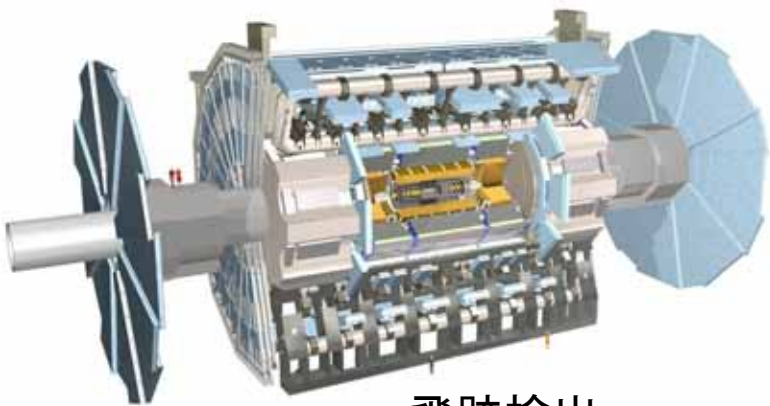
ATLASコンピューティングとグリッド

上田郁夫
東京大学素粒子物理国際研究センター

2008年3月24日
日本物理学会 シンポジウム
「LHC First Collisionに向けた実験準備」

コンピューティング

測定器



飛跡検出
粒子エネルギー測定

データ

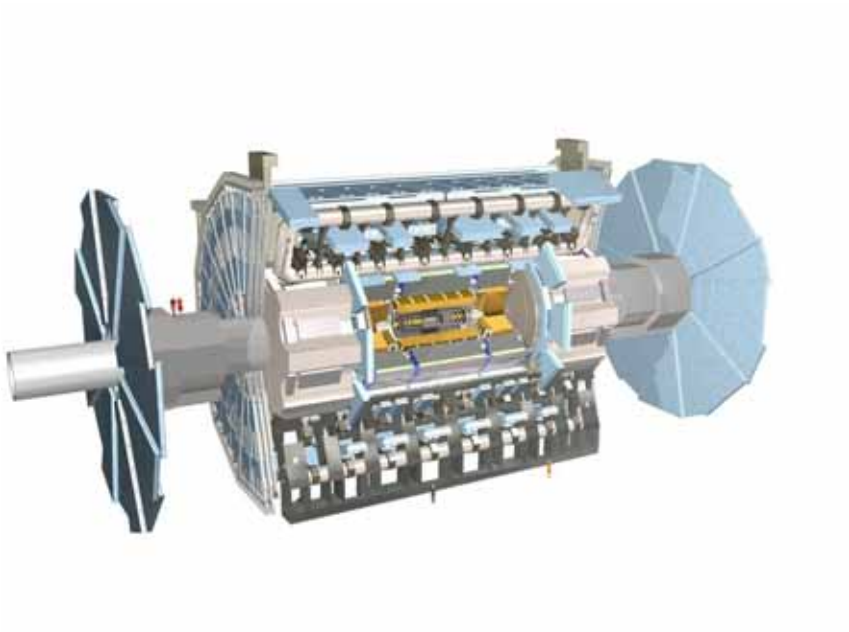
計算システム

データ処理
・保管・解析

物理事象

物理成果

ATLASのデータ量



- データサイズ:
  1.6 – 2 MB/event
- イベント数:
  200Hz、2×10^9/yr
- データ量:
  3.2 – 4 PB/yr

比較

- テープライブラリ
  400GB のテープ
  x 8000本
  ＝3.2PB

- LEP／OPAL
  ～3TB／yr

データ量が膨大

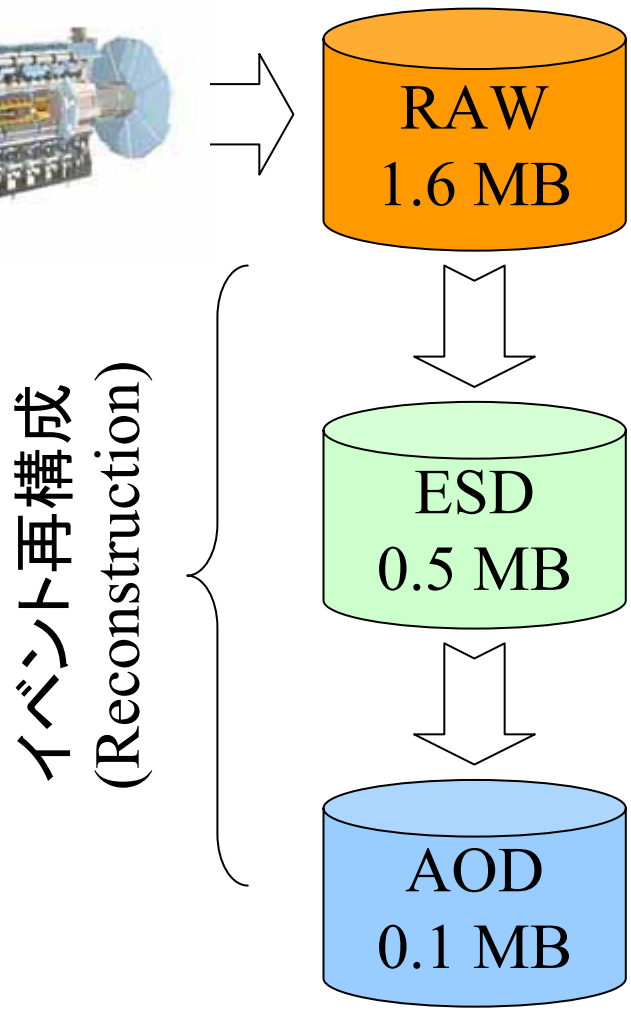
- 保管の困難
  （記録装置の容量）
- 計算量が膨大
  （計算機の台数）

→ CERN計算機センターだけでは足りない
実験Collaborationで分担
⇒「地域センター」

- テープに保存
  – アクセスの困難
- ディスク容量
- I/O負荷

⇒ サイズを絞ったデータを作る

データ階層

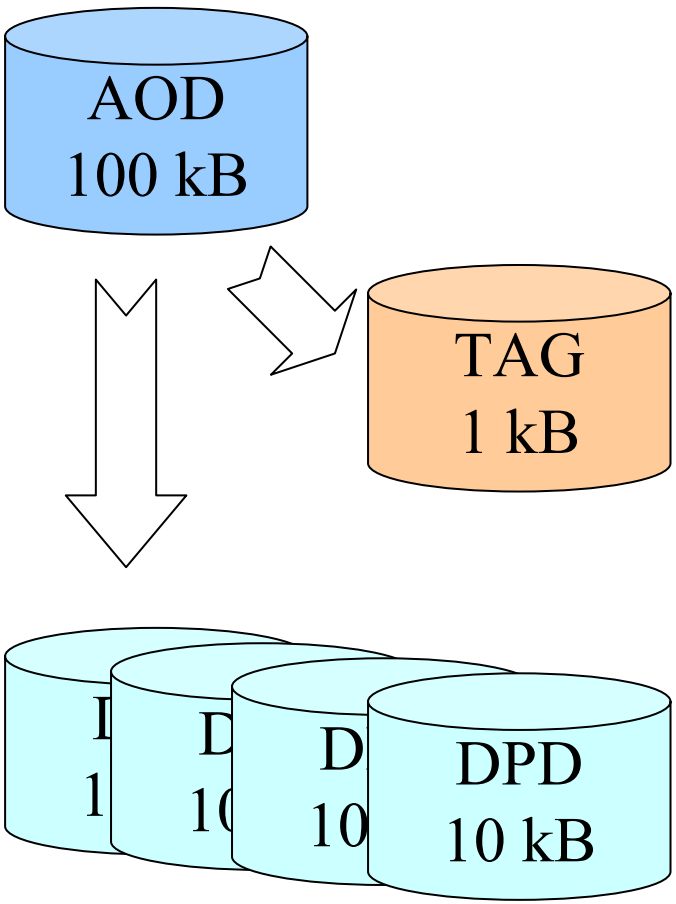


Raw data
- 測定器から読み出されたデータ
- Byte Stream

Event Summary Data
- データ較正、イベント再構成したもの
- Track, EnergyCluster, …

Analysis Object Data
- 物理解析用データ (0.1MB)
- Particle, Jet, Missing Energy, …

データ階層



Tag data

イベント選択用
データベース

Derived Physics Data

物理解析の目的毎
Histogramming

計算資源階層 (Tiers)

RAW記録

Tier0 (RAW)

RAW保管
初回Reconstruction

Tier1 (RAW, ESD, AOD, …)

各種データ保管
Reconstruction
Simulation

Tier2 (AOD, TAG, DPD)

物理解析
Simulation

Tiers & Sites

CERN
- Tier0： RAW記録・保管・送出
- CERN Analysis Facility： 物理解析、較正

地域センター (Regional Centers)
- Tier1： データ保管、Reconstruction、Simulation
  比較的大規模の計算機センター
  10 Sites
- Tier2： 物理解析、Simulation
  ある程度の資源を揃えた研究所、大学
  34 Sites
  （ATLAS日本は東大素粒子センターが Tier2 センター）

WLCG Collaboration

- Worldwide LHC Computing Grid
  - LHC実験のデータ解析に必要な計算機資源を整備するためのCollaboration
  - 必要なソフトウェア (middleware) を開発・配備
  - 各実験と各センターの協議の場
- WLCG MoU
  - 各Siteの供給資源の約束（実験毎）
  - 各Tierの役割とサービスレベルの取り決め

