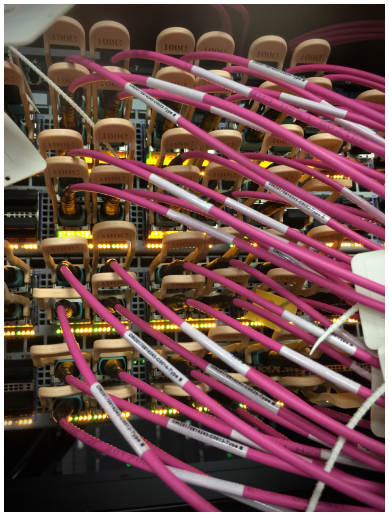
SINET5に支えられたHPCIにおける情報共有基盤、HPCI共用ストレージの紹介





理化学研究所　計算科学研究センター 原田 浩
筑波大学　計算科学研究センター　建部 修見
理化学研究所　計算科学研究センター 金山 秀智
東京大学　情報基盤センター　小瀬田 勇

2019/10/4

HPCI共用ストレージ

Overview of HPCI

出典：http://www.hpci-office.jp/pages/e_what_is_hpci

What is HPCI?

The innovative High-Performance Computing Infrastructure (HPCI) is a shared computational environment built with recommendations from the pre-establishment organization of the HPCI Consortium. Its operation has started since September 24, 2012.

HPCI connects the flagship "K computer" of RIKEN and other major supercomputers as well as storages of universities and research institutions in Japan via high speed networks. The mission of HPCI is to realize the scientific and technological computing environment where a wide range of HPC users in Japan can access national HPC resources efficiently.

To achieve this goal, HPCI seeks to answer to various users' needs and to promote the sharing of the innovative computational environment. We believe that these efforts will not only accelerate the scientific breakthroughs and development of the technologies, but also will contribute to increase the industrial competitiveness, development of human resources, and expansion of the user base.

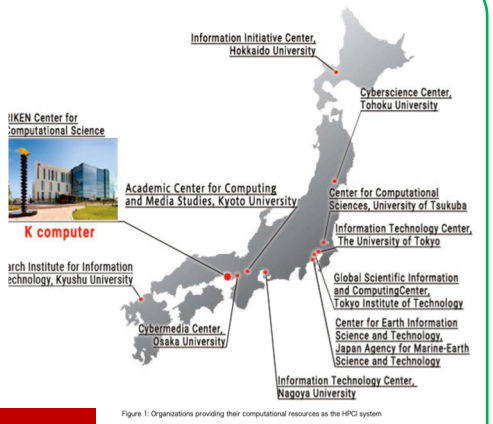
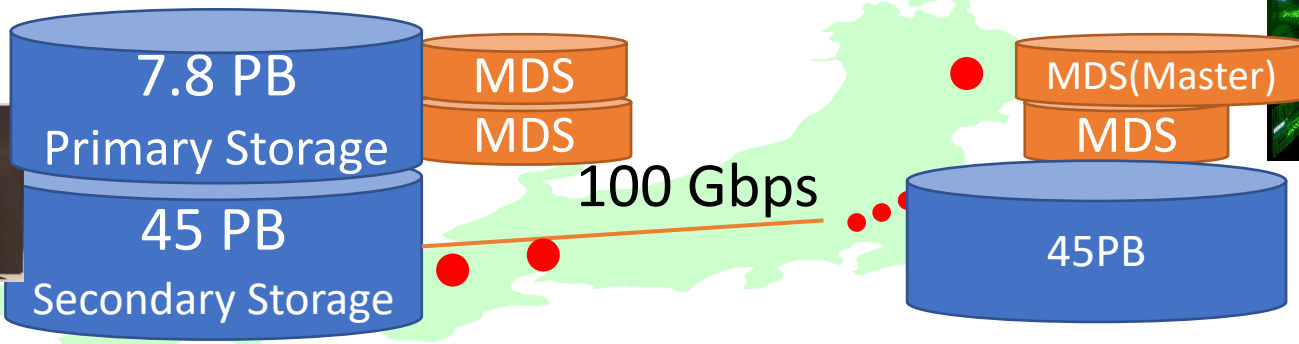


Figure 1: Organizations providing their computational resources as the HPCI system

