Scalable stochastic tracing of distributed data management events

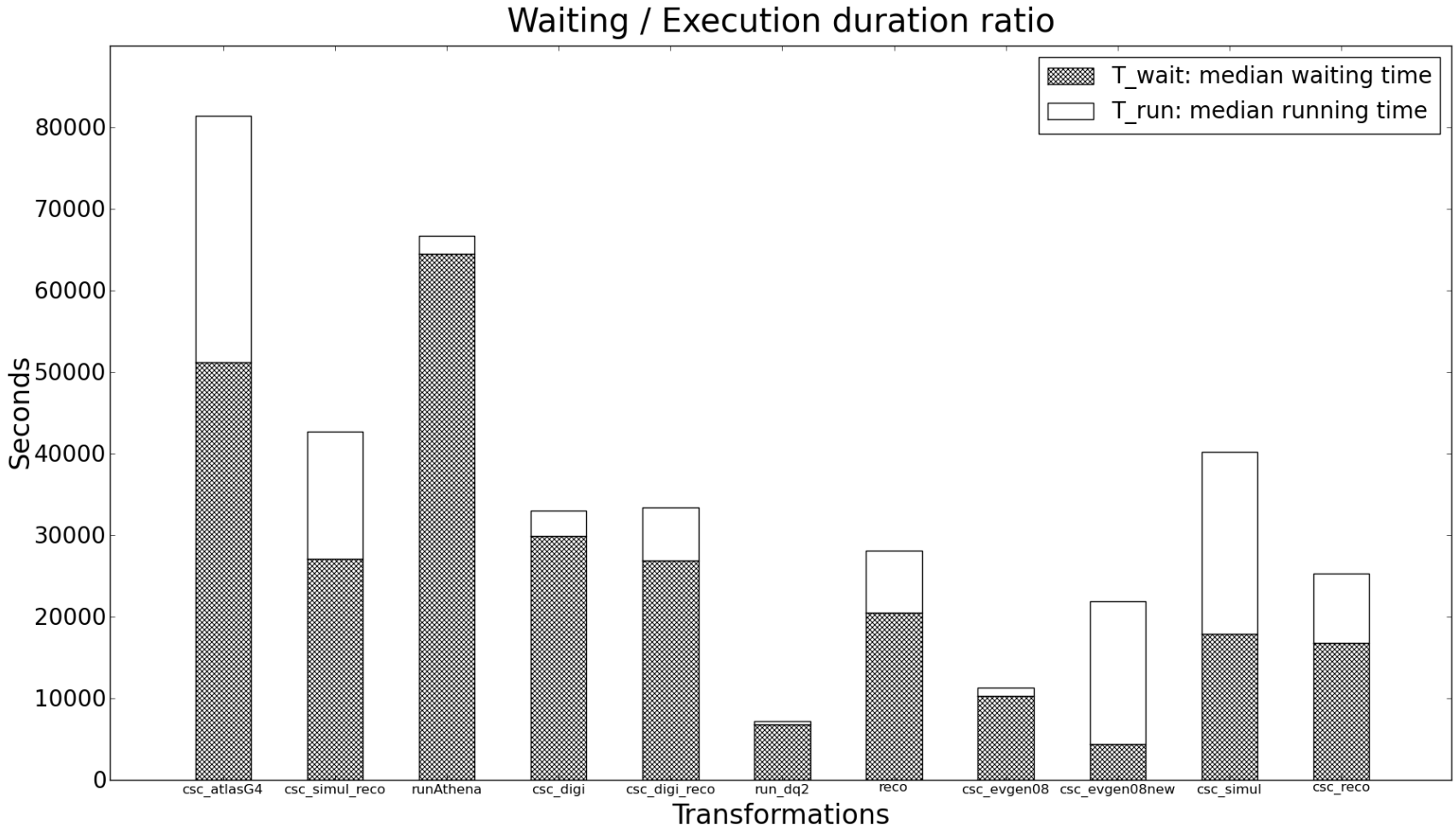
Mario Lassnig

mario.lassnig@cern.ch

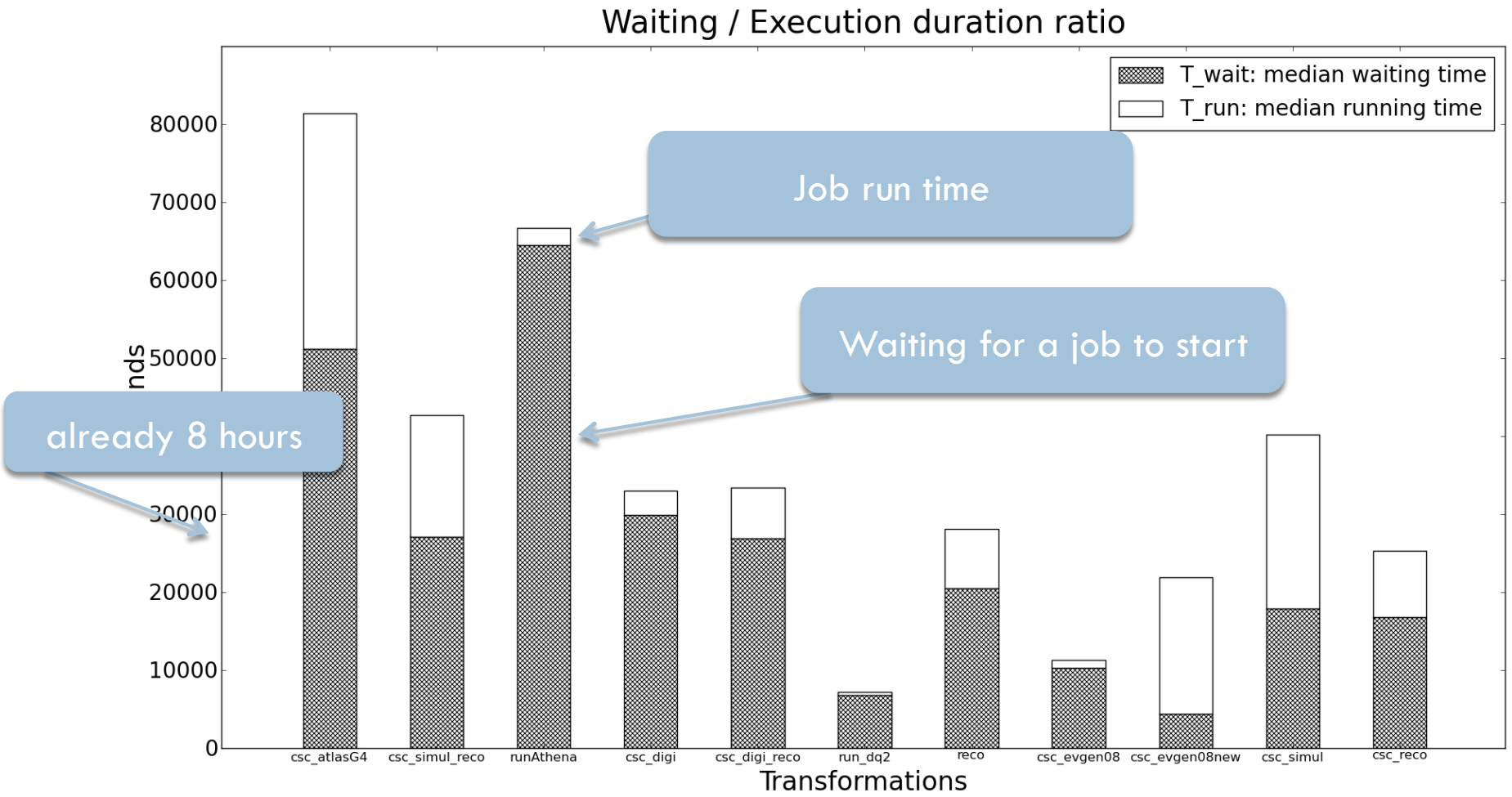
ATLAS Data Processing – CERN Physics Department
Distributed and Parallel Systems – University of Innsbruck

EGI Technical Forum, 14-09-10, Amsterdam

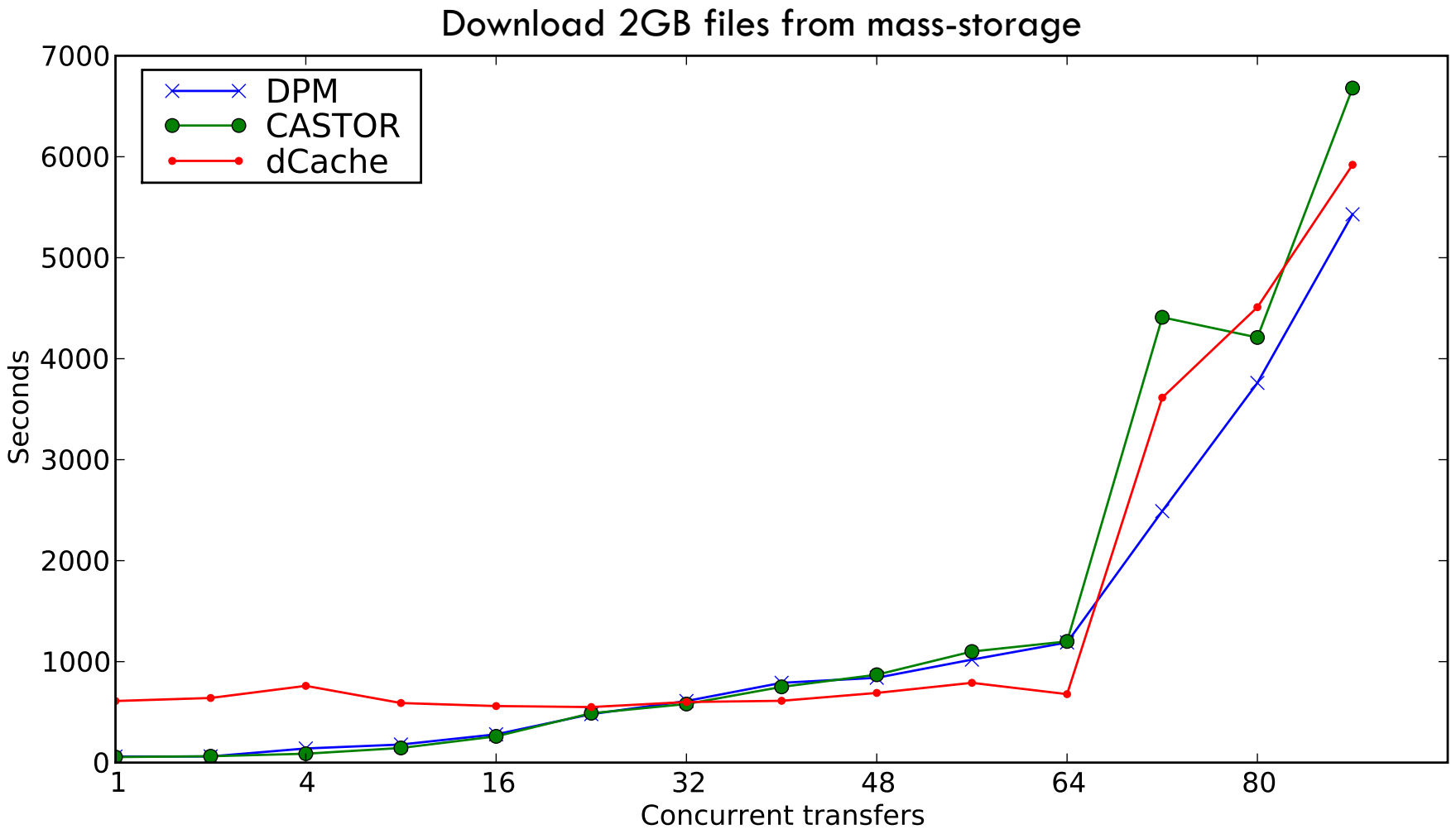
Why is data management important?



Why is data management important?

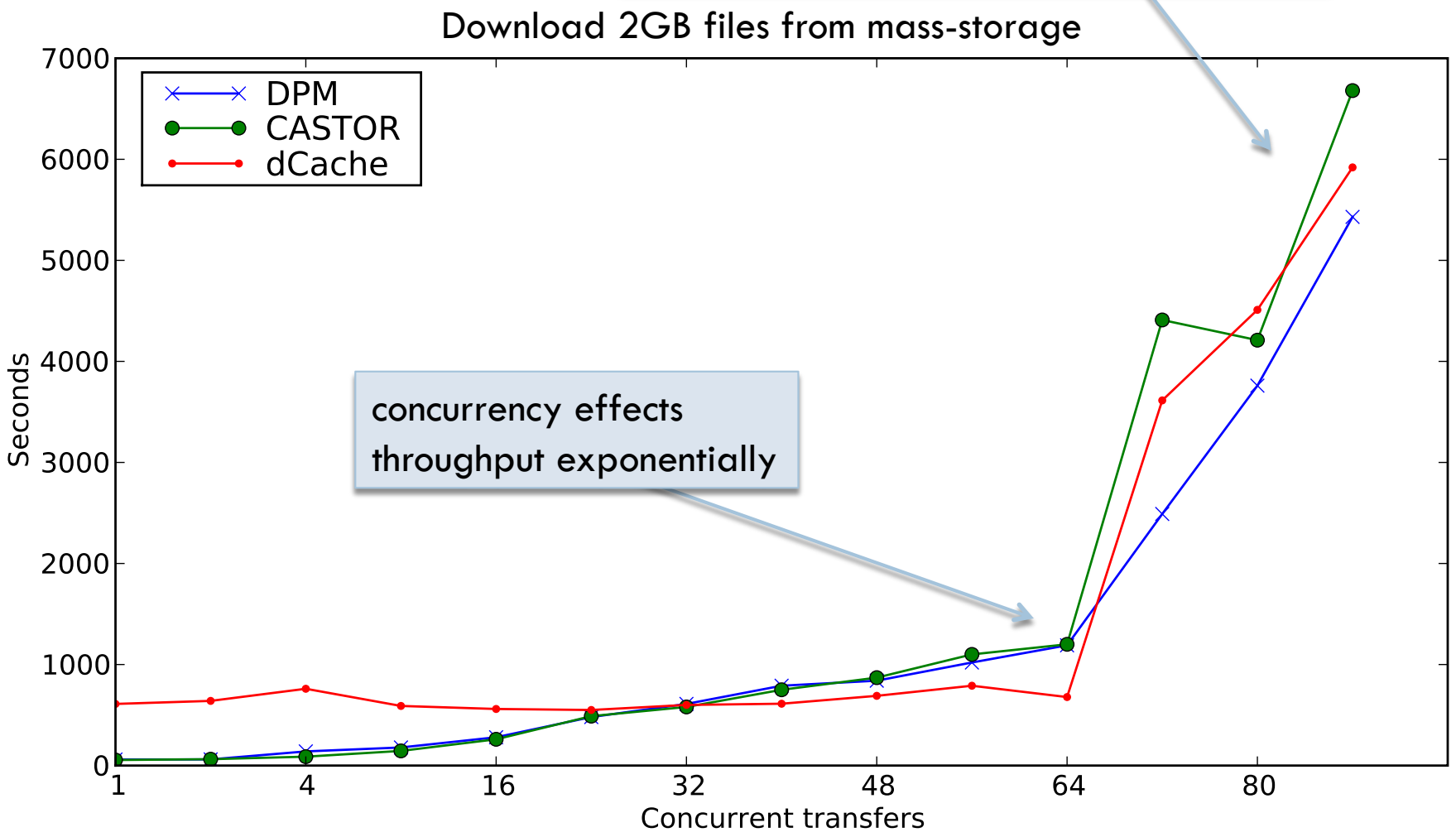


The basic data management problem



The basic data management problem

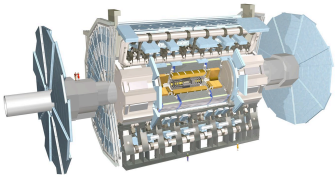
components do not break but
degrade until they fail QoS requirements



