

5.2 グリッドシステム

飯田好美、岩井剛、河井裕、佐々木節、鈴木聡、高瀬亘、松永浩之、渡瀬芳行

5.2.1 概要

計算科学センターではグリッドコンピューティングの仕組みを利用し、本機構と共同研究を行う研究機関との間で、データ及び計算資源を共有するためにグリッドシステムを構築し、運用を行っている。

混乱を避けるため、先に本節で用いられるグリッドシステムという用語に対する解釈を与える。図 1 は実際にインストールされた計算科学センターに於いて運用中の要素をサービスに着目し、まとめたものである。図 1 に示されるように多くの場合、グリッドに於けるサービスとは一つ以上の小サービスがお互いに作用することにより形成され、複数のサービスにより、グリッドに於ける管理単位に相当する「サイト」が形成される。これはグリッドの「複数機関に跨った計算基盤を利用する」という設計思想に基づく。このサイトを形成・構築するためのソフトウェアを特に区別してミドルウェアと呼ぶが、計算科学センターでは複数のミドルウェア (gLite、NAREGI、iRODS 等) を使用し、複数サイトを運用している。本節では計算科学センターで運用される、これら複数のサイトを全て含め「グリッドシステム」として扱うこととする。

次節以降でグリッドシステムを構成しているサイトの利用状況と、これらにより提供される計算基盤上での 2011 年度活動内容について報告する。

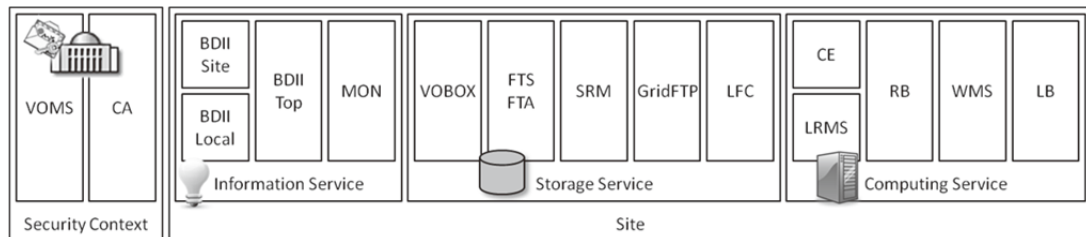


図 1 計算科学センターにて実運用中の Grid 計算基盤の構成。実際には複数のミドルウェア (NAREGI, iRODS, gLite 等) により複数のサイトが構成されているが、混乱を避けるため、図中には gLite ミドルウェアにより構成された LCG (LHC Computing Grid) 計算基盤のみを示した。

5.2.2 KEK Grid CA (Certificate Authority)

KEK Grid CA は世界に 80 程度ある IGTF¹の策定する規格に準拠・承認された認証局の一つである。2006 年 2 月に正式運用を開始して以来、主に国内のグリッド利用者、及びサービス提供ホストに対して電子証明書を発行している。

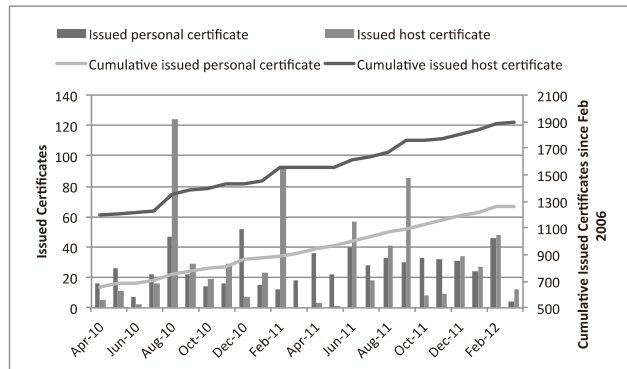


図 2 2010-2011 年度の電子証明書発行部数

図 2 に KEK Grid CA が 2010 年度、および 2011 年度に発行した個人証明書とホスト証明書の発行部数を示す。2011 年度は 359 部 (267 部 JFY2010) の個人証明書と 344 部 (358 部 JFY2010) のホスト証明書が発行され、2006 年の認証局運用開始以来の累計発行部数はそれぞれ 1266 部、1897 部に達した。