2020HPCI課題公募中
ストレージ単体でも応募可能

- HPCI 資源としてストレージ資源を提供
- 全国のHPCI計算機資源からデータ共有可能な高速・大容量の単一ファイルシステム
- シングルサインオンで全国のスパコンセンターから利用可能
- データ冗長化やデータの自動一貫性チェックによるデータ保護
- 広域分散ファイルシステムGfarmを利用
- 通信内容をGSI(Grid Security Infrastruncture)で暗号化



理研計算科学研究センター

東京大学情報基盤センター

HPCI共用ストレージの特徴

・2018年10月10日から、無停止運転継続中 来週に連続稼働1年間達成予定(?)
現行システムは２世代目。2017年度に導入。

2019年度から45PB+45PBを提供
SINET5を介して理研・東大間でデータを二重化

高信頼システム(ネットワーク、データサーバ、ストレージを冗長化)

メタデータは東大:2、R-CCS:2
に分散して4重化
東大/R-CCS間フェイルオーバ

データ完全性チェック
チェックサムの自動照合

高性能化　ストレージ、ネットワーク

高いサービス継続性を実現
東大またはR-CCSによる単独運用可能

MDS　MDS(master)

7.8 PB
Primary Storage

51 PB
Secondary Storage

理研計算科学研究センター

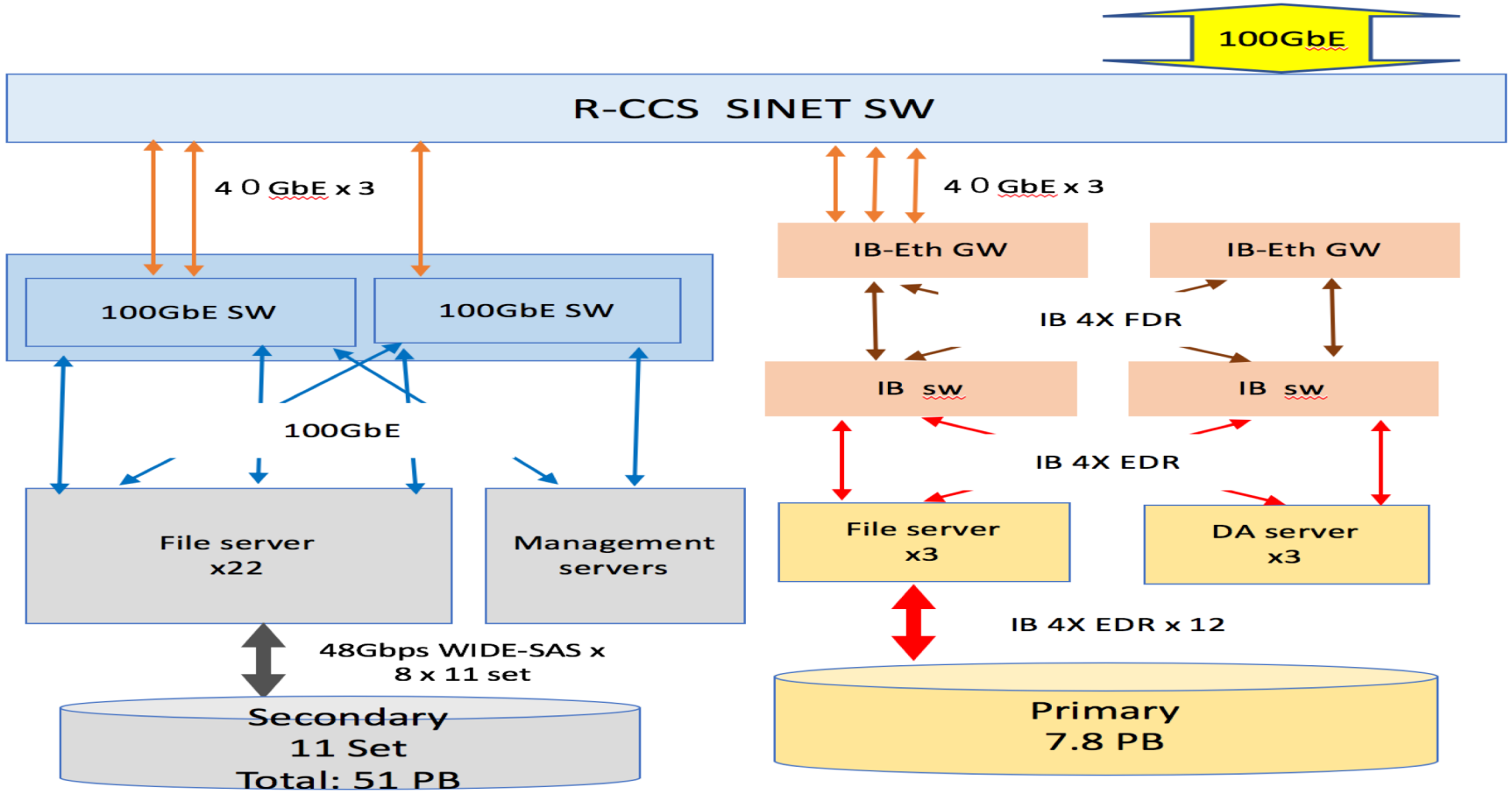
FT2018からデータ二重化サービス
東大、R-CCSにそれぞれ一つ複製を自動配置

