



Real big dataとクラウドサービス

高エネルギー加速器研究機構
計算科学センター 教授
佐々木節



高エネルギー物理学

- 物質の根源と宇宙の謎を実験的に解明する
- 加速器を用い、粒子を加速し、素粒子同士の相互作用を詳しく調べる
- 大きな施設を必要とするので、国際協力で実験行われている
- データ量の増大と共に、計算機、ネットワークの比重が高まってきた
 - World Wide Web, インターネット、グリッドコンピューティングなどの発展に寄与

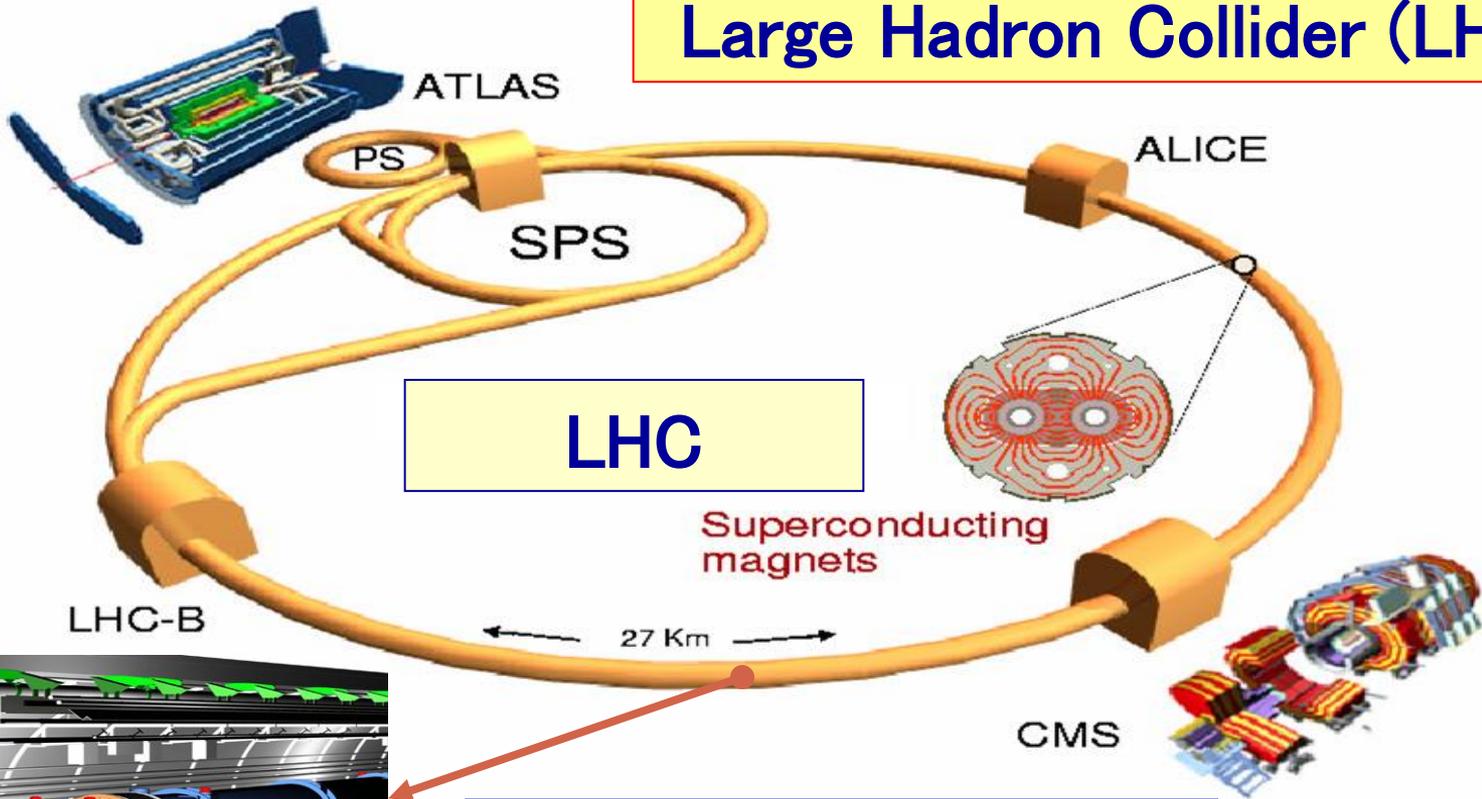


LHC

- スイス・ジュネーブに建設された加速器
- Higgs粒子の発見で有名に
- 多量のデータを国際協力で解析
- プライベートクラウドを採用
 - ハンガリーに第2計算センターを構築