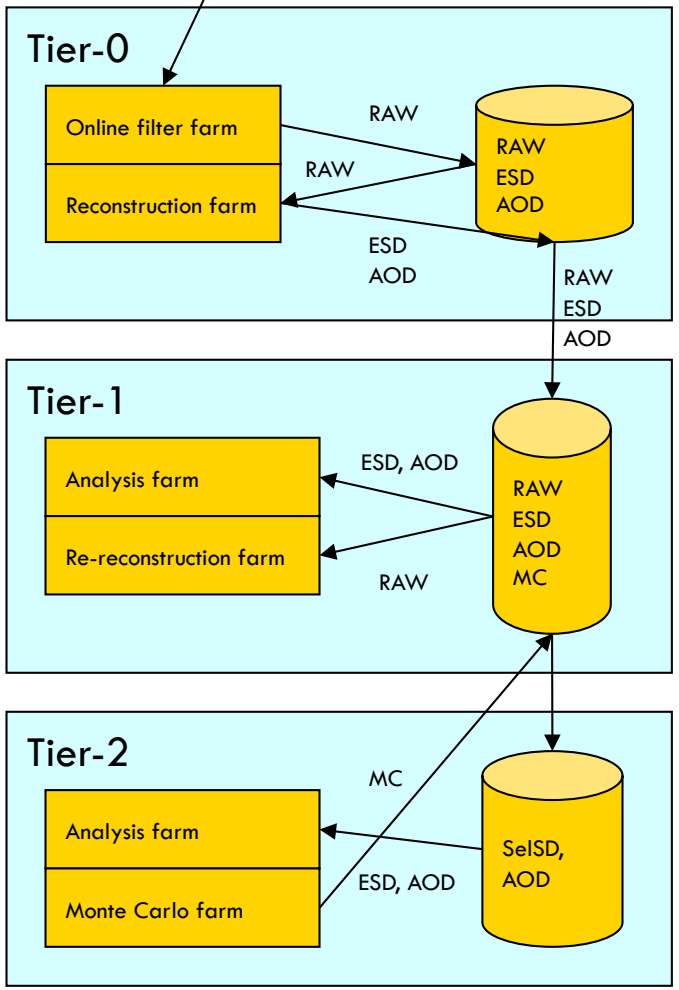
Don Quijote 2 (DQ2)

- Manage all ATLAS experiment data
  - provide data *query* / *transfer* / *access* / *provenance* capabilities for
    - users
    - analysis frameworks
  - between dedicated and on-demand resources (sites)
    - data centres
    - university installations & laptops
  - high performance
    - sustain 2000 MB/sec throughput aggregate, 1 mio file transfers daily (2010 estimates)
    - replication of data for parallel access
  - keep all data consistent
    - while still allowing high-latency distributed read-writes
  - easy to use
- Optimisation problem
  - $\forall$ user/framework *u*, request *r*: *min*(time-to-first-byte(u_n,r_n) $\wedge$ time-to-last-byte(u_n,r_n)
- Software stack is called *Don Quijote - Version 2* (*DQ2*)
  - managing all ATLAS data since 2005

Don Quijote 2 (DQ2)

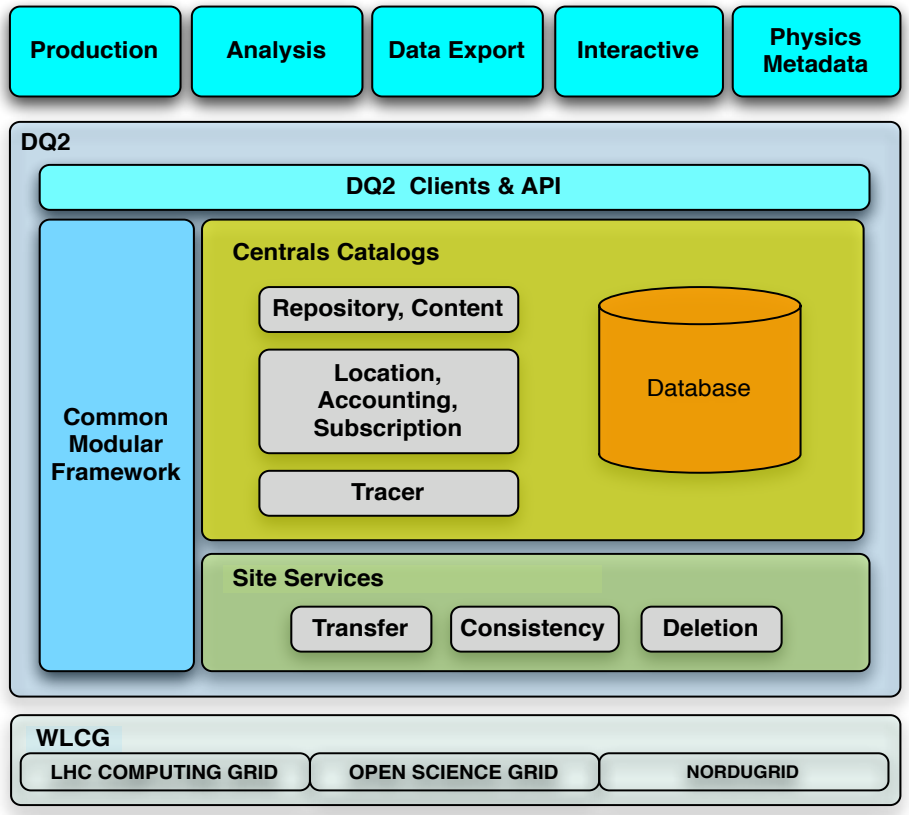


- Basic unit is a *data set*
  - logical collection of files
  - annotations of file data
  - subscriptions of data to sites
- Decentralised structure
  - make use of already deployed Grid technologies
- Sites are organised in Tiers (Computing Model)
  - hierarchical
  - each Tier has a specific role
    - Tier-0 (CERN)
      - record RAW detector data
      - distribute data to Tier-1s
      - calibration and first-pass reconstruction
    - Tier-1s (10 large data centres)
      - permanent storage
      - capacity for reprocessing and bulk analysis
    - Tier-2s (~100 institutes, some bigger, some smaller)
      - Monte-Carlo simulation
      - user analysis



Don Quijote 2 (DQ2)

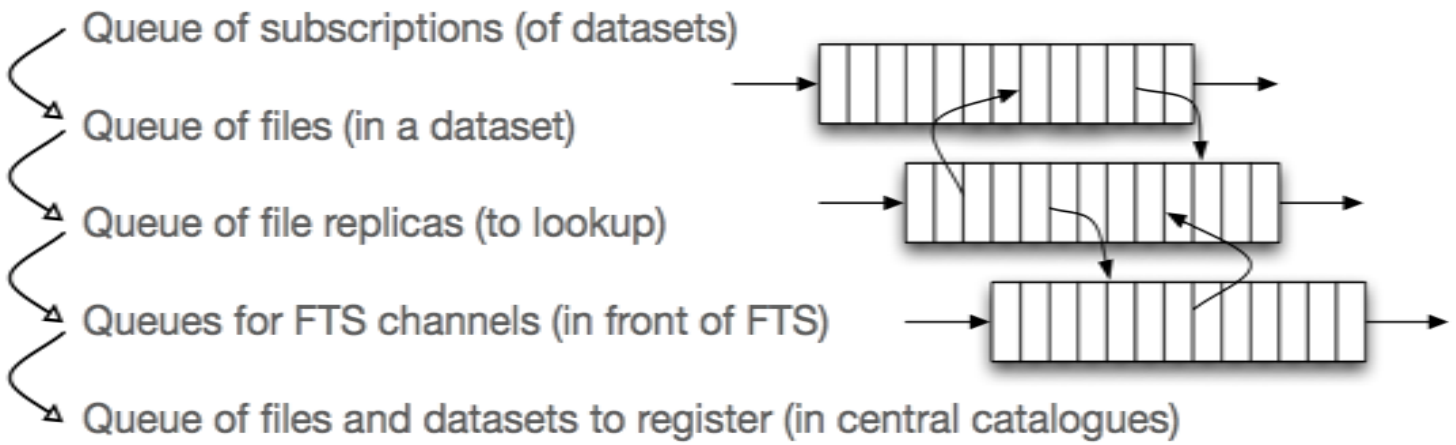
- Centralised catalogues (HTTP / Oracle)
  - Repository (datasets)
  - Content (files in datasets)
  - Location (datasets at sites)
  - Accounting (user on data)
  - Subscription (dataset to site)
  - Tracing (framework/user activity)
- Distributed site service agents
  - every site has one
  - enact / monitor transfers (dataset subscriptions)
  - consistency check and repair
  - deletion of data
- Clients and API
- File Catalogs (LFC) and Transfer Service (FTS)
  - not part of DQ2 but WLCG foundation
  - logical to physical mapping of files
  - physical transport of files



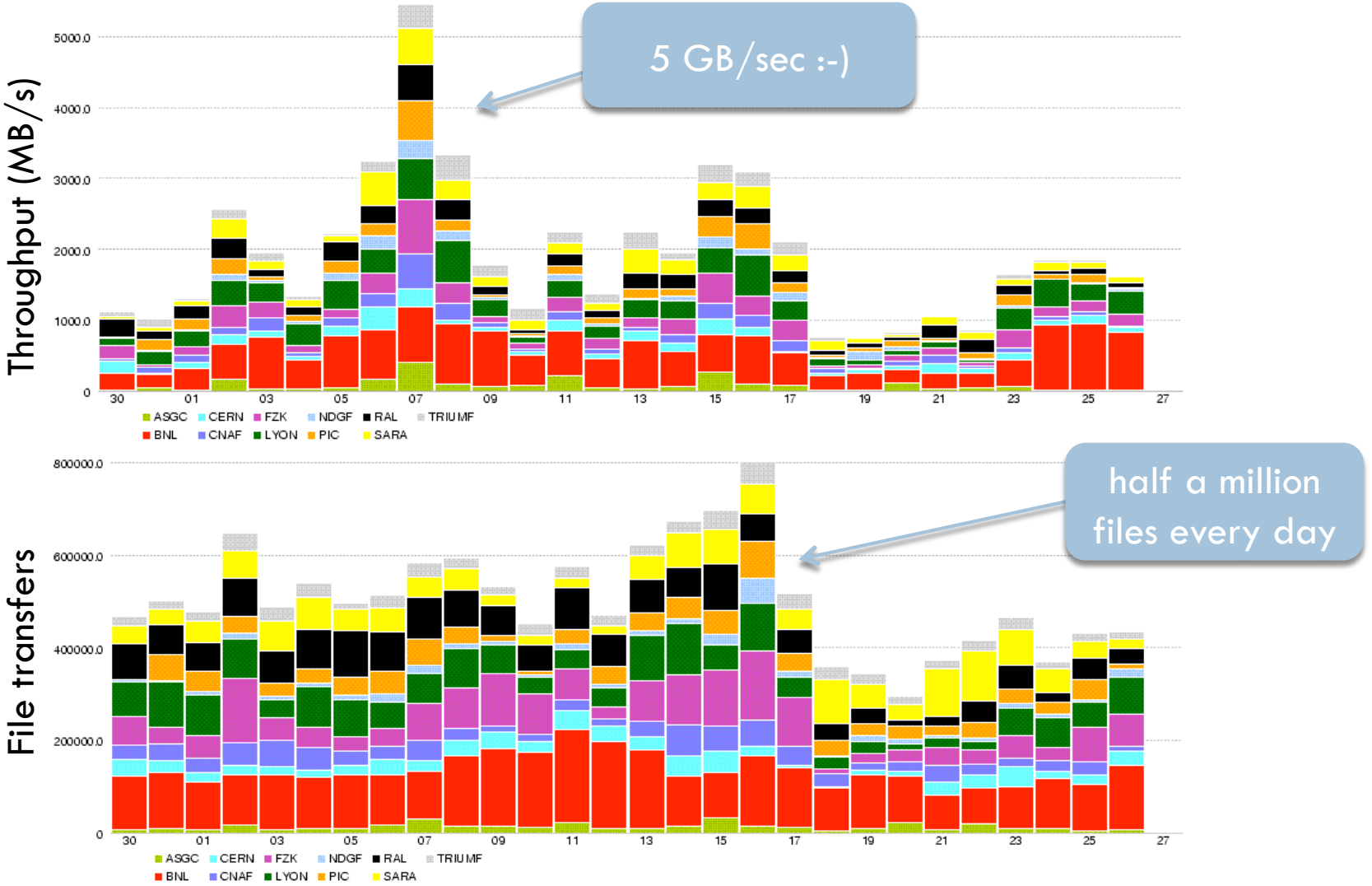

```
from dq2.clientapi.DQ2 import DQ2
dq2 = DQ2()
dict = dq2.listDatasets('test.xyz*')
```

Don Quijote 2 (DQ2)

- DQ2 is a parallel multi-queue-based system
  - site services schedule all transfers to achieve a configured min-max QoS
    - e.g.: complete dataset, channel throughput, site utilisation
  - late reshuffling of queues
    - lots of small files in ATLAS (avg. 100 000 files backlog per multi-queue)
    - (high,med,low)-priority datasets
  - exponential back-off retrial strategy
    - no prediction (too slow)
    - feedback-based only (faster to ask for forgiveness than permission)

