

## Computing RRB 4月12日 9:00

<https://indico.cern.ch/conferenceOtherViews.py?confId=128046&view=lhcrb&showDate=all&showSession=3&detailLevel=contribution>

前回2010年10月のミーティングの議事録の承認

### Status of the WLCG project I. Bird

2010年10月から2011年3月までの期間の報告。この期間は、ppランの最後の時期で、加速器のルミノシティがどんどん増加した時期と、重イオンのデータを数週間取った時期を含む。12月から3月の加速器が止まっていた時期も、WLCGは大変忙しく、全ての実験が2010の全データを再処理をした。

#### CERN T0 :

9月にCMSとALICEがそれまでの予定を超えるT0のリソースを要求した。ALICEは約2倍のデータ量（元々1.25GB/sのところ2.5GB/s）にして重イオンのデータを出来るだけ沢山とりたいということ。CMSは重イオンのデータをzero suppression無しで取ってオフラインでそれをやるという要望。そのためCMSではデータを小さくする処理が完了するまではT1にデータを分配することができず、一時的にコピーが存在しない状況になる。したがって、ALICEとCMSに十分なディスク容量を割り当ててT1にコピーをとるまで一時的なバックアップをCERNに作るようにした。このことは、T0の2011年用のリソースが既にインストールされていたため（たまたま）可能だった。ATLASは重イオンのデータもppと同じレートで取り、特別なことは必要なかった。

T0のストレージ (CASTOR)は実験からの大量のデータを問題なく受け取ることができた。テープに書くデータのピークは220TB/日に達し、データを受け取るスピードは最大11.5GB/sで、読み出しは25GB/sだった。これは実に当初の目標を上回るもので、システムはこれを問題なくこなした。ppランの期間、2PB/月の割合でデータがT0にストアされ、重イオンのときは4PB/月だった。

#### データ転送 :

LHCOPN上のトラフィックは平均13GB/s程度で、ピーク時には20GB/sを超えた。

#### 実験のリソースと負荷 :

2010年のランの間、T1とT2のリソースはちゃんと使われ、年の終わりに向かっては100%使用率は珍しくない。

#### WLCGサービス状況 :

大規模なサービス中断は報告してフォローアップすることになっている。これはトラブルの経験とそれから学んだことを各サイトで共有するためである。23のインシデントのうち、11はインフラストラクチャー（電源や空調など）に関係するもの、6はデータベースの問題、4はストレージに関係するもので、2つがネットワークの問題だった。これをみると分散システムであることに起因する問題は少ないようだ。CERNでの大規模な停電は12月18日に起こり、これはクリスマス休日の最初の日だった。電源がバックアップされている重要なサービスは維持されたが、優先度の低いサービスのいくつかは復旧に24時間かかった。ATLASのデータベースは数日間トラブルが続いた。

#### サイトの信頼性 :

この6ヶ月の信頼性のデータをみると、T1と大きなT2（リソースの大部分を提供している）については大変安定で、高いレベルである。いくつかの小規模なT2についてはまだ改善の余地があるようだ。

#### 計画 :

2011年のリソースの大部分は、2010年におこなわれた需要予測にもとづいてすでに設置されている。各実験は、2010年の経験にもとづいて新しい予測をおこなった。2012年と2013年に関しては、新しいLHCのスケジュールでは2012年を通してデータを取り、2013年にシャットダウンということになった。LHCのルミノシティ増強の方法は、パンチ間隔は50nsまでで、そのかわり大強度のパンチを強く絞ってぶつけるというものである。そのためイベントのパイルアップが大きくなり、その結果データ量が大きくなり、またデータ処理に必要なCPUの需要も大きくなる。このようなことが背景にあって、各実験のリソース要求量は一般に大きくなっている。更に2012年にもデータを取る事も理由となる。2013にも資源増加が必要なのは、データ解析やデータ処理のための計算機需要のためである。