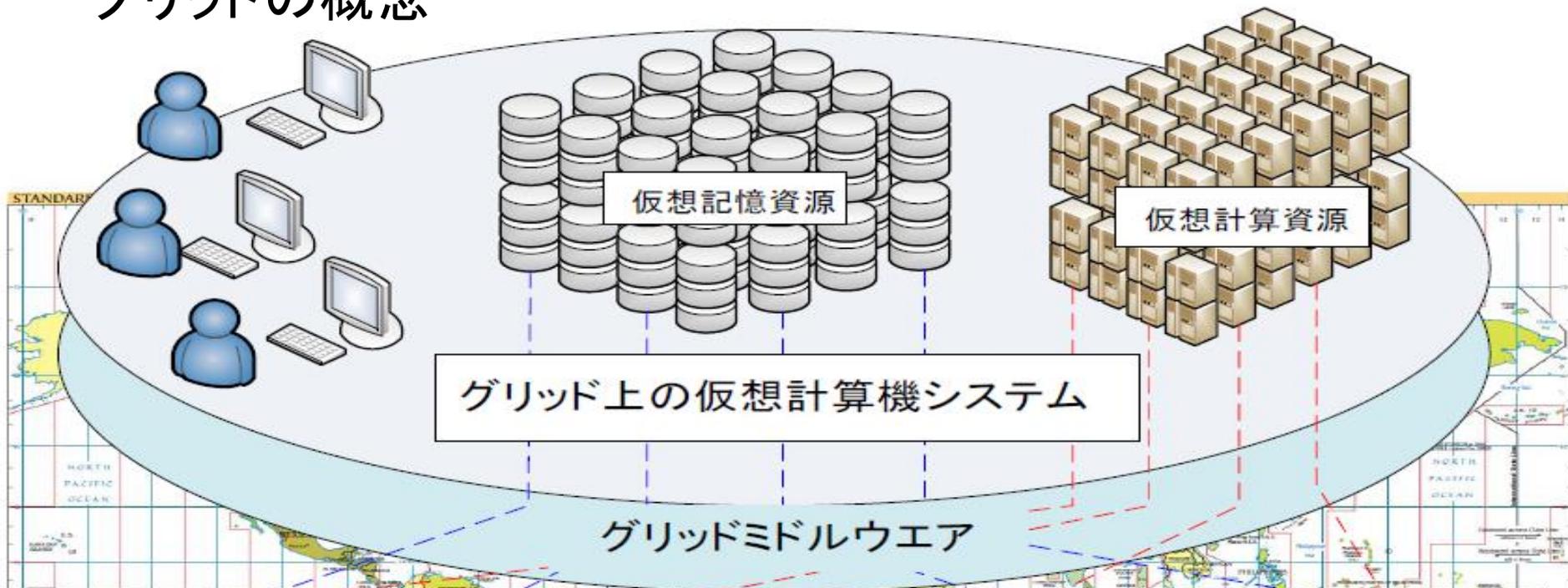
Large Hadron Collider (LHC)



