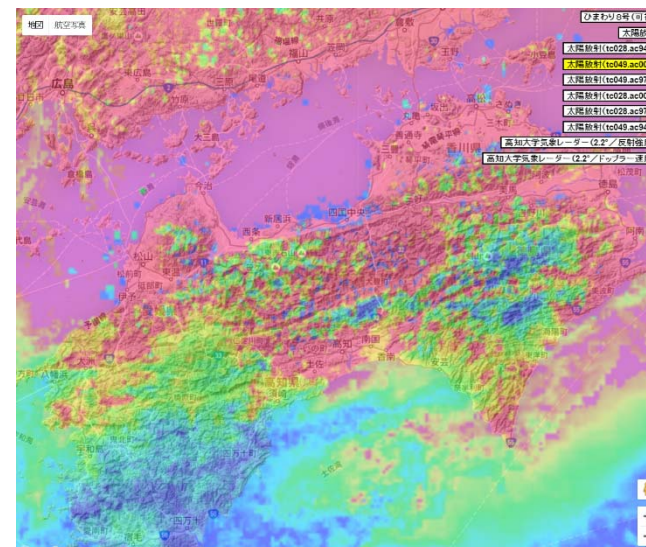
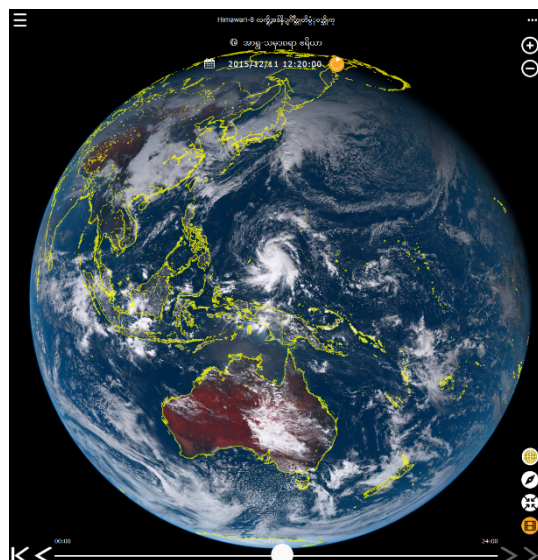


Gfarm/Pwrakeを活用した 様々な科学研究アプリケーションと それを支える高速データ通信プロトコルの開発

村田健史、山本和憲、長屋嘉明(情報通信研究機構)

鵜川健太郎、村永和哉、村山純一、鈴木豊(セック)

水原隆道、高木文博、福島啓介(クレアリンクテクノロジー)



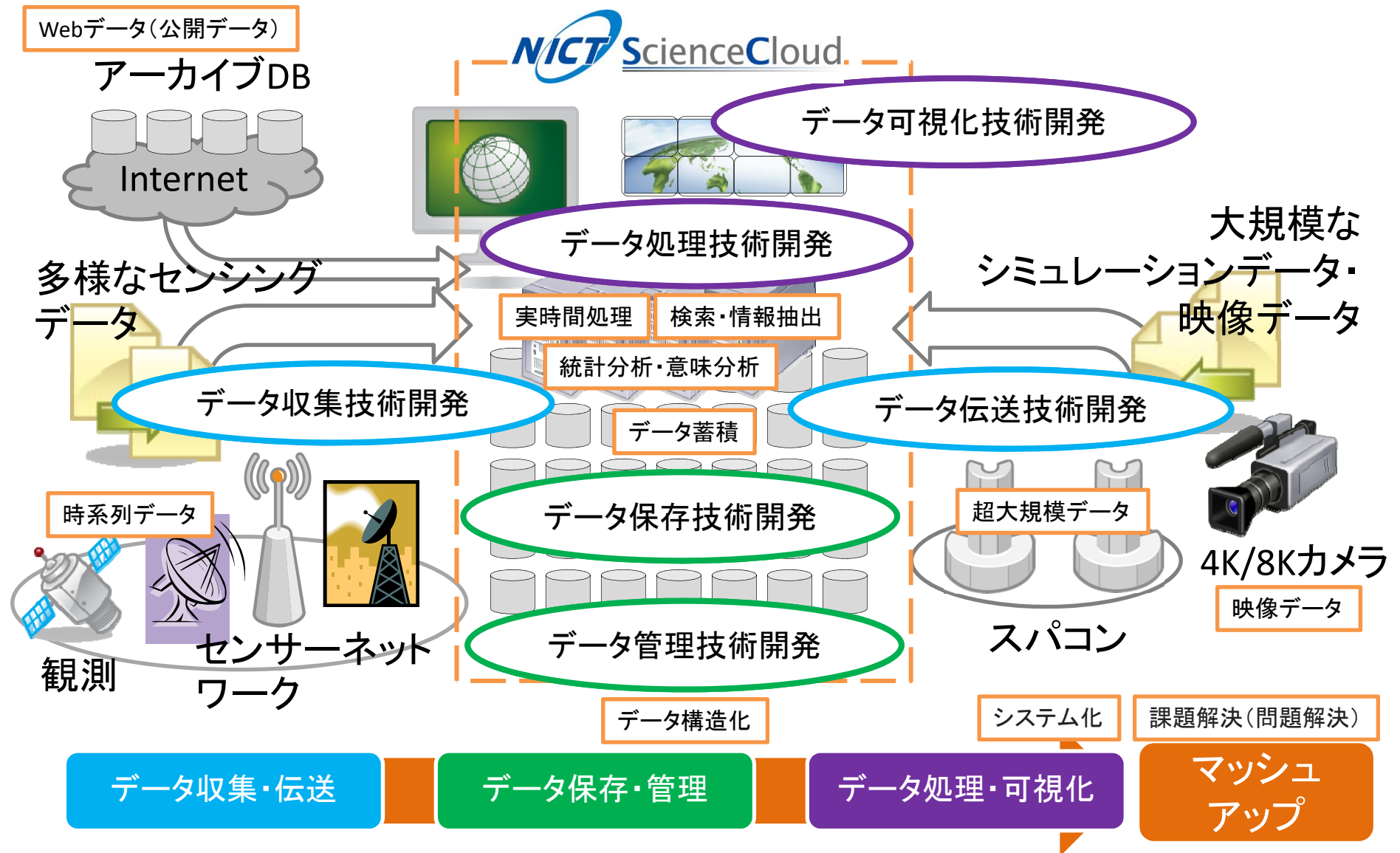
Gfarm/Pwrakeを活用した様々な科学研究アプリケーションと それを支える高速データ通信プロトコルの開発

