平成 28 年度スーパーコンピュータ HOKUSAI 運用報告書

情報基盤センター

1　理研スーパーコンピュータ・システム

1.1　はじめに

理化学研究所情報基盤センター（以下、「当センター」）で運用を行っているスーパーコンピュータ・システムは、理研における科学技術研究の推進と発展を目的とした大型共同利用計算機である。

当センターでは、2015 年 4 月からスーパーコンピュータ・システム HOKUSAI の運用を行っており、本年度は運用 2 年目であった。HOKUSAI システムからは、システムを 2 段階で導入することになり、第 1 段階システムである HOKUSAI GreatWave（HGW）がすでに稼働しており、第 2 段階システムである HOKUSAI BigWaterfall（HBW）は 2017 年程後半に稼働開始となる予定である。システムを 2 段階に分けたことにより、システムを全く使えない期間を短くし、日進月歩の勢いで進歩する計算機器やシステムに追随し、最新の計算資源を提供することが可能となった。

また、RICC システムの半分程度の計算資源を、HOKUSAI システムの一部として引き続き稼働している。HGW システムの主計算資源は RICC システムと計算機アーキテクチャが異なるので、アプリケーションによっては HGW の利用が非効率になる場合があり、利用者がアプリケーションに適したシステムを使い分けることができるようするためである。

1.2　HOKUSAI システムと運用スケジュール

HOKUSAI システムの構成を図 1 に示す。今年度の HOKUSAI システムは昨年度に引き続き、HGW システムと半分に削減された RICC システムからなる。フロントエンドは HGW と RICC で分かれており、利用者は HGW のフロントエンドには直接アクセスできるが、RICC のフロントエンドにアクセスする際には、一旦 HGW のフロントエンドにアクセスした後、RICC のフロントエンドにアクセスするという手順となる。ファイルシステムについては共通で、HGW システムのファイルサーバに RICC システムからもアクセス可能になっている。

運用スケジュールの概要を図 2 に示す。歴代のスーパーコンピュータ・システムで RICC システムまでは 1 つのスーパーコンピュータ・システムは 5 年間の運用を行ってきたが、HOKUSAI システムからは、全体システムの運用期間を 7 年間とし、2 段階のシステムを 2 年程度ずらして立ち上げ、各段階のシステムは 5 年間の運用を行う。

2 段階の運用にするのは、システム完全停止期間の最小化と新技術への迅速な対応の 2 つの目的があるからである。今までのシステム更新時期においては、ハードウェアの物理的な入れ換えやデータ移行などを行うために 1 か月程度の全停止を行う必要があったが、2 段階の立ち上げを行うことにより完全停止期間を 1 週間以内にできると考えている。また、2，3 年に一度新システムを導入することにより、より最新の技術の計算機を提供することが可能となる。

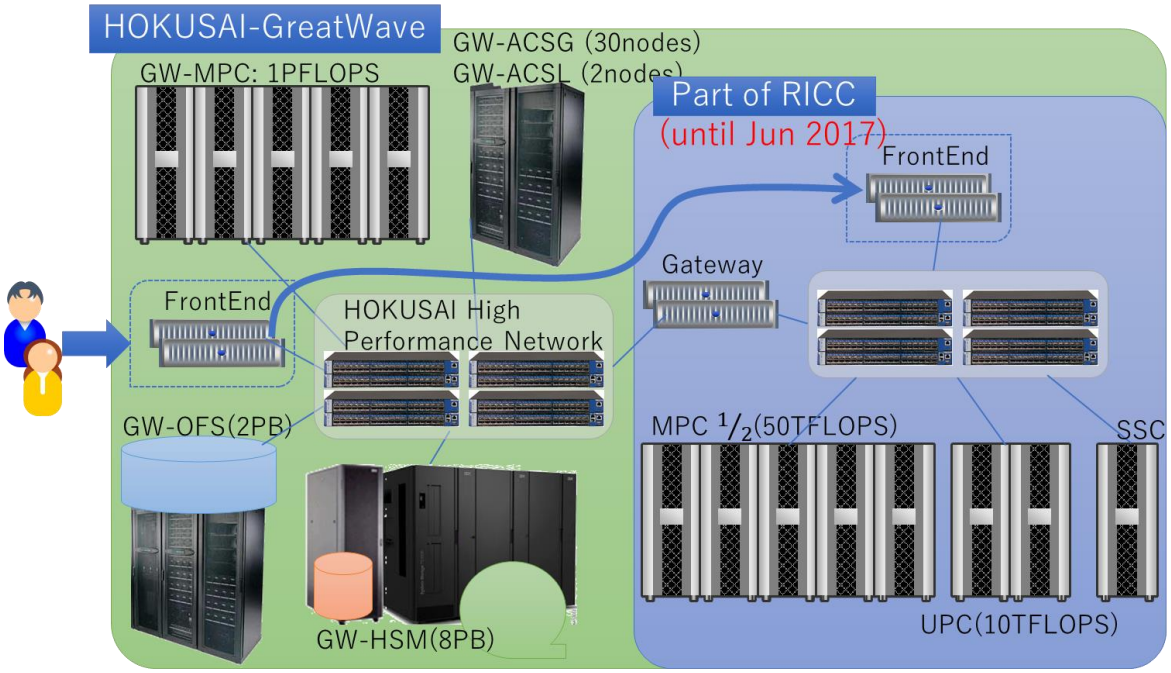

図 1　HOKUSAI システムの概要

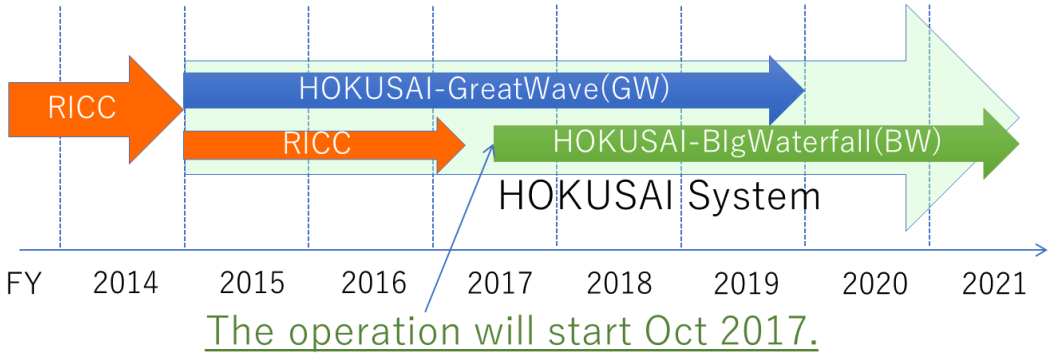

図 2　運用スケジュール概要

1.3　HGW システムの概要

HOKUSAI システムの第 1 期システムとして稼働開始された HGW システムは、3

種類の計算資源と 2 種類のストレージ、高速ネットワークのシステムとなっている。

1. GW-MPC：超並列演算システムとして、スーパーコンピュータ「京」と互換性のある Fujitsu PRIMEHPC FX100 を 1,080 ノード（1PFLOPS）
2. GW-ACSG : 4GPU を搭載可能な SGI C2010G-RP 演算サーバ 30 ノード（K20X をノード当たり 4 枚、合計 120 枚搭載）
3. GW-ACSL : 1TB のメモリを搭載した Fujitsu PRIMERGY RX4770M1 サーバ 2 ノード
4. Online storage : Fujitsu Exabyte File System(FEFS)による並列分散ファイルシステムにより 190 GB/s の広帯域を実現した、実容量 2.2PB の NetApp5600 からなるストレージ
5. HSM storage : IBM GPFS（General Parallel File System）+ TSM（Tivoli Storage Manager)により HSM 構成し、ディスク容量 300GB、テープ容量 7.9PB のシステム
6. High bandwitdth network : Infinband FDR を FBB（Full Bi-section Bandwitdh)で構成した高速ネットワーク

