

# Scientific Computing

Stefan Kluth  
MPP Project Review  
15.12.2015

# The third way

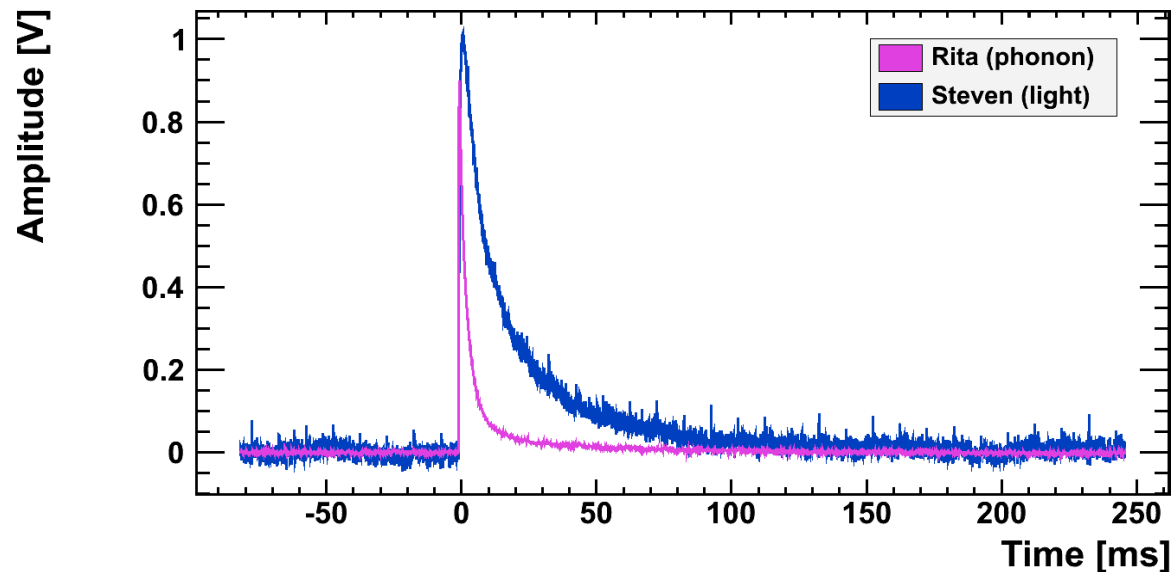
- The scientific method (simplified)
  - Experiment: design a setup and collect data, infer from data underlying principles; test theories
  - Theory: build up from fundamentals a mathematical framework to describe nature and make predictions; learn from experiment data
- With computers
  - Numerical simulation: translate abstract / unsolvable models into practical predictions, discover behavior
  - Find structures in (unstructured) data

# Overview

- Example applications
  - CRESST
  - ATLAS
  - (Theory see Thomas Hahn talk)
- Data Preservation
- Software development example
  - DataBrix
- Resources
  - MPP, MPCDF, LRZ, Excellence Cluster (C2PAP)

# CRESST continuous data stream

- Current approach: hardware trigger activates the DAQ and the samples are stored



