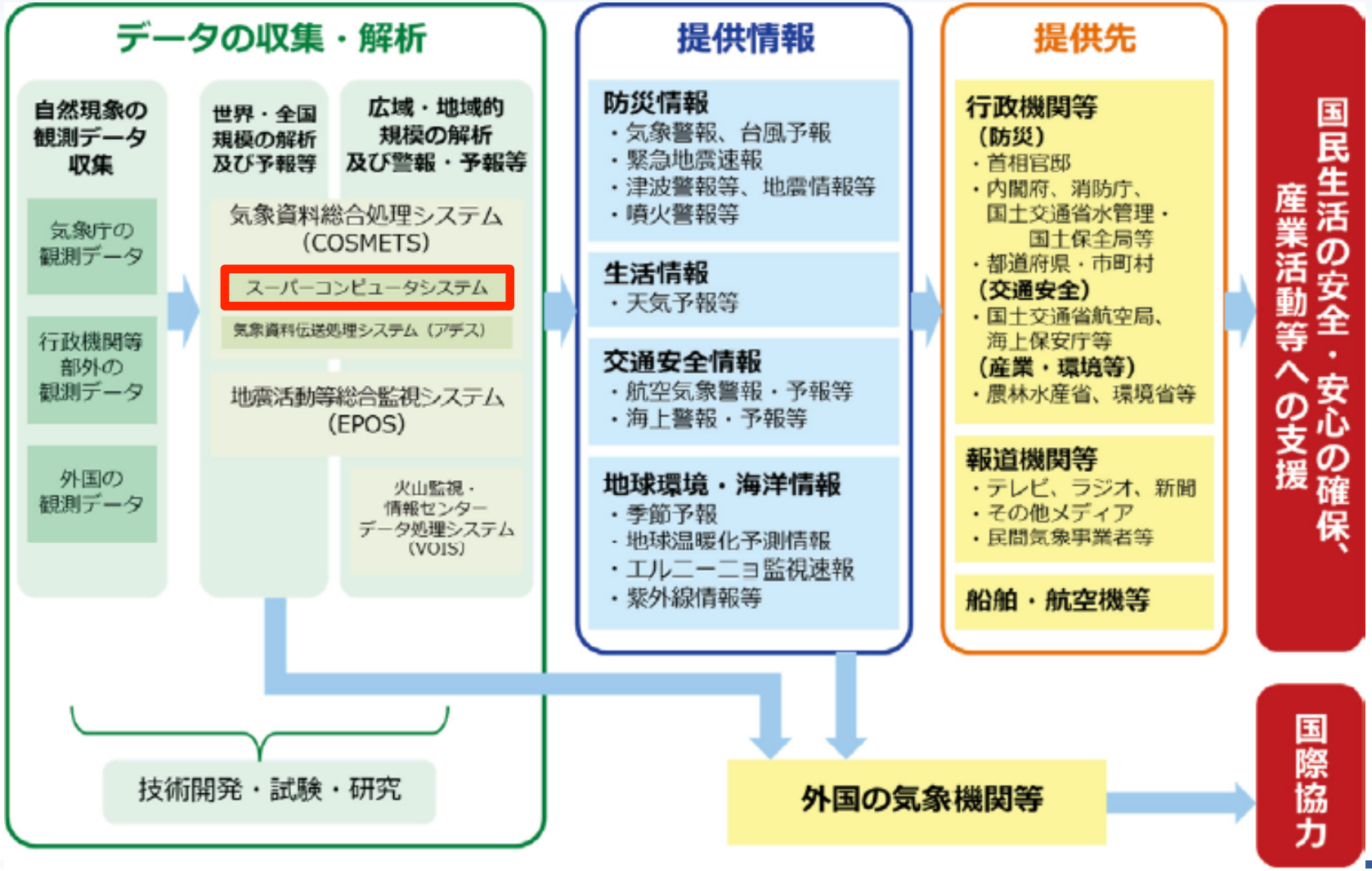
気象庁
スーパーコンピュータシステム

気象庁　予報部　情報通信課
牧　美照

2016.10.18
Japan Lustre User Group (JLUG) 2016

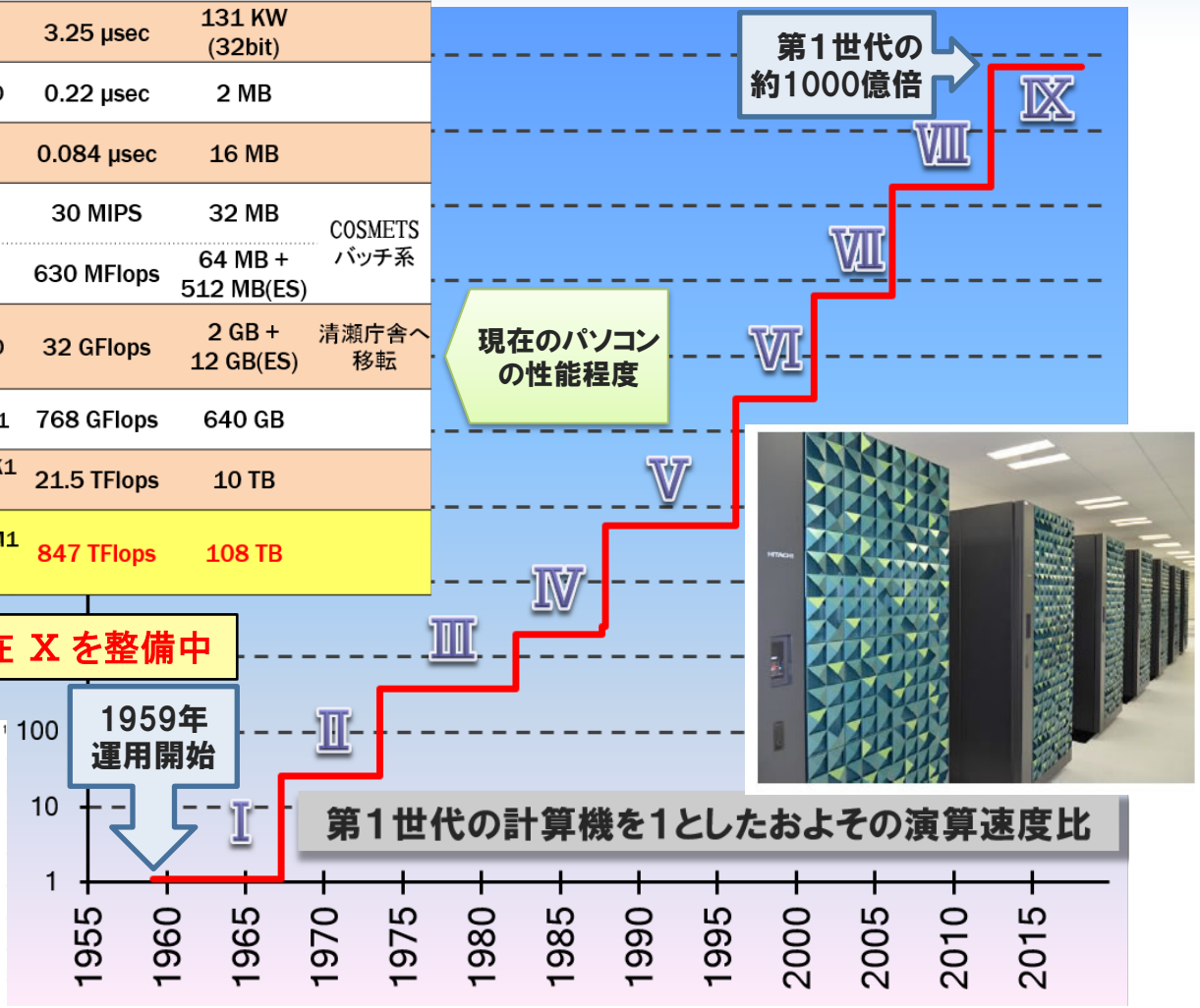
気象庁の業務


データの収集・解析
自然現象の観測データ収集
気象庁の観測データ
行政機関等部外の観測データ
外国の観測データ
世界・全国規模の解析及び予報等
広域・地域的規模の解析及び警報・予報等
気象資料総合処理システム（COSMETS）
スーパーコンピュータシステム
気象資料伝送処理システム（アデス）
地震活動等総合監視システム（EPOS）
火山監視・情報センターデータ処理システム（VOIS）
技術開発・試験・研究
提供情報
防災情報
・気象警報、台風予報
・緊急地震速報
・津波警報等、地震情報等
・噴火警報等
生活情報
・天気予報等
交通安全情報
・航空気象警報・予報等
・海上警報・予報等
地球環境・海洋情報
・季節予報
・地球温暖化予測情報
・エルニーニョ監視速報
・紫外線情報等
提供先
行政機関等
（防災）
・首相官邸
・内閣府、消防庁、国土交通省水管理・国土保全局等
・都道府県・市町村
（交通安全）
・国土交通省航空局、海上保安庁等
（産業・環境等）
・農林水産省、環境省等
報道機関等
・テレビ、ラジオ、新聞
・その他メディア
・民間気象事業者等
船舶・航空機等
国民生活の安全・安心の確保、産業活動等への支援
外国の気象機関等
国際協力

気象庁スパコンの変遷

世代	運用開始年月	主計算機	演算速度	記憶装置	備 考
I	1959/3	IBM-704	84 μsec	8 KW (36bit)	運用開始（本庁）
