



NICTサイエンスクラウド H.25年度成果報告

情報通信研究機構 統合データシステム研究開発室
村田健史・渡邊英伸



平成26年1月24日
WDS国内推進会議第5回会合

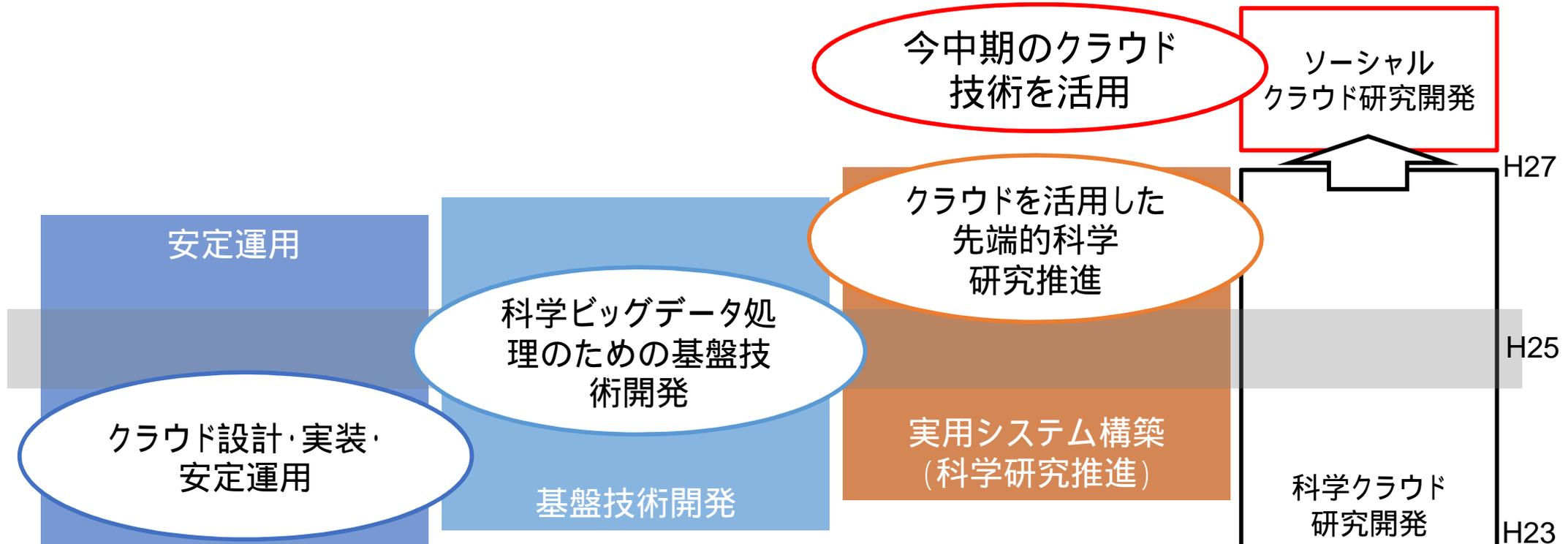
今中期計画の進め方と平成25年度の重点的研究開発課題

<平成25年度の重点的研究開発課題>

セキュリティー対応を含めたクラウド安定運用
統合的科学的データ解析のためのクラウド基盤技術開発
クラウドを活用した先端的科学研究の推進

<平成25年度の成果>

セキュアWebアプリケーション開発手順の確立・トレーサビリティとタイムスタンプによるクラウドストレージ
分散ストレージとタスクスケジューリングによる高スケーラビリティ
科学ビッグデータ処理技術開発・遠隔高速ストレージ技術開発・科学データとソーシャルデータを融合するLODベースのWebアプリ技術開発
次世代ドップラーレーダの準リアルタイム3次元可視化システム開発(連携PJ)・宇宙環境モニタリングのための広域センシング網(観測網)の構築



NICTサイエンスクラウドの基本的な考え方と実績

データ収集・伝送

遠隔高速ストレージ

- どこからでもクラウドストレージを高速に読み書きできる技術
- H.25年度に米国から7Gbpsでデータ読み出しに成功

広域観測網監視システム

- 世界中の観測拠点を監視するシステム
- 南極・赤道・アラスカの16拠点に設置し監視・データ収集

データ保存・管理

セキュアWeb開発手順

- Webアプリケーションをセキュアに開発する手順の確立
- H.25年度に6つのWebアプリを開発・安全に運用

クラウドストレージトレーサビリティ

- 広域分散ストレージのトレーサビリティシステムの開発と運用
- H.25に1.4億ファイルの監視

データ処理・可視化

バーチャルオーロラツール

- 宇宙天気の時系列3次元可視化を行うツール
- H.26年度に観測とシミュレーションの融合で活用予定

ビッグデータ並列処理

- 分散ストレージと分散処理の連携によりビッグデータを高速処理
- クラウド内で20GB/secのI/Oと単サーバの100倍以上の高速処理

個別の基盤技術をマッシュアップして初めて一つのシステム(アプリ)として機能する

NICTサイエンスクラウドの基本方針(中期計画)