- NICTサイエンスクラウド紹介（Gfarmを中心に）
 - データ収集・データ管理・データ処理
- 可視化アプリケーション紹介
 - なぜGfarm/Pwrakeが必要なのか
 - Gfarm/Pwrake活用事例
- 高速データ通信プロトコルとGfarm
 - HbVRS（高速仮想遠隔ストレージシステム）
 - HpFP（High-performance and Flexible Protocol）
 - HpFP版HbVRSの現状

Gfarm/Pwrakeを活用した様々な科学研究アプリケーションと それを支える高速データ通信プロトコルの開発

- NICTサイエンスクラウド紹介 (Gfarmを中心に)
 - データ収集・データ管理・データ処理
- 可視化アプリケーション紹介
 - なぜGfarm/Pwrakeが必要なのか
 - Gfarm/Pwrake活用事例
- 高速データ通信プロトコルとGfarm
 - HbVRS (高速仮想遠隔ストレージシステム)
 - HpFP (High-performance and Flexible Protocol)
 - HpFP版HbVRSの現状

NICTサイエンスクラウド(2010～)システムコンセプト

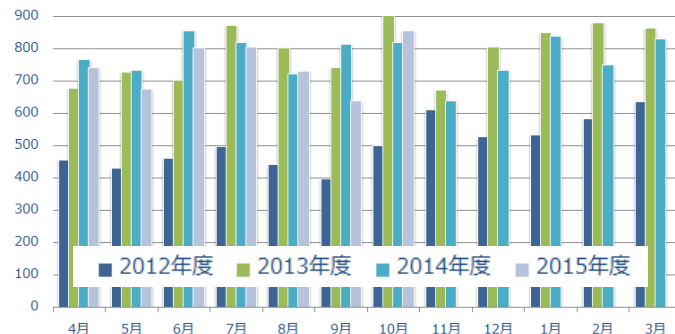
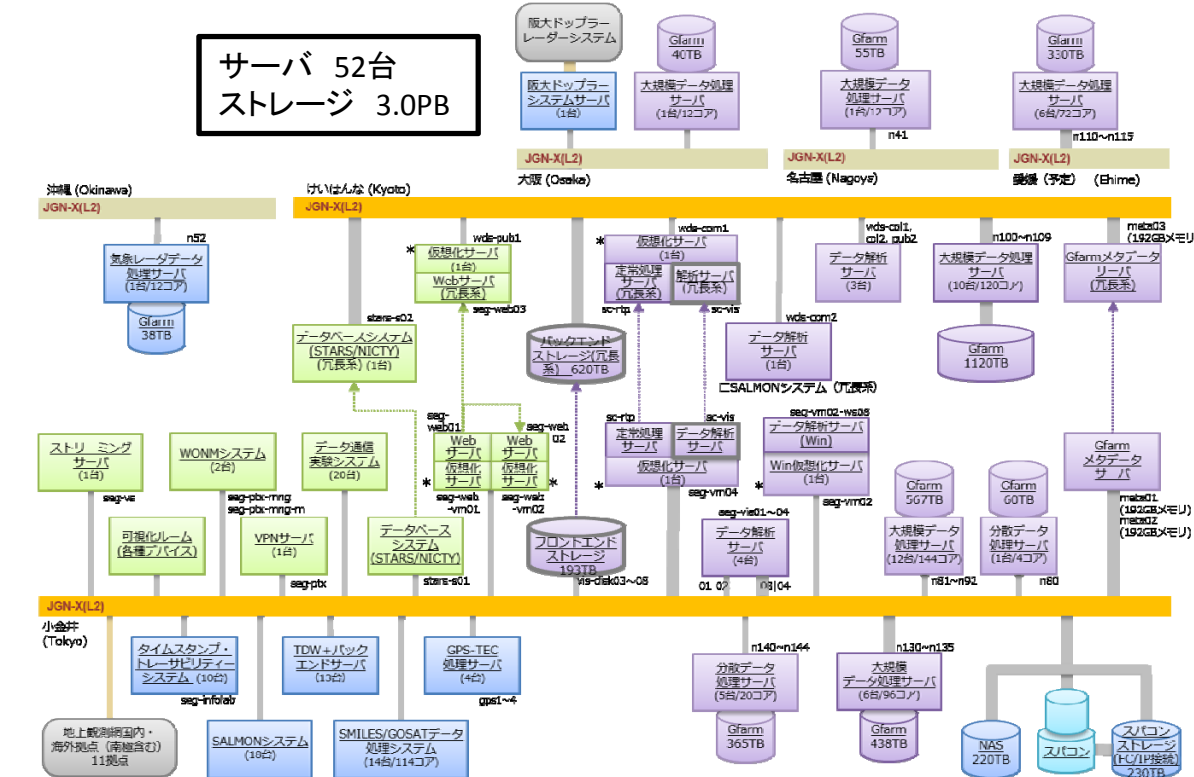