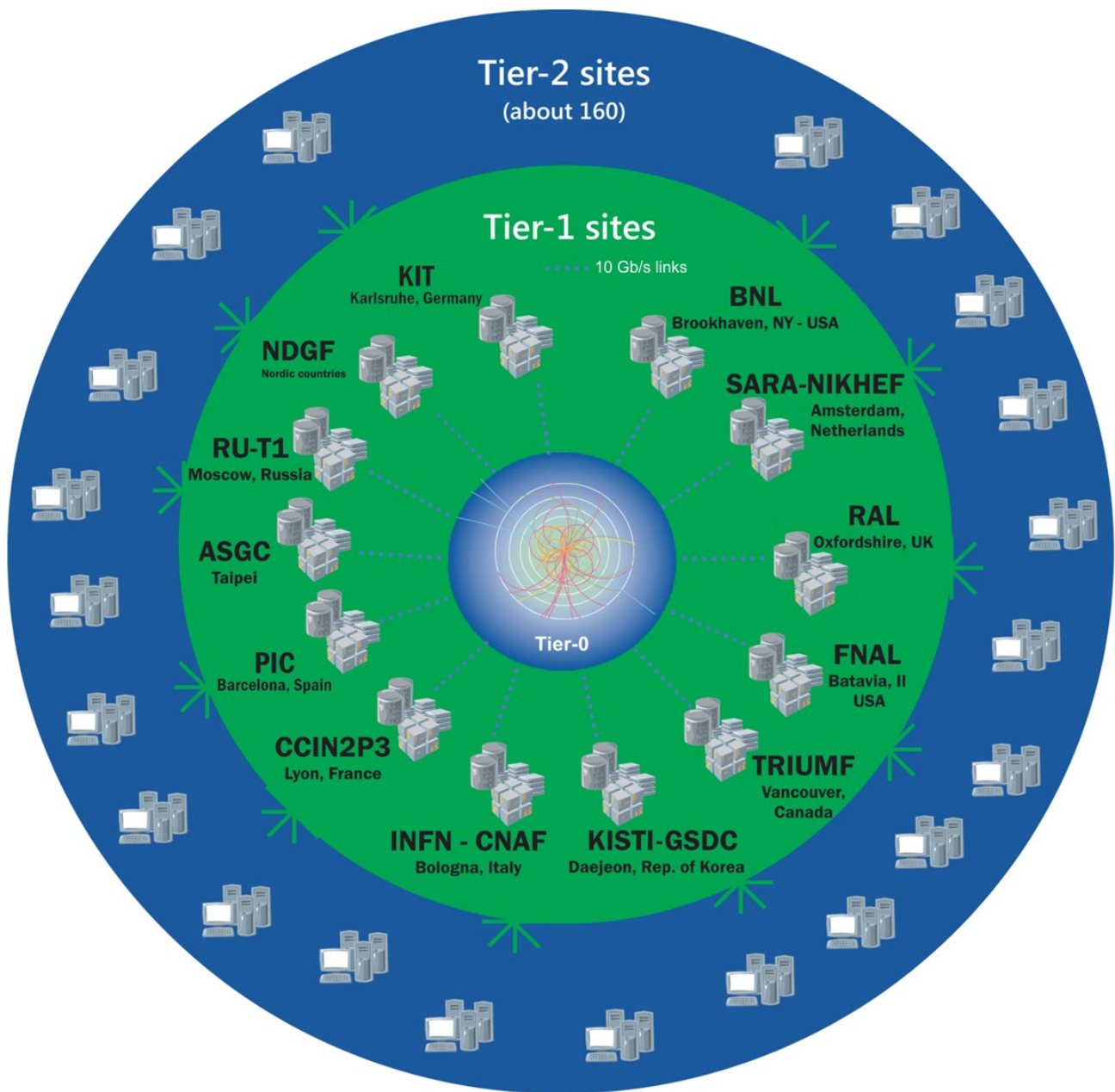
- New idea: sample and store entire baseline



# CRESST continuous data stream

- Low energy threshold is key to low energy dark matter search
- Software trigger can perform better than hardware
- Allows to re-think trigger decisions
- 40 Channels read at 25 kHz, 16 bit: ~120 MB/min or ~50 TB/yr
- ~50 core-years CPU for trigger calculation
  - ~18 days on 1000 cores

