Gfarmワークショップ2019
(2/1 10:45～11:25)

「共用ストレージ東拠点ポータル」の整備状況と今後の拡充方針

2019/2/1

中　誠一郎（東京大学）

発表趣旨

- 「共用ストレージ東拠点ポータル」に至るニーズとシーズ
  - ニーズ（用途）
  - シーズ（設置経緯）
- 「共用ストレージ東拠点ポータル」の構成
  - 位置付け
  - データハブの実装方式
  - データハブのサービス内容
- 「共用ストレージ東拠点ポータル」の拡充方針
  - 今後の拡充方針（課題）
  - 利用者持ち込みサーバの扱い

「共用ストレージ東拠点ポータル」に至るニーズとシーズ

- ニーズ（用途）
- シーズ（設置経緯）

ニーズ（用途）

共用ストレージの新たなニーズ

- 今までは、専らアーカイブ用途
- 共用ストレージの利用活性化の芽
  - 「京」のサービス停止に伴い、新たなニーズが浮上
  - HPCIの産業利活用の促進に伴う、新たなニーズの発掘

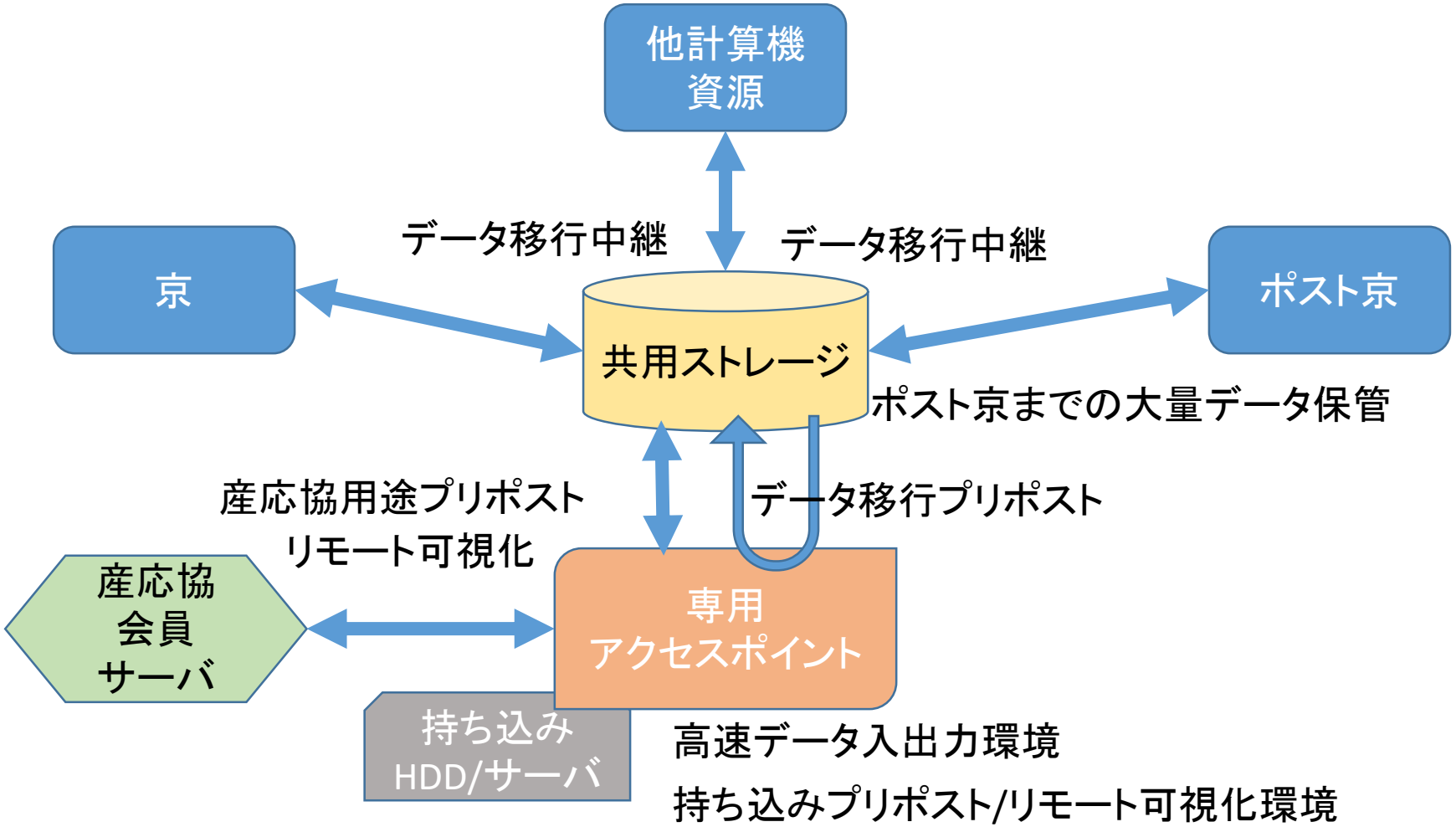
「京」のサービス停止に伴う新たなニーズ

- ポスト「京」運用開始までの保管用途
  - 共用ストレージの提供容量増は達成済み（21PB→90PB）
- 計算資源間のデータ移行用途
  - 2019年度から早々に開始