◆サイエンス(アカデミック)クラウドの取り組み

NICTサイエンスクラウド

宇宙科学 地球環境 人間科学 ICT

サーバ 52台
ストレージ 3.0PB



サイエンスクラウドを活用した
査読付き論文数 (H.24~H.27)
90件

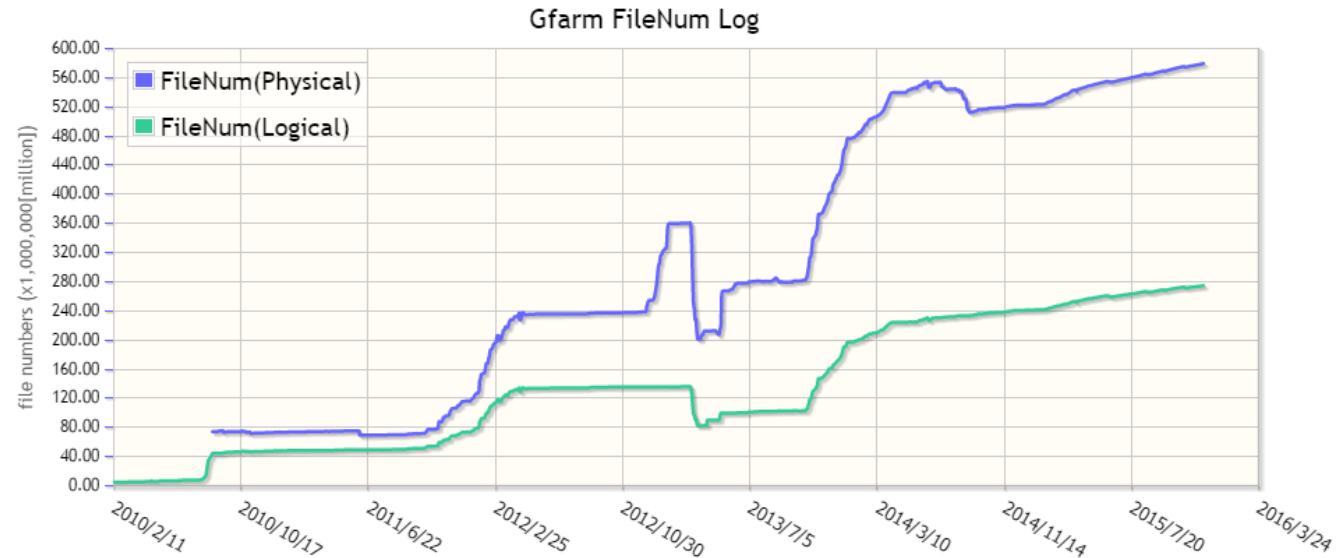
ユーザアクセス状況
(600~900アクセス/月)

ICSU/WDS事務局に選定
(2010年)