2　理研スーパーコンピュータ・システム HOKUSAI システムの運用報告

2.1　HOKUSAI システム 2 年目の運用概要

HOKUSAI システムは本年度 2 年目の運用を行った。4 月からの運用開始直後から順調に利用が進み、計算リソースが逼迫することとなった。特に、GW-MPC は昨年度から引き続き大変混雑していた。ただし、昨年度の反省を踏まえ、今年度からは課題審査の仕組みを大幅に変えたことにより、利用者は割り当てられた計算リソースを十分に使うことが可能になった。また、後述するように HGW システムのハードウェアの故障が、本年度始めは昨年度に比べ減少していたが、年度の後半になるにつれ多発するようになり、対策を練る必要が出てきた。

2.2　課題審査と利用者数統計

本年度はスーパーコンピュータ課題審査委員会による利用希望者の課題申請審査を 3 月と 9 月の 2 回実施した（表 1)。今年度から、一般・専有利用審査は大きく変更することとし、サブシステム毎に各課題の申請時間の上限を 20%とし、また全システムの 10%を超える課題については大規模利用課題としてより厳しい審査を行うことになった。結果として、大規模利用課題は 5 課題あり、2 課題については審査により不採択となった。簡易利用については年度を通じて随時申請を受け付けた。月毎の課題採択数を図 3 に、課題数の推移を図 4 に示す。2017 年 3 月末時点で 135 課題が利用していた。図 5 に分野別の課題数を示す。

表 1　課題応募・採択数

課題審査委員会	利用開始月		応募	採択
第 1 回	4 月	占有利用課題	1	1
		一般利用課題	35	33
第 2 回	10 月	占有利用課題	募集なし	0
		一般利用課題	2	2

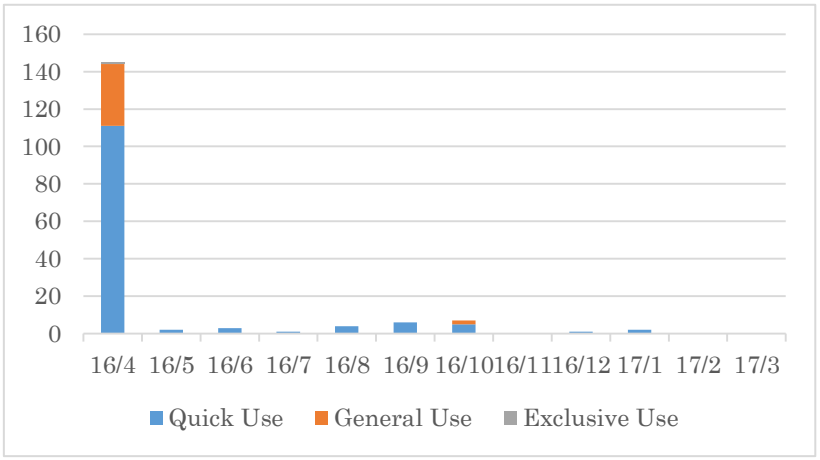

図 3　課題採択数推移

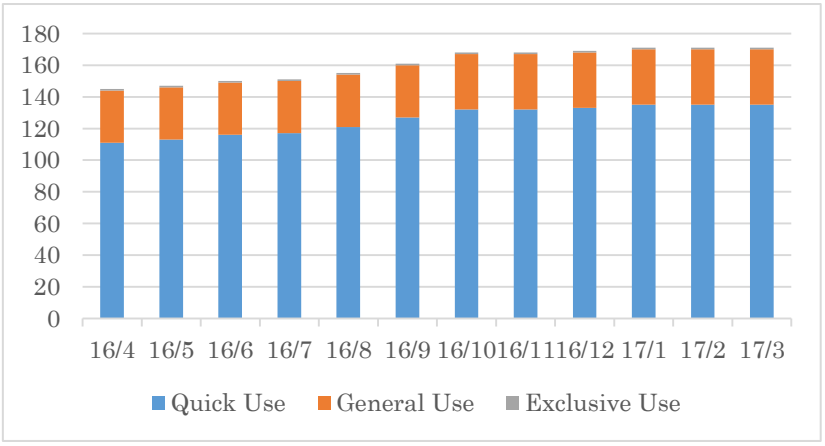

図 4　課題数累積推移

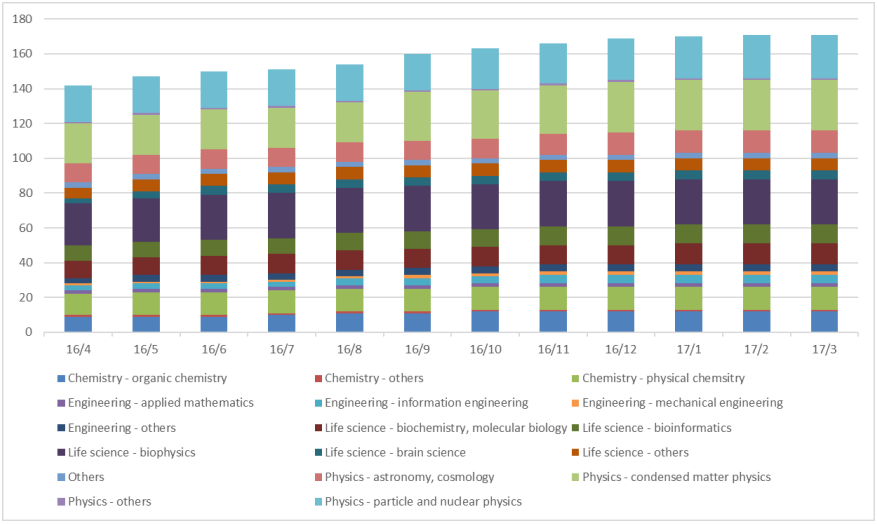

図 5　分野別課題数累積推移

図 6 に役職別のアカウント数を、図 7 にセンター別のアカウント数を示す。2017 年 3 月末時点で、利用者数は 385 人であった。利用者は様々なセンターに所属しており、ILs、仁科加速器研究センター、計算科学研究機構の順に利用者が多かった。

