Introduction to PHENIX offline parallel session

- agenda
- requirements
- plan
- manpower & cost-to-complete at end

Requirements

- Calibration
- Event reconstruction
- Event offline selection and physics analysis
- Core Offline
  - Infrastructure for software development and data production
  - Databases - geometry, calibration, detector element, selection “tag”
  - Event display
- Simulation
  - Detector geometry and response
  - Event generators
  - Magnetic field simulation
  - Production of events for design & background study
  - Tools for collaboration to produce events

Requirements (from S. Sorensen)

- Raw Data (steady state, not Day-1)
  - event size ~ 300-400 kBytes
  - recording rate = 20 MBytes/sec
  - data rate ~ 50 events/sec
  - 500-900M events/year & 200-300 TB/year

- DST (reconstruction output)
  - ~ 20 kB/event (20 TB/year) physics info (?)
  - keep intermediate information for some fraction of events
    - 800kBytes/event
    - 80 TB/year if keep for 10% of events
  - copy raw data to DST (ease of access) for some fraction of events
    - 0.1 - 1% would add 0.2-2 TB/year
    - (NB: accessing 0.1% would require mounting EVERY tape)
  - so ~ 100 TB/year of DST information

Reconstruction Requirements

- if 50 events/sec and 100 processors
  have 2 seconds/event

- require 4056 SpecInt95 at RCF to reconstruct
  MAIN consumer of RCF CPU cycles
  need at least 128 MB memory per processor

- resource limited to 1-1.5 passes through the whole data set
  primarily tapemount limitation

Requirements, ctd.

- Calibration
  - PHENIX has ~300K channels
    - if store 20 bytes of calib data
    - 4 time per hour (or 4 passes)
      - ~100 GB/year of calibration data
- Tag database (for data mining)
  - event characteristics
    - filled at reconstruction
    - includes trigger information
  - resident on disk
  - ~100 GB/year (100-200 bytes/event)
- Export
  - DST + Tag + Calibrations
    - 20.2 TB/year w/o intermediate info
      - sample 10% -> 80 TB/year
      - sample 1% -> 8 TB/year
  - must develop infrastructure to export
  - 30-100 TB/year

Data mining and mini/micro DST

- typically require significant fraction of events
  - RHIC physics is not a rare event search
- select (usually) subset of event data
- mining done by physics working groups
  - evaluate performance of
    - detector & reconstruction
  - calibrations using tracks
  - physics analysis
- disk space ~5TB at day-1
  - 7-10 ~0.5 TB miniDSTs resident on disk
  - frequency of mining I/O, tape mount limited
    - typically search through ~10-20 TB
- at steady state - 14 TB of disk
  - allow 2 1TB mini-DSTs per working group
- microDST starts with group level mini-DST
  - select 10-100 GB for specific analysis
  - export to user's desk top

Plan

- Calibration
- Bookkeeping
- Event Reconstruction & DST
- Analysis
- Simulations
- Framework to support it all
- Use Mock Data Challenge
    - test basic components & RCF interaction
    - assess architecture
    - assess implementation decisions
    - reevaluate plan

PHENIX

Next efforts (for Day-1)

- Implement calibration software and database
- Reconstruction
  - automation & bookkeeping
  - optimize performance
- Analysis
  - use of cluster (design not fixed yet)
  - archive analysis software
- DST logical, physical structure, contents
- Optimize data access vs. resource limitations
- Scope of objectivity usage
  - storage of DST, mini-, micro-DST's
- update Tag database contents, usage
- object-oriented data model
- event displays
- remote sites and data export strategy

Core offline milestones

Core Offline			
task	**January-98**	**April-98**	**Jul-98**
negotiate RCF contrib	**1/98**	**1/98**	**1/98**
ROOT installed	**2/98**	**2/98**	**2/98**
IDL compiler: more tables	**2/98**	**2/98**	**2/98**
IDL compiler: multidimensional arrays	**2/98**	**2/98**	**2/98**
STAF tutorial	**3/98**	**3/98**	**3/98**
put/fetch to database	**3/98**	**4/98**	**8/98**
database templates	*?/98*	*7/98*	*8/98*
software chain test for MDC	**4/98**	**5/98**	**6/98**
read raw data format	**5/98**	**5/98**	**7/98**
write DST	5/98	5/98	8/98
demonstrate DST into ROOT	**5/98**	**5/98**	**7/98**
Reconst farm: conceptual design	**5/98**	**5/98**	**7/98**
generic histogram interface		**5/98**	**7/98**
calibrated simple physics signal (multiplicity	**6/98**		**6/98**
develop distributed computing (DB) access	7/98?		11/98
run on multiCPU farm (for MDC)	8/98	8/98	8/98
Grand Challenge framework evaluate	8/98	8/98	9/98
GC: tag database	?/98		8/98
event display	9/98	6/98	8/98
STAF:multiple entry points in module	10/98		
Analysis cluster: conceptual design	11/98	11/98	11/98
physics signal (easter egg) hunt	3/99	3/99	3/99
Day-1 operation	9/99	9/99	9/99
Class libraries (CLHEP, STL decisions)	**?/98**	**?/98**	**6/98**
review MDC & update priorities, milestones			10/98
IDL compiler: nested structures	?/98	timely finish impossible	
STAF:TCL/TK or Java user interface	?/98	timely finish impossible	
OO data model	?/98		
Reconst farm: automate production	?/98		
Reconst farm: bookkeeping	?/98		
GC: query/fetch			9/98
GC: micro DST			
provide regression testing tools	?/98		
maintain scripts, environment, directory trees	continuous		

