NICT ScienceCloud

National Institute of Information and Communications Technology

第5回サービスコンピューティング研究発表会

NICTサイエンスクラウド広域分散ファイルシステムのセキュリティ機能拡張の要件

渡邉　英伸[1],
岩間司[1], 村田健史[1], 鵜川健太郎[2], 村永和哉[2], 鈴木豊[2],
木村映善[3], , 建部修見[4], 高杉英利[5], 亀澤祐一[5]

[1].情報通信研究機構, [2].株式会社セック, [3].愛媛大学大学院医学系研究科,
[4].筑波大学計算科学研究センター, [5].株式会社エヌ・ティ・ティネオメイト

◆発表の構成

NICT ScienceCloud

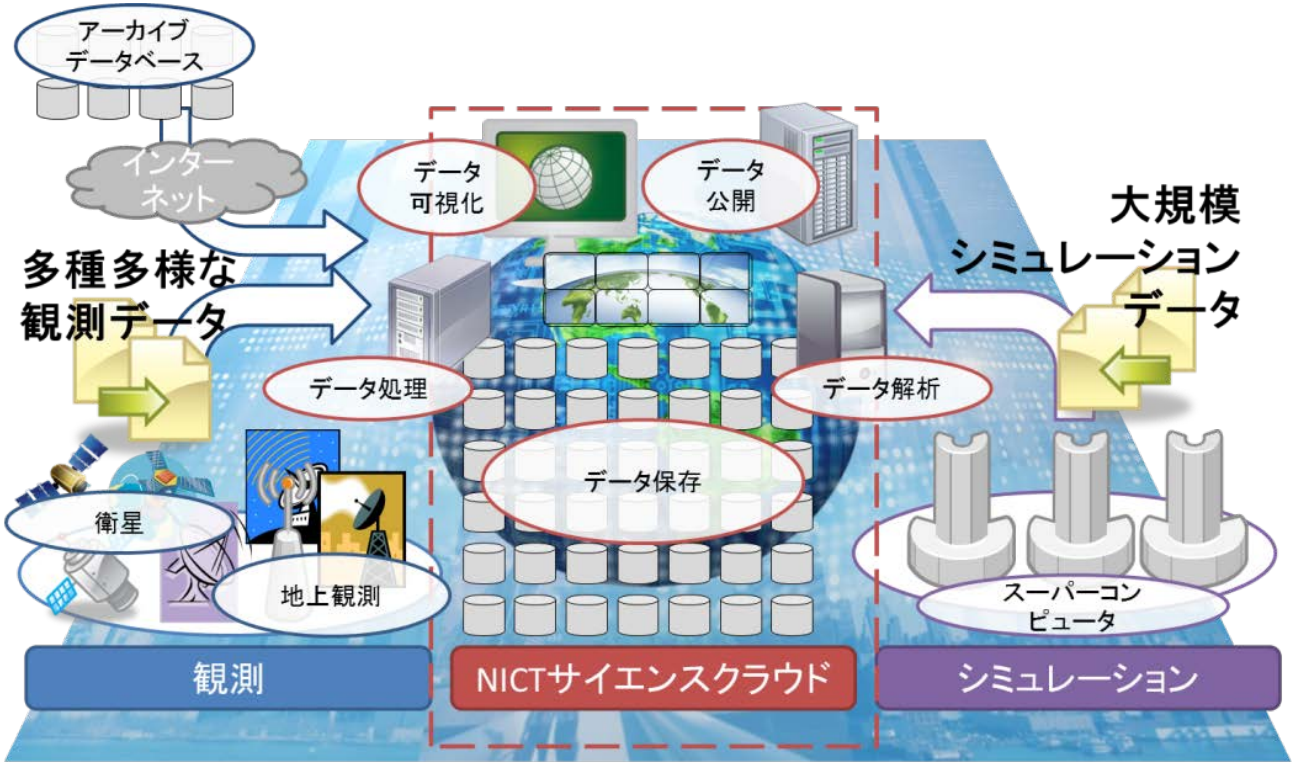
- 背景・発表内容
- NICTサイエンスクラウド大規模分散ストレージシステムの現状
- 広域分散ファイルシステムのセキュリティ要件
- 取り組みについて
- まとめ

◆背景　NICTサイエンスクラウドについて

NICT ScienceCloud

- 科学研究向けクラウドシステム（2010年よりテストベッド運用開始）
  - 約550コア、約3.2PB容量の分散並列処理環境（※）
  - 約5億ファイルの蓄積（冗長化込）（※）
  - 約230アカウント発行（※）

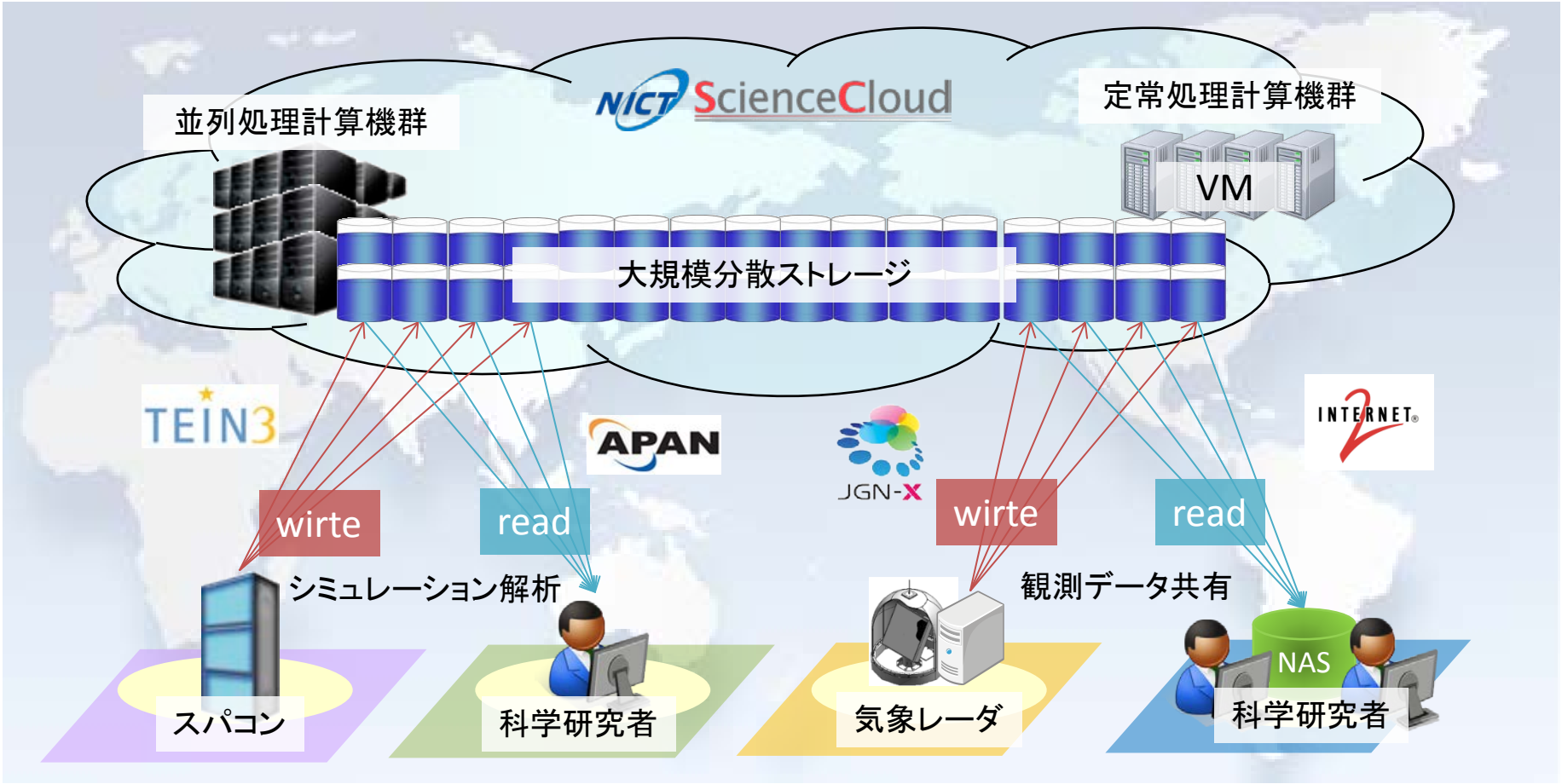
※H25年6月現在



あらゆるデータをクラウド上に！

◆発表内容

- NICTサイエンスクラウド大規模分散ストレージシステムの現状
- 広域分散ファイルシステムのセキュリティ機能拡張の取り組み



◆発表の構成

NICT ScienceCloud

- 背景･発表内容
- NICTサイエンスクラウド大規模分散ストレージシステムの現状
- 広域分散ファイルシステムのセキュリティ要件
- 取り組みについて
- まとめ