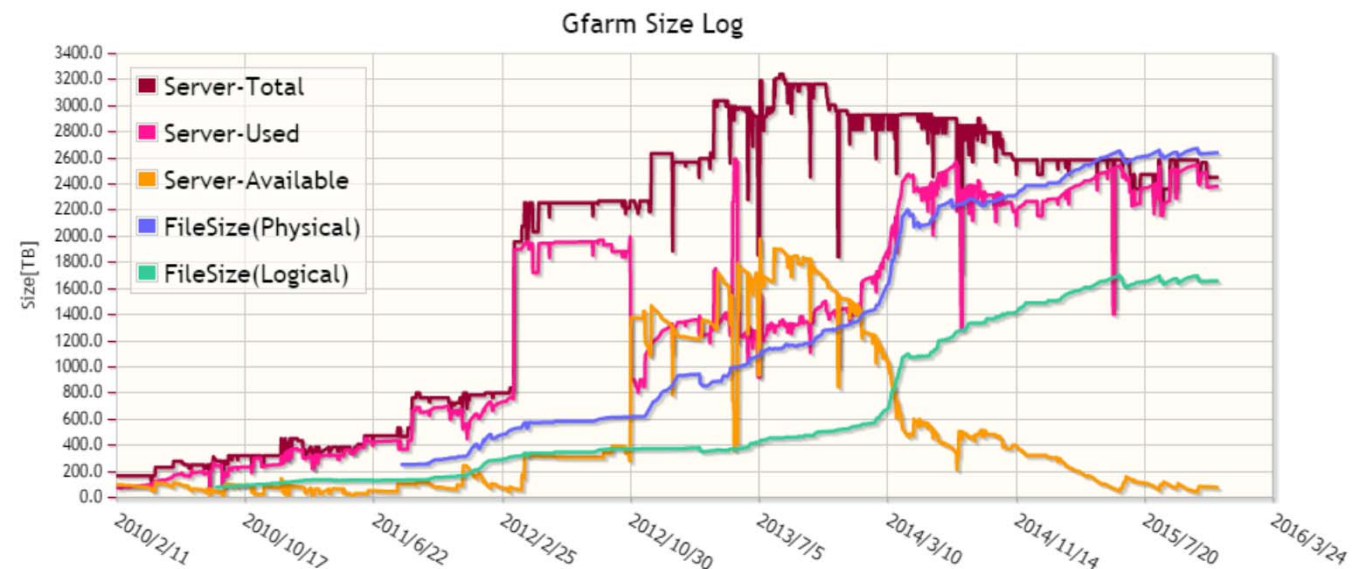
クラウド利用プロジェクト(計30)

分散ストレージシステム(Gfarm)履歴 (2010年2月～2015年12月)

- 管理ファイル数
(論理・物理)

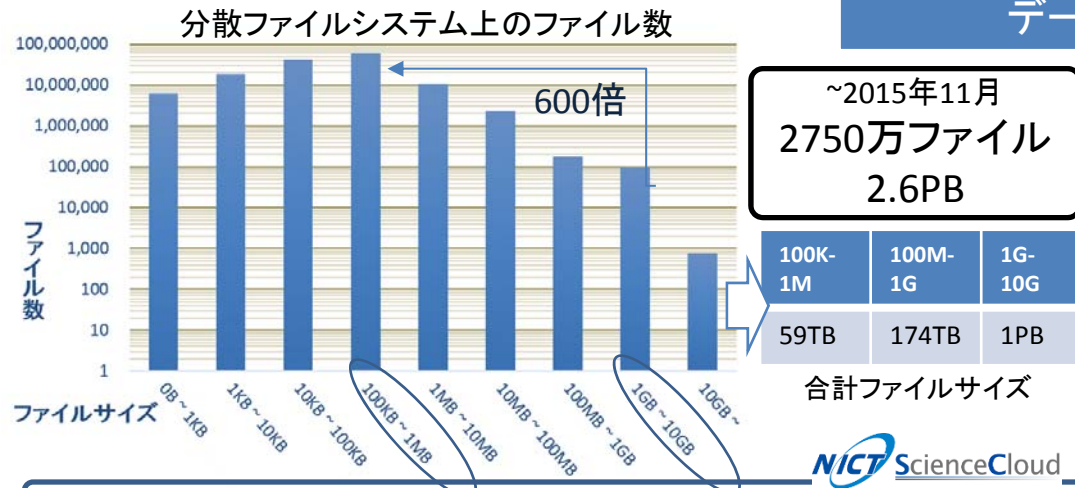


- 参考
- ストレージサイズ
 - 管理ファイルサイズ(論理・物理)



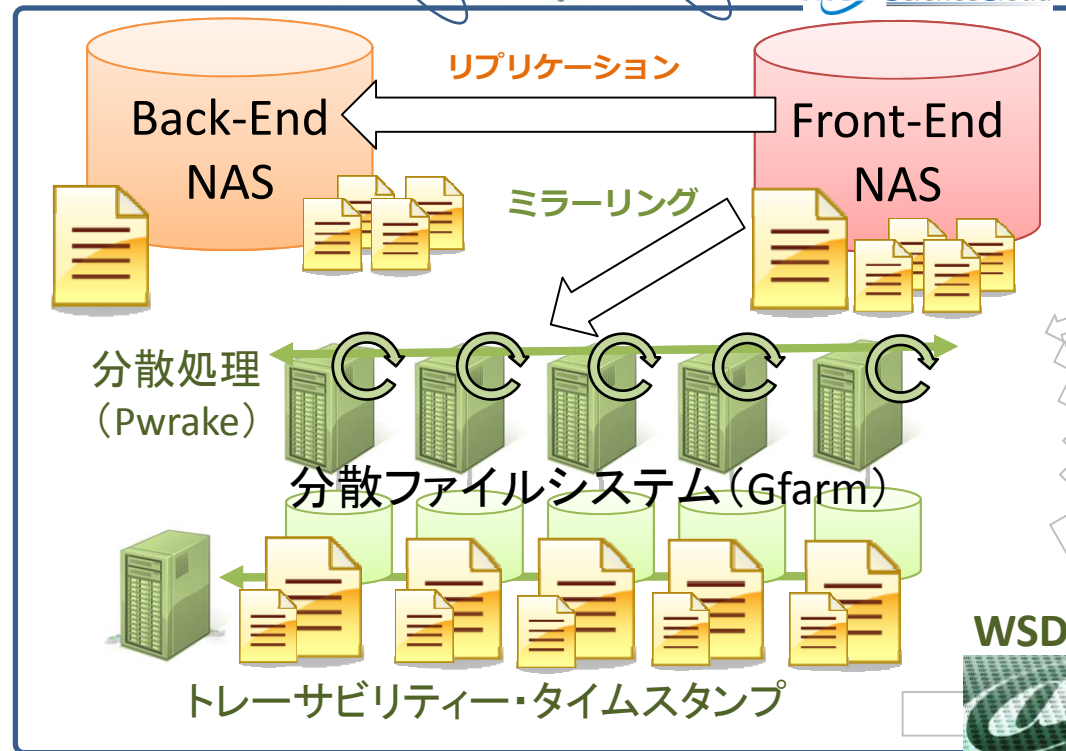
◆科学データ収集と保存(全体像)