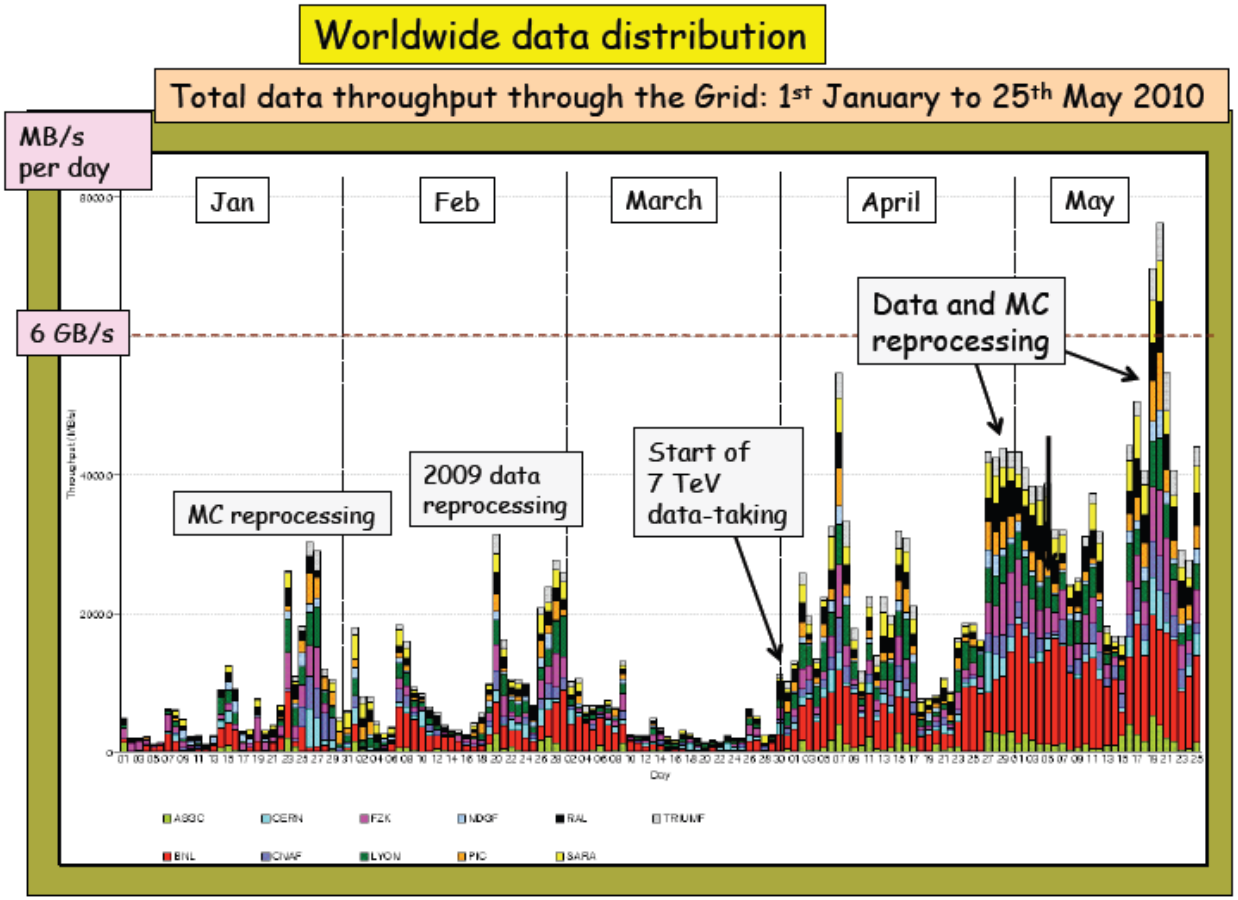


Directed transfers



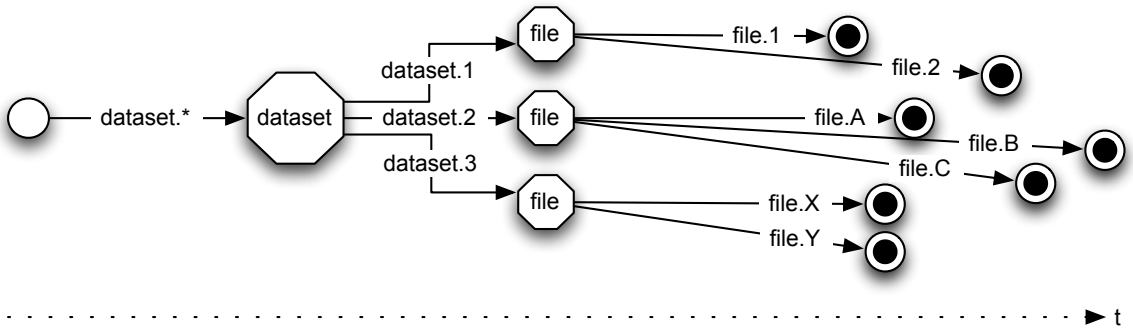
Directed transfers

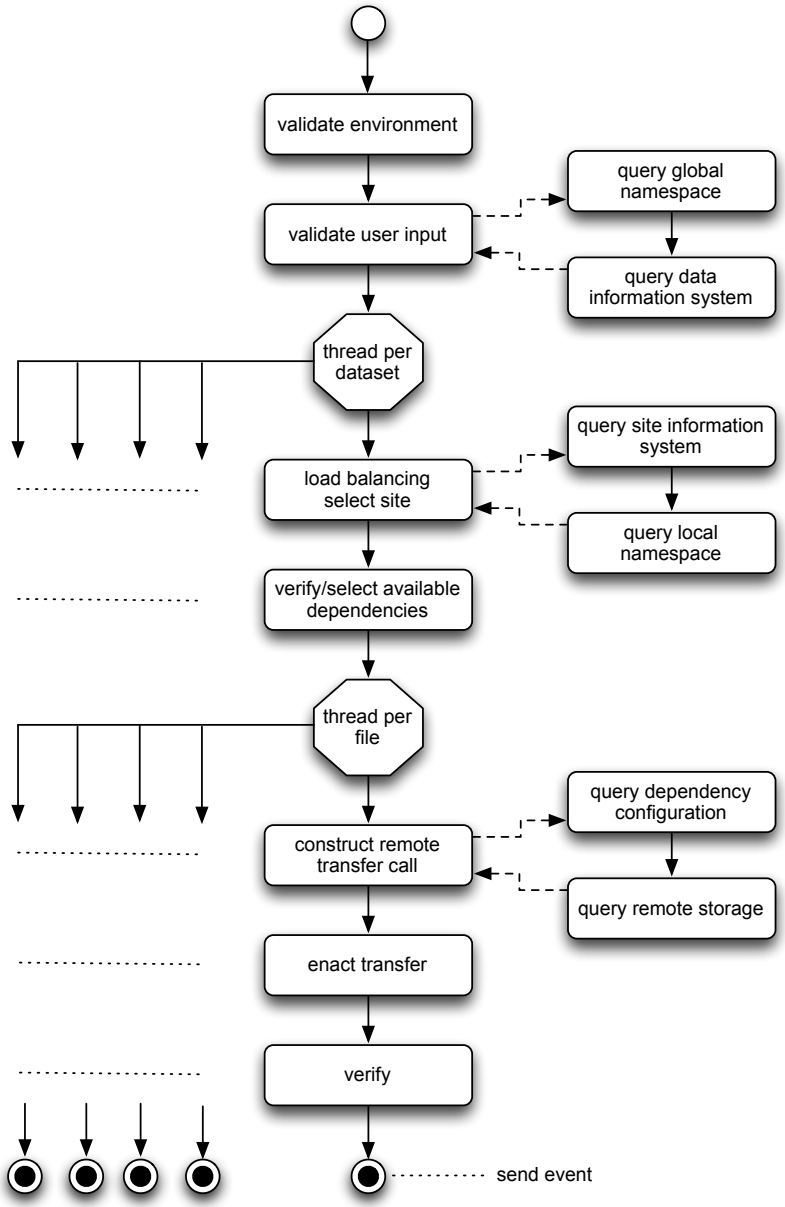
- burst behaviour everywhere, but we can still cope



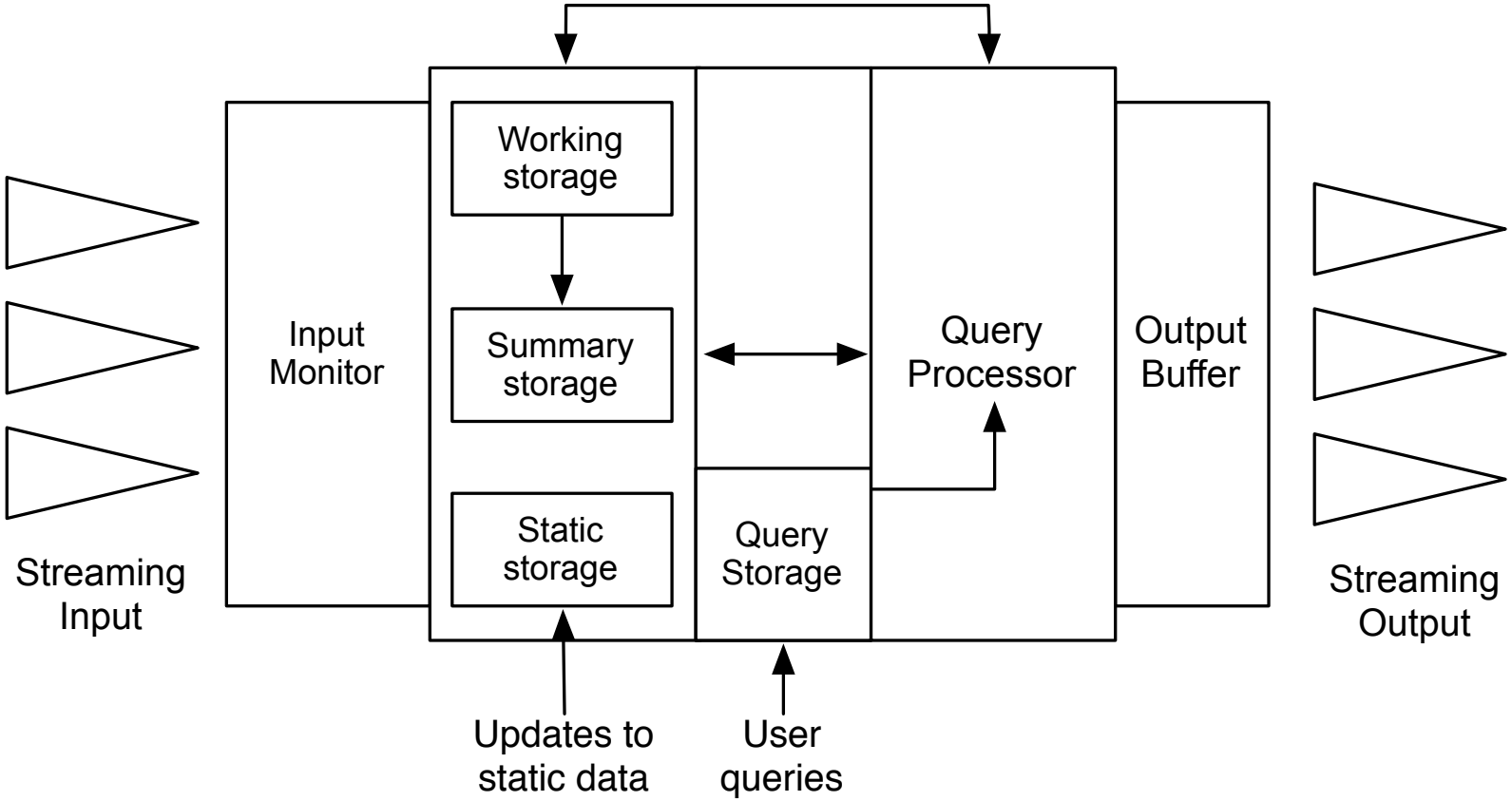
Capturing run-time information

- Users interact via DQ2Clients
  - create/retrieve datasets
  - immediate interaction, no schedules
- Original metric:
  - "angry email from a physicist"
- We needed to understand what's happening with the system, and the existing monitoring infrastructures simply couldn't keep up
  - arrival rate of events
  - information overload vs. not enough information
  - volunteered/locked-down infrastructure
- Build the monitoring directly in the DQ2 client application layer and capture workload streams

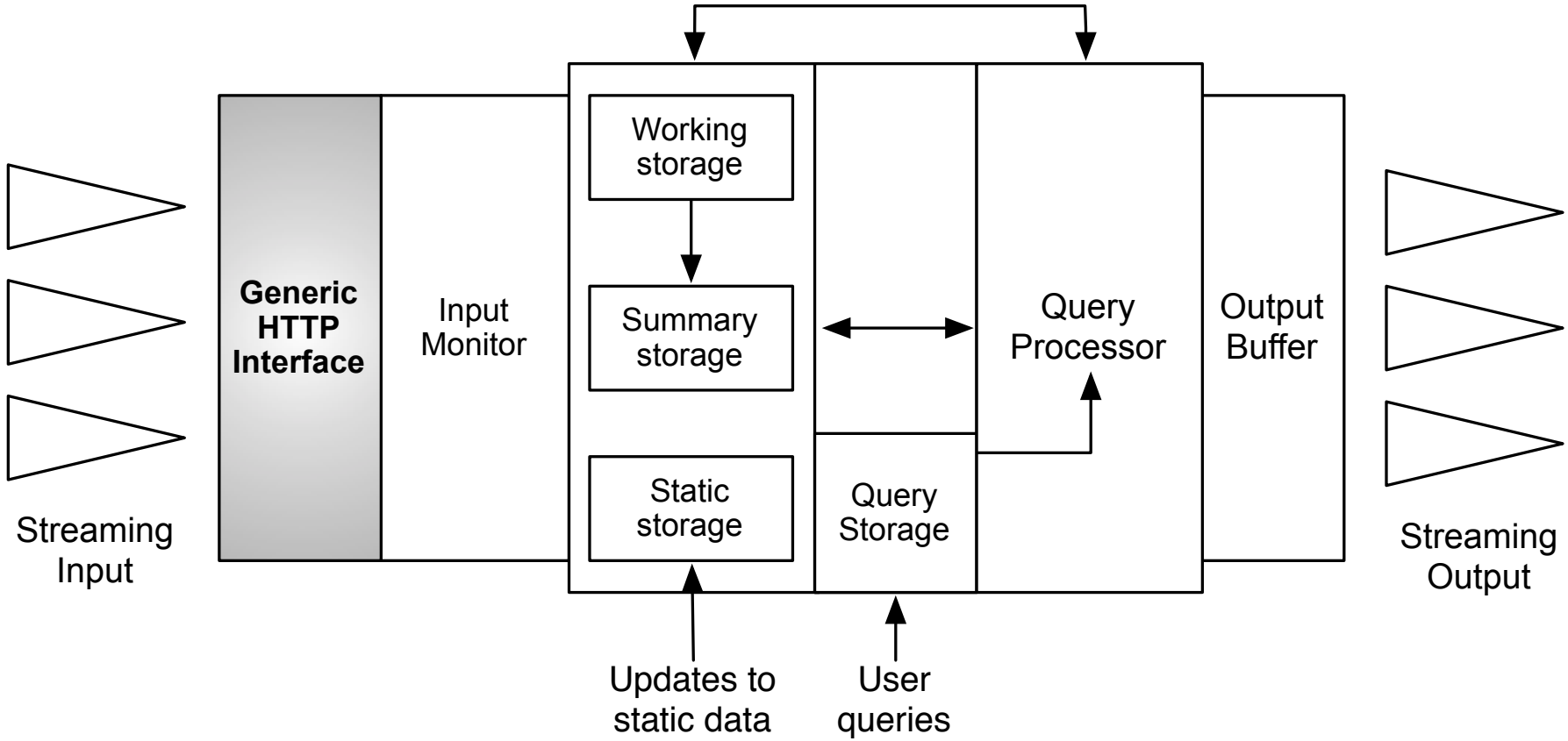




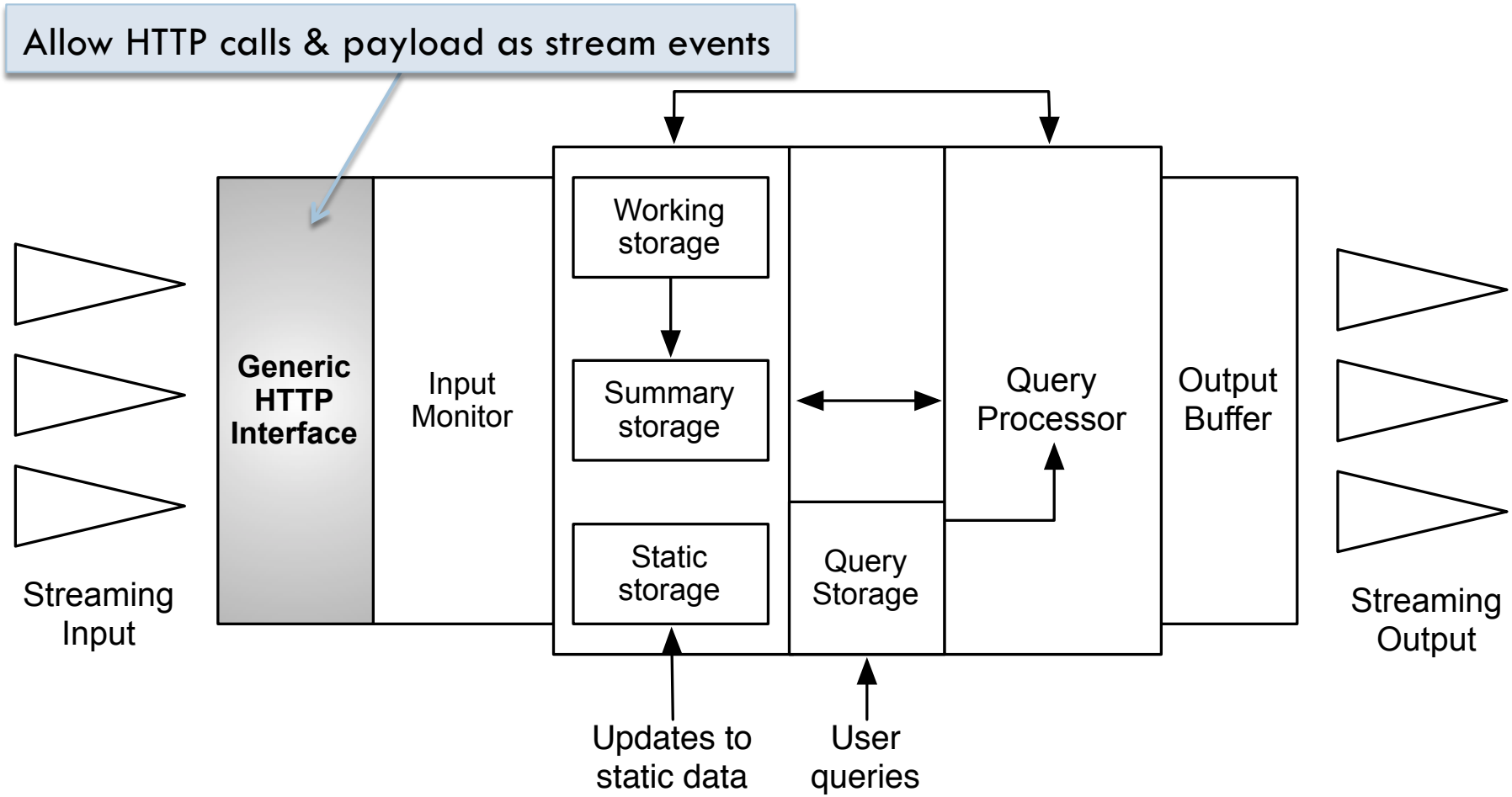
Design principles and reference architecture



