CDF Computing Project Status Report: CDF DH (non-SAM non-ops)

-or-

The CDF Run II Data Handling Road Map Part 1 Revisited
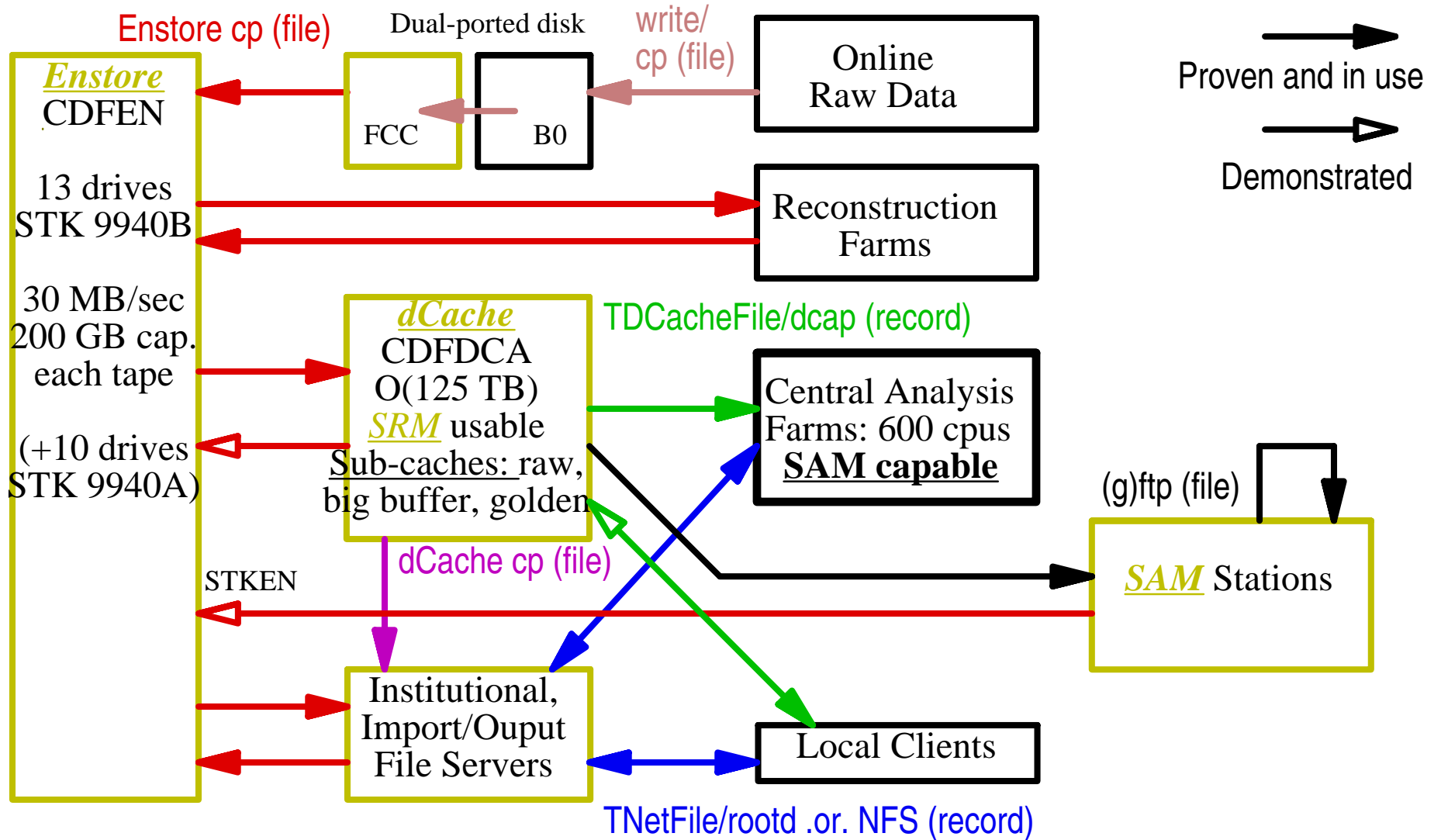
Robert D. Kennedy
FNAL Computing Division
for the CDF Data Handling Group
10 Dec 2003
(Roadmap from R2D2 Meeting - 06 May 2003)

