LHCb resources updates for 2012-2015

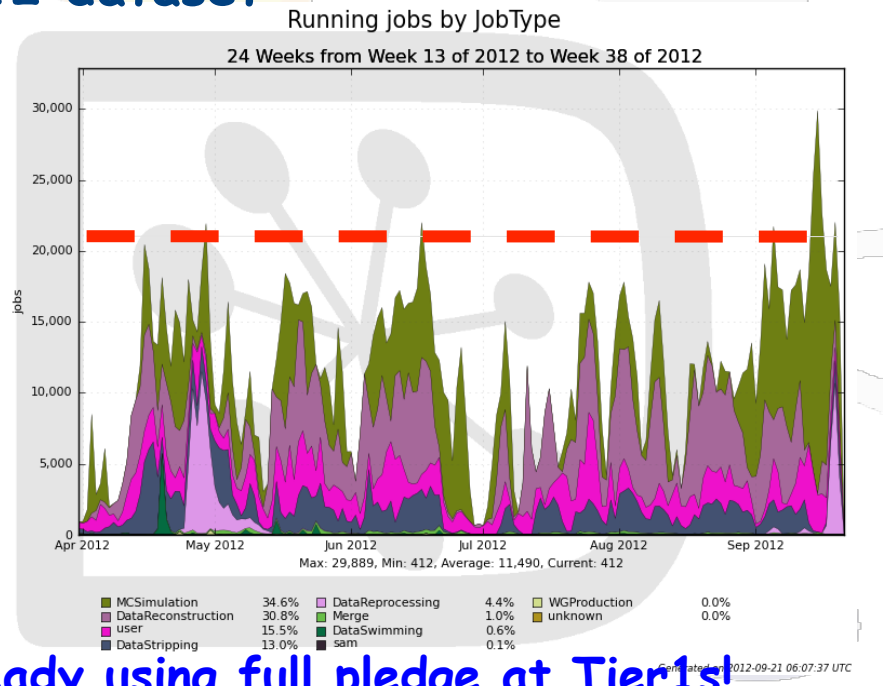
Marco Cattaneo

CERN – LHCb

On behalf of the LHCb Computing Group

LHCb computing production activities in 2012

- Prompt processing (reconstruction and stripping) of new data
- "Swimming" of 2011 stripped data
- MC production
  - Large samples for analysis of 2011 data
  - Preliminary samples for 2012 data
- Reprocessing of complete 2012 dataset
  - Started 17th September
- Resources OK until now:
  - CPU usage 50% of pledge
    - As expected for first half of year
      - now ramping to >100% with reprocessing
  - 2012 disk pledge ~sufficient also for reprocessed data
    - Active data management
  - Major shortage of tape, already using full pledge at Tier1s!
    - Pledge based on 2012 request, assuming shorter run
    - Major problem for the future (rest of 2012, but also 2013)



Changes since 2012 request

- New reconstructed data format:
  - Previously, Reco output (SDST) did not contain copy of RAW
    - Both RAW and SDST files had to be staged, synchronously, when restripping
    - Very demanding on tape systems, since in general on different tapes
    - Does not allow to achieve maximum throughput
  - Now write FULL.DST, containing RAW+SDST
    - Effectively an extra copy of RAW on Tape
- Data taking changes
  - Extension of pp run by 2 months
  - LHCb operation at higher instantaneous luminosity
    - 4×10^{32} $cm^{-2}s^{-1}$ (3.7 in 2011)
  - Higher HLT output rate
    - (2011: 3kHz, 2012 forecast: 4.5kHz, 2012 actual: 5kHz)
  - Participation in pA (heavy ion) run in early 2013

Effect of changes on computing resources

- RAW data: 1.7 PB
  - 43% increase with respect to previous estimates (1.2 PB).
    - Two copies on tape
- Reconstructed data (FULL.DST):
  - 3.1 PB, 120 % increase with respect the previous estimate (SDST, 1.4 PB)
    - One copy on tape
- Derived datasets (DST, MDST for physics analysis)
  - 43% increase
    - disk resident, plus tape archive
- Heavy Ion RAW data: 100 TB
- Higher CPU peak power for reprocessing
  - Originally foreseen after data-taking, now in parallel to be ready for Winter 2013 conferences
    - Clashes with prompt processing (data quality, fast analyses)
    - Cannot use resources of HLT farm

Mitigation to fit into 2012 pledges

- CPU:
  - Reduce prompt processing to 50% of new data
    - Only at CERN, 20% sub-contracted to Tier2s
    - Only for data quality etc., no physics analyses using these data
  - Reprocessing extended by two months
    - Only at Tier1s, 20% sub-contracted to Tier2s
    - Only ~60% of 2012 data available for Winter 2013 conferences
- Tape:
  - Removed NOW all output of 2012 prompt reconstruction
    - Needed for restripping
      - No new analyses until reprocessing completed
  - Reduce all archives (data preservation) to one copy
- Disk:
  - Aggressive reduction of 2012 prompt DST/MDST copies
    - In parallel with appearance of reprocessed data
  - New distribution policy taking into account unused pledge
    - In practice, replicas of reprocessed data will go mostly to CERN, RAL, CNAF, GRIDKA