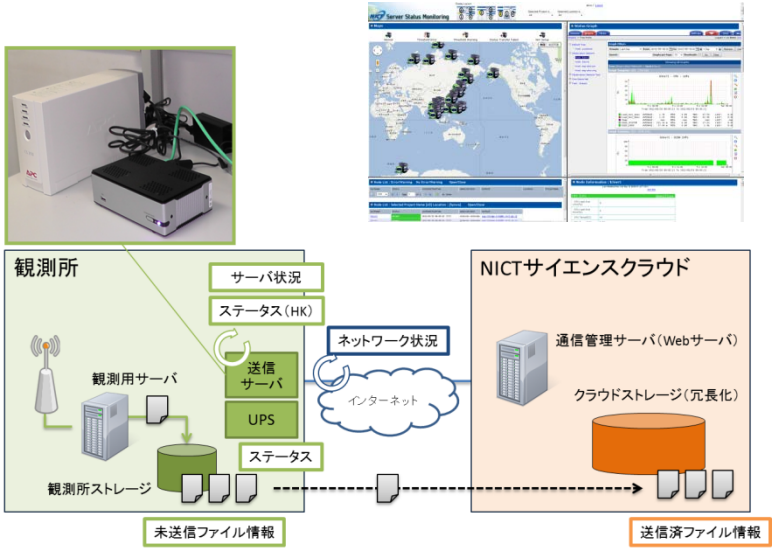
◆ 利用事例（1） NICT宇宙天気観測ネットワーク

Space Weather Monitoring Networks (NICT-SWM)

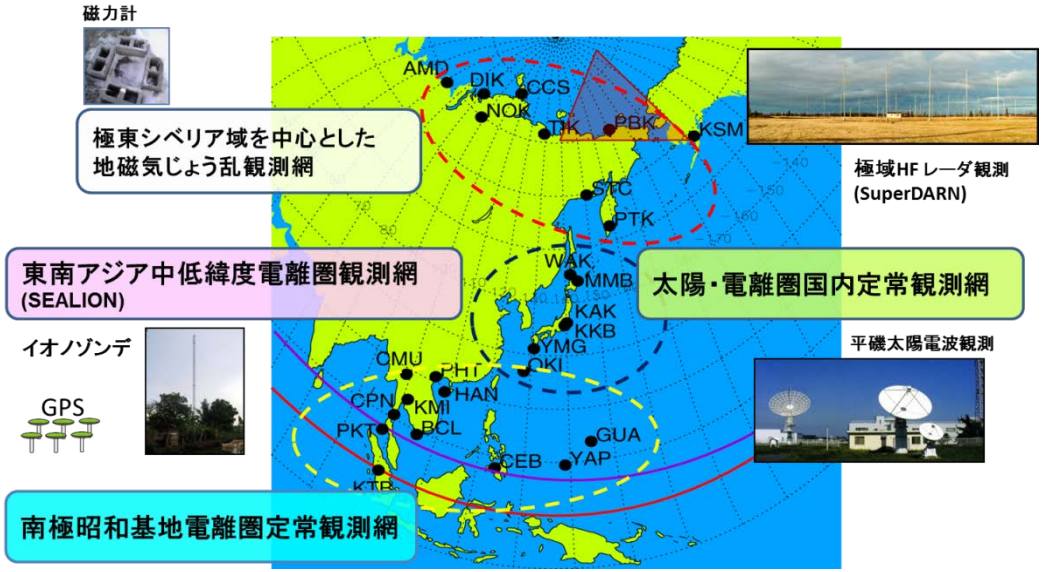
NICT ScienceCloud

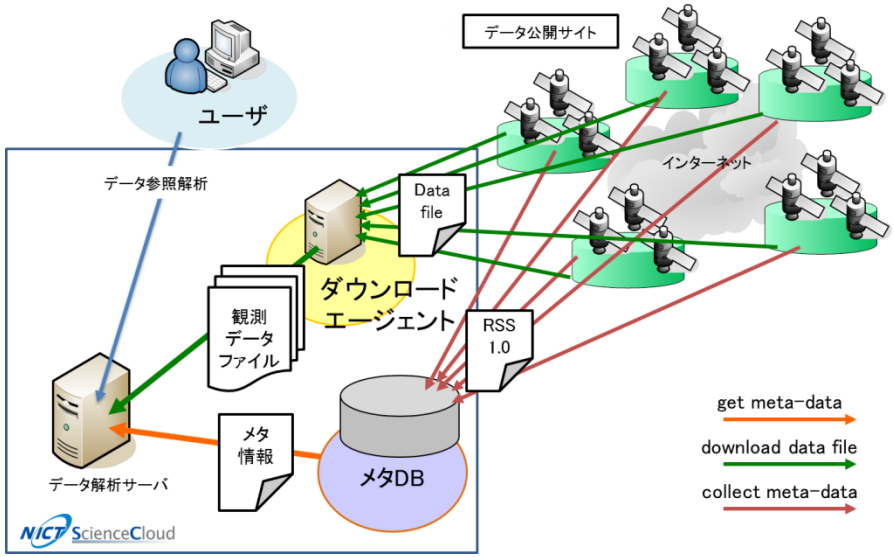
- 太陽宇宙環境の計測・予測，および電波伝播障害の研究開発の一環として宇宙天気モニタリングを実施
  - 世界中の27観測拠点から観測データを収集保存
  - 約70の科学データ公開サイトから自動でデータを収集保存
  - 分析・解析処理結果をWebサイトで公開

データのアーカイブと共有が重要



広域観測ネットワーク監視システム（WONM system）



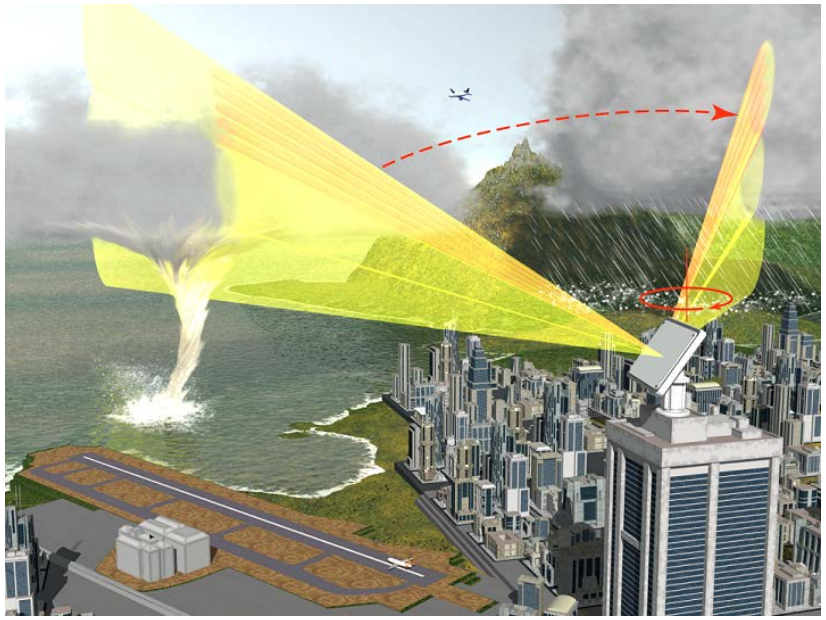


データ収集ツール（NICTY/DLA）

◆ 利用事例（2）
フェーズドアレイ気象レーダのデータ保存とデータ処理

NICT ScienceCloud

- 突発的、局所的気象災害の予測や災害対策のため、局地的大雨（ゲリラ豪雨）、竜巻突風等を数十秒間隔100mグリッドで、すき間のない3次元データとして観測
  - リアルタイムにデータの解析およびデータの公開
  - バックグラウンドで準リアルタイムに観測データをバックアップ
  - 将来的には、観測データをリアルタイム伝送・リアルタイム処理し、突発的、局所的気象災害の予測を行う

