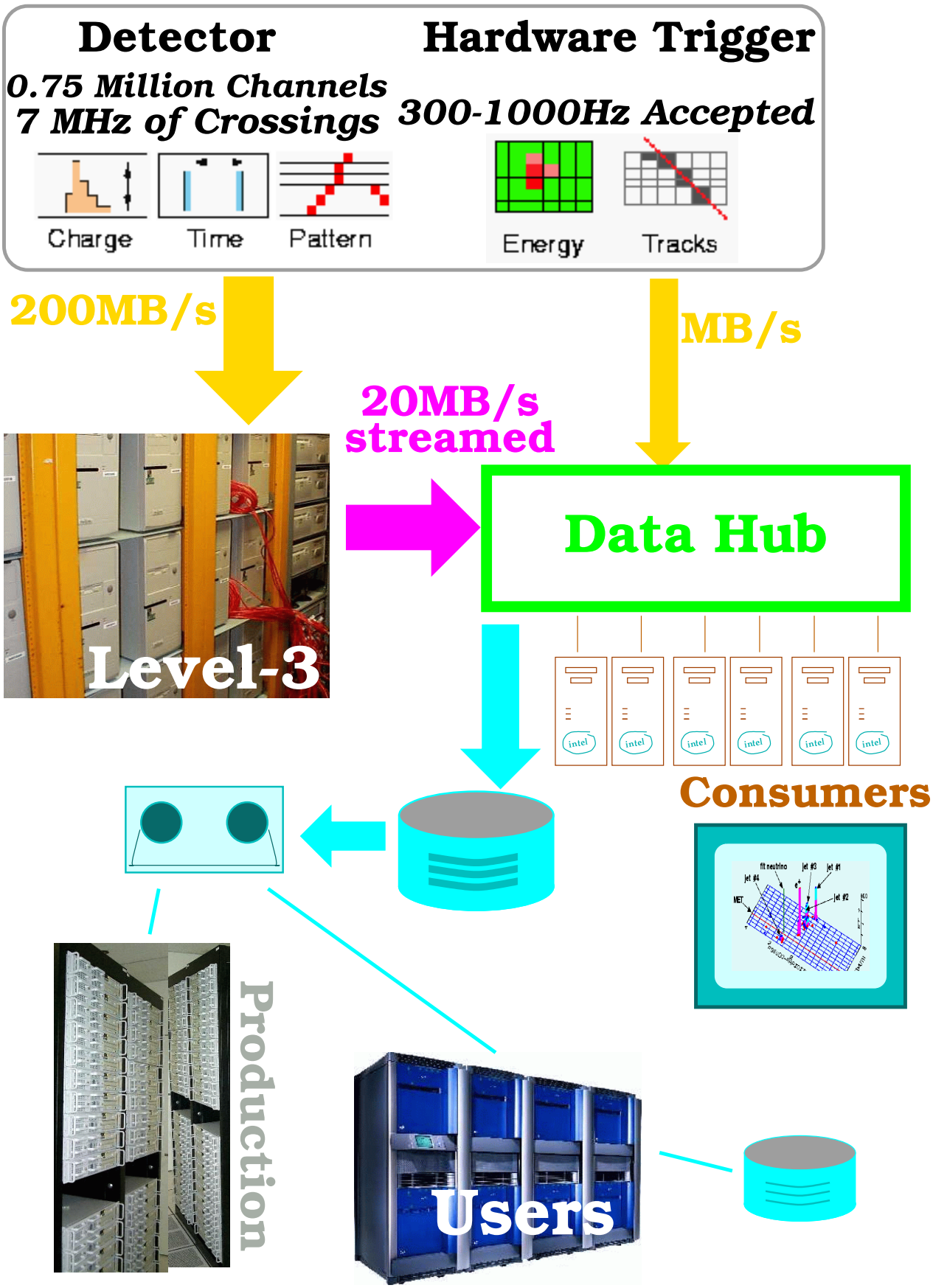
Data Handling

Robert Harris
Kevin McFarland
CDF Offline Operations
September 25, 2001

---

1. Scope of the system
   - Analysis Needs, *the endgame*
   - Commissioning Needs, *the nowgame*
2. Operational Issues
   - Deployment and integration
   - Commissioning
   - Operations
   - Transition to the *endgame*


Detector
0.75 Million Channels
7 MHz of Crossings
Charge
Time
Pattern
Hardware Trigger
300-1000Hz Accepted
Energy
Tracks
200MB/s
MB/s
20MB/s
streamed
Level-3
Data Hub
Consumers
Production
Users