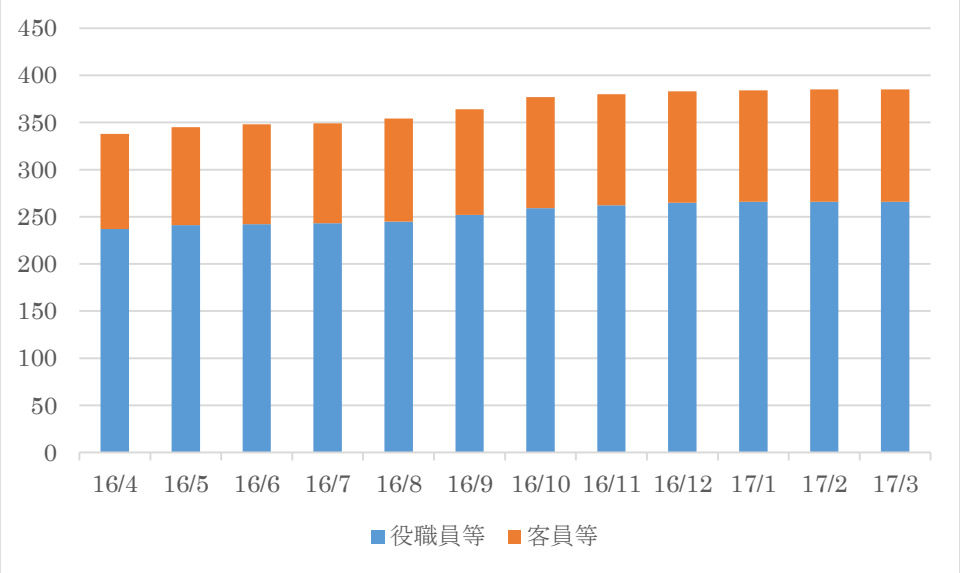

図 6　役職別アカウント数累積推移

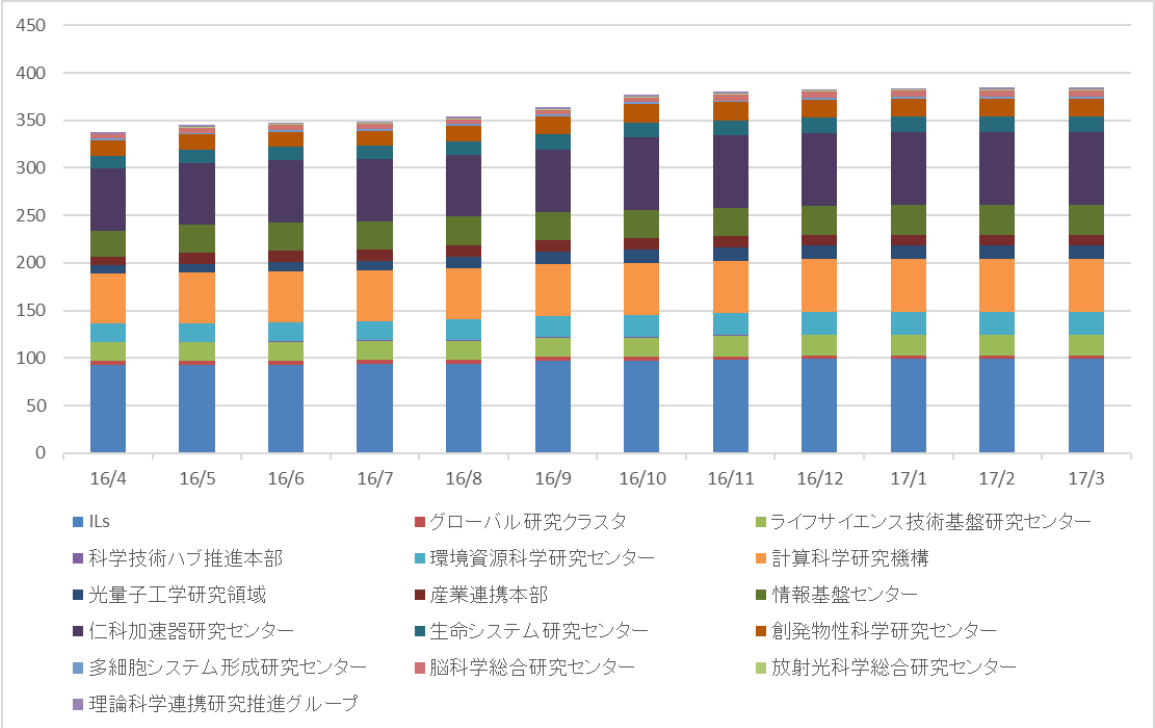

図 7　センター別アカウント数累積推移

2.3　HGW システムの稼働状況

HGW システム全体の利用状況を図 8 に示す。簡易利用は赤で、一般利用は緑で色分けしてある。4 月の運用開始当初から 90%程度と高い稼働率で、その後も 90%を超える高い値で推移した。また、簡易利用は 10-20%で収まるような設計を行っているが、ほぼその通りの利用状況となった。

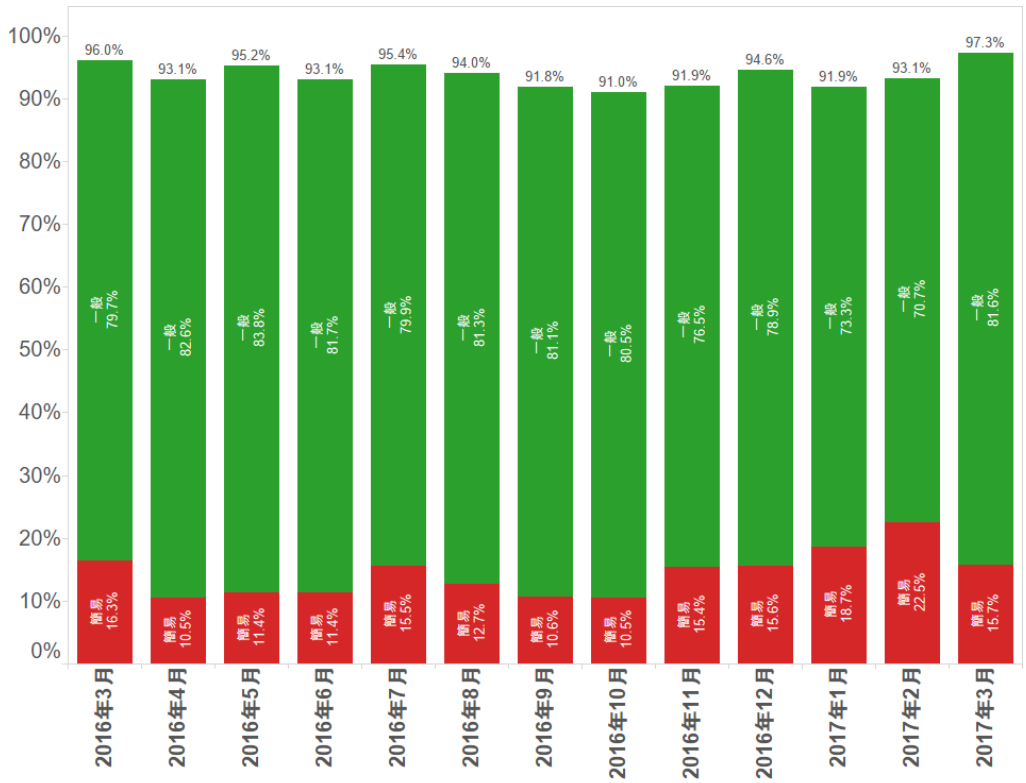

図 8　HGW 利用状況

HGW のコア時間消費状況を分野別で示したものを図 9 に示す。素粒子・原子核物理で総コア時間の 3 分の 1 を消費しており、物性物理、生物物理、物理化学、天文の分野の利用が次に多くなっている。