II	1967/4	HITAC-5020F	3.25 μsec	131 KW (32bit)	
III	1973/8	HITAC-8700/8800	0.22 μsec	2 MB	
IV	1982/3	HITAC-M200H（2台）	0.084 μsec	16 MB	
V	1987/9	HITAC-M680	30 MIPS	32 MB	COSMETS バッチ系
	1987/12	HITAC-S810	630 MFlops	64 MB + 512 MB(ES)	
VI	1996/3	HITAC-S3800_480	32 GFlops	2 GB + 12 GB(ES)	清瀬庁舎へ移転
VII	2001/3	HITACHI-SR8000E1	768 GFlops	640 GB	
VIII	2006/3	HITACHI-SR11000K1（2台）	21.5 TFlops	10 TB	
IX	2012/6	HITACHI-SR16000M1（2台）	847 TFlops	108 TB	

※平成30年6月に運用切替へ向け、現在 X を整備中



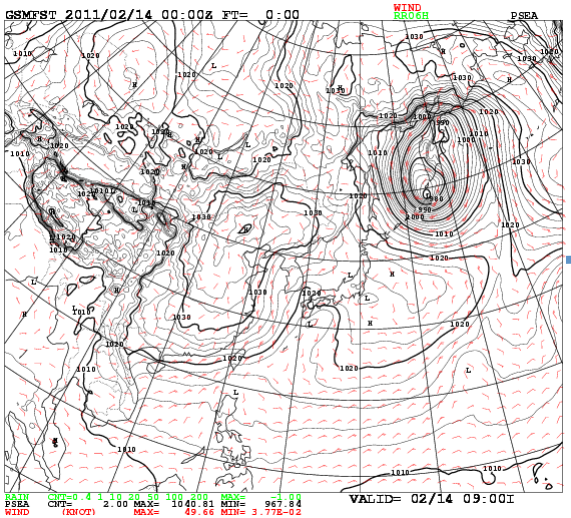
※上記、気象情報作成用のスーパーコンピュータの他
気象研究所（つくば市）にも研究用スーパーコンピュータを整備

数値予報とは

- 観測値を基に、ある特定の時刻の大気の状態を数値的に解析（初期値を作成）し
- 流体力学や熱力学などの物理法則に基づいて、その大気状態を時間発展させ、将来の大気の状態を予測すること

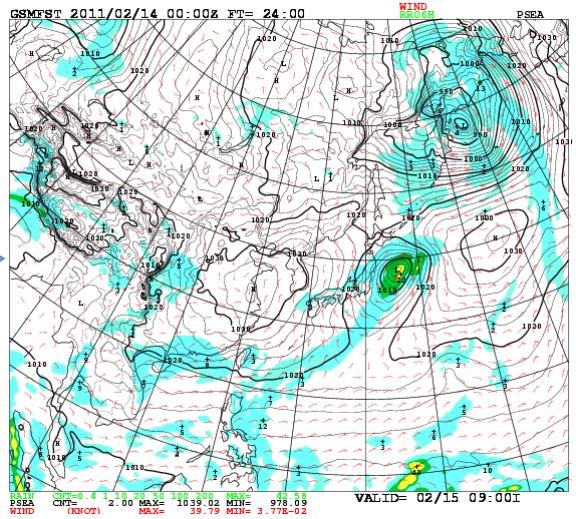
ある時刻の、大気の状態を表す要素（気圧,気温,水蒸気量,風向・風速など）の観測値

