Intelが提供する新機能と今後の開発プラン

Oleg Drokin

Senior Staff Engineer

2014年10月17日

Agenda

Features of note in Lustre 2.6 release

Distributed Namespace phase 2

LFSCK phase 3

16M RPC

Data on MDT.

Questions

Lustre* 2.6

2014年7月にリリース

追加された新機能

- LFSCK MDT-OST　整合性チェック(LU-1267)
- シングルクライアントのIO性能向上 (LU-3321)
- DNEストライプディレクトリ(LU-3531) のプレビュー

新しいカーネルのサポートの基礎

Feature Release Only - メンテナンスリリースの予定はなし

Lustre* 2.7

2015年2月にリリース予定

予定されている新機能

- UID Mapping and Shared Key (LU-3527)
- LFSCK MDT-MDT　整合性(LU-4788)
- Dynamic LNET Configuration (LU-2456)

RHEL7のクライアントのサポートを追加

- SLES12のクライアントも追加予定

2.5.xと2.6リリースと相互運用性、アップグレード可能

Feature Release Only - メンテナンスリリースの予定はなし

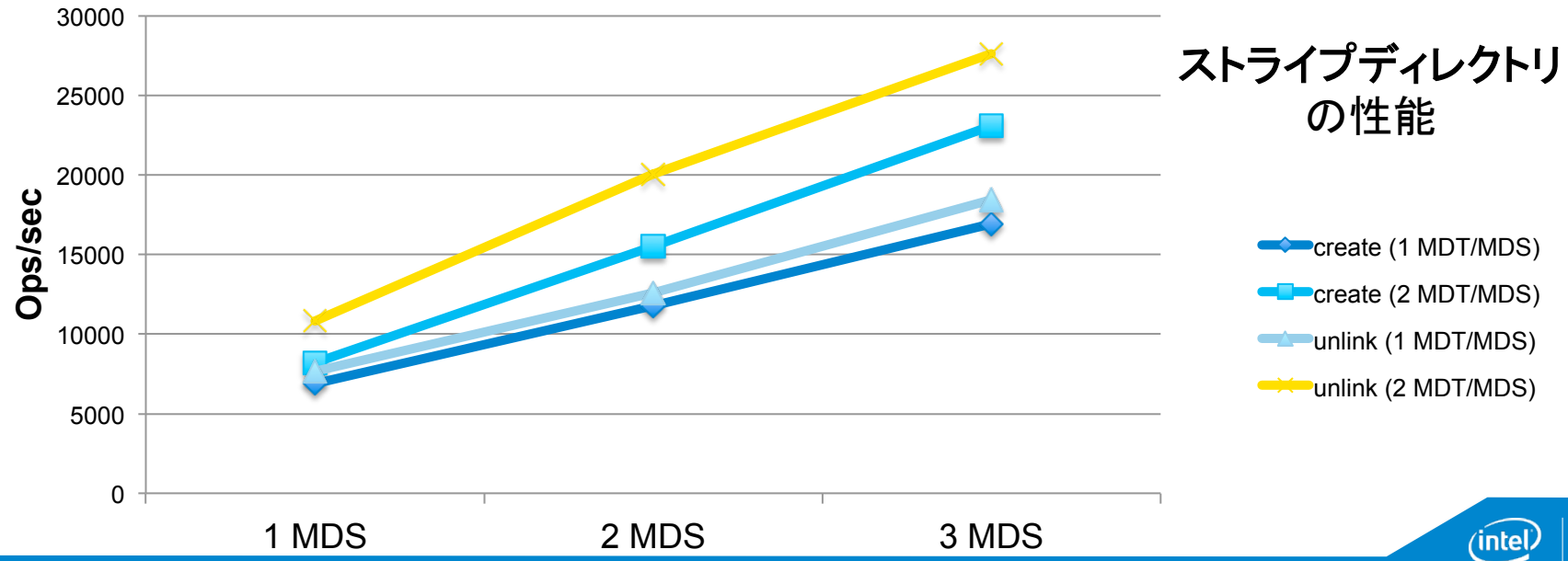
DNEストライプディレクトリ Phase 2

ストライプディレクトリとinodeマイグレーションツール(2.6)

- シングルディレクトリ内にて複数のMDTに分散配置
- データコピーなしでMDT間でファイル/ディレクトリの移動が可能

非同期のリモートアップデート(2.7)

- 複数のMDT下のオペレーション (mkdir, rmdir, striped dir, …)に対する性能向上
- 他のMDTとのrenameおよびハードリンクのサポート



Lustreの整合性チェック(LFSCK) ― Phase 3

現在までに実装済みのLFSCK機能(lustre-2.3-2.6)

- Phase 1: OI Scrub for local inode iteration and OI reconstruction
- Phase 1.5:ローカルnamespaceの整合チェック (name->FID, linkEA->parent)
- Phase 2: MDT-OST整合性チェック(LOV EA check, orphans)

MDT-MDT間での整合性チェック(lustre-2.7)

- リモートディレクトリとファイルリンクを検証
- 親のないリモートディレクトリは「lost+found」に移動
- ディレクトリエントリでinodeの参照先がないものを修復