- ・ 14 TeV の陽子・陽子衝突型加速器
- ・ LEPTンネルを利用
- ・ 建設に14年
- ・ 総建設費は約5000億円

- * CERNのLEP(e^+e^-)は 200 GeV
- * 米国フェルミ研究所の Tevatron(pp)は 2 TeV

グリッドの概念



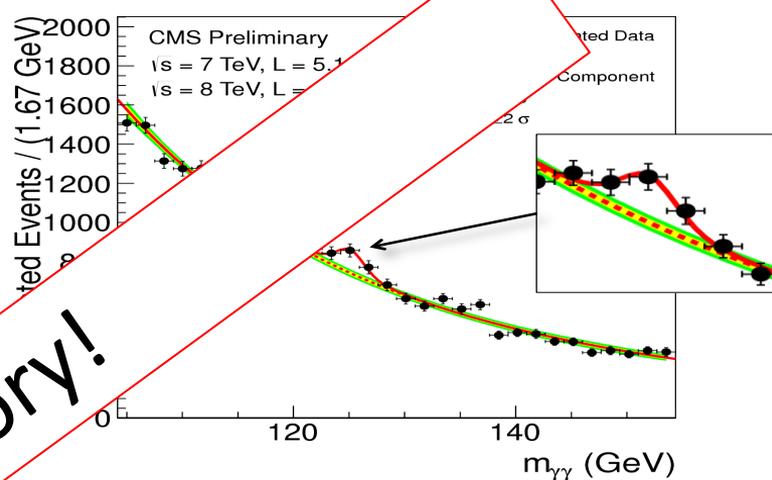
Global Effort → Global Success

Results today only possible due to extraordinary performance of accelerators – experiments – Grid computing

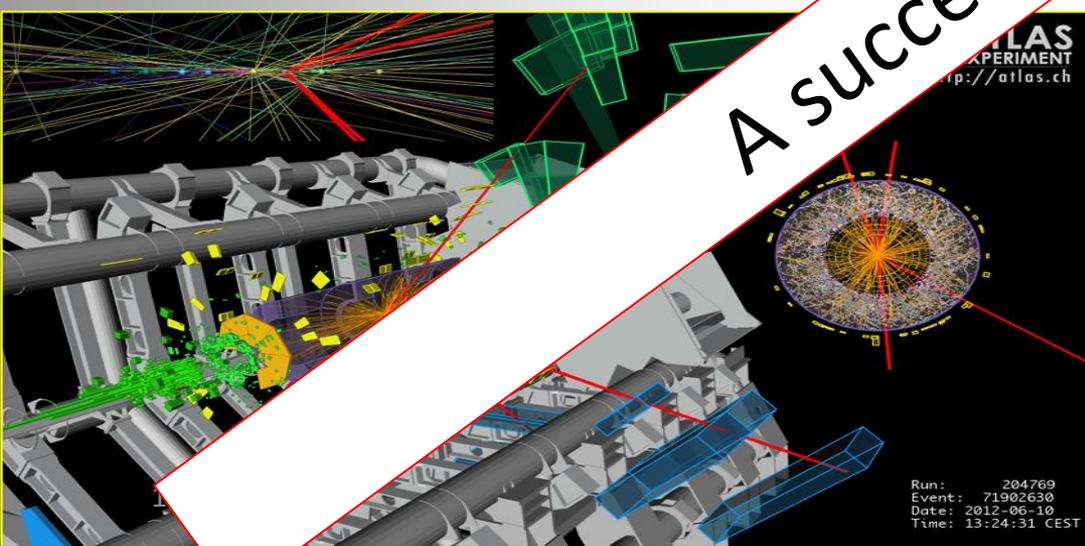
Observation of a new particle consistent with a Higgs Boson (but which one...?)

Historic Milestone but only the beginning

Global Implications for the future



A success story!



Higgs boson-like particle discovery claimed at LHC

COMMENTS (1865)

By Paul Rincon

Science editor, BBC News website, Geneva



The moment when Cern director Rolf Heuer confirmed the Higgs results

Cern scientists reporting from the Large Hadron Collider (LHC) have claimed the discovery of a new particle consistent with the Higgs boson.