MDS　MDS

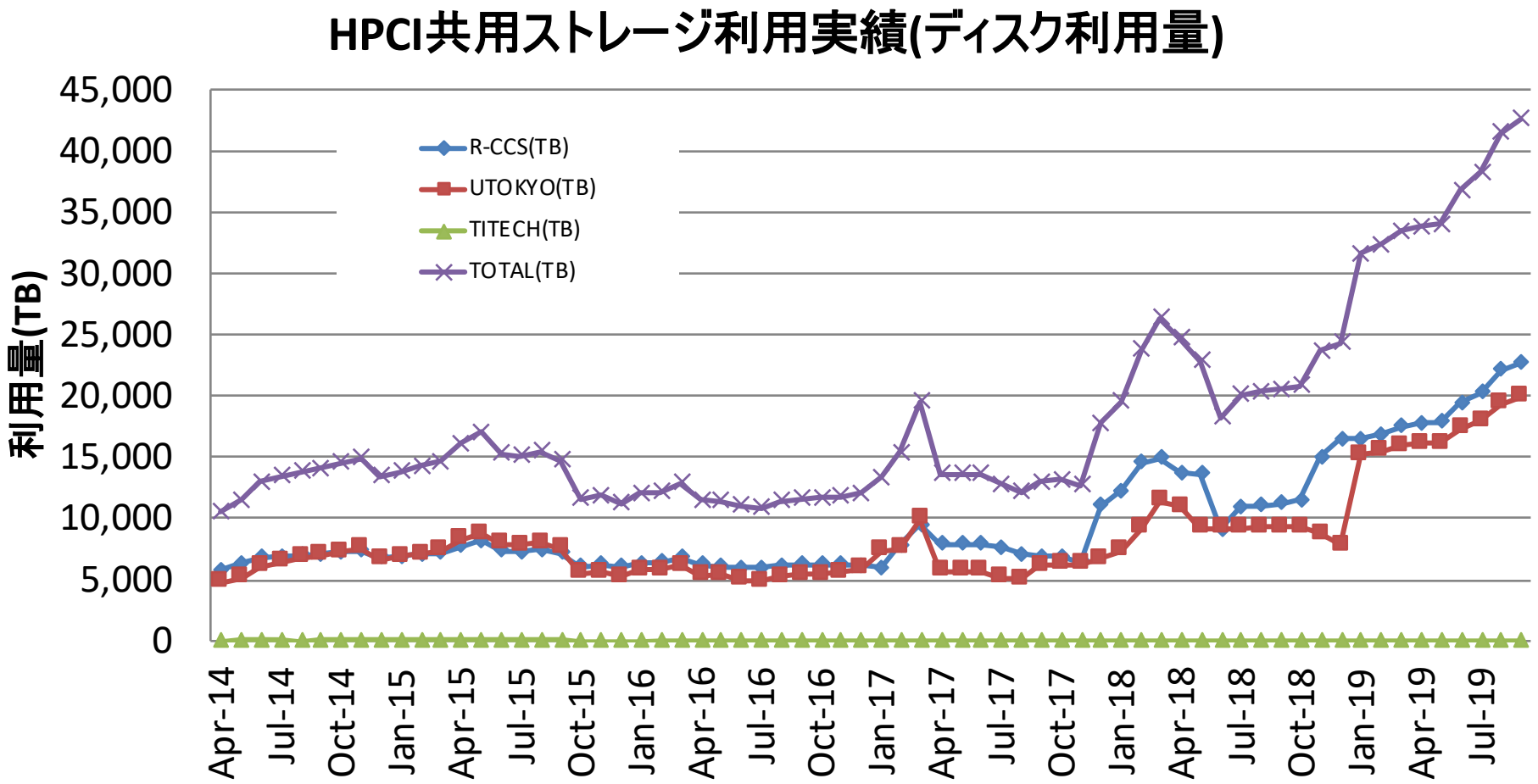
45 PB

東京大学　情報基盤センター

HPCI共用ストレージ



HPCI共用ストレージ

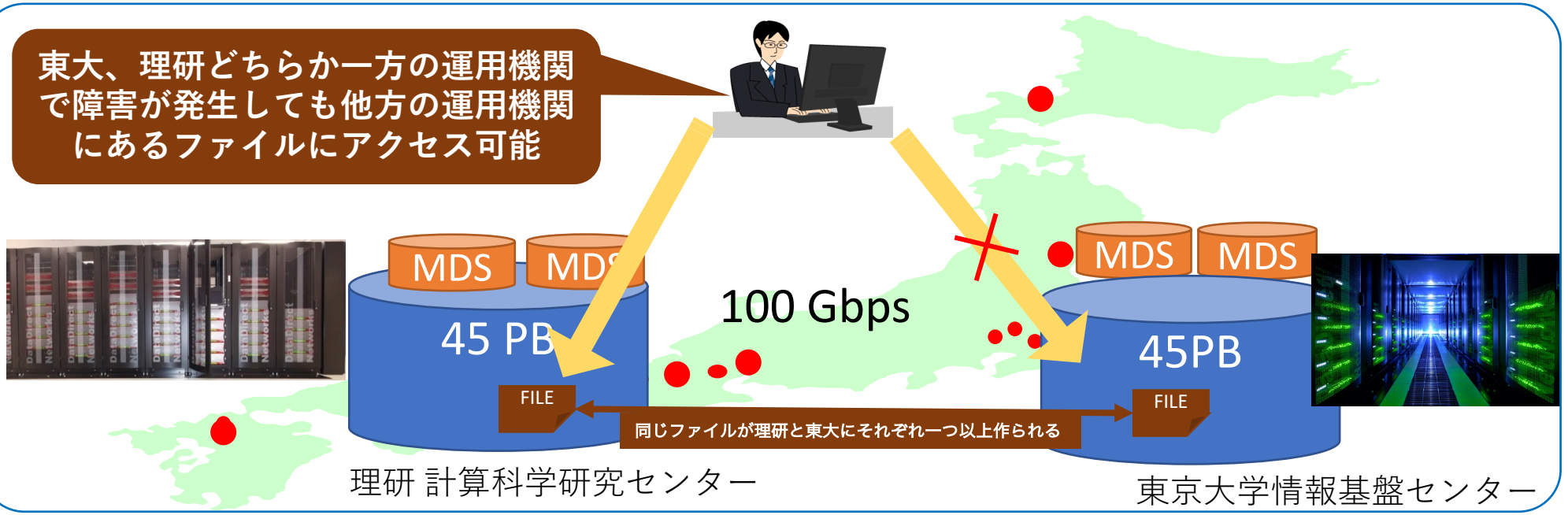


データ二重化運用

- 2018年度からファイルデータを東京大学、理研にそれぞれ一つ以上自動配置する運用を本格的に開始
  - 運用側によるファイルデータの複製管理(運用側が、ファイルのロケーション、多重度を管理)
  - ユーザは複製管理不要(ユーザは、ファイルのロケーション、多重度を意識せず利用可能)
  - 旧機材からのデータ移行も完了

- 稼働率の向上
  - メタデータの四重化＋データ二重化により、東京大学、R-CCSいずれか片方の運用機関だけでも継続運用可能
  - 計画停電、メンテナンス等によるサービス停止時間の削減
  - 2018年度以降は、稼働時間の大幅な向上を達成
  - 2018年10月10日から連続稼働継続中

- 強固なデータ保護
  - 東京大学、R-CCSいずれかで重大障害が発生してもデータは保全されます
  - データ書き込み時にチェックサムを自動照合
  - ファイルデータ二重化時にチェックサム照合も自動実行
  - データ書き込み時のチェックサム照合と合わせ、二重、三重のデータ完全性チェック