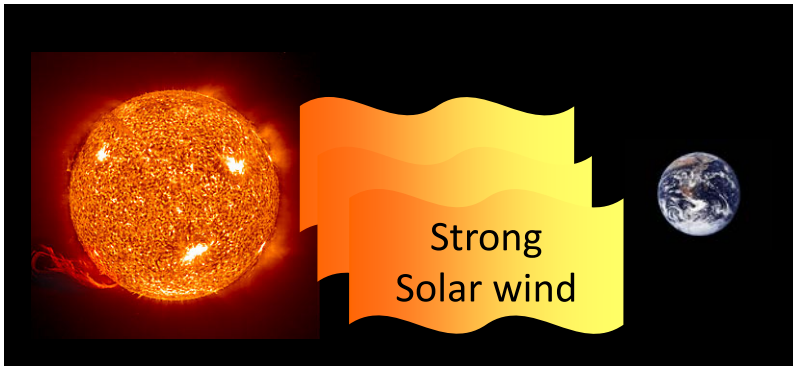


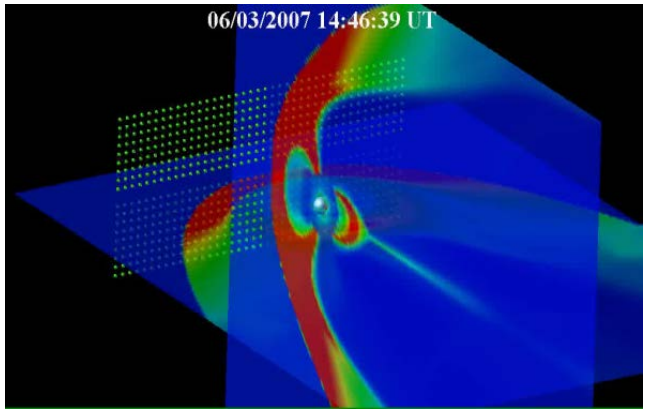
データの高速なバックアップが重要

◆ 利用事例（3） Global MHDシミュレーションの3次元可視化

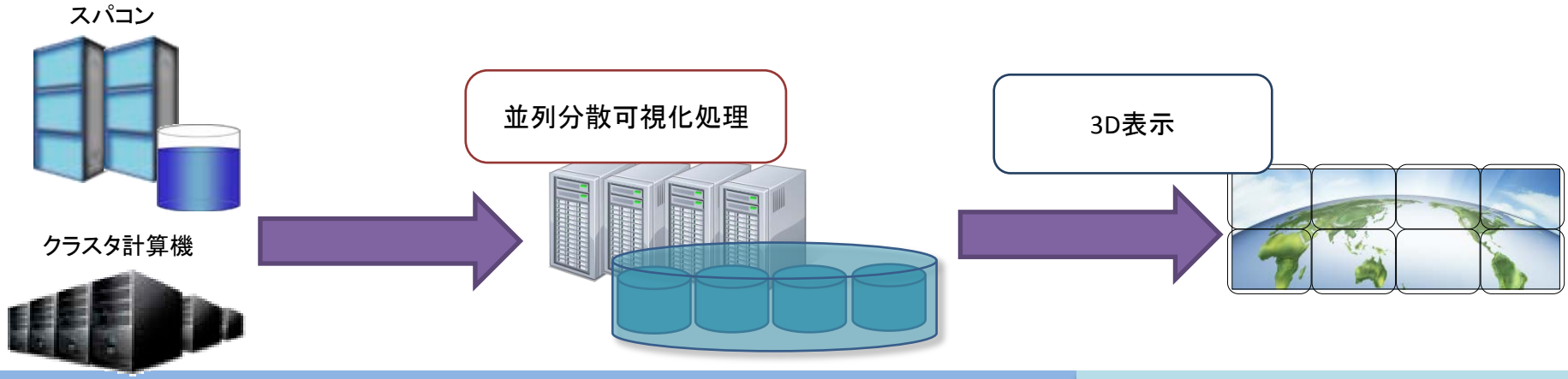
NICT ScienceCloud

- 太陽風と地球磁気圏相互作用などの複雑なシミュレーション結果を理解するため、3次元可視化によって磁気圏の流線、磁力線及び電流構造の特徴を明らかにする試み
  - NICTのスパコンなどから数値シミュレーション結果を一時保存
  - 3次元可視化処理を並列分散処理
  - 3次元アニメーションの表示