Relat
Q&A



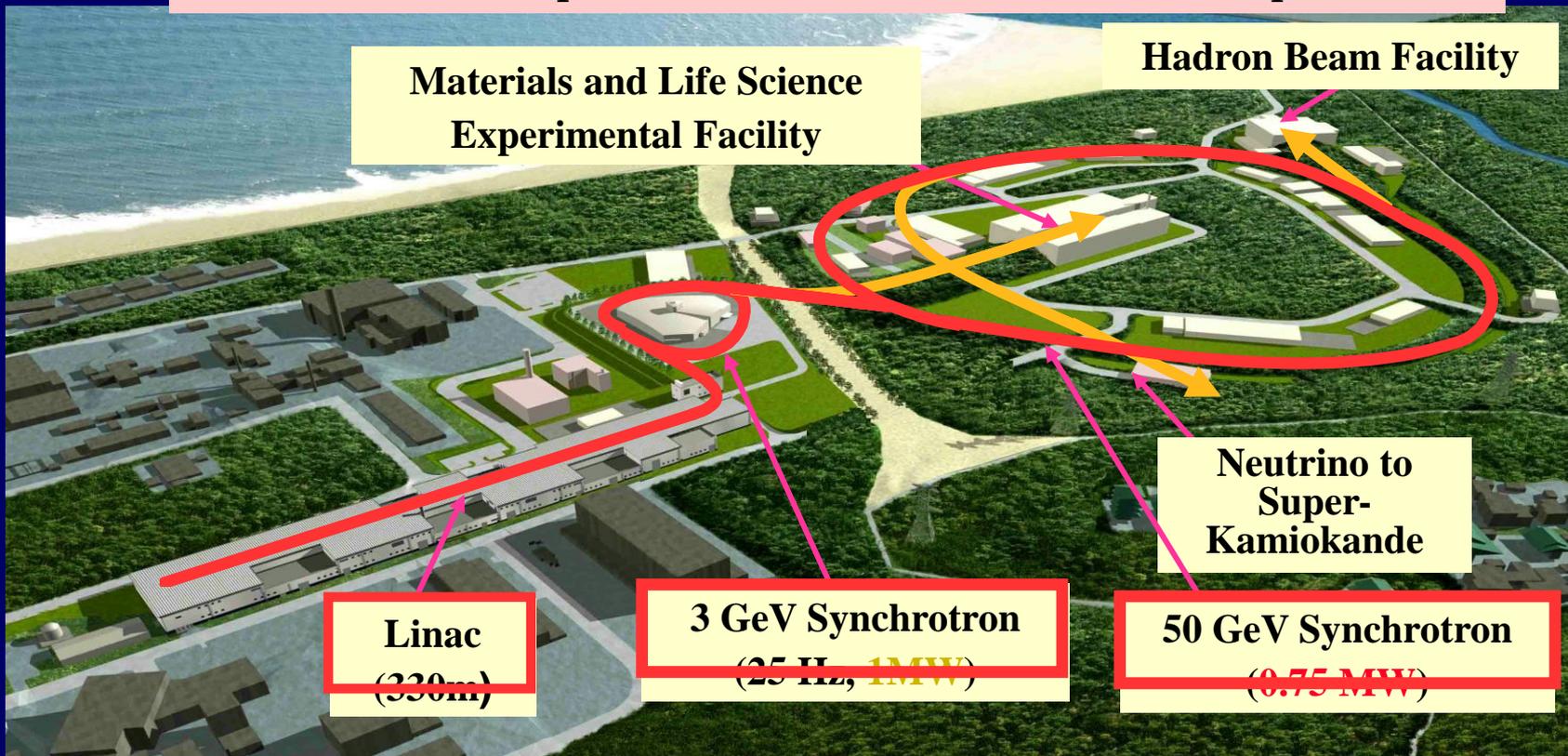
KEKの加速器

- J-PARC
 - 東海村で稼働中
 - 素粒子、原子核実験、物質、材料、ライフサイエンス
 - T2K実験他
 - 年間2PB程度のデータ
- フォトンファクトリー
 - 物質、材料、ライフサイエンス
- Super-KEKB
 - KEKBをアップグレード
 - BelleII実験
 - 世界21の国と地域からの約80の大学・研究機関に所属する500人以上の研究者が参加
 - Belle実験で蓄積されたデータの50倍のデータを収集・解析



J - PARC (Tokai)

J-PARC = Japan Proton Accelerator Research Complex



Joint Project between KEK and JAEA



J-PARC (T2K Experiment)