WLCG MoU

東京大学素粒子センターはTier2 site としてWLCGに参加

Japan, ICEPP, Tokyo	Pledged	Planned to be pledged				
	2007	2008	2009	2010	2011	2012
CPU (kSI2K)	1000	1000	1000	3000	3000	3000
Disk (Tbytes)	200	400	400	600	600	600
Nominal WAN (Mbits/sec)	2000	2000	2000	2000	2000	2000

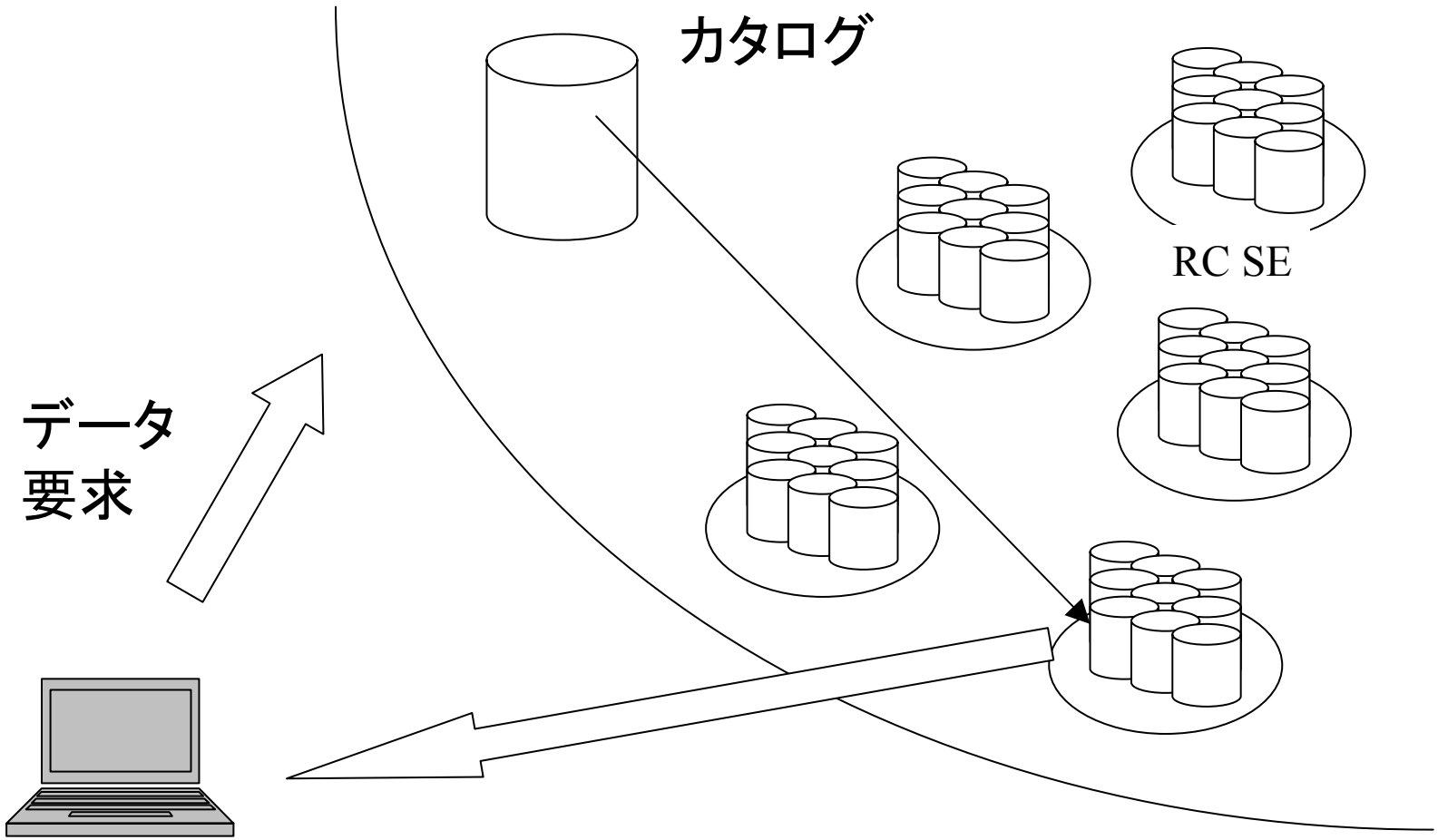
ATLAS日本の地域解析センター
東京大学素粒子センターに設置された計算機システム

グリッド

Grid：各地の計算機センターを統一的に利用出来る様にしたもの

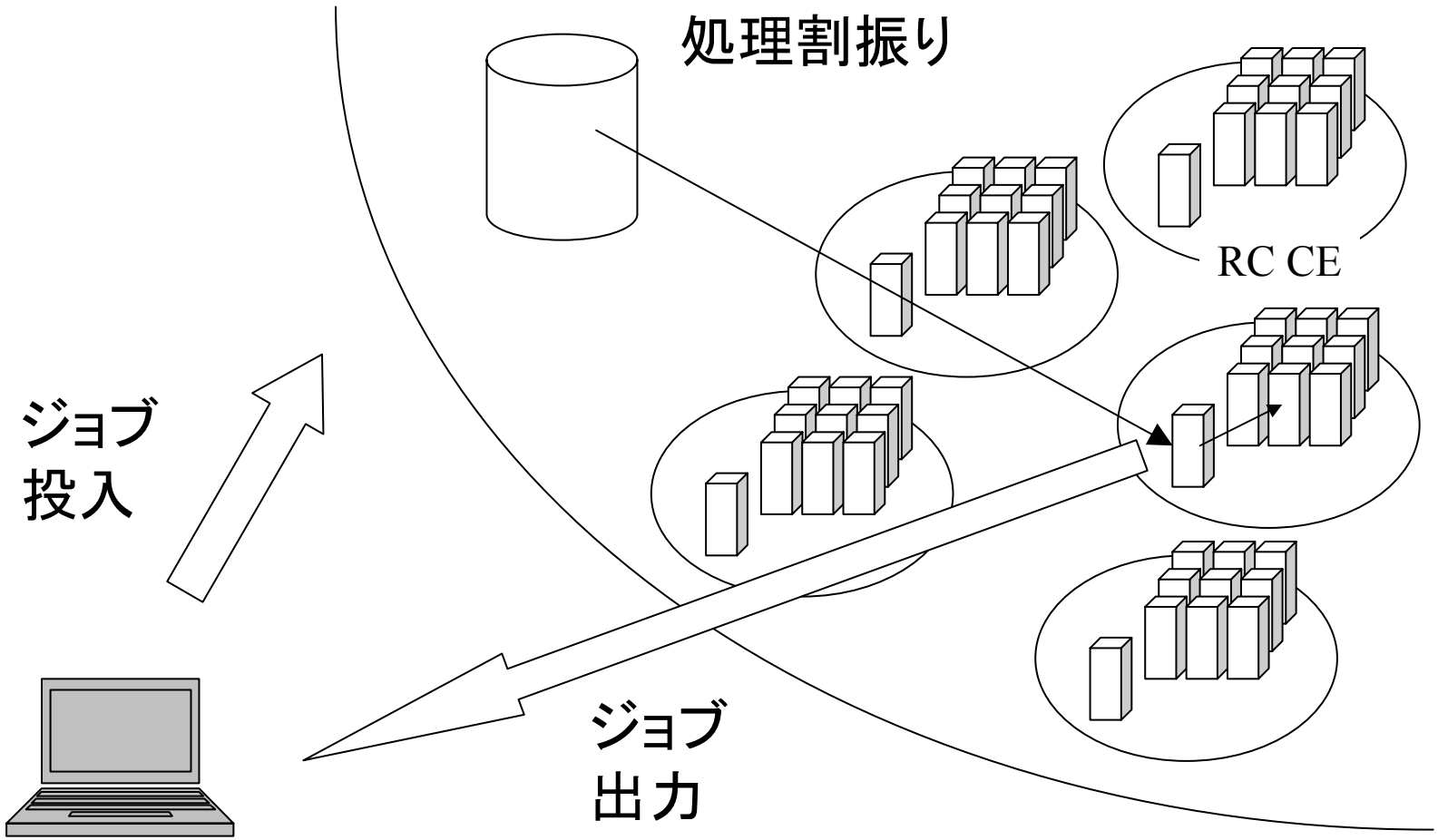
- データ管理：分散したデータを統一管理
  - データカタログ
  - データ転送
- データ処理：各センターのバッチを統合
  - どのセンターに割り振るか自動判定
    - データのある所、ジョブの少ない所

グリッド（データ管理）



SE = Storage Element

グリッド（データ処理）



CE = Computing Element (= batch system)