# ATLAS WLCG



Tier-0: CERN

Tier-1: GridKa

Tier-2: MPPMU

Originally hierarchical,  
moving to network of  
sites

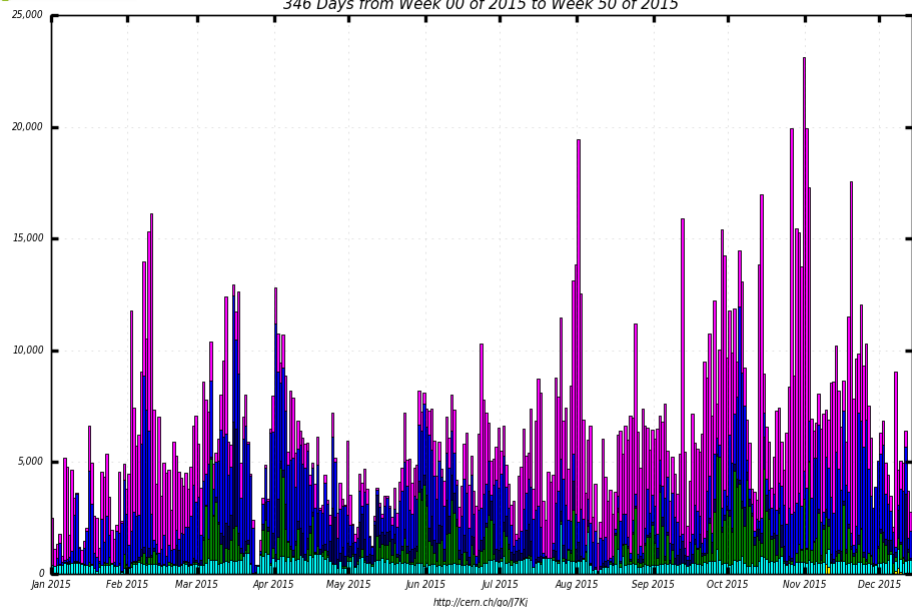
MAGIC, CTA, Belle 2  
following this model,  
our Tier-2 supports this

# ATLAS MPP Tier-2



Completed jobs

346 Days from Week 00 of 2015 to Week 50 of 2015



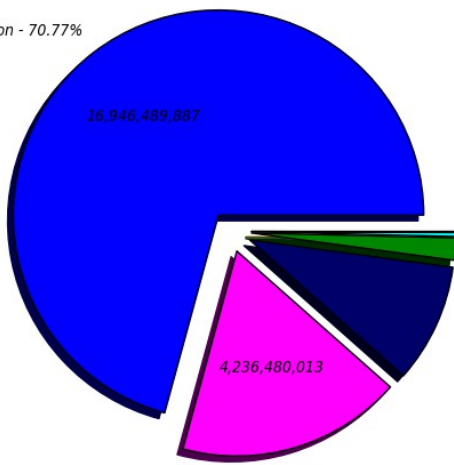
<http://cern.ch/go/7Kj>

■ Analysis  
■ Data Processing  
■ MC Simulation  
■ MC Reconstruction  
■ Group Production  
■ Others

Maximum: 23,134, Minimum: 372.00, Average: 6,325, Current: 2,779

CPU consumption Good Jobs in seconds (Sum: 23,946,248,892)

MC Simulation - 70.77%

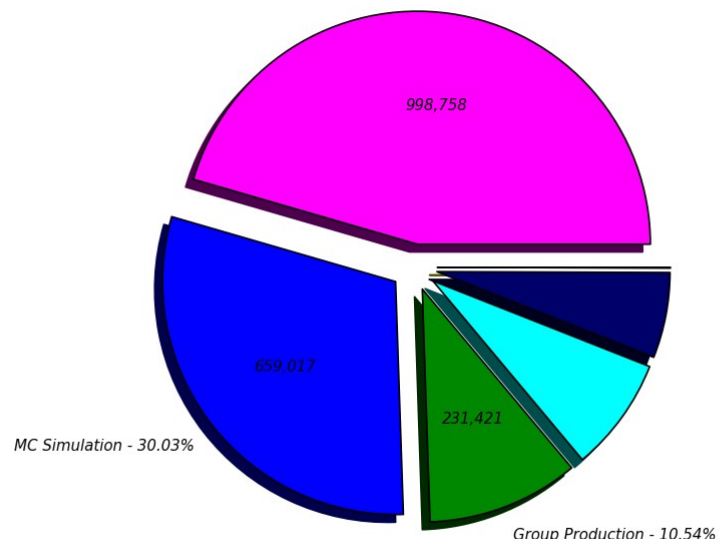


<http://cern.ch/go/pMx7>

■ MC Simulation - 70.77% (16,946,489,887)  
■ MC Reconstruction - 9.40% (2,252,095,700)  
■ Data Processing - 0.31% (75,377,750)  
■ Analysis - 17.69% (4,236,480,014)  
■ Group Production - 1.76% (420,354,659)  
■ Others - 0.06% (15,450,882)