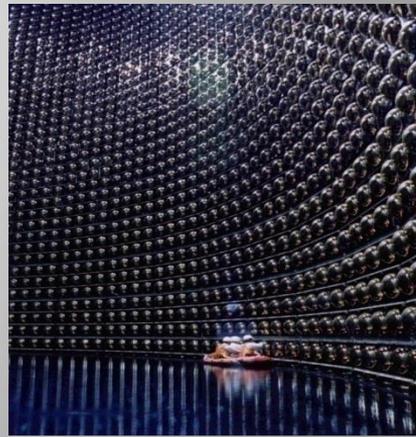
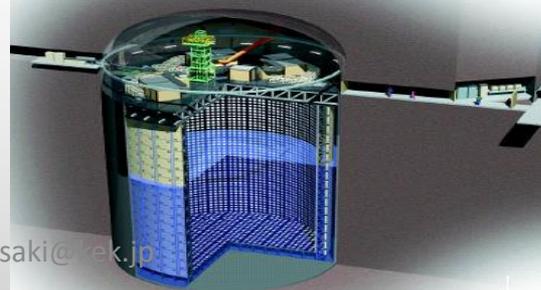


Neutrino Experimental Facility

295 km
West



Super-Kamiokande





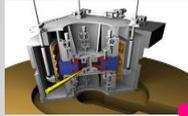
J-PARC(Materials & Life Science Facility)

Tentatively Approved Instruments

Versatile powder diffractometer
- JAEA



Protein Dynamics Analysis Instrument (DIANA)
- JAEA



High-Resolution Chopper Spectrometer (HRC)
- KEK

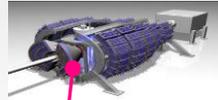
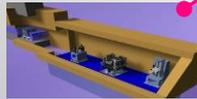


Cold Neutron Double Chopper Spectrometer (CNDCS)
- JAEA

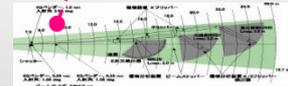
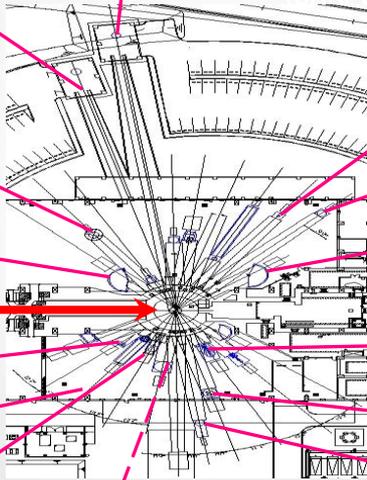


High-intensity SANS (HI-SANS) - JAEA

Neutron Reflectometer with Horizontal-Sample Geometry
- KEK



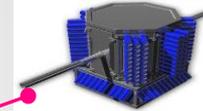
Super High Resolution Powder Diffractometer (SHRPD) - KEK



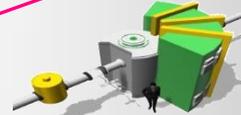
Neutron Resonance Spin Echo spectrometers
Takashi.Sasaki@kek.jp
KEK, Kyoto University



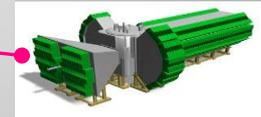
IBARAKI Biological Crystal Diffractometer
- Ibaraki Prefecture



Diffractometer for Biological X-ray diffraction (BIX-PN) - JAEA



4d Space Access Neutron Spectrometer (4SEASONS)
- Grant-in-Aid for Specially Promoted Research, MEXT
High-intensity Versatile Neutron Total Diffractometer
- KEK