  - 保管スペースが当然必要だが、単なる保管機能だけでは不十分
  - 異なる計算資源へのデータ移行に伴うプリポスト処理が必要
  - クラウド的な共用ストレージ上のプリポスト処理実行環境を提供できれば理想的だが、、、
  - 共用ストレージに隣接したプリポスト処理実行環境が現実的

産業利活用の促進に伴う新たなニーズ

- 2016年度からHPCI全体で産業利活用の検討開始
  - 産応協の提言活動チーム発足
- 産業界は情報セキュリティ、データハンドリングを懸念
  - 情報の社外持ち出し時のセキュリティ対応
  - データのダウンロード速度
- 現状、アクセスポイント神戸・東京が実質的な接点
  - アクセスポイント利用時のHDDハンドキャリーの煩雑さ
  - 社内持ち込み時のセキュリティ対応
- 共用ストレージ自体の産業利活用の可能性
  - 共用ストレージ活用の観点が欠如
  - 共用ストレージ直結の専用アクセスポイントの提供

ニーズの整理


他計算機
資源
データ移行中継
データ移行中継
京
ポスト京
共用ストレージ
ポスト京までの大量データ保管
産応協用途プリポスト
リモート可視化
データ移行プリポスト
産応協
会員
サーバ
専用
アクセスポイント
持ち込み
HDD/サーバ
高速データ入出力環境
持ち込みプリポスト/リモート可視化環境