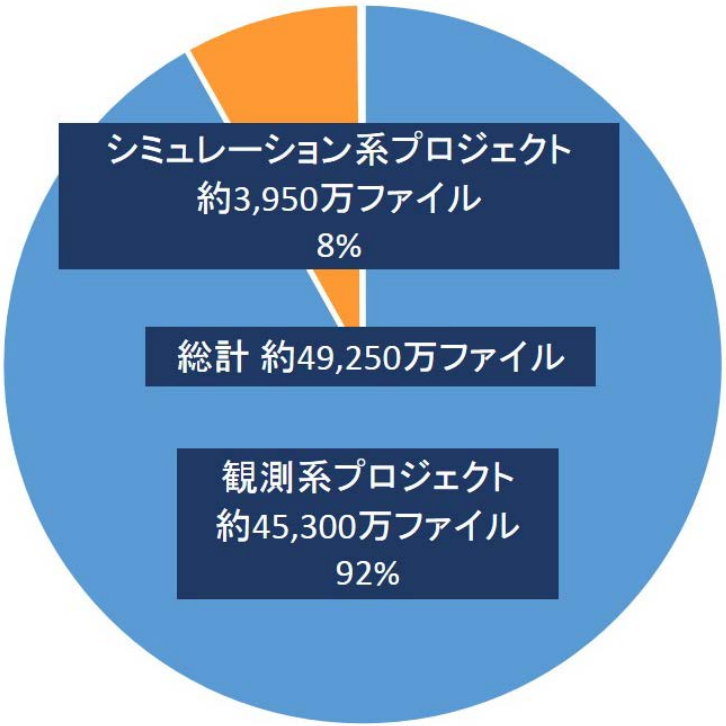
高速なデータI/Oとデータアクセスの継続性が重要



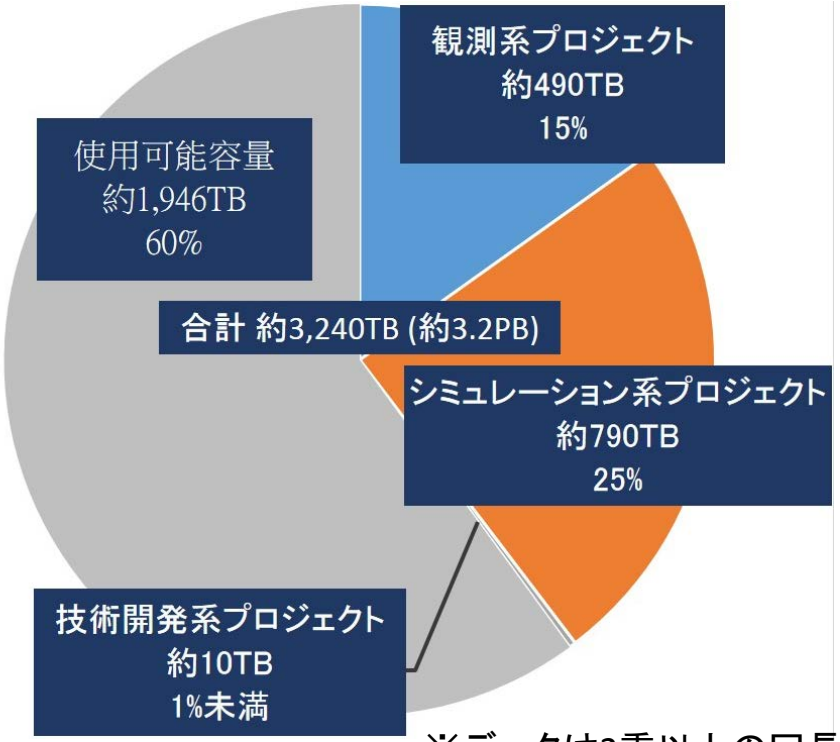
◆科学データの保存状況

NICT ScienceCloud

- 保存ファイル数
  - 約4億9千万ファイル
  - 観測系プロジェクトのファイル数が9割を占める
    - ✓ 観測データは再現性なし



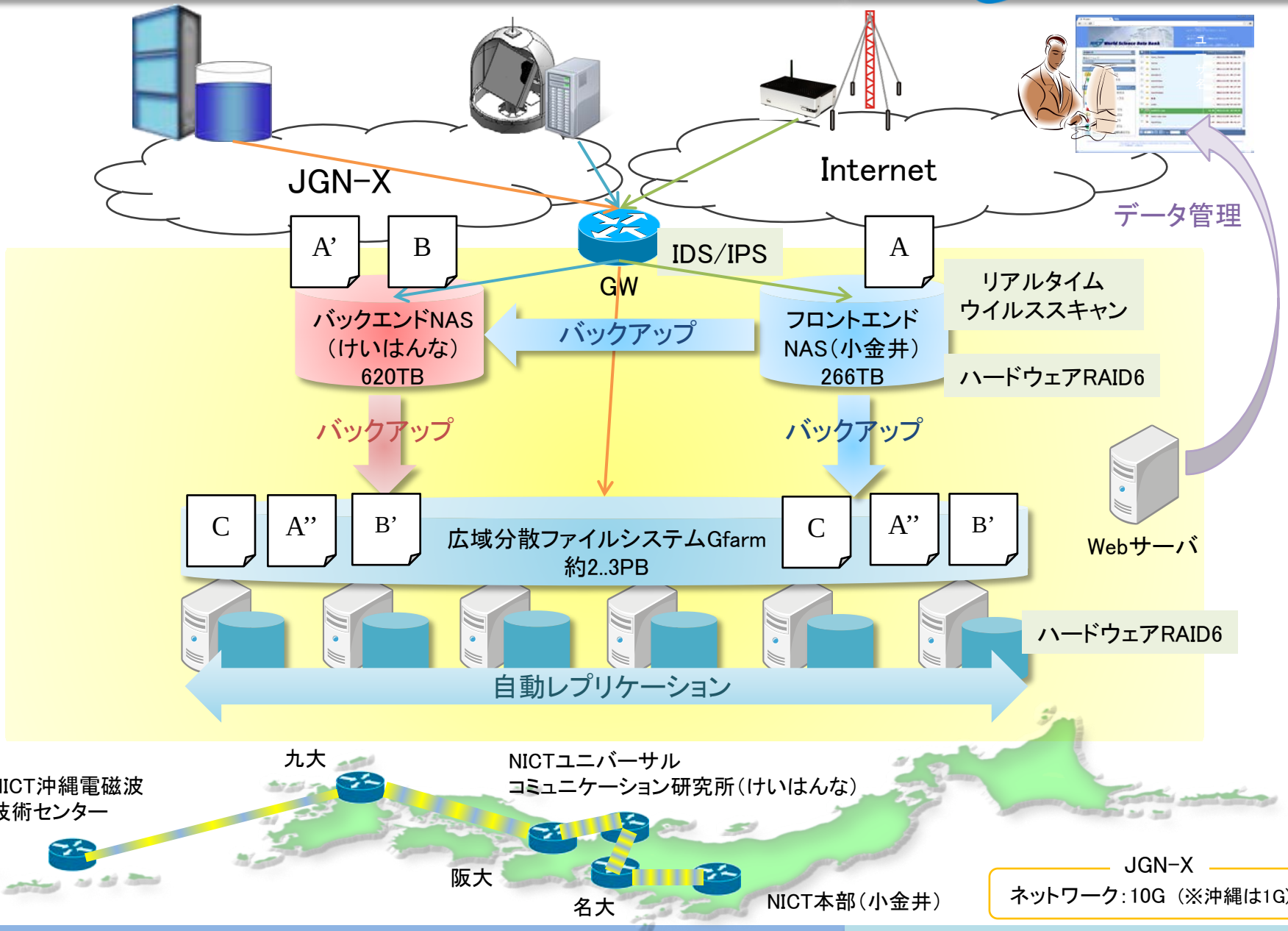
- ディスク使用状況
  - 約1.28PB
  - シミュレーション系プロジェクトのディスク使用率が多め
    - ✓ シミュレーションデータは再現性あり



※データは2重以上の冗長化

◆ NICTサイエンスクラウド大規模ストレージシステム構成

NICT ScienceCloud



◆発表の構成

- 背景･発表内容
- NICTサイエンスクラウド大規模分散ストレージシステムの現状
- 広域分散ファイルシステムのセキュリティ要件
- 取り組みについて
- まとめ

◆ 広域分散ストレージサービスのメリット・デメリット

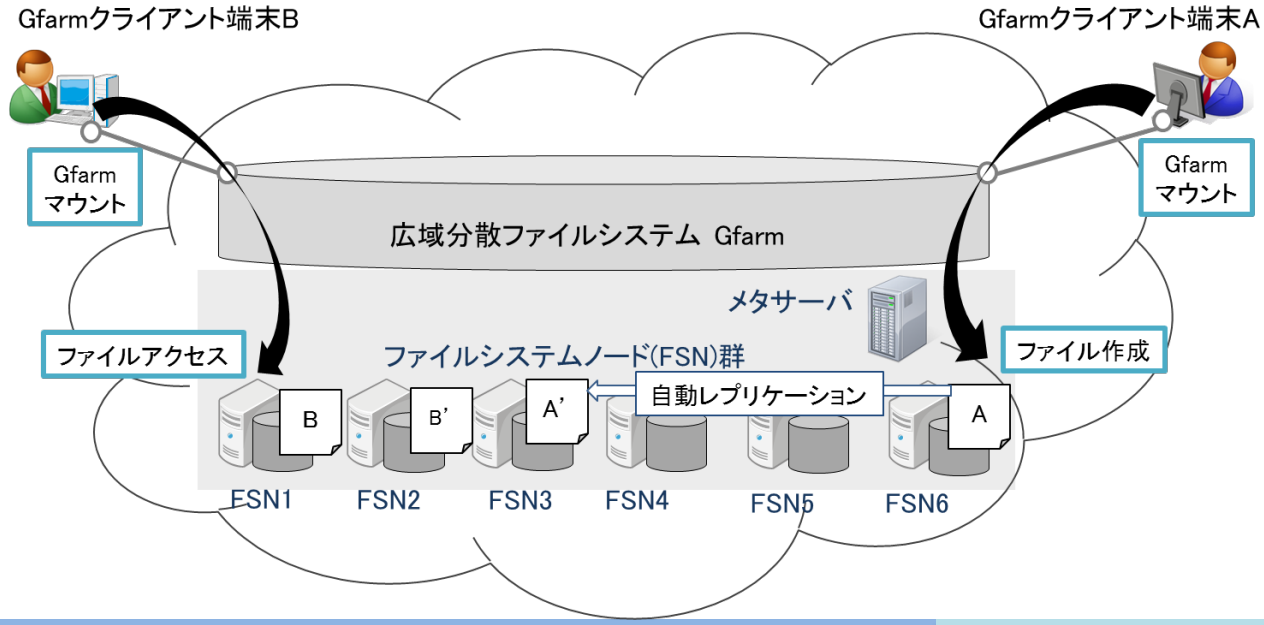
NICT ScienceCloud

広域分散ストレージサービスの利点

- スケーラビリティ
  - ディスク容量の拡大が容易
- 可用性
  - データの地域分散保存によりデータの損失率を抑え、いつでもアクセスできる
- 作業の省力化
  - サーバ管理、調達など科学研究者の研究以外の作業の省力化が期待できる

