

# Belleデータ損失 調査委員会報告

山中 卓  
大阪大学

Belleデータ一部損失に関する検証委員会  
2013年2月15日@東京

2

## 委員会

- \* 山中卓 阪大 J-PARC実験 (委員長)
- \* 坂本宏 東大 ATLAS実験大型計算機サーバー
- \* 堺井義秀 KEK Belle実験
- \* 佐々木節 KEK 計算科学センター

# 方針

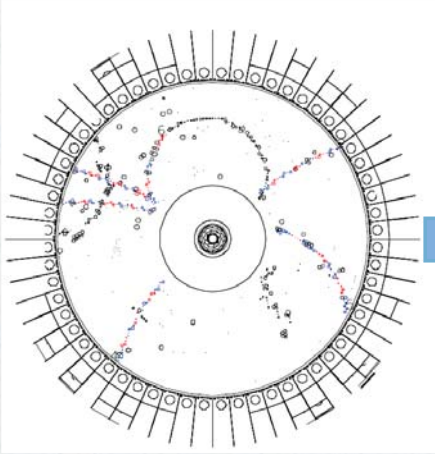
- \* 何がなぜ起きたのか、根本から調べる
- \* 同様な事故の再発を防ぐため

# 調査経過

- \* 2012年9月26日
  - \* Belleと計算機科学センターから報告
  - \* 議論
- \* 2012年10月30日
  - \* 担当SEにインタビュー
  - \* その他
    - \* ファイルの解析、プログラムの試行
- \* 2012年11月6日
  - \* 報告書提出

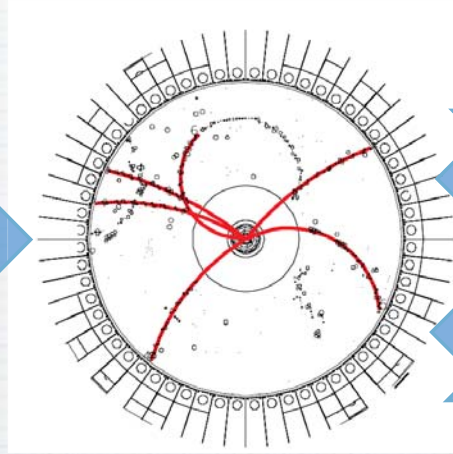
# データの種類

生データ



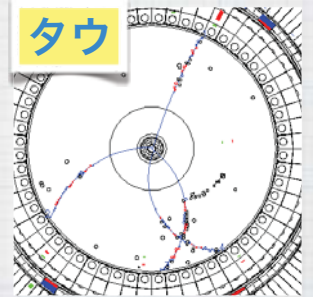
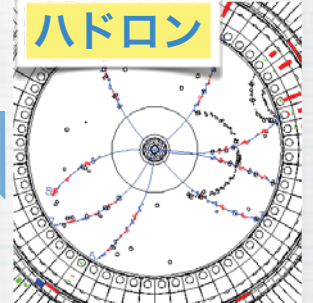
粒子の通過位置、  
時間など