ニーズの集約⇒専用アクセスポイント

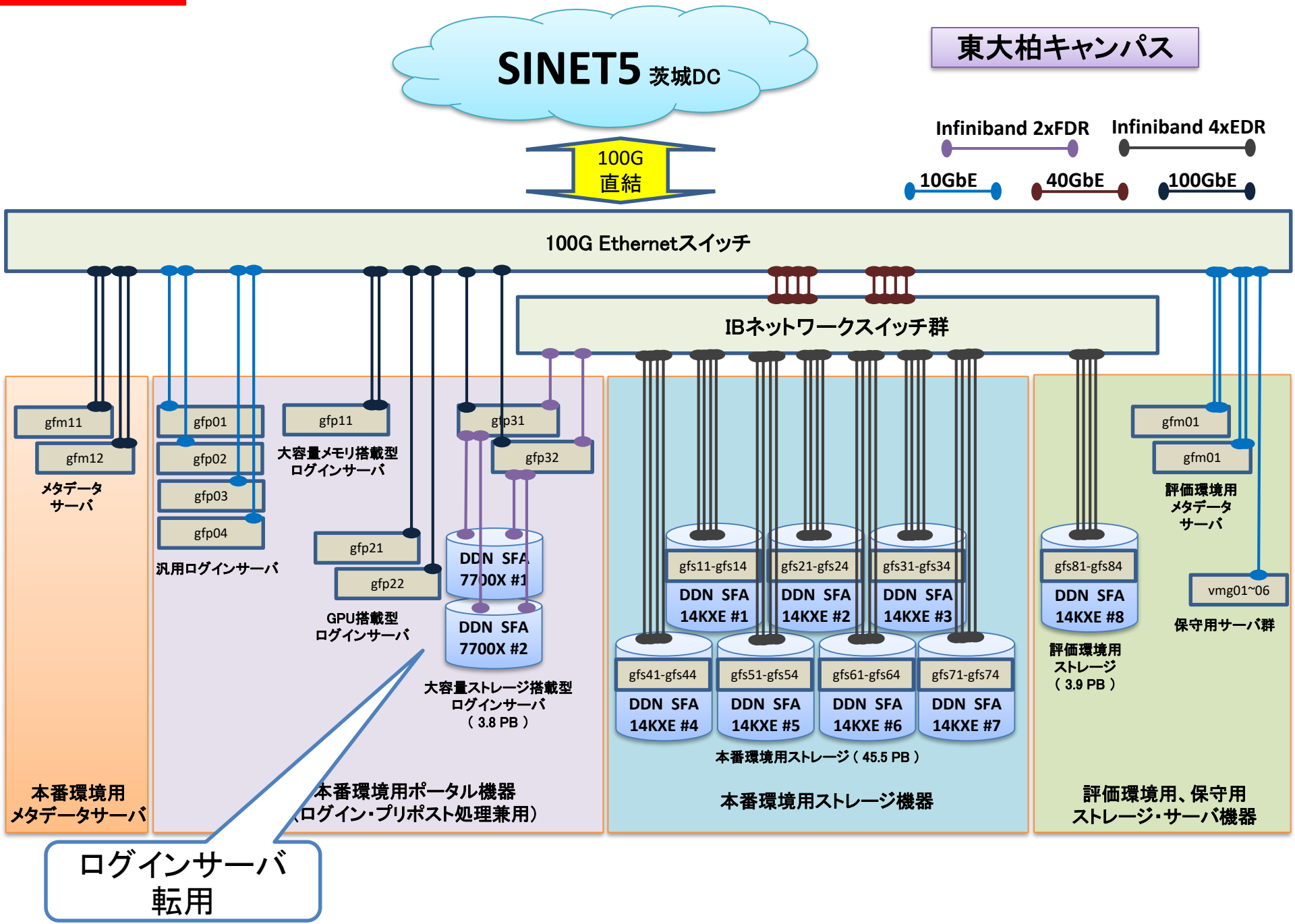
- プリポスト処理環境の提供
  - 計算資源間のデータ移行促進のためのプリポスト処理
  - 産応協会員用途向けのプリポスト処理
  - データハンドリング効率化のためのリモート可視化環境
  - 持ち込みサーバによるプリポスト処理・リモート可視化環境
- データ入出力環境の提供
  - 持ち込みHDD/サーバへのデータ入出力環境

シーズ（設置経緯）

データ保存用ストレージの転用

- 当初の想定用途（2016年度）
  - SFA7700X 3.8PB
  - データ保存課題の専用格納領域の提供
  - 万一、共用ストレージが第一期のみで終了した場合の保険
- 共用ストレージ第二期提供決定後の用途（～2017年度）
  - データ保存課題の専用格納領域の提供は不要
  - 第二期機材更新時のデータ中継用途として大活躍
- 共用ストレージ第二期運用開始後の用途（2018年度～）
  - 更新ストレージ機材（SFA14KXE)とのスペック差が大きいため、本番環境での混在利用を断念
  - 大容量ストレージ搭載型ログインノードへの転用により、データハブとして利活用