Completed jobs Pie (Sum: 2,194,786)  
Analysis - 45.51%



<http://cern.ch/go/7Kj>

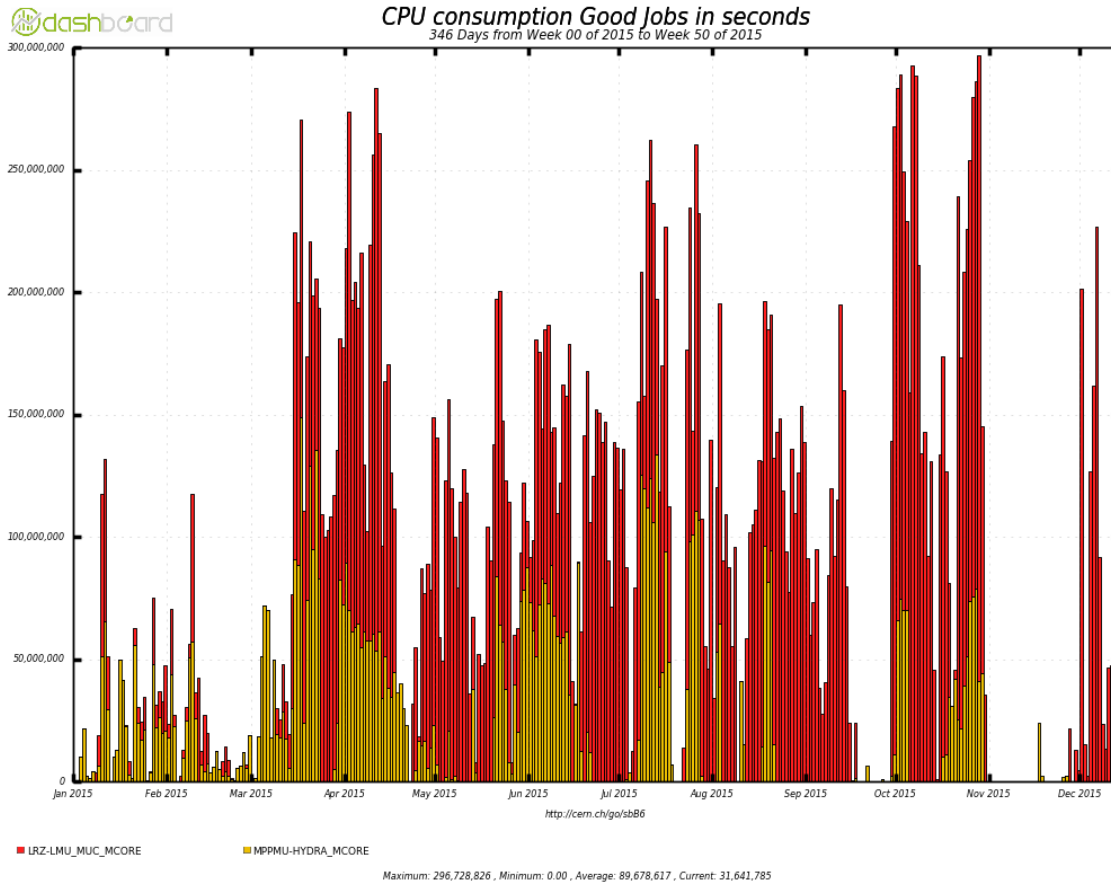
■ Analysis - 45.51% (998,758)  
■ MC Simulation - 30.03% (659,017)  
■ MC Reconstruction - 5.94% (130,289)  
■ Group Production - 10.54% (231,421)  
■ Data Processing - 0.06% (1,231)  
■ Others - 7.93% (174,070)