Scope of DH project

- Raw Data Volume
  - √ CDF data logging: 75 Hz peak of 250 kB raw events
    - ⋆ Canonical steady state efficiency factor of π^{-1}
  - ↪ Raw data volume is approximately 20 TB/month in steady operation (*endgame*)
    - √ Approximately 5–10 TB/month *in nowgame*
    - √ Half (or less) is destined for analysis
  - ↪ Will write into 8 or 9 datastreams
    - √ Can live with fewer (*nowgame*)
  - √ Require fast access to small subsets of data (*game-invariant*)
    - √ "Consumer" online monitoring (5 MB/s peak) (DAQ group)
    - √ "LOOK" area (files) (1 MB/s DC)
      (pioneered by DH, now operated by DAQ group)

(√ indicates empirically verified statement)

Scope of DH project (cont'd)

- Production
  - ↪ Output format is PADs, 60–100 kB, *in endgame*
    - ⋆ Implies need for substantial re-production (raw data not present)
  - √ DH Project handles I/O for Production farms
    - ⋆ Eventual I/O capability is 2–3 times raw data DC rate (not always concurrent with raw data taking)
    - ⋆ Want capability to partition farms into 3–5 subfarms (includes one for MC)
    - √ In commissioning, we don't reprocess
    - √ BUT write RAW+PADs+... (*nowgame*)
    - √ Require final I/O capability early
  - √ Need to write many datasets (Level-3 selected) out of production
    - √ Highly granular datasets reduce analysis I/O load
    - √ Need is *game invariant* since we are running at peak logging rate

(√ indicates empirically verified statement)

Scope of DH project (cont'd)

- User Analysis
  - √ These requirements are the most nebulous
  - √ Users can fill whatever bandwidth is provided
    - √ Coordinated data access is necessary
  - ↪ DAMNAG goal of 50 MB/s I/O rate (DC) is reasonable
    - ⋆ Largest Run II datasets from production are 10 TB (*endgame*)
      - · 5% of data, PAD format
      - · Run through in 3 weeks with 10% of I/O resources
  - √ Need large, flexible pools of disks to cache data
  - √ Require robust I/O with error checking and recovery
    - √ Must at least catch data corruption!
  - √ Require simple user definition of secondary, tertiary datasets
    - √ With ability to archive to tape
    - ⋆ Version control
  - ↪ Require export facility to remote sites
    - √ Significant computing available offsite
    - ⋆ Requires integration of network/tape copied files with exportable DH infrastructure
    - √ Key during commissioning; can assist partially completed CAF

(√ indicates empirically verified statement)

The Design of the System

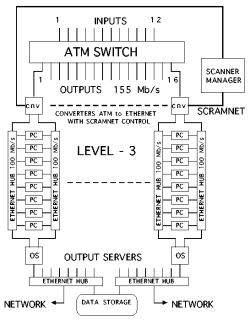
We believe the Data Handling system as designed is capable of meeting the specifications for Run II.

- Tape I/O, FCC logger, Robot, Disk are all capable of meeting bandwidth targets
- Central analysis systems are a major uncertainty
  - ↪ CAF committee may come/is coming to the conclusion that DAMNAG specifications are inadequate
    - ⋆ Software is expected to be slower than projected (large uncertainty)
    - ⋆ There is a desire to expand capabilities to the high end of what was expected (driven by hadronic B physics, low p_t dileptons)
  - ↪ Solution may impact the DH system, or it may not
  - ↪ *We will identify, and attempt to minimize, the impact of any needed changes on the design of the DH system*
- Run IIb will be another story if data volume is $10\times$ larger

n.b., Design $\notin$ {Implementation, Integration, Operation}

Requirements for Deployment

- Data Handling must ensure resources for users during installation, commissioning, steady operation of CDF
  - ↪ Premise: data handling is a "service" project
    data handling requirements driven by customers
  - ↪ contrast with SVXII
  - ↪ Therefore, commissioning of service systems starts well before TeVatron delivers luminosity
  - ↪ DH required to provide full logging capabilities, significant fraction of access capabilities "up front"



- Example: Level-3 PC farm
  - ↪ ATM, I/O nodes commissioned 1997-1999
  - ↪ Processors not fully installed yet
  - ↪ Met internal/external needs
- After the above is satisfied, delay purchases to take care of technological advances. Stage if possible.
  - Example: CDF II Tape Technology
    - ↪ AIT1 for commissioning.
    - ↪ Delayed deployment of full tape capacity was crucial for allowing 2nd generation technology (AIT2)