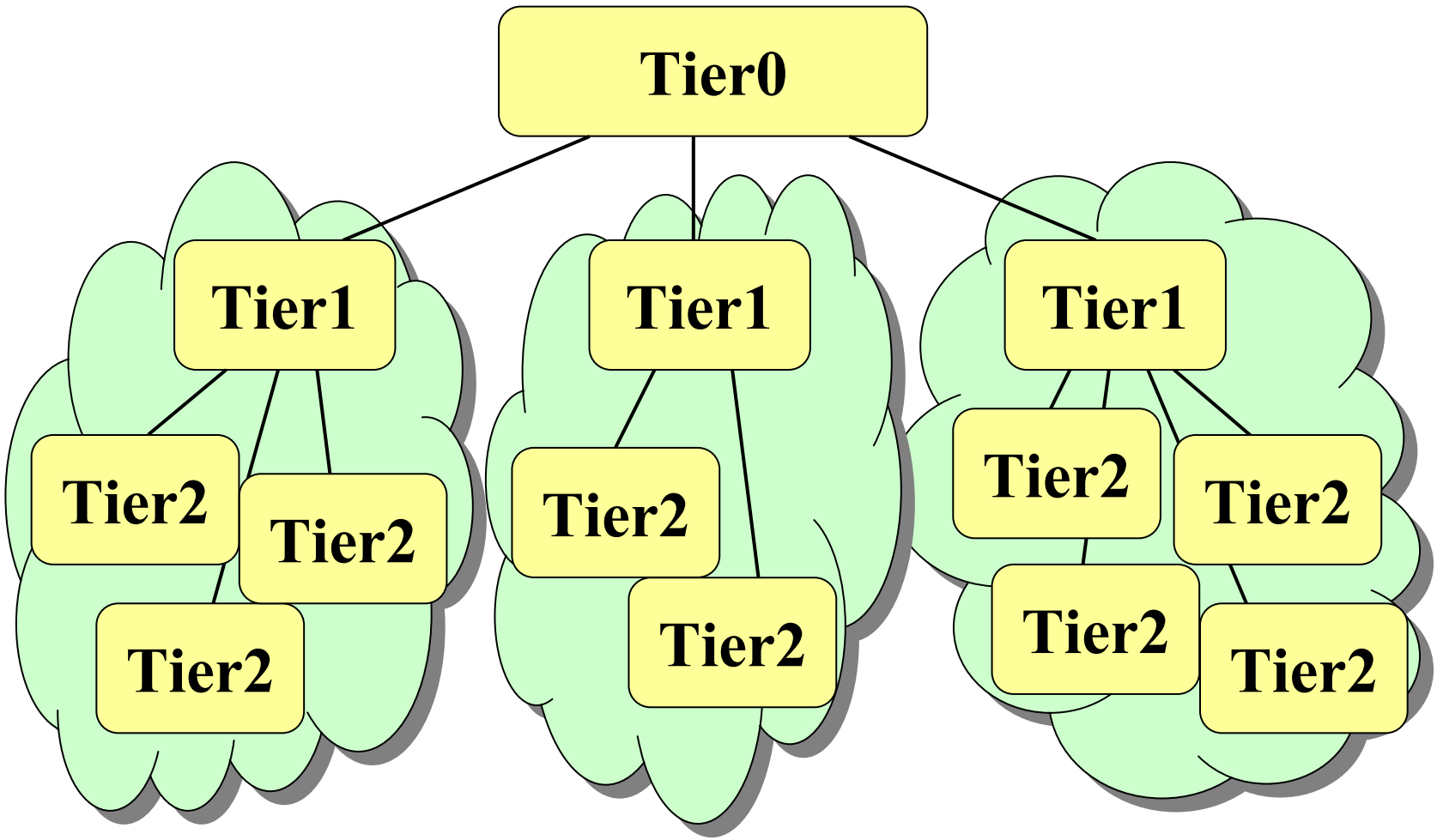
データ配置の秩序化

- グリッド = 混沌？
  - 無原則な利用 → データが無秩序に分散
- データ処理の困難
  - 1ジョブに複数の入力ファイル
    → ジョブ毎にデータ転送が生じて非効率

  ⇒ 同種のデータはまとめて配置（データセット）
- データ管理の困難
  - 膨大なデータ → データカタログの負担

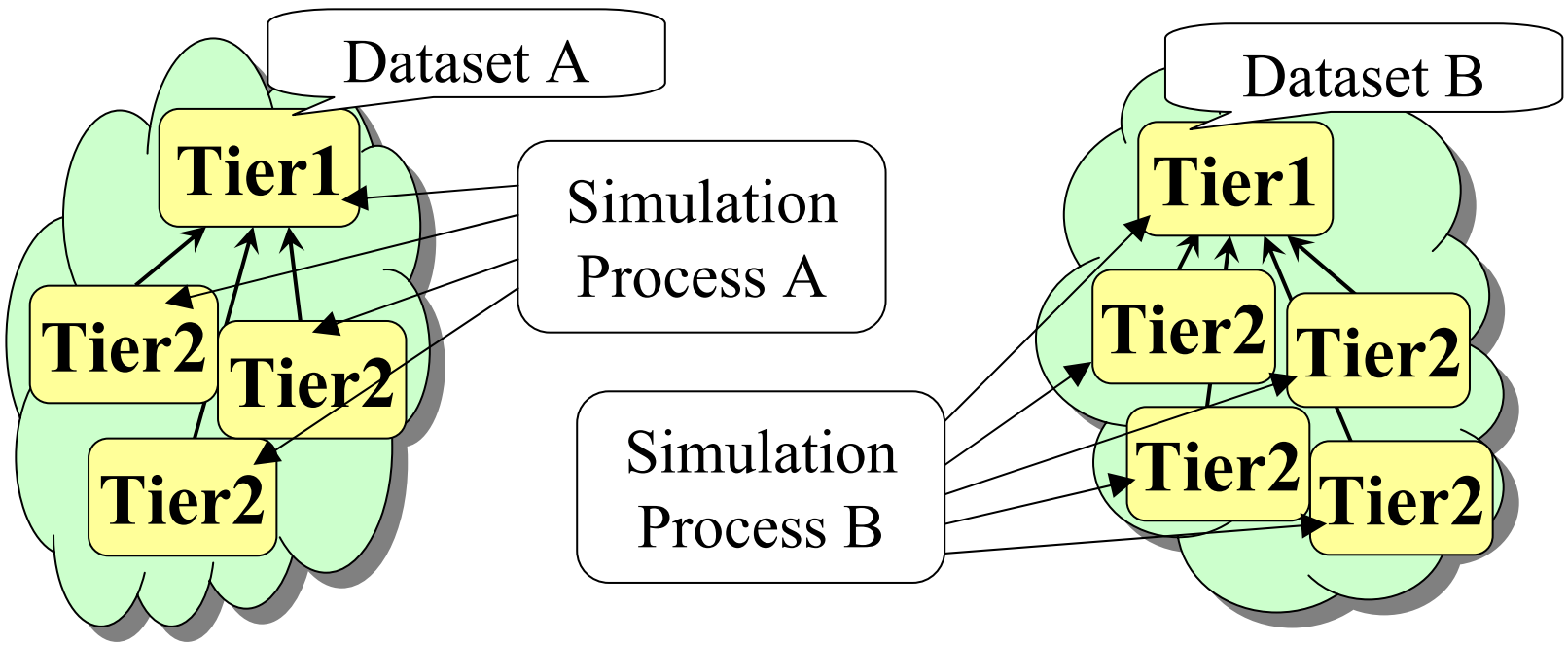
  ⇒ データカタログの階層化
    - 中央カタログ： データセットの所在を記録
    - 地域カタログ： データファイルの所在を記録