客観解析



ある特定の時刻の大気状態（初期値）

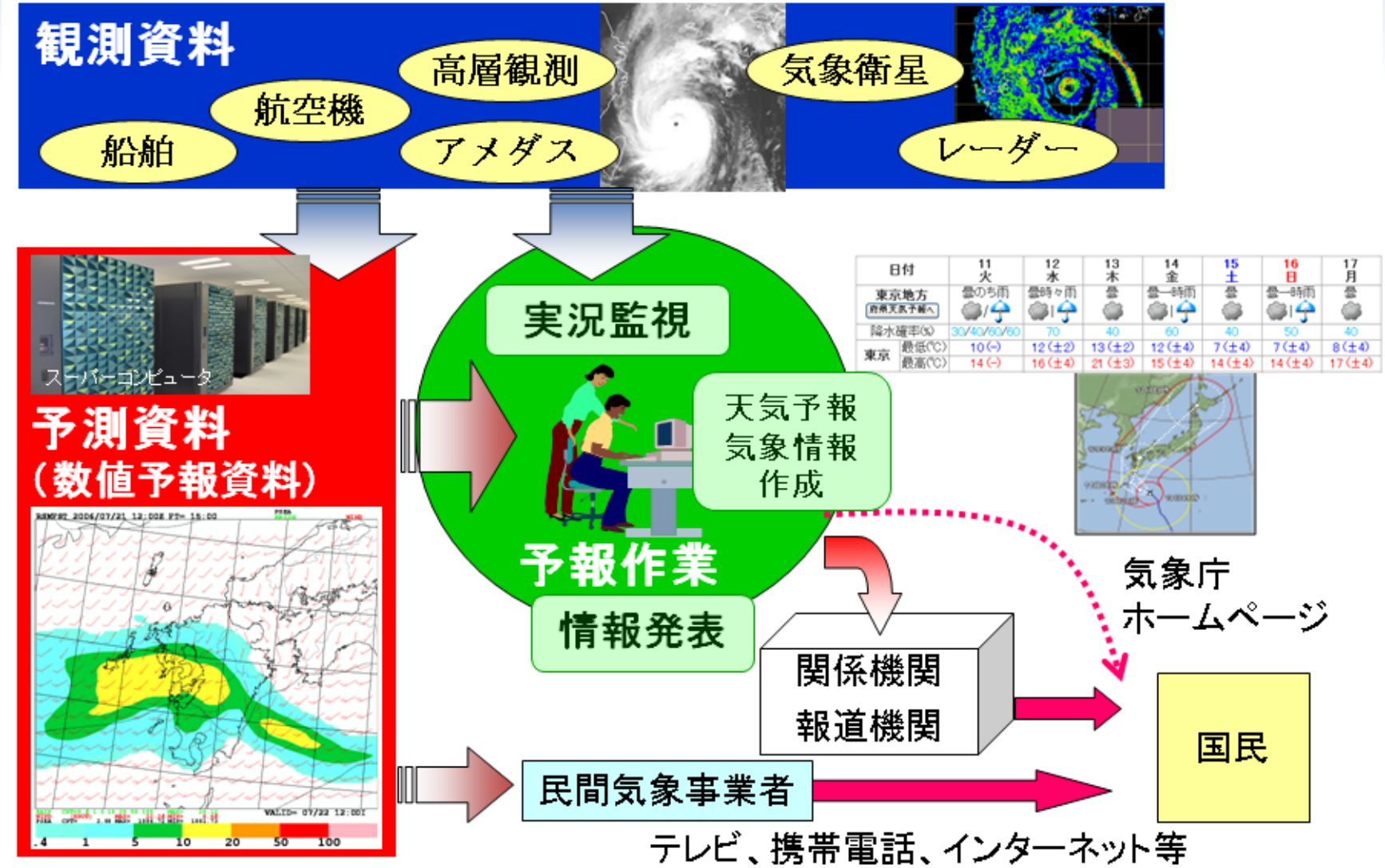
数値予報



将来の大気状態（予測値）

数値予報の役割(1)

気象業務における数値天気予報の位置づけ

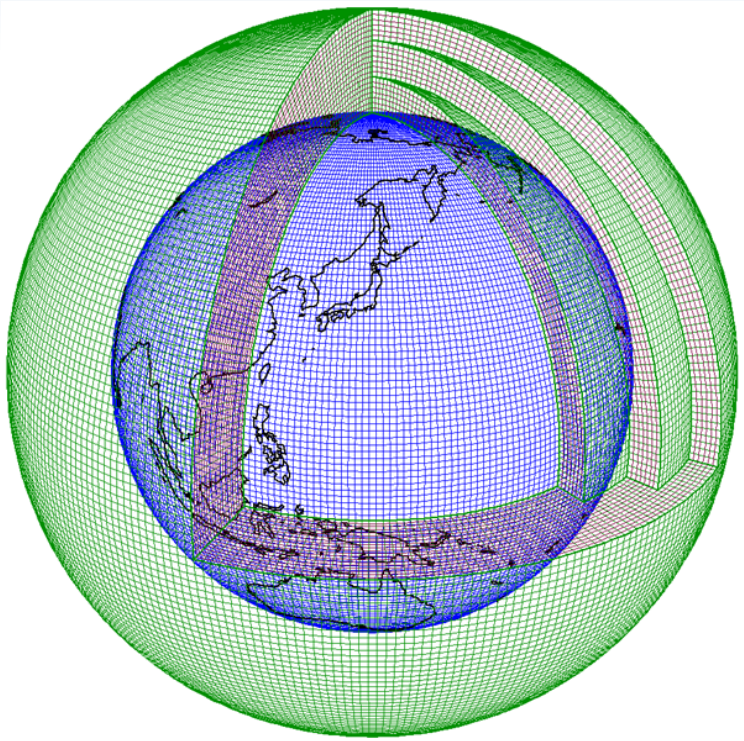


数値予報の役割(2)

数値予報の顧客

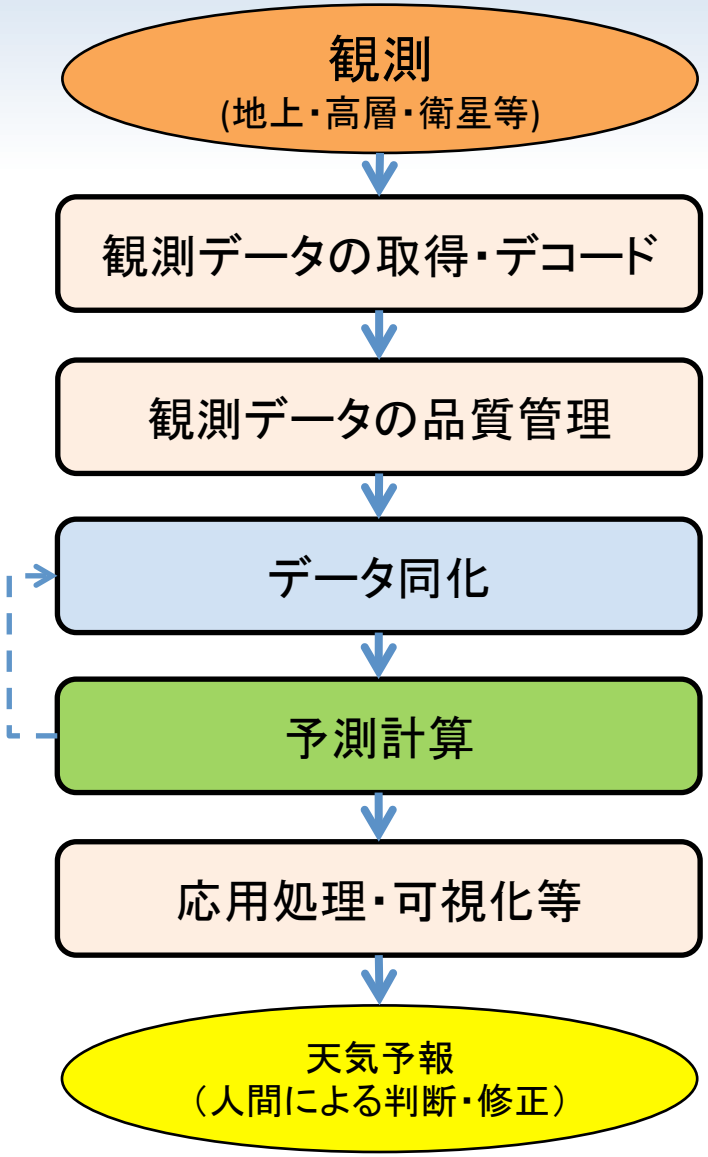
- 全国の予報現場
- 民間気象事業者
- 地方公共団体等の防災機関
（二次利用が中心）
- 国民（二次利用が中心）
- 海外の気象局
- 航空分野
（航空管制、民航、パイロット）
- 海上交通分野（外洋、沿岸）
- 防衛分野
（航空・海上・陸上自衛隊）
- 防災・水資源管理等
- エネルギーマネジメント（電力）
- 陸上交通・農業等

数値予報計算



・大気の空間を規則正しく並んだ格子で区切る (※格子数は全球モデルで約1億3千万個にも及ぶ)
・各格子は気圧、気温、風などの大気の状態を表す物理量を当てはめる（初期値作成）
・物理法則により、各物理量の変化率を求め、次の時刻の状態を求めていく（積分計算）
※原理的には、格子間隔が細かいほど大気の状態をより正確に表現可能

