

Computing Directions

Ian Fisk

July 2, 2013

What follows

- ✦ Jim Shank and I are tasked with developing the Computing Report for the Energy Frontier
 - ✦ It's not entirely clear what deep insights there are to make about this Computing Frontier
 - ✦ Looking at the machine plans. They are high luminosity machines with potentially very high trigger rates and complicated events
 - ✦ Constrained budgets and little magic
- ✦ What follows are some observations to spawn discussion

Looking Back

- ✧ I decided to look back 10 years before trying to look forward 10
 - ✧ Tevatron was in the 3rd year of Run2 in 2003
- ✧ Compare to 2012
 - ✧ The third year of LHC

Complexity and Collaborations

- Trigger rate, event size, and reconstruction time all rise by a factor of 10

Metric	<u>Tevatron (2003)</u>	<u>LHC (2012)</u>
Trigger rate	50Hz	500Hz
Prompt Reconstruction rate/week	13M Events	120M events
Re-reconstruction rate	100M events per month	800M – 1B events per month
Reconstructed size	200kB	1-2MB
AOD size	20kB	200-300kB
Reconstruction time	1-2s on CPUs of the time	~10s on CPUs of the time

- Collaborations increase by a factor of 3

Metric	<u>Tevatron(2003)</u>	<u>LHC(2012)</u>
Collaboration Size	800	2000-3000
Number of individual analysis submitters per day	100	300-400
Number of total analysis submitters	400	Greater than 1000

Resources

- Resources and challenges increase at different rates

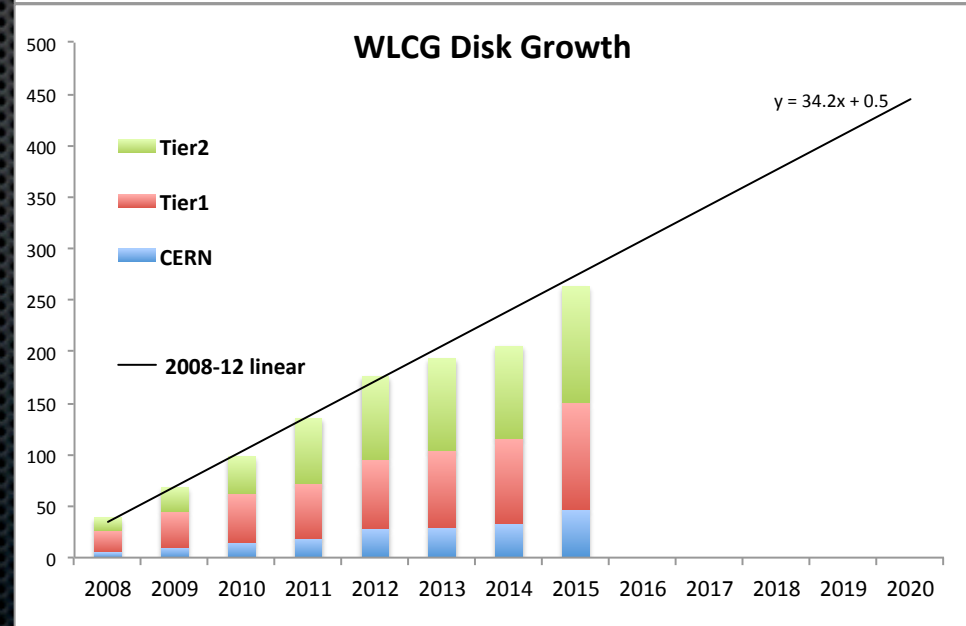
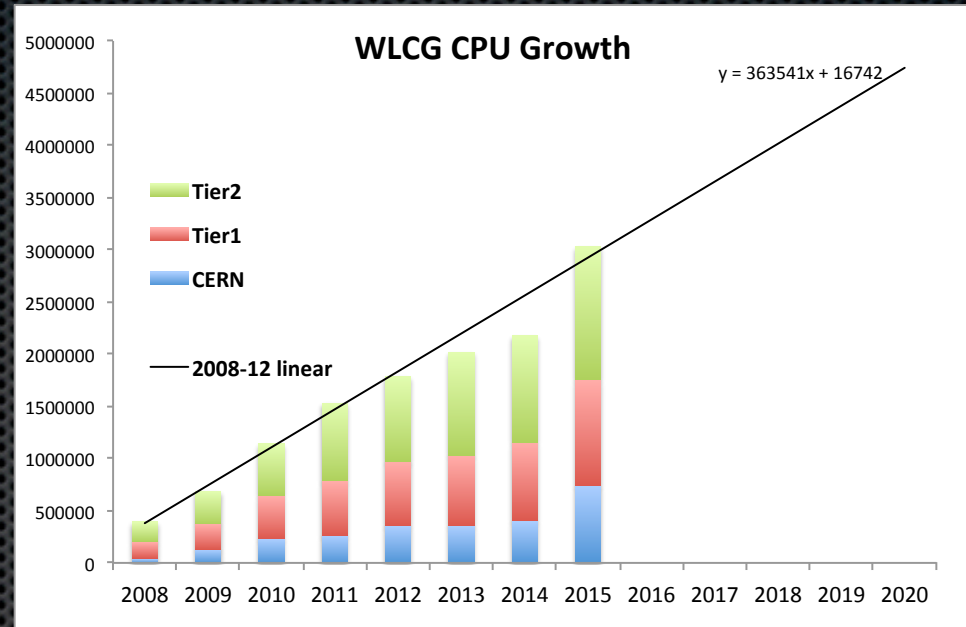
Metric	<u>Tevatron(2003)</u>	<u>LHC(2012)</u>
Remote Computing Capacity	15kHS06 (<u>DZero</u> Estimated)	450kHS06 (CMS)
User Jobs launched per day	10k per day	200-300k jobs per day
Disk Capacity per experiment in PB	0.5PB	60PB
Data on Tape per experiment	400TB	70PB
MC Processing Capacity per month for Full Simulation	3M	300M
Data Served from <u>dCache</u> at FNAL per day	25TB per day	10PB per day
Wide Area networking from host lab	200Mb/s	20000Mb/s
Inter VO transfer volume per day	6TB (<u>DZero</u> SAM)	546TB (ATLAS)