データ転送・収集・保存システム概要

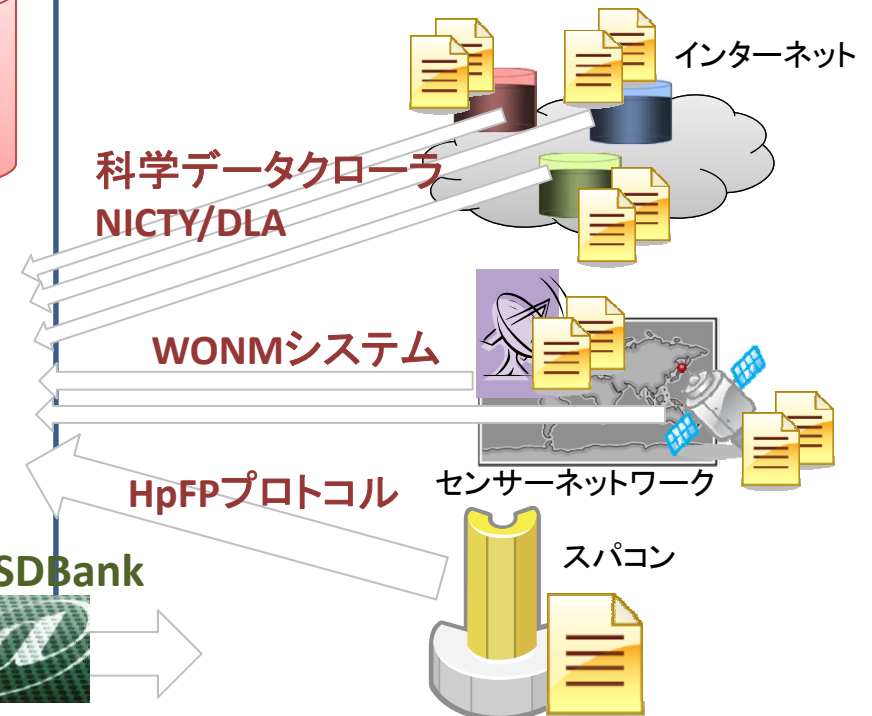


データ種別	データ提供組織名	収集ツール	データファイル数	総データサイズ
GNSS観測	UNAVCO, 国土地理院など	NICTY+独自ツール	23,506,753	9.6TB
SEALION	NICT	WONM	60,787	1.6TB
GOSAT衛星	JAXA, ECMWF	独自ツール	3,737,123	58.4TB
フェーズドレイ 気象レーダー	大阪大学 (NICT)	独自ツール	2,358,677	217.1TB
SMILES	JAXA	独自ツール	564,253	1.4TB
南極 (電磁気観測)	NICT	WONM	280,859	424GB
地磁気データ	NICT, 京都大学	NICTY+独自ツール	277	1.5TB
ひまわり衛星	気象庁	NICTY+(HDD搬送)	26,868	26.8GB
GEOTAIL衛星	京都大学 生存圏研究所	NICTY	29,840	20GB
GOES衛星	NOAA	NICTY	1,638	5GB
ACE衛星	NOAA	NICTY	759	20GB
こだま衛星	JAXA	NICTY	3,890	1.8GB
KAGUYA衛星	金沢大学	-(伝送なし)	3,227	221GB
宇宙天気リアルタイム シミュレーション	NICT	NICTY (メタデータのみ)	63	3.5TB