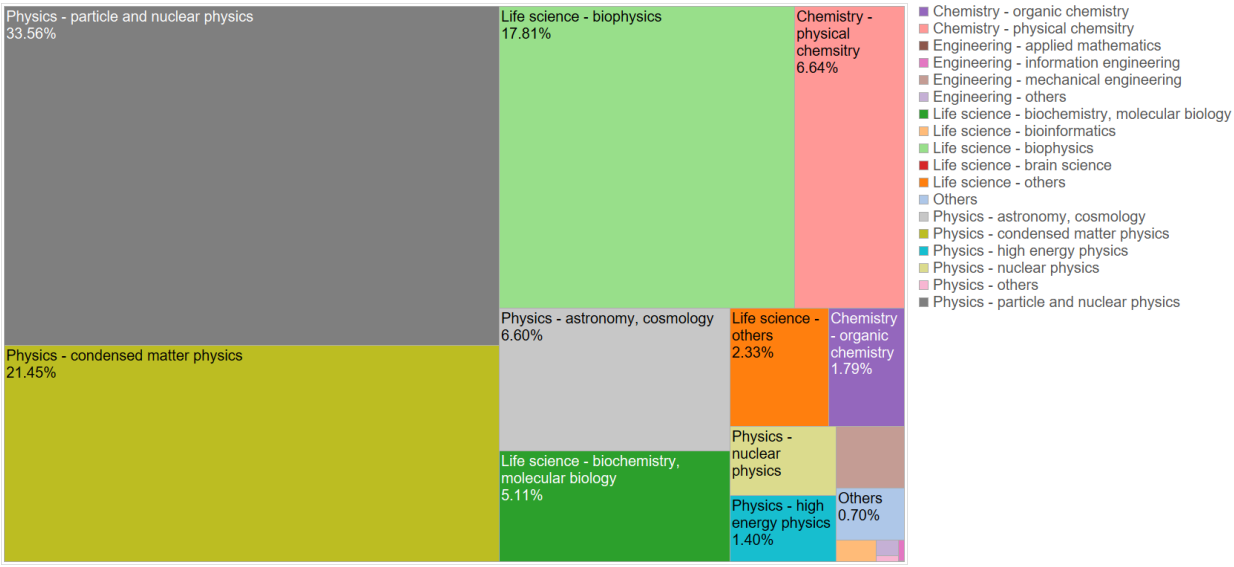

図 9　分野別での HGW の総コア時間消費の概要

GW-MPC での課題毎の割当コア時間に対する消費割合を、一般利用について図 10 に、簡易利用について図 11 に示す。一般利用については、年間を通じて継続的に利用していた課題については 80%を超えるような消費率となった。これは課題審査の方法を変えたことによって、大きく改善された点である。簡易利用課題については、10 課題程度が 40%以上の高めの消費率となり、次の 20 課題程度が 10%以上の消費率で、残りの課題は 10%以下の低い消費率となっていた。

一般利用の一部の課題については 100%を超える消費率になっているが、これは消費コア時間を計算する際の計算ミスに由来する。この計算ミスは 2004 年に稼働した RSCC（RIKEN Super Combined Cluster）の頃から発生したようであるが、一般利用課題については実際に使った量を 0.8 倍し、簡易利用課題については、1.2 倍するという処理がされていた。この間違った処理は課題毎の消費されたコア時間を計算する際だけに行われており、システムのジョブ充填率など他の統計量については影響がない。

この間違った処理により、一部の一般課題で 100%を超える計算リソースが消費され、簡易利用課題については、許可された時間の 83.3%までしか実際は使うことができなかった。大部分の課題については、資源利用のフェアシェア（公正分配）制御で実際に使える計算時間が決まってくるので、このことによる影響は大きくはないと考えられる。ただし、長期間に渡って利用者にご迷惑をお掛けしたことを深く反省し、このようなミスを繰り返さないことが重要であると考えている。

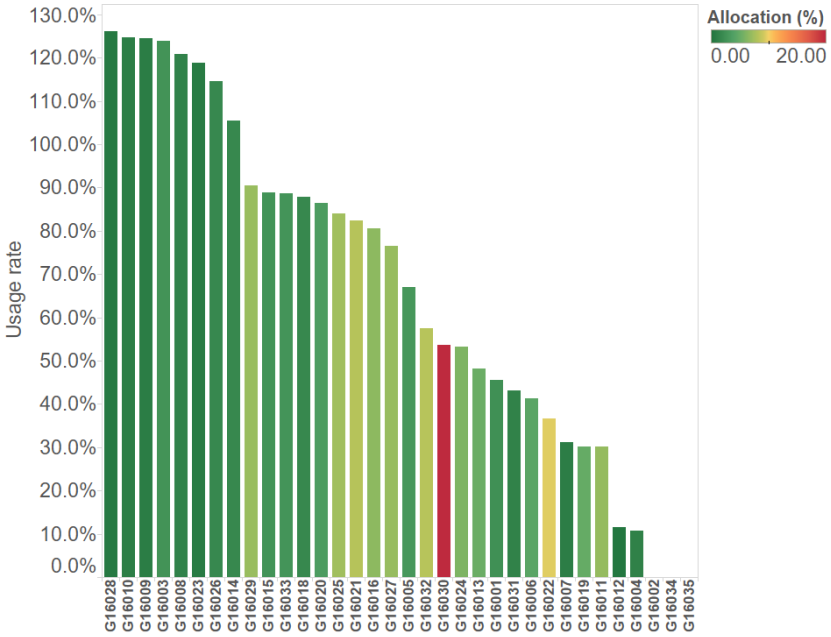

図 10　課題毎の割当コア時間消費割合(一般利用)

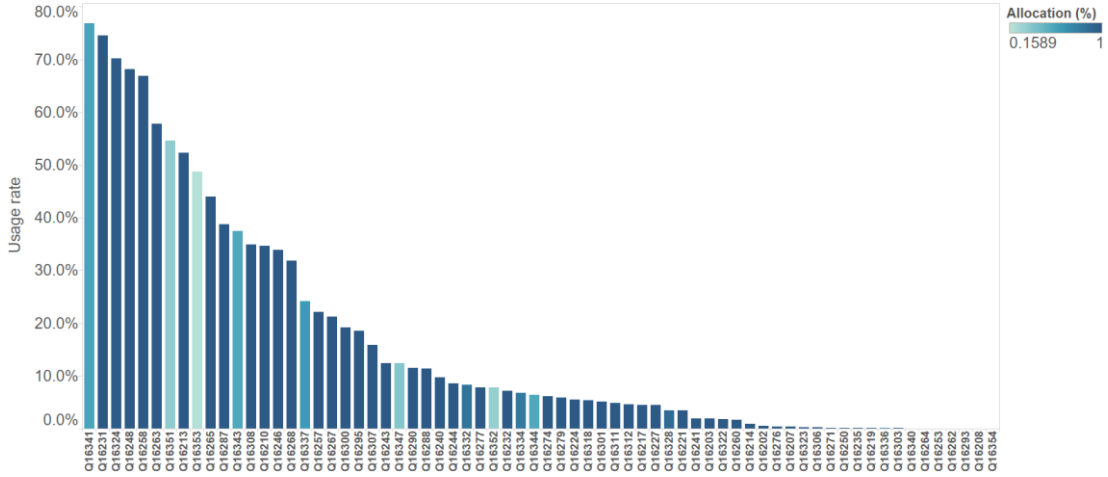

図 11　課題毎の割当コア時間消費割合(簡易利用)