16MB Bulk RPC – Intel/DDN

4MB OST RPC サポートの追加 (Lustre-2.4)

- ストリーミングRAIDのWrite/Readの性能向上
- ネットワーク上で同時に複数のLNET RDMAの送受信が可能
- クライアントのI/O リクエストエンジン自体の変更はほとんどなし

16MB+ Bulk RPCをサポートし更なる性能向上

- クライアントのIO リクエストエンジンを改良し性能劣化を最小化
- Large RPCのメモリハンドリングの改善
- Large RPCでのランダムI/Oの性能向上

Data on MDT (2.x)

MDTで効率的に小さいファイルを保存

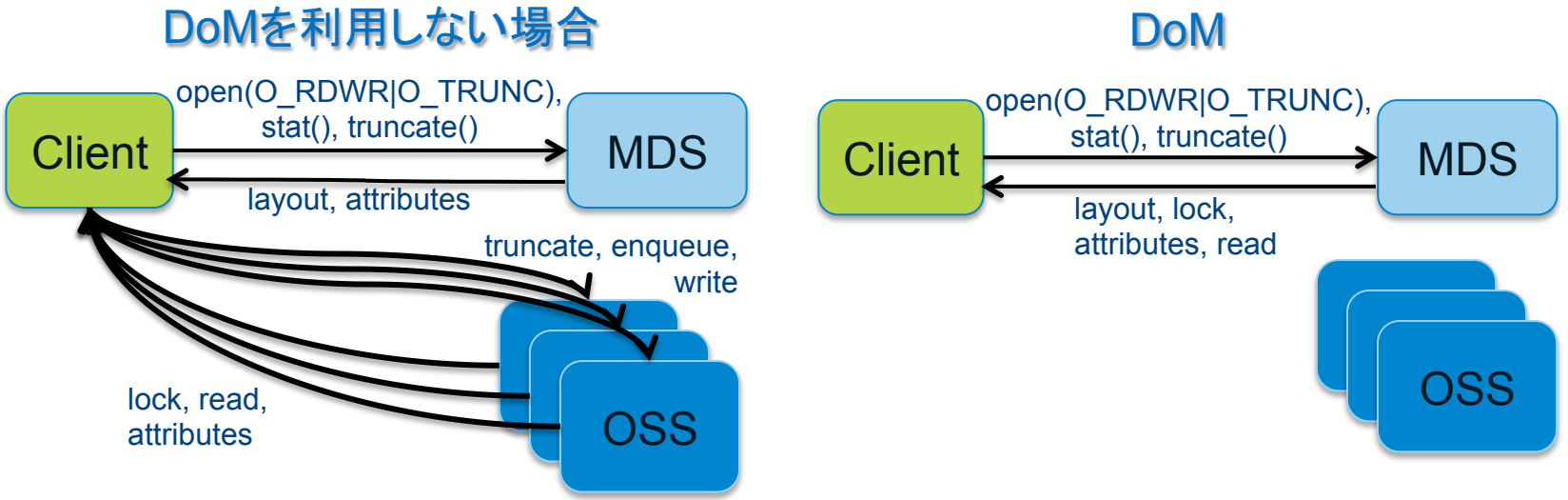
- ファイルアクセスの時OST RPCは不要
- ファイルアクセスの時OST ロックも不要
- 小さいファイルに対するストレージの最適化(RAID10/SSD/NVRAM)
- RAID-5/6の小さいファイルに対する性能劣化を回避

MDTの使用量は、クォータによって管理

ファイルはファイルレイアウトによって配置が決定

- 一番大きさMDSファイルはmin(ユーザ, 管理者)
- 一般的に <= 1MB、MDTの空き領域に応じて
  - Phase1 : 制限を超えるファイルはMDTで保存出来ない (EFBIG)
  - Phase2: 制限を超えるファイルはOSTに移す

Data on MDT(DoM)

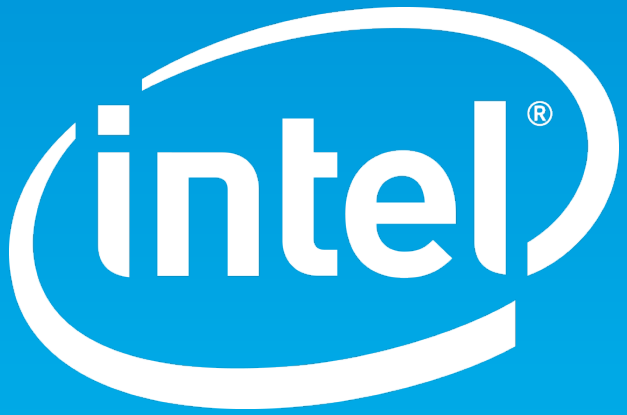


DoMはファイル生成時のみ有効

- 既に存在するオブジェクト対しては利用できない
- ディレクトリ内でストライピングの設定も可能に(Phase2)
- lfs setstripe **--stripe-pattern=mdt** [--stripe-size=<size>] new_file

http://cdn.opensfs.org/wp-content/uploads/2014/04/D1_S10_LustreFeatureDetails_Pershin.pdf
http://wiki.opensfs.org/Contract_SFS-DEV-003

Question?


intel