5.4 中央計算機システム

飯田 好美(*)、村上 晃一、佐々木 節、鈴木 次郎、高瀬 亘

5.4.1 中央計算機システムの概要

2012 年に導入した中央計算機システム(新 KEKCC)は 2011/9/9 に開札が行われ、IBM より調達することとなった。新 KEKCC のデータ解析システムは、表 1 に示すサーバにより構成されている。

表 1 新 KEKCC データ解析システムの構成

磁気ディスクシステム	IBM GPFS (General Parallel File System)
	DDN SFA10000 6 台 (3.84PB:GPFS, 3.07PB:GHI)
	IBM System x3650 M3 32 台(GPFS/GHI サーバ)
HSM システム	HPSS (High Performance Storage System)
	IBM System Storage TS3500 (実効総容量 16PB)
	IBM System Storage TS1140 60 台 (テープドライブ)
ワーク・計算サーバ	IBM System x iDataPlex dx360 M3 340 台(4080 コア)
	Intel Xeon X5670×2, SL5.5
	4GB/core 282 台, 8GB/core 58 台
並列サーバ	IBM System x3550 M3 8 台
	Intel Xeon X5690×2, 48GB/node
gLite システム	IBM System x3550 M3 30 台
iRODS システム	IBM System x3650 M3 4 台, x3550 M3 2 台
NAREGI システム	IBM System x3550 M3 7 台

1) ストレージシステム

磁気ディスクシステムは、複数ノードからアクセスが可能な分散共有ファイルシステムである GPFS により、ワーク/並列/計算サーバでのユーザのホーム領域、グループ領域、ライブラリ領域を提供している。研究グループのホストから磁気ディスク領域へのアクセスは、CIFS、NFS v4 にてアクセス可能である。ただし、NFSv4 についてはセキュリティ上注意が必要なので、特別な理由がある場合に限り利用を認める。

HSM システムは、実験データ等の大容量のデータを保存するための領域であり、GHI ファイルシステムと HPSS により実装されている。GHI ファイルシステムは磁気ディスクで使用している GPFS と HPSS とを連携するインターフェイスを搭載しているファイ

ルシステムで、POSIX 標準 API により通常のファイルシステムと同様に操作可能である。また、GHI ファイルシステムに書き込まれたファイルは全て HPSS システムへデータが移動される。磁気テープとしては 3592 テープメディアを使用し、非圧縮時の実効総容量は 16PB である。

2) ワーク・計算サーバ

ワークサーバはユーザがリモートログインして、インタラクティブな解析処理、プログラム開発および計算サーバへのジョブ投入を行う。計算サーバはバッチジョブ処理専用のサーバで、ワークサーバ及び Grid 環境からジョブスケジューラを介してジョブが投入される。ジョブ管理ソフトウェアとしては Platform Computing LSF を導入する。ワークサーバ及び計算サーバは、ストレージシステムとの間を InfiniBand 4xQDR で接続する。

ワークサーバ、計算サーバの一部では、インターナルクラウドサービスを展開し、多様化するユーザグループのニーズに対応する環境を提供する。クラウドソフトウェアとしては、Platform Computing ISF を導入する。

3) Grid システム

Grid ミドルウェアとして gLite、NAREGI 及び iRODS を稼動しており、Grid 環境に必要な電子認証局の運用も行っている。gLite-UI コンポーネントはワークサーバにも導入されており、ワークサーバから gLite のジョブを計算サーバに投入可能である。また、gLite、NAREGI 及び iRODS のデータストレージとして KEKCC のストレージシステムが利用可能である。

4) その他

新システムより、夏場の電力事情による電気使用量制限や地震災害時における運用継続性を保つため、システムの自動運転、自動停止の機能を実装した。また、省エネ対策の一環として「電力見える化システム」の導入と消費電力キャップの設定を可能なシステムとした。

システムの自動運転、自動停止については、3 台の UPS 監視に関わる管理サーバを導入し、システムの自動停止、起動の制御を行う。計画停電や有事の際には、UPS からの信号を受けてサーバをシャットダウンする shell プログラムを管理サーバ上で開始する。電源復旧後、システムへの電源投入を管理サーバ経由で運用担当者が手動にて行う。

「電力見える化システム」として、IBM System Director サーバ、Active Energy Manager、クランプ式センサ監視用サーバにより、システムの電力消費量を一元的に監視し、電力消費のレポート機能を有する。また、計算サーバにおいて消費電力キャップを設定でき、設定値内で自動的に運用可能である。

