Grand Challenge Status

D. Zimmerman, LBNL

RCF Meeting, 1 April 1999

Outline

- MDC2 status & results
- Plans for scalability tests
- Integration with experiments

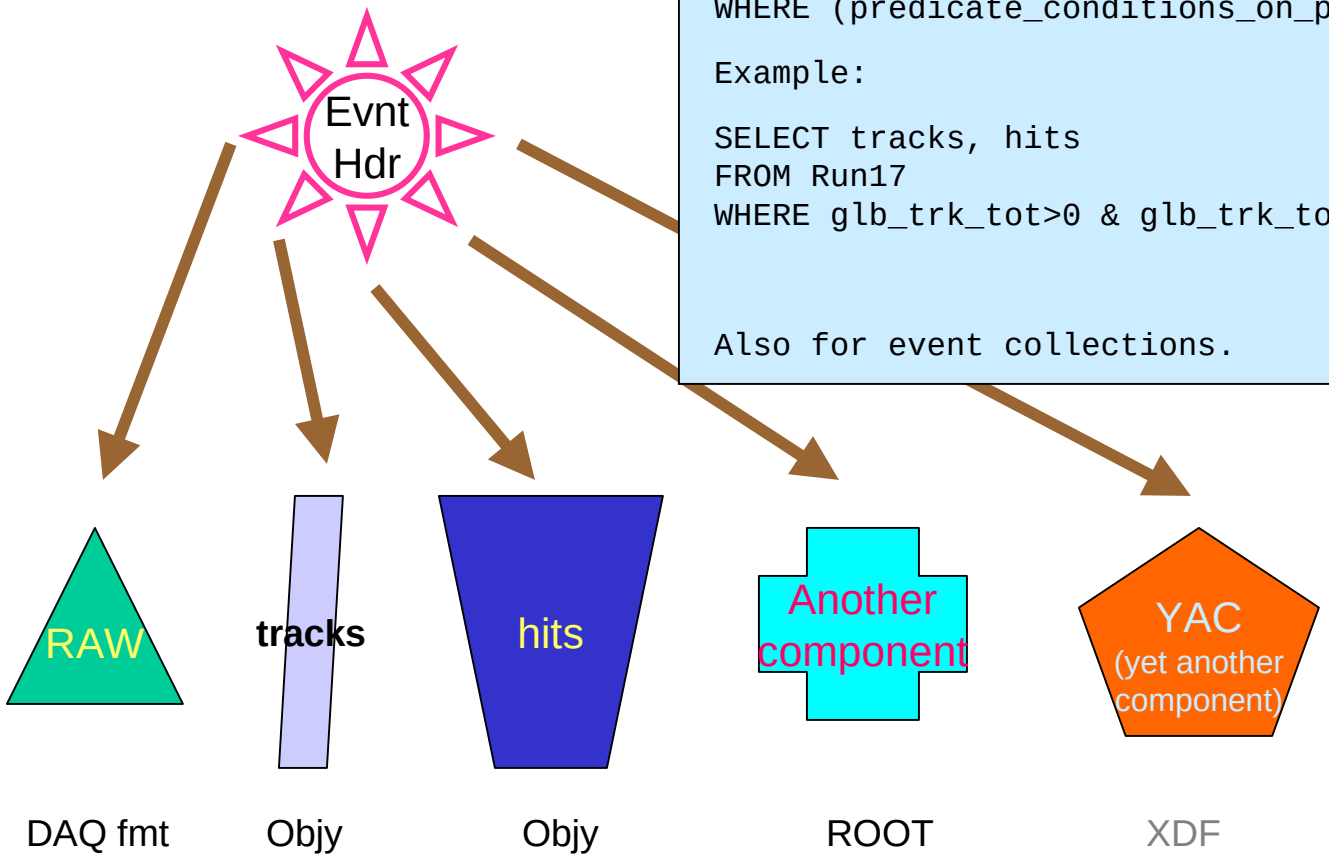
Key change for MDC2: Multiple named components

```
SELECT (component_name, component_name, ...)
FROM dataset_name
WHERE (predicate_conditions_on_properties)

Example:

SELECT tracks, hits
FROM Run17
WHERE glb_trk_tot>0 & glb_trk_tot<10 & n_vert_total<3


Also for event collections.
```