Cloudモデル


Tier0
Tier1
Tier1
Tier1
Tier2
Tier2
Tier2
Tier2
Tier2
Tier2
Tier2
Tier2

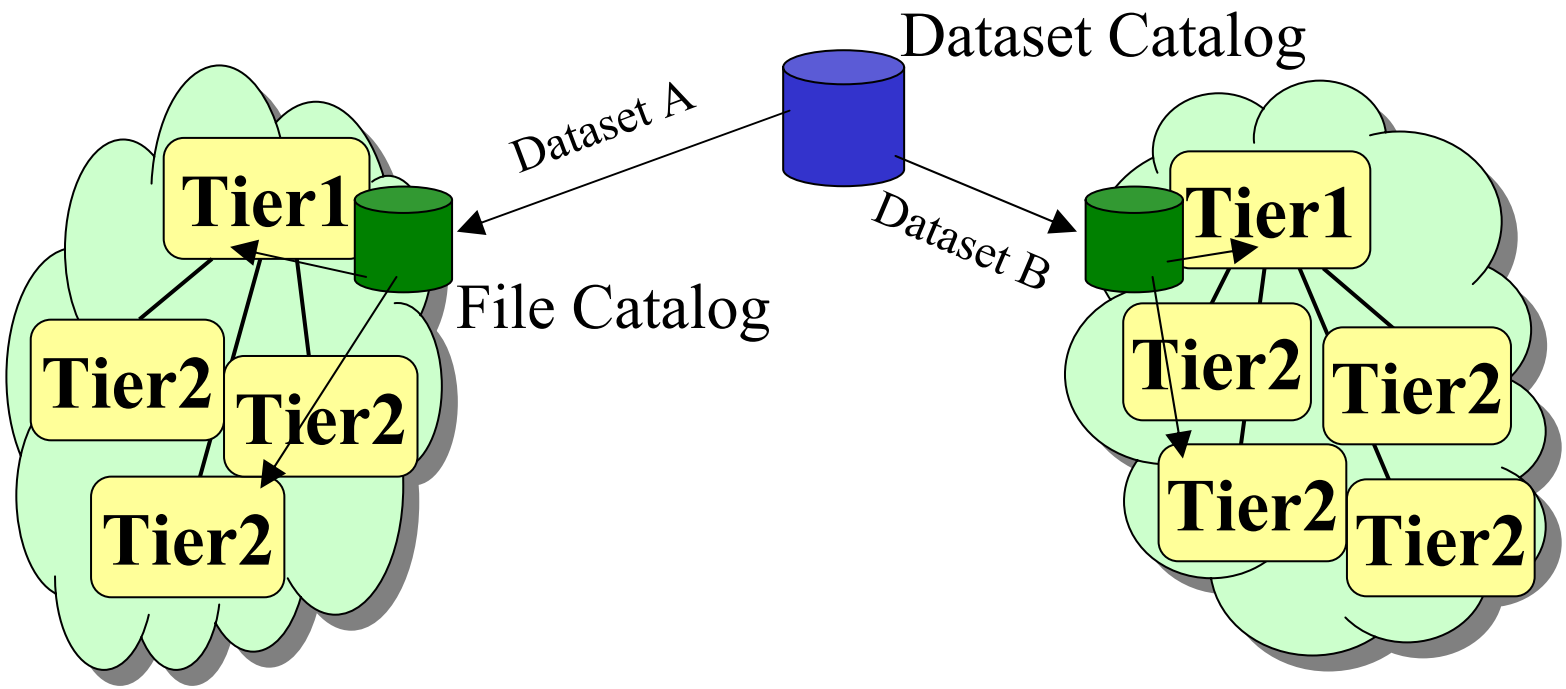
Cloud毎のデータ処理

- Simulation Production は Cloud 単位
  - 各データセットはCloud内で完結
    （ひとつのTier1でまとめて保管）



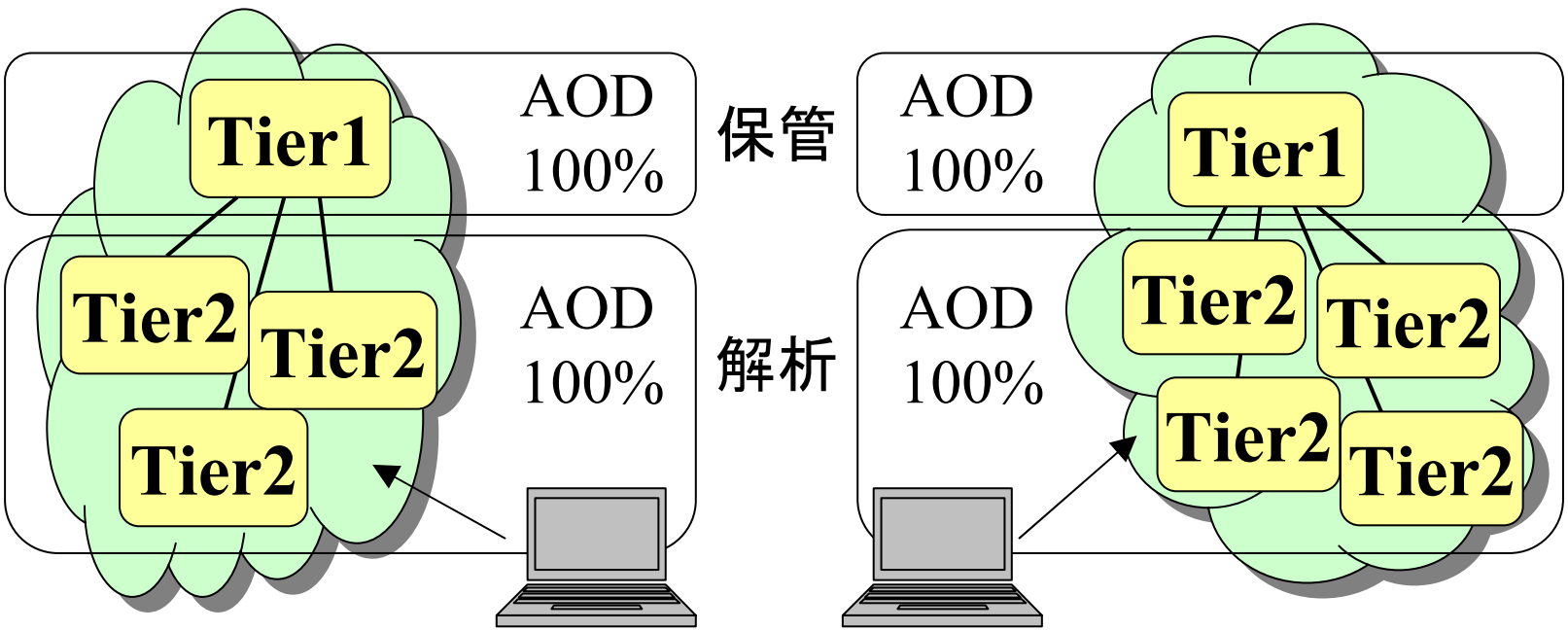
Cloud毎のデータ管理

- データカタログ
  - データセットカタログ：全データセットを管理
  - ファイルカタログ：Cloud内のファイルを管理

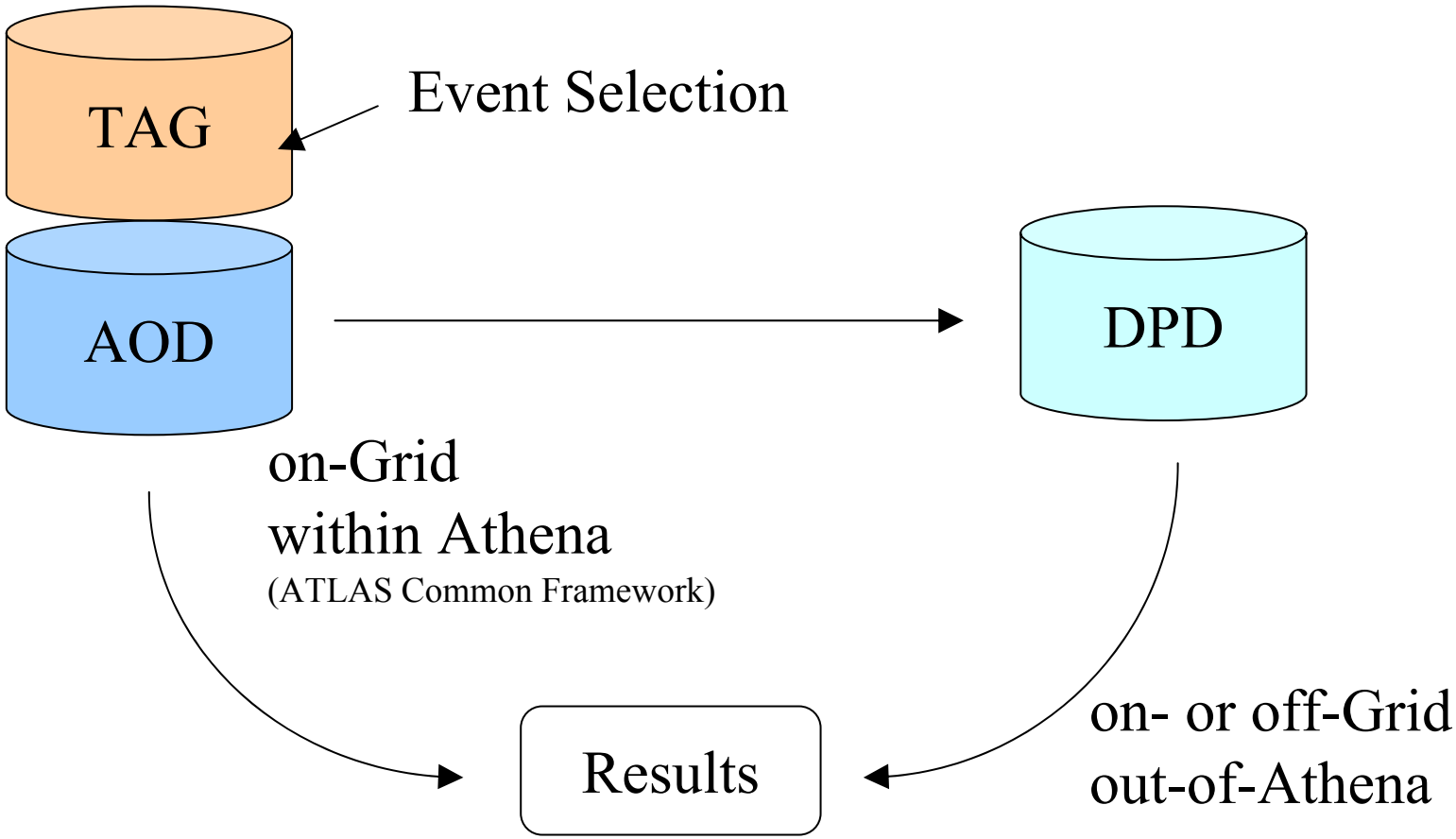


Cloud毎のデータ分散

- データ分配
  - Tier1 に AOD フルセット（データ保管）
  - Tier2 は AOD の複製を分散保持（物理解析）
    - Cloud 内の Tier2 全体でフルセット ⇒ Cloud 内で全解析可