all-mdst



粒子の運動量など

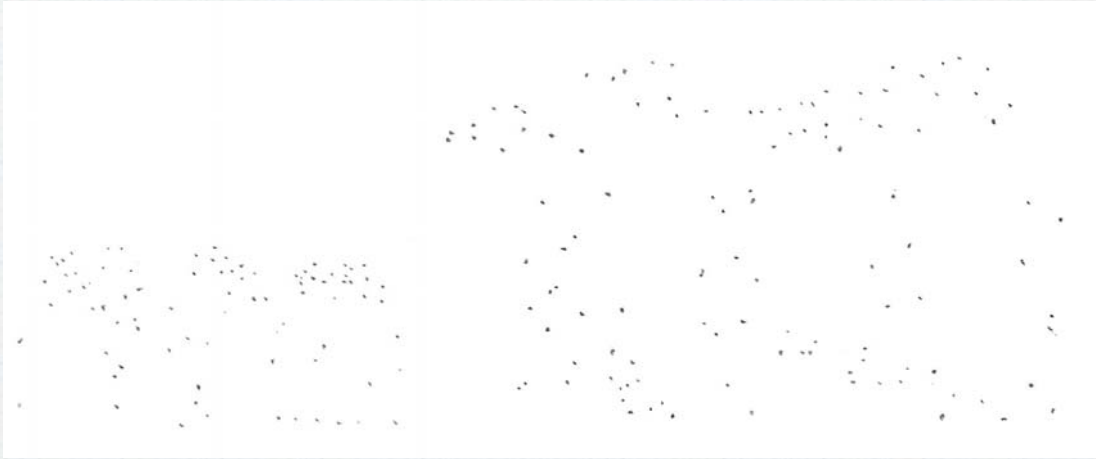
skim-mdst



2012年12月 堺井氏のスライドより

# データの種類の たどる

# 生データ = 点



# all-mdst = 線と測定値



# skim mdst = 特定の種別を選別

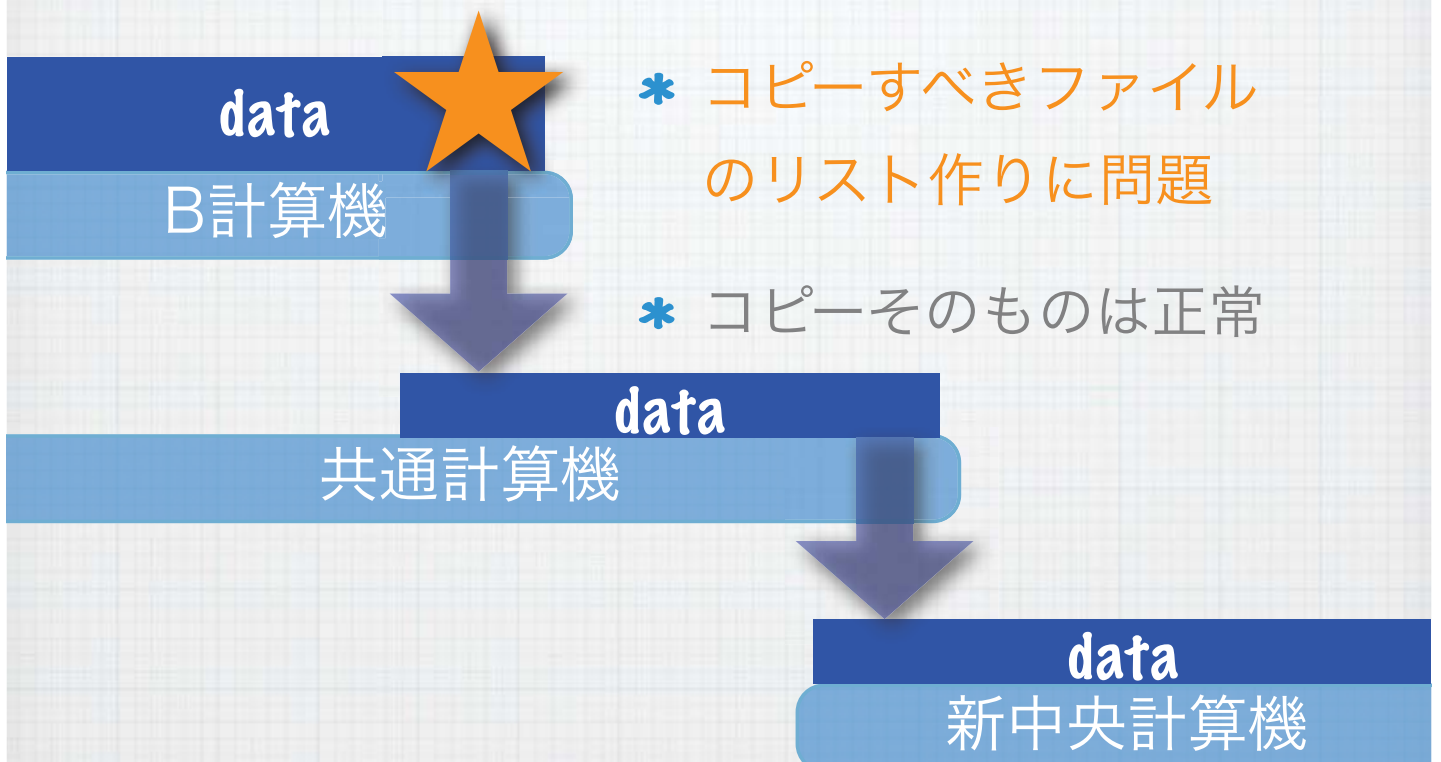
子どものペンギン



大人のペンギン



## 計算機とデータの移行



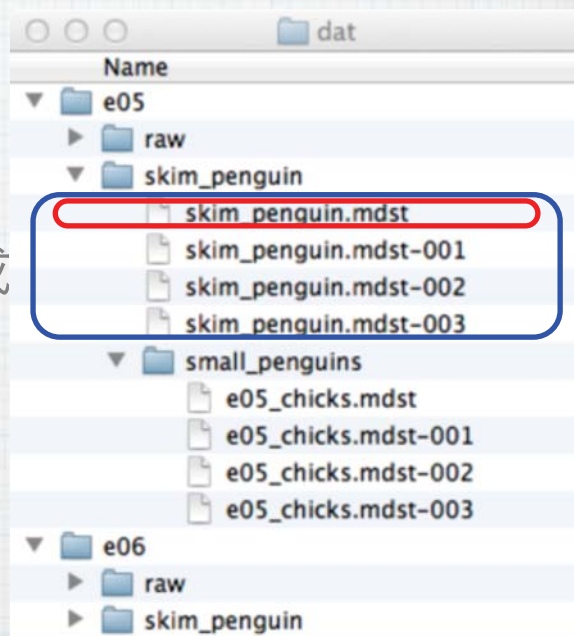
# 何が起きたのか

- \* Belle: コピーするフォルダーを指定



e05/skim\_penguin

...

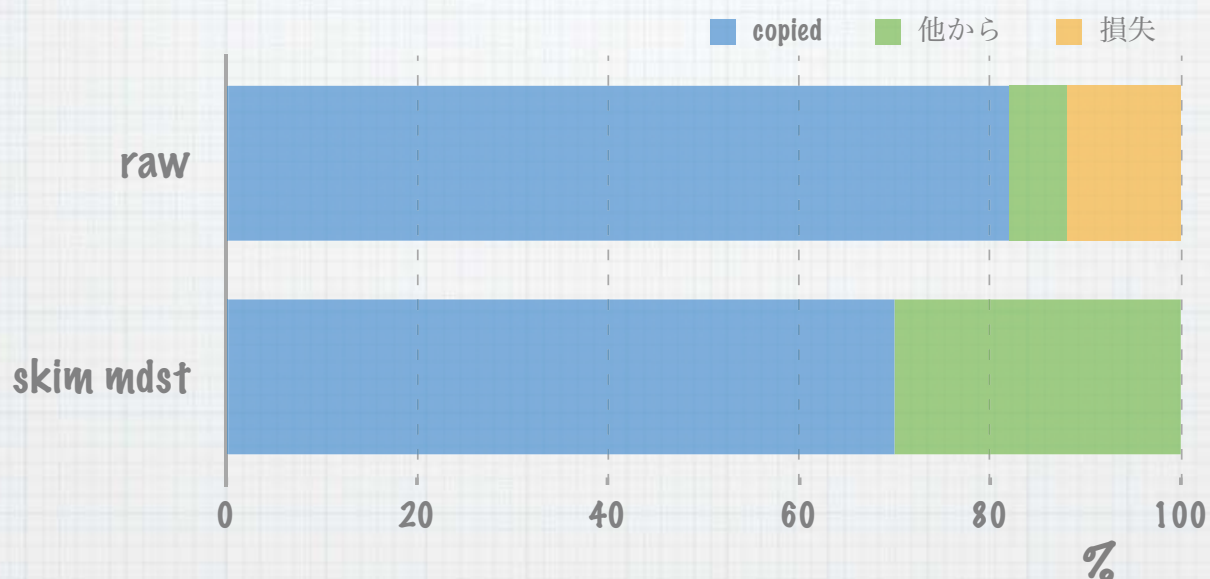


- \* 業者: ファイルのリストを作成

- \* コピーすべきファイル

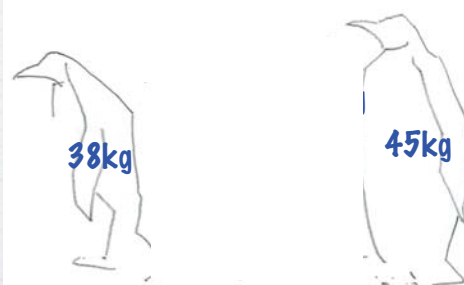
- \* 作られたリスト

# コピーされた割合(容量)



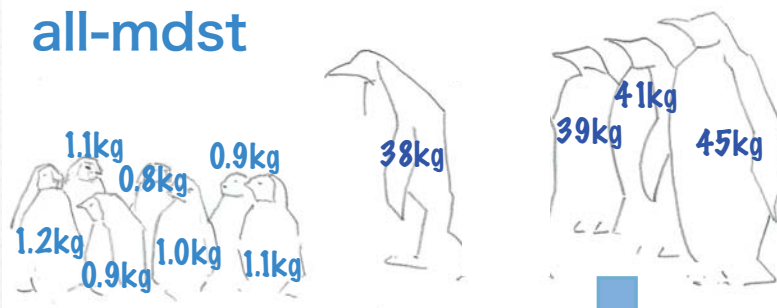
# 最終的 損失

skim-mdst

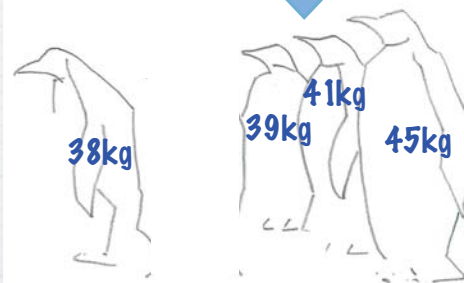


# 最終的 損失

all-mdst



skim-mdst



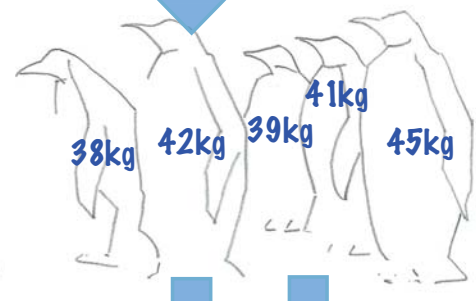
# 最終的 損失

=どのレベルにも  
残っていない  
事象

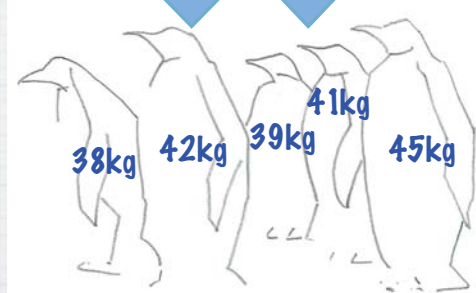
生データ



all-mdst



skim-mdst



14

## 最終的な損失

- \* ypipi 以外 : 5%
- \* ypipi (scan) : 7%
- \* ypipi (Y(5s)) : 2.5%
- \* 詳しくは、山内氏の発表に



# 原因

- \* 技術的な問題
- \* 体制の問題
- \* 計算機の切り替え方の問題

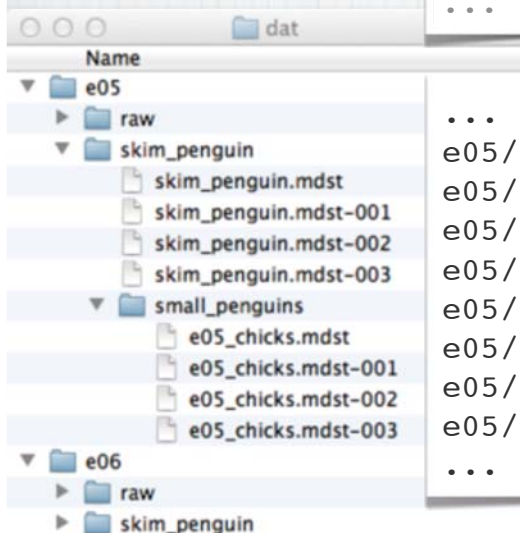
## 技術的な問題

### コピーすべきフォルダーのリスト

```
...
e05/raw
e05/skim_penguin
e06/raw
e06/skim_penguin
...
```

### 全ファイルのリスト

```
...
e05/skim_penguin/skim_penguin.mdst
e05/skim_penguin/skim_penguin.mdst-001
e05/skim_penguin/skim_penguin.mdst-002
e05/skim_penguin/skim_penguin.mdst-003
e05/skim_penguin/small_penguins/e05_chicks.mdst
e05/skim_penguin/small_penguins/e05_chicks.mdst-001
e05/skim_penguin/small_penguins/e05_chicks.mdst-002
e05/skim_penguin/small_penguins/e05_chicks.mdst-003
...
```



# 技術的な問題

```

...          コピーすべきフォルダーのリスト
e05/raw
e05/skim_penguin
e06/raw
e06/skim_penguin
...

```

フォルダー名を抜き出し、比較 全ファイルのリスト

```

e05/skim_penguin/skim_penguin.mdst
e05/skim_penguin/skim_penguin.mdst-001
e05/skim_penguin/skim_penguin.mdst-002
e05/skim_penguin/skim_penguin.mdst-003

```

一記号などが名前に入ったファイルが選ばれなかった

# 技術的な問題

- \* 複雑な処理はしていたが、  
方法に過ちは見つからない
- \* **フォルダー名を抜き出す** コマンドが  
正しく動いていなかった可能性はある

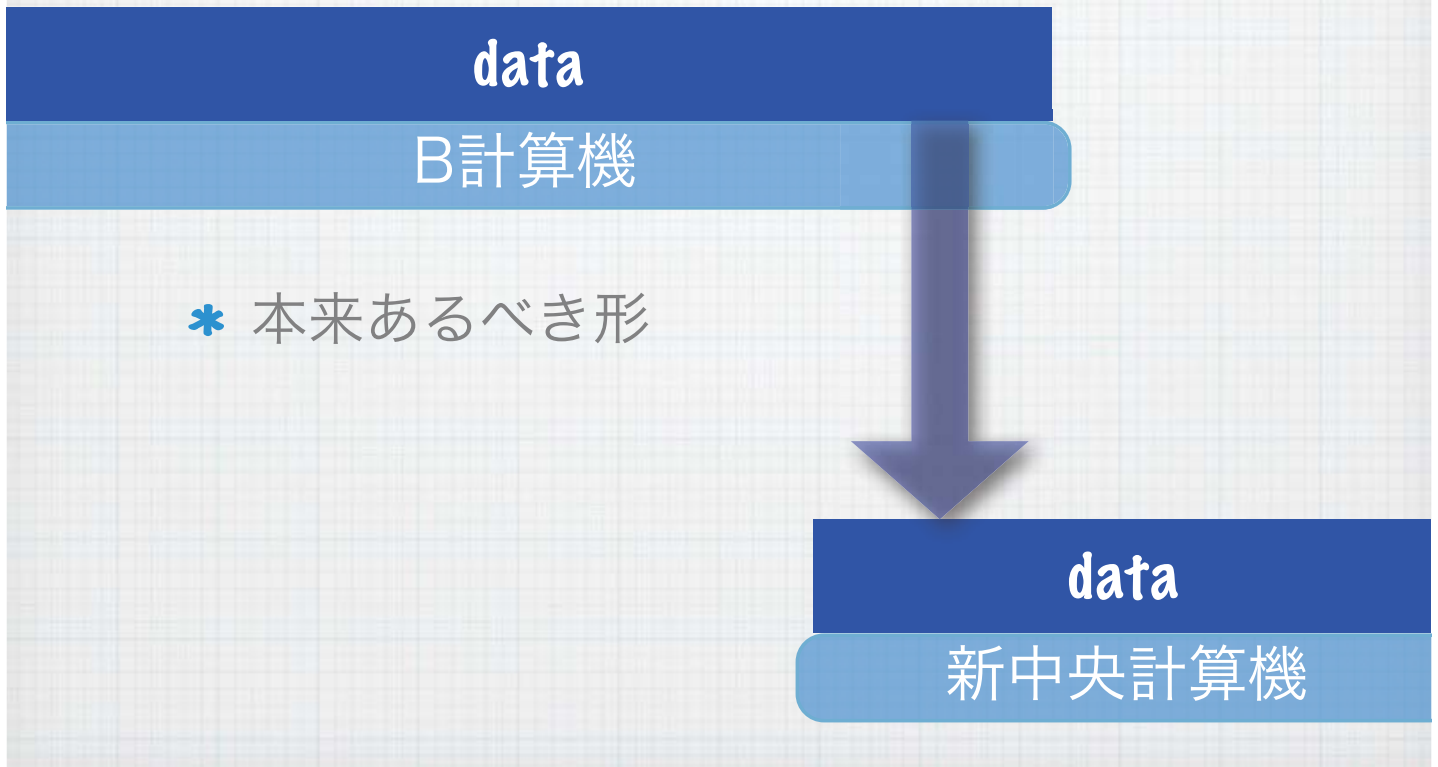
# チェック体制

- \* プログラムにバグはつきもの
- \* SE、Belle側によるファイルリストのチェックが不十分
- \* Belle側もコピーされるファイルの数を把握していない

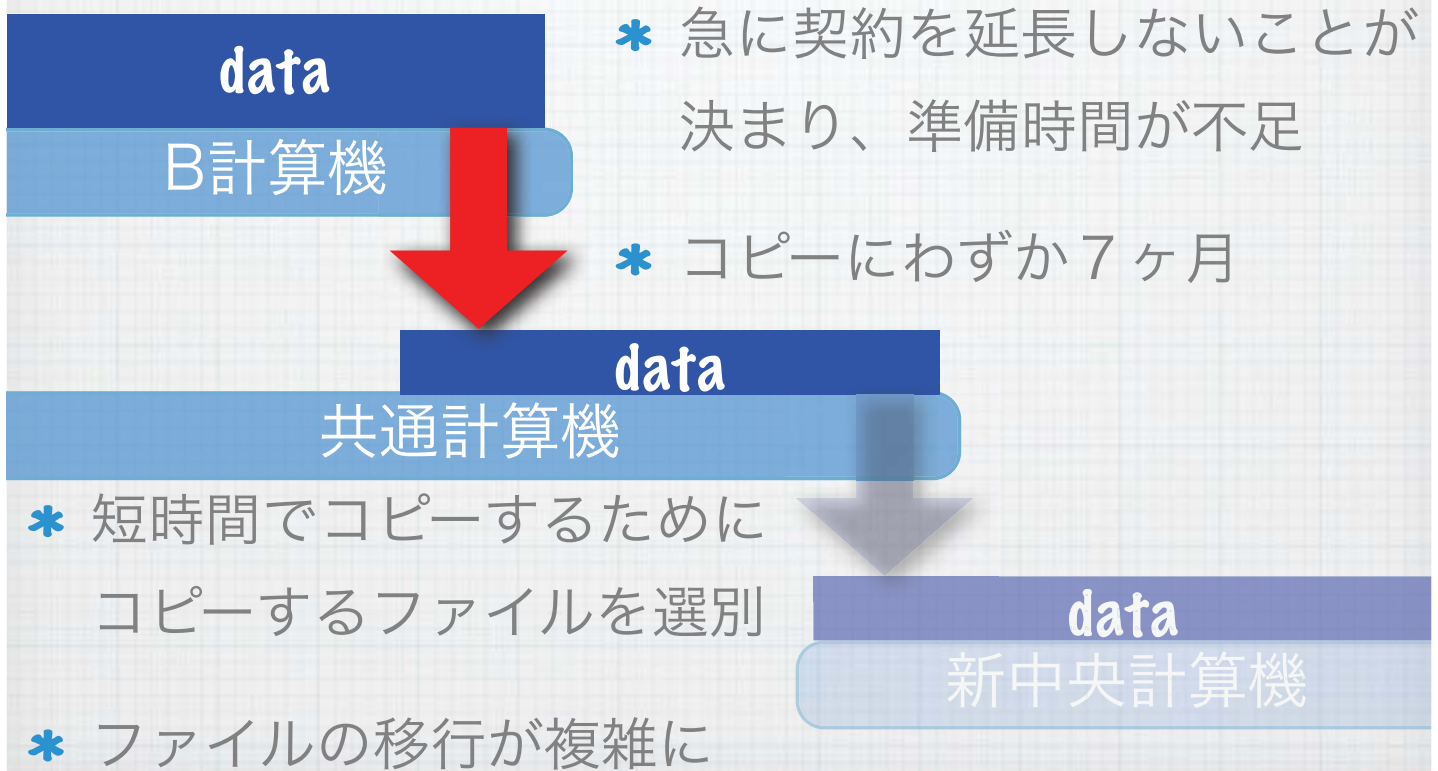
# 責任体制

- \* 契約書
  - \* ... データの後継システムへの移行に協力すること。
- \* あくまでKEK側の作業
- \* ファイルのチェックやコピーの責任は  
計算科学センター? Belle?

# 新旧の計算機に オーバーラップがない



# 時間の問題



# 金と人

- \* Belle-IIの予算確保のために、綱渡り的な  
計算機切り替え
- \* データ移行に深くまで注意を払ったKEK  
の人がいなかった

# 同様の事故を 繰り返さないために

- \* 当事者以外のチェック
- \* 十分な人員と時間
- \* 作業の責任の明確化
- \* データの保護の重要性を認識
- \* 計算機やソフトのための人材と予算の重要性  
を素核研と機構は認識すべき

# まとめ

- \* 人は過ちを犯す
- \* それを前提とした
  - \* チェック体制
  - \* 復旧ルートの確保（命綱）
- \* 機動性と柔軟性は保つべし