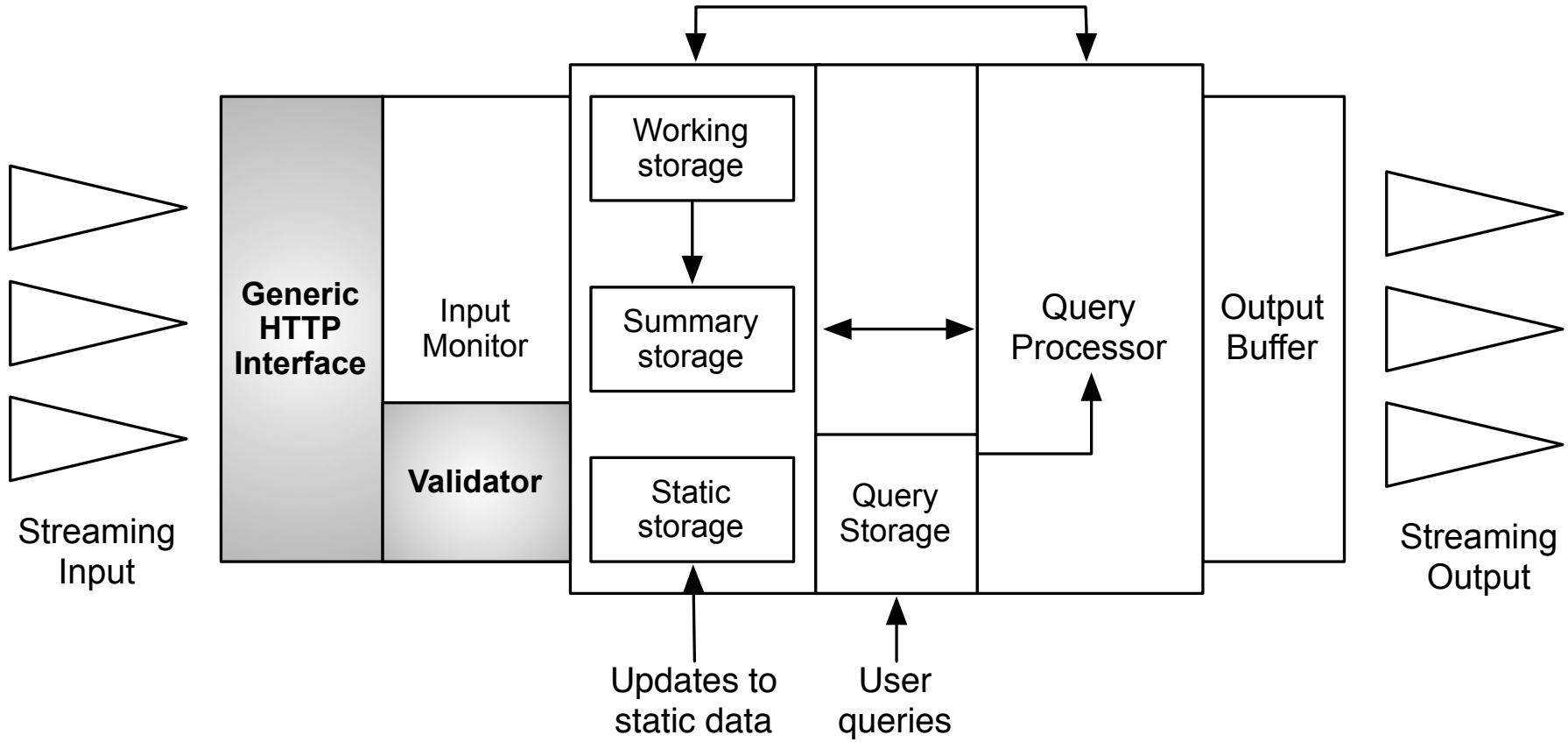
Extensions for concealed environments



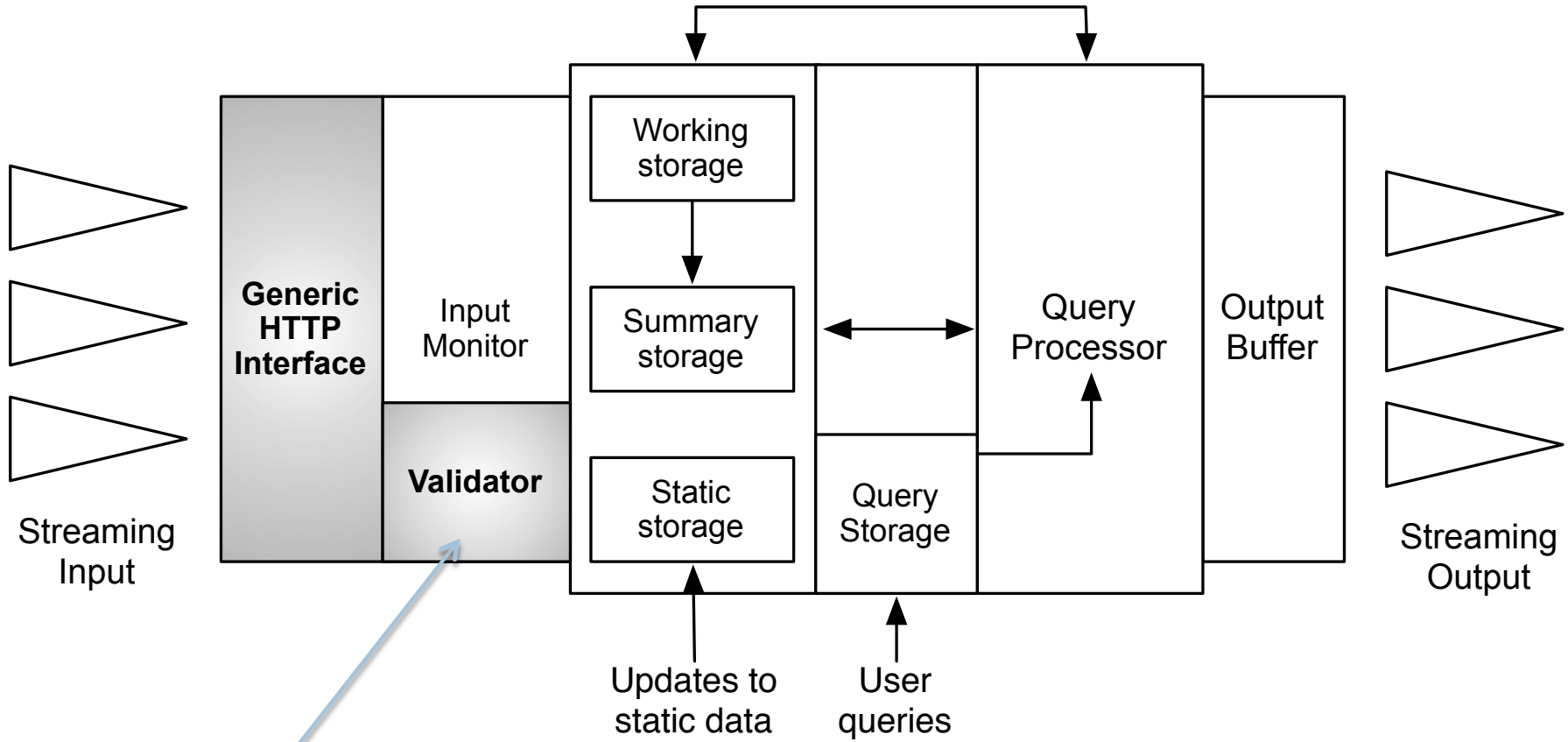
Extensions for concealed environments



Extensions for concealed environments

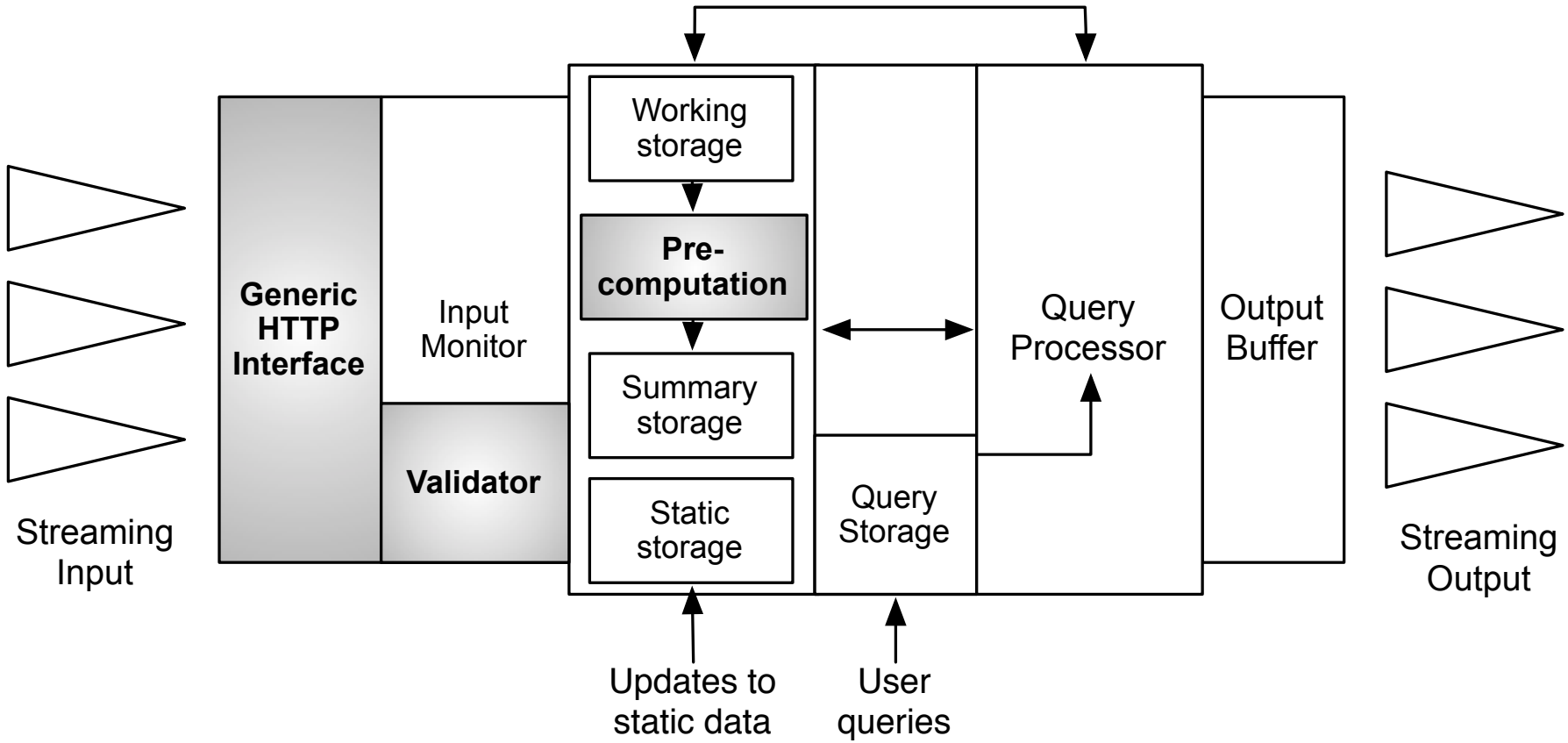


Extensions for concealed environments

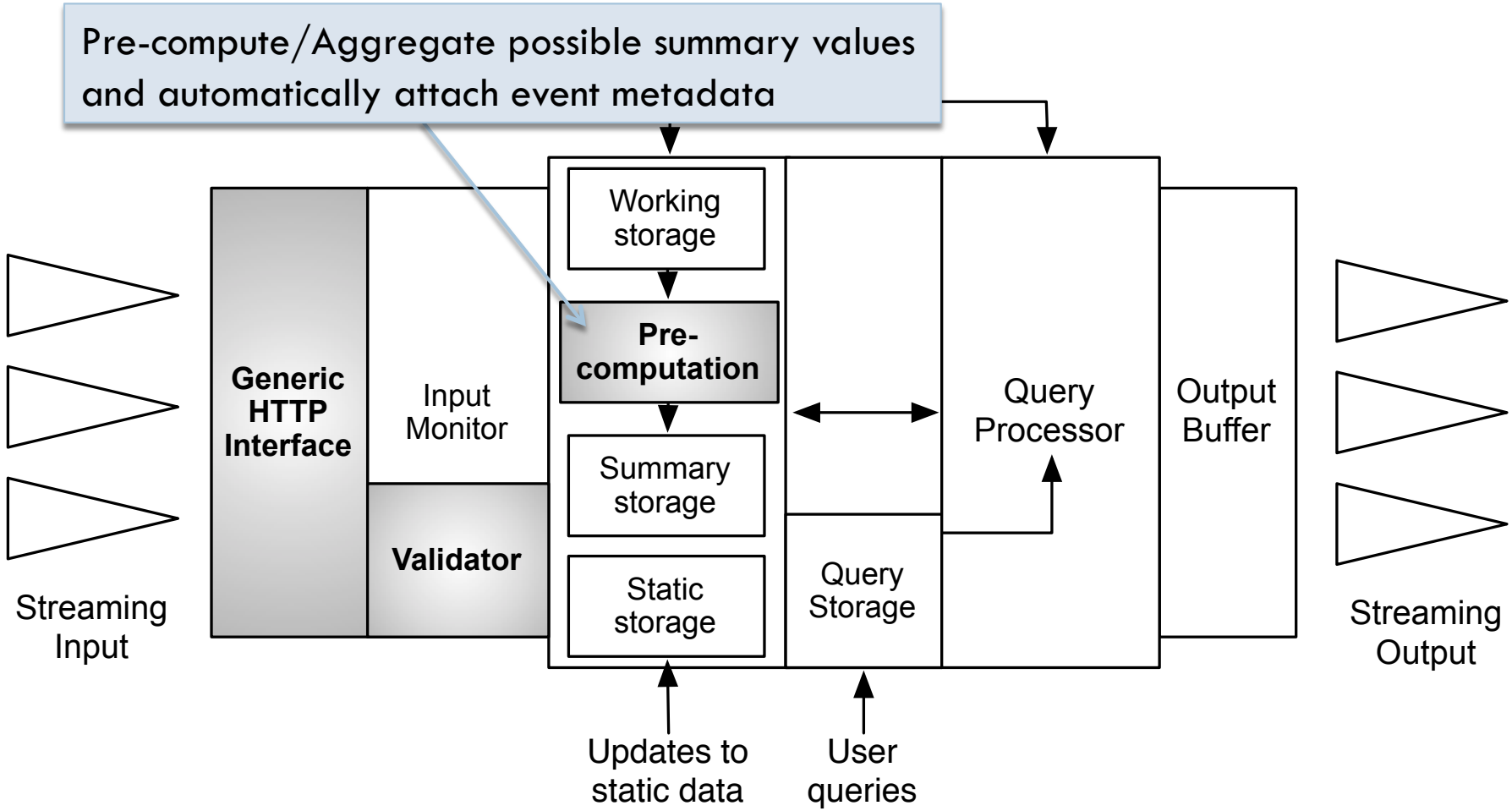


Validate the payload (basic data types) and the order of events