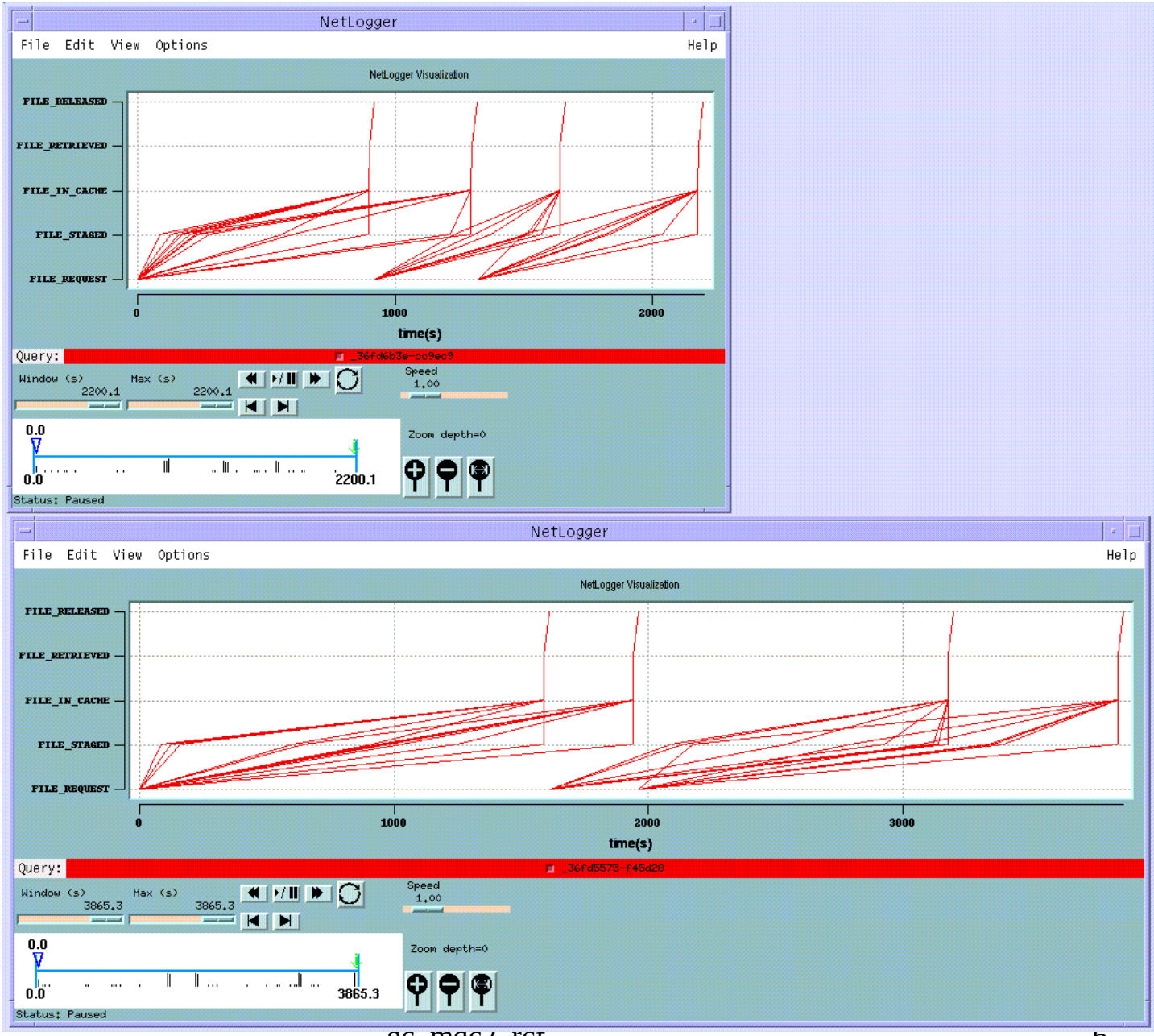
MDC2 status & results

- The MDC2 plan: utilize multi-component event data in Objectivity or with Objectivity event headers.
- We had no experiment data like this so we built a sample federation with
  - 20k events, each event with 8 components
  - 330 GB
  - 100 Objy files, 400 ROOT files