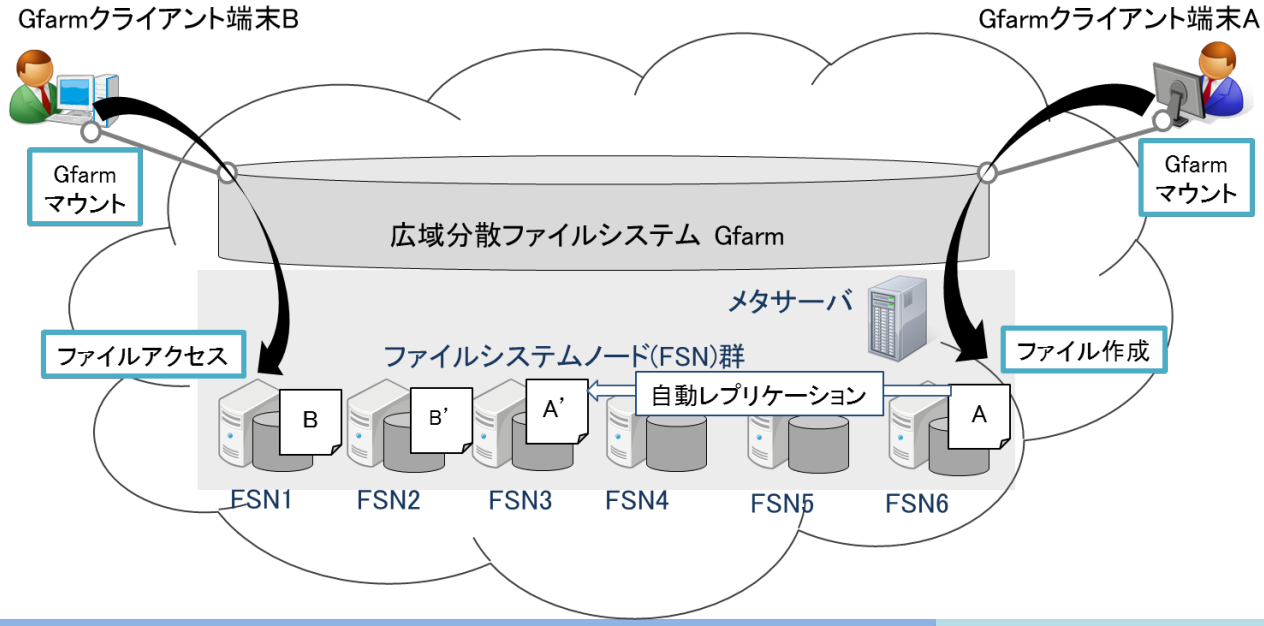
広域分散ストレージサービスの課題

- セキュリティ
  - 情報漏えいなどセキュリティの心配
- 性能
  - 大容量データや長距離に対する低い転送効率
- 見える化
  - ブラックボックス化による情報・状況の不透明さ

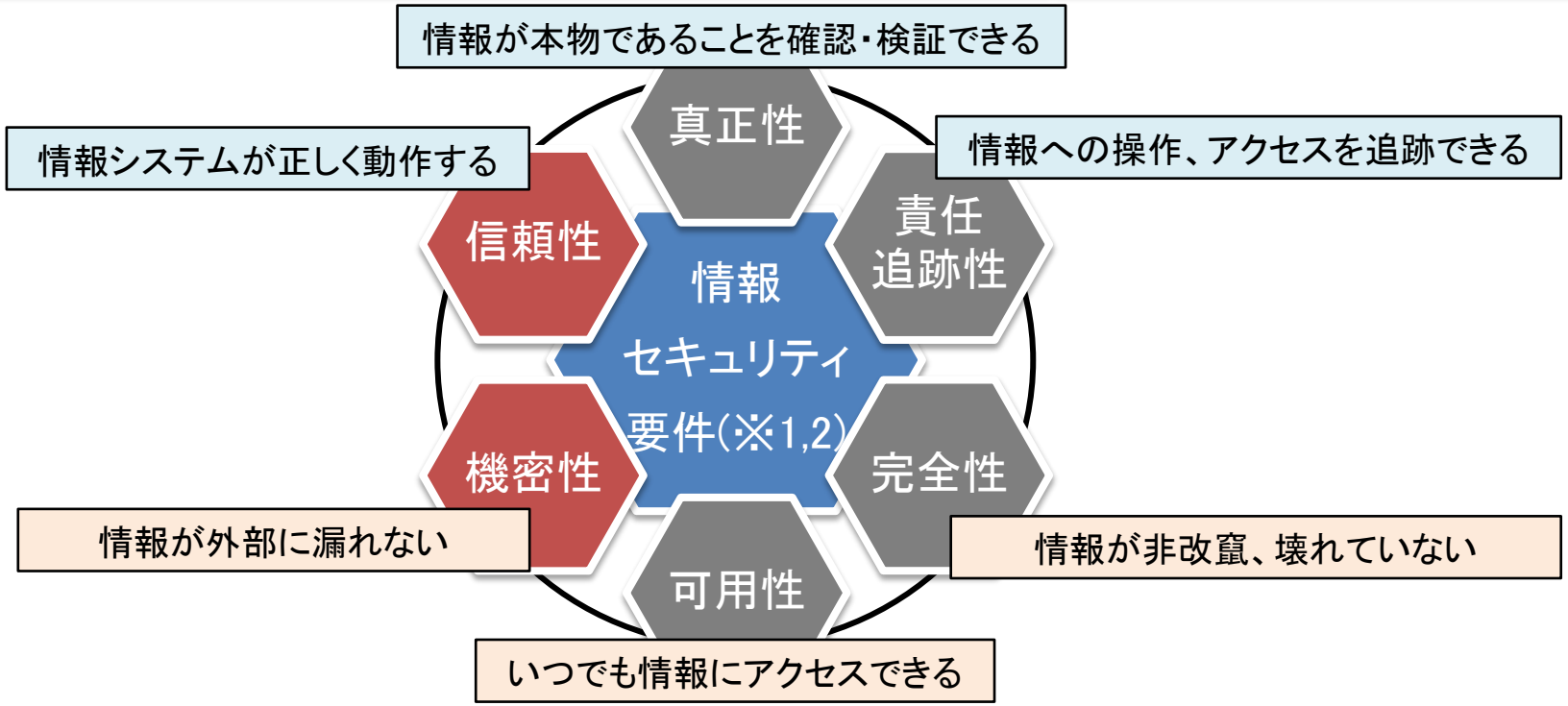


◆Gfarm　(Grid Datafirm)

- 日本国内で開発されたオープンソース（OSS）の広域分散ファイルシステム
  - 大量のデータ保存と高速データI/Oによる並列処理環境も提供
  - SC05(SuperComputer2005) StoreCloud Challenge にて，"Most Innovative Use of Storage In Support of Science" Award を受賞
  - HPCIの共有ストレージに採用
  - 民間企業のメールアーカイブやファイル共有のインフラにも採用事例あり