5.2.3 LCG 計算基盤

JP-KEK-CRC-01 及び JP-KEK-CRC-02 (以降 KEK-1 及び KEK-2 と表記) は EGEE²が開発を推進した gLite ミドルウェアにより構築された計算資源を提供するサイトである。約 50 カ国・200 機関から 300 程度のサイトが参加することによって、LCG と呼ばれる全世界規模の計算基盤が展開され、2011 年度中、当センターではそのうちの 2 サイトである KEK-1 及び KEK-2 が運用中であった。

KEK-1 は研究開発用途に、KEK-2 は KEK-1 で培われた経験に基づき、より安定したサービスを提供するために分離して運用がなされている。以降は特に注釈のない限り、KEK-2 のみに関して報告する。

サービスの品質

図 3 に 2010 年度、および 2011 年度中に発生したダウンタイム、およびサービス稼働率を示す。2011 年度は 1200 時間程度 (200 時間 JFY2010) のダウンタイムが発生し、システムの年間稼働率は 73% (88% JFY2010) であった。これは 3 月に発生した東日本大震災によるシステム全停止と 8 月の計画停電によるシステム全停止、および 2011 年 1 月から始められた新システム移行への影響に依るものである。LCG に於いて定められた月間最小サービス稼働率は 70% に設定されており³、これは図 3 に見ることが出来るように概ね達成された。また、2011 年度末は新システムへの移行作業により、非常に長いダウンタイムが設定され、結果として稼働率が下がっているが、全停止に依るものではなく、縮退運転に依るもので

¹ The International Grid Trust Federation: <http://www.gridpma.org/>

² Enabling Grids for E-science: <http://public.eu-egee.org/>

³ 算出式を含む完全な資料は <https://edms.cern.ch/document/860386/> から取得可能。

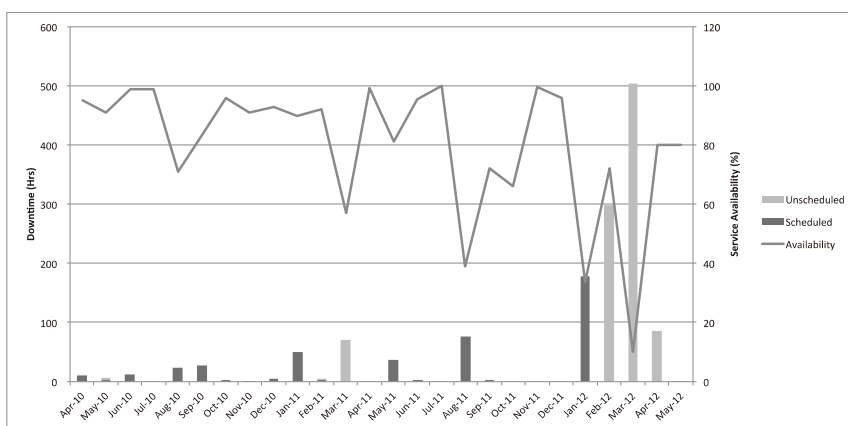


図 3 2010-2011 年度に発生したダウンタイムとサービス稼働率

ある。

利用者とシステム管理者が幅広く分散する LCG インフラではコミュニケーションチャンネルのひとつとして KIT (Karlsruhe Institute Technology) が主導する GGUS (The Global Grid User Support) と呼ばれるチケットングシステムが利用される。全体では年間 1 万件以上のチケットが発行されるが、2011 年度中に当センターに関連したチケットは 20 件 (45 件 JFY2010) であった。ほとんどのチケットは 24 時間以内に初期対応が行われ、1 週間以内でチケットはクローズされた。

利用状況

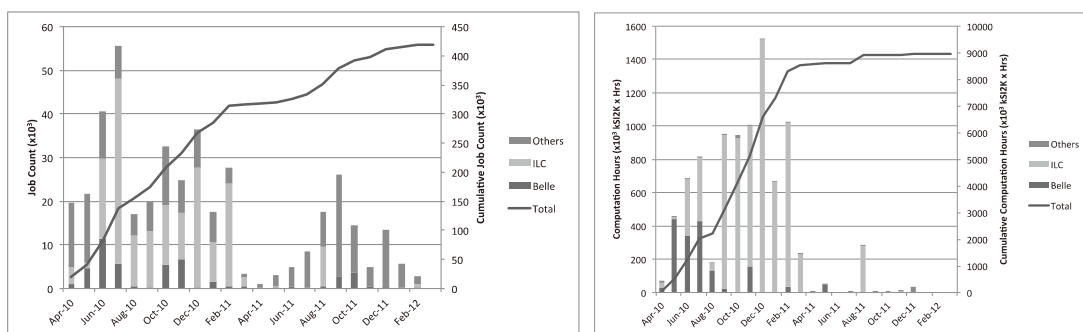


図 4 2010-2011 年度のジョブ処理数 (左) と CPU 消費時間 (右)