- 中期計画前半はクラウド安定運用と基盤技術開発を中心に行う。
- 基盤技術は、データ収集・転送技術、データ管理・保存技術、データ処理・可視化技術の3本柱で構成する。
- 中期計画後半は3つの基盤技術のマッシュアップにより(サイエンスクラウドの特長を活かした)様々な分野の実用的なシステムを構築・運用する。

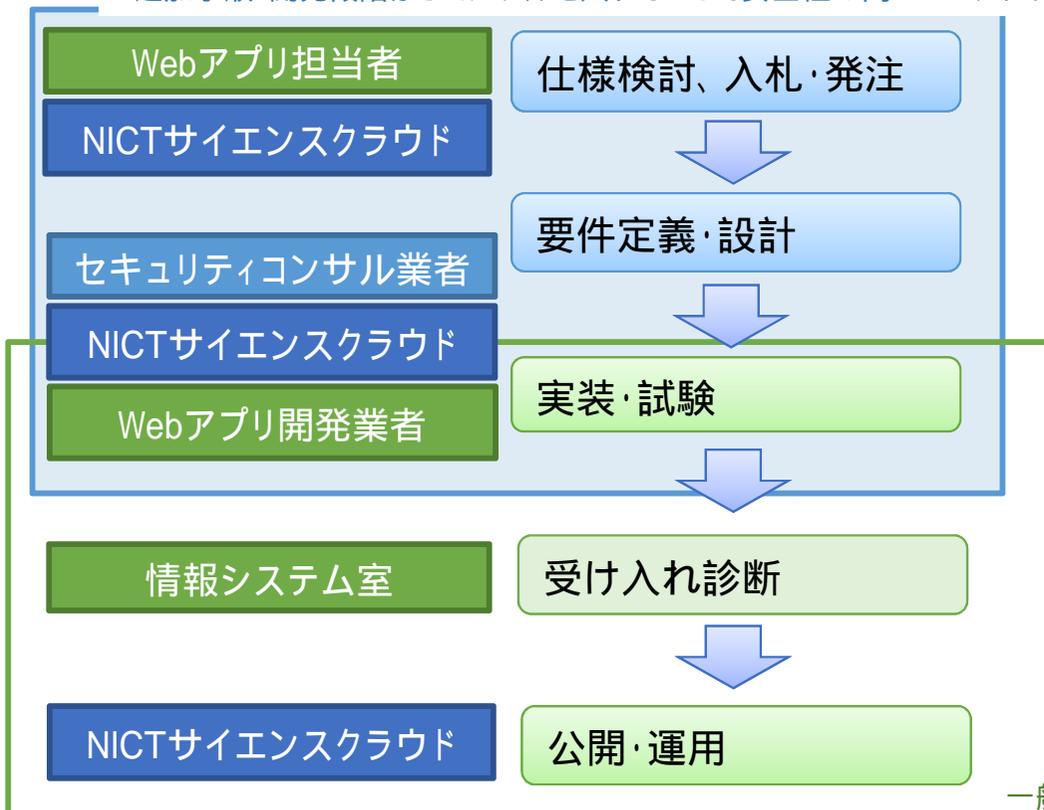
平成25年度の重点的研究開発課題とその成果

セキュアWebアプリケーション開発手順の確立・トレーサビリティとタイムスタンプによるクラウドストレージ

セキュアWeb開発手順確立

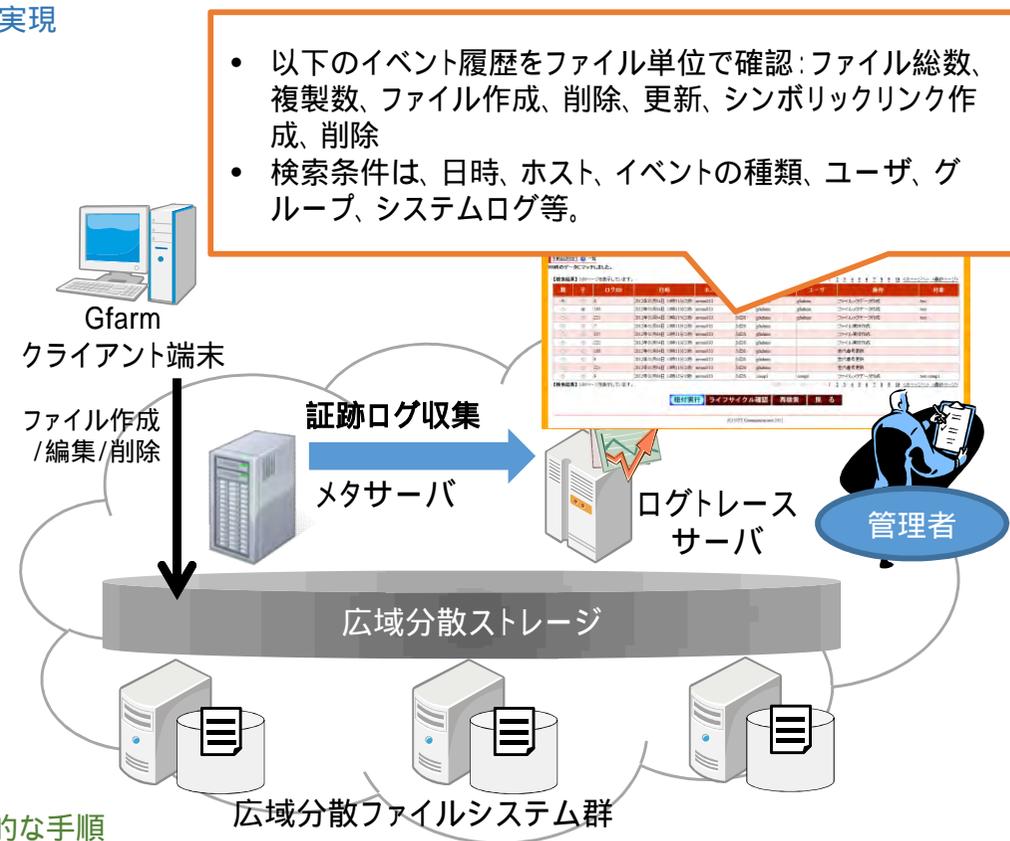
H.23年9月にNICTサイエンスクラウド内で発生したインシデントを受けて、「高いセキュリティ対応のためのWebアプリ開発手順(書)」を確立した。(6つのWebアプリ開発で実証した。)

追加手順: 開発段階からコンサルを入れることで安全性の高いWebアプリを実現



クラウドストレージトレーサビリティ

広域分散ファイルシステムでシステム管理者が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)を追跡できるシステム開発を行い、約1.4億ファイルを監視している。



平成25年度の重点的研究開発課題とその成果

異分野(時系列)データを同時閲覧できる
Webアプリ(開発中)

科学データとソーシャルデータを融合するLODベースのWebアプリ技術開発

NICTY/DLAによるデータ収集

- 気象庁: ひまわり衛星*
- NICT/JAXA: SMILES/GOSAT衛星
- 各国: GPS(GNSS)データ
- 各国: 宇宙天気モニタリング衛星
- JAXA/ISAS: 各種科学衛星*
- 極地研: 全天カメラ画像*

* は(一部)計画中

WONMシステムによるデータ収集

- NICT: 阪大ドップラーレーダ
- NICT: SEALION観測網
- NICT: 地磁気観測網

データ収集と処理の現状

- 観測データ収集・転送実装済み・運用(開始)
- 新聞データRDF化(OWL作成完了・実装未定)
- RDFデータベース構築(一部完了)
- RD公開ツールによるデータ公開(未定)