- T0 - 2012,13年に続いて、2014年にも大幅なT0リソースの需要が見込まれる。CERN内部での増強は限界があるので、外部にもデータセンターを構築する構想にたいして、昨年11月までに28のオファーを受け取った。いくつかのサイト訪問や議論をおこなったが、正式な手続きにはまだ入っていない。

- ネットワーク - 昨年の6月に、特にT2、T3においてデータアクセスの信頼性や性能をどのように改善するかを議論するワークショップが開かれた。その結論の一つは、既存のネットワーク環境の範囲内で実験グループがデータのマネージメントの方法を改良する努力を続けるとともに、少なくともT2においてはネットワーク資源が十分与えられて、現在の階層構造のネットワーク構成ではなく、任意のサイトから任意のサイトへのデータアクセスが可能になるようなアーキテクチャが重要である。LHCOPNグループ

は次の数年間の実験の要求を理解し、それを満たすようなネットワーク構成を提案するタスクを担う。最初の提案が提出されて現在ネットワーク機関との間で議論中である。提案は、LHCONE(LHC Open Network Environment)という名前で、これによると、他の分野のためのトラフィックと分離しつつT2やT3が任意のT1やT2からデータを持ってくる事ができるようなる計画である。

Report from the C-RSG E. Espriu

=====

詳細なレポートが提出された。2012, 2013年の実験からの要求の精査はまだ途中段階で、秋のRRBで更新する。そのため詳細な情報を、実験グループは9月1日までに提出する必要がある。

2010年の全体的なWLCGリソース利用実績：

2010年の初期には、存在するリソースが実験の必要量を上回り、余裕があった（2010年の間にルミノシティが5桁増大したことに注意）。実験グループは、これをモンテカルロの量を増やしたり、データの再処理を頻繁におこなったり、T1, T2に配備するデータを増やしてアクセスを容易にするなどに活用した。このことは、なにかと手探り状態の実験初期には役に立つ。特にT2は良く使われた。これに対してT1の使用は少なめだったが、実験グループは仕事の割り振りを調整して状況は改善した。年の終わりから2011の初めにかけて状況は変わり、2011年の承認されたリソースにマッチするためには実験のコンピューティングモデルのパラメータを調整する必要があった。これは、2011年にはルミノシティが更に増加すること、また、その増加は、強いパンチを強く絞ってルミノシティを稼ぐという方法であるため、衝突あたりの反応数が大きくなり、事象のパイルアップが増加するためである。このため、事象あたりのデータ量が増え、また処理に必要なCPUも増加する。

ATLASとCMSはCERNとT1のCPUを然るべき割合で使った。ALICEとLHCbはCERNのリソースへの依存が大きかった。2010年秋の精査で、T1の使用量が少ないことを指摘したが、ATLASとCMSをモンテカルロ生成をより多くT1でやるように調整するなどして状況を改善した。Funding agenciesが準備したリソースをちゃんと使う事が重要である。

数字でみると、T2のCPU使用率は大変高い。2010年10月の利用率は92%。2011年4月の使用量はpledgeに対して122%となっている。つまり、約束したものよりも多くのリソースが実際に稼働している、あるいは状況に応じて他のリソースをWLCGに回したりした結果だろう。また、利用の効率も期待してたものより高い。効率に関しては、もともとT2での多数のユーザーによるランダムな利用においては効率は60%だと仮定し、要求量の基にしていたが、実績はもっと高かった。これに基づいて、今後はT2の効率を66%と仮定することを支持する。

約束された資源量と実際に提供された資源量をみると、資源の種類やサイトの種類によるが、100%から129%の間に分布しており、大変満足のいくものである。しかし、部分的には不足しているサイトもある。オランダT1は79%しかディスクを提供できなかった。台湾T1はディスクは71%、テープは82%にとどまった。