Storage: forecast for March 2013

- Disk OK but big imbalance between sites (backup slide)

Disk	Pledge	2012	
	PB	PB	%
Tier0	3.5	3.5	36
Tier1	7.3	6.3	64

- Serious shortfall in tape, no solution yet
  - Already cleaned up all 2012 prompt SDST to fit in existing tape
  - No room at Tier1s for second copy of new RAW, and for ONLY copy of FULL.DST

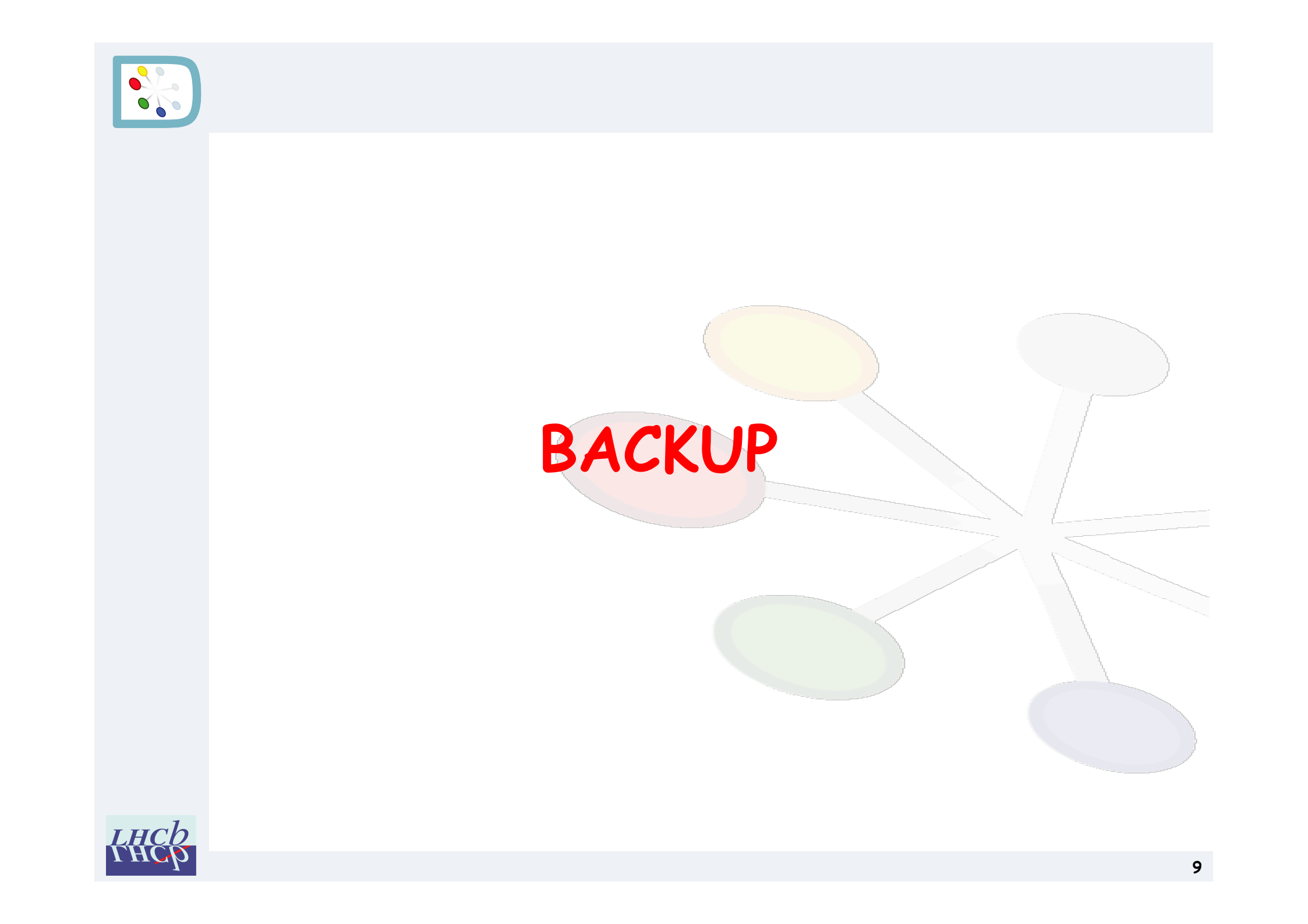
Current Tape (PB)	Tier0	Tier1
RAW	2296	2020
SDST	152	1025
Archive	1544	2251
Total Current	**3992**	**5296**
Tape March 2013 (PB)		
RAW	3096	2820
FULL.DST	1350	4449
Archive	1675	2775
Total March 2013	**6121**	**10044**
Pledges	**6400**	**5324**