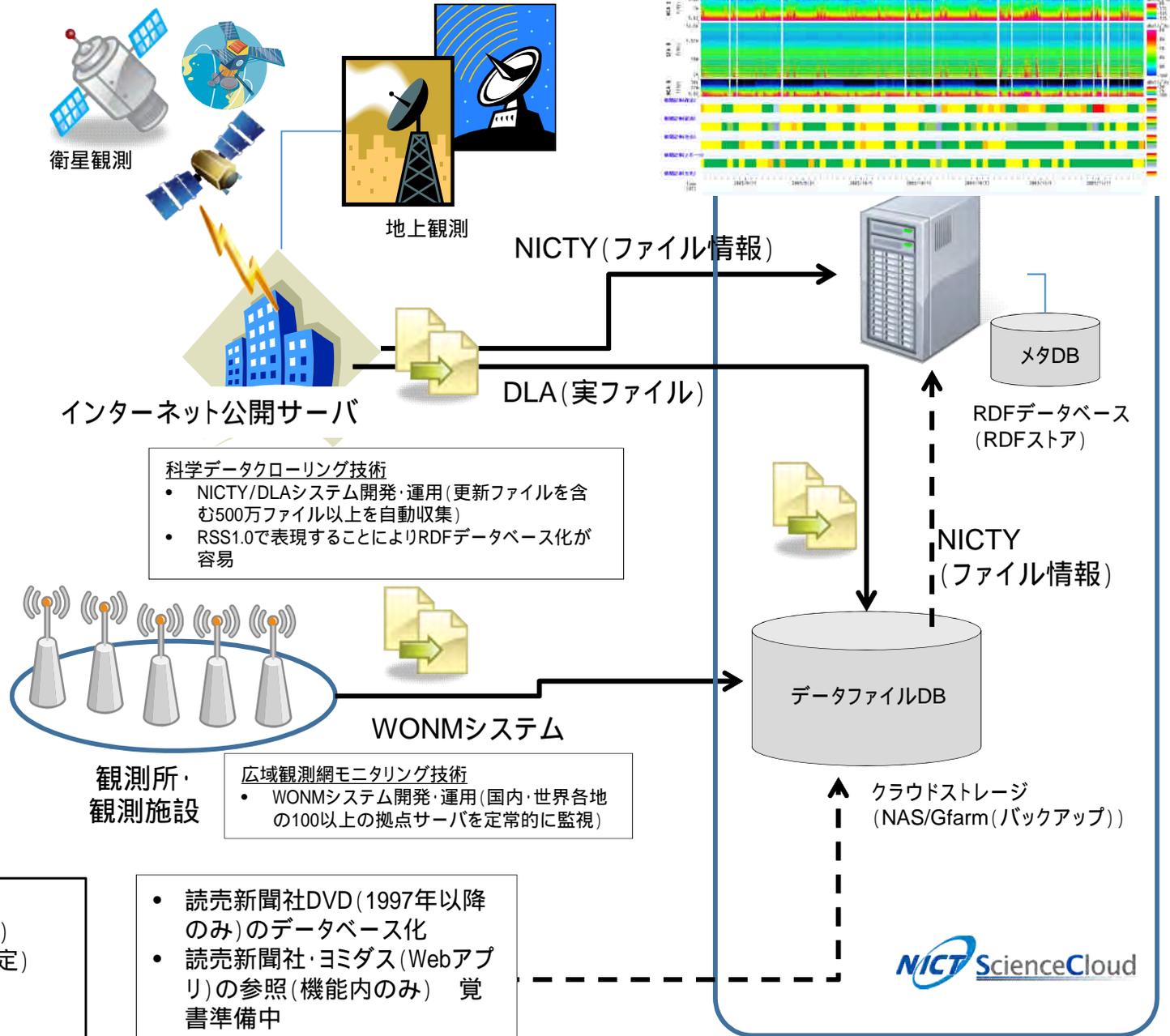
科学データクローリング技術

- NICTY/DLAシステム開発・運用(更新ファイルを含む500万ファイル以上を自動収集)
- RSS1.0で表現することによりRDFデータベース化が容易

広域観測網モニタリング技術

- WONMシステム開発・運用(国内・世界各地の100以上の拠点サーバを定期的に監視)