2010年の各実験のリソース利用実績：ATLASについて

ATLASは10億イベントの事象を記録した。年の初期は測定器のコミッシュニングによりフォーカスしてしたが、すぐに物理解析の重みが増した。モンテカルロ生成や、データの再処理、またコンファレンス前の解析のピークなど、計算資源をおおいに使った。

事象再構成に必要なCPU時間は、資源要求の計算で想定した事象あたりの時間より50%多かった。これは、再構成するトラックのpT下限値をより低くしたり、そのほかコミッシュニングの段階におこなわれたいろいろな変更の結果である。

ATLASは14億事象のモンテカルロを生成した。これは、予定より多めである。これは、2010年初期の事象のシミュレーションはパイルアップが少ないので計算時間が少なめで、それに比べて十分はCPUリソースが使えたからである。

ESDのデータサイズが20%小さくなった。生データのサイズも資源要求に使った値より小さくなった。

T1のディスクはフルに使われた。データのコピーを余分に配備したり、モンテカルロデータの量が増えたことなどによる。グループやユーザーの解析のためのディスク量は全体として期待値よりも少なかった。CERN CAF（解析用ファーム）のリソースが十分あったので、T1のリソースの利用は少なめで、それによってT1での解析用のディスクの利用も少なかった。T1のテープも明らかに必要量を過大評価していた。

2012, 2013の実験からの要求の精査：ATLASについて

ATLASとCMSの要求量は共通のレフェリーによって精査された。ATLASから提供された情報はクリアーで詳細で大変良かった。2012年のATLAS要求量は、パイルアップの増加が効いている。この影響は、データ配備戦略の変更により緩和されている。

ATLASコンピューティングモデルへの中期、長期の継続可能性についての批判に対して、ATLASはこれまでにデータサイズや処理時間、データフォーマットの種類やその配備計画などについて改善をおこなってきた。CRSGは今後もこのような努力が継続することを期待する。

しかし、ATLASの要求量はやはり大きくて、2012年にLHCシャットダウンがおこなわれると想定されていたときの2013年の要求量（14 TeVでの運転）に匹敵する量を2012年に要求している。CRSGはATLASに次のことを検討するよう奨める。

- ディスクとCPUを制限するために、より中央管理なデータの準備をおこなうこと
- 2013年のCERN CPUの利用を2012年並みに押さえる事。

また、T2のディスクの使用実績は割当より少なめなので、T2ディスク要求が必ずしも正当化されてないのではないと思う。

ATLAS（CMSも）当初の想定よりもたくさんデータを記録する計画である。これは新物理への感度を低いエネルギー領域でも維持するためであり、2011-12年の物理ゴールを達成するためにLHCCからサポートされている。しかし、これを限られた計算機資源のなかで実現するためにはコンピューティングモデルの更なる見直しが必要だろう。

リソースの増加を緩和するために：

実験グループは、これまでの経験を基にして、長期的に維持可能なコンピューティングモデルを探ることが必要である。パイプアップの増加（その結果としてのデータサイズの増加、計算時間の増加）などによる必要資源の更なる増加を防ぐためのいんな最適化が必要だろう。

- T1, T2に配置するコピーの数の縮小
  - シミュレーションの最適化
  - いろいろなデータフォーマットの役割の見直し
  - T1, T2の使い方の最適化
- など。

各実験への勧告と要求：

- WLOGGのリソースの割当をできるだけ使わなければならない。実験によっては、CERNに強く依存しすぎて、T1のリソースを十分利用していない。
- 一方、効率の観点から、CERNのリソースは、もし割当が十分使われていない場合は、それを他の実験の低いプライオリティのジョブなどでシェアすることを奨める。
- ネットワーク帯域の改善とデータのダイナミックな配備にもとづく最適なリソースの利用について検討すること。
- 実験グループは、パイプアップの影響を正確に評価すること。10月のRRBではその結果を期待する。
- 現在60%と想定されているT2の効率パラメータを66%に変えることを勧告する。これはfunding agencyにとって10%の節約に対応する。いろいろな使い方モードの影響の正確な評価をし、10月のRRBで最終的に決定する。
- T2の利用収支レポートは、改善しているが、まだ不十分である。