⇒ **SAM is integral part of CDF DH, but talk scope pre-defined**

⇒ **Review of the CDF DH Roadmap**

⇒ **Some new components to CDF DH project**

CDF DH Today (legacy systems now retired)

Enstore cp (file)

Dual-ported disk

write/
cp (file)

Enstore
CDFEN

FCC   B0

Online
Raw Data

13 drives
STK 9940B

Reconstruction
Farms

30 MB/sec
200 GB cap.
each tape

(+10 drives
STK 9940A)

dCache
CDFDCA
O(125 TB)
SRM usable
Sub-caches: raw,
big buffer, golden

TDCacheFile/dcap (record)

Central Analysis
Farms: 600 cpus
SAM capable

(g)ftp (file)

SAM Stations

STKEN

dCache cp (file)

Institutional,
Import/Ouput
File Servers

Local Clients

TNetFile/rootd .or. NFS (record)

Proven and in use

Demonstrated

Underlying *Meta-data*: PNFS (Enstore/dCache), CDF Data File Catalog, SAM meta-data
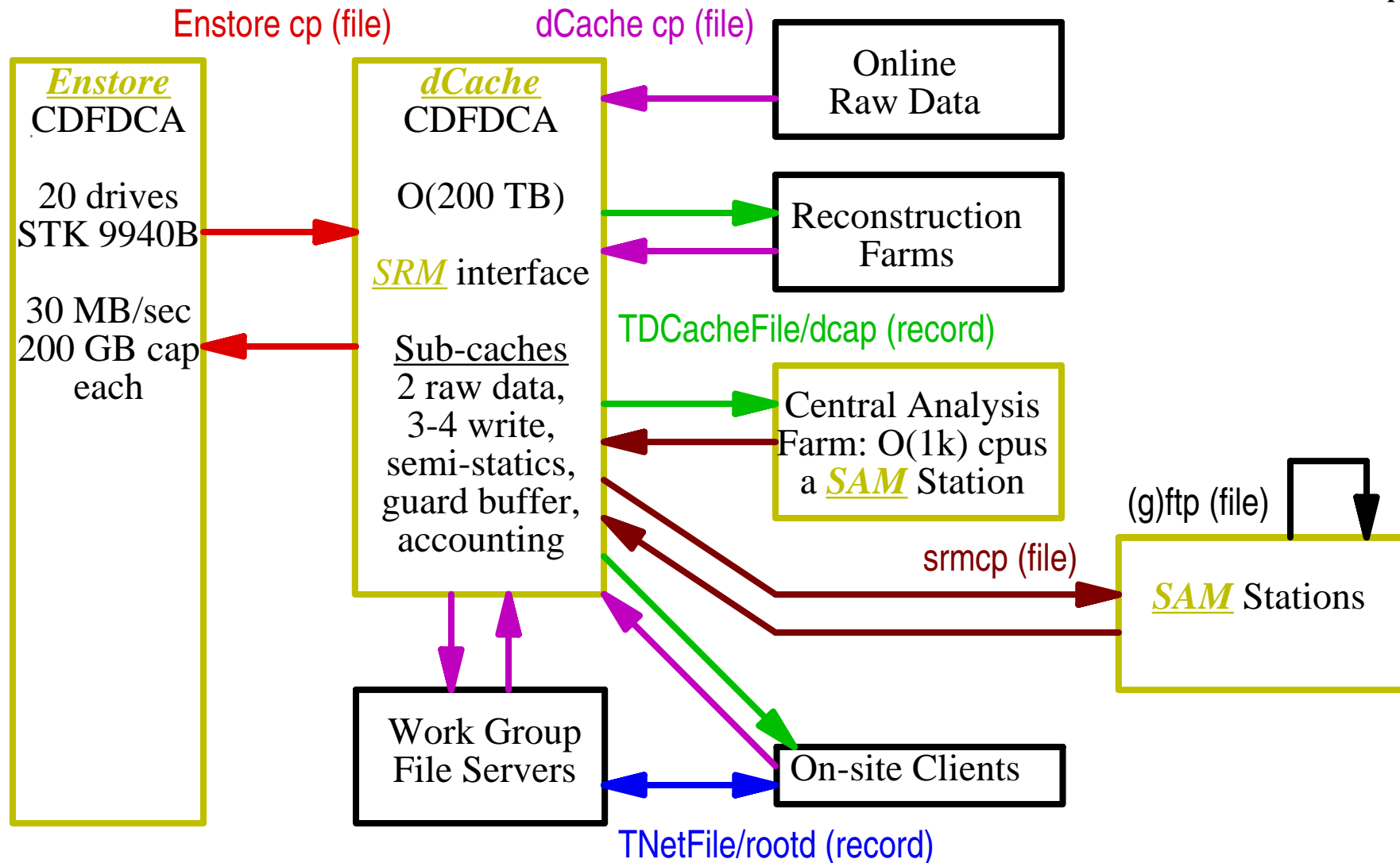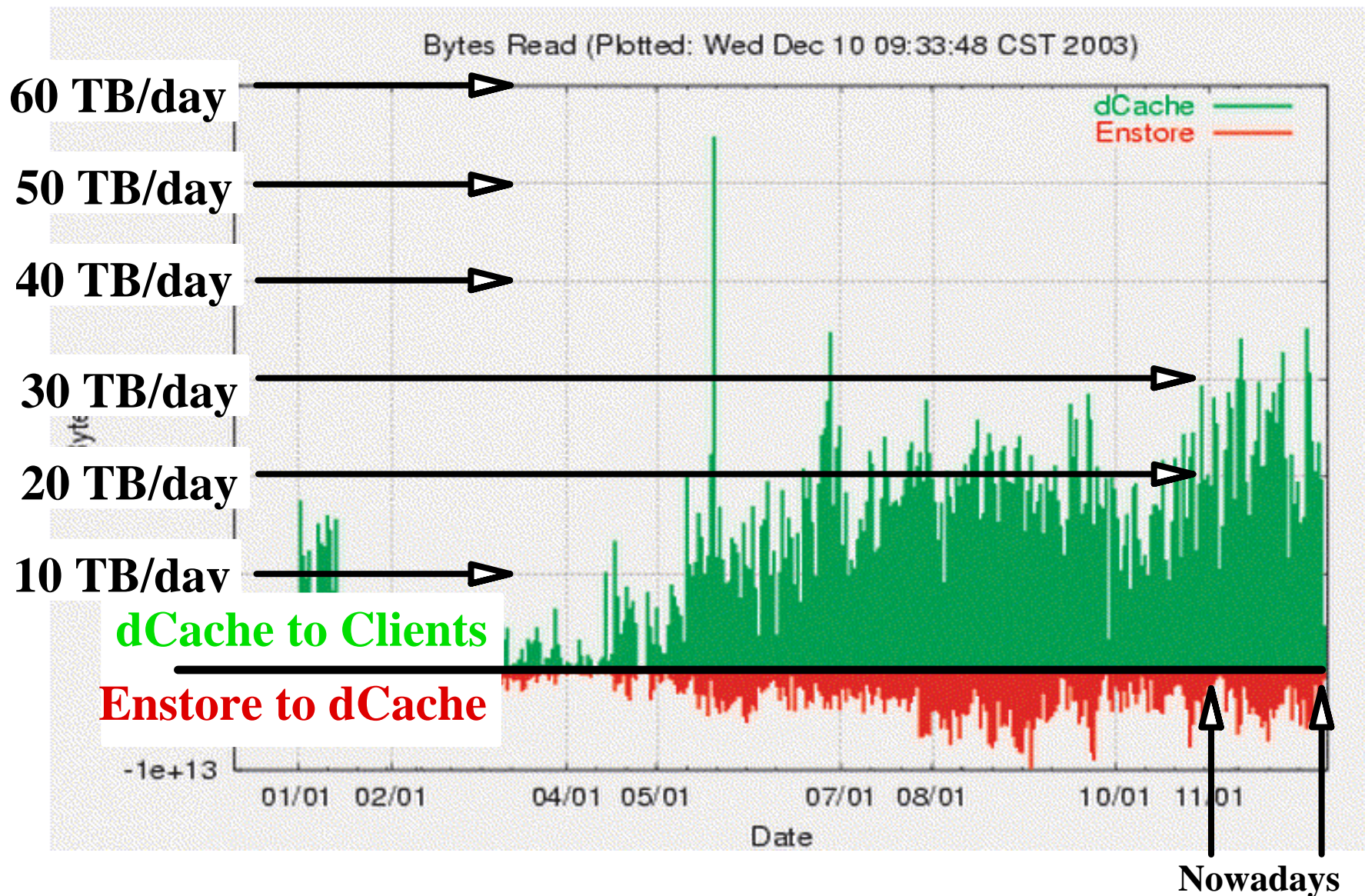
Overview: Baseline Goal FY2004