- 読売新聞社DVD(1997年以降のみ)のデータベース化
- 読売新聞社・ヨミダス(Webアプリ)の参照(機能内のみ) 覚書準備中

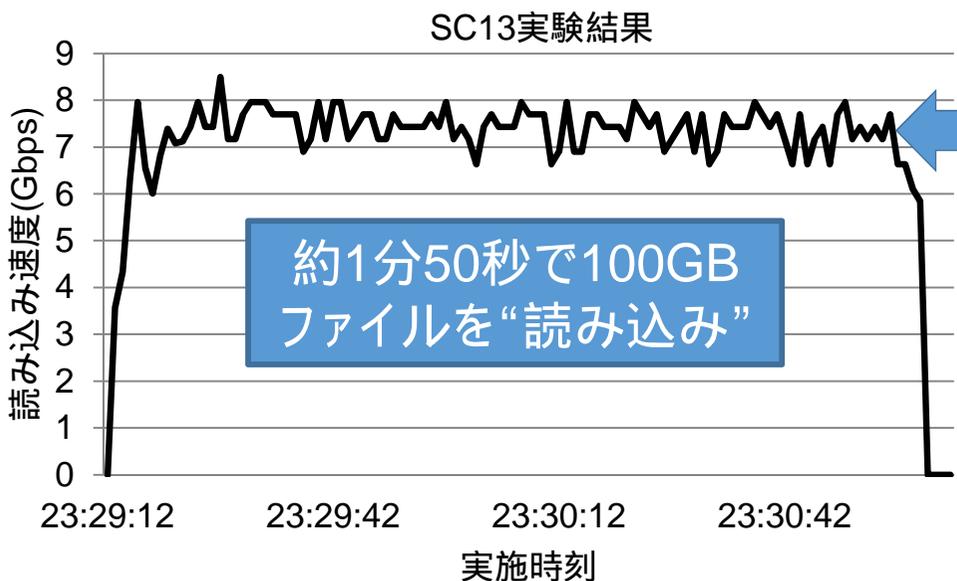
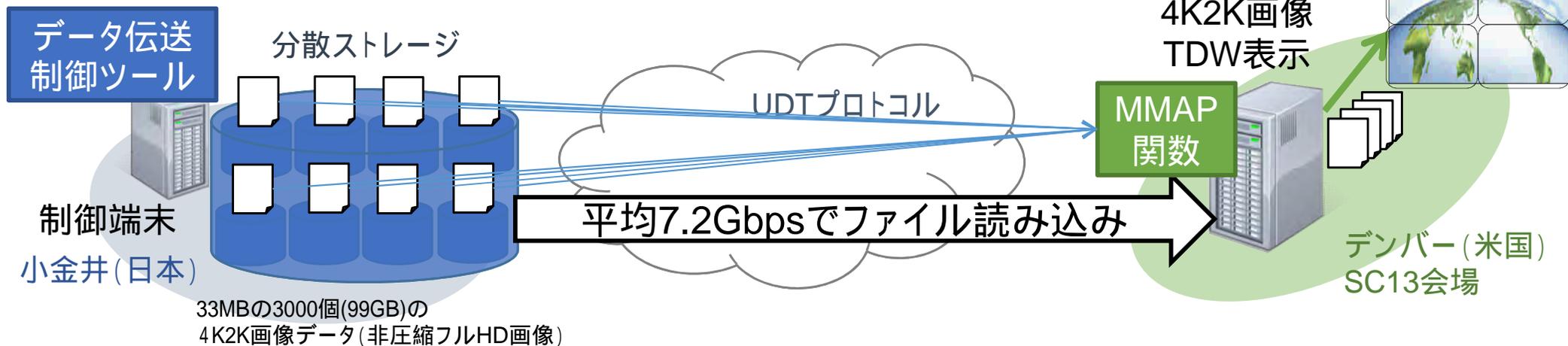


NICT ScienceCloud

平成25年度の重点的研究開発課題とその成果

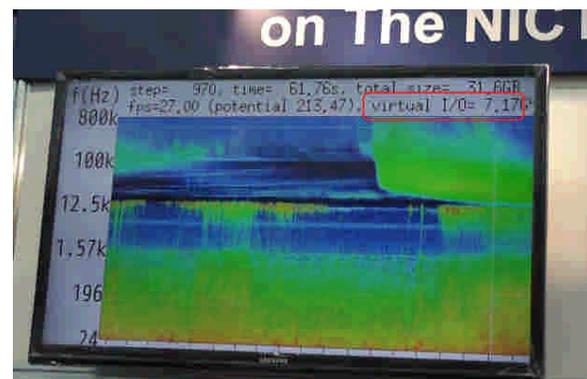
クラウド上のビッグデータにクラウド外からアクセスするための遠隔高速ストレージ技術開発

遠隔地からクラウドストレージデータを利用するための遠隔高速ストレージ技術開発を行った。一般的なローカルストレージより高速な遠隔ストレージを実現した。



7.2Gbps (平均読み出し速度)

参考: 単体I/O (Read) 性能
SSD: 約2.79Gbps
SATA3・RAID5: 約5.48Gbps



SC13での実験の様子(読み出したフルHD画像を米国でリアルタイム表示)