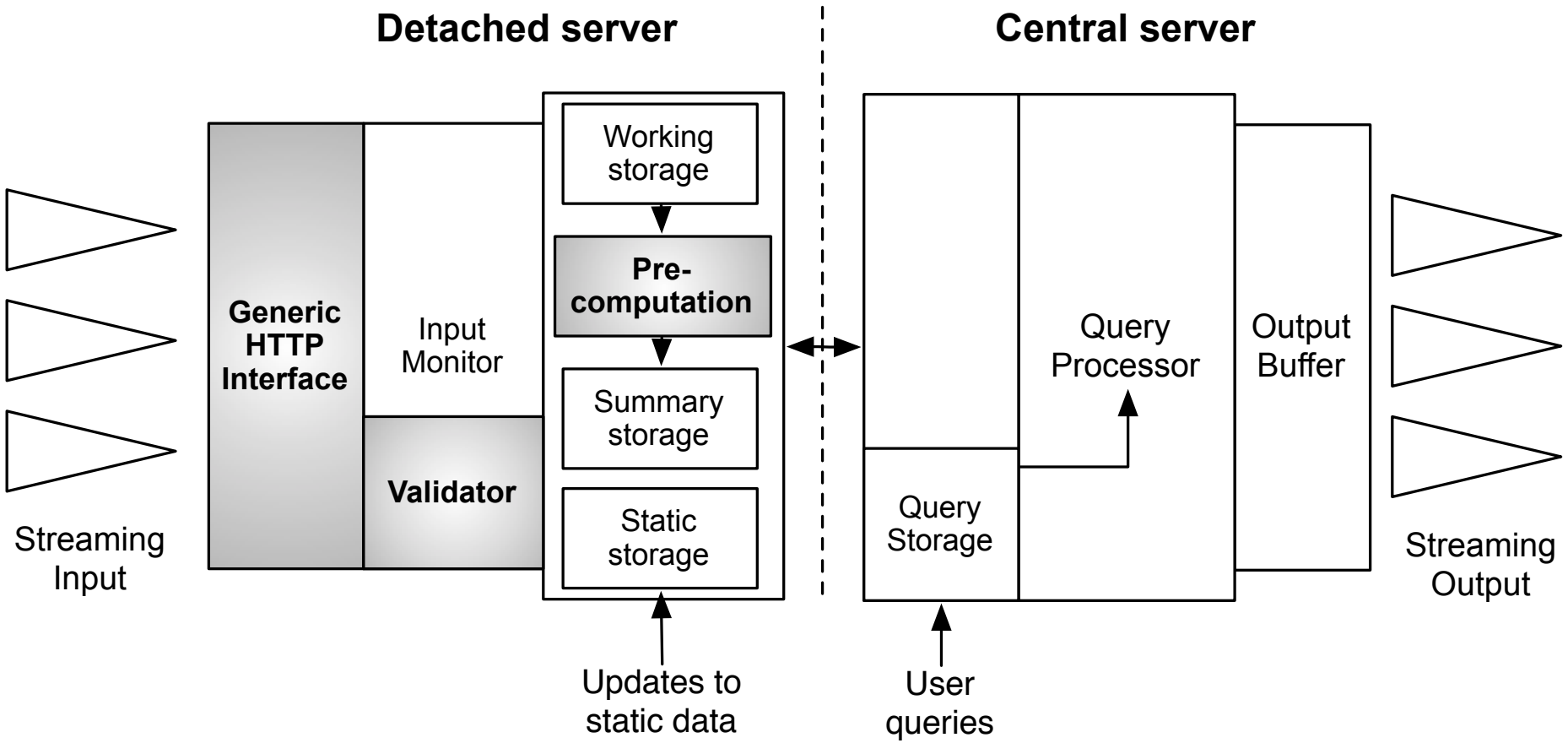
Extensions for concealed environments



Extensions for concealed environments

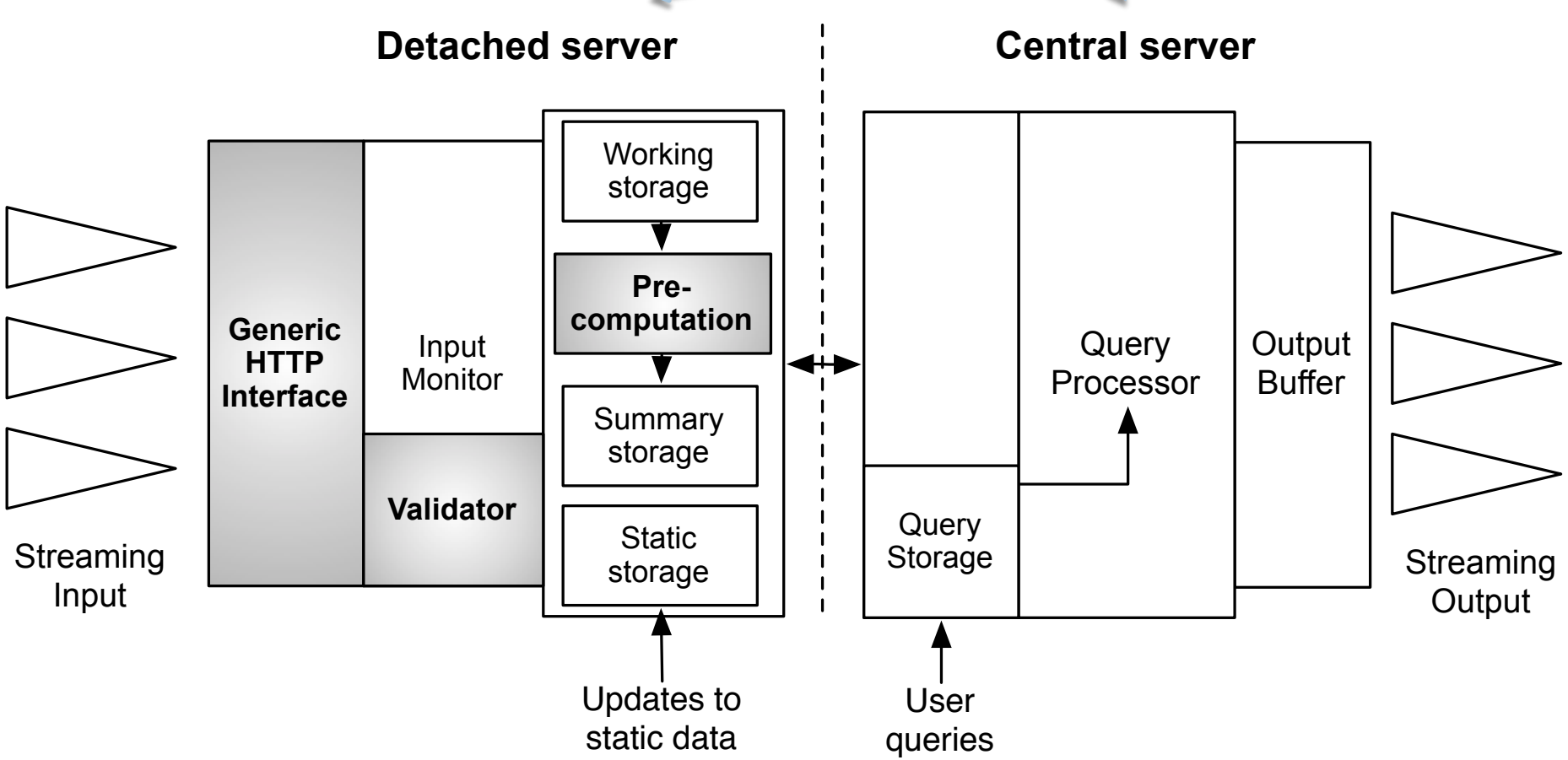


Extensions for concealed environments

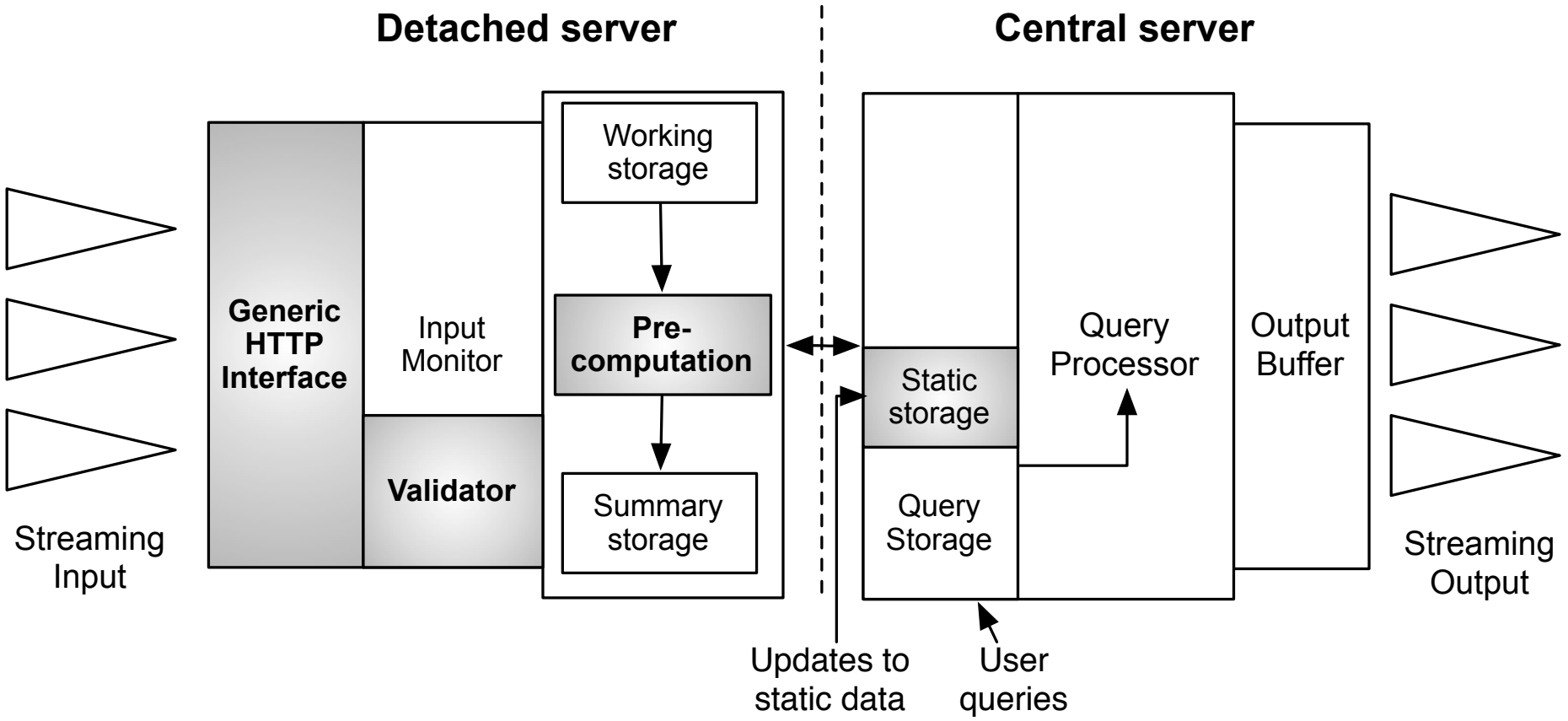


Extensions for concealed en

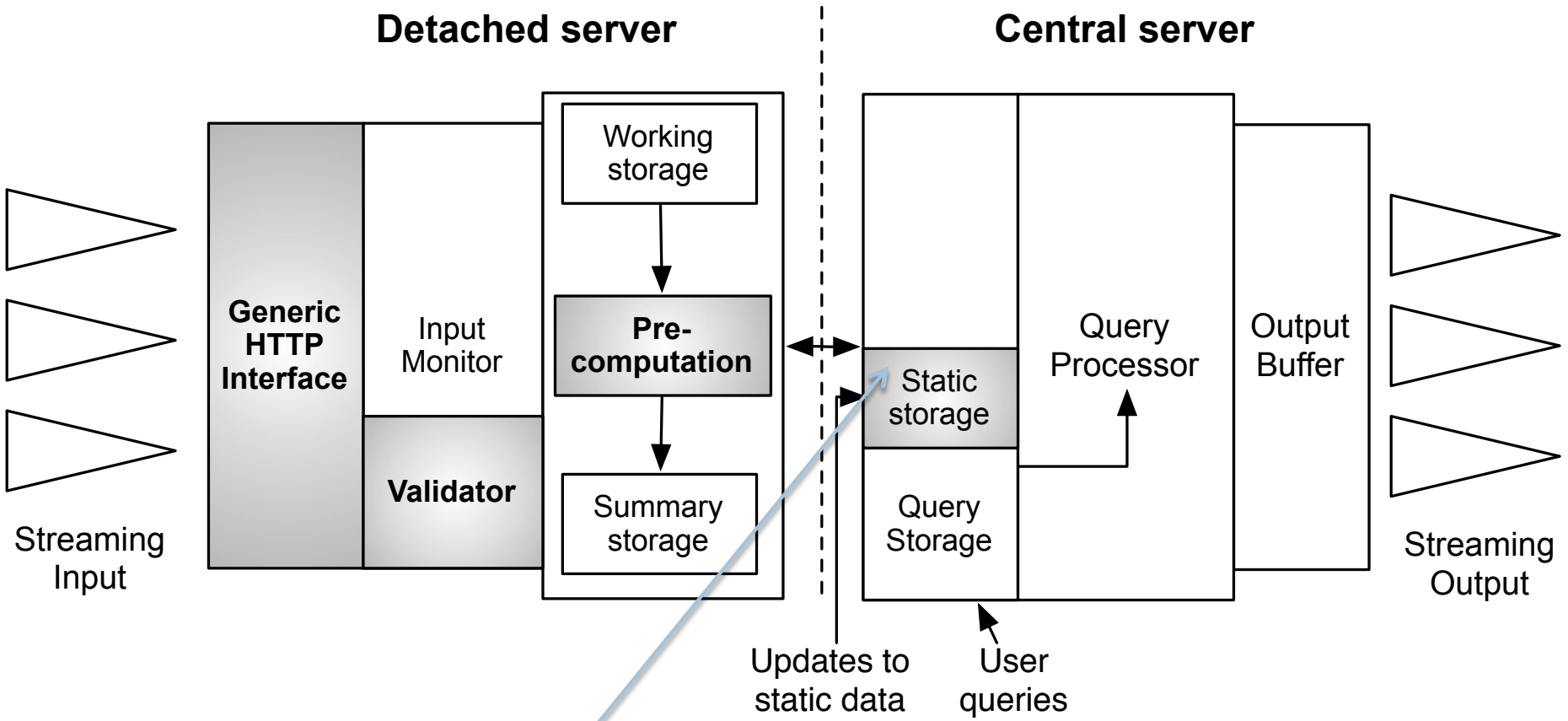
Split into many detached working areas and a centralised data mining area