数値予報の流れ



世界中で観測されたデータを収集し、数値予報で利用可能な形式に変換する

観測データの品質を確認し、誤差の大きいデータの除去・補正を行う

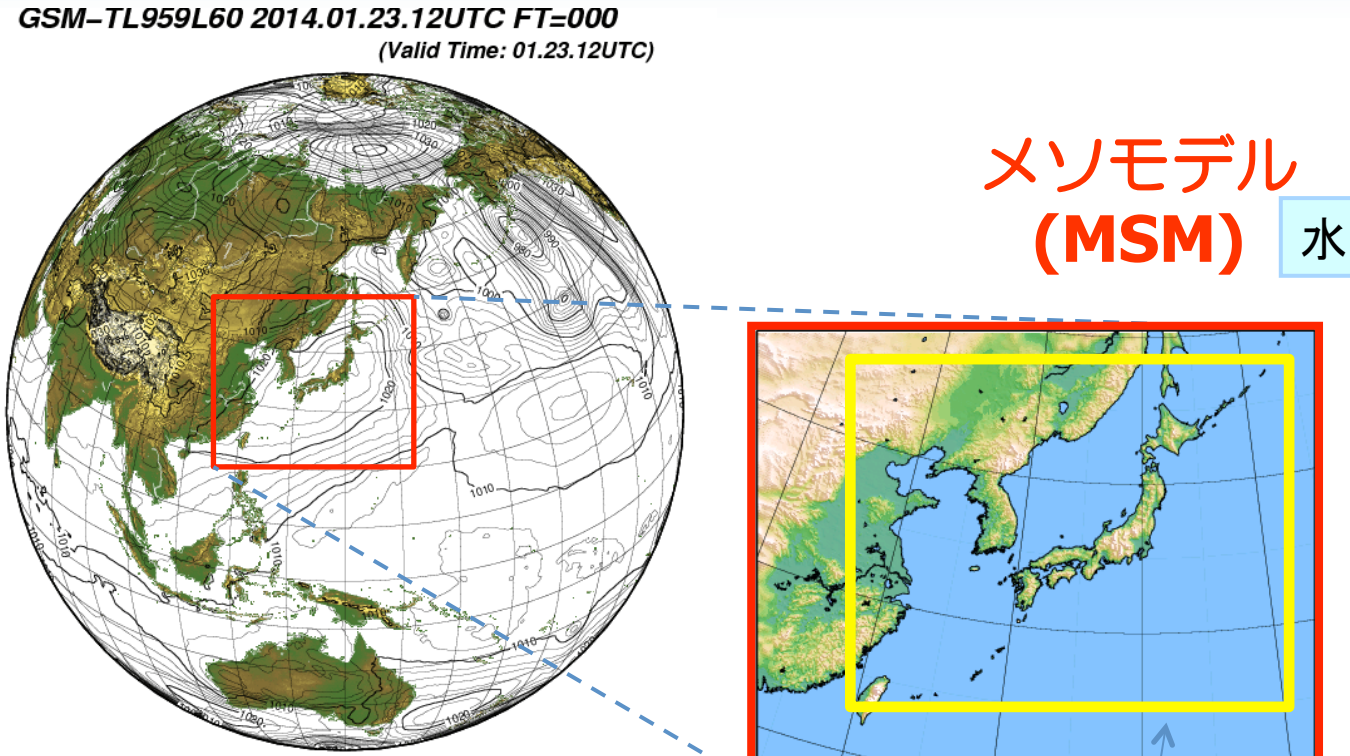
データ同化システムを用いて、観測データに基く精度の高い初期値を作成する

数値予報モデルを用いて、将来の大気状態を表す

各種物理量の予測値を算出する。
数値予報モデルによる予測結果の補正を行うとともに、予報作業に必要な情報への翻訳・可視化を行う

一連の処理を高速且つ確実に行うために
スーパーコンピュータが必要

気象庁の短期予報向け数値予報モデル



全球モデル
(GSM)

水平格子間隔 ~20km

メソモデル
(MSM)

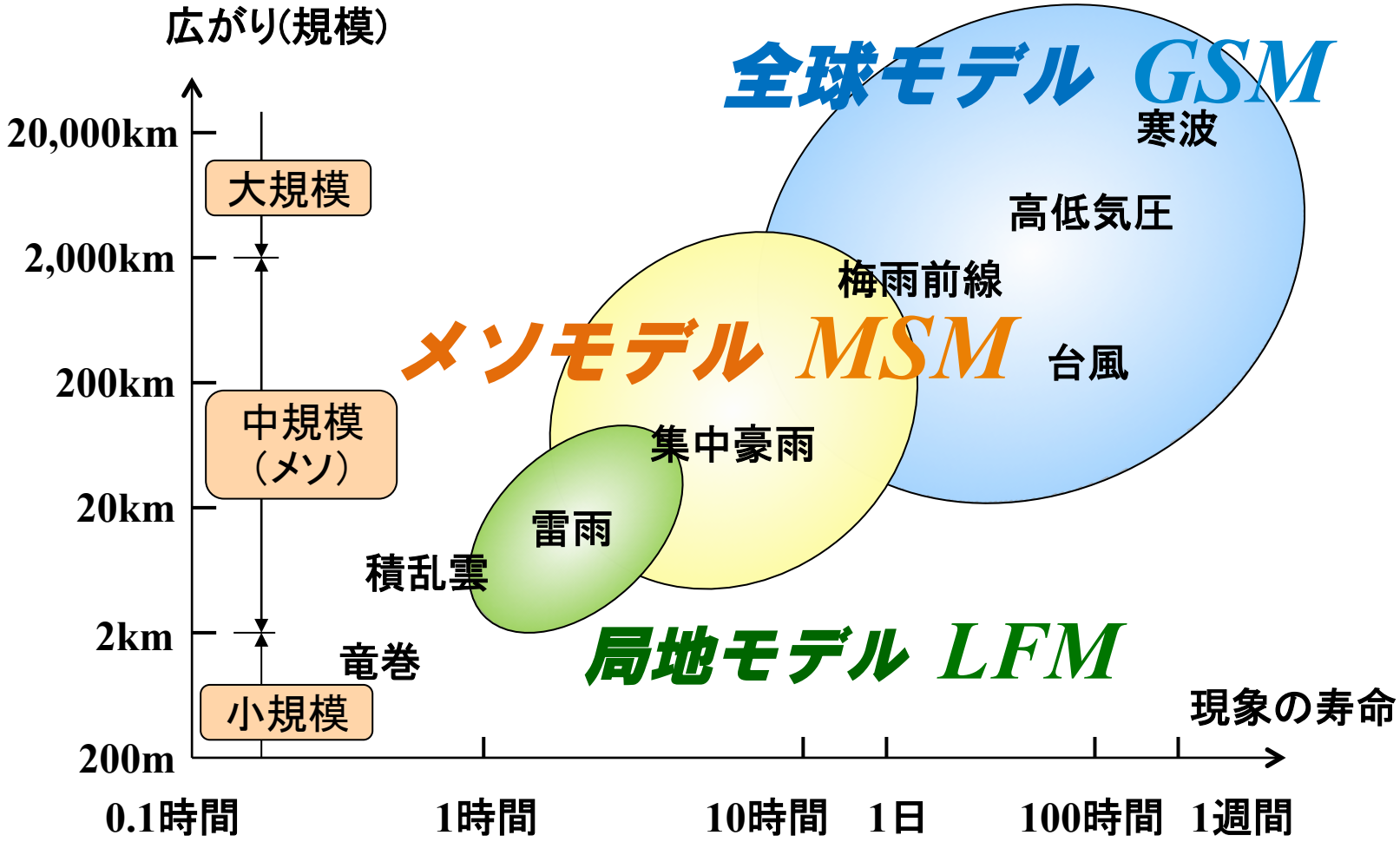
水平格子間隔 5km

局地モデル
(LFM)