Requirements for Integration

- Data Handling system is naturally the centerpiece of integration testing
  - ↪ Data logging is *de rigeur* to verify "event building" functionality of DAQ
  - ↪ Full functionality should be tested:
    - ⋆ Data splitting: streams, datasets
    - ⋆ Sustained full bandwidth
      - · *or* **rigorous** *scaling demonstration...*
    - ⋆ Multi-user access to high volumes of data
    - ⋆ Plurality of interfaces to test
  - ↪ Incomplete testing of DH in Mock Data Challenges was a missed opportunity
- Integration means adaptation
  - ↪ Examples abound in CDF hardware, DAQ, software, where initial integration of sub-systems involves modification of designed interfaces
  - ↪ Test full integration as one would vote in Chicago
    - ⋆ *Integrate early; integrate often*
- "Plethora" of interfaces implies Data Handling should be among the *earliest* systems to be available for integration

Special Needs During Commissioning: *Commissioning CDF*

- Commissioning requires full capabilities of system
  - ↪ E.g., trigger problems may not appear except at full rate
  - ↪ E.g., readout glitches may be rare events
  - ↪ Configuration changes may be rapid
    - ⋆ Advance verification of configurability required
- Commissioning requires operation of system in modes that are not the *endgame*
  - ↪ E.g., event synchronization testing requires quick access to consecutive events
    - ⋆ Providing single stream of data to user-accessible disk is suddenly highest priority
    - ⋆ Production, even data logging become secondary
  - ↪ Systems that allocate bandwidth must be configurable
    - ⋆ In *nowgame* to be flexible
    - ⋆ In *endgame* to tune DH system

Special Needs During Commissioning: *Commissioning Data Handling*

- Commissioning data handling itself may require only "vertical slices" and scaling arguments
  - ↪ Many subsystems can and have been successfully tested this way
  - ↪ Tape, disk, I/O nodes
- *However*, this doesn't test everything
  - ↪ Interfaces
    - ⋆ Are all interfaces fully functional?
    - ⋆ Do related systems receive needed service in real operations?
  - ↪ User interaction
    - ⋆ Are user tools in place?
    - ⋆ Can users *use* them?
  - ↪ Performance on CAF hardware/software
    - ⋆ Do other parts of system adversely affect operations?
    - ⋆ Difficult to foresee all problems
  - ↪ Failure cases
    - ⋆ How do users see hardware failures?
    - ⋆ Misuse/abuse by user interfaces
    - ⋆ Missing handshakes in high bandwidth systems

⇐ Can't commission without full integration

Operating DH

- Operation of DH requires significant resources
  - ↪ New experts must be trained and brought into system after development phase
  - ↪ Shift crews are desirable
- Visible, comprehensive monitoring is critical
  - ↪ Allows diagnosis by shift crews, users
  - ↪ Key ingredient for resource evaluation for collaboration
- Operations, inevitably, feeds back into development
  - ↪ Identifies missing functionality, flexibility
  - ↪ Particularly true while commissioning interfaces
  - ↪ Fresh perspectives from new experts
- Flexibility of system should be exercised and taxed by operations
  - ↪ Outside-driven configuration as needed
  - ↪ Operators must be able to tune system parameters

Operating While Installing

- Simultaneous operation and development is a requirement for DH
- Each require large resources
  - ↪ Significant influx of personnel as operations begin
- Operations, the real-time element, must have instantaneous priority
- System must be "partitionable"
  - ↪ Isolated sub-system development can only go so far
  - ↪ Cannot jeopardize operations (*nowgame*) during expansion to *endgame*
- Comprehensive staging and purchasing plan necessary
  - ↪ Resource starvation leads to crises in installation
  - ↪ Installation in "crisis mode" risks operational glitches
    - ⋆ Cannot drive expansion with view of *nowgame* instead of *endgame*

Conclusions

1. DH requirements make this among the most challenging of CDF II projects
   - Requires a large, effective team for design, implementation, operations
2. Stakes for collaboration are very high
   - Collaboration must assume ownership and responsibility for project
3. DH is more than hardware and bandwidth
   interfaces are the key
   - Integrate early; integrate often
4. DH system will evolve with new hardware, usage patterns, etc.
   - Operation and development must be closely integrated
   - "Modular", flexible system necessary