ユーザ約100名
平均6000DL/月

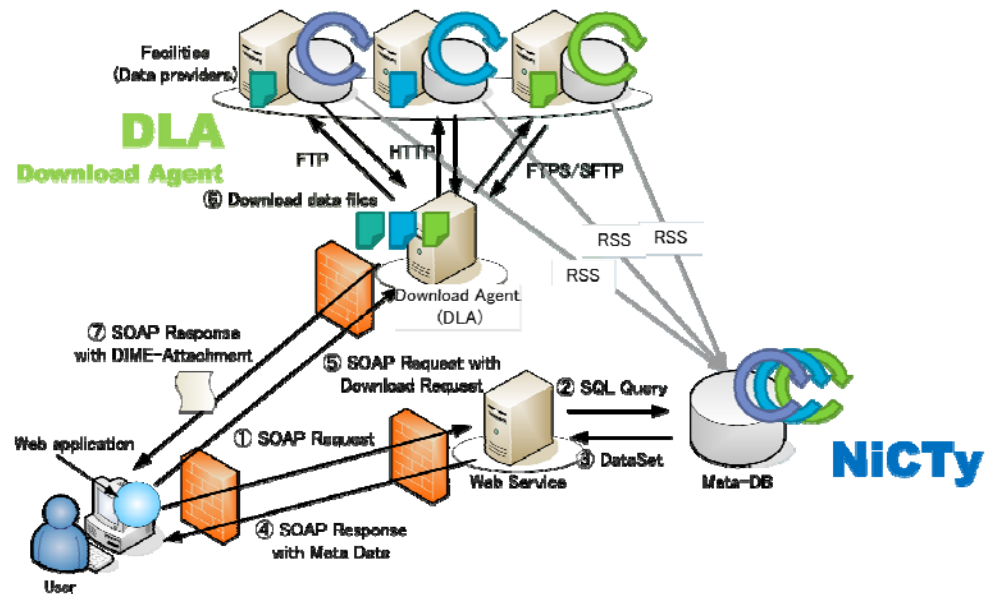
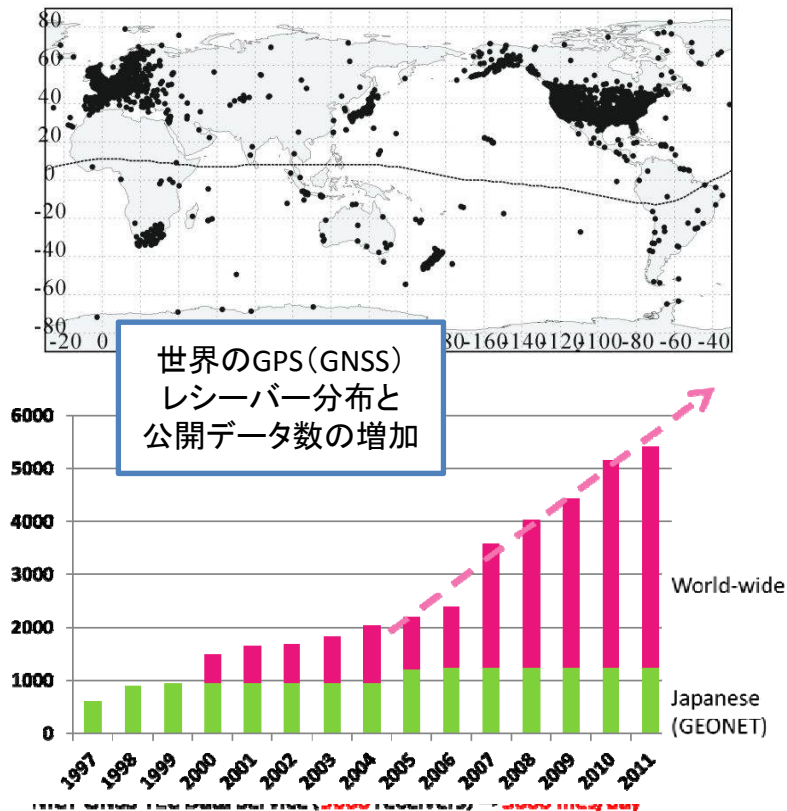