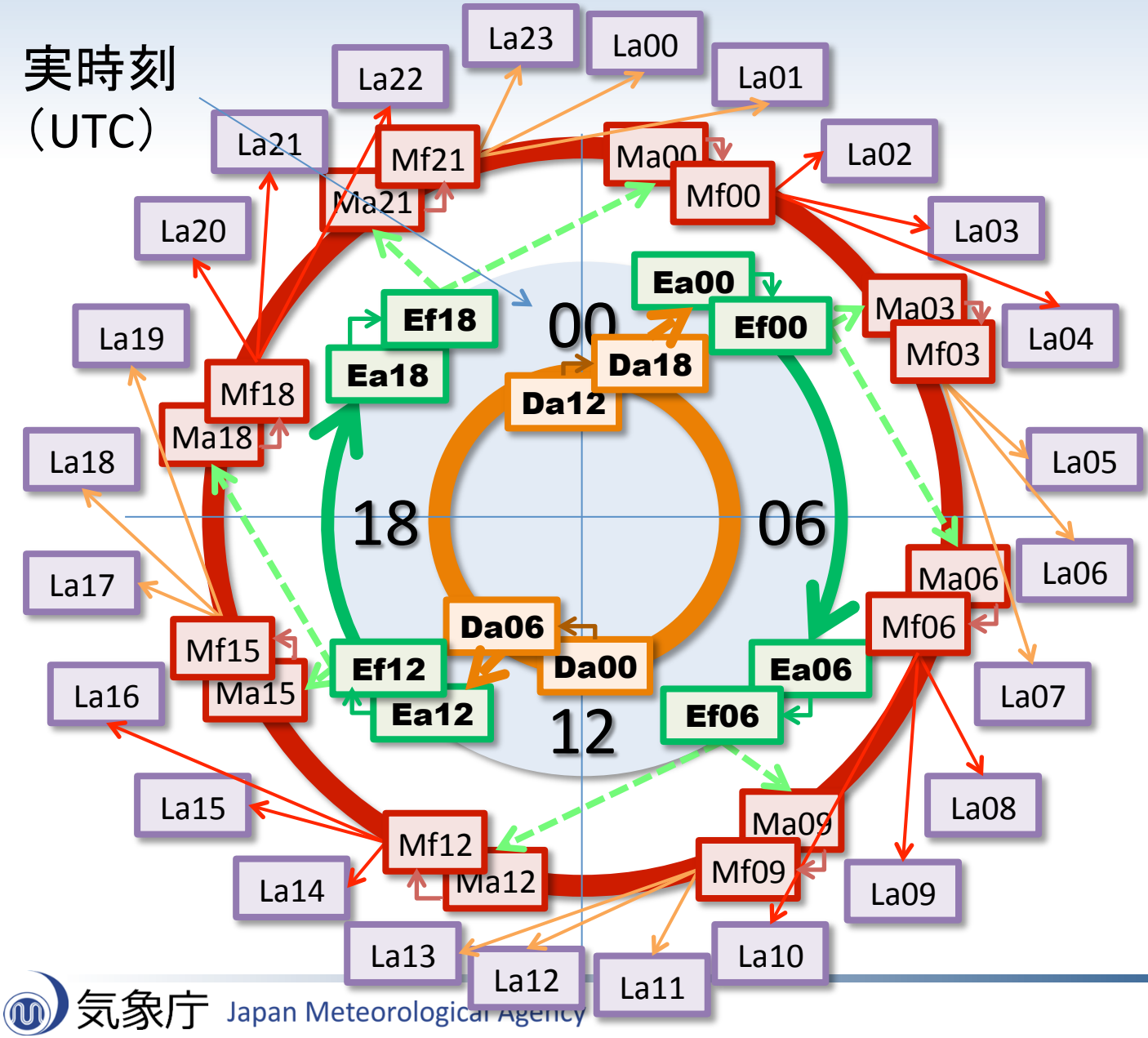
水平格子間隔 2km

現象と予測可能性

予測可能な時間の限界：予測対象や初期状態によって異なる



解析予報サイクルスケジュール



全球サイクル解析 (Da)

- 一番内側で自己完結したサイクルで精度維持のための解析
- 観測データの待ち時間を十分にとった後処理を実行

全球速報解析 (Ea)

- 全球予報のための解析
- Ea00,12はDa18,06解析値から第一推定値を作ることで、より高い精度の維持を図っている

全球予報 (Ef)

メソ解析 (Ma)

- メソ予報のための解析
- 全球解析からはほぼ独立しているものの、側面境界値を全球予報からもらっている

メソ予報(Mf)

局地解析 (La)