⇒ **CDF Enstore fully migrated to use of 9940B tape drives**
   About 3/4 done by data volume - work by enstore-admins

⇒ **CDF dCache fully implemented, 100 TB+, with sub-caches**
   Done: Krb5 auth required for reads and writes. Stable ops.

⇒ **All Enstore access goes through dCache (read .and. write)**
   Write pools deployed, ready for bulk testing.

⇒ **Tapeless data paths for produced (and raw?) datasets**
   Ready for use when write pools validated at load.

⇒ **Retire "legacy" systems, once baseline proves reliable**
   Done. Static file servers absorbed. DIM/Kahuna turned off.

⇒ **Fully adapt to SAM schema; drop DFC schema, keep API**
   In progress. (but this is "non-SAM" talk).

⇒ **CAF adapted to become a SAM station w/direct dCache access**
   In open beta testing. (but this is "non-SAM" talk).

⇒ **Simplified "entry-level" SAM input for users (carrot, not stick)**
   Work on SAM optimization of file delivery w/dCache ESM.

⇒ **Robust operations, doc'd procedures, user experience issues**
   Investigating rare errors, corrupt files going into Enstore, ...

⇒ **Groundwork: CAF output catenation, unified data processing**
   "scratch dCache" (FKW), CAF+SAM+Grid, ...

CDF DH Baseline Goal FY2004

Enstore cp (file)

dCache cp (file)

Enstore
CDFDCA

20 drives
STK 9940B

30 MB/sec
200 GB cap
each

dCache
CDFDCA

O(200 TB)

SRM interface

Sub-caches
2 raw data,
3-4 write,
semi-statics,
guard buffer,
accounting

Online
Raw Data

Reconstruction
Farms

TDCacheFile/dcap (record)

Central Analysis
Farm: O(1k) cpus
a *SAM* Station

(g)ftp (file)

srmcp (file)

SAM Stations

Work Group
File Servers

On-site Clients

TNetFile/rootd (record)

Underlying *Meta-data*: PNFS (Enstore/dCache) and SAM meta-data

dCache: Bytes Read/day in 2003

Bytes Read (Plotted: Wed Dec 10 09:33:48 CST 2003)

dCache ⎯⎯
Enstore ⎯⎯

60 TB/day
50 TB/day
40 TB/day
30 TB/day
20 TB/day
10 TB/day

dCache to Clients
Enstore to dCache

-1e+13

01/01 02/01   04/01 05/01   07/01   08/01   10/01 11/01

Date

Nowadays

FY2004: New Components

⇒ **Simplify treatment of NOACCESS files in Enstore**
    Now involves e-mail comm, manual meta-data changes
⇒ **dCache Features: Auto-stage, open timeout**
    Simplify tapeless path mgmt. DB disconnects during tape pause.
⇒ **SRM interface to dCache**
    Revive component work to simplify SAM dCache ESM.
⇒ **Scratch dCache, CAF+SAM+Grid, global DH deployment**
    SAM Install Fest Jan 2004 (UF). Work in SAM context to refine.
⇒ **Test dCache as CAF and data volume scales increase.**
    Look for weak or slow points to maintain current robustness.
⇒ **Improved error reporting, monitoring for user point-of-view**
    Help users understand state of system, requests w/o experts
⇒ **Consult on old and new DH applications**
    Example: ntuple-based datasets need "taming" and file catalog.
⇒ **Integrate operations with CDF Offline, CDF SAM support.**
    Use time now spent triaging reports to work on DH/SAM dev.
⇒ **Integrate non-SAM development with SAM development**
    In progress: CDF DH is SAM (w/dCache, Enstore thrown in)