2018/12/21


SINET5 茨城DC
東大柏キャンパス
100G
直結
Infiniband 2xFDR
Infiniband 4xEDR
10GbE
40GbE
100GbE
100G Ethernetスイッチ
IBネットワークスイッチ群
gfm11
gfm12
メタデータ
サーバ
gfp01
gfp02
gfp03
gfp04
汎用ログインサーバ
gfp11
大容量メモリ搭載型
ログインサーバ
gfp21
gfp22
GPU搭載型
ログインサーバ
gfp31
gfp32
DDN SFA
7700X #1
DDN SFA
7700X #2
大容量ストレージ搭載型
ログインサーバ
(3.8 PB)
gfs11-gfs14
DDN SFA
14KXE #1
gfs21-gfs24
DDN SFA
14KXE #2
gfs31-gfs34
DDN SFA
14KXE #3
gfs41-gfs44
DDN SFA
14KXE #4
gfs51-gfs54
DDN SFA
14KXE #5
gfs61-gfs64
DDN SFA
14KXE #6
gfs71-gfs74
DDN SFA
14KXE #7
本番環境用ストレージ (45.5 PB)
gfs81-gfs84
DDN SFA
14KXE #8
評価環境用
ストレージ
(3.9 PB)
gfm01
gfm01
評価環境用
メタデータ
サーバ
vmg01~06
保守用サーバ群
本番環境用
メタデータサーバ
本番環境用ポータル機器
(ログイン・プリポスト処理兼用)
本番環境用ストレージ機器
評価環境用、保守用
ストレージ・サーバ機器
ログインサーバ
転用