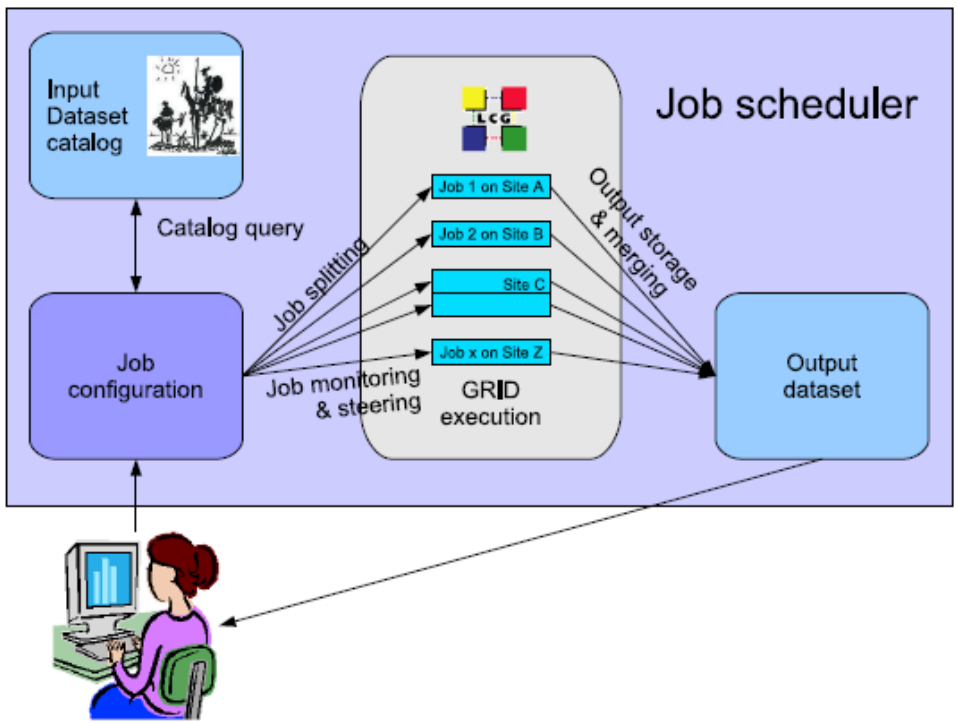
ATLAS Analysis Model



Distributed Analysis

Distributed Analysis

How to combine all these: Job scheduler/manager: GANGA

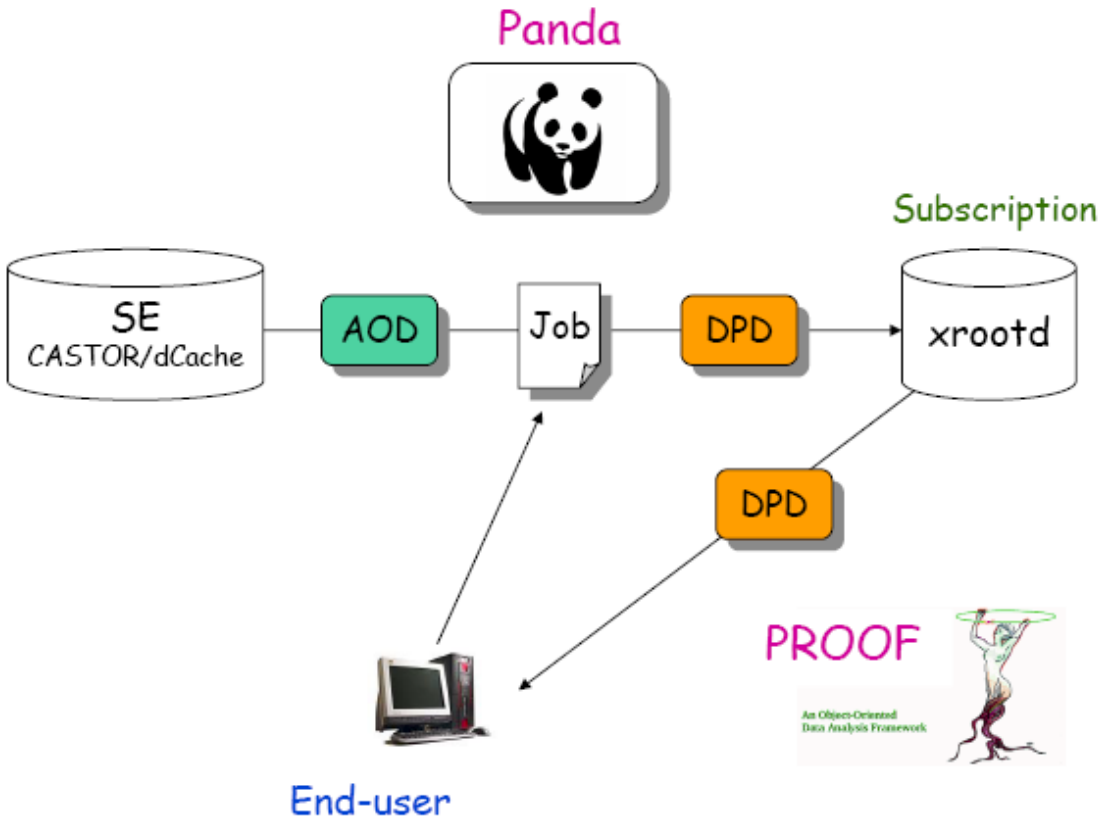


Johannes Elmsheuser (LMU München) ATLAS Distributed Analysis Meeting 24/10/2007 4 / 14

Atlas software & computing workshop (2007.Oct.) Johannes Elmsheuser

Distributed Analysis

ROOT-based Analysis (2/3)



8

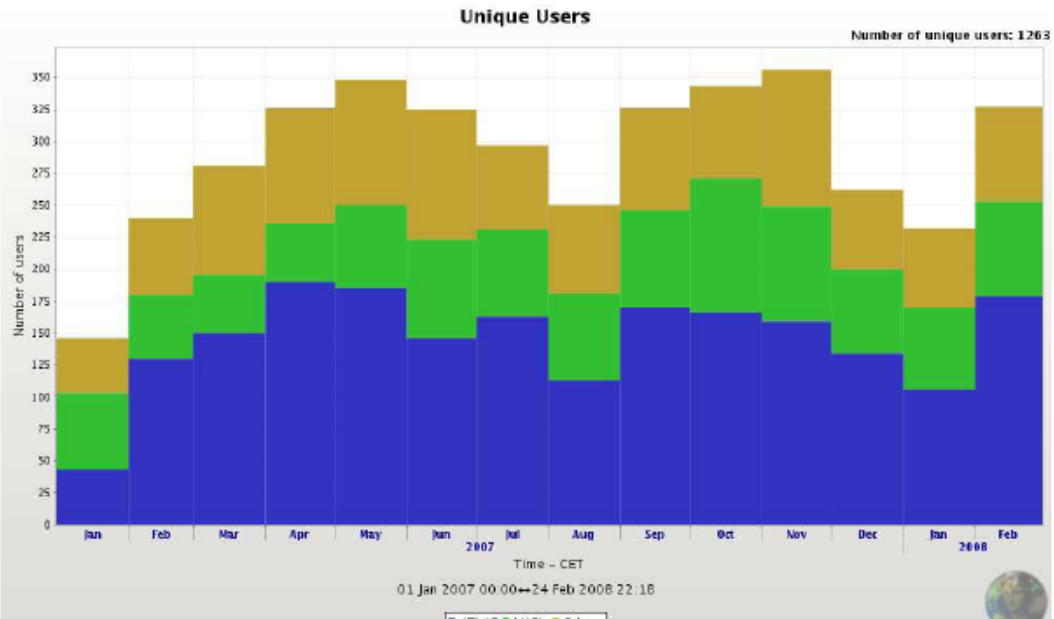
Atlas software & computing workshop (2007.Jun.) Tadashi Maeno

Distributed Analysis

Ganga Usage Statistics

- Over 1260 unique users since beginning of 2007
- Over 740 ATLAS users have tried Ganga at least once
- About 60 ATLAS Ganga users per week