2019障害・メンテナンス分析(R-CCS機器)

HDD障害

一次ストレージHDD：MTBF 250万h、12TB

二次ストレージHDD ：MTBF 250万h 、10TB

運用期間	運用時間	運用HDD数	障害HDD数	MTBF
2019/4～2019/8	3672時間	800+6380	5	527.3万時間 (二次ストレージは全ディスクをSMARTで定期検査)

ネットワークSW障害

メタデータサーバ・データサーバは全て100GbE二重化、スイッチも二重化

片系リンクダウン	13	• 多くは、QSFPの抜き差しで復旧、原因不明 • 停止期間中に対策をしておきたい • 両系同時ダウンな発生していないので、なんとか運用

メンテナンス

実施期間	内容	サービス
2019/4/1～2019/4/3	• Gfarm更新 • 電子証明書更新(GfarmはGSI認証)	無停止

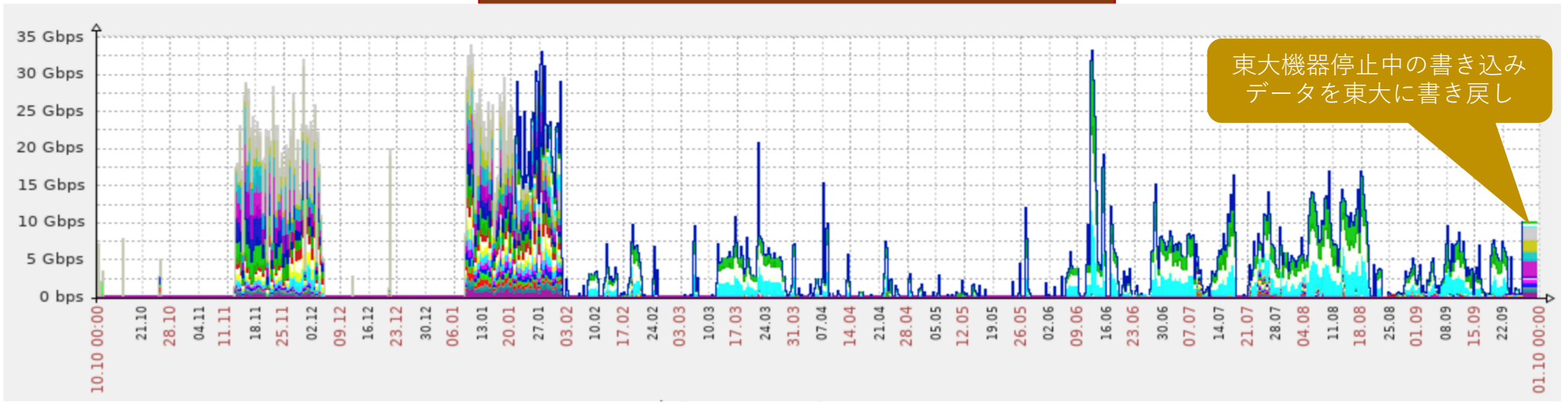
2019/7/8～2019/7/12	• 東大のマスタメタデータサーバ試験運用(R-CCS機器長期停止対策)
	• R-CCS ローカルメンテナンス
	  • メタデータサーバUPSの電源ケーブル交換
	  • 各種ファームウェア更新
		無停止
		「京」運用終了に備えて、データをHPCI共用ストレージに書き込み多数

2019/9/27～2019/9/30	• 東大設備点検(東大機器停止中は、R-CCS機器で片肺運転) • R-CCS機器停止 • R-CCS機器停止中、復帰は11月はじめを予定 • 東大機器で片肺運転中。　マスタメタデータサーバにフェイルオーバ	