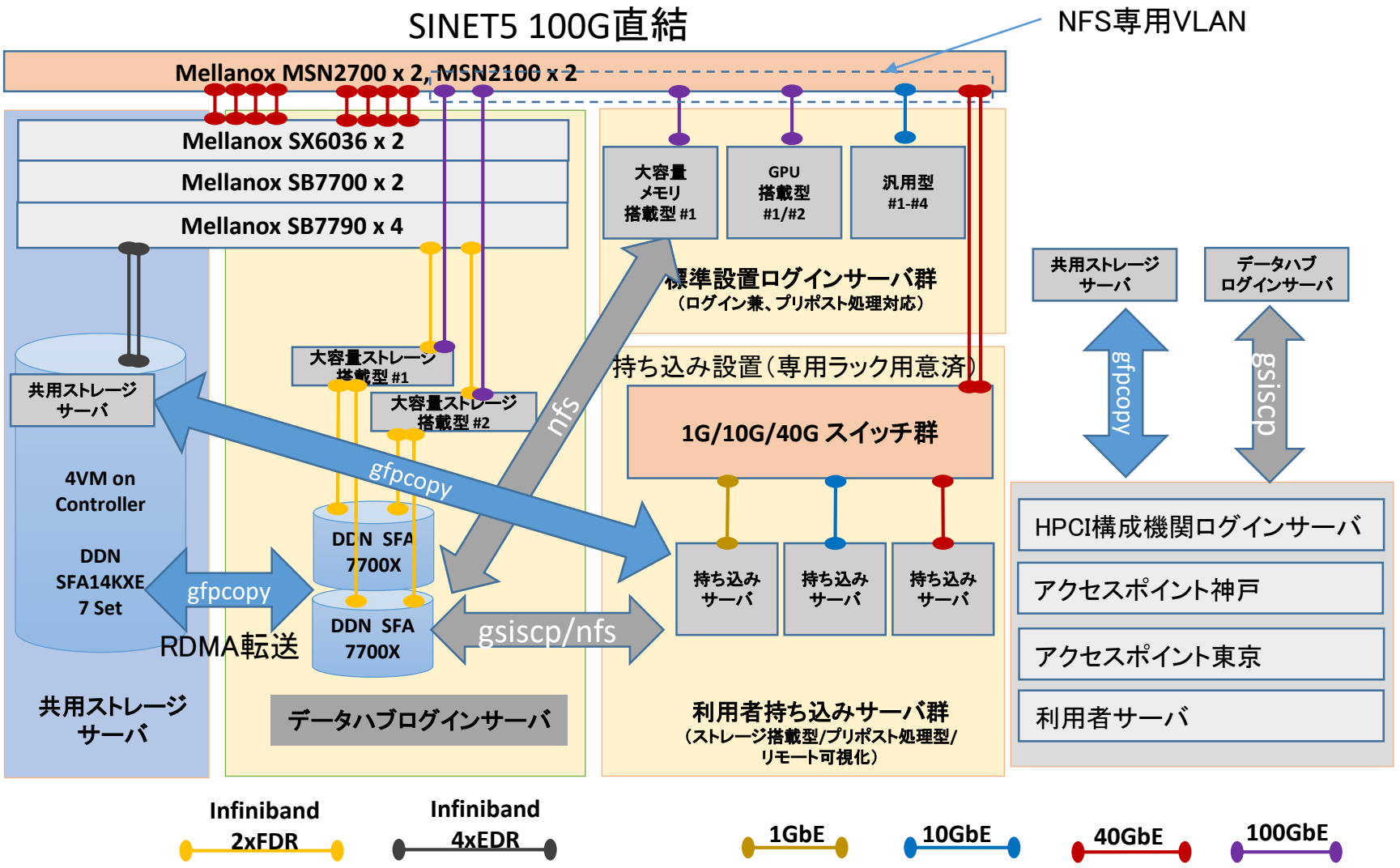
「共用ストレージ東拠点ポータル」の構成

- 位置付け
- データハブの実装方式
- データハブのサービス内容

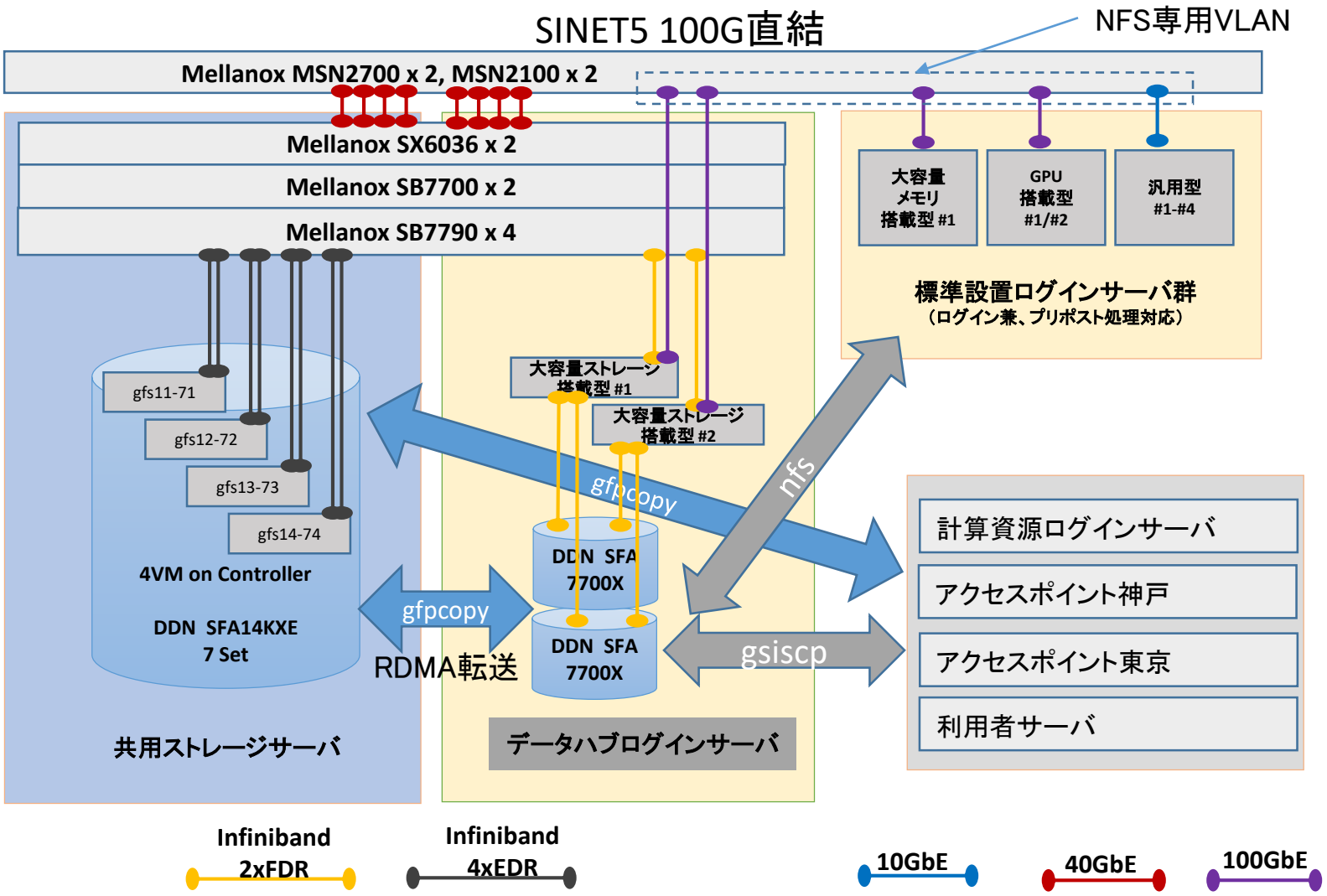
位置付け

- 専用アクセスポイント（アクセスポイント柏）が最終構想
- 当初はデータ入出力環境の提供を保留
  - アクセスポイントの役割であり、文部科学省、HPCIコンソーシアムの承認要
  - アクセスポイント神戸・東京との役割分担が未調整
    - 神戸・東京：HDD持ち込み、柏：サーバ持ち込みの分担が共存案
  - 持ち込みサーバのセキュリティ担保の実装方式・運用方法が未確定
- 「共用ストレージ東拠点ポータル」＝
  「アクセスポイント柏」 ー （データ入出力環境）