Status of resources and financial plan S. Foffano

MoUの状況：

前回のRRB以降の新しいMoUはないが、現在2つのMoUが署名にむけて準備中である：ギリシャ（CMS T2）とアメリカLBNL（ALICE T2）。その他の国からのEoIや、既存のメンバーから内容の変更の申請なども審査中である。

前回のRRBでの要請をうけて、MoU annex 1,2に書かれているいくつかのサイトについて、サイト名、略称、責任者の名前などが更新された。この内容は常に更新される必要があると改めて要請された。

CERNのWLOGG予算：

前回のRRBでは、例外的に、CERNの予算計画や執行状況の報告はなかった。これは、計画に関係する多くの変更（LHCのスケジュールや計算センター増強計画）があり、正確な情報を準備するのが困難だったからである。2010年の会計を締めた結果は：

人件費：13,578kCHF

設備：20,770kCHF

2010年夏の新しい予算カット計画は2011-2013のWLOGG計画にも影響するが、戦略を改めて、2010年には設備費を節約して2011年に繰り越し、予算の減少を補足することにして、5.9MCHFが2011年のWLOGG予算として繰り越された。資料の表3に2011-2016の予算計画が示されている。人件費は大丈夫だが、設備については不足する年がある（LHCが運転されることになった2012年など）。いくつかの注意事項としては：

- 計画は暫定的な要求量に基づくものである
- LHC加速器の運転計画は2011 Chamonix meetingの結果に基づき、2013と2016年は運転しないという想定
- T0の計画は外部サイトの計画がまだまとまらないので概算である
- 将来の設備費の増加やそのふらつきに対応するために2010-2011年でやったような予算繰り越しを含めて柔軟な計画が必要である

リソース使用の収支：

- CERNおよびT1

資源量とその使用量はLCG webで報告されている。 <http://lcg.web.cern.ch/LCG/accounts.htm>

- T2

現在ではほとんどのT2がその収支状況を報告している。いくつかのT2サイトがまだレポートができずにおり、今後も技術的なフォローをおこなう。

実験のリソース要求量とリソースの提供：

前回のRRBで、実験の要求量と資源提供量の報告をしたが、その一部は暫定的なものであった。また、いくつかのT2では2011年の資源量が2010年のものを下まわるケースが見られた。改訂版を2010年12月15日に公表した。

2011年の提供資源が2010年より少ない5つのT2については、以下の説明がされた。

- オーストリア：インフラの制限により約束した2011年の設備をインストールできないことがわかった。新しいサイトを探すなど対策を考える。

- ブラジル サンパウロ：CPUの性能を過大評価していたために2010年の約束を実は満たすことが出来なかった。2011年の資源量はそれを修正したことを反映している。

- フランス クレルモン：予算不足に対応して下方修正した。

- 韓国：2010年のディスク提供量は約束に達しなかった。2011年の量は、それを反映して修正した。その後2011年1月になって、新しくディスクを増設することができたので、2012年の提供量は大丈夫そうである。

- スペイン LHCB連合：2011年の量は、LHCBの要求量の減少に単に対応して減らしたものである。

3月にフランスから予算状況に対応して改訂したT1, T2資源量のレポートがきた。T1は6%の削減。T2では、7つのうち3つがCPU削減 (7-21%)とディスク量の削減 (5-12%)。

実験の要求量(CSRGで精査した後のもの)と最新の資源提供量を比較すると、ALICEのT1, T2資源の不足が目立つ (30-40%)。ほかの実験はだいたいバランスしている (20%以内)。

2012, 2013の資源提供量データ：

秋のRRBでは、2012年の確定値と2013年の計画を承認する。それに先立って、CSRGが承認する実験の要求量が発表される。各サイトからのインプットの締め切りは2011年9月30日である。

川本辰男 (東大ICEPP)