図 4 に 2010 年度、および 2011 年度中に投入されたジョブ数と CPU 消費時間⁴を示した。投入ジョブの総数は 102,819 (316,756 JFY2010) で 428,012CPU 時間 (8,550,869CPU 時間 JFY2010) であった。

ストレージサービスは共通情報システム HPSS をバックエンドに VFS マウントした領域を GridFTP 及び SRM プロトコルでサービスを提供している。記憶領域の増設については、

⁴ SPEC (Standard Performance Evaluation Corporation) が策定したシステムの性能評価を行うベンチマークスコアにより規格化した CPU 消費時間。高エネルギー実験分野で使用されるプログラムは SPEC Int にスケールすることが多いため、ここでは 1000 Spec Int 2000 で CPU 消費時間を規格化している。

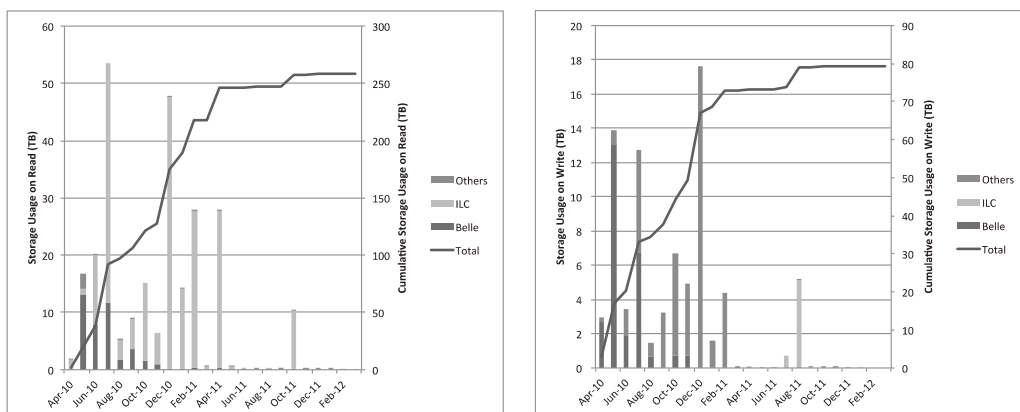


図5 2010-2011年度ストレージ利用量

グループ毎にテープカートリッジを購入してもらおう方針で運用している。2009年後期までは現行の共通情報システム導入時に割り当てた初期状態の領域を各V0で共有していたが、この領域がBelleとILCの要求容量には見合わないため、テープを購入してもらい独立した領域を割り当てている。2011年度は73TB(218TB JFY2010)の読み出しと6TB(40TB JFY2010)の書き込みが行われた。(図5)

2010年度に比べて著しく利用率が低下した原因は「5.2.5 震災の影響」でも触れるが、2011年3月の地震以降、電力使用制限が課せられたため、平時の25%程度に資源規模を縮退して運転した事に依る。

5.2.4 ユーザー対応・他

グリッドにより構成される計算基盤の運転維持には多くの人的資源が必要とされるため、大学研究室の規模で継続的に運用をまかなうことは困難である。当センターでは2007年3月に国内の共同研究機関(東北大・筑波大・名古屋大・神戸大・広工大)にシステム一式を構築し、拠点間を遠隔相互診断するための計算基盤としてKEKを含む6拠点間で科学的目標に依存しない加速器科学仮想組織を整備した。2011年度は富山高専を新たな支援対象に加え、支援作業を行った。

