Estimated 2011 & 2012 data volume

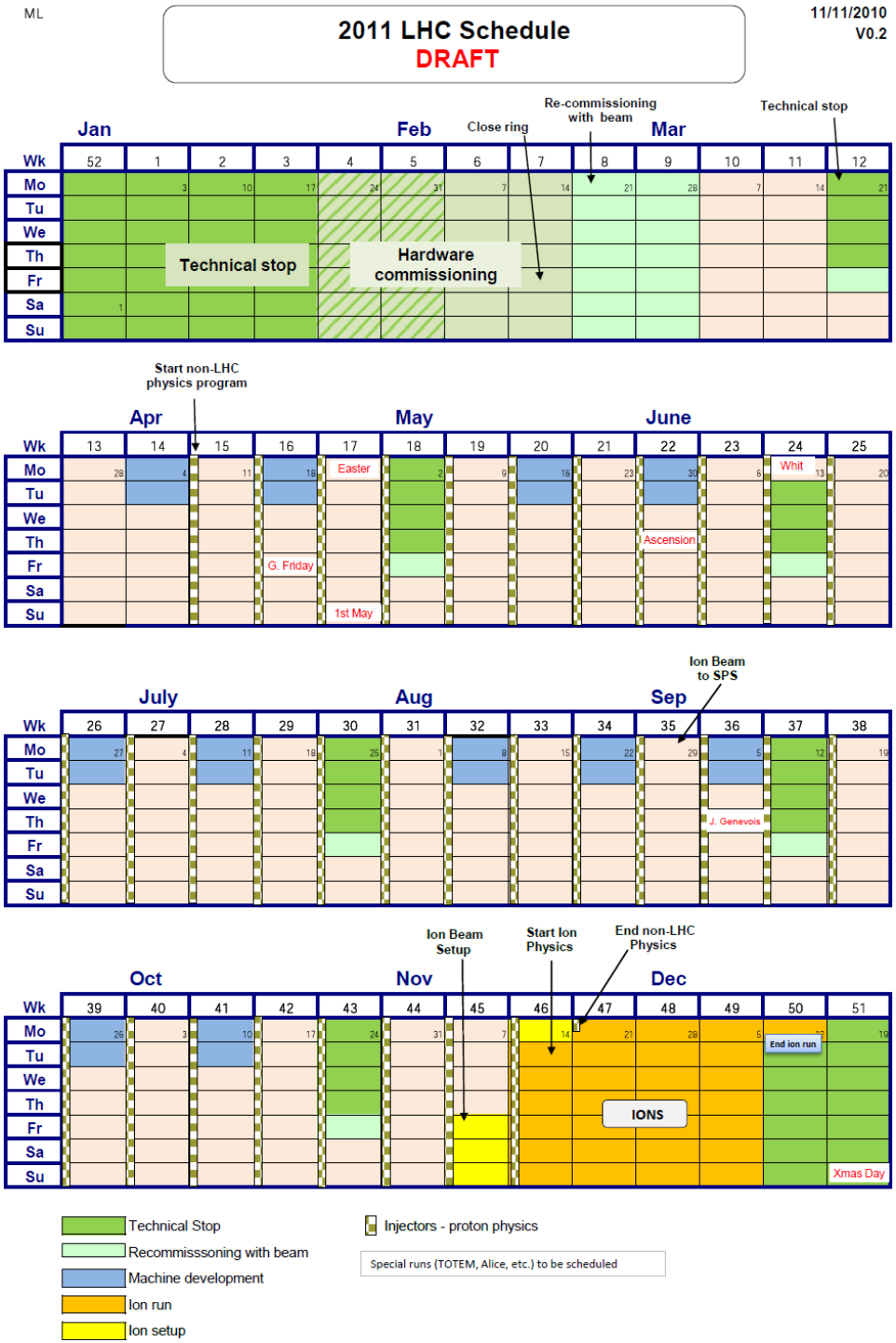
(*for the moment only data, not MC)*

For Napoli

February 3 – 5, 2011

(new version with BNL SRM values corrected)

Kors Bos



2011

- Beam back around 21st February
- 2 weeks re-commissioning with beam (at least)
- 4 day technical stop every 6 weeks
- Count 1 day to recover from TS (optimistic)
- 2 days machine development every 2 weeks or so
- 4 days ions set-up
- 4 weeks ion run
- End of run – 12th December

~200 days proton physics
~28 days HI physics

T1 resources & pledges

Summary Ext. Tier1s	2010	2011	2012	Split 2011	ALICE	ATLAS	CMS	LHCb	SUM 2011
CPU (HEP-SPEC06)	402416	524284	540807	Offered	71471	250208	132173	70432	524284
				Required	117000	226000	150700	65000	558700
				Balance	-39%	11%	-12%	8%	-6%
Disk (Tbytes)	43577	52467	59646	Offered	5527	26869	16254	3817	52467
				Required	7900	24800	19500	3500	55700
				Balance	-30%	8%	-17%	9%	-6%
Tape (Tbytes)	50570	88297	115501	Offered	8013	[illegible]	44392	3933	88297
				Required	13000	30100	[illegible]2400	3470	98970
				Balance	-38%	6%	[illegible]%	13%	-11%

Ext. Tier1 Requ. 2011	ALICE	ATLAS	CMS	LHCb	SUM
CPU (HEP-SPEC06)	117000	226000	150700	65000	558700
Disk (Tbytes)	7900	24800	19500	3500	55700
Tape (Tbytes)	13000	30100	52400	3470	98970

TIER 1 Notes

Note 1: France. The final budget allocation will only be known in January 2011, depending on the outcome there is a risk that these pledges may decrease.

Note 2: Netherlands. These pledges are unconfirmed due to incomplete negociations with the Funding Agency. The current assumption is that they will be accepted before end 2010 enabling procurement to commence however the compute and disk resource deployment will be delayed, available by summer 2011.

Note 3: UK. Tape is provisioned on demand. The full pledge will not be deployed until required.

T2 resources & pledges

Summary Tier2s with Split in 2011	2010	2011	2012	Split 2011	ALICE	ATLAS	CMS	LHCb	Sum 2011
CPU (HEP-SPEC06)	502367	725324	776203	Offered	80932	281228	315202	47962	725324
				Required	121000	278000	319500	36000	754500
				Balance	-33%	1%	-1%	33%	-4%
Disk (Tbytes)	39255	60454	71998	Offered	5738	34203	20219	295	60454
				Required	6600	37600	19900	20	64120
				Balance	-13%	-9%	2%	1377%	-6%

Requirements 2011	ALICE	ATLAS	CMS	LHCb	SUM
CPU (HEP-SPEC06)	121000	278000	319500	36000	754500
Disk (Tbytes)	6600	37600	19900	20	64120
Number of T2s	17	39	33	16	

TIER 2 Notes

Note 1 - Austria: The pledge for 2011 had to be reduced due to infrastructure limitations which will be addressed by relocating the computing facility at a new site.

Note 2 - Australia: numbers are currently constrained by data centre power and cooling limitations. New DC is due for completion early mid-2011. Numbers may be increased.

Note 3 - Belgium: the FNRS pledges for 2011 are subject to approval of the funding request for 2011 by the FNRS funding agency

Note 4 - Brazil: The HS06 of the processors was overestimated therefore the 2010 pledge was not met. The 2011 pledge reflects the current installation while waiting for agreement on funding requests to finance the 2012 pledges.

Note 5 - Canada : These pledges are unconfirmed due to the timing of the Call for Proposals process in Canada. The current assumption is that the proposal will be accepted and all new resources will be available by 1 April 2011.

Note 6 - France: All T2's - the final budget allocation will only be known in January 2011; depending on the outcome there is a risk that these pledges may decrease.

Note 7 - France LPC Clermont: Local resource funding was lower than expected in 2010 therefore the 2011 pledge was adjusted accordingly.

Note 8 - Korea KISTI: the 2010 disk pledge was not met and due to funding uncertainty the 2011 pledge has been reduced to match the disk resources currently deployed.

Note 9 - Spain LHCb: The 2011 pledge is based on the decreased CPU requirement from LHCb that this site has agreed to provide 6.5% of, which translates to 2340 HS06.

Note 10 - Sweden: The Swedish Research Council confirmed the 2011 Tier2 pledges at its meeting on 4th November 2010

Note 11 - Ukraine: The 2011 pledge values correspond to resources already deployed or planned to be deployed by April 2011. The 2012 pledges will depend on budget allocation to be confirmed during 2011.

T1 disk resources

[PB]	2011 pledged	now available	2012=2010+30%
Total	24.8	26.9	36 (?)
55% for data	9.9	14.8	19.8
30% for MC	7.4	8.1	10.8
15% for users & groups & buffers	7.4	4.0	5.4

- Not much more can be expected for 2011
  - We already have more than we asked
- In 2011 we've got a 28% increase relative to 2010
  - Can we expect a 30% increase for 2012 ?

The split right now (29/01/11)

- Data 55%
- MC 31%
- *Data & MC will be combined*
- User etc. 4%
- Buffers 10%

TIER1s

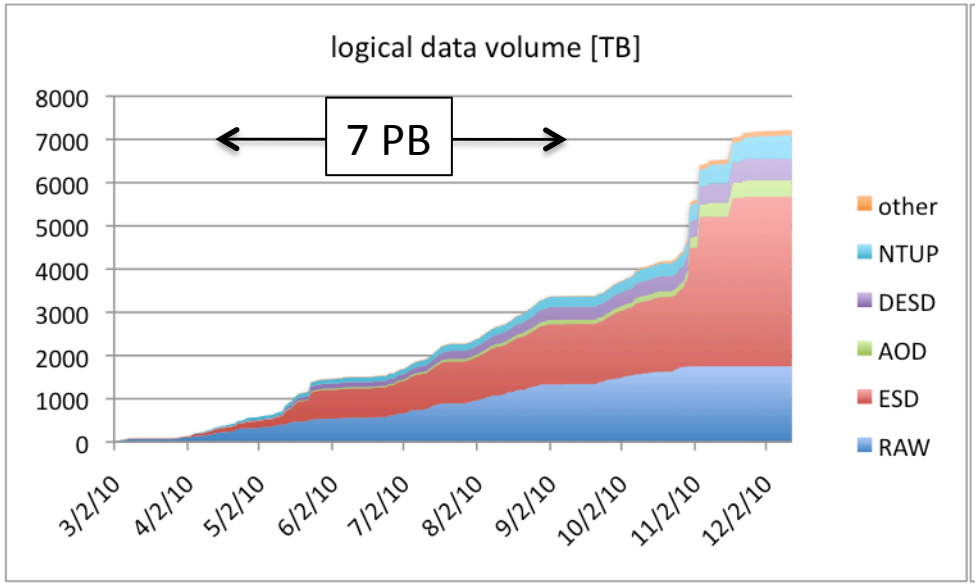
SPACETOKEN	FREE(TB)	USED(TB)	TOTAL(TB)
DATADISK	3175	11752	14927
DATATAPE	533	139	672
GROUPDISK	434	531	965
HOTDISK	170	28	198
LOCALGROUPDISK	133	159	292
MCDISK	1866	6716	8583
MCTAPE	214	113	327
PPSDATADISK	0	0	0
PRODDISK	47	5	53
SCRATCHDISK	386	559	945
USERDISK	20	509	529
TOTAL	6980	20510	27490

2010 data

Stats derived from the DDM database

***Logical Volume** means without replication*

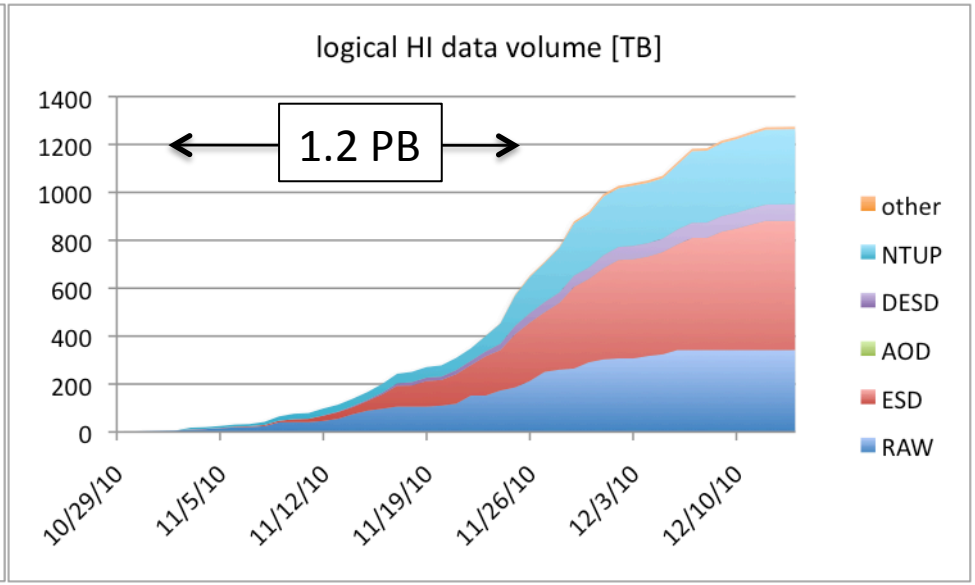
pp data (version n and n-1)



- 1.8 PB RAW on tape
- 5.5 PB on disk

RAW	ESD	AOD	DESD	NTUP	other
1751	3920	384	498	543	120

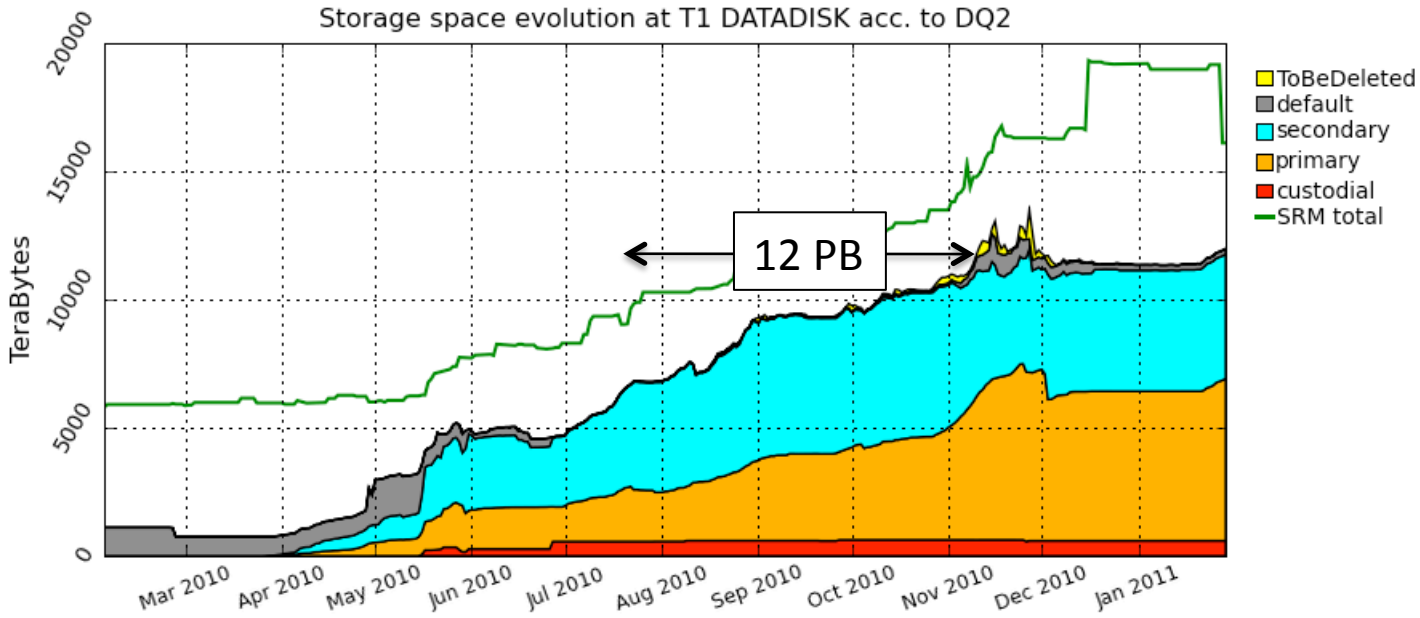
HI data (only 1 version)

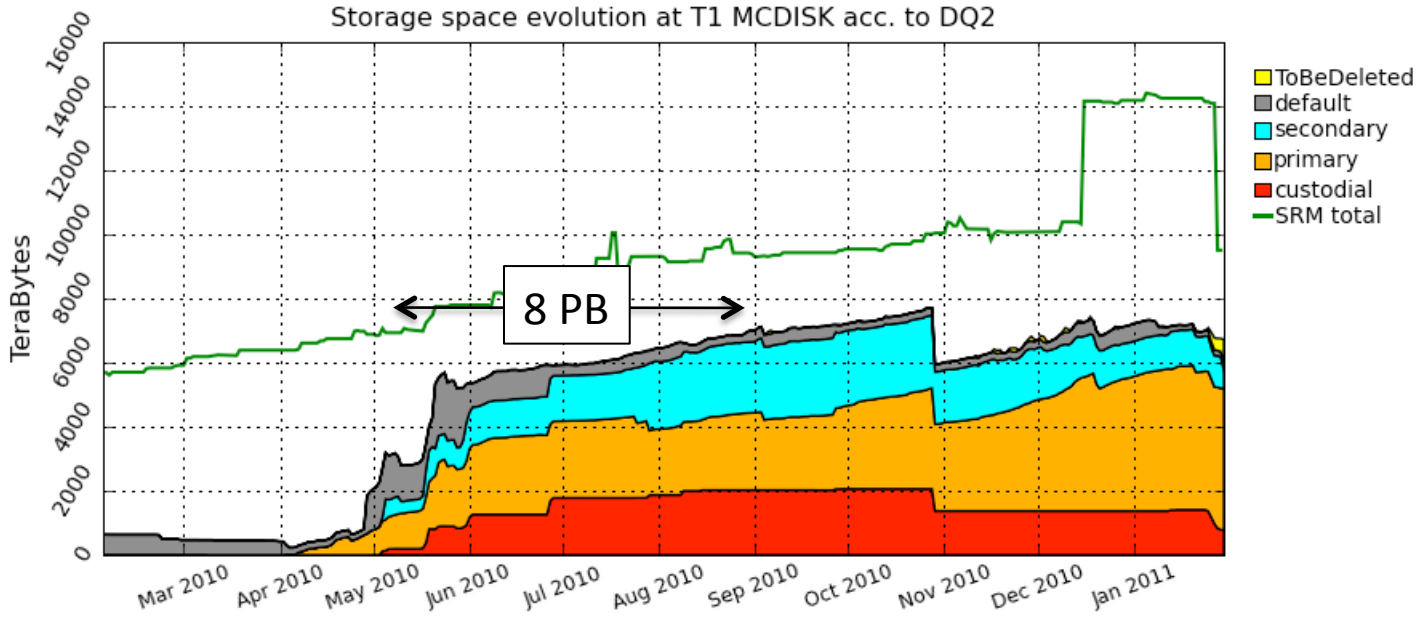


- 0.3 PB RAW on tape (20%)
- 0.9 PB on disk (24%)

RAW	ESD	AOD	DESD	NTUP	other
342	539	0	70	313	10

T1 disk including all replica's





Nota bene 2010 data

- The amount of data on disk now is measured using the central catalogue so only registered data is taken into account. RAW data is not considered anywhere as it is assumed to only go to tape in the T0 and the T1s.
- The fraction is calculated of the total 2010 volume for DESD and AOD. This is supposed to be 1 according to the computing model. It is actually 1.3. This fraction is then used to predict how much DESD there will be produced in 2011 and 2012. The same is done for NTUP and “other”. This is an optimistic guess in case ESD will be suppressed because that will cause more and different DESD to be made available.
- Heavy Ion data is listed separately. There is no AOD produced. It is not yet re-processed, so there is only 1 version. Moreover only 1 copy of the data exists among the T1s. Just as above the 2010 fraction of DESD and ESD is calculated and used to predict how much DESD will be produced in 2011. The same for NTUP and others.
- HI data is supposed to be 10% of our pp data. It can be seen that it is actually 20% for RAW and 17% for all derived data. This is due to the fact that we took data at a much higher trigger rate than the foreseen 60 Hz in the computing model. For 2011 the trigger rate for HI has been set to 200 Hz. This way almost a factor 2 less will be produced than in 2010 assuming the same event size. This may change as it expected we will be able to trigger better on central events that are much bigger.
- It is assumed that HI data will be re-processed and version n and n-1 will be retained. It is still assumed there is only 1 copy of all HI data.

Nota bene 2011 data

- Various scenario's are shown using a simple spreadsheet. More scenario's can easily be calculated. The trigger rate and the fraction of ESD kept on disk is varied by varying the event size of the ESD. ESD for HI can be treated separately.
- After re-processing the data in the summer and in the fall version n and n-1 are supposed to be kept.
- The model assumes 2 copies of all derived data on T1 disk. This is needed to correct for lost files. Moreover 1 copy of everything in the T1s is thought to be not sufficient to efficiently do analysis in all T2s on 3 continents. The limited transatlantic bandwidth would make PD2P dynamic data placement inefficient.
- No derived data is pre-placed in the T2s any more, all data is pulled dynamically when needed. Pulling data is mostly done from the T1 and increasingly also from other T2s when the network allows this. It can be expected that more and more across-cloud T2 – T2 data transfer will be possible in the course of the next 2 years as the new LHCONE network will be rolled out.
- At this moment most of the 2011 T1 disk resources are already in place and only a small increase may still be expected. Three quarter of those resources are already used for primary data and 2011 data taking still has to start

Nota bene 2012 data

- It is assumed that 2012 will be a year identical to 2011. This is conservative as operational experience may make the LHC machine more efficient than 30%.
- It is assumed that the disk resources will increase by 30% compared to 2011. In 2011 the disk resources increase by 28% compared to 2010. This may be optimistic as several funding agencies have already announced cuts.

Default, Scenario 0: 200 days, 30%, 200 Hz

2011

2011 14 installed | 2012 20 requested

running days	**200** pp		**28** HI	
overall eff	**30%**		**30%**	
Trigger rate	**200** Hz		**200** Hz	

pp		original versio in 2011 [PB]	2 versions in 2011 [PB]	2 copies 2 versions in 2011 [PB]	2 copies 2 versions 2010+2011 [PB]	pp + HI 2 copies (pp) 2 versions 2010+2011 [PB]	pp + HI 2 copies (pp) 2 versions 2010+'11+'13 [PB]
RAW size	1.4 MB (to tape)	1.5	1.5	1.5	3.2	3.8	5.4
ESD size	**1.48** MB (was 1.48)	1.5	3.1	6.1	10.1	11.7	18.4
AOD size	0.18 MB	0.2	0.4	0.7	1.1	1.1	1.9
DESD size	*1.3 x AOD*	0.2	0.5	1.0	1.5	1.7	2.7
NTUP size	*1.4 x AOD*	0.3	0.5	1.1	1.6	2.6	4.0
other	*0.3 x AOD*	0.1	0.1	0.2	0.4	0.4	0.6
sum on disk		*2.3*	*4.6*	*9.1*	*14.6*	*17.5*	*27.6*

HI		original versio in 2011 [PB]	2 versions in 2011 [PB]	1 copy 2 versions 2010+2011 [PB]
RAW size	1.48 MB (to tape)	0.2	0.2	0.6
ESD size	**2.01** MB (was 2.01)	0.3	0.6	1.7
AOD	0 MB	0.0	0.0	0.0
DESD	*0.1 x ESD*	0.0	0.1	0.2
NTUP	*0.6 x ESD*	0.2	0.3	1.0
other	*0.0 x ESD*	0.0	0.0	0.0
sum on disk		*0.5*	*1.0*	*2.9*

2010

2010 ~11 PB available

	pp 2 versions in 2010 [PB]	fraction of pp AOD	HI 1 version in 2010 [PB]	fraction of HI ESD	HI fraction of pp	(2)pp+(1)HI in 2010 [PB]
RAW	1.8		0.3		0.20	2.1
ESD	**3.9**	(was 3.920)	**0.5**	(was 0.539)	0.14	4.5
AOD	0.4		0.0		0.00	0.4
DESD	0.5	1.3	0.1	0.1	0.14	0.6
NTUP	0.5	1.4	0.3	0.6	0.58	0.9
other	0.1	0.3	0.0	0.0	0.08	0.1
sum on disk (not RAW)	*5.5*		*0.9*		*0.17*	*6.4*

Scenario 1: 200 days, 30%, 400 Hz

2011

	2011 14 installed	2012 20 requested

running days	**200**	pp	**28**	HI				
overall eff	**30%**		**30%**					
Trigger rate	**400**	Hz	**200**	Hz				

pp			original versio in 2011 [PB]	2 versions in 2011 [PB]	2 copies 2 versions in 2011 [PB]	2 copies 2 versions 2010+2011 [PB]	pp + HI 2 copies (pp) 2 versions 2010+2011 [PB]	pp + HI 2 copies (pp) 2 versions 2010+'11+'13 [PB]
RAW size	1.4	MB (to tape)	2.9	2.9	2.9	4.7	5.2	8.3
ESD size	**1.48**	MB (was 1.48)	3.1	6.1	12.3	16.2	17.9	30.7
AOD size	0.18	MB	0.4	0.7	1.5	1.9	1.9	3.4
DESD size	*1.3 x AOD*		0.5	1.0	1.9	2.4	2.7	4.7
NTUP size	*1.4 x AOD*		0.5	1.1	2.1	2.7	3.6	6.1
other	*0.3 x AOD*		0.1	0.2	0.5	0.6	0.6	1.1
sum on disk			*4.6*	*9.1*	*18.3*	*23.7*	*26.6*	*45.9*

HI			original versio in 2011 [PB]	2 versions in 2011 [PB]	1 copy 2 versions 2010+2011 [PB]
RAW size	1.48	MB (to tape)	0.2	0.2	0.6
ESD size	**2.01**	MB (was 2.01)	0.3	0.6	1.7
AOD	0	MB	0.0	0.0	0.0
DESD	*0.1 x ESD*		0.0	0.1	0.2
NTUP	*0.6 x ESD*		0.2	0.3	1.0
other	*0.0 x ESD*		0.0	0.0	0.0
sum on disk			*0.5*	*1.0*	*2.9*

2010

2010
~11 PB available

	pp 2 versions in 2010 [PB]	fraction of pp AOD	HI 1 version in 2010 [PB]	fraction of HI ESD	HI fraction of pp	(2)pp+(1)HI in 2010 [PB]
RAW	1.8		0.3		0.20	2.1
ESD	**3.9**	(was 3.920)	**0.5**	(was 0.539)	0.14	4.5
AOD	0.4		0.0		0.00	0.4
DESD	0.5	1.3	0.1	0.1	0.14	0.6
NTUP	0.5	1.4	0.3	0.6	0.58	0.9
other	0.1	0.3	0.0	0.0	0.08	0.1
sum on disk (not RAW)	*5.5*		*0.9*		*0.17*	*6.4*

Scenario 2: 200 days, 30%, 400 Hz, 10% of ESD

2011

running days	**200**	pp	**28**	HI	
overall eff	**30%**		**30%**		
Trigger rate	**400**	Hz	**200**	Hz	

2011 — 14 installed

2012 — 20 requested

pp		original versio in 2011 [PB]	2 versions in 2011 [PB]	2 copies 2 versions in 2011 [PB]	2 copies 2 versions 2010+2011 [PB]	pp + HI 2 copies (pp) 2 versions 2010+2011 [PB]	pp + HI 2 copies (pp) 2 versions 2010+'11+'13 [PB]
RAW size	1.4 MB (to tape)	2.9	2.9	2.9	4.7	5.2	8.3
ESD size	**0.148** MB (was 1.48)	0.3	0.6	1.2	1.6	1.8	3.1
AOD size	0.18 MB	0.4	0.7	1.5	1.9	1.9	3.4
DESD size	*1.3 x AOD*	0.5	1.0	1.9	2.4	2.7	4.7
NTUP size	*1.4 x AOD*	0.5	1.1	2.1	2.7	3.6	6.1
other	*0.3 x AOD*	0.1	0.2	0.5	0.6	0.6	1.1
sum on disk		*1.8*	*3.6*	*7.2*	*9.2*	*10.5*	*18.3*

HI		original versio in 2011 [PB]	2 versions in 2011 [PB]	1 copy 2 versions 2010+2011 [PB]
RAW size	1.48 MB (to tape)	0.2	0.2	0.6
ESD size	**0.201** MB (was 2.01)	0.0	0.1	0.2
AOD	0 MB	0.0	0.0	0.0
DESD	*0.1 x ESD*	0.0	0.1	0.2
NTUP	*0.6 x ESD*	0.2	0.3	1.0
other	*0.0 x ESD*	0.0	0.0	0.0
sum on disk		*0.2*	*0.5*	*1.4*

2010

2010 — ~11 PB available

	pp 2 versions in 2010 [PB]	fraction of pp AOD	HI 1 version in 2010 [PB]	fraction of HI ESD	HI fraction of pp	(2)pp+(1)HI in 2010 [PB]
RAW	1.8		0.3		0.20	2.1
ESD	**0.4**	(was 3.920)	**0.1**	(was 0.539)	0.14	0.4
AOD	0.4		0.0		0.00	0.4
DESD	0.5	1.3	0.1	1.3	0.14	0.6
NTUP	0.5	1.4	0.3	5.8	0.58	0.9
other	0.1	0.3	0.0	0.2	0.08	0.1
sum on disk (not RAW)	*1.9*		*0.4*		*0.23*	*2.4*

Scenario 3: 200 days, 30%, 400 Hz, no ESD

2011

running days	**200**	pp	**28**	HI
overall eff	**30%**		**30%**	
Trigger rate	**400**	Hz	**200**	Hz

pp		original versio in 2011 [PB]	2 versions in 2011 [PB]	2 copies 2 versions in 2011 [PB]	2 copies 2 versions 2010+2011 [PB]	pp + HI 2 copies (pp) 2 versions 2010+2011 [PB]	pp + HI 2 copies (pp) 2 versions 2010+'11+'13 [PB]
RAW size	1.4 MB (to tape)	2.9	2.9	2.9	4.7	5.2	8.3
ESD size	**0** MB (was 1.48)	0.0	0.0	0.0	0.0	0.0	0.0
AOD size	0.18 MB	0.4	0.7	1.5	1.9	1.9	3.4
DESD size	*1.3 x AOD*	0.5	1.0	1.9	2.4	2.6	4.5
NTUP size	*1.4 x AOD*	0.5	1.1	2.1	2.7	3.3	5.4
other	*0.3 x AOD*	0.1	0.2	0.5	0.6	0.6	1.1
sum on disk		*1.5*	*3.0*	*6.0*	*7.6*	*8.3*	*14.3*

HI		original versio in 2011 [PB]	2 versions in 2011 [PB]	1 copy 2 versions 2010+2011 [PB]
RAW size	1.48 MB (to tape)	0.2	0.2	0.6
ESD size	**0** MB (was 2.01)	0.0	0.0	0.0
AOD	0 MB	0.0	0.0	0.0
DESD	*0.1 x ESD*	0.0	0.0	0.1
NTUP	*0.6 x ESD*	0.0	0.0	0.6
other	*0.0 x ESD*	0.0	0.0	0.0
sum on disk		*0.0*	*0.0*	*0.8*

2010

2010
~11 PB available

	pp 2 versions in 2010 [PB]	fraction of pp AOD	HI 1 version in 2010 [PB]	fraction of HI ESD	HI fraction of pp	(2)pp+(1)HI in 2010 [PB]
RAW	1.8		0.3		0.20	2.1
ESD	**0.0**	(was 3.920)	**0.0**	(was 0.539)	1.00	0.0
AOD	0.4		0.0		0.00	0.4
DESD	0.5	1.3	0.1	0.130	0.14	0.6
NTUP	0.5	1.4	0.3	0.581	0.58	0.9
other	0.1	0.3	0.0	0.019	0.08	0.1
sum on disk (not RAW)	*1.5*		*0.4*		*0.25*	*1.9*

2011
14 installed

2012
20 requested

Lessons learned

scenario	**0**	**1**	**2**	**3**
Rate [Hz]	200	**400**	400	400
ESD [MB]	1.48	1.48	0.148 **(10%)**	**0 (0%)**
2011				
Needed [PB]	17	27	11	8
Available [PB]	15	15	15	15
Diff. [PB]	-2	-12	+4	+7
2012				
Needed [PB]	28	46	18	14
Request [PB]	20	20	20	20
Diff. [PB]	-8	-26	+2	+6

Caution

- It does not take into account other data on disk than 2010 data
- It assumes the 55/30/15% shares for data/mc/users&buffers remains
- It does not allows space for the n-2 version to stay while version n is made
- Disks cannot be filled more than ~80 - 90% for operational reasons
- It does not take into account extra buffer space needed for PD2P

- Have not thought about CPUs yet
- Do we scale down the trigger if the LHC is more than 30% efficient?
- How do we cope with pile-up (increases cpu time)?
- Can we increase the Pt cut to decrease cpu time?

- Have not thought about tapes yet
- Nor about network and disk I/O bandwidth

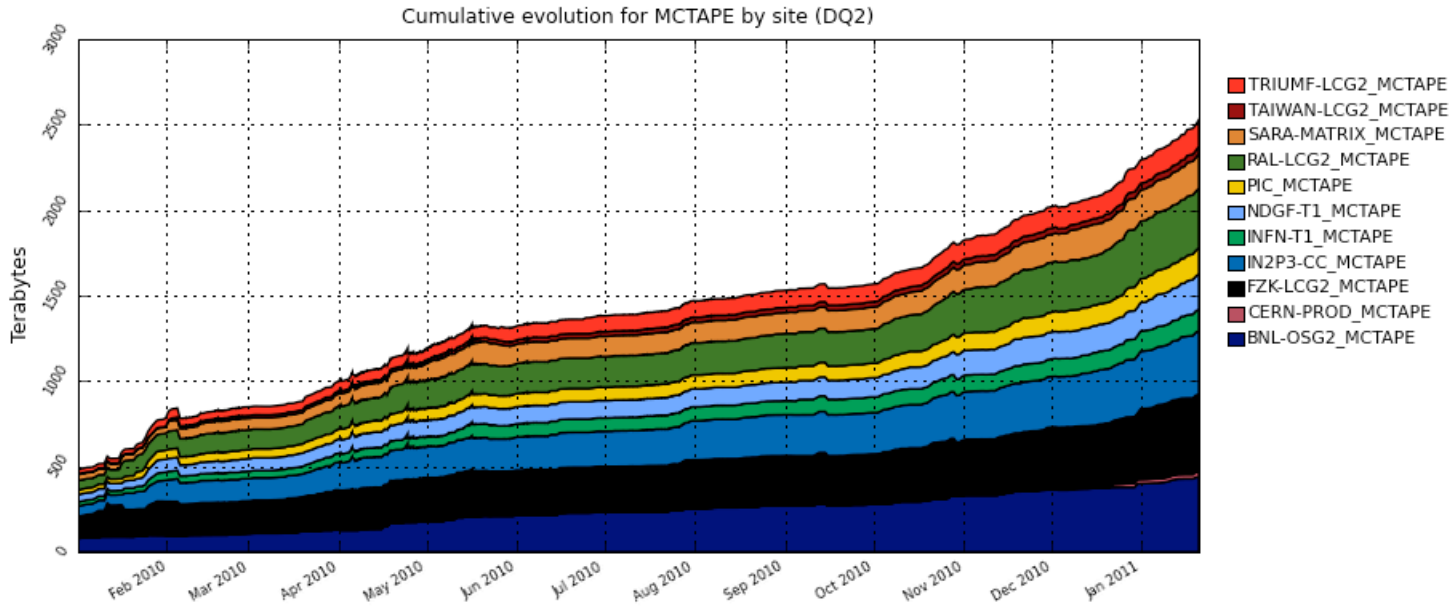
Conclusion

very preliminary !!

- The only 400 Hz scenario's that look possible with zero or 10% ESD
- If the machine turns out to be more efficient we need to lower the rate
- Running at >400 Hz would require even more drastic measures and we may run risks that things don't work as expected or less efficient than expected
- One such measure could be to keep MC at the T2s and use the MC space in the T1s for data. This needs a lot more thinking if we do it …
- Another such measure could be to remove all 2010 data from T1 disks at some point in time. Putting it on tape is no guarantee that it can be recovered later as the tape systems are overloaded during data taking or re-processing. It may be possible for 2013 during the machine stop

Backup slides

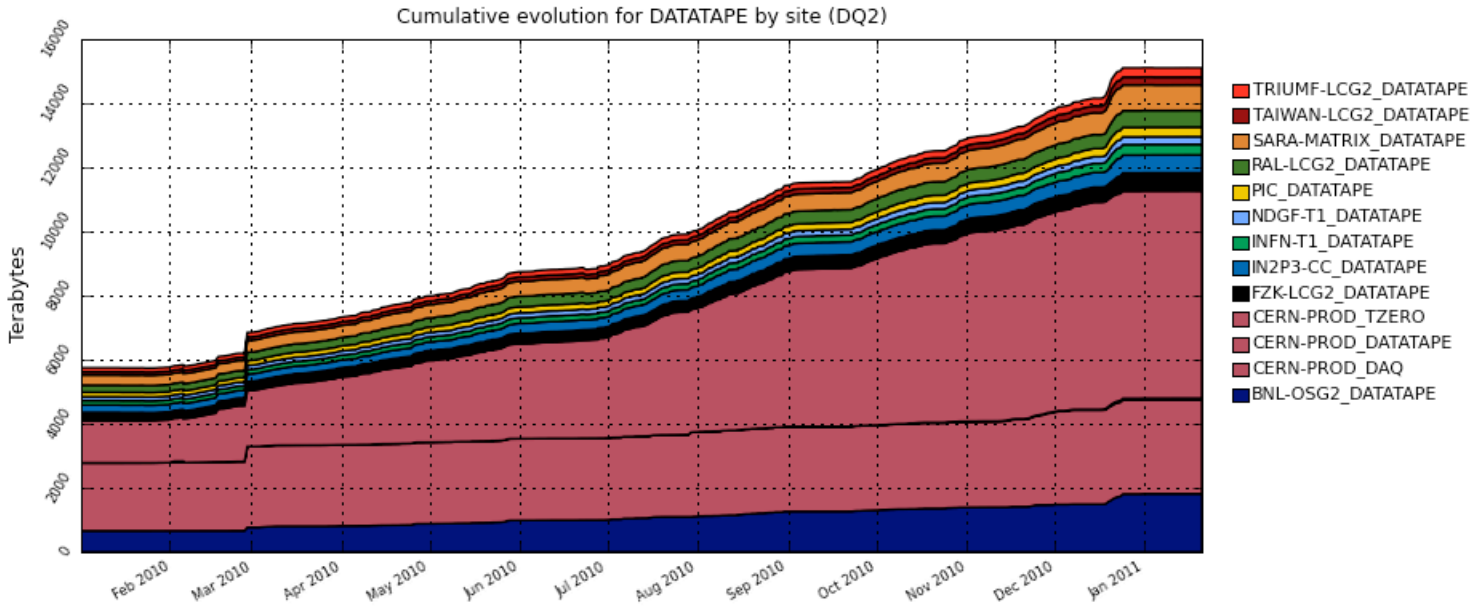
MC data on tape @T1s



Written in 2010

- All T1s 2 PB
- CERN 0 PB

MC data on tape @T1s



Need to re-make this graph without CERN contribution

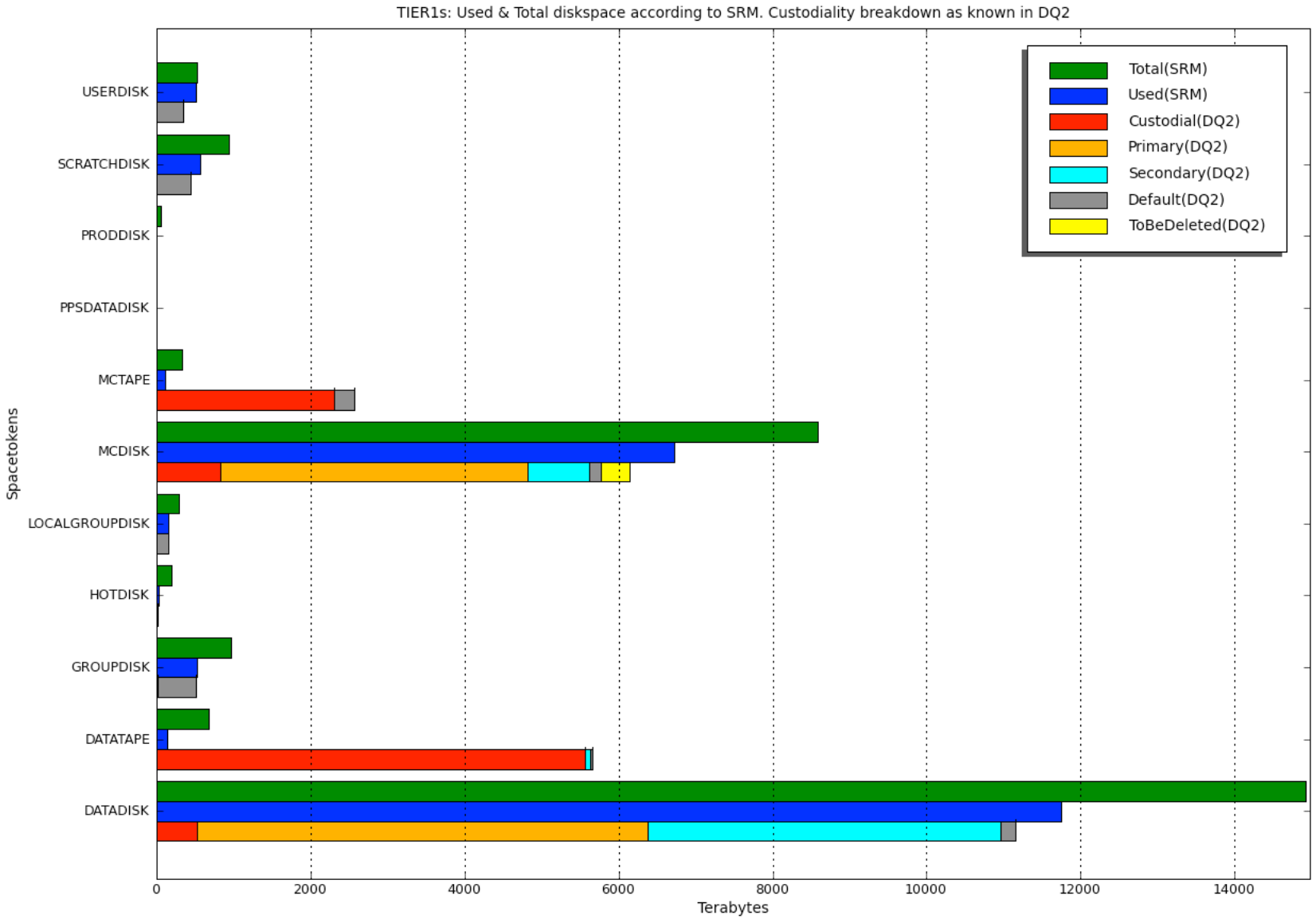
Written in 2010

- CERN 5.5 PB
- All T1s 1.5 PB (from previous slide 2.0 PB from MC)

Pledges for 2010 for data + MC

- CERN 9 PB (increase of 4)
- All T1s 18 PB (increase of 6)

Current T1 disk usage

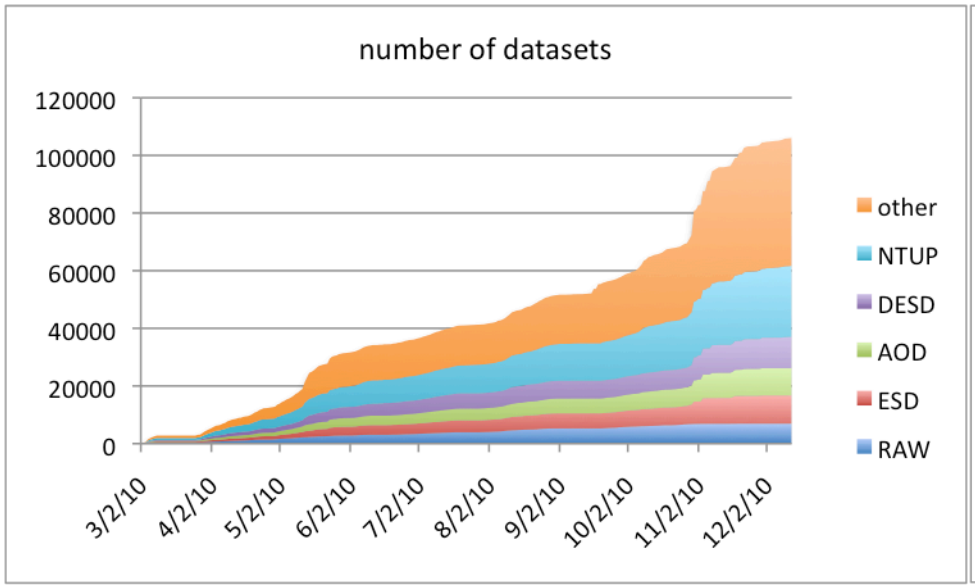


2010 data

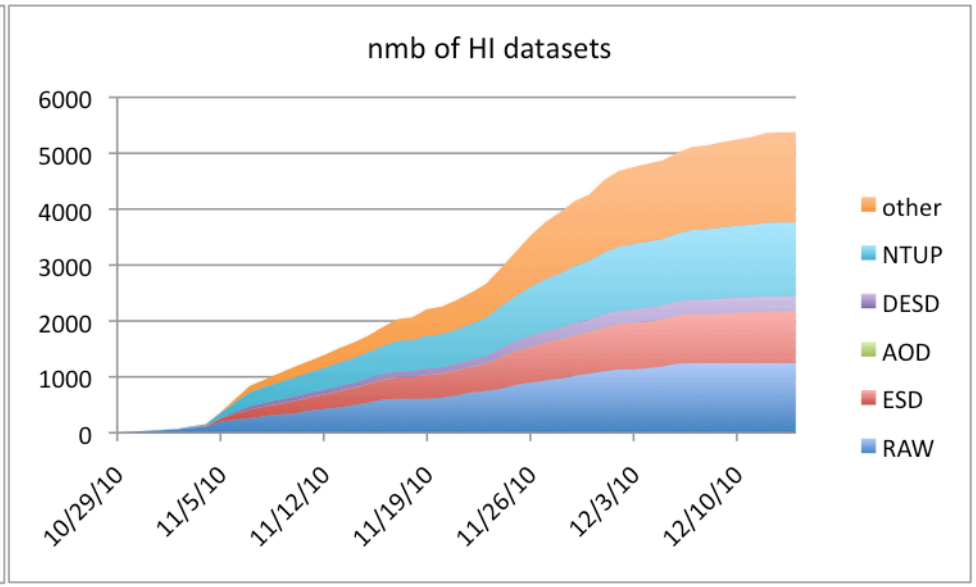
Stats derived from the DDM database

*number of **datasets***

pp data



HI data



- 120k datasets

RAW	ESD	AOD	DESD	NTUP	other
6927	9792	9492	10792	24661	44430

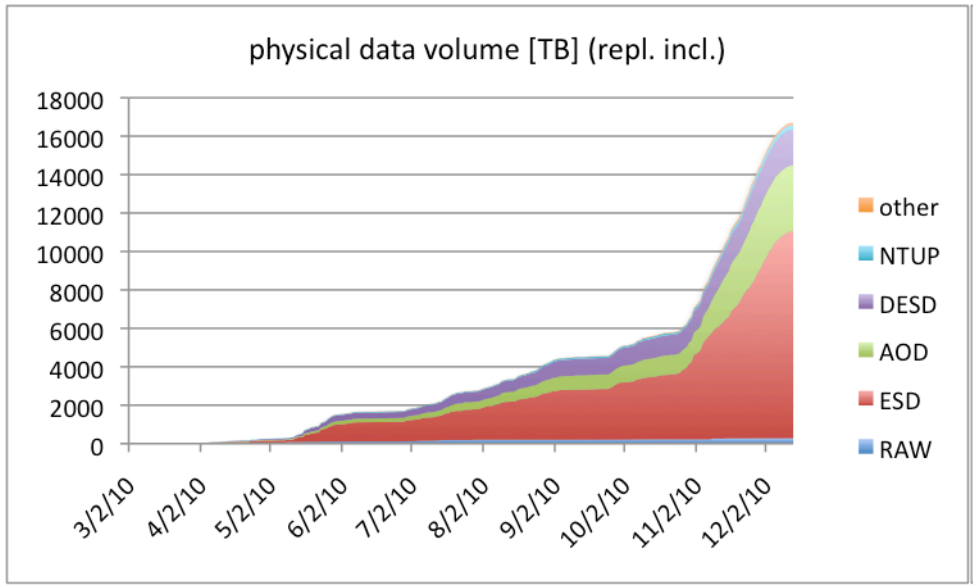
- 6k datasets (5%)

RAW	ESD	AOD	DESD	NTUP	other
1246	924	5	269	1318	1615

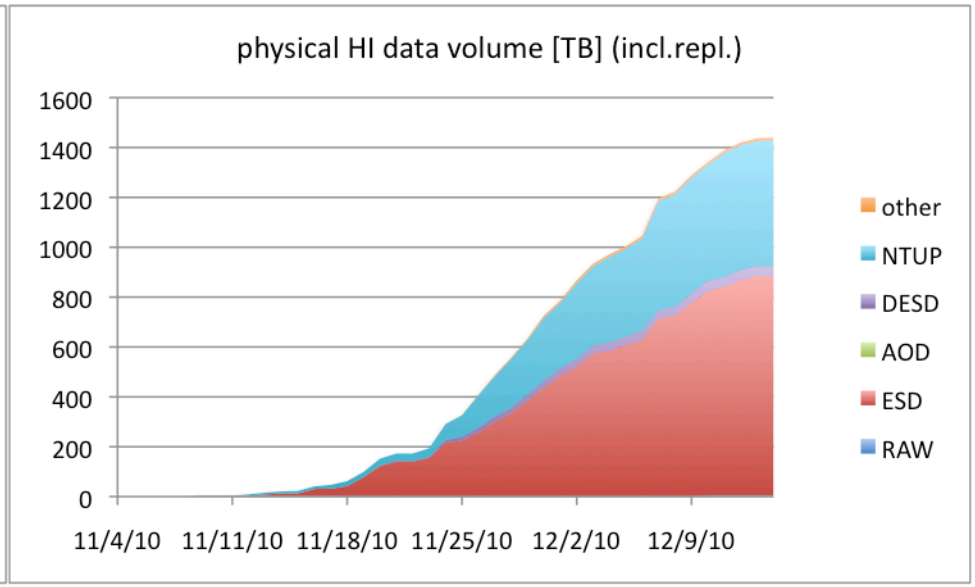
2010 data

Stats derived from the DDM database
Physical Volume means including replication to T1s and T2s

pp data



HI data



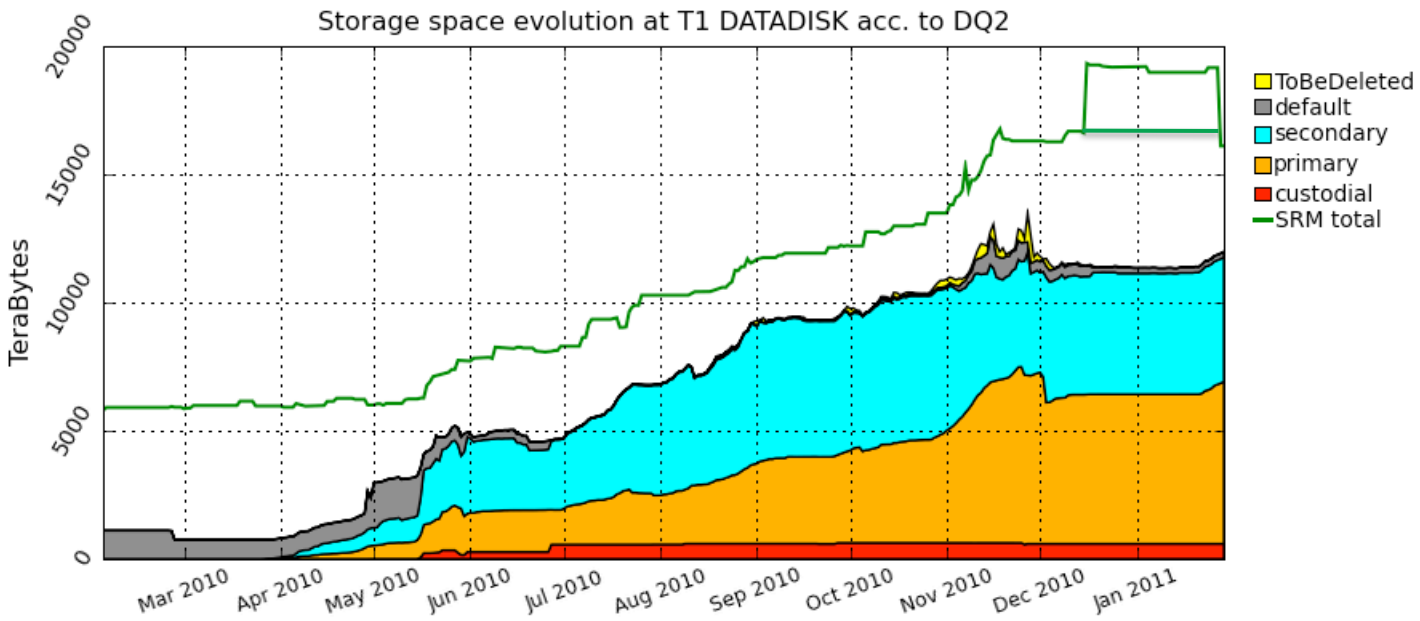
- 0.3 PB RAW on disk (DATRI)
- 16.7 PB on disk in T1s and T2s

RAW	ESD	AOD	DESD	NTUP	other
279	10772	3433	1835	264	112

- 0 PB RAW on tape (2%)
- 1.4 PB on disk in T1s and T2s (9%)

RAW	ESD	AOD	DESD	NTUP	other
6	879	0	39	505	10

- http://bourricot.cern.ch/dq2/accounting/custodiality_plot/T1s/DATADISK/395/



- http://bourricot.cern.ch/dq2/accounting/custodiality_plot/T1s/DATADISK/395/