◆情報セキュリティ要件とGfarmの対応状況



Gfarmの対応済セキュリティ要件

- **信頼性**: 徹底した機能試験によって信頼性を達成
- **機密性**: GSI(Grid Security Infrastructure) によりユーザ認証やデータ暗号化を達成

Gfarmのセキュリティ課題

- **完全性**: TCPチェックサムレベルでの担保のみ
- **責任追跡性**: 未実装
- **真正性**: 未実装（完全性と責任追跡性を担保することで真正性を確保することが可能）
- **可用性**: レプリケーション機能によりデータの冗長化は達成済だが、SLA(Service Level Agreement)は未定義

※1, JIS Q 27002 (ISO/IEC 27002)
※2, 経済産業省, 高度情報化社会における情報システム・ソフトウェアの信頼性及びセキュリティに関する研究会, 情報システム・ソフトウェアの信頼性及びセキュリティの取組強化に向けて～豊かで安全・安心な高度情報化社会に向けて～-中間報告書

◆発表の構成

- 背景･発表内容
- NICTサイエンスクラウド大規模分散ストレージシステムの現状
- 広域分散ファイルシステムのセキュリティ要件
- 取り組みについて
- まとめ

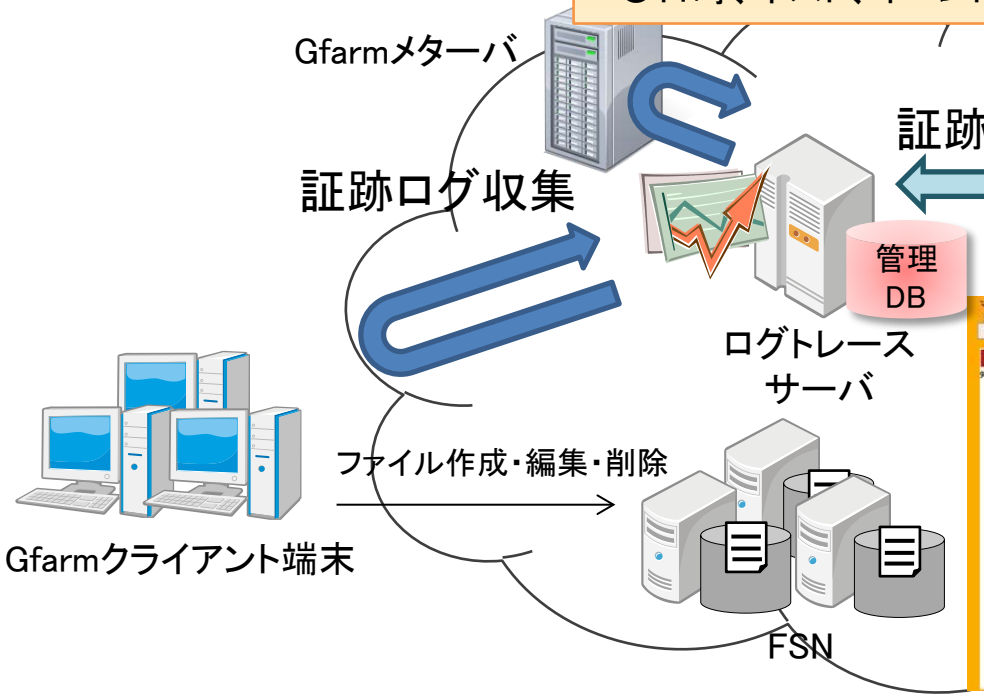
◆ 責任追跡性への取り組み：ログトレースシステム（1）

NICT ScienceCloud

- システム管理者に対して、データのライフサイクルのイベント（作成、更新、参照、複製、削除など）追跡可能なシステム(※)

凡例
- 証跡ログの流れ
- ユーザの操作

以下の情報を確認することができる。
●ファイルの総数、複製数、ファイルの作成、削除、更新、シンボリックリンクの作成、削除等に関するイベントの履歴
以下の条件をもとに検索できる
●日時、ホスト、イベントの種類、ユーザ、グループ、Gfarmシステムログ等



管理者ポータル(NICT内部からのみアクセス可能)

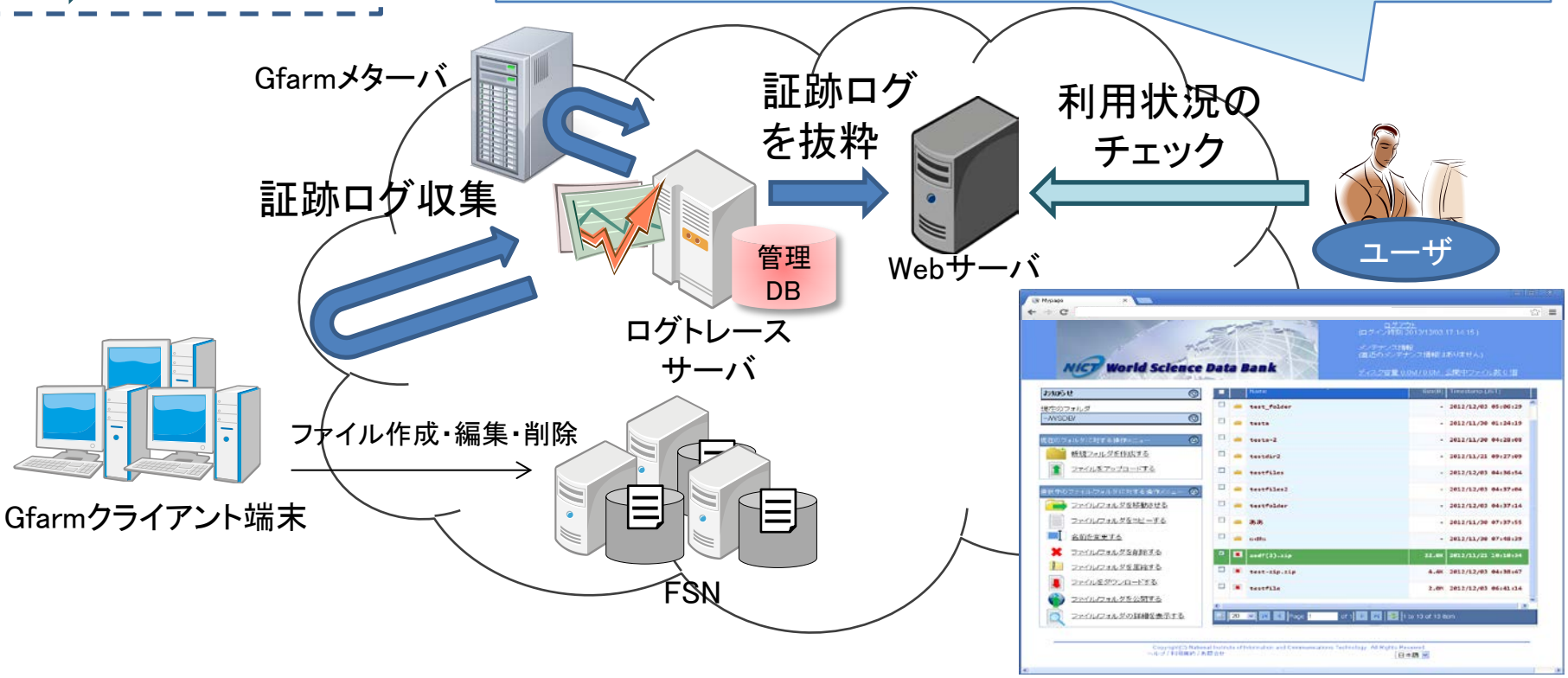
※ 大西 健司ら, "クラウドにおけるデータトレーサビリティ機能の検討および大規模分散ストレージシステム上での実装と評価", 情報処理学会研究報告,HPC-130(35), PP.1-8, 2011.