January 2008 statistic not complete



Atlas software & computing workshop (2008.Feb.) Johannes Elmsheuser

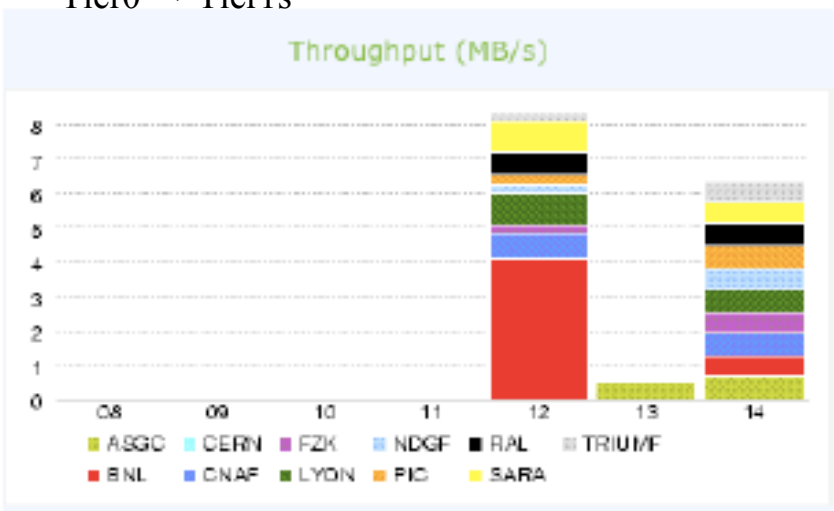
実験開始に向けた最終試験

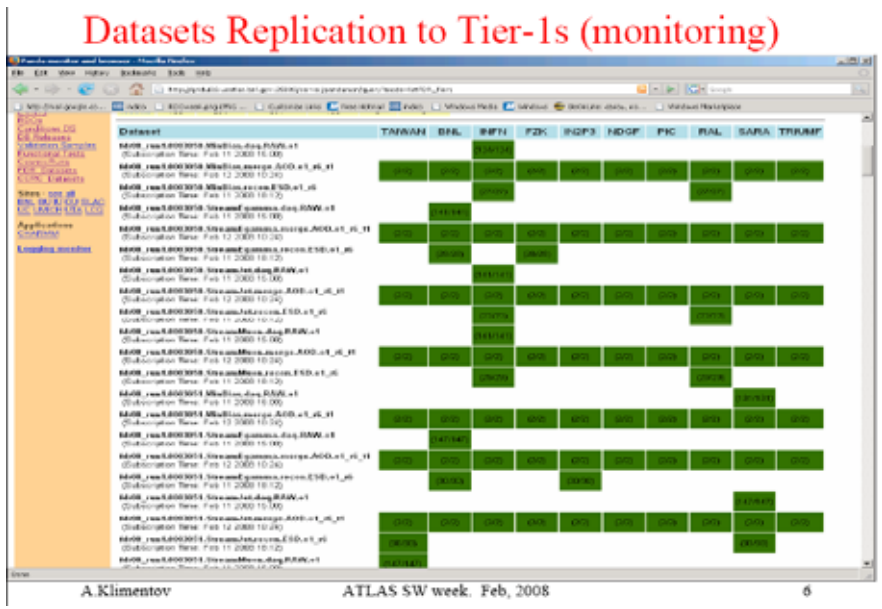
- FDR (Full Dress Rehearsal)
  - ATLAS の Software + Computing System の最終試験
  - FDR-1 (Feb. 2008), FDR-2 (May 2008)
- CCRC'08 (CommonComputingReadinessChallenges)
  - 実験個別の試験だけでなく、LHC四実験同時の試験が必要（主にデータ転送）
  - Phase-I (Feb. 2008), Phase-II (May 2008)

Full Dress Rehearsal

- FDR-1 (Feb. 2008)
  - Online System →(10 hrs @200 HZ)→ Tier0 → Tier1 → Tier2
  - Tier0: 較正、Data quality、Reconstruction、Grid: データ解析
  - 基本的な動作は確認出来たが各所に修正が必要

Tier0 → Tier1s

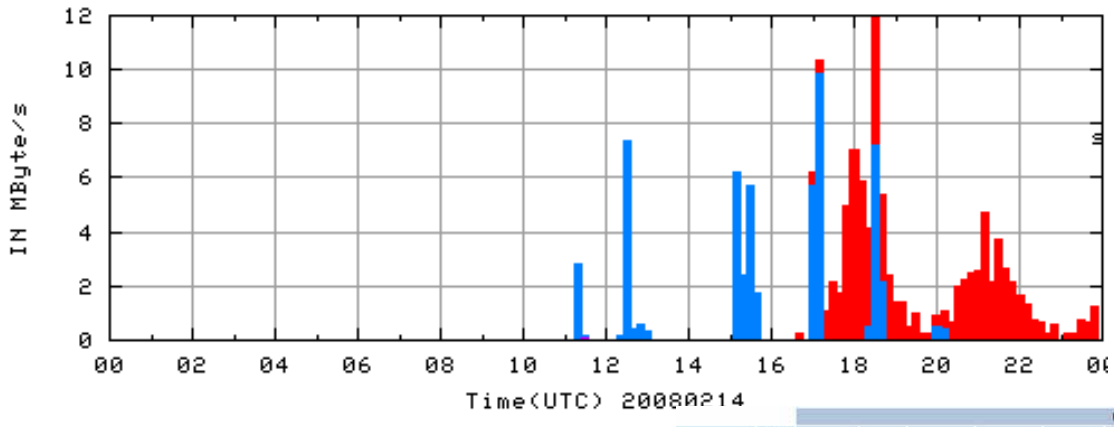




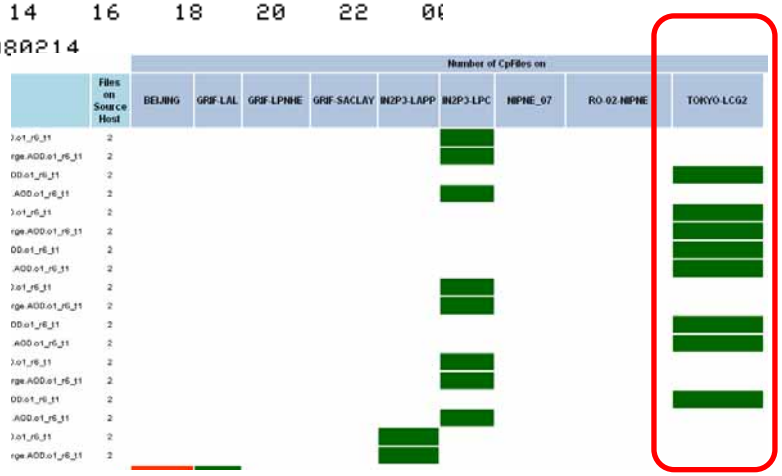
Tokyo in FDR-1

- 割り当て分を全て受信 (13 datasets, 26 files)

(青) Tier1s → Tier2 (Tokyo)



Tier2	Datasets	Total Files in datasets	Total CpFiles in datasets	Completed	Transfer	Subscribed
BEIJING	12	24	10	5	0	7
GRIF-LAL_DATADISK	12	24	24	12	0	0
GRIF-LPNHE_DATADISK	12	24	24	12	0	0
GRIF-SACLAY_DATADISK	12	24	24	12	0	0
IN2P3-LAPP_DATADISK	12	24	24	12	0	0
IN2P3-LPC_DATADISK	12	24	24	12	0	0
NIPNE_07	12	24	24	12	0	0
RO-02-NIPNE_DATADISK	12	24	24	12	0	0
TOKYO-LCG2_DATADISK	13	26	26	13	0	0

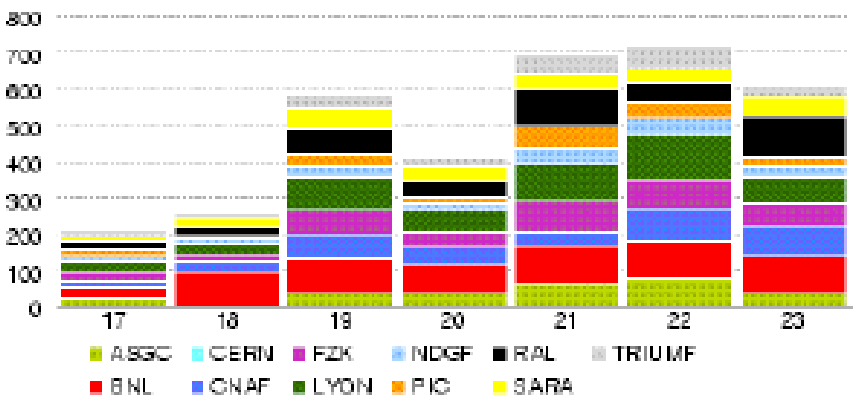


CCRC'08

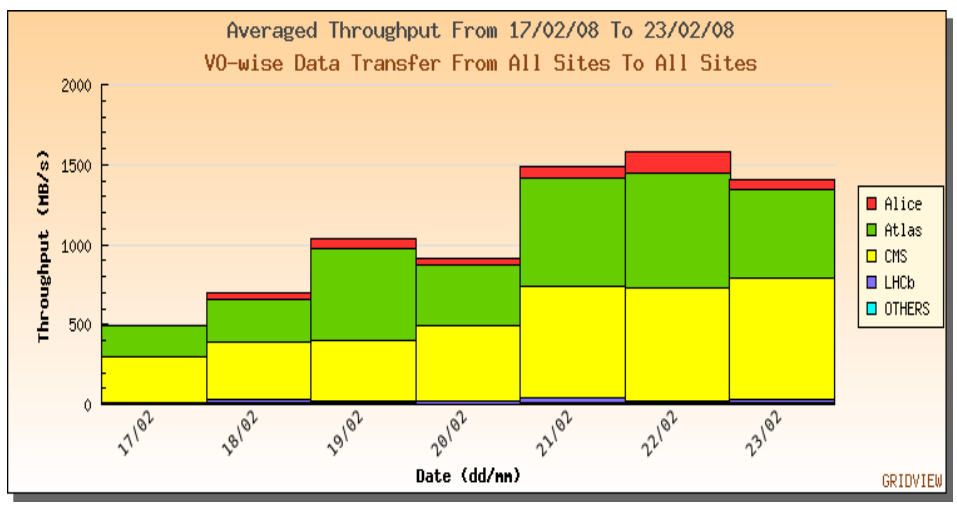
- Phase-I (Feb. 2008)
  - Tier0 → Tier1
  - ATLAS: 700 MB/s for 2 days, peak > 1 GB/s