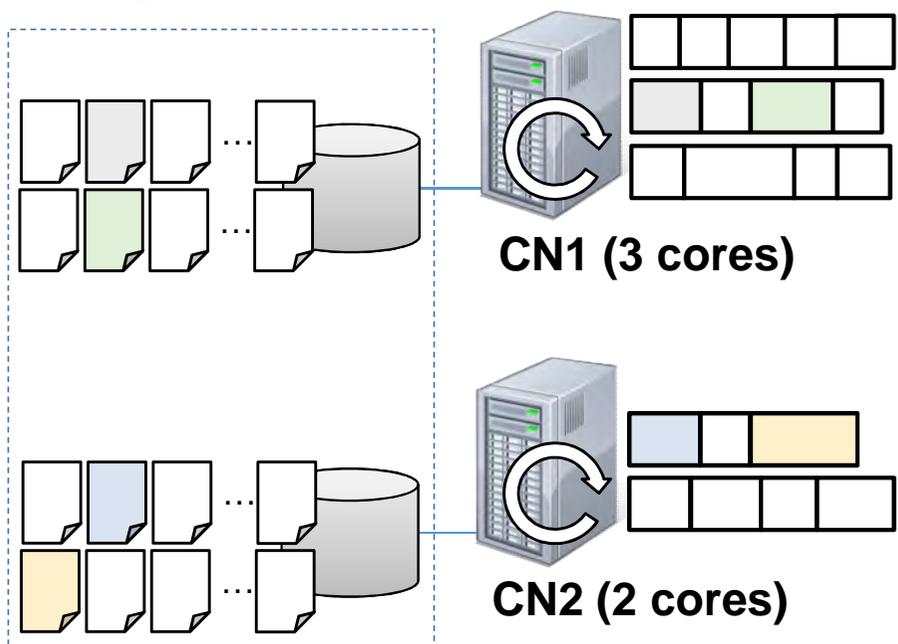
平成25年度の重点的研究開発課題とその成果

分散ストレージとタスクスケジューリングによる高スケーラビリティ科学ビッグデータ処理技術開発

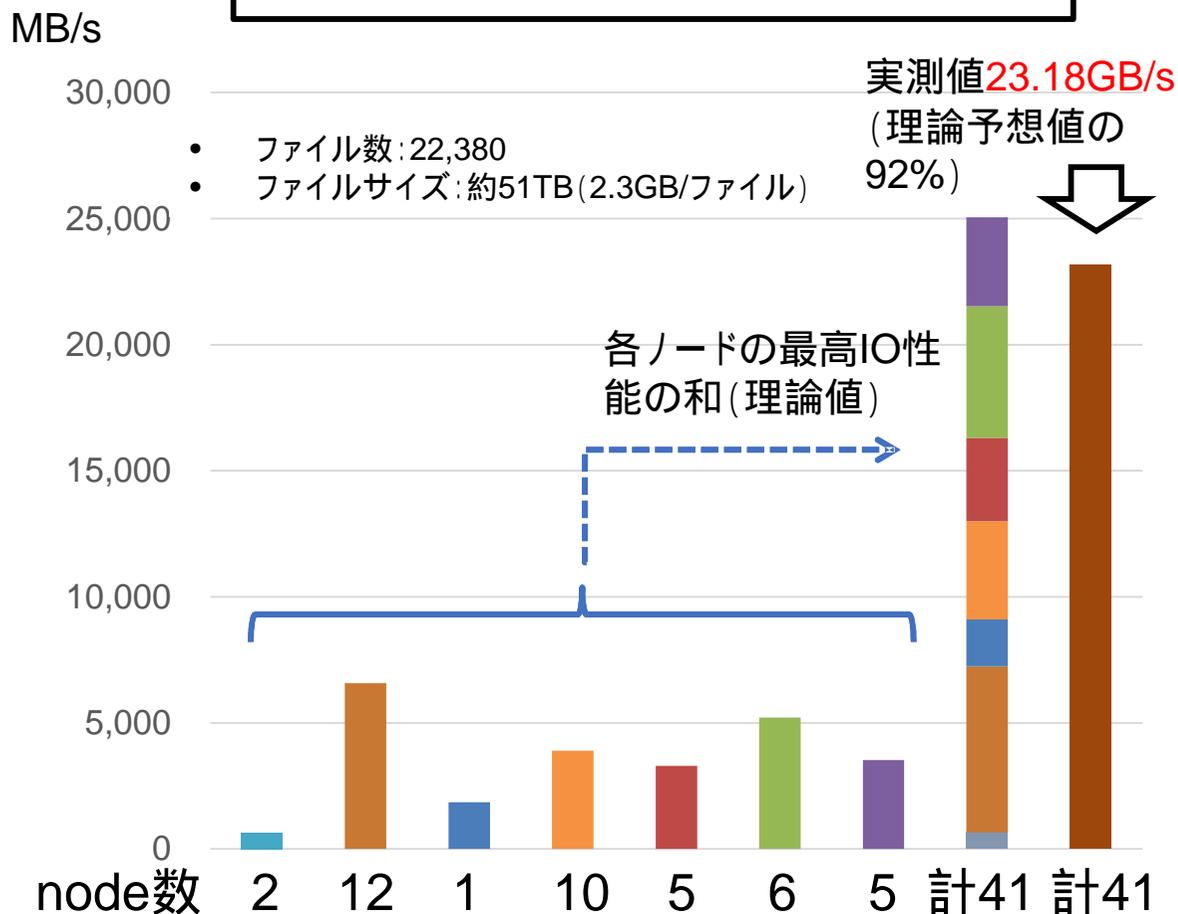
課題: ビッグデータ処理でボトルネックになるのは処理速度よりもI/O速度

分散ストレージと分散処理を組み合わせることで高速I/Oを実現

分散ストレージ