2.4　　個別システム稼働（ジョブ稼働）状況

HGW（図 12-14）と RICC（図 15-16）の個別のシステムの稼働率を以下の図に示す。主要な計算資源である GW-MPC も RICC-MPC は、本運用開始直後から概ね 90%を超える高い稼働率を示している。その他の GW-ACSG、GW-ACSL、RICC-UPC は月によっては、80%以下の稼働率となり比較的空いていることもあったが、基本的には高い稼働率で推移した。

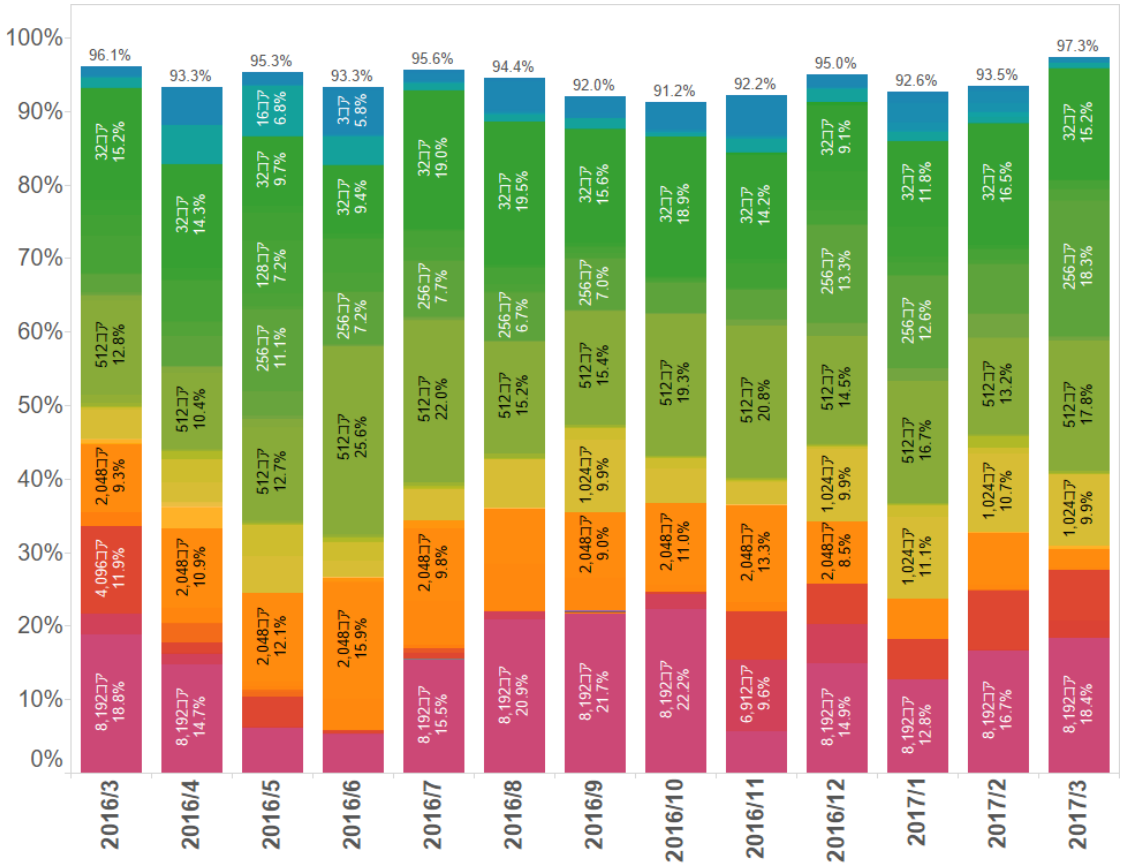

図 12　GW-MPC 稼働率（ジョブ充填率）

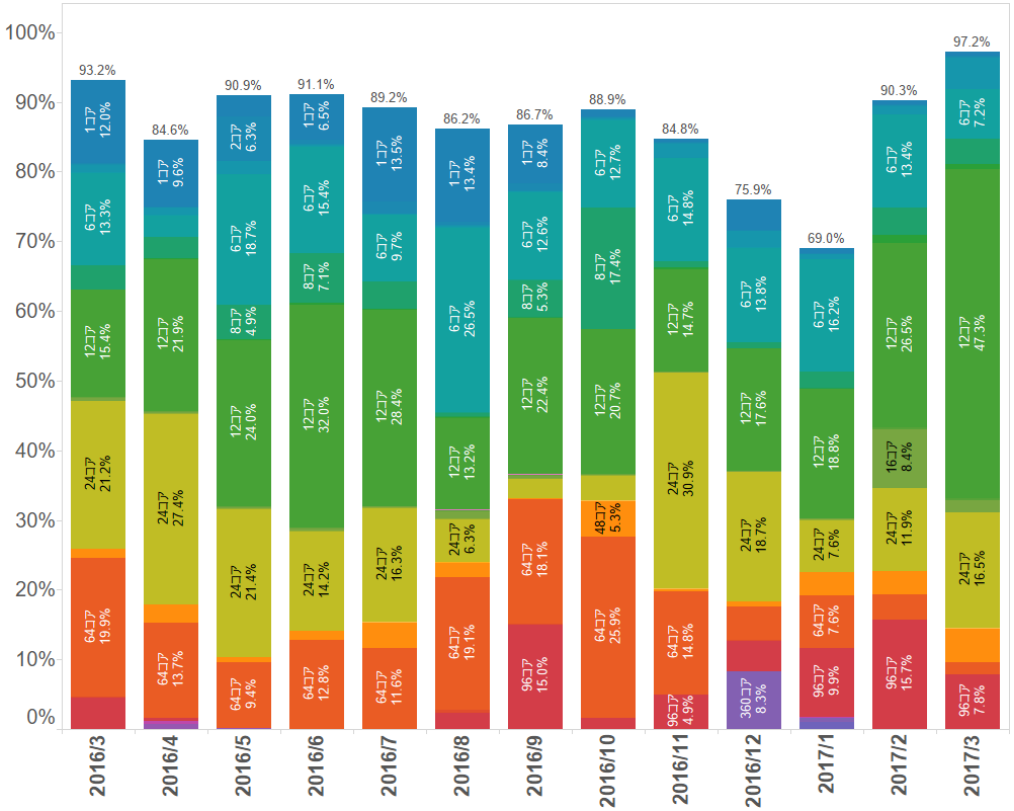

図 13　GW-ACSG 稼働率（ジョブ充填率）

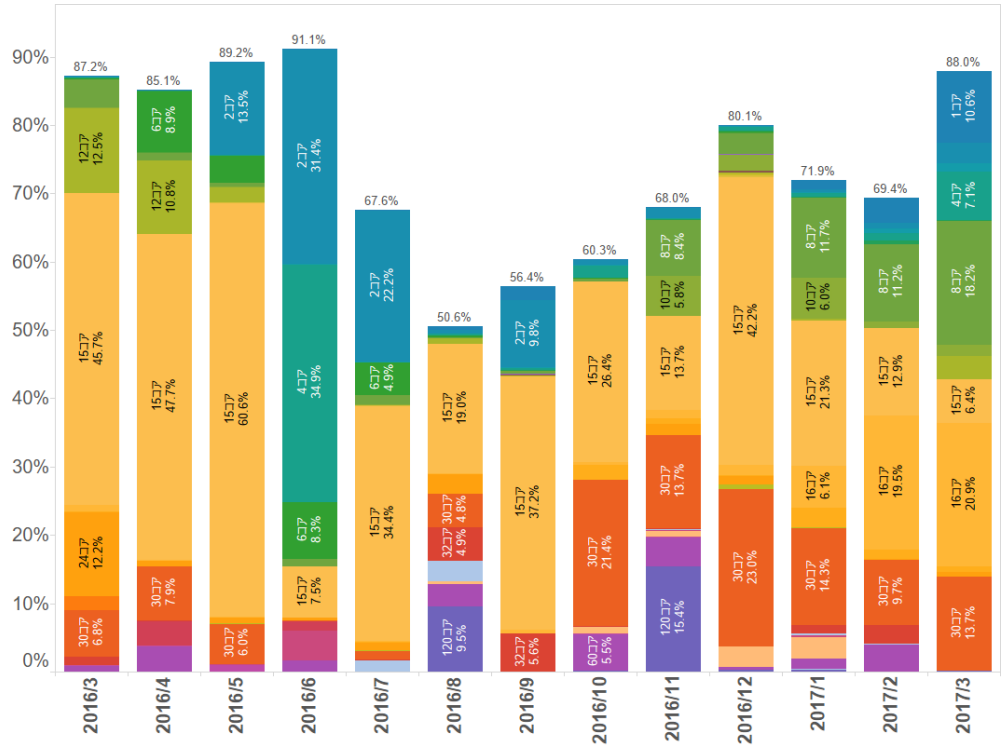

図 14　GW-ACSL 稼働率（ジョブ充填率）

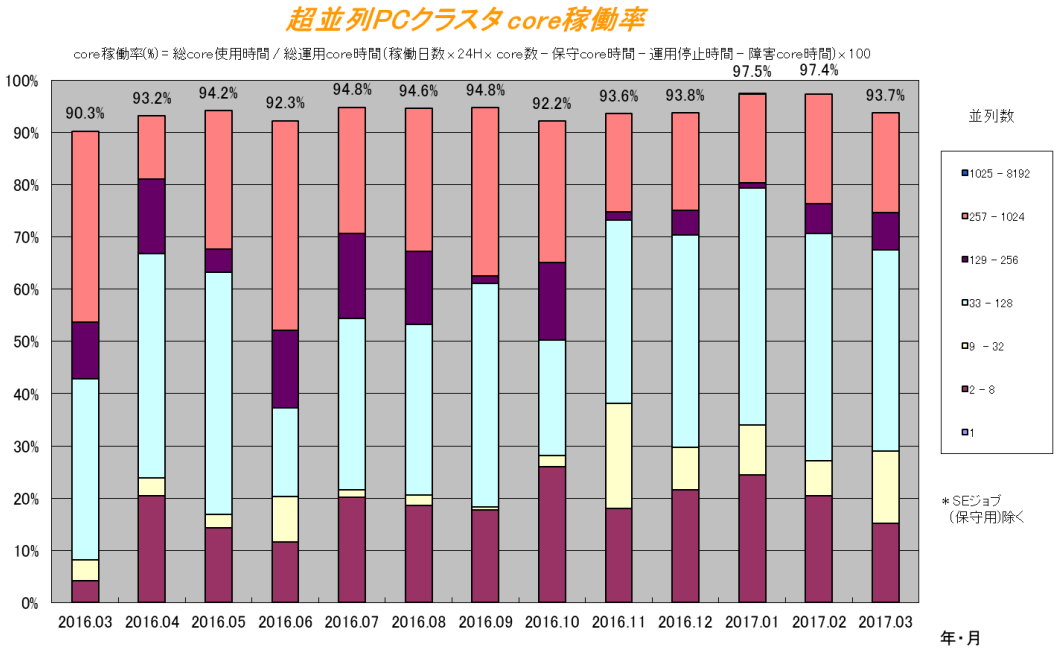

図 15　RICC-MPC 稼働率（ジョブ充填率）