◆ 責任追跡性への取り組み：ログトレースシステム（2）

- ユーザに対して、データのライフサイクルのイベント（作成、更新、複製、削除など）を確認可能なシステム

凡例
- 証跡ログの流れ
- ユーザの操作

以下の情報を確認することができる。
●ファイル作成、削除数、更新数の推移、ファイル冗長化数



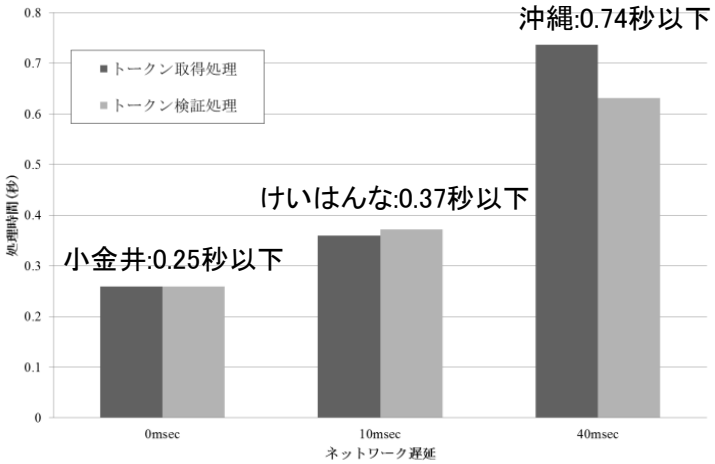
ユーザの情報表示画面(WSDBank Webアプリ)

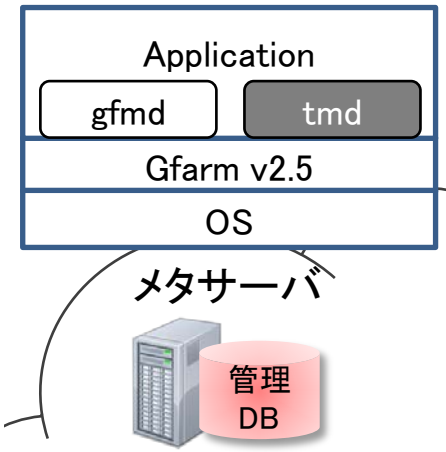
◆ 完全性への取り組み：時刻認証を用いたファイル検証システム

NICT ScienceCloud

- データのライフサイクルのイベント（作成、更新、参照、複製、削除など）発生時にデータの完全性と存在証明を保証する（プロトタイプ）システム

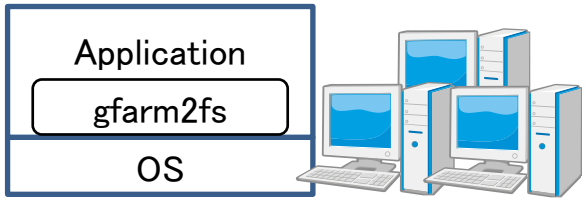
システム処理性能は**1秒以下**
（200KBデータが地域分散保存された場合）





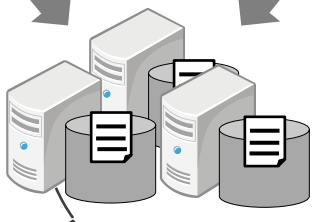
該当ファイル情報の提供

該当ファイルの時刻認証

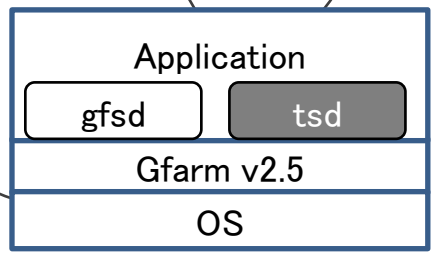


Gfarmクライアント端末

ファイル作成・編集・削除



FSN



◆可用性への取り組み（1）：SLA定義の現状

NICT ScienceCloud

・ほとんどのサービスが契約文書レベルの内容のみ（SLA定義方法を記載しているものはない）
・ストレージサービスのSLAの評価項目および定義方法は不透明・非公開

仮想化サーバに係わるSLAサービス（※）

月間稼働率＝(1－累計障害時間÷月間総稼働時間)×100

- 月間総稼働時間＝仮想サーバ台数×24（時間）×30（日）
- 累計障害時間＝仮想サーバの障害時間
  - （当社の責めに帰すべき理由により、仮想サーバが利用不可能になった状態が継続して発生した時間とし、1分未満の時間は切り捨てるものとする）
- 適用除外事項
  - メンテナンス（緊急メンテナンスを含む）による停止の場合
  - 本サービスの機能としての中断（HA機能）による場合
  - サービス利用者または第三者からの攻撃、妨害等による場合
- 稼働率を下回った場合
  - 99.99%に満たなかったと当社が判断した場合、当月分の料金の10%に相当する金額を返還する。

オブジェクトストレージサービスに係るSLAサービス（※）

月間稼働率＝(1－累計障害時間÷月間総稼働時間)×100

- 月間総稼働時間＝60（分）×24（時間）×30（日）
- 累計障害時間＝オブジェクトストレージの障害時間
  - （当社の責めに帰すべき理由により、仮想サーバが利用不可能になった状態が継続して発生した時間とする）
  - 利用不可能な状態とは、特定の5分間のエラー率（エラー数／総リクエスト数）が100%である状態とする
- 適用除外事項
  - 本サービスの利用の中止による場合
  - 当社サービスの範囲外でのインターネットアクセス、ネットワークの障害、端末の障害による場合
  - お客様システム側の特定のハードウェア・ソフトウェアに依存した、当社サービスとの接続性による場合
  - コントロールパネルの不具合による場合
  - サービス利用者または第三者からの攻撃、妨害等による場合
- 稼働率を下回った場合
  - 99.9%に満たなかったと当社が判断した場合、当月分の料金の10%に相当する金額を返還する。

<u>標準的な指標は無く、サービス提供側で過去の経験や社内サービス規定に基づいて定義されているのが現実</u>

※出典：NTTコミュニケーションズCloudnSLA規定

◆可用性への取り組み（2）：GfarmのSLA定義

NICT ScienceCloud

1．分散ストレージシステムのSLA評価モデルを作成

- 冗長化構成を採用する等、各要素の信頼性要求条件を定めていくことで、不稼働率が不稼働率曲線内に納まるようにする

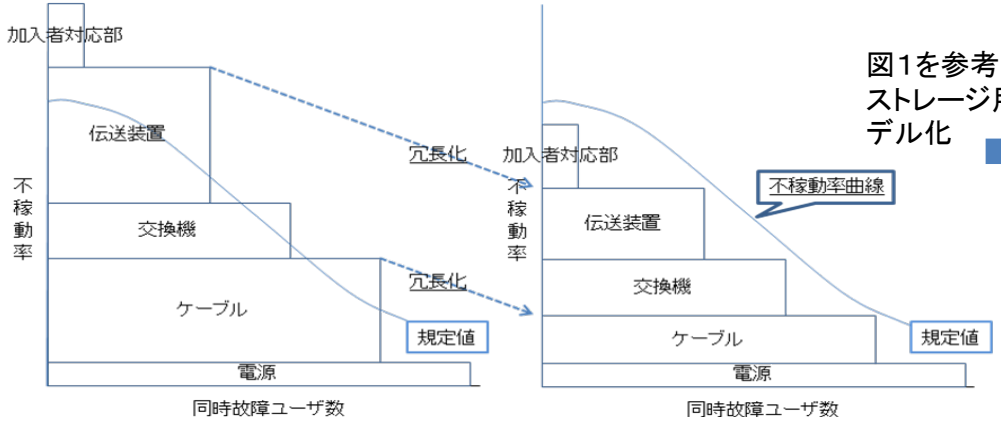


図1．通信網における不稼働率曲線活用方法

図1を参考に分散ストレージ用にモデル化

前提条件

- ネットワークのSLA99.99以上であること
- メタサーバ(MDS)は冗長構成かつロケーション分散を図っていること
- ファイルサーバ(FSN)はロケーション分散を図っていること
- レプリケーション機能がファイルを2つ以上複製し、ロケーション分散して保存する設定であること