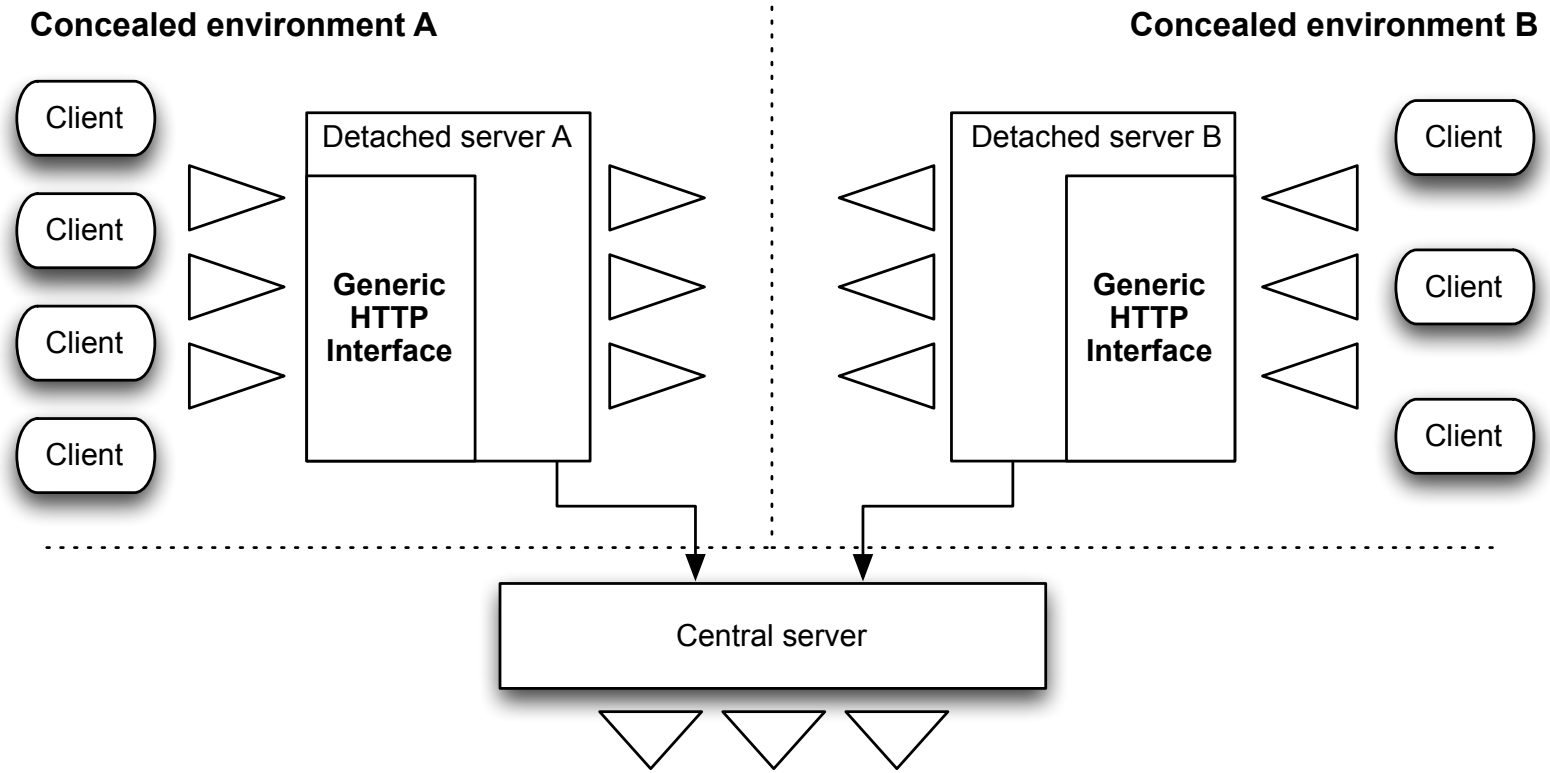
Extensions for concealed environments



Extensions for concealed environments



Capturing run-time information



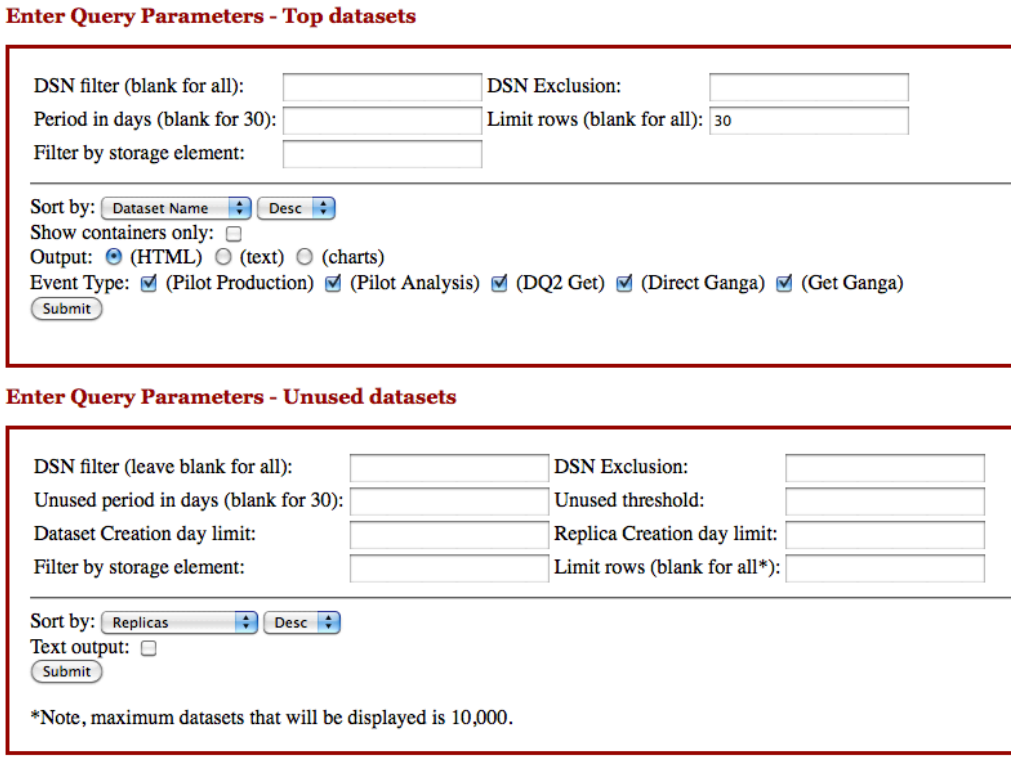
Stochastic tracing

- Stochastic tracing data yields useful implicit information
  - system component behaviour (storage elements, middleware, …)
  - user behaviour (access patterns, data popularity, …)
  - without resorting to specialised monitoring infrastructures

- As of September 2010
  - tracing file events since May 2008, ramp up in November 2008
  - 580.000.000 tracer events
  - *event arrival rate* of 25 ± 3 Hz
  - *detached server* computational overhead 0.03 ± 0.01
  - *central server* computational overhead 0.01 ± 0.01
  - *network* throughput overhead 0.04 ± 0.02

Popularity service

- First application to use this data
  - helps operators decide what to replicate and what to delete
  - immense amount of data (~35Hz, 24/7)
  - incremental generation of daily, weekly, ... reports
  - basic data mining support for association rules in trace events



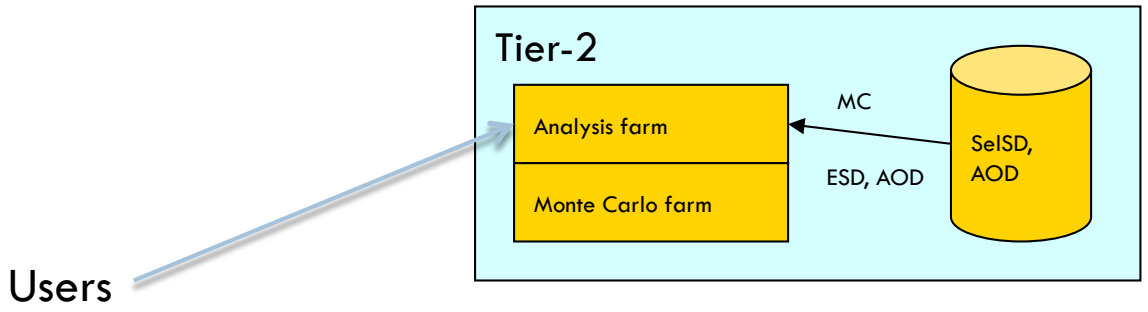
Popularity service

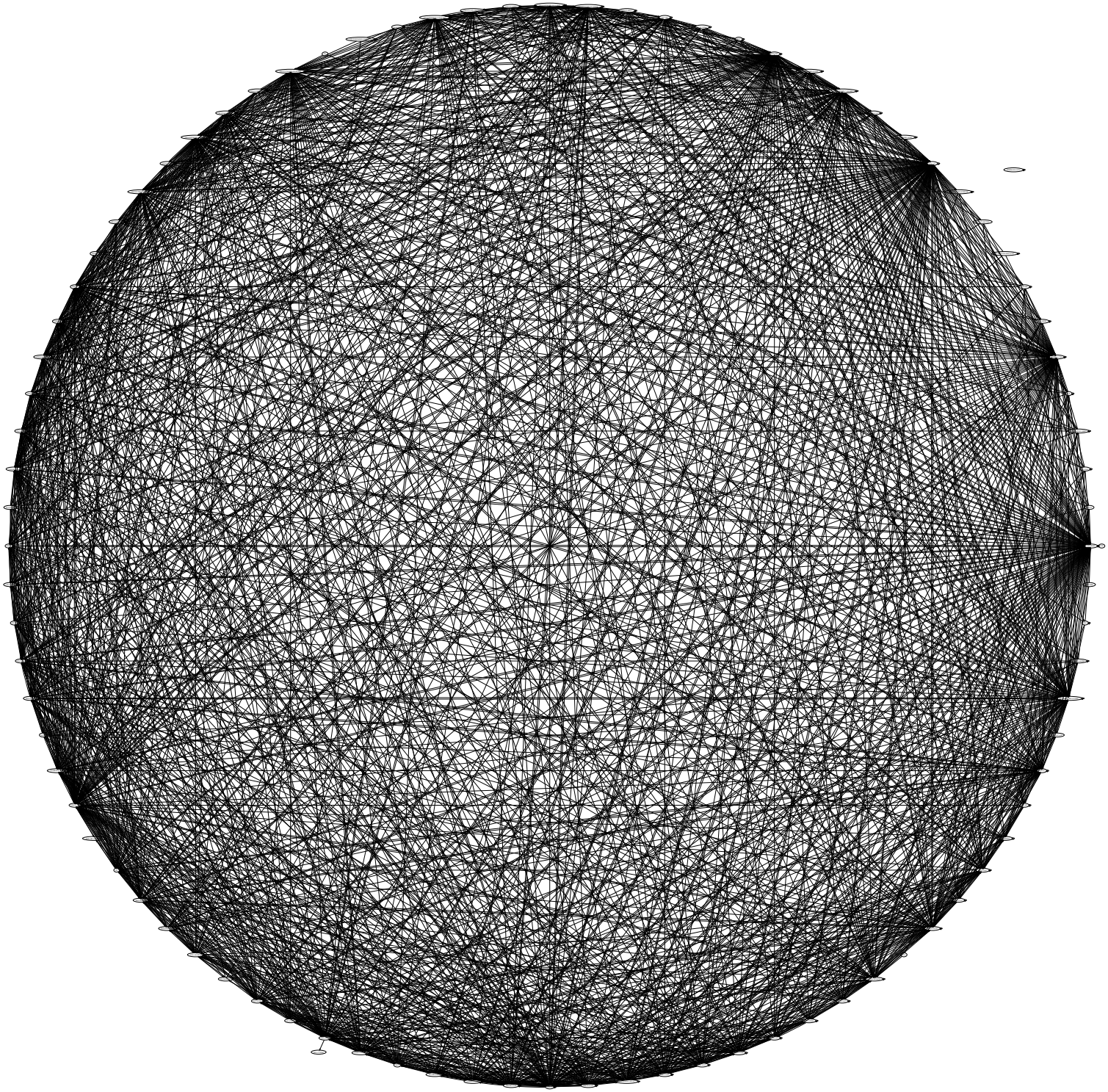
Dataset Name	Sites	Locals	Users	File Requests	File L-Requests	Ops	Local Ops	Replicas	File No	Dataset Sizes	Last Read
ddo.000001.Atlas.Ideal.DBRelease.v100701	86	86	304	1536713	1536710	1536050	1536047	97	1	399333120	2010-06-21
ddo.000001.Atlas.Ideal.DBRelease.v090001	82	82	7	470554	470554	470554	470554	98	1	348095259	2010-06-21
ddo.000001.Atlas.Ideal.DBRelease.v110001	78	78	110	230441	230441	230441	230441	97	1	401298508	2010-06-21
ddo.000001.Atlas.Ideal.DBRelease.v100801	69	69	113	175533	175530	175533	175530	97	1	401285007	2010-06-21
ddo.000001.Atlas.Ideal.DBRelease.v100401	67	67	28	145089	145089	145089	145089	97	1	392037719	2010-06-21
ddo.000001.Atlas.Ideal.DBRelease.v100301	76	76	48	108628	108628	108628	108628	96	1	386660845	2010-06-21
mc09_7TeV.105805.filtered_minbias6.evgen.EVNT.e530_tid144946_00	6	6	1	104040	104040	104040	104040	4	2000	111859161013	2010-06-21
mc09_7TeV.105805.filtered_minbias6.evgen.EVNT.e530_tid144947_00	9	9	1	96426	96426	96426	96426	4	2000	111862745357	2010-06-14
mc09_7TeV.105805.filtered_minbias6.evgen.EVNT.e530_tid144948_00	12	12	1	82475	82475	82475	82475	4	2000	111828609599	2010-06-11
mc09_7TeV.105001.pythia_minbias.evgen.EVNT.e517_tid107743_00	24	24	1	79764	79764	79764	79764	7	2000	70827198081	2010-06-08
data10_7TeV.00153565.physics_MinBias.merge.ESD.r1297_p161_tid143466_00	4	4	26	169910	169842	79089	79058	6	5486	20678992906406	2010-06-21
ddo.000001.Atlas.Ideal.DBRelease.v090103	49	49	2	74926	74926	74926	74926	94	1	350687721	2010-05-31
data10_7TeV.00155697.physics_L1Calo.recon.ESD.f261	5	5	44	171798	169733	60133	59566	8	3518	8601359214389	2010-06-21
data10_7TeV.00155112.physics_L1Calo.merge.ESD.r1299_p161_tid143352_00	5	5	37	126671	126573	54923	54839	7	2917	10149828227066	2010-06-21
ddo.000001.Atlas.Ideal.DBRelease.v060501	6	6	14	47129	47129	47129	47129	108	1	290367871	2010-06-21
ddo.000001.Atlas.Ideal.DBRelease.v10070102	45	45	23	86526	86524	45534	45532	98	55	12143271984	2010-06-21
mc09_7TeV.105001.pythia_minbias.evgen.EVNT.e517_tid125551_00	21	20	2	42892	42891	42892	42891	4	2000	70828562612	2010-06-08
ddo.000001.Atlas.Ideal.DBRelease.v100601	32	32	14	37756	37756	37756	37756	97	1	399034287	2010-06-18
data10_7TeV.00153565.physics_MinBias.merge.AOD.r1297_p161_tid143466_00	12	12	25	213141	207596	37620	37612	21	5486	1167788162982	2010-06-21
mc09_7TeV.105001.pythia_minbias.recon.ESD.e517_s764_s767_r1302_tid1364	1	1	34	275965	272743	36236	33014	3	5000	3989751863375	2010-06-21
data10_7TeV.00155112.physics_L1Calo.recon.ESD.f255	8	8	33	99425	97736	35769	35731	9	3072	9783220355396	2010-06-16
data10_7TeV.00153565.physics_MinBias.recon.ESD.f251	7	7	16	134795	121441	35677	33996	9	11475	20034730996361	2010-06-21
ddo.000001.Atlas.Ideal.DBRelease.v100502	55	55	30	35365	35364	35365	35364	97	1	394404968	2010-06-21
data10_7TeV.00155112.physics_L1Calo.merge.AOD.r1299_p161_tid143352_00	20	17	57	429547	422786	33735	33716	23	2917	756685139434	2010-06-21
mc09_7TeV.105001.pythia_minbias.recon.ESD.e517_s764_s767_r1302_tid1364	1	1	34	251153	246222	33729	28798	1	5000	3990465019429	2010-06-21
data10_7TeV.00155112.physics_L1Calo.recon.ESD.r1299_tid141285_00	1	1	23	104282	99954	33715	33465	1	3072	10146774491151	2010-06-21
ddo.000001.Atlas.Ideal.DBRelease.v100101	46	46	21	33655	33655	33655	33655	97	1	376728135	2010-06-21
data10_7TeV.00155697.physics_MuonswBeam.recon.ESD.f261	7	6	34	98812	92866	33642	28080	8	1821	1509944843191	2010-06-21
data10_7TeV.00155634.physics_L1Calo.recon.ESD.f260	6	6	44	93735	92606	32921	32915	7	1269	2774118899042	2010-06-21
data10_7TeV.00155569.physics_L1Calo.recon.ESD.f260	4	4	48	118203	115444	30630	30605	5	1582	2529422476523	2010-06-21