    cf. 1.6 MB x 200 Hz = 320 MB/s
  - Total: >1.5 GB/s, peak >2 GB/s

ATLAS Tier0 → Tier1s



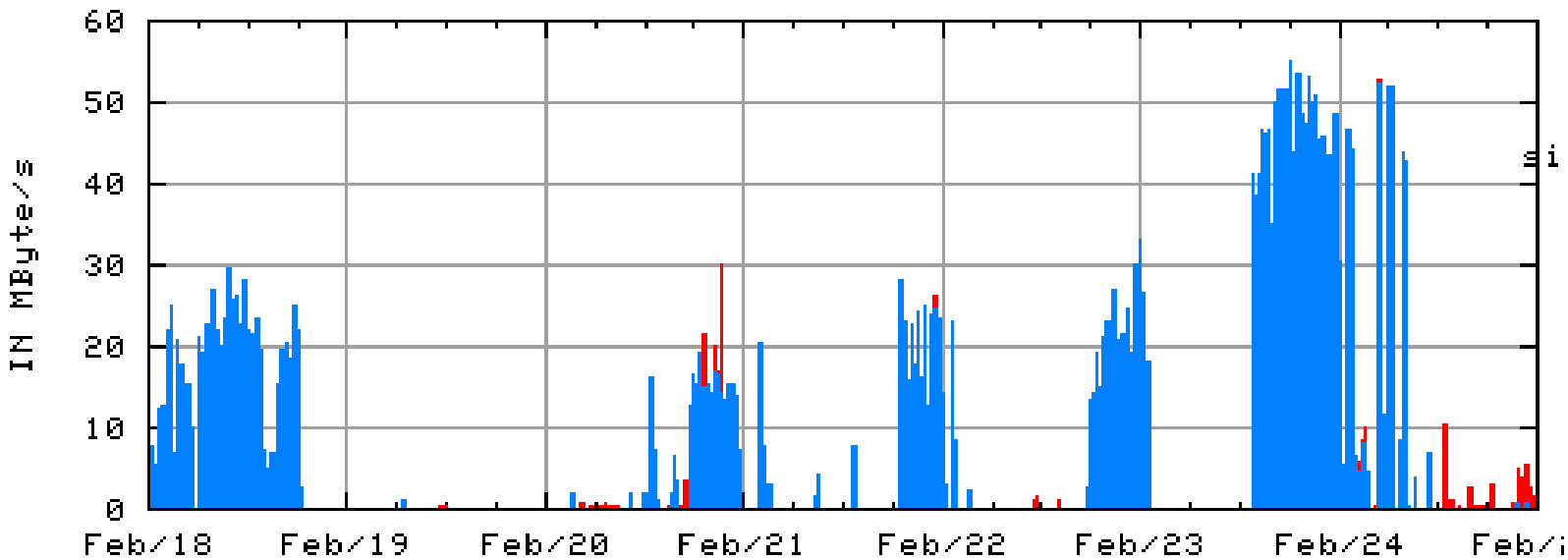
WLCG All



Tokyo in CCRC'08

- ATLAS は Tier0 → Tier1 の転送試験に集中していたが、自発的に Tier1 → Tier2 (Tokyo) 試験を実施

(青) Tier1s → Tier2 (Tokyo)