2013-2014

- Activities during 2013-2014
  - Several restrippings, ~every 6 months
    - ~ neutral in disk usage since older strippings get replaced
    - moderate increase in tape (archive copy)
  - Full reprocessing of 2011-2012 data, during 2014
    - Ultimate reprocessing of this dataset
    - Duration stretched to fit within existing CPU resources
    - ~neutral on tape (FULL.DST replaces previous version)
  - Continuing MonteCarlo production at constant rate
    - Higher rate than previous estimates to match increased real data sample
      - But no increase in Tier2 CPU request, use HLT farm instead
    - Linear increase in disk and tape
- Requests for 2013-2014 (details in backup slide)
  - CPU: 20% increase at Tier1 in 2013, otherwise flat
  - Disk: 20% increase per year at Tier1, 25% per year at CERN
  - Tape: 100% increase in 2013 at Tier1, 10% in 2014
    - Tape increase actually needed ASAP, April 2013 is too late.

Resources after LS1

- Date of startup for physics, and number of physics days in 2015
  - LHCb nominal luminosity reached rather quickly (lumi levelling). Assume:
    - Any data taking before April 2015 has insignificant impact on 2014 resources
    - $5*10^6$ seconds in 2015, at nominal data rates (i.e. 'normal' year)
- Changes in data-taking rate (not precisely predictable yet)
  - Average multiplicity of pp collisions
    - Higher than at 8 TeV
  - Effect of pileup on RAW size
    - With 25ns bunch spacing, lower multiple interaction rate but greater spillover from out of time events
  - Operating luminosity at LHCb IP
    - Considering up to 50% higher than in 2012, may lead to higher HLT rate
  - HLT bandwidth allocation to different physics channels
    - To be studied during LS1, may lead to significant changes in rate
- Assume all above lead to doubling of data rate (MB/s) from HLT
  - Currently limited by DAQ to 700 MB/s
  - Doubles also all derived (reconstruction, stripping) data formats
  - CPU time scaled with data volume
- 2015 equivalent to 2011+2012 combined.
  - Resources needs are ~doubled compared to 2014 (see backup slide)


BACKUP

Resources imbalance

- Evolution of disk pledges at Tier1s

Disk pledge	FR-CCIN2P3	DE-KIT	IT-INFN_CN/	NL-T1	ES-PIC	UK-T1-RAL
2011 (PB)	1010	600	450	757	244	651
2012 (PB)	1090	1610	1400	810	485	1767
% change	**8%**	**168%**	**211%**	**7%**	**99%**	**171%**
T1 Disk share						
2011 (%)	27.2%	16.2%	12.1%	20.4%	6.6%	17.5%
2012 (%)	15.2%	22.5%	19.5%	11.3%	6.8%	24.7%

- Changes balance of resources among sites
  - Preferentially clean old data from sites below share
  - Available space used as weight for placing replicas of new data
- E.g. 2012+2011 reprocessed data distribution will be:

DST distribution at Tier1s (TB)						
CERN	CNAF	GRIDKA	IN2P3	PIC	RAL	SARA
614	**455**	**491**	**113**	**53**	**450**	**111**

- Consequence: analysis of 2011 and 2012 reprocessed data will not use all Tier1s according to their CPU share

2013-2014 requests (LHCb-PUB-2012-014)

Power	Pledge 2012	2012		2013		2014	
	kHS06	kHS06	%	kHS06	%	kHS06	%
Tier0	34	34	18	34	16	34	16
Tier1	91	110	58	110	52	110	52
Tier2	47	46	24	46	22	46	24
HLT farm				20	10	20	10

Table 5-2: Estimated CPU power needed at the different Tier levels.

Disk	Pledge	2012		2013		2014	
	PB	PB	%	PB	%	PB	%
Tier0	3.5	3.5	36	4.4	34	5.5	35
Tier1	7.2	6.3	64	8.6	66	10.4	65

Table 5-5: Resulting Disk estimates for each Tier level.

Tape	Pledge	2012		2013		2014	
	PB	PB	%	PB	%	PB	%
Tier0	6.4	6.2	38	6.5	38	7.3	38
Tier1	5.3	10.0	62	10.8	62	11.9	62

Table 5-6: Resulting Tape estimates for each Tier level.

- N.B. 2012 column is what is needed to complete processing of 2012 data, after all possible mitigations

First estimate of resources in 2015 (LHCb-PUB-2012-015)

Power	2015	
	kHS06	%
Tier0	87	22
Tier1	232	58
Tier2	80	20

Table 3-1: Estimated CPU power needed during 2015.

Disk	2015	
	PB	%
Tier0	10.8	34
Tier1	21.1	66

Table 3-2: Estimated Disk Storage requirement for 2015.

Tape	2015	
	PB	%
Tier0	16.4	43
Tier1	21.3	57

Table 3-3: Estimated Tape Storage requirement for 2015.