アクセスポイント柏（最終構想）



共用ストレージ東拠点ポータル構成（2018年度）

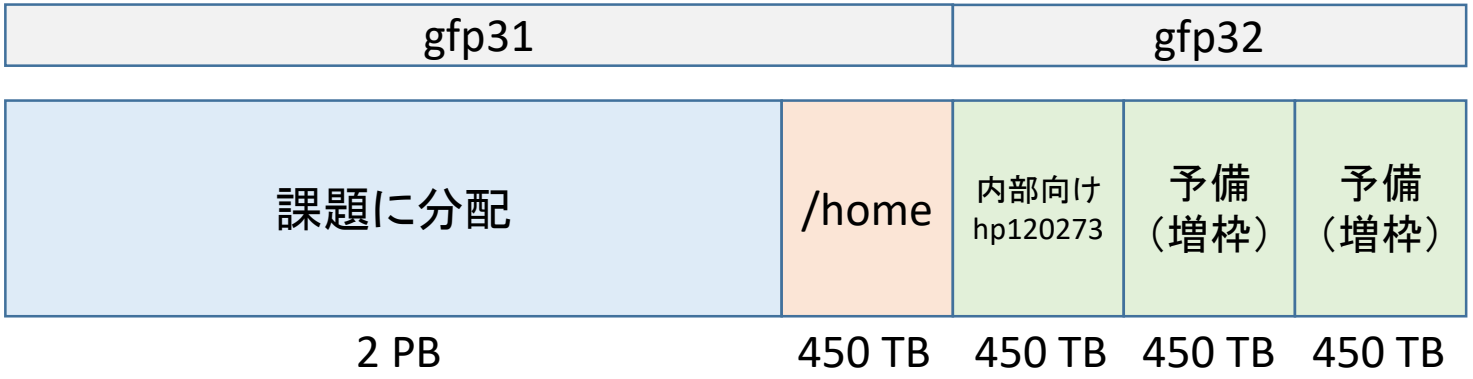


データハブの実装方式

- データ保存用ストレージサーバの転用
  - 3.8PBの大容量ローカルストレージ搭載
- 共用ストレージとの転送
  - 東拠点の共用ストレージサーバと同一IBネットワークに接続
  - 共用ストレージサーバとのRDMAによる高速転送
- 対外サーバとの転送
  - 対外サーバ（計算資源ログインサーバ、アクセスポイント神戸・東京、利用者サーバ）
  - 対外サーバとのgsiscpによる高速ファイル転送
- プリポスト処理サーバ間のデータ共有
  - ログインサーバ間は100Gbps直結
  - NFSマウントによるデータ共有
- NFS専用VLAN設置
  - NFSトラフィック分離
  - 持ち込みサーバ（サービス提供未決定）に対するセキュリティ対策の布石

データハブのサービス内容

- gfp31：一般の課題利用者向けのデータハブ
  - 合計2PBを課題毎共有領域として提供
  - 割当容量（クォータ設定値）は、該当課題への共用ストレージ割当容量の10分の1とする（ただし、最小 1TB、最大 50TBとする）
  - ホームディレクトリ領域（合計450TB）も提供
    利用者毎の割当容量（クォータ設定値）は、500GB
- gfp32：テストアカウント(hp120273)課題を含め、予備用（増枠等）のデータハブ
  - 予備の課題毎共有領域として提供
  - hp120273課題向け：主に共用ストレージ内部向けに提供
  - 一般課題向け：gfp31の一般割当枠以上の増枠を要望している課題等、状況に応じて提供



「共用ストレージ東拠点ポータル」の拡充方針

- 今後の拡充方針（課題）
- 利用者持ち込みサーバの扱い

今後の拡充方針（課題）

- 今年度（2018年度）は、「共用ストレージ東拠点ポータル」のハードウェアのみの提供に留まっている
  - 計算資源間のデータ移行でどのようなプリポスト処理が必要か
  - 産応協向け用途でどのようなプリポスト処理・リモート可視化環境が必要か
  - そのためのソフトウェアとして何が必要か
  - どこまでを標準機能として提供すべきか
- 「アクセスポイント柏」へ昇格するために、利用者持ち込みサーバをどう扱うかで、３つのレベル
  - 利用者持ち込みサーバは、一切許可しない（レベル0）
  - 利用者持ち込みサーバには、東拠点内の高速ネットワークへの接続を許可するだけ（レベル１）
  - 利用者持ち込みサーバには、NFSアクセスまで許可する（レベル２）

利用者持ち込みサーバの扱い

- レベル1（東拠点の高速ネットワーク接続のみ）
  - ストレージ内蔵の持ち込みサーバとの高速なデータ転送が可能であり、HDDのハンドキャリーの煩雑さが無い
  - セキュリティ面では、持ち込みサーバ上のデータ保全の課題有り
- レベル2（＋NFSによるデータ共有）
  - プリポスト処理・リモート可視化のソフトウェアは利用者側で自由に選定、利用可能
  - セキュリティ面では、持ち込みサーバ上のデータ保全に加え、利用者持ち込みサーバのRoot権限問題（他課題データアクセス）有り

関連サイト

http://oss-tsukuba.org/event/gw2019

https://www.hpci-office.jp/info/display/cnt00011