50% nominal Tier-2

1/60 of total ATLAS Tier-2

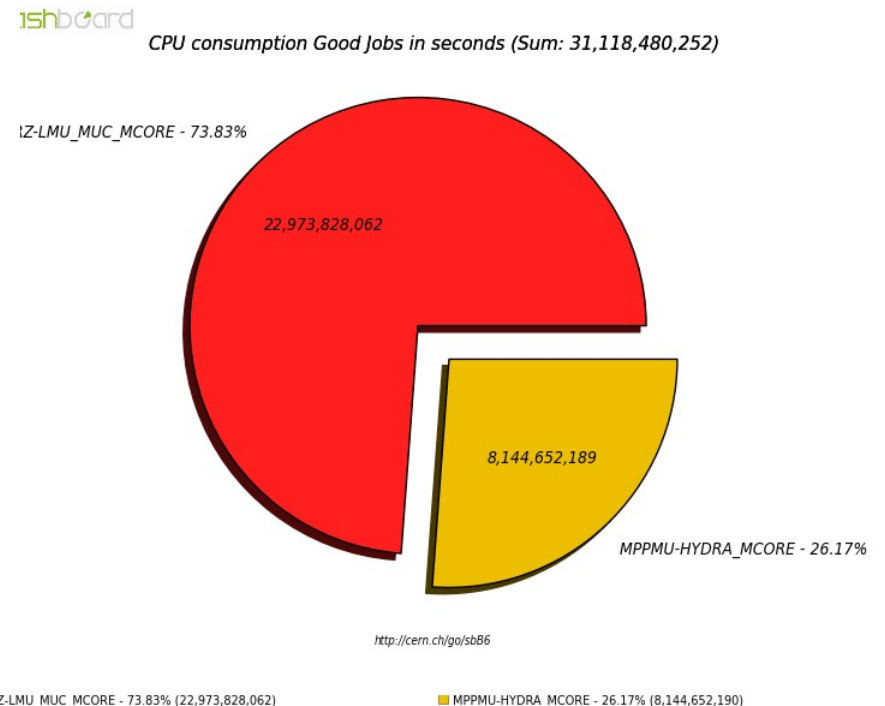
Incl. above pledge contributions

# ATLAS @ Hydra



Hydra (MPCDF)  
SuperMUC (LRZ)

About 1 nominal Tier-2  
Almost 600k jobs



Hydra full with other (big) jobs,  
opportunistic use only

SuperMUC grant for ATLAS



# DPHEP

Andrii Verbytskyi

- MPP has several experiments with valuable data and ongoing analysis activity
- H1 and ZEUS @ HERA
- OPAL @ LEP and JADE @ PETRA
- See Andrii Verbytskyi talk
  - and previous project reviews since 2000

# DPHEP

- Save bits: copy data to MPCDF
  - Provide access via open protocols (http, dcap)
  - Use grid authentication (X509)
  - About 1 PB (H1, ZEUS, OPAL, JADE), goes to tape library
- Save software: installation in virtual machine
  - Provide validated environment (SL5, SL6, ...)
- Save documentation: labs, inspire, ...
  - Older experiments: scan paper-based documents

# ZEUS event display@MPP connected to MPCDF

The screenshot displays the ZEUS event display software interface. The main window, titled "Zeus Run 52284 Event 107169", shows event data for a date of 7-12-2004 at 07:00:36. The data is organized into two columns of parameters:

$E=36.8$ GeV	$E_1=8.05$ GeV	$E-p_z=31$ GeV	$E_1=20.5$ GeV	$E_2=1.59$ GeV
$E_2=14.7$ GeV	$p_1=0.84$ GeV	$p_x=-0.193$ GeV	$p_y=-0.818$ GeV	$p_z=5.85$ GeV
$\phi_1=-1.80$	$t_1=1.12$ ns	$t_2=3.29$ ns	$t_3=0.243$ ns	$t_4=-0.425$ ns
$E_{SIRA}=8.52$ GeV	$Q_{z,SIRA}=3.02$	$\phi_e^{SIRA}=-0.12$	$\text{Prob}_e^{SIRA}=0.982$	$x_{e,DA}^{SIRA}=0.00$
$y_{e,DA}=0.72$	$Q_{e,DA}^{SIRA}=3.095$ GeV <sup>2</sup>			

Below the data table, there are two views of the detector: "XY View" (top) and "ZR View" (bottom). The XY View shows a top-down view of the detector's octagonal structure with particle tracks. The ZR View shows a side view of the detector with particle tracks. The interface also includes a terminal window on the left showing command-line output, a system monitor window at the bottom left, and a desktop background with the CentOS logo.

HW independant: VirtualBox on 64-bit CentOS7 runs 64-bit CentOS6.  
Outside of DESY: ZEVIS in MPP reads via dCap ZEUS data from MPCDF.

10 / 12

# Current status of H1&ZEUS DP

Data/MC	ZEUS		H1		
DESY archive DESY available online DESY access MPCDF/MPP online+archive MPCDF/MPP access	Processed data/MC ntuples		Raw data/MC, processed data/MC up to 80%		
	Everything				
	NFS, from 2 machines in DESY+BIRD		NFS, from 2 machines in DESY+BIRD		
	As in DESY+raw data Multiprotocol, worldwide with ZEUS VO cert.		As in DESY (online) Multiprotocol, worldwide with H1 VO cert.		
Software					
DESY reconstruction DESY MC generation DESY analysis DESY user storage DESY environment	No		Yes		
	No		Yes		
	Yes		Yes(up to 5y)		
	Yes, limited, on 2 machines in DESY 2 machines in DESY+BIRD(up to 5y)		Yes, limited, on 2 machines in DESY 2 machines in DESY+BIRD(up to 5y)		
MPCDF/MPP reconstruction MPCDF/MPP MC generation MPCDF/MPP analysis capability MPCDF/MPP user storage MPCDF/MPP environment	Yes		Planned		
	Yes		Planned		
	Yes		Planned		
	Yes, unlimited, MPCDF SE+CephFS CentOS7 virtual machine available		Yes, unlimited, MPCDF SE+CephFS CentOS7 virtual machine planned		
Documentation					
DESY analysis primer/manual DESY legacy notes, drafts etc. DESY preservation paper/note	Archived web-server InSpire+DESY library		Archived web-server InSpire+DESY library		
	No		No		
MPCDF/MPP analysis primer/manual MPCDF/MPP legacy notes, drafts etc. MPCDF/MPP preservation paper/note	Relies on DESY InSpire+DESY library		Relies on DESY InSpire+DESY library		
	First draft is available (+A.G.)		Planned		
DESY	Finished	Finished, but not optimal	Significant advance	Moderate advance	Will not be done
MPCDF/MPP	Finished	Finished, but not optimal	Significant advance	Moderate advance	Will not be done