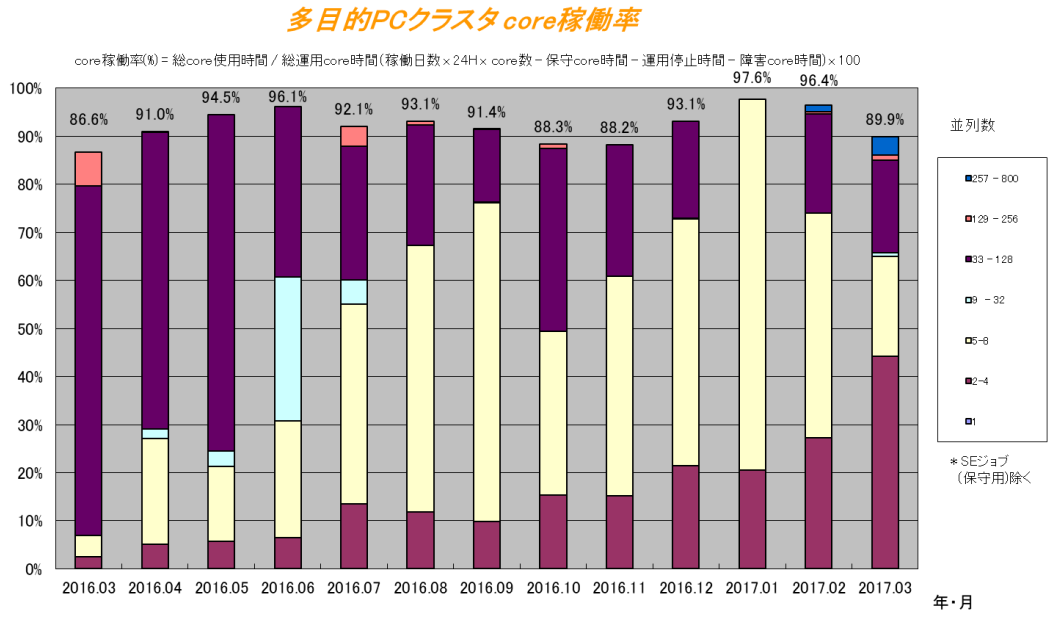

図 16　RICC-UPC 稼働率（ジョブ充填率）

2.5　HOKUSAI データ預かりサービス

HOKUSAI データ預かりサービスは，スーパーコンピュータ・システム HOKUSAI のテープ・ライブラリ装置を利用したサービスで，スーパーコンピュータ利用者以外の理研の研究者にも大容量データのバックアップを提供するものである。旧 RICC システムでの同様のサービスから利用要件とインターフェースを変更し，一部の利用者を引き継いで 2015 年 12 月にサービスを開始した。

本サービスでは基本的にデータは研究室単位で管理され，容量は初期値 4TB から最大で 52TB まで拡張可能である。データの通信は，高速データ転送が可能であり，かつデータの再送機能をもつ専用ツールをインストールしておこなう。

図 17 に利用者数の推移を示す。2015 年 12 月のサービス開始時は旧サービスから移行した 14 名であったが，2017 年 3 月末では利用者は 21 研究室 47 名である。