- 局地予報のための解析
- メソ解析・予測値を最初の第一推定値として使っている

※局地予報はスペースの関係上、記載を省略

数値予報ルーチン

- 数値予報関連処理を定常的に実行する
  - 数値解析予報サイクルが中心
  - 各処理には依存関係がある … 順序をまもる

- 天気予報等は、毎日・決まった時刻に
  - 予定通りに処理を開始、終了させねば…
  - 後続の処理に遅延が生じる
  - 情報の提供が遅れる

データの欠落など、さまざまな異常にも耐える

数値予報ルーチン実行タイムライン

→ 時間の流れ

データ取得
デコード処理

全球解析
予報処理

メソ解析
予報処理

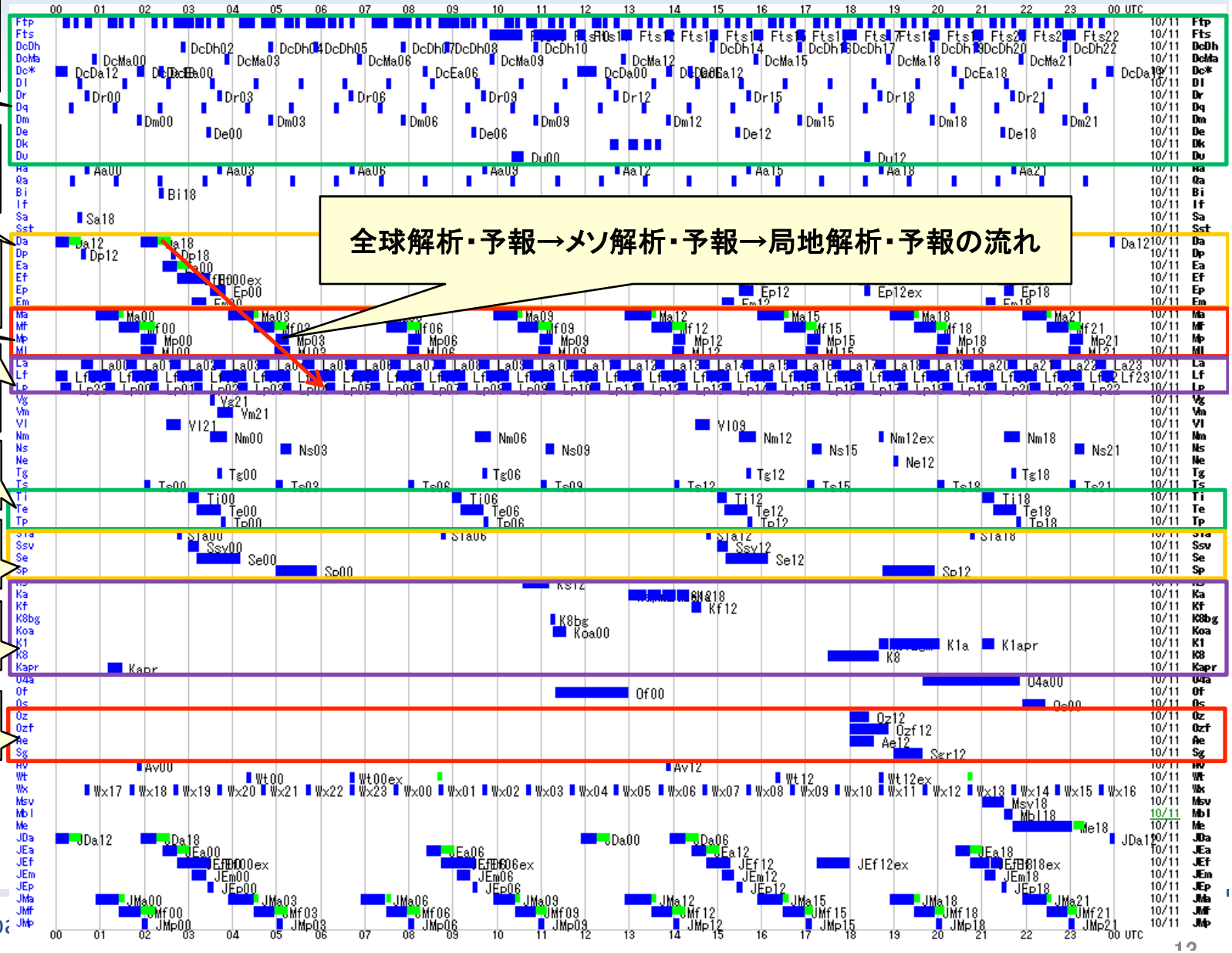
局地解析
予報処理

台風予報処理

週間予報処理

気候解析
予報処理

オゾン、黄砂
解析予測等

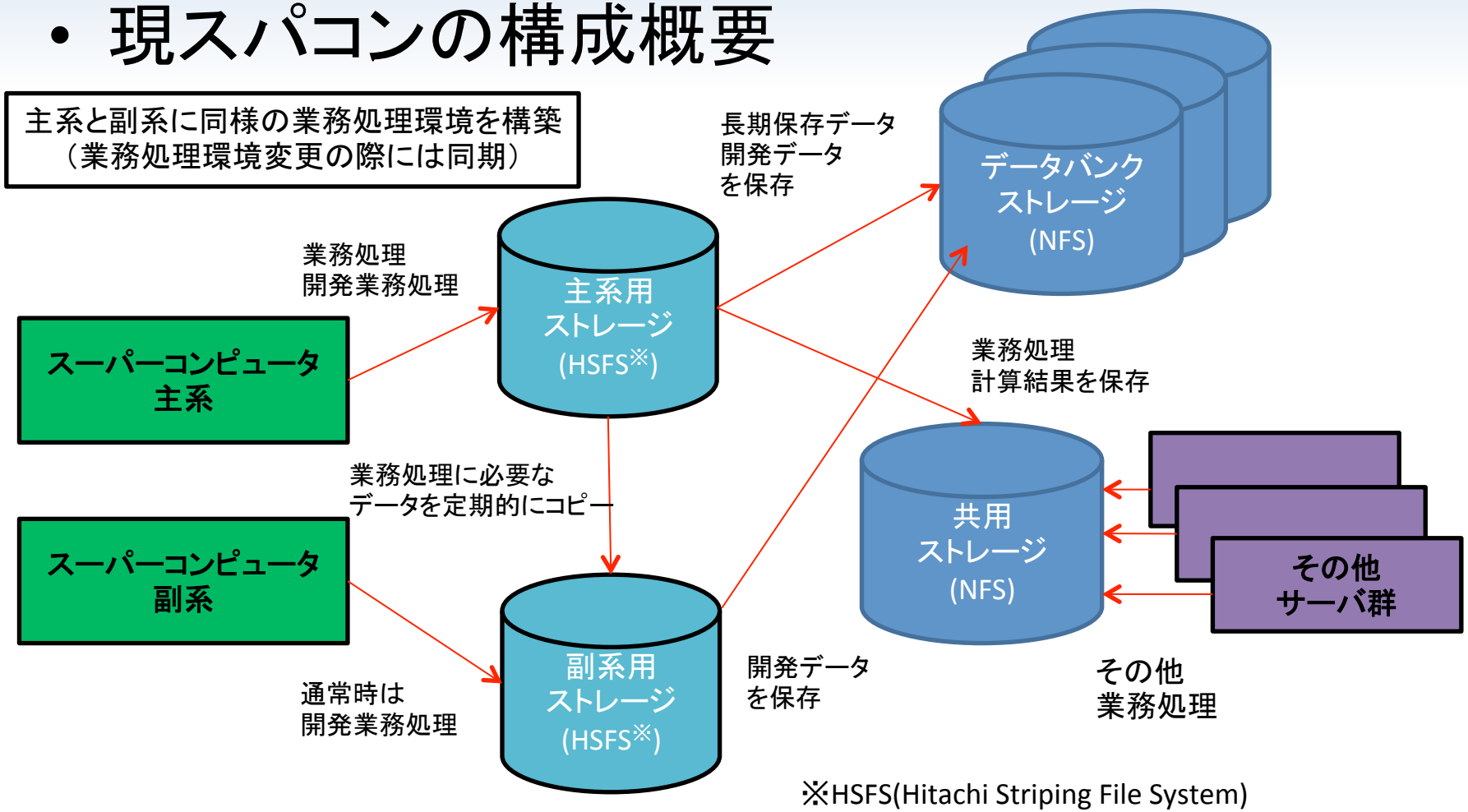


気象庁が求めるスパコン

- 高性能
  - （より高度な計算で）もっと予報を当てたい
- 可用性・安定性
  - 24時間通年の安定運用が必要
- 堅牢性
  - 定常業務と同時に、常に開発（改善）作業も