# Current status of OPAL&JADE DP

Data/MC	OPAL, Host=CERN		JADE, Host=DESY		
Host data	Raw/processed, data/MC on CASTOR/EOS		Probably		
Host access	Multiprotocol, CERN		No		
MPCDF/MPP available online	Raw/processed, data/MC		Raw/Processed Data/MC		
MPCDF/MPP archive	Raw/processed, data/MC		Raw/Processed Data/MC		
MPCDF/MPP access	Multiprotocol, worldwide with OPAL VO cert.		Multiprotocol, worldwide with ZEUS VO cert.		
Software					
Host reconstruction	Yes		No		
Host MC generation	Yes		No		
Host analysis	Yes		No		
Host user storage	CERN users only		No		
Host environment	Default CERN		No		
MPCDF/MPP reconstruction	Yes (M.S.)		Planned		
MPCDF/MPP MC generation	Yes (M.S.)		Update/Planned		
MPCDF/MPP analysis	Yes		Planned		
MPCDF/MPP user storage	Yes, unlimited, MPCDF SE+CephFS		Not needed		
MPCDF/MPP environment	CentOS7 VM available		AIX → Fedora17 PPC or CentOS7 x86_64 VM planned		
Documentation					
Host analysis primer/manual	CERN web-server		No		
Host legacy notes, drafts etc.	InSpire+CERN library		InSpire+DESY+J.O.		
Host preservation paper/note	No		No		
MPCDF/MPP analysis primer/manual	Relies on CERN		Yes (Update!)		
MPCDF/MPP legacy notes, drafts etc.	InSpire+CERN library		InSpire+DESY library+		
MPCDF/MPP preservation paper/note	Yes, early stage update		Yes (Update!)		
Host	Finished	Finished, but not optimal	Significant advance	Moderate advance	Will not be done
MPCDF/MPP	Finished	Finished, but not optimal	Significant advance	Moderate advance	Will not be done

# Software: DatABriCxx / dbrx

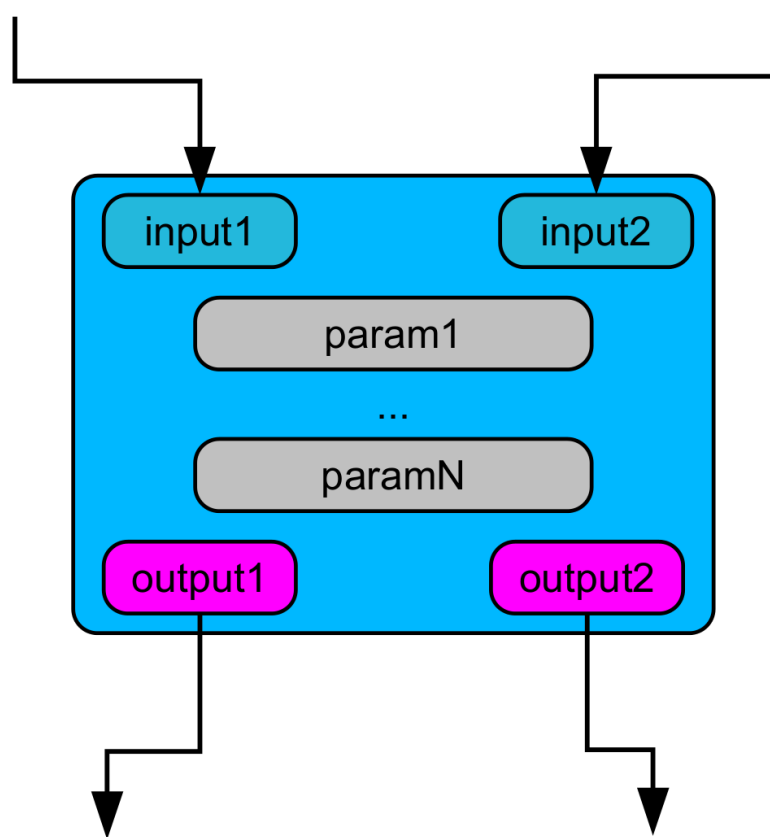
- Many small experiments have organically grown computing

Oliver Schulz

- Does not scale to bigger collaborations
  - Does not scale to much bigger data sets
  - Does not scale to much longer lifetimes
- Provide a framework
    - Selection and organisation of algorithms
    - Configuration and parameter handling
    - Services (IO, calibrations, ...) provided

# dbrx

Algorithm: receive input from upstream, hold parameters (configuration), provide output to downstream (transformation)



1 to n: Mapper (event loop),  
1 to 1: transform (hits to tracks,  
fits, ... )  
n to 1: reducer (histgrammer,  
Output, final calculations, ...)