図 18 に本サービスで保存されているデータ量とファイル数の推移を示す。サービス開始時は, 94TB, 約 67,000 ファイルであったが, 2017 年 3 月末では, 154TB, 約 62,000 ファイルである。2017 年 1 月から 3 月にかけてデータ量が大きく増減しているのは，一部の旧ユーザのデータを新規のアカウントに移動するためのコピーが発生したためで，移動完了後は元のデータ量に戻っている。

本サービスではテープを使用しているのでサイズの小さいファイルを扱うと非常に効率が悪くなる。そこで，1 ファイルのサイズを大きくして保存するように利用者には繰り返しアナウンスした。その結果，ファイル数については図 18 に見られるように比較的増加が緩やかになり，システムの効率的な運用につながった。

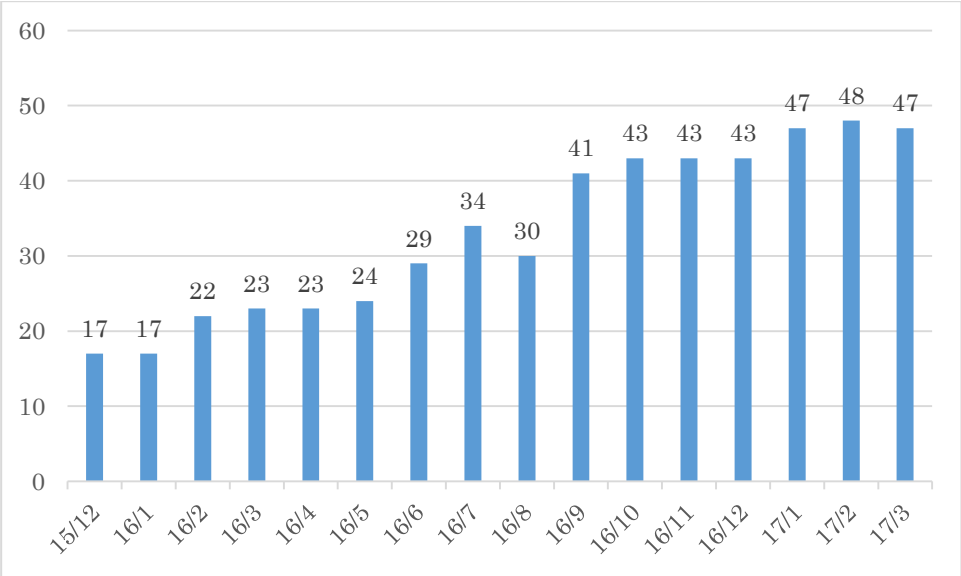

図 17　データ預かりサービス利用者数累計推移

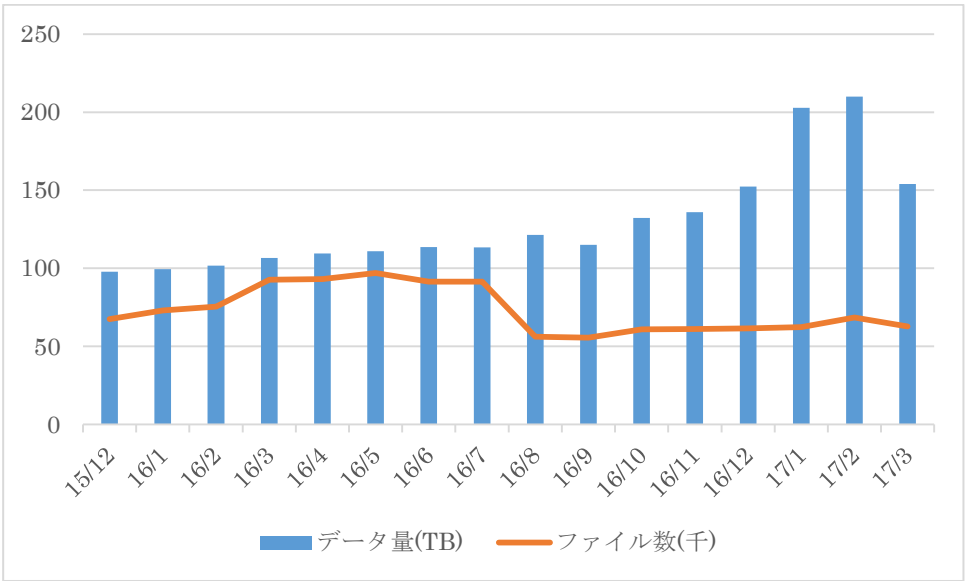

図 18　データ量とファイル数推移

2.6　障害情報

HGW システムのハードウェアの障害件数を図 19 に示すが、GW-MPC が故障件数のほとんどを占めていた。GW-MPC の障害の内訳を図 20 に示すが、メモリ障害(HMC)とネットワーク関連(Tofu)の障害件数が多く、特にネットワーク関連の障害が、年度後半にかけて増加した。GW-MPC の障害対応は、引き続きハードウェアベンダーと協議し、障害発生を減らしたいと考えている。また、RICC システムのハードウェアの障害件数を図 21 に示すが、RICC については年度を通じて障害件数は低かった。