IBARAKI Materials Design Diffractometer
- Ibaraki Prefecture



Engineering Diffractometer
- JAEA

KEK Tsukuba campus



SuperKEKB and BelleII

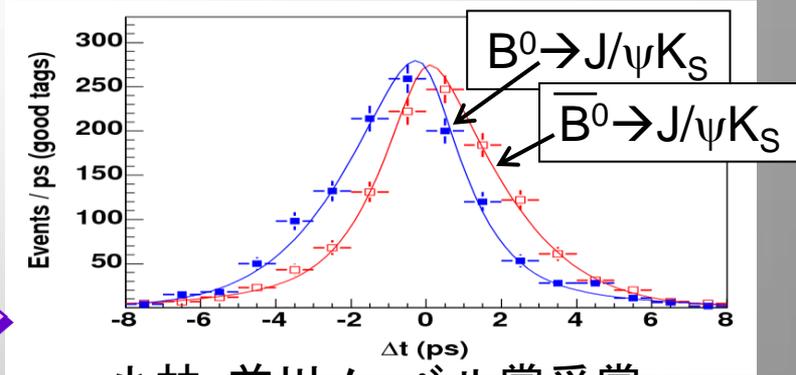
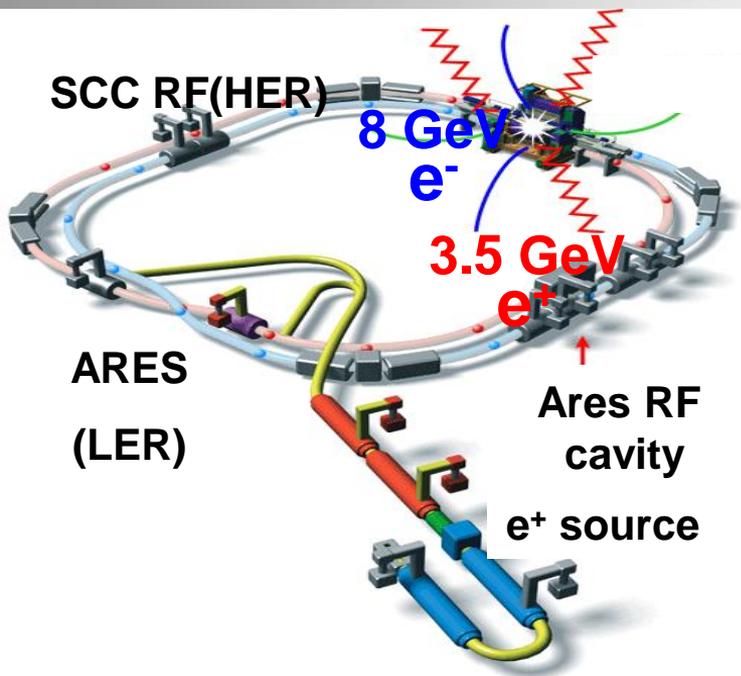




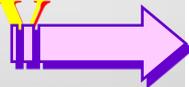
KEKB $e^+ e^-$ Collider

Belle Experiment

13 countries, 57 institutes, ~400 collaborators



Observation of CPV
in the B meson system



Takashi.Sasaki@kek.jp

小林・益川ノーベル賞受賞

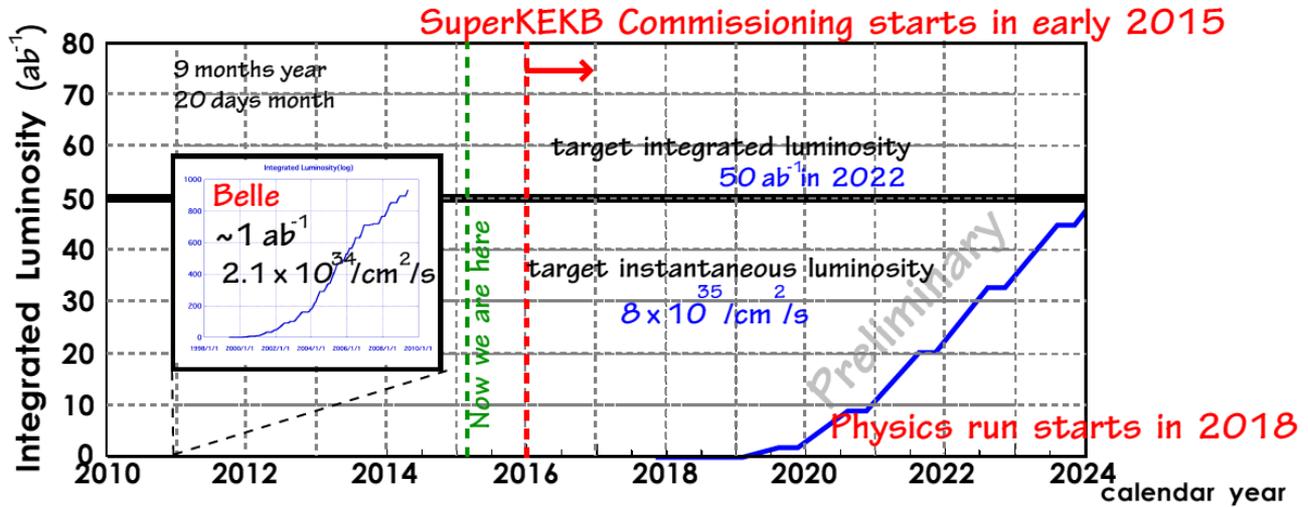


BelleII

- 2018年にデータ収集開始
- 生データだけで100PB程度
 - 同じ程度のモンテカルロデータも作る
 - ユーザはリダクションし、物理系統ごとに分けたサブセットを解析する
- LHCのために開発されたグリッドの枠組みを借用して国際分散計算機環境を構築