プロジェクトごとのデータ収集状況

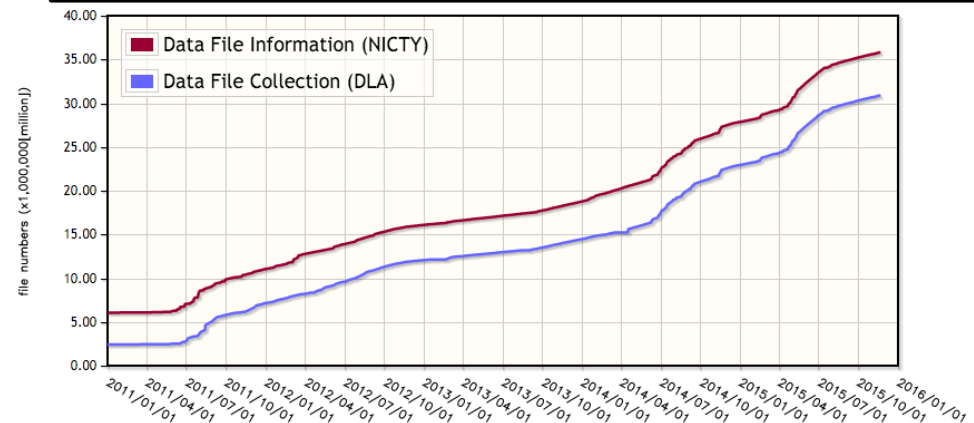


◆インターネット公開データのクローリング

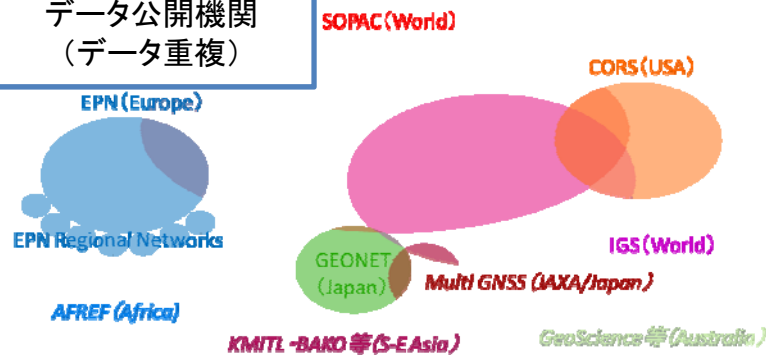
NICTY(時系列データクローラ)によるGPSデータ収集



クローラ=インターネット上のデータが日々自動収集される



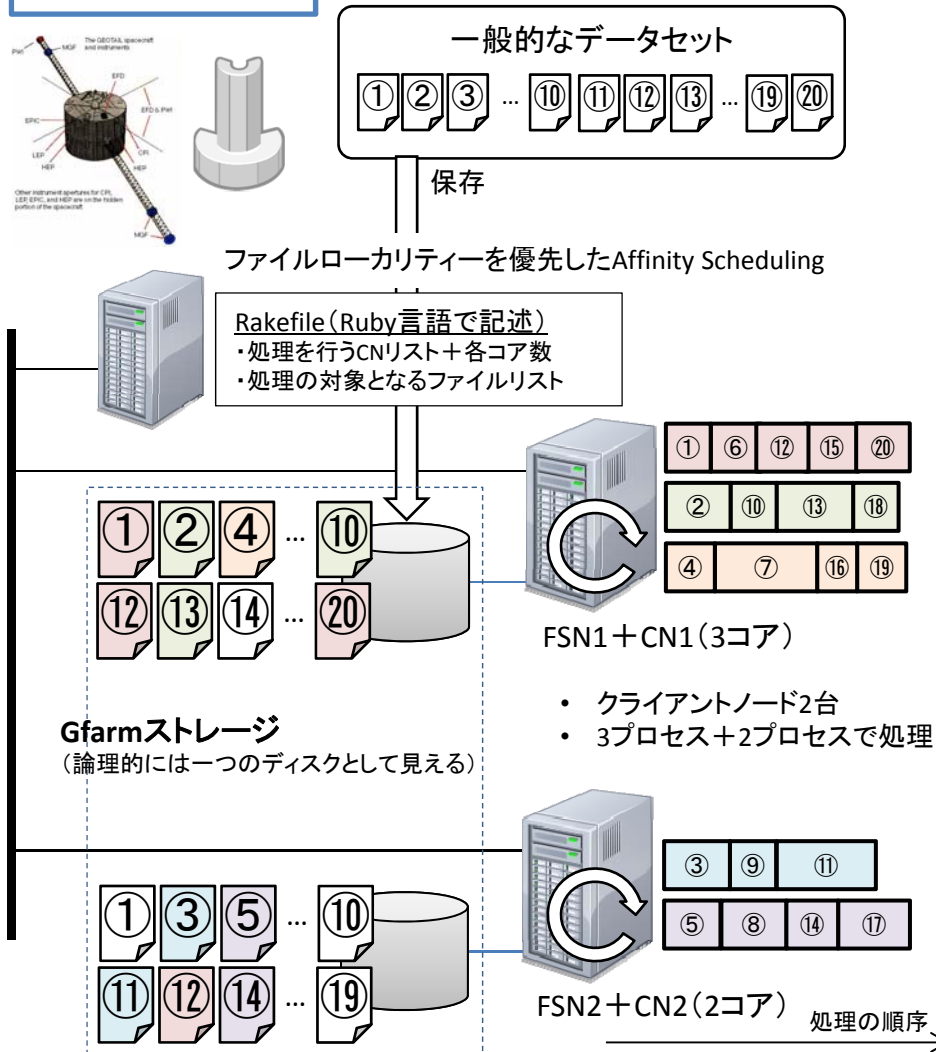
データ公開機関
(データ重複)