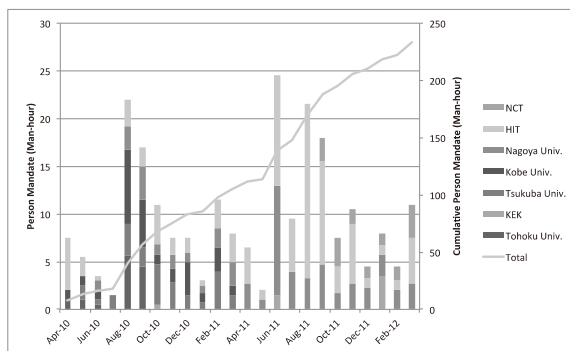


図6 2010-2011年度に実施された加速器科学仮想組織に於ける運用支援作業時間

各拠点のソフトウェア・アップデート、QA対応といった運用支援も継続的に行われている。2011年度は128人・時間(106人・時間 JFY2010)の運用支援が各拠点に対して行われた。(図6)

機構内の研究グループ(ILCとBELLE)への支援も継続的に行われている。定期的にミーティングを開催し、性能向上・環境改善に役立てられた。

5.2.5 震災の影響

2011 年中のグリッドシステムにて利用可能な計算資源量を図7に示した。2011年3月11日に発生した東日本大震災の影響により、地震発生以後、約2週間にわたってシステムが完全停止状態となった。グリッドの性質上、1サイトがダウンしていても、世界に分布したグリッド全体に深刻な影響を及ぼすことはないが、当センターでは Belle V0 向けのクリティカル・サービスをいくつかホストしているため、復旧が急がれた。

地震発生から約2週間経過した3月25日、V0の構成に必須となるサービス (VOMS, VOMRS, LFC, site BDII) の再稼働を開始した。以降は逐次、電気容量の許す範囲で資源割り当てを行い、4月末には地震発生前の25%程度の規模でシステムが安定的に動作することが確認され、以降、規模縮退した状態で運転が継続された。

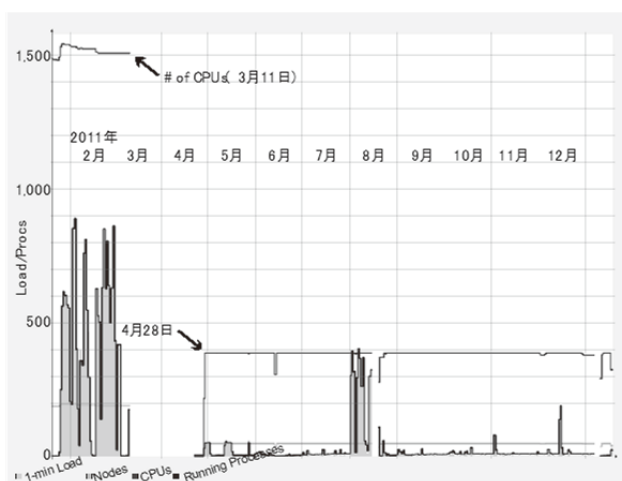


図7 2011年3月11日に発生した東日本大震災以降の計算資源量推移

論文

なし

国際会議プロシーディングス

- [1] Y. Kawai, G. Iwai, W. Takase, T. Sasaki, and Y. Watase, “Universal Grid User Interface (UGI) for Multiple Grids and Cloud”, PoS – Proceedings of Science, 2012. (査読有)
- [2] Y. Kawai, A. Hasan, G. Iwai, T. Sasaki, and Y. Watase, “A method for reliably managing files with RNS in multi Data Grids”, Procedia Computer Science, Vol. 4, No. 1, pp. 412-421, 2011. (査読有)
- [3] Y. Kawai, G. Iwai, T. Sasaki, and Y. Watase, “Managing distributed files with RNS in heterogeneous Data Grids”, The 11th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing (CCGrid) 2011, Vol. 1, No. 1, pp. 494-503, 2011. (査読有)
- [4] Y. Kawai, G. Iwai, T. Sasaki, and Y. Watase, “SAGA-based application to use resources on different Grids”, PoS – Proceedings of Science, Vol. 1, No. 1, 2011. (査読有)
- [5] G. Iwai, Y. Kawai, T. Sasaki, and Y. Watase, “A development of lightweight

Grid interface” , Journal of Physics: Conference Series (JPCS), Vol. 331, No. 6, pp. 062035, 2011. (査読有)

会議発表

- [6] Go Iwai. “Asia-PoP : A proposal of the high throughput data sharing testbed” , AFAD 2012, Feb 2012. (100 名)
- [7] Go Iwai. “GRID status and future” , ILC Tokushin - Kickoff Meeting, Sep 2011. (50 名)

その他

- [8] “UGI - Universal Grid Interface” , Research exhibition, SC11, Seattle, WA, US, Nov 2011.