Luminosity Prospect



Experiment	Event size	Rate @ Storage	
		[kB]	[event/sec]
Belle II	300	6,000	1,800 (@ max. luminosity)
ALICE (Pb-Pb)	50,000	100	4,000
ALICE (p-p)	2,000	100	200
ATLAS	~ 700	600	400
CMS	$\sim 1,000$	500	several 100s
LHCb	55	4,500	250

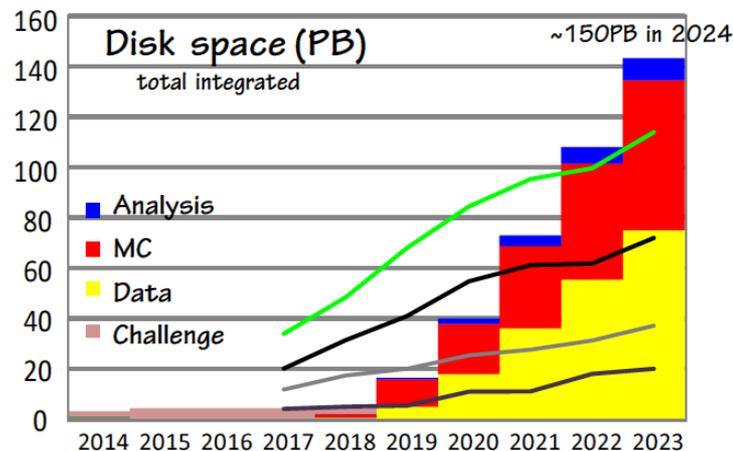
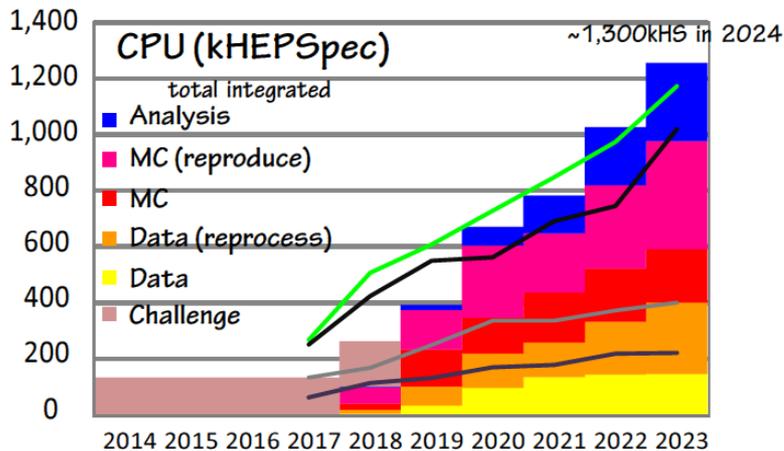
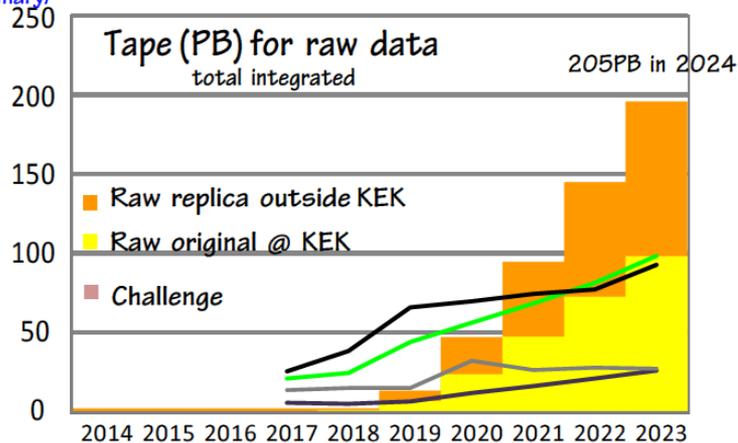
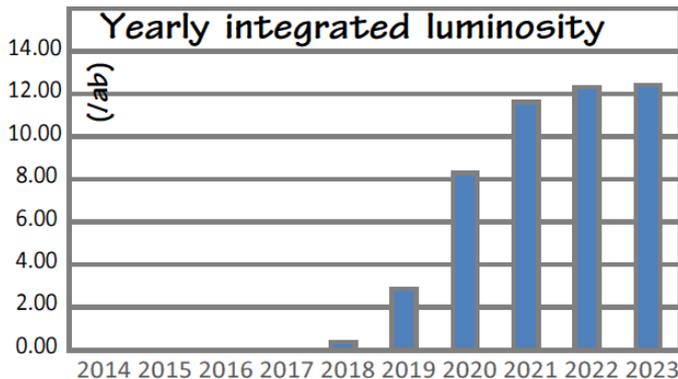
(LHC experiments : as seen in 2011/2012 runs)