実験開始に向けた試験

- Cosmic Ray Runs
  - 宇宙線を利用したATLAS測定器の試験
  - 同時にコンピューティングシステムも試験

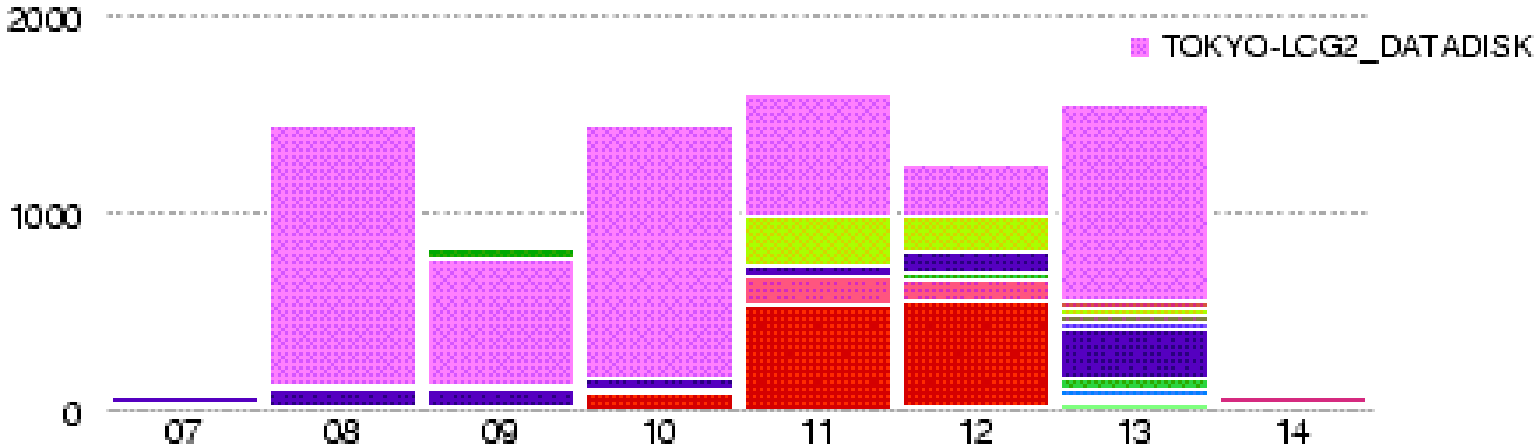
  ⇒ATLAS実験の全システムの試験

  測定器 ⇒ Online ⇒Tier0 ⇒ Tier1 ⇒ Tier2

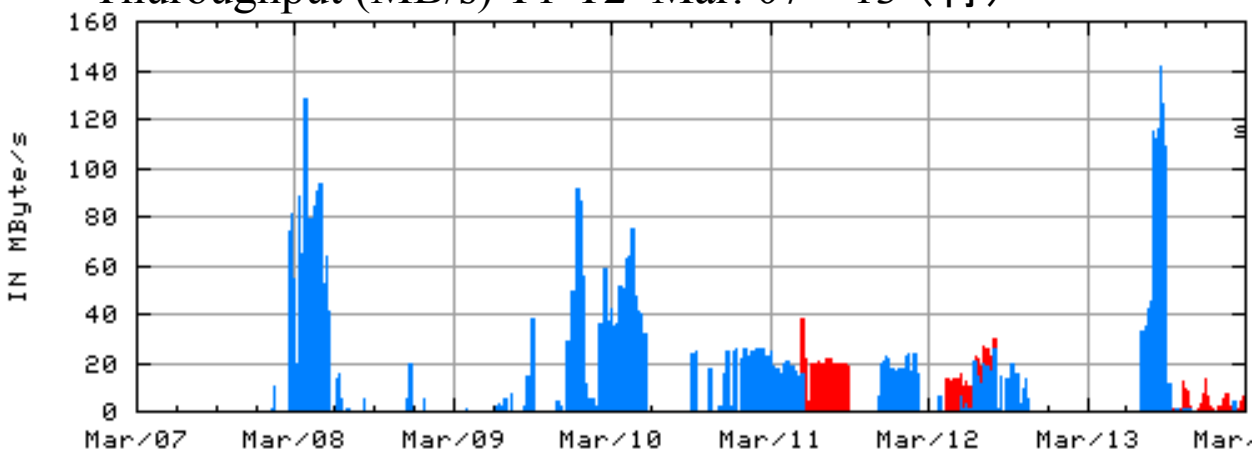
  cf. FDRはOnline systemから、CCRCはTier0から

Tokyo in Cosmic Ray Run

Data transferred (GB) T1-T2 Mar. 07 - 13



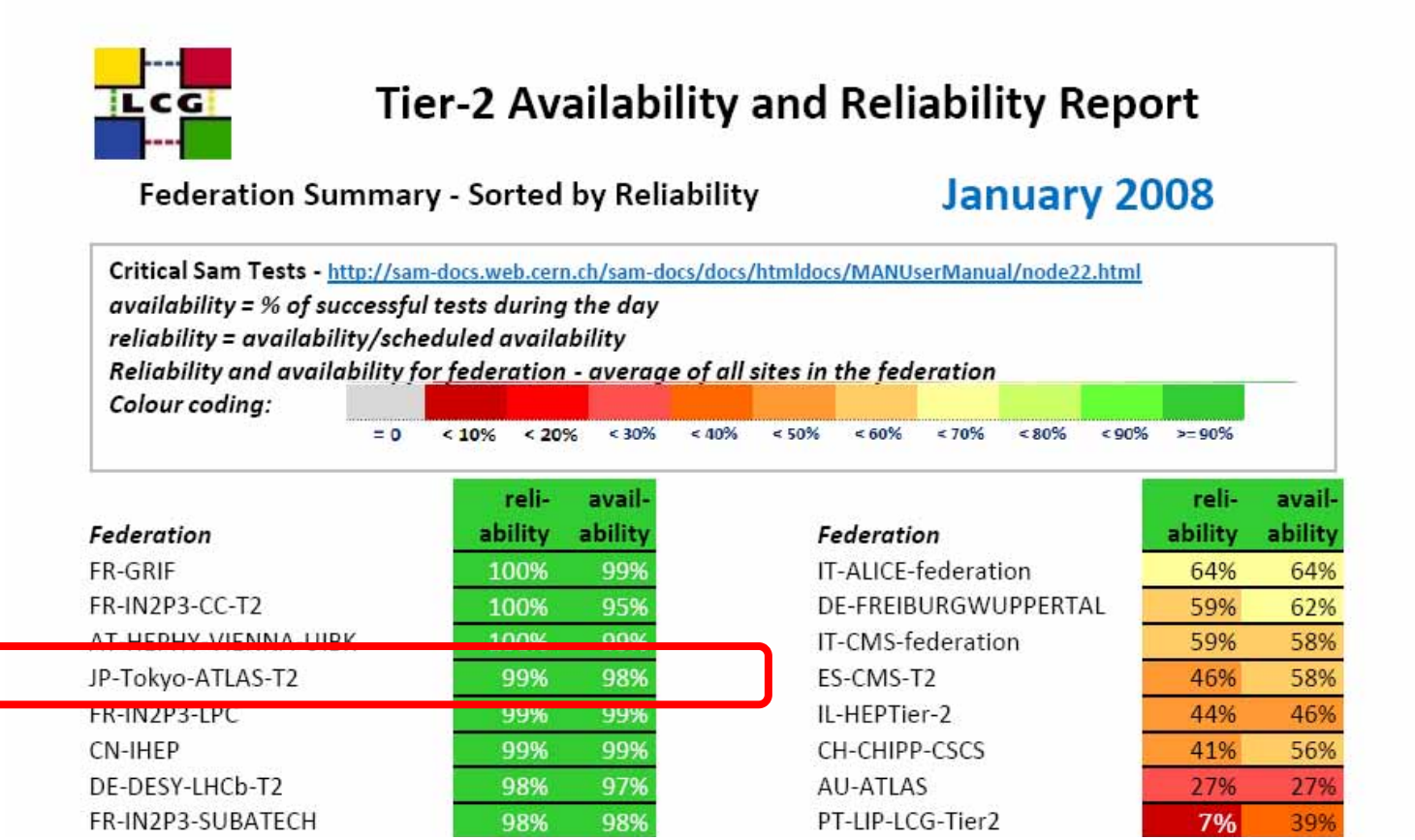
Thuroughput (MB/s) T1-T2 Mar. 07－13（青）



モニタリング

- 短期的な機能試験・性能試験だけでなく、長期的動作・安定性も確かめる必要がある
- Availability + Reliability
  - 各サイトの安定性
- Accounting
  - 使用量が供出量に比較して適正か確認
  - Fair share の実現のために必要

WLCG T2 Reliability



Tier-2 Availability and Reliability Report

Federation Summary - Sorted by Reliability — **January 2008**

Critical Sam Tests - http://sam-docs.web.cern.ch/sam-docs/docs/htmldocs/MANUserManual/node22.html

availability = % of successful tests during the day

reliability = availability/scheduled availability

Reliability and availability for federation - average of all sites in the federation

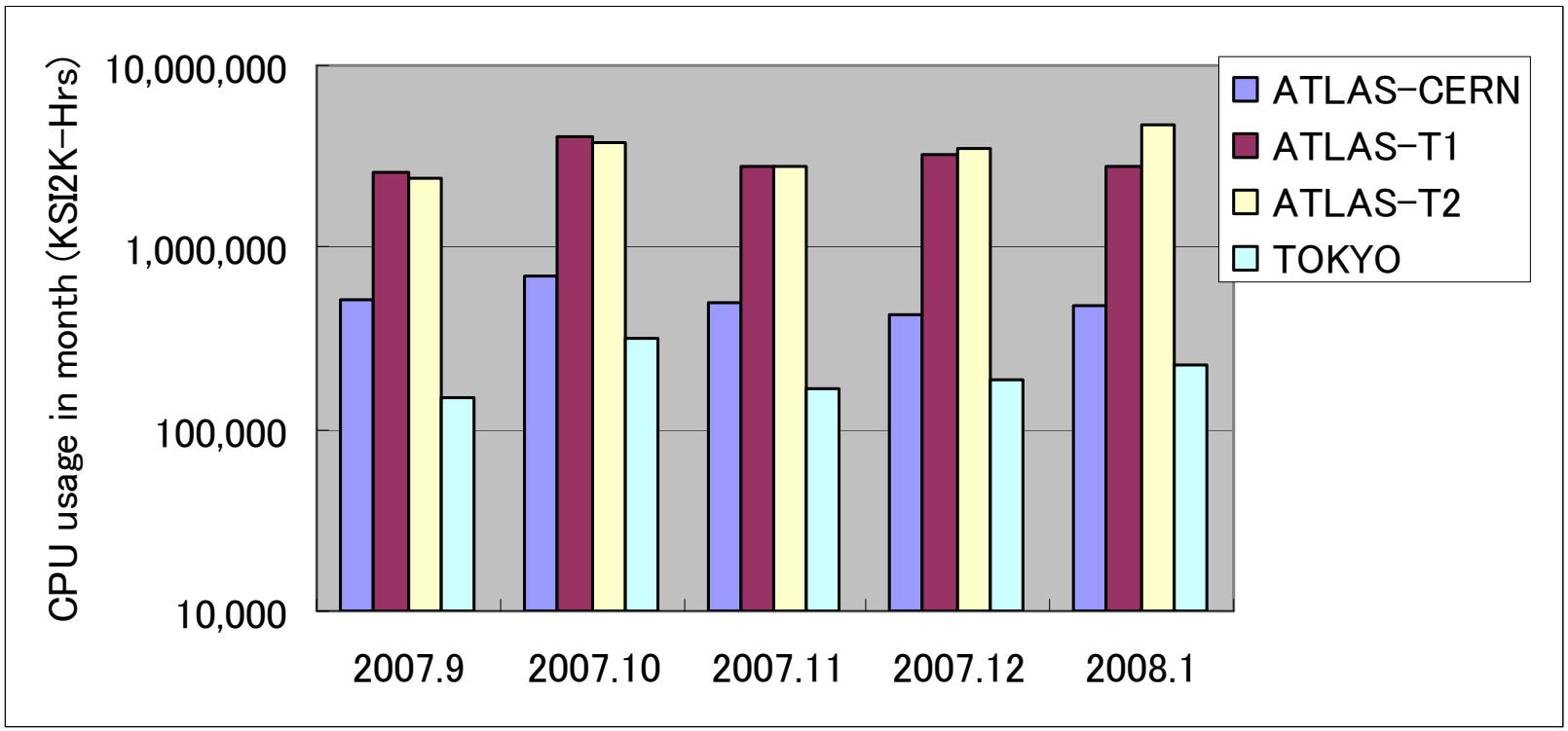
Colour coding: = 0 | < 10% | < 20% | < 30% | < 40% | < 50% | < 60% | < 70% | < 80% | < 90% | >= 90%

Federation	reli-ability	avail-ability
FR-GRIF	100%	99%
FR-IN2P3-CC-T2	100%	95%
AT-HEPHY-VIENNA-UIBK	100%	99%
JP-Tokyo-ATLAS-T2	99%	98%
FR-IN2P3-LPC	99%	99%
CN-IHEP	99%	99%
DE-DESY-LHCb-T2	98%	97%
FR-IN2P3-SUBATECH	98%	98%

Federation	reli-ability	avail-ability
IT-ALICE-federation	64%	64%
DE-FREIBURGWUPPERTAL	59%	62%
IT-CMS-federation	59%	58%
ES-CMS-T2	46%	58%
IL-HEPTier-2	44%	46%
CH-CHIPP-CSCS	41%	56%
AU-ATLAS	27%	27%
PT-LIP-LCG-Tier2	7%	39%

http://cern.ch/LCG/MB/accounting/Tier-2/January2008/Tier-2%20Accounting%20Report_January2008.pdf

ATLAS CPU Usage



(Values from http://cern.ch/LCG/planning/planning.html)

ATLAS Jobs per Site

2007.10. http://dashb-atlas-prodsys.cern.ch/dashboard/request.py/summary?grouping=site

site	success	failure	efficiency
TRIUMF-LCG2	44564	11956	78.8%
IN2P3-CC	27935	21288	56.8%
BNL	33067	15859	67.6%
TOKYO-LCG2	29391	8457	77.7%
UKI-NORTHGRID-MAN-HEP	33412	3432	90.7%
INFN-T1	27613	7529	78.6%
FZK-LCG2	26241	5361	83%
None	23219	8098	74.1%
RAL-LCG2	25179	4375	85.2%
Titan A (UiO/USIT)	23184	3251	87.7%
MidwestT2	18218	5175	77.9%
SLAC	17062	3545	82.8%
GRIF	14019	4913	74%
BostonU	14680	3203	82.1%
IN2P3-CC-T2	10967	6650	62.3%
IN2P3-LPC	11360	5189	68.6%
CERN-PROD	13773	1511	90.1%
VICTORIA-LCG2	7111	7025	50.3%
SiGNET	10478	3657	74.1%

まとめ

- ATLASコンピューティングシステムはグリッドを利用して整備され、実験開始に向けて準備が整いつつある
- グリッドを利用したデータ解析も既に可能となっている
- 測定器からのデータをCERNで記録し、同時に各地域センターに配布する試験を実施し、動作することを確認した。また、必要な転送レートも達成している。

- ATLAS日本の地域解析センターとして、東京大学素粒子センターは積極的に試験に参加し、実験データを確実に、かつ高い転送レートで受けられることを確認した
- 同センターは長期間安定して動作しており、Simulation data の生産でも着実に貢献している