User behaviour

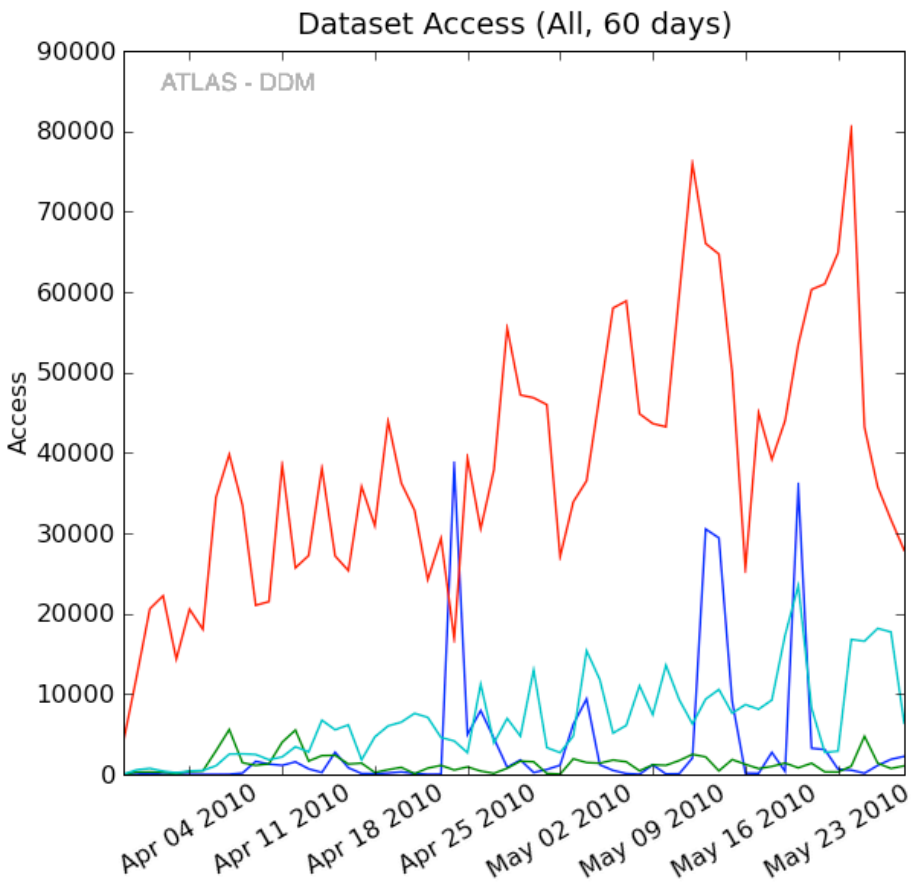
- We cannot control what our users do
  - and in some way, that's both good and bad

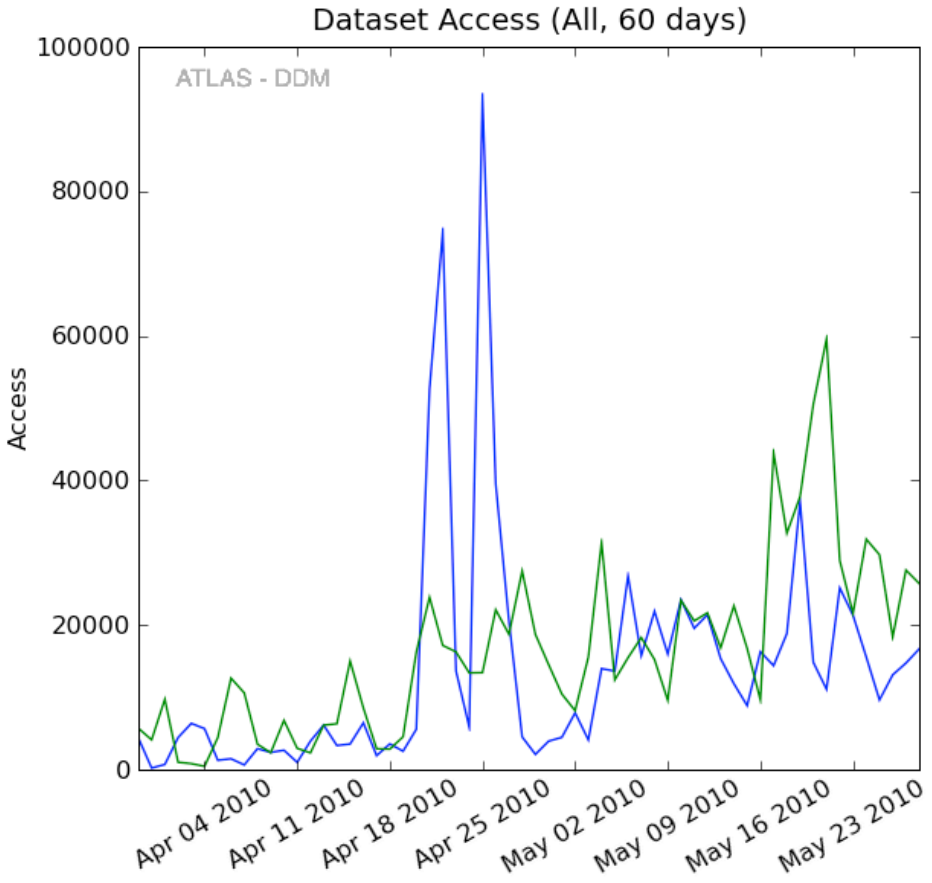
- remember the Computing Model
  1. data is moved by DQ2 automatically
  2. user submits analysis job
  3. scheduler selects a Tier-2 with required data for the job
  4. job runs at Tier-2 ("*send the job to the data*")



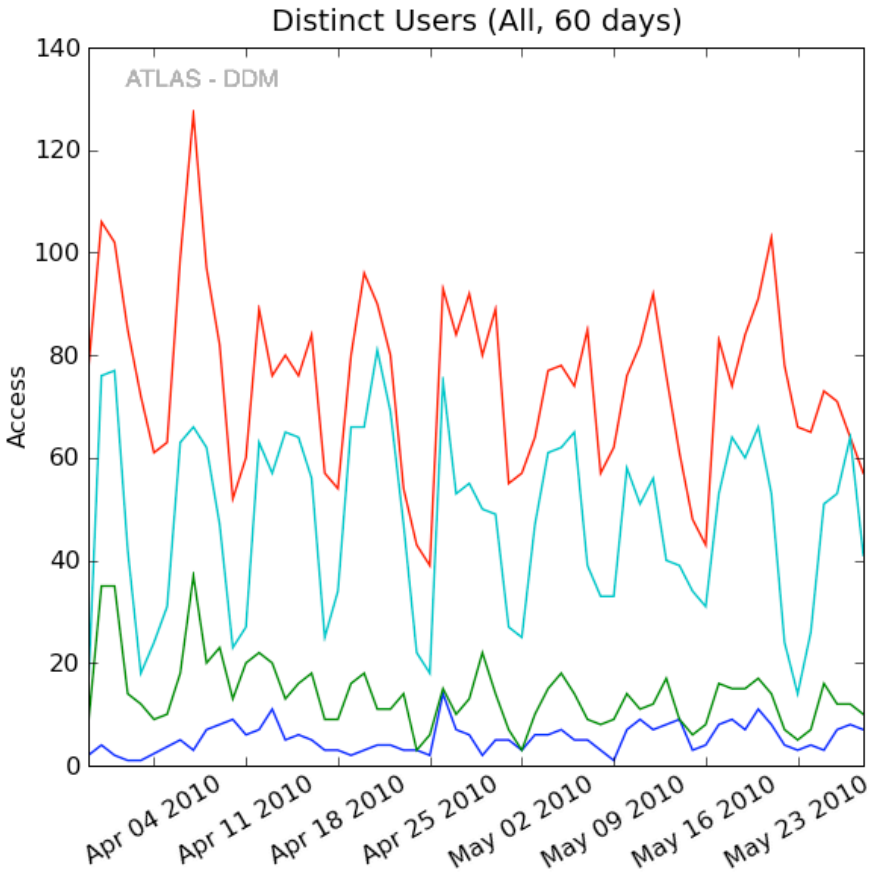


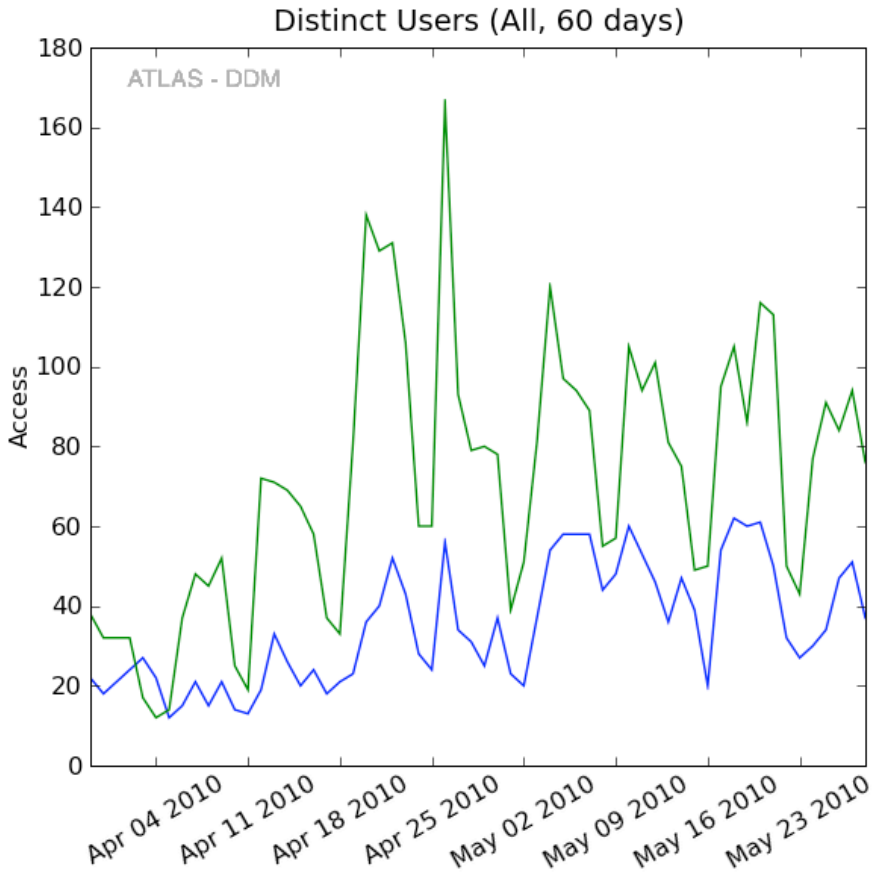
Dataset access (real vs. sim data)



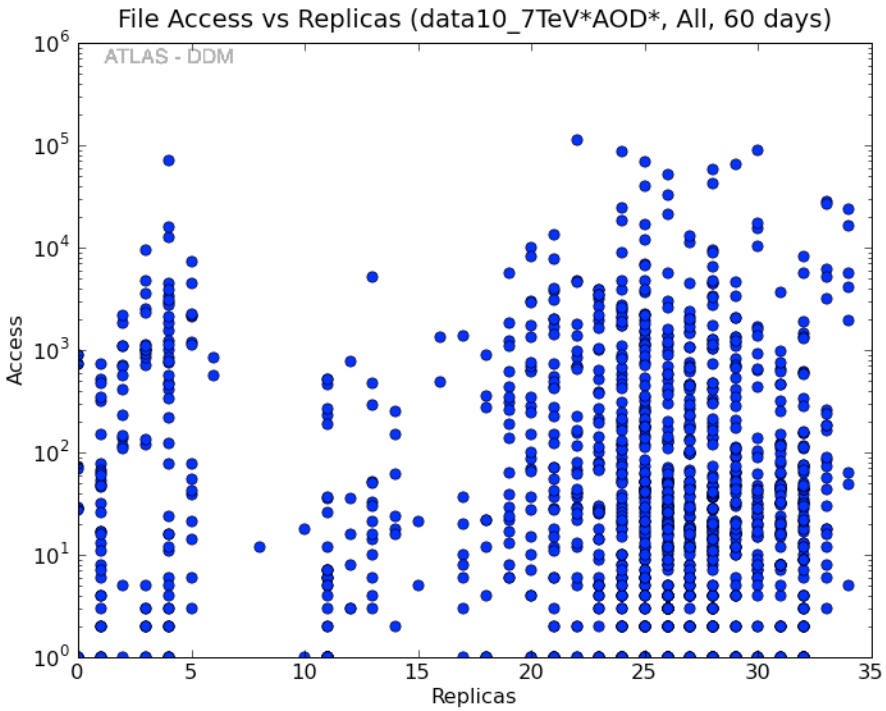


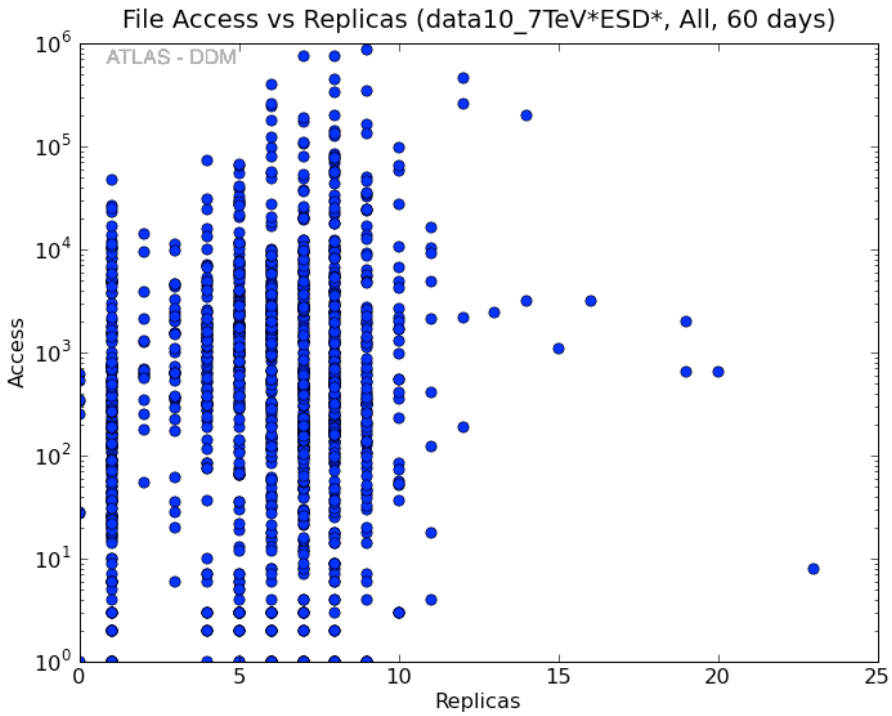
Distinct users (real vs. sim data)