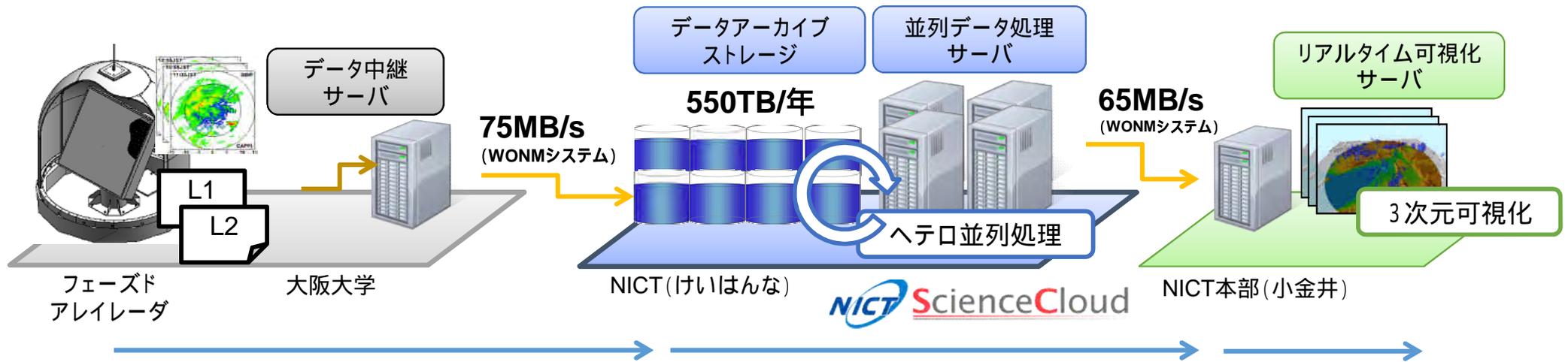


実験結果: NICTサイエンスクラウド(41ノード・348コア)によるI/O速度測定



平成25年度の重点的研究開発課題とその成果

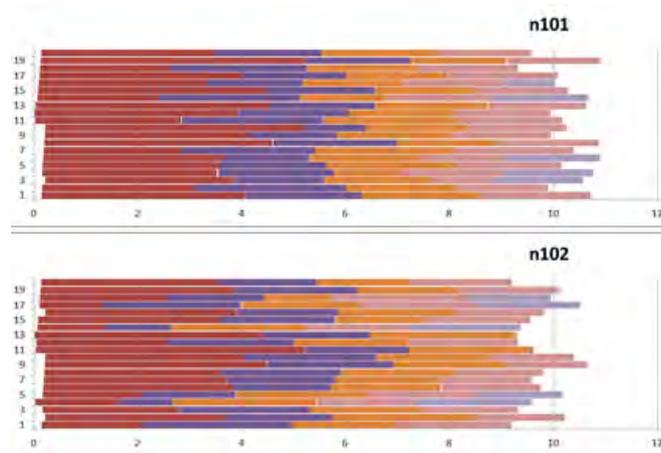
次世代ドップラーレーダのリアルタイム3次元可視化システム開発



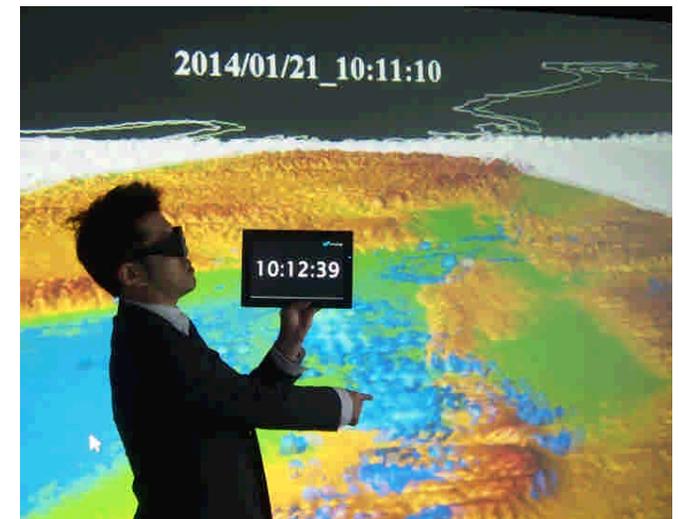
■3次元フェーズドアレイレーダデータを**1分半以内**に伝送・処理・可視化するシステム