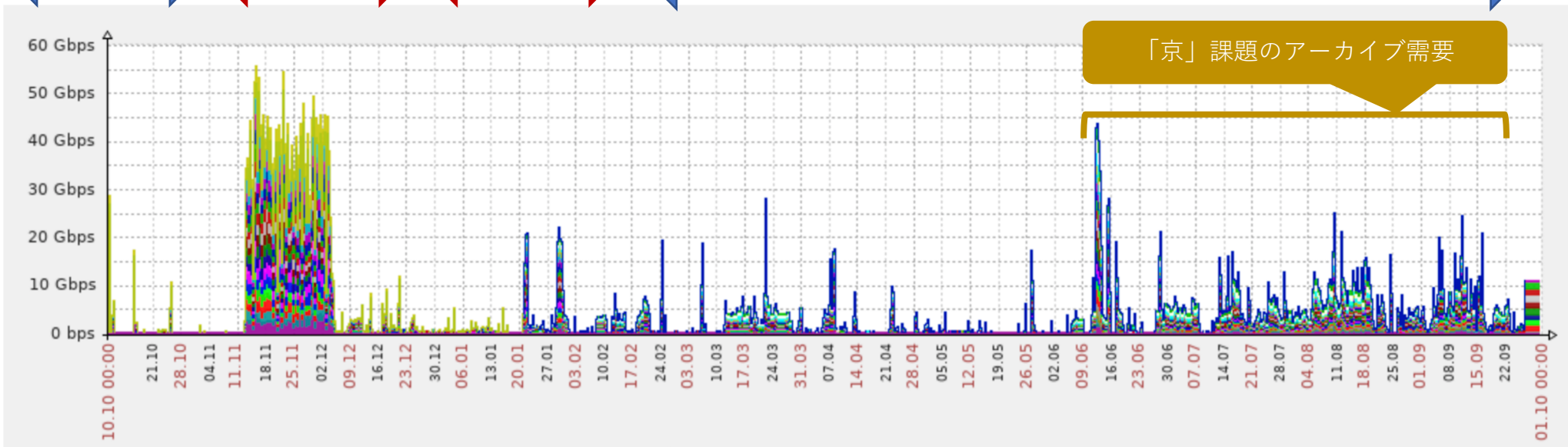
SINET障害・自然災害ともにゼロ

R-CCSネットワークトラフィック(過去１年間)

Outbound



東大:1+R-CCS:1 | 東大:1+R-CCS:2 | 東大:2+R-CCS:2 | 東大:2+R-CCS:2



Inbound

直近の自然災害
台風にもマケズ、雷にもマケズ

- 2018年8月23日深夜：台風20号の影響でR-CCS設備(電源供給)障害
- 翌2018年8月24日に緊急で電力の一時停止(事前通知あり)
- マスタメタデータサーバを東大にフェイルオーバして、機材停止
- R-CCS機材停止中はReadOnly運用
- 復電後に機材起動、マスタメタデータサーバをフェイルバック

- 2018年8月31日18時：落雷の影響で東大柏キャンパス情報基盤センター停電発生
- 東大メタデータサーバ、ファイルサーバダウン
- 復電後、全サーバ起動
- 幸い、マスタメタデータサーバはR-CCSで運用中

- 2018年9月26日～10月10日、東大柏、R-CCS、２週続けて週末計画停電のため長期メンテナンス
- AICSからR-CCSへドメイン変更
- サービス停止は、ファイルサーバのドメイン名、IPアドレス変更(東大のみ)のための３日間だけ
- データ破損、消失も無し