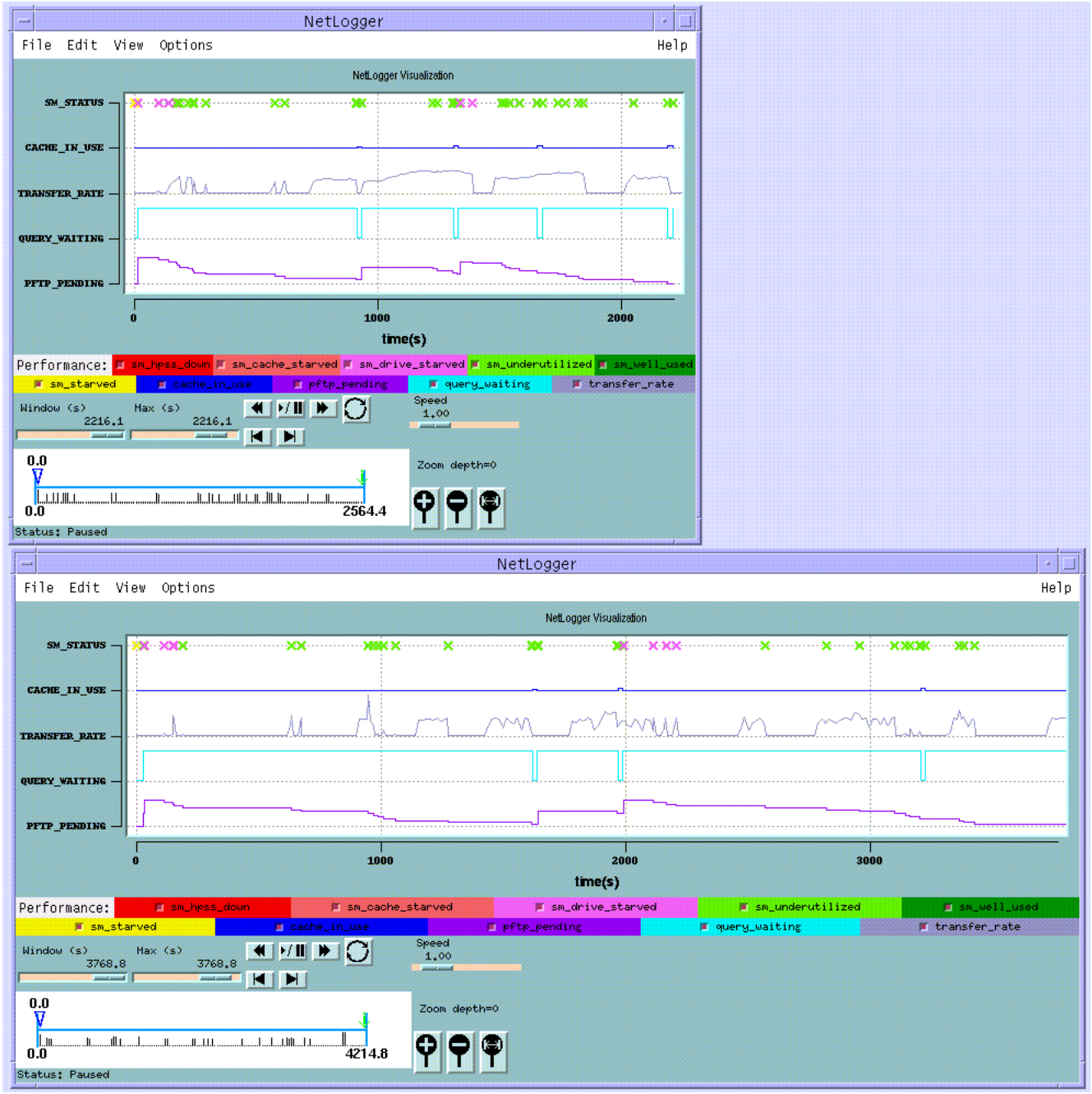
What we did

- Run tests with various numbers and types of queries and measured effects of policy parameters and number of tape drives.
- Ran simple client code on rmds03 and rsun00.
- Made a simple Objy-aware client code that ran on linux but we had lock server problems and could not run it with our federation.

2 drive vs. 1 drive file tracking



2 drive
vs.
1 drive
performance



Partial summary of runs (30 runs total)

This is a summary of the tests so far. Note: some problems were noticed till run 7, none caused a crash, except run 7. All fixed and ran OK before Run 8.

RUN	Query, Code used	purpose	#bundles X comp = files	policy on	Cache Size (GB)	Proc Time (sec)	Total Run time (sec)	comments
1	SII,Q1	Robustness, small files	32 x 4 = 128	yes	5	0.1	3167	Ran mostly ok Problem: one file purge too early
2	SII,Q2	Robustness, Large files	15 x 4 = 60	yes	5	0.1	12086	Interrupted by HPSS down twice (40 min + 10 min)
3	SII,Q3	Overlap (same query twice)	12 x 3 = 36	Yes	6	0.1	3316	15 min delay between queries. Policy on. Problem: staged file release too early
4	SII,Q3	Overlap (same query twice)	12 x 3 = 36	NO	6	0.1	5925	15 min delay between queries. Policy off. Problem: skipping queries incorrectly
5	SII,Q4	3 queries, 50% overlap bundles	15 x 3 = 45	Yes	12	0.1	6208	Problem: no purging occurred, bug in cache size setup
5a	SII,Q4	3 queries, 50% overlap bundles	15 x 3 = 45	Yes	12	0.1	6615	Problem: bundle request made prematurely