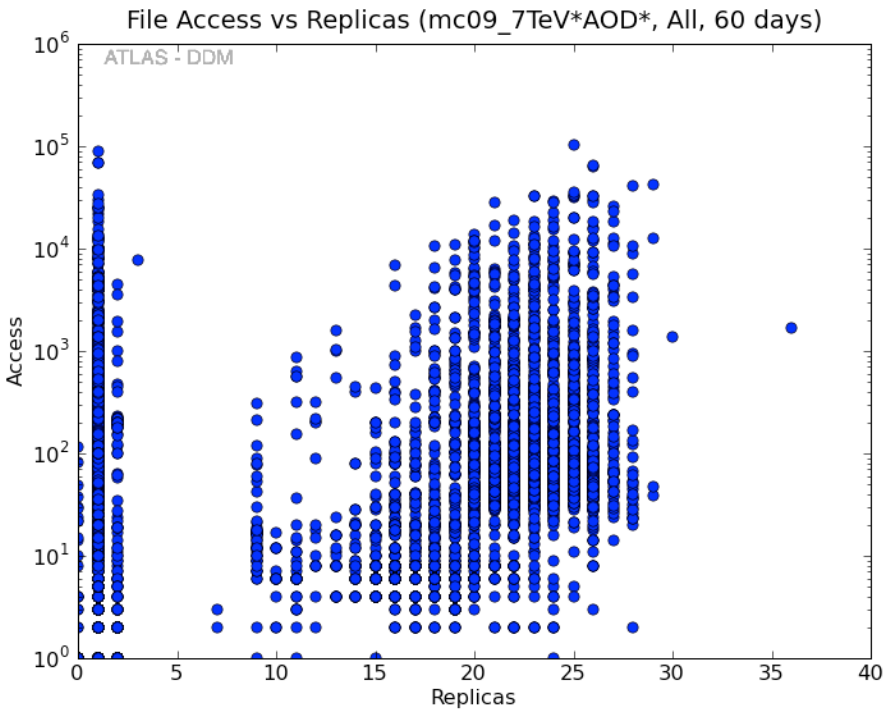


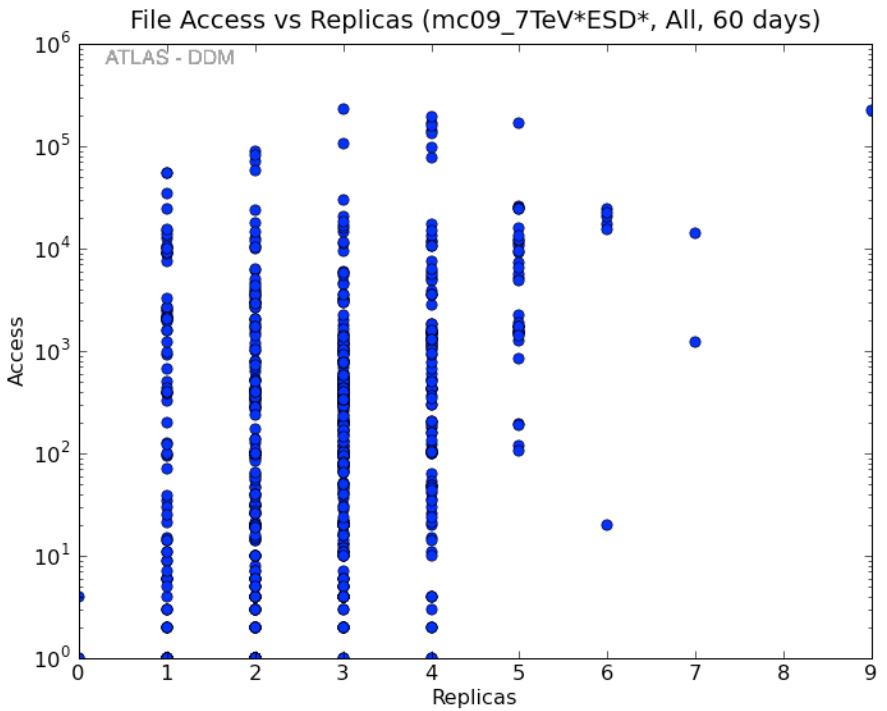
Access vs Replicas (real data)



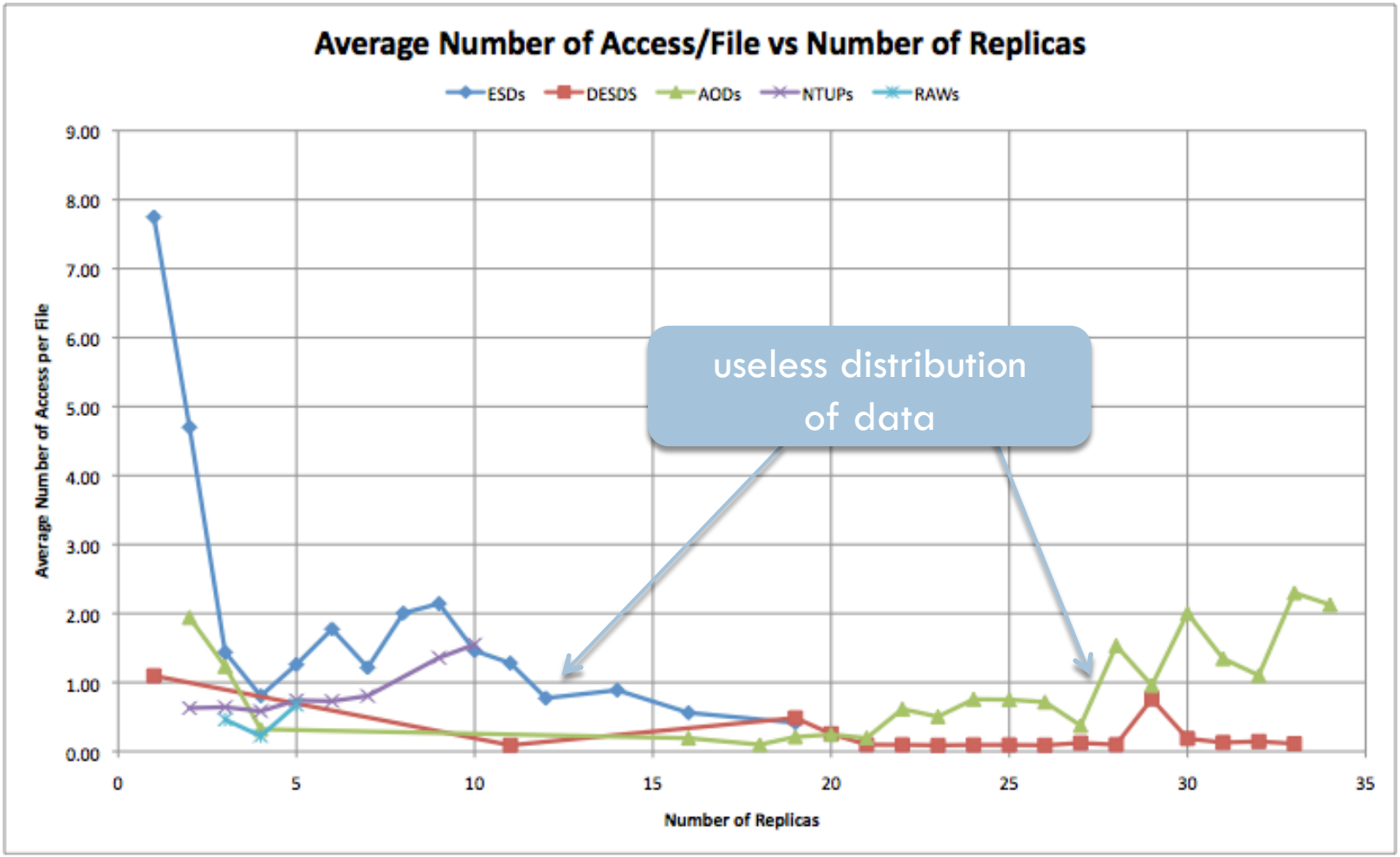


Access vs Replicas (sim data)

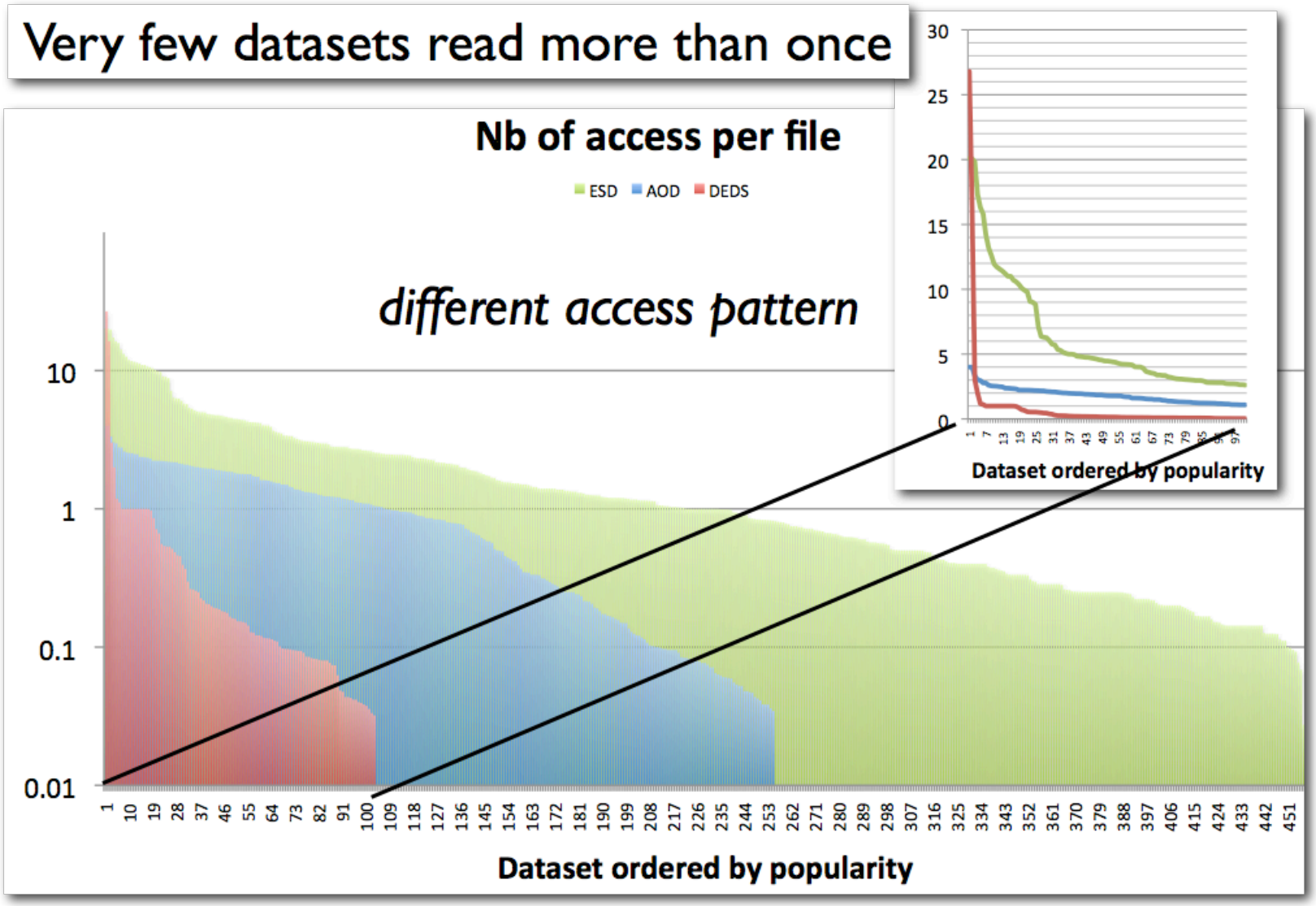




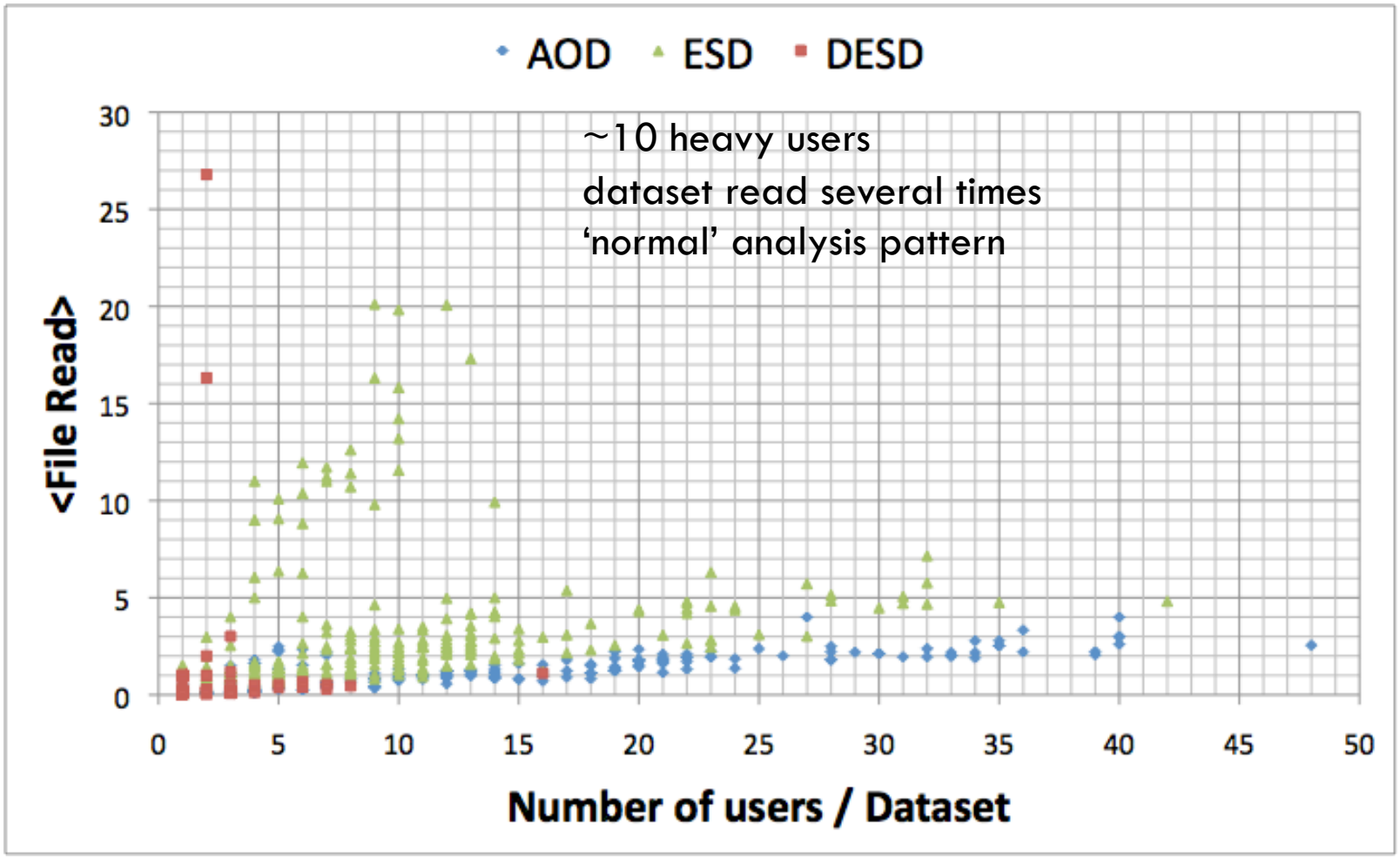
Replication factor



Dataset access



Dataset access



Summary

- We are just starting to understand data management at this scale
  - "it works"™
    - still, but we constantly grow
    - therefore we need to understand exactly what's going on in almost real-time
    - rebuilding full system in simulation (sim currently in validation)
  - many different monitoring infrastructures, yet none call tell us what the users really do
    - too much heterogeneous data
    - consolidation of data too time-consuming
    - we had to build this tracer directly in all our software
    - make it easy for 3rd party apps to hook in via simple HTTP calls
    - provide realistic workload for the simulation studies

- So, if you want to remember just one thing from this work:

  "Monitoring of user behaviour needs to be done explicitly, and in a very simple way."

Scalable stochastic tracing of distributed data management events

Mario Lassnig

mario.lassnig@cern.ch

- M. Lassnig, T. Fahringer, V. Garonne, A. Molfetas, M. Branco, ``Identification, modelling and prediction of non-periodic bursts in workloads'', Proc. *10th IEEE Int. Conference on Cluster, Cloud and Grid Computing, IEEE, 2010*
- M. Lassnig, T. Fahringer, V. Garonne, A. Molfetas, M. Branco, ``Stream monitoring in large-scale distributed concealed environments'', Proc. *5th IEEE Int. Conference on e-Science, IEEE, 2009*
- M. Branco, E. Zaluska, D. De Roure, M. Lassnig, V. Garonne, ``Managing very large distributed datasets on a Data Grid'', *Concurrency and Computation: Practice & Experience, in press (early view available), Wiley, 2009*
- M. Branco, E. Zaluska, D. De Roure, P. Salgado, V. Garonne, M. Lassnig, R. Rocha, ``Managing Very-Large Distributed Datasets'', *Lecture Notes in Computer Science, Vol. 5331(2008), Springer, 2008*

EGI Technical Forum, 14-09-10, Amsterdam