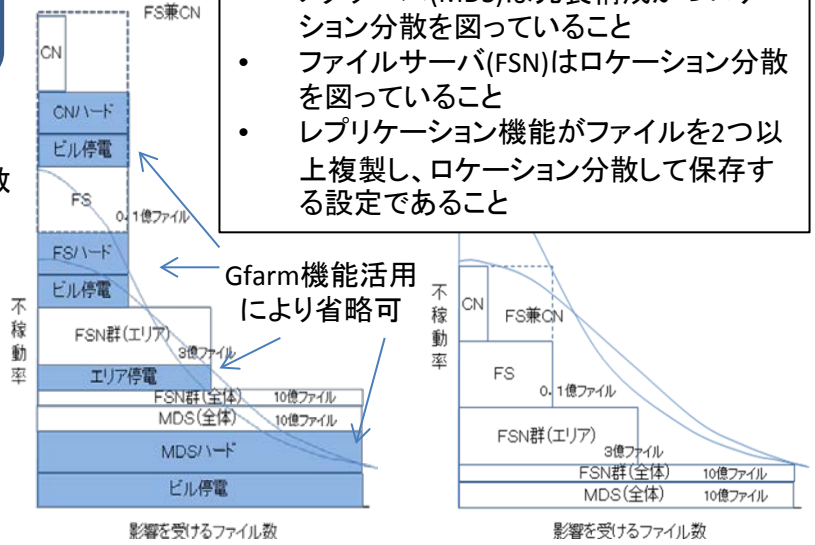


図2．Gfarmを用いた分散ストレージシステムのSLA評価モデル

2．稼働率の算定

分散ストレージシステムの稼働率はaとbの稼働率から算定する

a.　システム構成から算出する稼働率　⇒　システムの稼働率

b.　運用データから算出する稼働率　⇒　ファイルアクセスの成功率

システム稼働率算出方法

単体稼働率A= MTBF(※1) / (MTBF + MTTR(※2))

直列接続の稼働率As= A * A

並列稼働率Ap＝1 - (1 - A) * (1 - A)

a. メタサーバおよびファイルの冗長化構成におけるシステム稼働率算出式

稼働率A(system)＝Ap(MDS) * Ap(FSN)

例.　MTBF=50000H, MTTR=24Hとした場合、0.9995202となる

b. ファイル数をFとした場合の稼働率(ファイルアクセス成功率)算出式

稼働率F(success)＝（F(all)－F(error)）／F(all)

分散ストレージシステムの稼働率算出式

稼働率A(Gfarm)＝A(system) * F(success)

※1. MTBF:平均故障間隔

※2. MTTR：平均復旧時間

◆ 可用性への取り組み(3):
NICTサイエンスクラウドで試行したGfarmのSLA評価試験結果

a. システム稼働率

機器のMTBF(平均故障間隔)とMTTR(平均復旧時間)条件

- ルータ:MTBF(R)=200000H
- L2SW:MTBF(SW)=120000H
- メタデータサーバ:MTBF(MDS)=50000H
- ファイルサーバ:MTBF(FSN)=50000H
- クライアント:MTBF(CN)=50000H
- MTTR=24H

①MDS稼働率: MDS・L2SW・Rの冗長化構成(小金井・けいはんな)

稼働率Ap(MDS)＝1 - (1-A(MDS)*A(SW)*A(R)) * (1-A(MDS)*A(SW)*A(R))
＝0.9999866

②FSN稼働率: FSN・L2SW・Rの2箇所分散構成(小金井・けいはんな)

稼働率Ap(FSN)＝1 - (1-A(FSN)*A(SW)*A(R)) * (1-A(FSN)*A(SW)*A(R))
＝0.9999866

③システム稼働率

稼働率A(system)＝Ap(MDS) * Ap(FSN)＝0.9999732

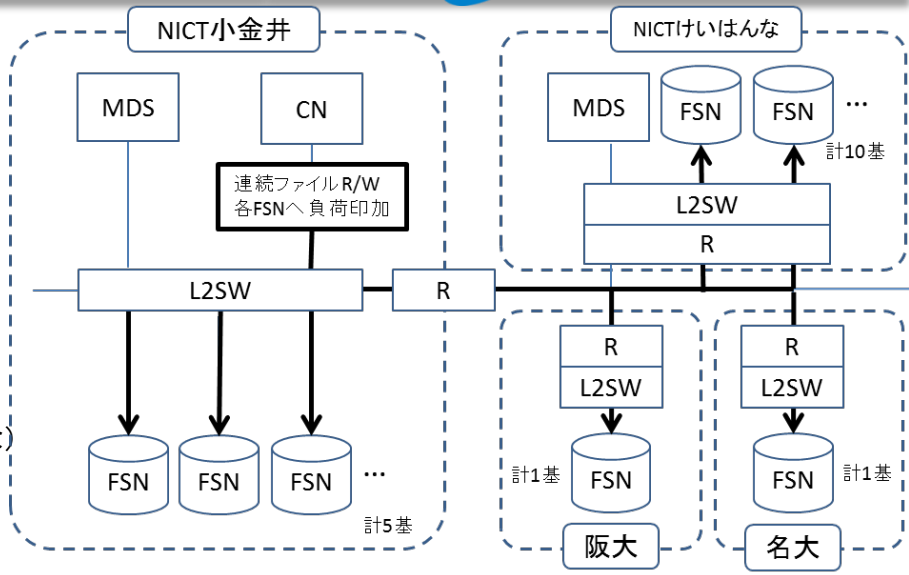


図3. SLA評価するシステム構成モデルと負荷印加イメージ

b. ファイルアクセス成功率

- ネットワークトラフィック量を徐々に増大させる負荷印加試験(16Mbps - 200Mbps)
- 負荷印加試験期間:2/8 - 3/1 (22日間)
- 総読み書き回数: 2,424,615回
- エラー発生件数: 84件
  - 80件: SSHログイン失敗(LDAP認証エラー)
  - 4件: ファイルアクセス失敗(ネットワークボトルネックによるタイムアウト)

表1. ファイルアクセス負荷印加条件(NICTサイエンスクラウド過去アクセス履歴参考)

ファイルサイズ	1KB	50KB	500KB	5MB	50MB	500MB	5GB
ファイル作成間隔(秒)	3.0	1.0	2.0	2.5	30.0	40.0	600.0
1分あたりのファイル作成数	20.0	60.0	30.0	24.0	2.0	1.5	0.1
負荷(Mbps)	0.003	0.31	1.93	16.00	13.33	100.0	68.67

④ファイルアクセス成功率

稼働率F(success)＝(F(all)－F(error))／F(all)＝ (2,424,615-4)/2,424,615 = 0.9999983

Gfarmを用いた分散ストレージシステムの稼働率= A(Gfarm)= A(system) * F(success)= 0.9999715

◆まとめ

- 広域分散ファイルシステムのセキュリティ要件とGfarmの対応状況
  - セキュリティ要件として、機密性、可用性、完全性は必須、信頼性、真正性、責任追跡性も対応すべき
  - Gfarmは信頼性、機密性、可用性に対する機能は有しているが、完全性、責任追跡性、真正性は未対応ならびにSLAは未定義
- Gfarmのセキュリティに対する取り組み
  - ログトレースシステムの開発・導入により責任追跡性を保証可能
  - 完全性を保証するための時刻認証を用いたファイル検証プロトタイプシステムを評価中
    - ✓ 真正性は、ログトレースシステムならびにファイル検証システムの組み合わせで対応可能
  - 標準的なSLAの定義・評価基準がないため、GfarmのSLAを新規に定義し評価。
  - NICTサイエンスクラウド環境を実例としてSLA99.99%を達成できることを確認
- 今後の課題
  - ログトレースシステムの準リアルタイム化
  - ファイル検証システムの実用性の検証
  - NICTサイエンスクラウドのSLA評価

ご清聴ありがとうございました