# dbrx

Brics are analysis building blocks

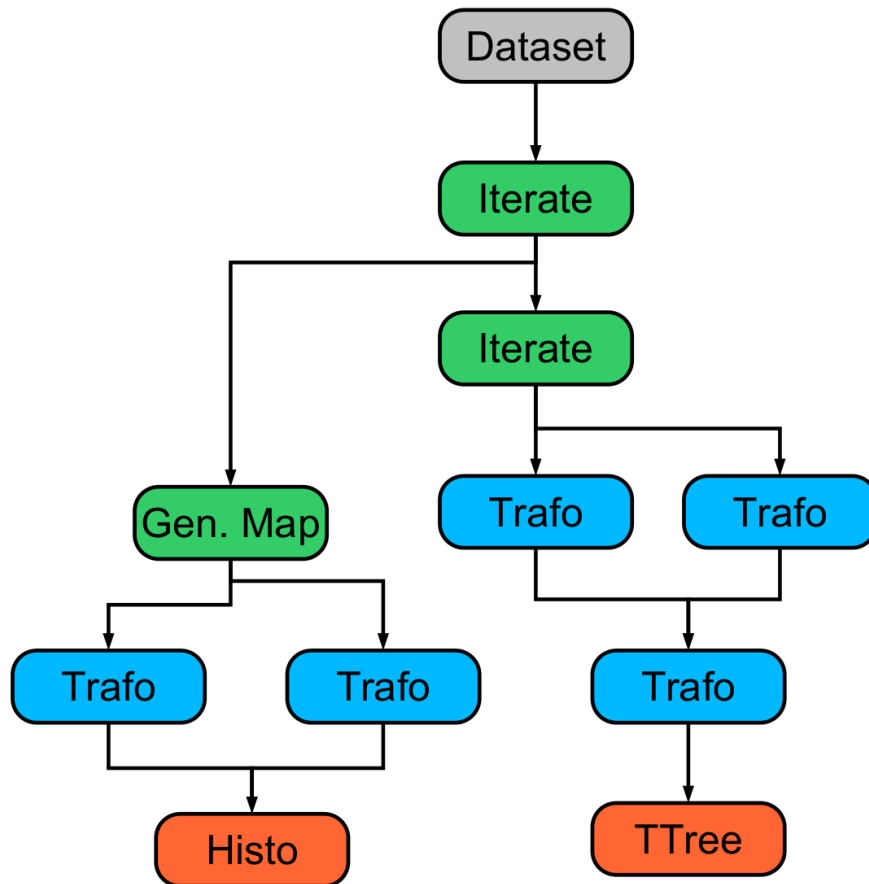
Brics form a directed acyclic graph  
a.k.a. tree

Framework schedules brics according  
to data flow dependencies,  
naturally parallel

GUI possible via http

GeDet, CRESST, GERDA (partially),  
COBRA, ...

LGPL, get it from github





# Resources

- **MPCDF**
  - MPP Linux cluster: 84 nodes, ~1.000 cores, > 2.5 PB storage
  - Hydra: 4.110 nodes, ~83.000 cores, fast interconnect, 338 nodes with dual Nvidia Tesla K20X
- **LRZ**
  - SuperMUC: >12.000 nodes, 241.000 cores, fast interconnect
- **Excellence Cluster Universe**
  - C2PAP: 128 nodes, >2000 cores, fast interconnect, SuperMUC integration

# Resources

- MPP
  - > 200 desktop PCs via condor batch system (need to sort out kerberos login)
  - Brand new PC with Intel Xeon Phi 5100
    - 60 core, 8 GB, 1 GHz
    - Theory group

# Summary

- Scientific computing essential for our success
- Many activities at MPP
  - From software development to data preservation
- Resources: MPP, MPCDF, LRZ, C2PAP
- All centers provide application support
  - Porting to parallel platforms, performance tuning, ...
- Transition to HPC in many of our research areas