Hardware Resources for Belle II

Pledge summary of LHC experiments: <http://wlcg-rebus.cern.ch/apps/pledges/summary/>

— ATLAS — CMS — ALICE — LHCb



ICTにおける課題

- データの保全
 - KEKのみにデータを保存するのは危険
 - 外部にコピーを置きたい
- 地域解析センター
 - 国際協力によるデータ解析
 - 多量のデータを送受信
- 需要見込みが困難
 - 加速器は困難な技術的課題を解決しつつ開発され、運転される
 - 見込み以上の性能を発揮することも
- KEK内の他のプロジェクトとの共存



データ保全

- データのコピーを別のサイトにおいて、非常時に備えたい
 - 3.11からの教訓
 - データ損失事故にも備える
- Data preservation
 - データとソフトウェアを長期(永久に?)に動態保存すべきという考えもある
- コマーシャルクラウド
 - データを読み出す必要が出来た時に、多額の送出料を払えるか
 - 災害や事故など、予算的に困難な時期に外部からの復旧が必要となる
- アカデミッククラウド
 - 持続性が課題



地域解析センター

- 国によっては、アカデミッククラウドサービス（無料とは限らないが）が展開されており、グリッドに代わり、クラウド上で地域解析センターを運用
 - データ転送のプロトコルはGridFTPで決め打ち
 - ファイルカタログは、自前のもの



クラウドサービスの活用

• パブリッククラウド

- コマーシャル、アカデミック両方が候補
- Amazon EC2等を試用
- ネットワークの転送量が問題
 - コマーシャルクラウドは、送出料が高価

• プライベートクラウド

- 需要が大きすぎて欲しいだけ計算機を買えない
 - プロジェクト毎の優先順位付けが必要
- 計算機を最大限に効率よく使いたい
 - 大きなプロジェクトの空き時間は、別のプロジェクトに割り振りしたい

パラダイムシフト

- 計算機は必要なだけ買えない
 - 世界中のお友達に手伝ってもらわなければならない
- データは使い捨てではない
 - 再現性、科学遺産などの観点から永久保存が望まれている
- ネットワークは無料ではない
 - 受益者負担が求められる時代に
- お金を払えば代行してくれるサービス(クラウド)はある
 - 予算との兼ね合い

今後の見通し

- 年々、KEKの予算は苦しくなっており、計算機およびネットワークへの需要の増大に答えること難しくなりつつある
 - 施設建設維持経費
 - 加速器の運転に必要な電気代
- 計算機を調達し、内部で運用するコストと、外部委託のコストは十分吟味する必要がある
 - クラウドサービスの調達であれば、政府調達の18か月ルールに掛からない??
- 全てを外部委託は現時点では不可能
 - 検出器からの1.8GB/secの転送を担保する必要

まとめ

- Total costsを最小限に最適化するためにクラウドサービス(アカデミック/コマーシャル)の積極的な活用を視野に入れている
 - データ転送料が鍵
- 所有する計算機システムの効率的利用のために、プライベートクラウド化を推し進める