
IO500 10 node challenge (2)



Locality-aware MPI-IO [M1 杉原]



- MPI-IOで分割セグメントの内部形式で格納して、N-1アクセス→N-Nアクセスに変換
- IOR N-1アクセス



まとめ

- Gfarmファイルシステム
  - NPO法人つくばOSS技術支援センターによるサポート
  - Gfarm 2.7.16を11/30にリリース
  - https://github.com/oss-tsukuba/
- Gfarm/BBバーストバッファ
- InfiniBand RDMA、ディレクトリクオータ機能
- データ完全性、サイレントデータ損傷対応
- HPCI共用ストレージ、JLDGなど実運用実績
- 進行中
  - IPv6対応（Gfarm 2.8）
  - 暗号化対応