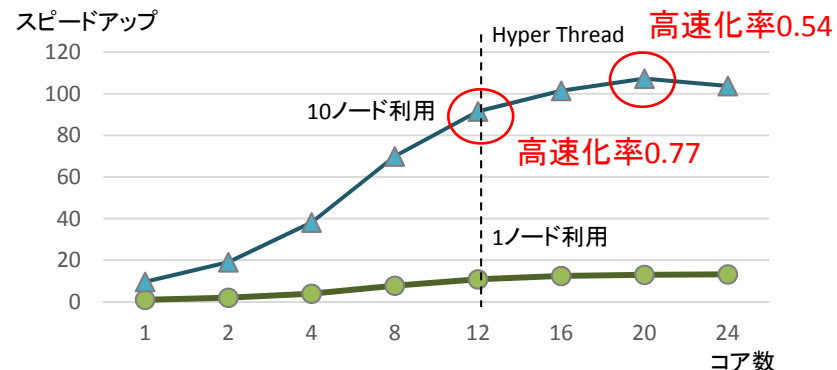
◆観測・シミュレーションデータ解析技術開発

Gfarm/Pwrakeによる並列分散処理

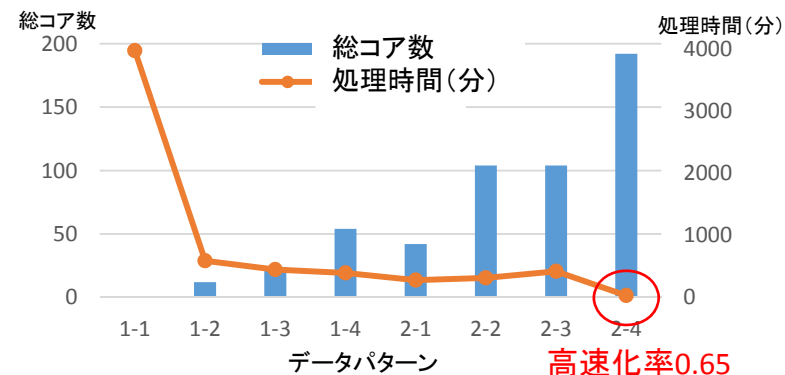
時系列観測



■GEOTAIL科学衛星データ処理(合計サイズ190GB、ファイル数27576)



■地球磁気圏シミュレーション可視化(合計サイズ2.3TB、ファイル数1000)



その他のGfarm/Pwrake利用実績

- フェーズドアレイレーダ3Dデータリアルタイム処理
- ひまわり8号衛星データ可視化
- 太陽放射データ可視化
- 140年新聞記事データ可視化
- エピゲノムデータ可視化

ミドルクラス
で有効

◆セキュアで信頼性が高いクラウド技術

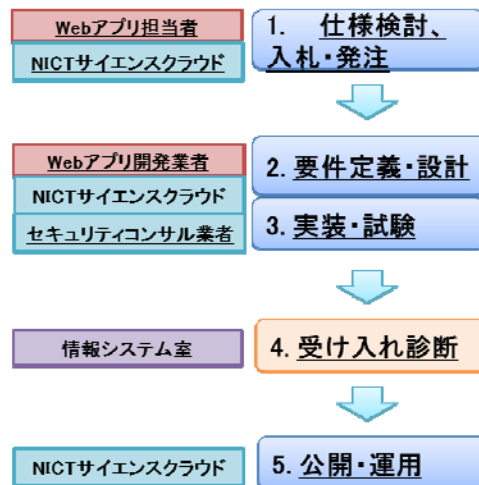
セキュアWeb技術

仕様検討、入札・発注

- ✓ Webアプリケーション製作仕様書サンプルを提供
- ✓ Webアプリケーション脆弱性検査仕様書サンプル提供

要件定義・設計

- ✓ セキュリティコンサルティングが必要か否かの相談
- ✓ セキュリティコンサル業者との打ち合わせを実施
- ✓ セキュリティコンサル実施計画を提案
- ✓ 参考となるセキュリティ要件定義書を提供
- ✓ セキュリティ要件定義書を点検しアドバイス（レビュー）
- ✓ 設計書のセキュリティ部分を点検しアドバイス（レビュー）
- ✓ セキュリティ試験内容を点検しアドバイス（レビュー）
- ✓ 言語・フレームワークの選定に対するアドバイス



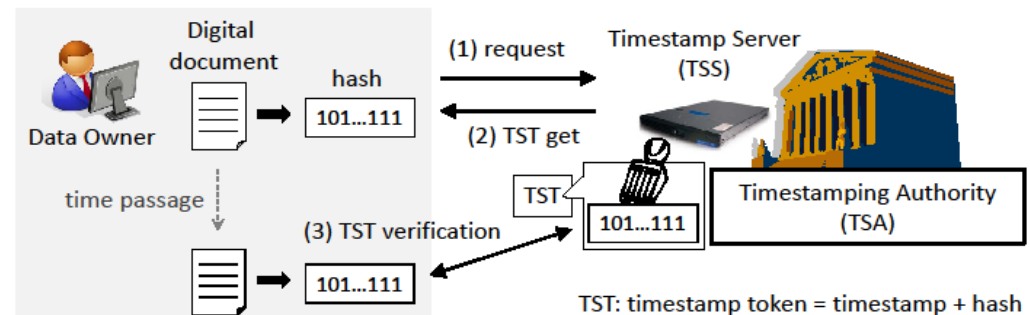
分散ファイル管理技術(タイムスタンプ)

背景

- 時刻認証局(TSA)は日本では4社のみ(アマノ、セイコーソリューションズ、NTTデータ、北海道総合通信網)
- 現状では各社とも自社のTSAサーバでのタイムスタンプしか認めておらず**クラウドの高速・大容量のデータベースには未対応**(300スタンプ/秒が限界)
- 今後はこのようなクラウドに直結(内包)タイムスタンプ方式の実用化を視野に入れたビジネスモデルが必須

サイエンスクラウドの成果

- 分散ストレージシステムとタイムスタンプ・サービスの協調機能によりデータトレーサビリティを実現
- 重要なデータの完全性と真正性を担保



完全性 (integrity) と真正性 (authenticity)

- ・ 文書の作成者・作成時期、紙文書などと電子化した文書が同一であることが確認できること。
- ・ 保存義務期間中に文書が改ざん・消去されないこと、改ざんされたことが確認できること。

総合的信頼性認証機能(トレーサビリティ+タイムスタンプ)