5.4.2 新 KEKCC の導入

新システム導入のための入札関連作業として、2011 年度は仕様確定のために計 7 回の

仕様策定委員会を開催し、2011/7/6 に入札説明会を実施した。その後、3 回に渡って技術審査会を開催し、提案の評価を行った。新システムは日本 IBM が落札し、設計は日本 IBM の担当者と計算科学センターの担当者で行った。作業分担と計算科学センター側の担当者は以下の通りである（ただし、＊は計算科学センター外部者）。

データ解析システム

ファイルシステム	飯田好美、村上晃一、高瀬亘、佐々木節
ワーク・計算・並列	村上晃一、鈴木次郎、飯田好美
GRID	岩井剛、松永浩之、高瀬亘、飯田好美、佐々木節
運用	飯田好美、高瀬亘
電子メールシステム	橋本清治、押久保智子
Web システム	柴田章博、八代茂夫、飯田好美、村上晃一
ネットワーク	鈴木聡

その他のシステム

機構 Web	藤本順平＊
文献情報サーバ	鴨志田美由喜＊
ソフトウェアサーバ	柿原春美
JACoW SPMS	古川和朗＊
PC 端末	濁川和幸＊
設置、工事	柿原春美、飯田好美、村上晃一、金子敏明、中村貞次

設計に関するシステム全体の打ち合わせは 2011/9/28 のキックオフミーティングから始まり、2011/10/12 から 2012/3/21 まで合計 11 回、全体進捗報告会を行った。それ以外に、システム毎に担当者間で数回から十数回の打ち合わせと、システム毎のメーリングリストによる密なコンタクトが持たれた。

システムの移行作業はユーザの移行とデータ移行があり、ユーザの移行に関しては、ワーク・アクセス・CIFS サーバのアカウントと所属グループの確認、サブグループの再編成などを行った。B 計算機のユーザについては、Belle 実験グループの管理者に移行ユーザのリストアップを依頼し、KEKCC との重複ユーザについてはグループ管理者もしくはエンドユーザ本人と調整しプライマリグループを設定した。パスワードについては、B 計算機、旧 KEKCC ともパスワードの認証方式、管理方法が異なっているため移行することができず、全アカウントのパスワードを新規設定することとなった。パスワードの配布については、サブグループ毎に管理者もしくは Belle 秘書室から配布を依頼した。

データの移行については基本的にはセンターで一括して実行した。HPSS については、2012/1/4 にサービスを停止し、メタデータの移行、テープカートリッジの移動などを行った。1/30 からは T2K と HAD グループに HPSS のサービスを再開したが、再開後に書き込んだデータについてはユーザ自身で新システムへ移行してもらうこととなった。

HPSS のデータを GHI ファイルシステムで読むためのデータ変換については旧 KEKCC の領域を 2012/2/17 から開始して 2012/4/11 に完了し、旧 B 計算機の領域は 2012/4/17 から開始して 2012/7/20 完了の予定である。磁気ディスクについては、2012/1/7 からオンライン移行を行い、2/22-27 の運用停止期間に最終的なオフライン移行を実施した。B 計算機の磁気ディスク領域については、2012/2/3 から段階的にオフライン移行を実施し、移行終了した領域は再度ユーザへ開放された。解放後は 2/22 の運用終了までオンラインで差分の移行を実施した。

システム移行に伴う 2012/2/22 から 2012/4/2 までの運用停止期間中も、計算資源、ストレージ資源を必要とするユーザのために、2012/2/28 から 2012/4/2 までの間、旧 KEKCC を暫定システムとして再稼動した。暫定システムで書き込んだデータについては、ユーザ自身でデータ移行を行うこととした。また、2012/3/1 から新 KEKCC を仮運用システムとして運用を開始し、予めサブグループ管理者に利用希望ユーザのとりまとめとパスワードの配布を依頼した。

エンドユーザに対する新システムの説明会は 2 回開催された。新中央計算機システム説明は 2011/12/5 に小林記念ホールで開催され、データ解析システム、メールシステム、Web システムなどシステム全体の説明を行い、約 40 名の出席があった。2012/2/2 には 4 号館セミナーホールでデータ解析システム説明会を開催し、新 KEKCC のデータ解析システムとシステム移行概要の説明を行い、約 30 名が参加した。