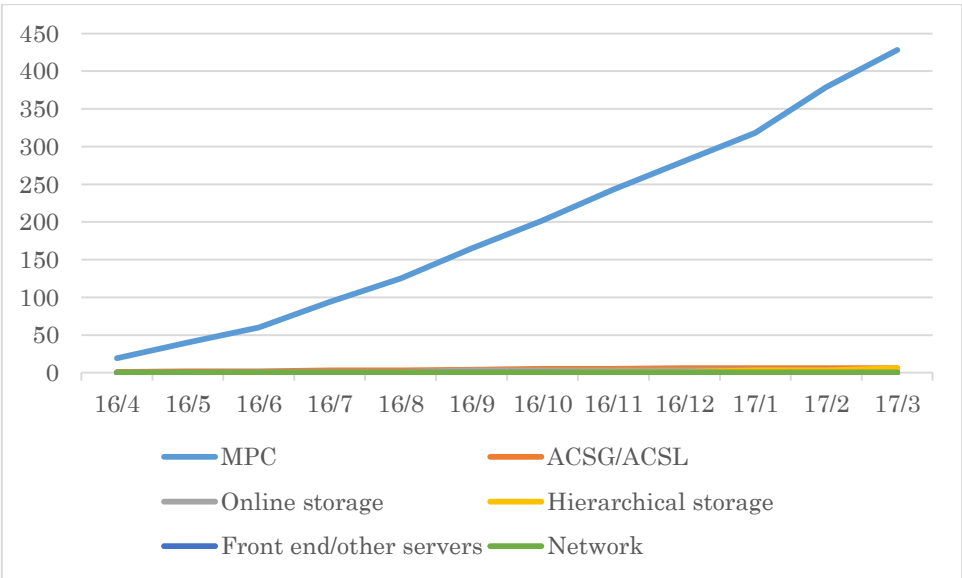

図 19　2016 年度 HGW ハードウェア障害件数累計推移

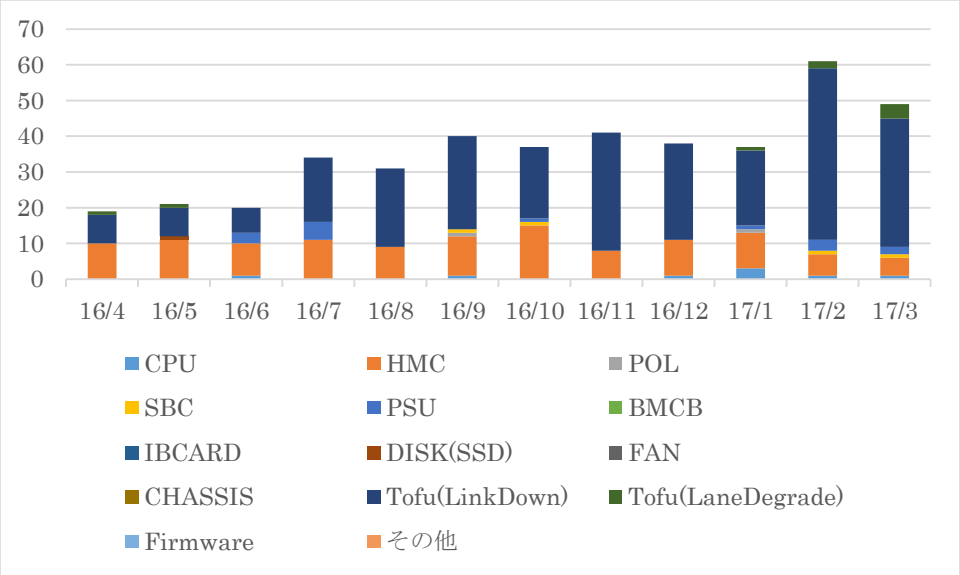

図 20　2016 年度 GW-MPC ハードウェア故障詳細推移

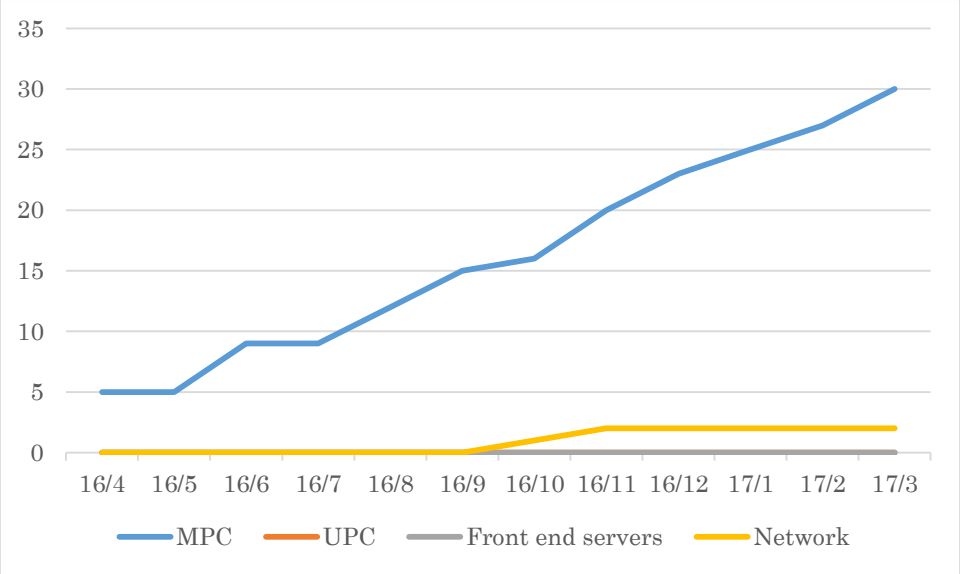

図 21　2016 年度 RICC ハードウェア障害件数累計推移

3　講習会の実施

本年度は従来行っていた少人数で行う座学によるプログラミングなどの基本的な内容の講習会の開催を行わなわなかった。来年度は新たに導入するシステムの構成を踏まえて利用者のニーズにあった講習会を企画する予定である。

4　アウトリーチ・成果公開

4.1　　見学

本年度のスーパーコンピュータ視察・見学は 28 回（973 名）であった。1 月以降 4D シアターの運動計測用設備（フォースプレート）の設置工事や計測があり見学を受け入れ可能な時期が限られたが例年と同程度の見学者を受け入れることができた。次年度は公募見学についての枠組みを見直す計画があり年度の早い時期の公募の見学会は行われない予定である。

表 3　視察・見学者数

日付		訪問者	人数
4 月 23 日	一般公開日	整理券取得者	300
5 月 18 日	見学･取材	富士通主催海外メディアツアー	30
5 月 28 日	見学	第 4 回 wacode（わこうど）参加者	20
6 月 3 日	見学	木更津工業高等専門学校	40
6 月 7 日	見学	山梨県立吉田高校	40
6 月 10 日	見学	公募件学会	16
6 月 21 日	見学	みよし市立三好丘中学校	13
7 月 14 日	見学	埼玉栄高校	45
7 月 19 日	見学	エネルギーハーベスティングコンソーシアム	20
7 月 28 日	見学	東京商工会議所	20
8 月 1 日	見学	Phillips Exeter Academy	2
8 月 3 日	見学	シリコンスタジオ	9
8 月 3 日	見学	東京商工会議所主催中学生団体	25
8 月 4 日	見学	宇都宮短大附属高校	28
8 月 8 日	見学	昌平高校	34
9 月 8 日	見学	群馬県立高崎高等学校	45
9 月 14 日	見学	群馬県立富岡高等学校	26
9 月 30 日	見学	宝仙学園中学校	13

10 月 5 日	見学	広島県立大門高等学校	15
10 月 12 日	見学	石川県立七尾高等学校	45
10 月 14 日	見学	公募見学会	20
10 月 28 日	見学	精密工学会画像応用委員会	20
11 月 1 日	見学	科学技術者フォーラム	40
11 月 8 日	研修	裁判官研修	2
11 月 29 日	見学	栃木県立宇都宮高校	20
12 月 8 日	見学	香川県立観音寺第一高校	35
12 月 9 日	見学	公募見学会	20
12 月 16 日	見学	海城中学	30
		合計	973

4.2　学会での展示

2016 年 11 月 14 日から 18 日まで米国ユタ州ソルトレイクシティで開催されたハイパフォーマンスコンピューティングに関する国際会議である SC16（The International Conference for High Performance Computing 2015）展示会場にて計算科学研究機構および GRAPE プロジェクトと共同で理研ブースを設け、新システム HOKUSAI GreatWave の説明および運動計測に関する映像展示を行い各国から集まった多くの研究者、技術者と情報交換を行った。

4.3　利用者向けウェブサイトの運用

平成 27 年度に刷新した情報基盤センターウェブサイトをより見やすくなるよう手を加えながら運用を行った。

今後もセキュリティ対策をより強化しつつ効率よく必要な情報が得られるウェブサイトになるよう心がけたい。

5　問い合わせ対応件数

4 月から徐々に問い合わせなどが減っていくのは見込み通りであった。特に大きな問題もなく予定通りの運用ができていた効果と考える。2 月は年度の報告書提出や次年度の課題申請があるため対応のメールが増えたが 3 月には収束するという例年通りの問い合わせ数の推移傾向であった。年度末の報告書については督促を自動化したため質問対応および不備の修正依頼のやり取りが大多数を占める。可能な限りピークを平準化できるよう報告書様式の改善などで疑問や要修正箇所を減らせるよう今度も努力し

たい。