Simulation milestones

Simulations				
task	**January-98**	**April-98**	**July-98**	
GEOM: PC update	**1/98**	**1/98**	**1/98**	
update PISA instructions on web	*1/98*	*4/98*		
incorporate EMCAL response code	**1/98**	**2/98**	**2/98**	
update PISA web pages	**2/98**	**4/98**	**4/98**	
DIO:merge function	**2/98**	**2/98**	**2/98**	
MDC plan - who writes what	**2/98**	**3/98**	**3/98**	
needs assessment for fast MC	**2/98**	**2/98**	**2/98**	
formulate plan for geometry database	*3/98*	*4/98*	*8/98*	
DIO:modify for TOF, BBC	**3/98**	**3/98**	**3/98**	
DIO:fix header	**3/98**	**3/98**	**3/98**	
separate PISA read from common DIO		*4/98*	*8/98*	
ID/key #'s to collect info from different files; maintain association	*3/98*	*5/98*	*8/98*	
Geom in/out of STAF tables	*4/98*	*4/98*	*8/98*	
MDC:make raw data tables/check with subsystems	**4/98?**	**4/98**	**7/98**	
automate event generation process; DB for run info - release to PHENIX	4/98	5/98	8/98	
merge digitizations in STAF	**4/98**	**4/98**	**5/98**	
MDC:create parallel GEANT,data files	*5/98*	*5/98*	*8/98*	*nearly done*
instructions for PISA production runs	6/98	6/98	8/98	
3D magnetic field	7/98	7/98	?	J. Lajoie help
MDC:generate 100K events	**7/98**	**7/98**	**7/98**	
PISA into GEANT4	6/00			
GEOM: MVD update	**?/98**	**1/98**	**1/98**	
verify muon response	**?/98**	**3/98**	**3/98**	
integrate fast MC-PISA with switches	?/98			
get fast MC parameterizations/ modules from subsystems	>8/98			
lepton generator for PHENIX	>8/98			
event generator data into STAF tables	>8/98			

Central reconstruction milestones

Central reconstruction				
task	**January-98**	**April-98**	**Jul-98**	
Track:all fill raw data tables	**1/98**	**1/98**	**1/98**	
Track:GEANT geometries as built	**1/98**	**1/98**	**1/98**	
Track:all in STAF as C modules	**1/98**	**1/98**	**1/98**	
EMCAL in STAF (fortran)	**1/98**	**1/98**	**1/98**	
Track:TEC intra planar eff>95%	**2/98**	**4/98**	**6/98**	
Track:TEC inter planar eff>95%	**2/98**	**3/98**	**3/98**	
Track:DC x-wire eff>95%	**2/98**	**3/98**	**3/98**	
EMCAL in STAF (C)	**2/98**	**3/98**	**6/98**	
Calib: get subsystem status info	**2/98**	**2/98**	**2/98**	
Track:DC UV eff>85%	**3/98**	**3/98**	**3/98**	
Track:PC cluster eff>95%	**3/98**	**4/98**	**5/98**	
Track:TEC slow simul complete	**4/98**	**4/98**	**4/98**	
Track:Global eff>90%; <10% ghosts	**4/98**	**4/98**	**7/98**	
measure track reconst time in central Au + Au	*2/98?*	*4/98*	*8/98*	
match track to TOF, EMCAL eff>90% with <10% ghosts	*4/98*	*4/98*	*8/98*	
Mom:reconstruction eff at CDR leve	**4/98**	**4/98**	**7/98**	
match track to RICH; eff>90% with <10% ghosts	*5/98*	*5/98*	*8/98*	
Mom:kinematics reconst.@CDR leve	*5/98*	*5/98*	*7/98*	
TEC PID eff>90%	*5/98*	*5/98*	*8/98*	
complete central arm PID for e,K; >90% efficiency (define efficiency?)	*6/98*	*6/98*	*9/98*	
freeze chain for MDC	*6/98*	*6/98*	*8/98*	
Calib: database interface used; parameters applied	7/98	7/98	10/98	
Calib: alignment spec complete	7/98	7/98	10/98	
Calib: alignment adjust sw complete	9/98	9/98	12/98	
Monit:subsystem unpack complete	?/98		8/98	
Monit:subsystem software complete	1/99	1/99	1/99	
integrated event display complete	1/99	1/99	1/99	
Calib: generation sw for calib params complete	?/98		1/99	

Muon reconstruction milestones

Muon reconstruction task	February-98	April-98	Jul-98	
DetResp:muTR in STAF	1/98	1/98	1/98	
DetResp:muID in STAF	2/98	2/98	2/98	
Geom:muTR & muID in	2/98	2/98	2/98	
Cal:muTR use calibs in	2/98	2/98	2/98	
Track reconstruction sw	2/98	2/98	2/98	
module to create dimuon	2/98	2/98	2/98	
Geom:STAF utils for muID	3/98	3/98	4/98	
DetResp:update	3/98	3/98	3/98	
improve muID trigger	*3/98*	*5/98*	11/98	K. