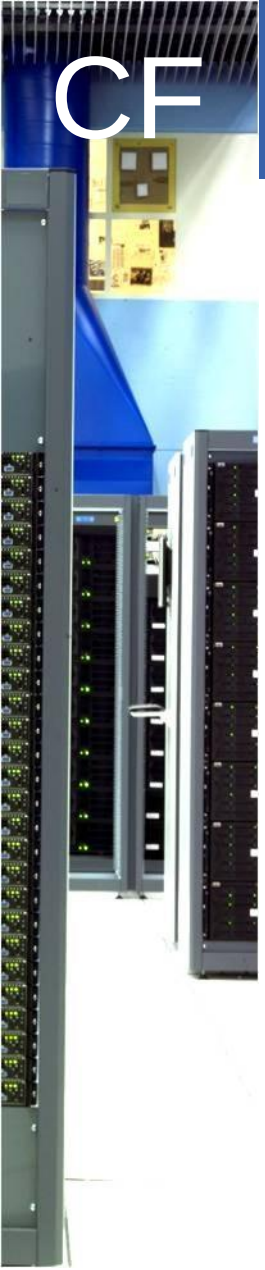
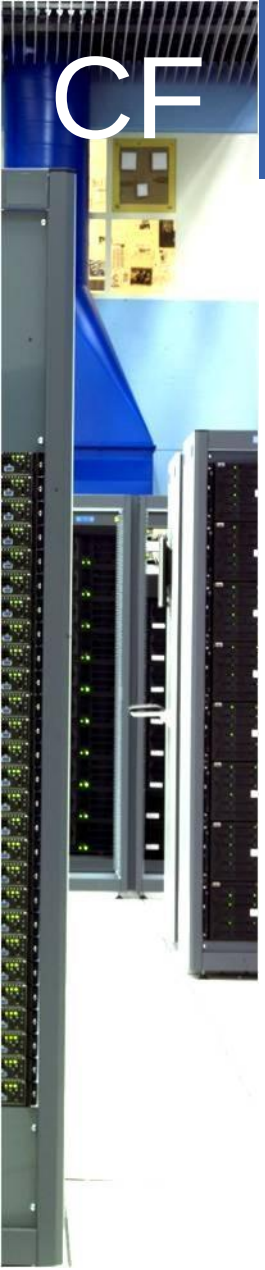


Hardware failures

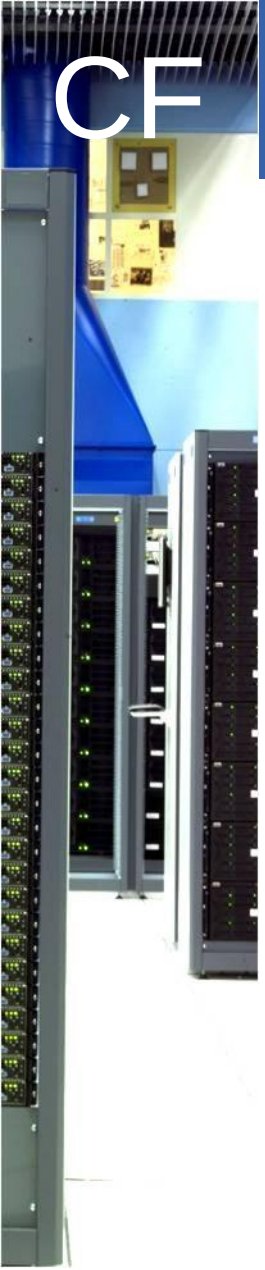
Wayne Salter
on behalf of Olof Bärring

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Outline



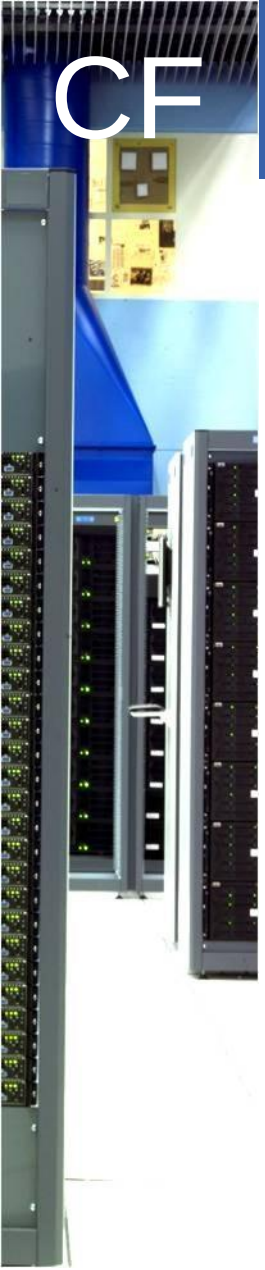
- Failures
  - What fails?
  - How often?
  - When?
- Repairs
  - How?
  - By whom?
  - How quickly?
- Conclusions



What fails? and how do we know?

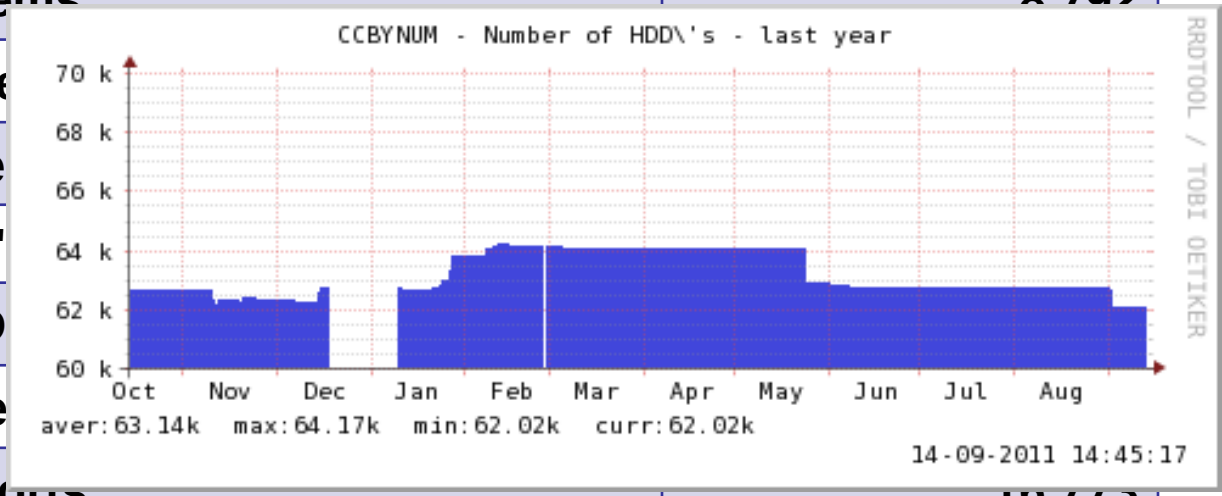
- The only things we know for sure about hardware are:
  1. It will fail
  2. Some of it fails more often than other…
     - disk drives for instance
- Monitoring failures
  - Disks: assume fail-stop but reality more complex
  - At CERN we base our decision on SMART counters and failed media scans
- Monitoring ‘repairs’ rather than ‘failures’:
  - Vendor tickets (~4k 2010-11)
  - Changes in serial numbers inventory (~10k 2010-11)

Failure space



- CERN IT by numbers (14/9/2011)

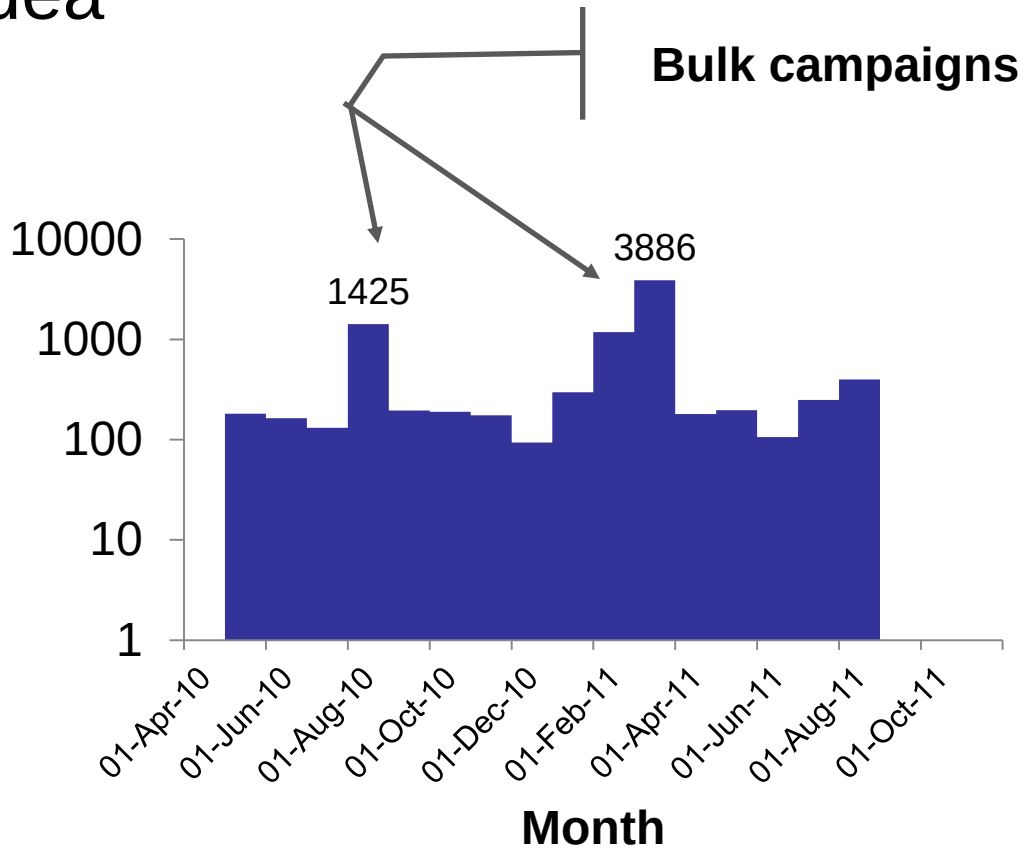
Number of systems	**8,792**
Number of proce	
Memory module	
Number of HDD'	
Number of RAID	
Number of Fibre	
Number of 1G ports	**16,773**
Number of 10G ports	**622**



CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

How often?

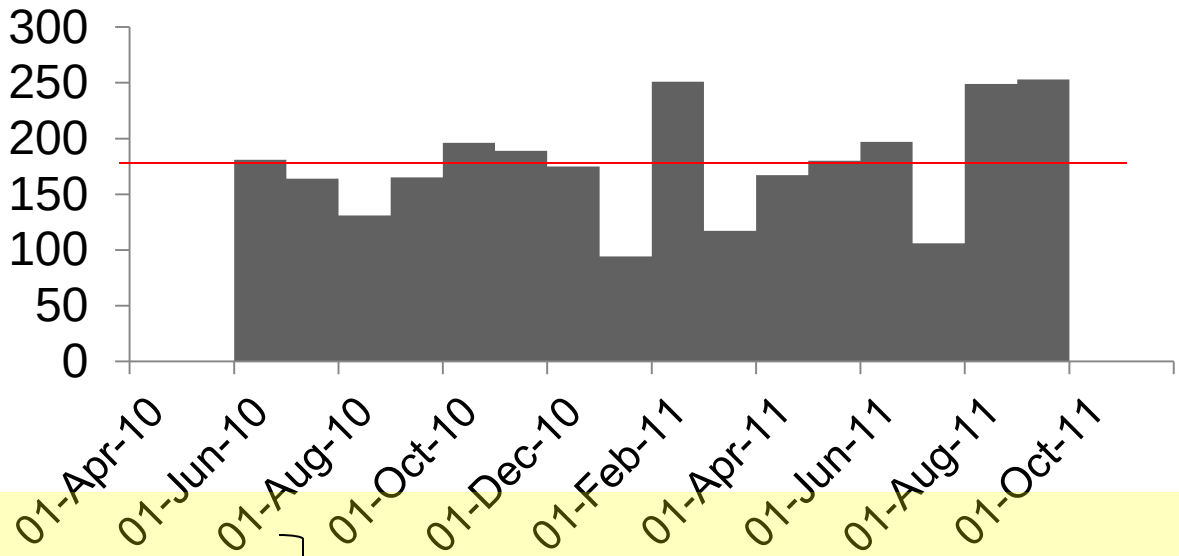
- Monitoring changes in serial numbers gives an idea



CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

How often?

- Monitoring changes in serial numbers gives an idea
  - Excluding campaigns ~170 disks /month (5 /day)



HDD failures/day:5
Hours/day: 24
→ ~1 fail per 5hrs

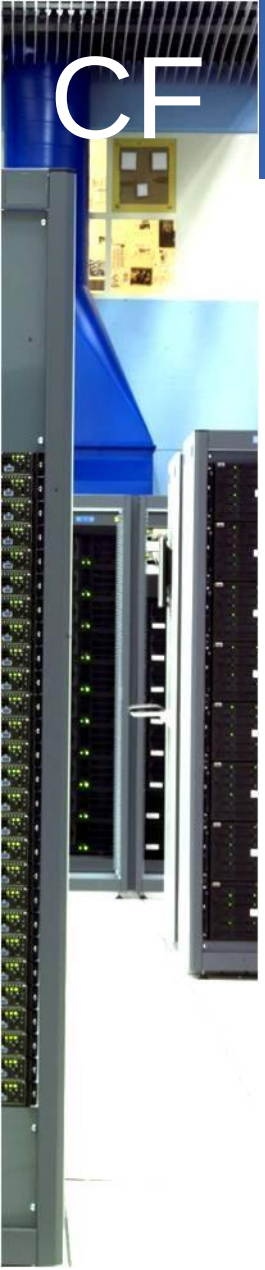
64,000 drives in the centre

→ MTTF = 320,000 hrs

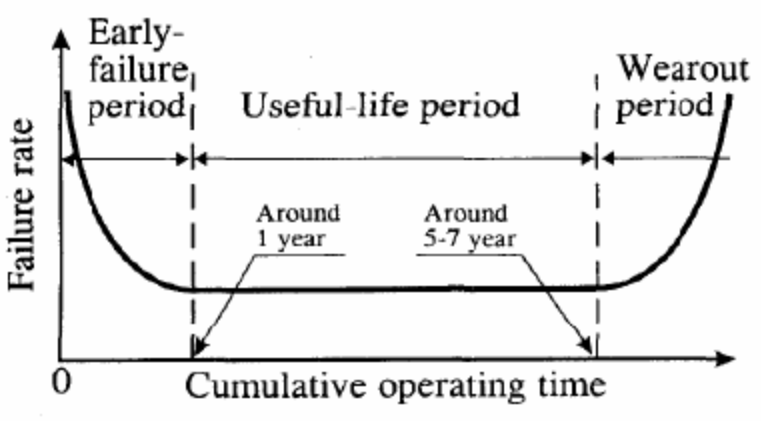
(Spec: 1.2Mhrs)

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

When?

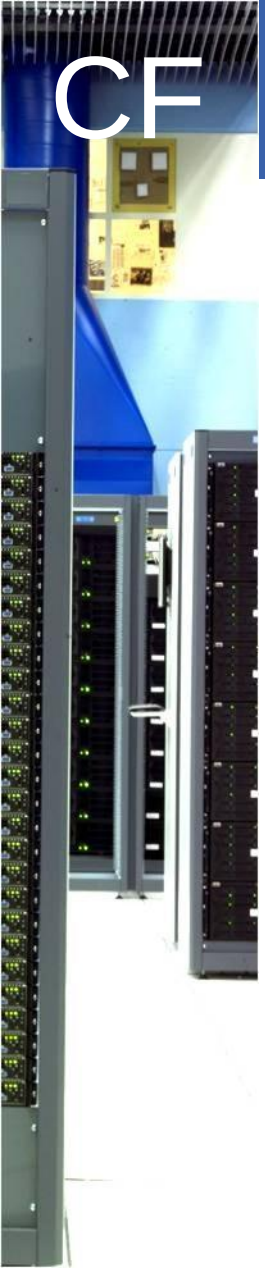


Failure rates of hardware products typically follow a “bathtub curve” with high failure rates at the beginning (infant mortality) and the end (wear-out) of the lifecycle[1].

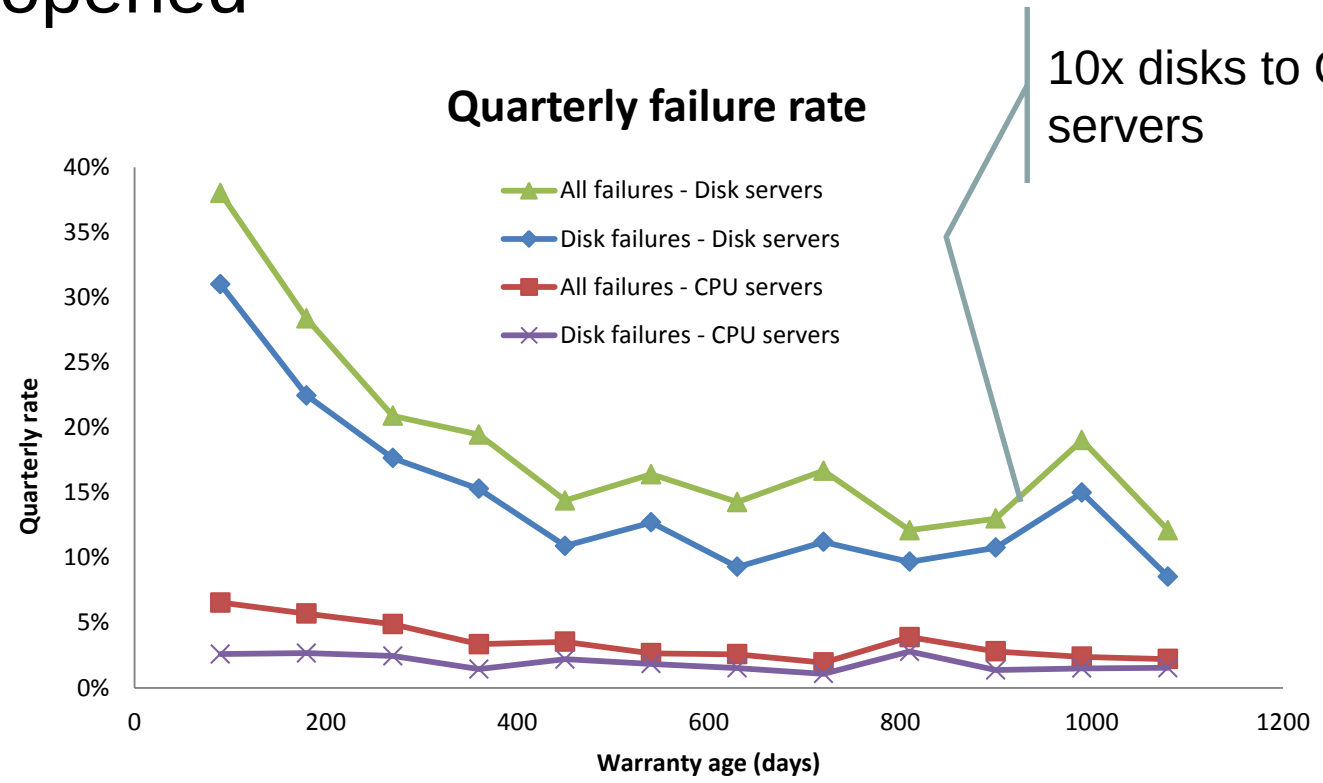


[1] http://www.usenix.org/events/fast07/tech/schroeder/schroeder.pdf

When?



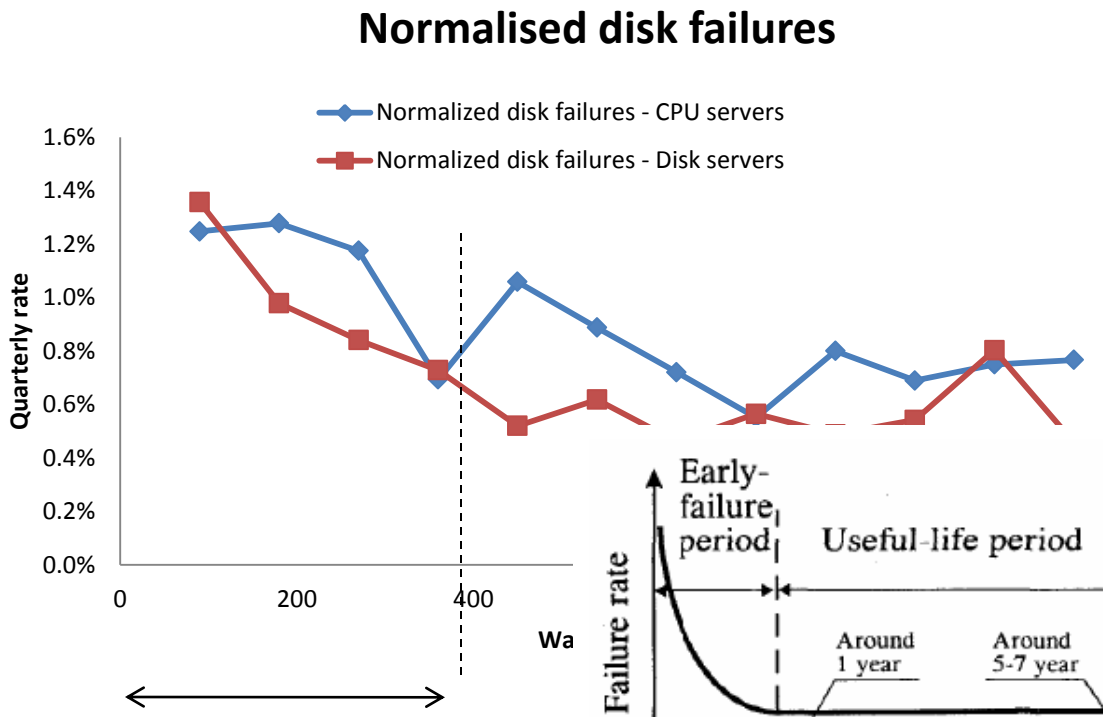
Process and categorize 2010-11 vendor calls according to 'Warranty age' when call was opened

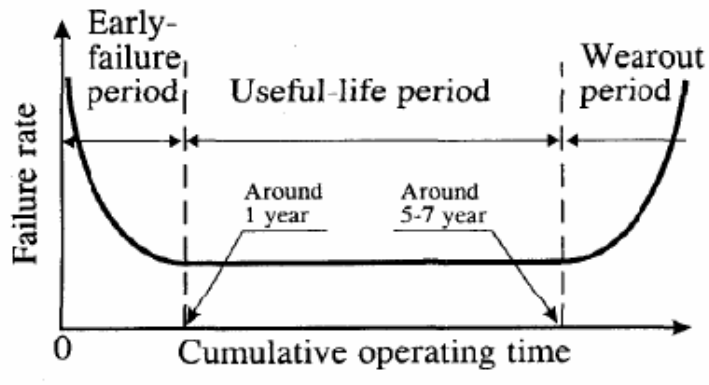


CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

When?

Quarterly disk failure rate normalized to number of disks

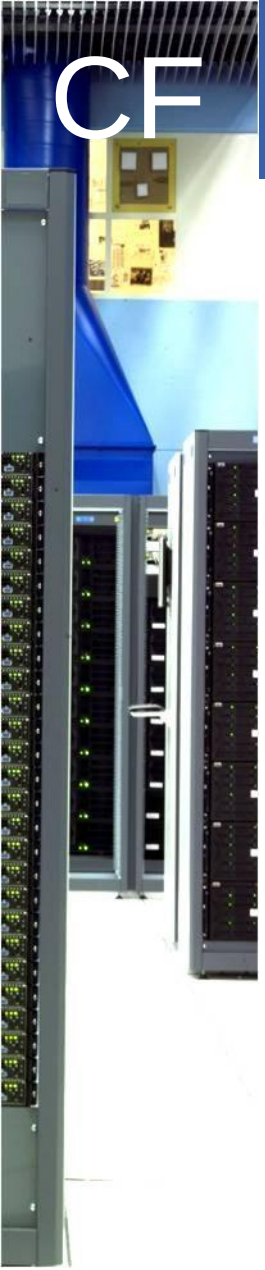




Early failures
(infant mortality)

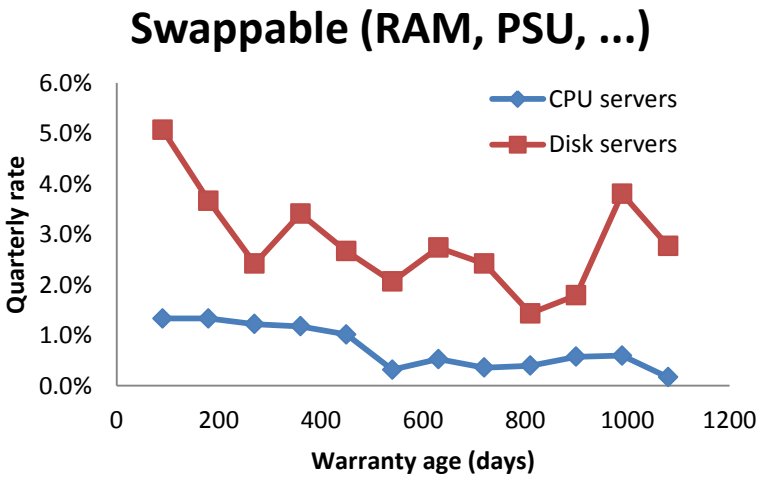
CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

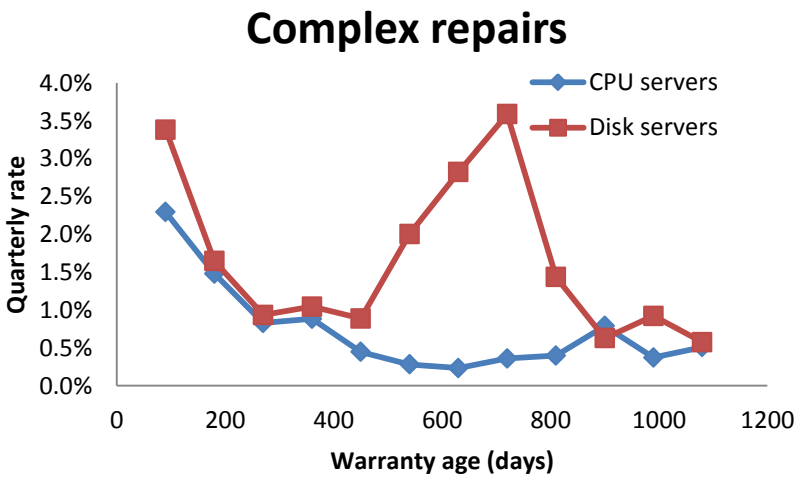
When?



Other failure types

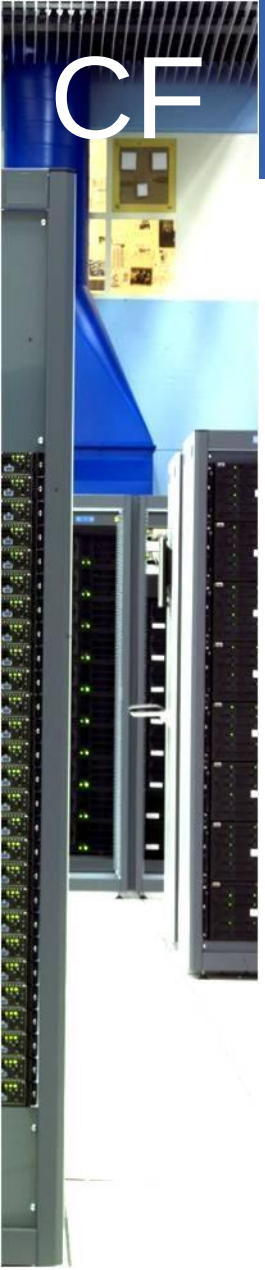
- Swappable: RAM, PSU, BBU, BMC, …
- Complex repairs: cabling, backplane, main board, … no clue…





CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

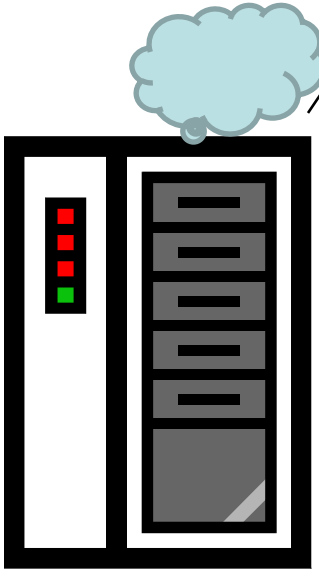
Repairs



Vendor call

Alarm

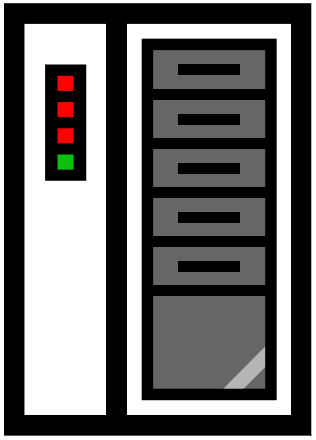
New sn: WD3342ABC

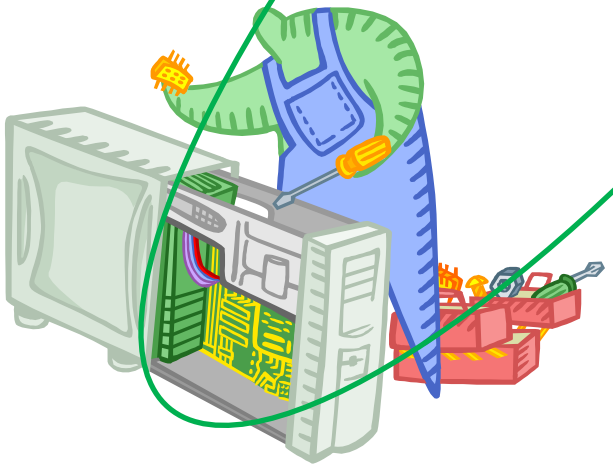


CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

By who,?

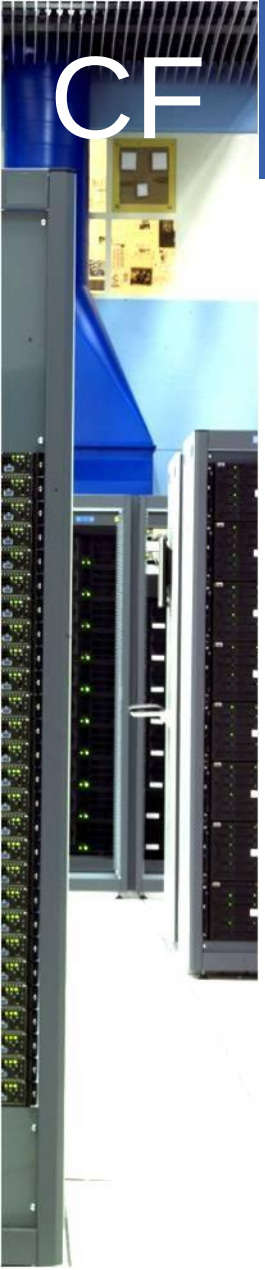
Vendor





CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

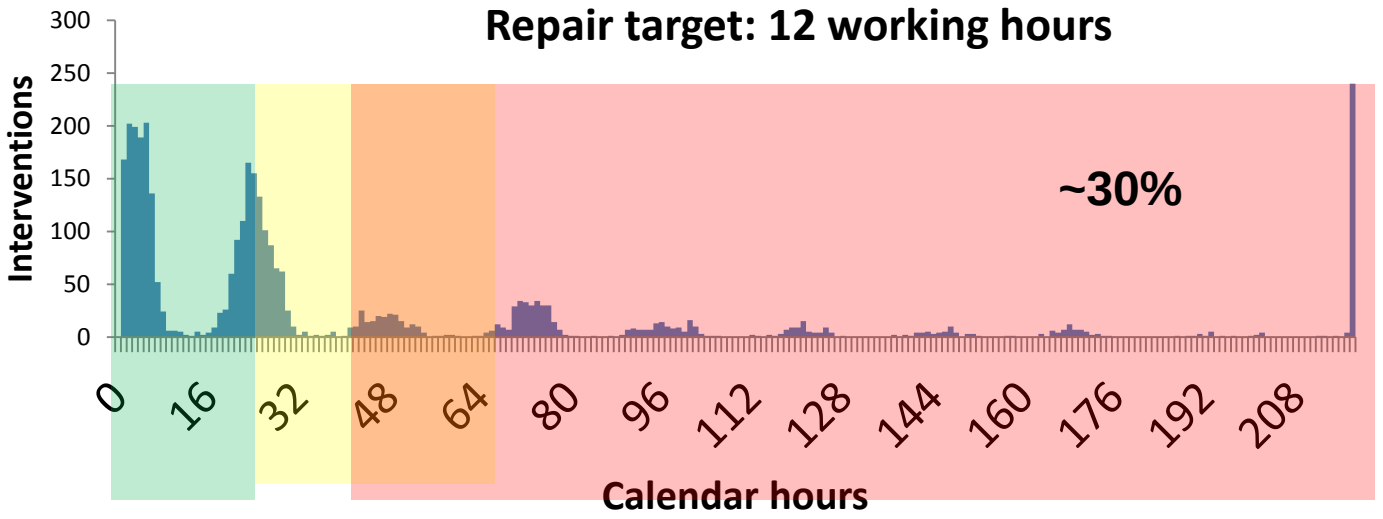
How quickly?



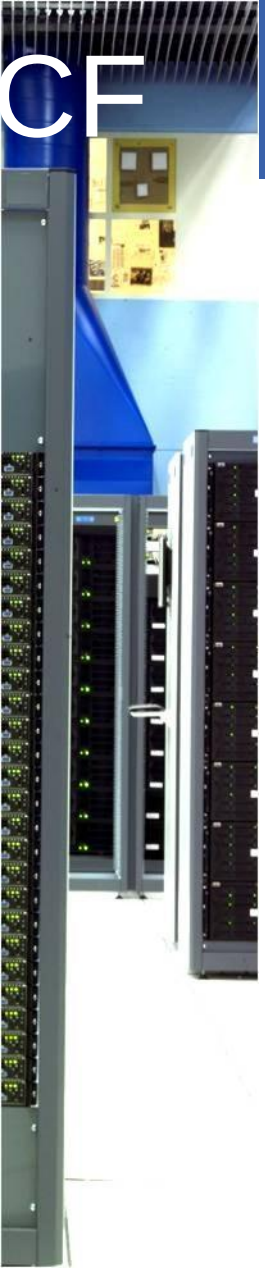
- Two contract types

Type	Time to intervene	Repair time
Normal	24 working hours	40 working hours
Fast	4 working hours	12 working hours

- ‘Normal’ only used for CPU servers



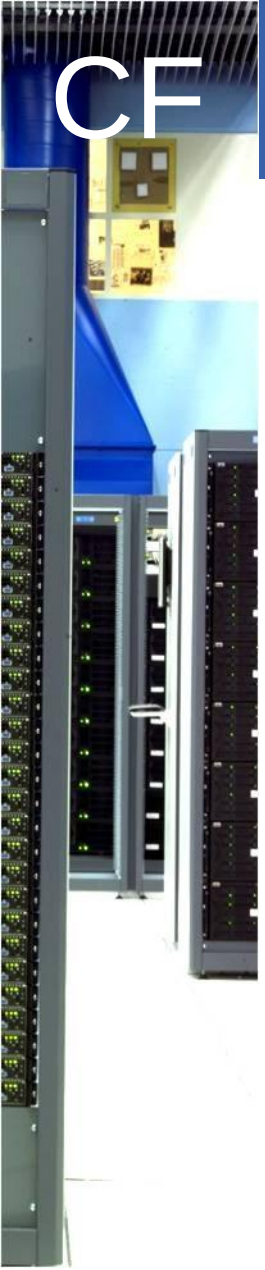
CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it



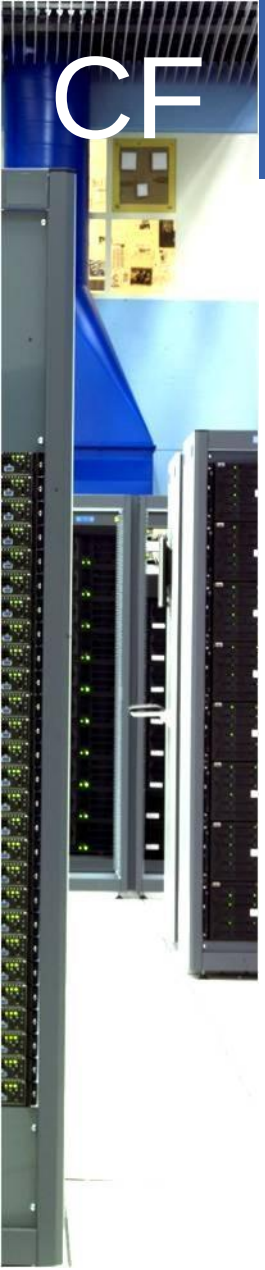
Ongoing Improvements

- Tracking changes to servers
  - Keep current tools that report HW info
  - Will store each server's HW info as a document (HW inventory)
    - Key is unique id stored in the BMC when hardware is purchased
    - Change log, e.g. replaced parts, for each server
  - Goals:
    - Better accessibility and usability of data
    - Provide base for a more comprehensive HW inventory tool
    - Systematic tracking of parts replacement due to failure
    - Trending and potential action (e.g. #disk replacements in last month > X

Conclusions



- Hardware fails
  - As expected
  - More often than expected
    - MTTF ~320khours rather than 1.2Mhours
  - When expected:
    - Effect of early failures (infant mortality) in first year
    - No sign of wear-out at the end of the 3 years warranty
- Repairs are currently carried out by vendor
  - Missed repair targets in ~30% of cases
  - Looking at a different model…



Questions?