年間無停止・連続稼働に立ちふさがる壁？

共用ストレージ

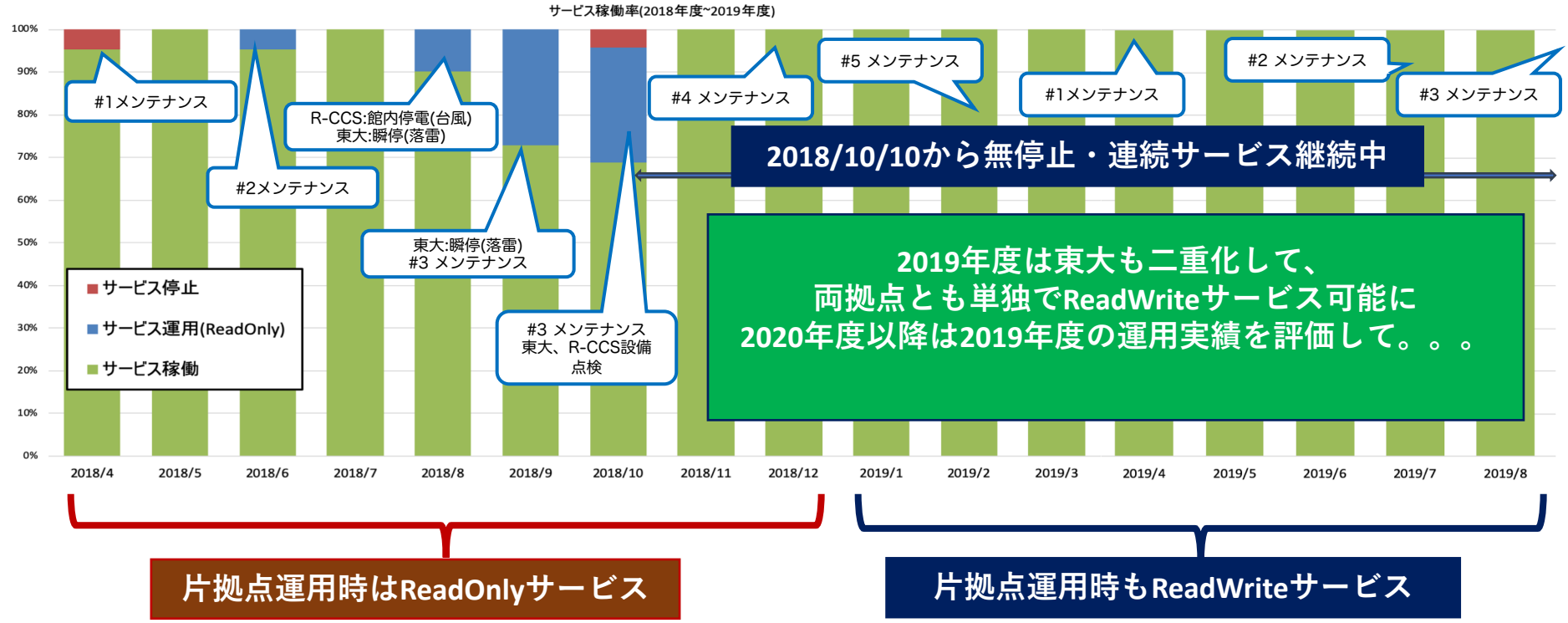
【共用ストレージ】R-CCS設備工事(サービス無停止)のお知らせ（2019年10月ころ)

作成者 理化学研究所 計算科学研究センター、最終変更日2019/01/18

お世話になっております、HPCI共用ストレージ担当です。
理化学研究所　計算科学研究センターでは2019年10月頃に大規模な設備工事を予定しております。
工事期間中は、R-CCS設置の共用ストレージ機器は運用を停止する予定です。
R-CCS設置機器の停止期間中も、東京大学設置機器による片拠点運用を予定しているため、
通常通りHPCI共用ストレージをご利用いただける予定です。
現在、R-CCS設置機器停止に備え東京大学設置のストレージ機器におけるデータ二重化を進めております。
停止期間などの詳細につきましては、決まり次第このページに記載する予定です。
ご協力のほど、よろしくお願い申し上げます。
--
HPCI共用ストレージ担当

- ReadOnly運用時間を削減して、過去最高の稼働率を目指したい
- 2019年度も9月~10月は鬼門
- R-CCS長期設備工事のため一ヶ月ほど機器停止を予定
- 東大片肺運転でサービス継続できそう→現在東大機器で片肺運転中(データは二重化)
  - マスタメタデータサーバの東大運用
  - 東大設置ストレージによるデータ多重化
- 9月10月をしのげれば、さらなる稼働率向上を目指せる
- 2019年度でデータ多重化運用の評価も固まると見込まれる。2020年度以降のデータ多重化方針も決められる

HPCI共用ストレージの稼働実績



- 2019年度は、稼働率100%継続中(無停止)
- 2018年度は、時間割合で99.3%サービスを提供(ReadOnlyサービス：6.84%)
  - 92.3%ReadOnlyの原因はR-CCS 台風による館内停電、東大　落雷による瞬停、第3回メンテナンス
  - サービス停止は計画した二回だけ。不意のサービス停止は発生していない
  - 2018年12月4日にR-CCS内でもデータ二重化し、R-CCS単独運用でReadWriteサービス提供可能に
  - 2018年12月14日開始の第４回メンテナンスでは、ReadWriteサービスを継続
- サイト間データ多重化運用の効果は絶大
- メタデータ四重化＆データ二重化はSINET5に支えられている