スパコンの可用性確保

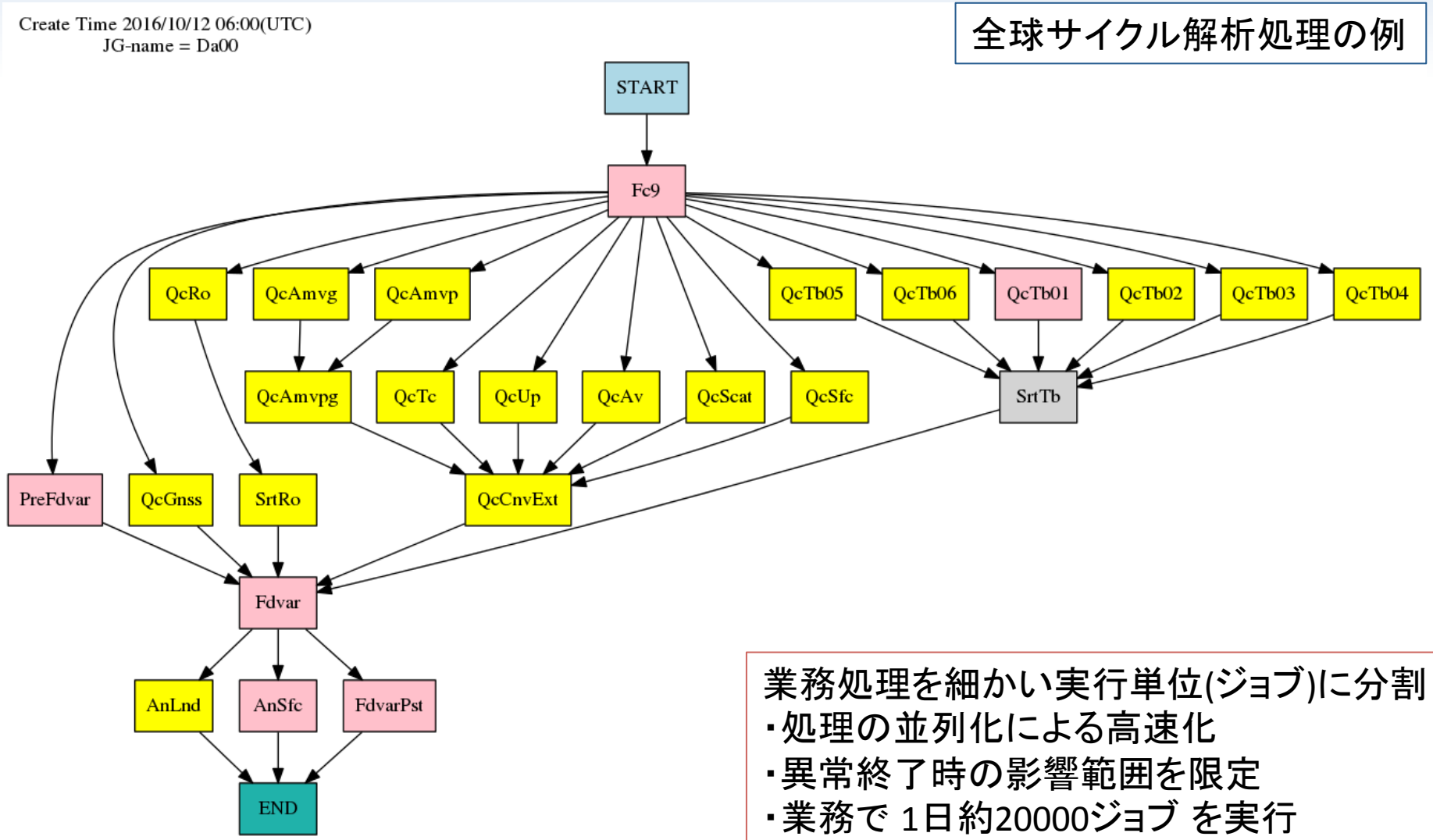
- 現スパコンの構成概要



異常時・メンテナンス時には副系で業務継続可能

気象庁業務処理の特徴

全球サイクル解析処理の例



業務処理を細かい実行単位(ジョブ)に分割
・処理の並列化による高速化
・異常終了時の影響範囲を限定
・業務で 1日約20000ジョブ を実行

気象庁がスパコン業務で求めるファイルシステム(1)

- 一意性
  - ジョブ終了後、その出力を入力とする後続ジョブを速やかに開始できる必要がある。
    →ジョブ間では、ファイルでデータ受け渡しを行うため、
    どのノードから見てもファイルシステムが同じ状態に見えることが重要。

- 安定性
  - 決められた時間内に処理を終了させる必要がある。
    →均一な性能が重要(**一時的な性能低下は困る**)
  - 処理時間を短縮するため、リスタートファイルは極力使わない。
    →異常終了したら最初から計算する
    異常を検知した場合、早目にジョブを異常終了させ、再実行させることが重要。

気象庁がスパコン業務で求めるファイルシステム(2)

- 高速性
  - 可用性のため、ジョブを投入するノードをあらかじめ予約することはしない（ノード確保自体は行う）。
    →ジョブ実行前に実行するノードを特定してローカルの高速ストレージにデータを準備する**（ステージング）等は行わない**。
    ※先行ジョブからの**入力データがファイルで渡される**（実行直前に入力ファイルが作成される）ことからもステージングはできない

  - 精度の良い数値予報には観測データが重要
    世界中からの観測データの到達(入電)は遅い
    このため、**数値予報計算に費やせる時間が制限される**。
    →実は数値予報モデルにおいて、**計算時間と比べてもI/Oの時間は無視出来ない**

計算時間に占める I/O時間の割合の例（参考）（1）

- 全球解析・予報の試験例

処理内容	実行ノード数	入出力処理	通信処理	計算処理
全球解析	128	8.13%	33.85%	58.03%
全球予報 (132h)	324	19.27%	38.39%	42.34%

※次期スパコン調達に向け、次期数値予報モデルを想定した
数値予報実験を行った際のプロファイル結果。
定常業務で動作させている処理とはモデルの解像度、実行ノード数等は異なる。

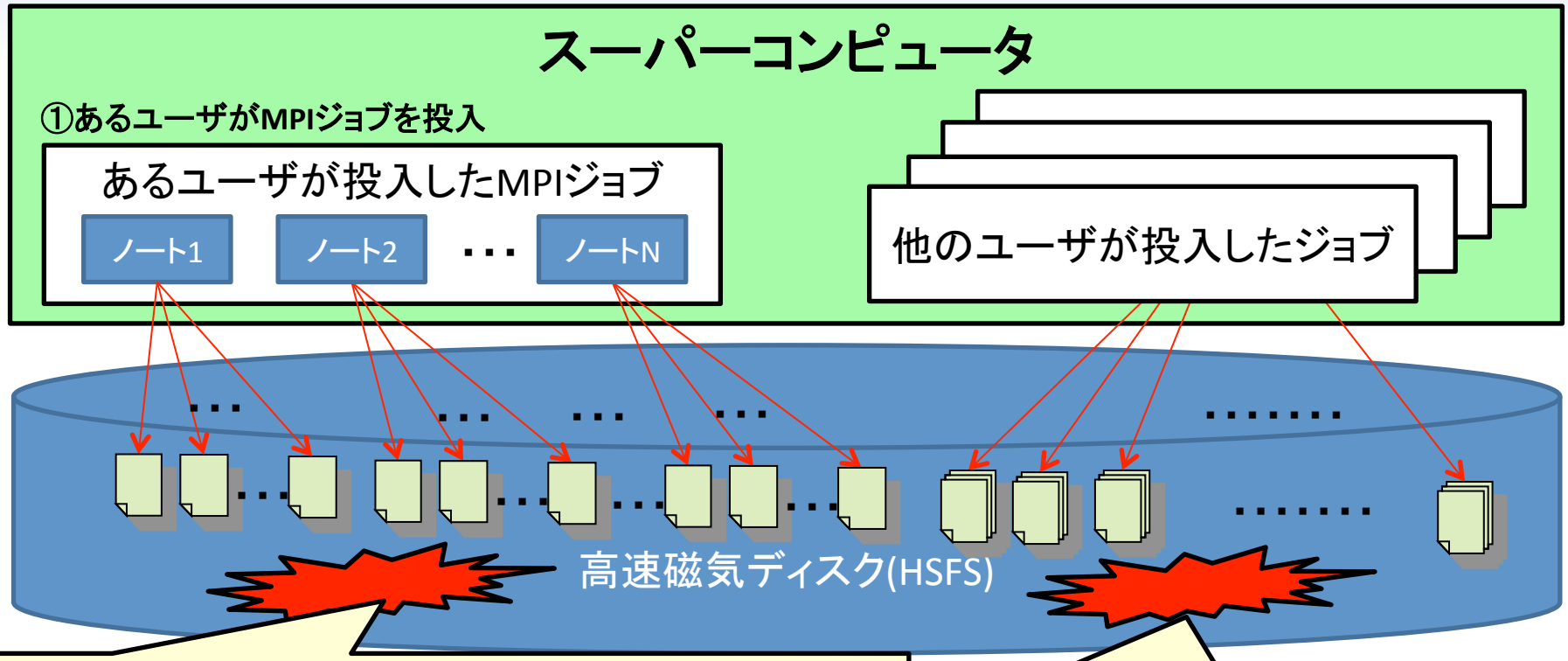
計算時間に占める I/O時間の割合の例（参考）（2）

- 局地解析の試験例

スパコン	ファイルシステム	実行ノード数	入力時間	全体の処理時間	備考
NAPS9	HSFS	85	93.8秒	329秒	初期値の入力時間が全体の約28.5%を占める
気象研究所	FEFS	144	65.4秒	902秒	参考

※気象研究所での実行結果は、NAPS9に最適化したものをそのまま使用したもので、気象研究所のシステムに最適化していないため、全体の処理時間はあくまで参考

現行ファイルシステムで発生した問題点（1）



②多数プロセス（報告では、832プロセス）が同時に個々のログファイルに書き込みを行ったことにより、メタデータディスクが高負荷となり、入出力性能が低下してジョブが遅延