Increases

- ✦ The processing has increased by a factor of 30 in capacity
 - ✦ This is essentially what would be expected from a Moore's law increase with a 2 year cycle
 - ✦ Says we spent similar amounts
- ✦ Storage and networking have both increased by a factor of 100
 - ✦ 10 times trigger and 10 times event size

For LHC Increases per year

- ✧ LHC Computing adds about 25k processor cores a year
- ✧ And 34PB of disk
- ✧ The cost and complexity of the storage is much larger than the processing

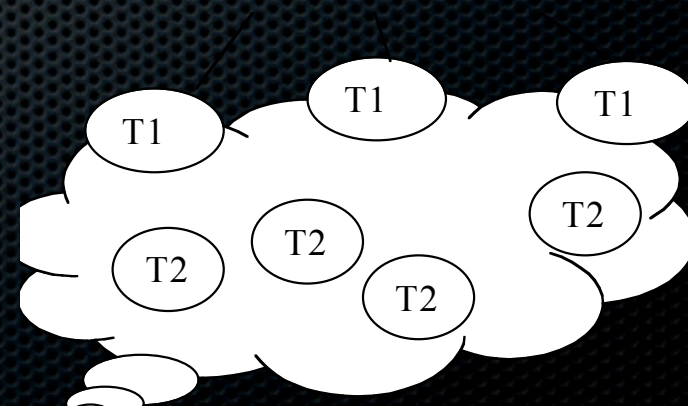
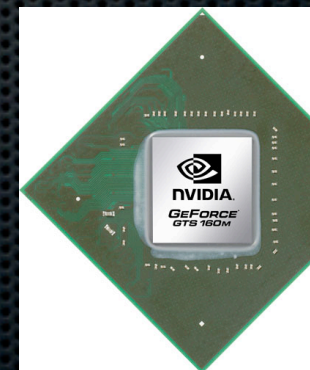


Increases

- ✦ Extrapolating the increase for LHC it doesn't look so alarming
 - ✦ In 10 years there is a factor of ~ 3
- ✦ What's more concerning is a new machine or deviation from the linear growth that would see the huge increase in one of the quantities.

Looking Forward

- ✧ Computing is at something of a cross roads
 - ✧ In one direction are clouds
 - ✧ Generic computing services that are bought, shared, or contributed
 - ✧ Computing as a service
 - ✧ In the other direction are very specialized systems
 - ✧ High performance, low power
 - ✧ Massively multi-core
 - ✧ GPUs



Clouds and Provisioning

- ✦ Commercial clouds are still very expensive for resources we use a lot
 - ✦ Small sites without a history of computing will probably be the first to simply buy capacity
- ✦ More opportunistic and academic resources will move to cloud provisioning methods
- ✦ Even sites we control will move to cloud provisioning tools because it simplifies the operations and places more expectations on the supported community to define and operate services
- ✦ We should expect our current service architecture will change to new provisioning tools

Current Hardware

- ✦ Currently energy frontier computing lives in a homogenous but non optimal environment
 - ✦ Looking back we have typically supported many more platforms
 - ✦ Most of the industry development is not in the chips that make up the bulk of our computing
 - ✦ Cores are added, but individual cores tend to stay at similar speed
 - ✦ We are not well optimized and we don't tend to use the full capacity of the hardware

Specialized Hardware

- Specialized chips like GPUs and co-processors have the potential for big improvements in performance, but are challenging to program and introduce a lot of heterogeneity
- Specialized machines like very high core count low power systems look like super computers
 - And have the programming challenges associated

Specialized Hardware Steps

- 1.) Places we completely control will get specialized gear
 - Trigger farms
- 2.) We buy/get access to specialized gear like a super computer allocation
- 3.) We own specialized gear
 - This will need to be a sufficient improvement to replace and entire class of systems

Becoming More Selective

- ✦ We have not really changed how we think about Energy Frontier events ever
 - ✦ Data is collected and protected
 - ✦ Trigger rates continue to rise
 - ✦ Most events are uninteresting background
- ✦ Energy Frontier may need to adopt more detailed and selective online reconstruction including some of the elements previous associated with analysis
 - ✦ Not necessarily all the events need to be treated the same

Energy Frontier

- ✦ As trigger rates proposed for Energy Frontier approach rates we would typically associate with Intensity Frontier we may need to adopt similar techniques
 - ✦ ALICE is already planning for this post LS2

Data Management

- We will have a mix of local, cloud, opportunistic, and specialized resources and we will need a data management system that deals with all
 - On cloud the concept of data locality begins to lose a lot of meaning
 - We cannot really afford another factor of 100 increase in storage, so we need to find ways of being more efficient in the use of the space

Connectivity

- ✦ Given the connectivity of our clusters and the expectations of the users, I believe we will have to evolve to content delivery networks
 - ✦ Data Management resources that deliver data on demand
 - ✦ Will be cached and replicated and intelligent about the placement, but large independent local storage systems connected to clusters is probably not the most efficient
- ✦ The data federations already being deployed are a first step, but work is needed

Networking

- ✦ Data delivery systems give a lot of flexibility in terms of how to make use of diverse computing systems, but they put strong requirements on networking
 - ✦ Currently a 10k core cluster (typical for 2020) would require 10Gb/s networking for organized processing like reconstruction
 - ✦ Analysis would require 100Gb/s

Outlook

- ✦ LHC moving forward may be sustainable with an evolution of how we work
- ✦ A big increase in luminosity and complexity would lead to a big jump that would be potentially very expensive
 - ✦ To handle this we need to change how we work by being more selective
 - ✦ Move to be able to run on fast hardware
 - ✦ Solve the data management problem