■開発技術

- 広域観測網監視システムを開発、データ伝送・処理等のシステム監視
- 高度によりデータ処理量が異なる3次元格子データに対し、タスクスケジューラを活用したヘテロ並列処理技術により、180コアで60倍の高速化を達成



180コア(20コア×9ノード)によるリアルタイム並列処理(図は2ノード分):縦軸はコア、横軸は時間、色はタスク(格子点に対応)を示す。

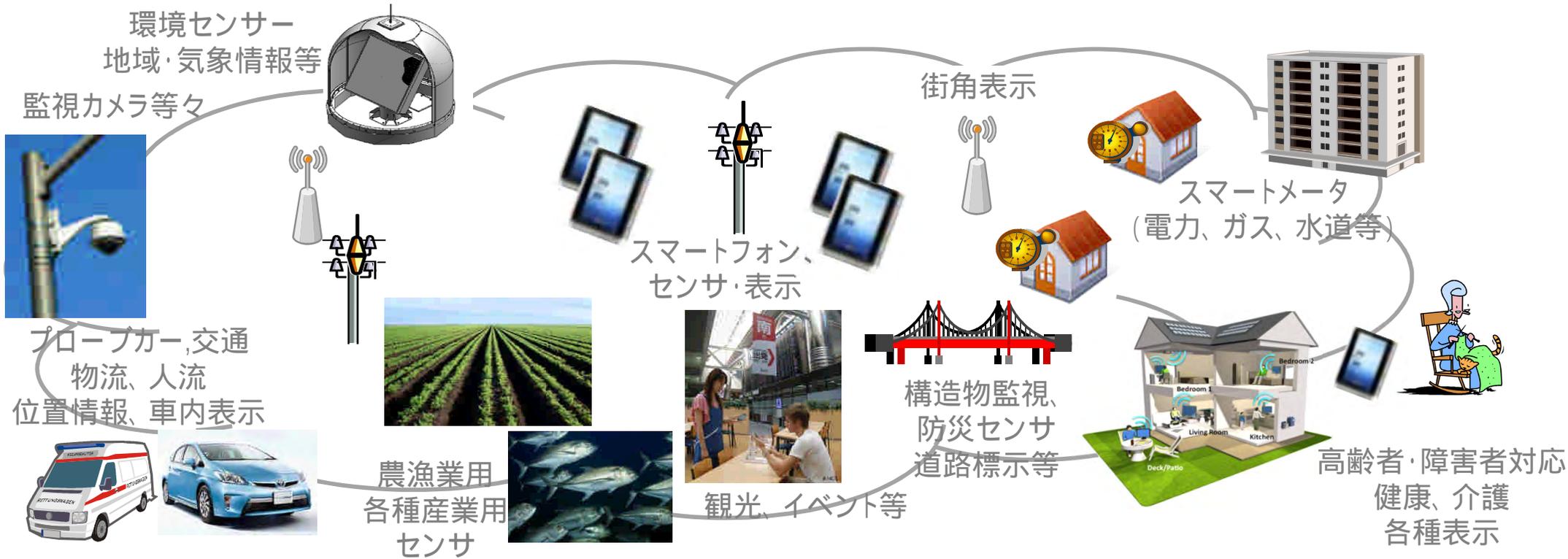


大阪平野の降雨状況ステレオ立体視の様子(1分29秒遅れで小金井本部でステレオ立体視)

まとめと今後の予定

- 本中期計画で、クラウドの安定運用、基盤技術開発を進め、これをもとにクラウドを活用した先端的科学研究推進を目指す。
 - H.25年度までに はおおむね達成、 は様々な基盤技術を構築した。
 - 今後は、 の融合(マッシュアップによる)システム構築を行い、科学分野の先端的成果を目指す。
- 将来的には、サイエンスクラウド基盤技術をソーシャルクラウドに応用することを検討中。

(参考) ソーシャルクラウドに発展するサイエンスクラウド開発技術



本中期計画技術のソーシャルクラウド技術への利用

広域観測網モニタ
リングシステム

広域分散ストレージ(DR対応・トレーサビ
リティー・タイムスタンプ)

TDW用高精細動
画像表示システム

- 遠隔要約筆記システム(聴覚障害者)
- ネット点字作成・印刷システム(視覚障害者支援)

次世代ドップラーレーダリア
ルタイム3D表示システム

センシング・観測ビッグデータ並列分散処理
技術(高速I/O利活用)

医療機関横断的電子カル
テ管理システム

時系列データ
表示Web技術

基盤技術

システム
開発