③スパコンで実行している他の実行中ジョブの実行も同様に入出力性能が低下し遅延が発生

I/O高負荷の原因を特定できるツールが欲しい

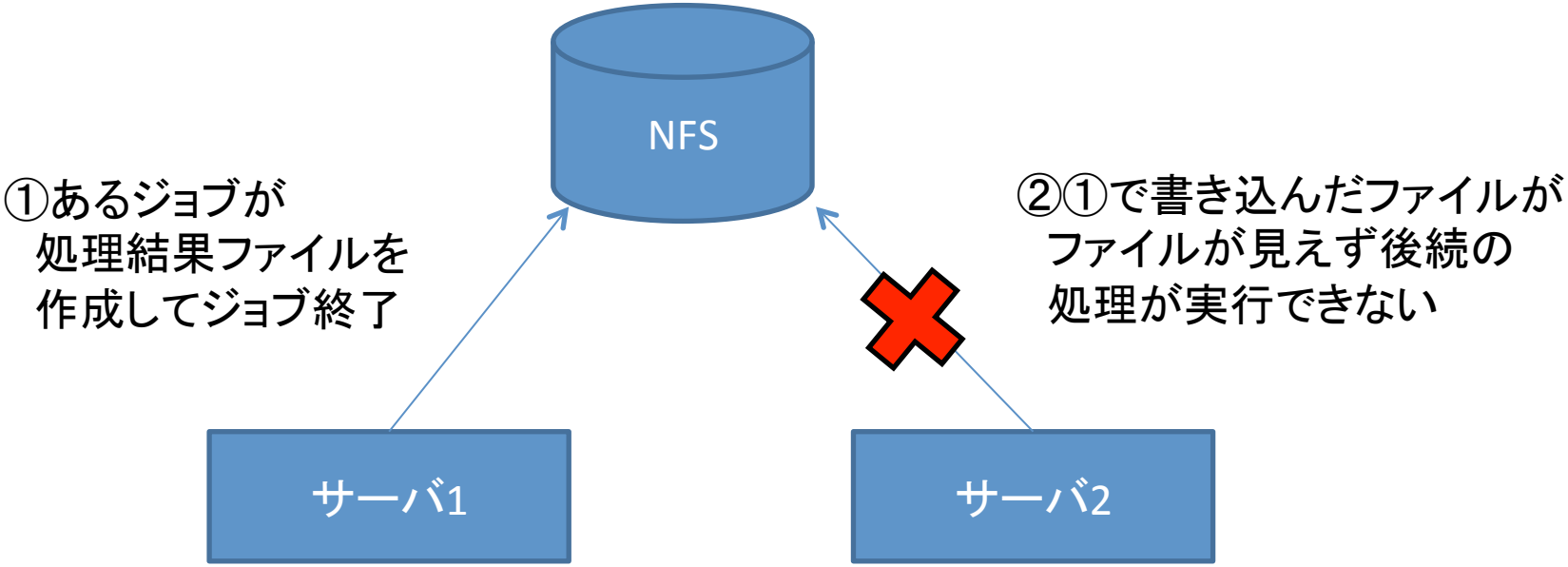
現行ファイルシステムで発生した問題点（2）

- CSES(Cluster Shared Extended Storage)
  - 複数ノードの主記憶を仮想的に連結した高速ストレージ（CSESデバイス）を提供する機能。クラスタ内の各ノードから共用アクセス可能
  - CSESデバイス上にESファイルシステムを作成、専用APIにより、ESファイルシステムにアクセス。

- CSESの問題点
  - ES作成、削除のオーバーヘッドが大きい
  - 後続のジョブが先行ジョブの出力データに高速にアクセスするには、先行ジョブと同じノードに投入する必要がある
    →可用性が失われる

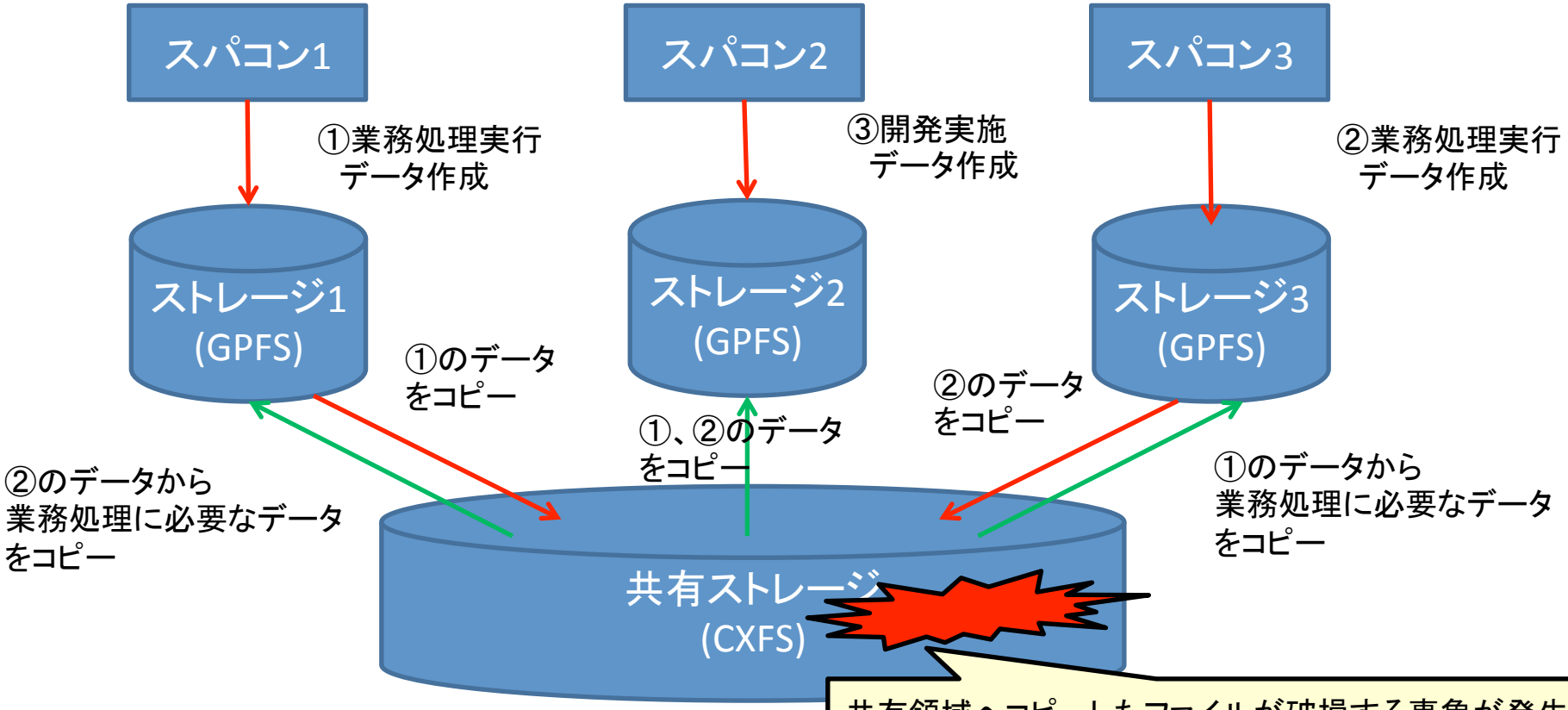
現行ファイルシステムで発生した問題点(3)

- ファイルの一意性の問題



過去に使用したファイルシステムとその問題点

- 共有領域にコピーしたファイルが破損



共有領域へコピーしたファイルが破損する事象が発生（コピー処理自体は成功しており、読み込んで初めて事象が発覚する）。
※異常発生時には検知できる仕組みが必要

最後に

- 次期スーパーコンピュータの高速ストレージにはLustreが採用される予定
- 可用性確保の仕組みだけでなく、様々な機能・性能面の開発も積極的に進められており数値予報処理発展への寄与を大いに期待
- 新機能にも期待しているが、気象庁としては安定性が最重要
- 大規模共有ファイルシステムは気象庁業務の根幹を成すスーパーコンピュータシステムに必要不可欠
- 今後も活発に開発を継続していただくことが気象業務への貢献にもつながっている