6	SII,Q4	3 queries, 50% overlap bundles	15 x 3 = 45	NO	12	0.1	6411	Problem: 3 extra files requested at end of run and not pushed to SII.
7	SII,Q5	Robustness, overnight	102 x 6 =612	Yes	12	0.1	4822	20 queries, 10 min apart, 10 cycles Problem: QM crashed at end of 3rd cycle
8	SII, Q7	1 drive	4 x 7 = 28	Yes	80	0.1	3865	Bundles selected with components on 5 tape; 7 components
9	SII, Q7	2 drive	4 x 7 = 28	Yes	80	0.1	2200	Same as above
10	SII, Q7	3 drive	4 x 7 = 28	Yes	80	0.1	20	Same as above, but forgot to empty cache
11	SII, Q7	3 drive	4 x 7 = 28	Yes	80	0.1	1993	Noticed: low pending PFTPs, not enough files requested betwn bundle processing.
12	SII, Q7	4 drive	4 x 7 = 28	Yes	80	0.1	1699	Same as above
13	SII, Q7	5 drive	4 x 7 = 28	Yes	80	0.1	1432	Same as above

Plans for scalability tests

- Some issues related to scalability (> 10^7 events) are known to not be resolved yet in the current software implementation.
  - Event list as query
  - Stress tests of components
- We will discuss plans in meeting at ANL, April 27,28.
- Expect to carry out scalability tests in June.

Some questions to consider

- For continued development & testing at RCF should there be a “development” class of service to use?
- Is June a good time frame for integration with experiments’ real event model?
  - We plan to make general non-Objy specific interface to Event Identifier and File Catalog.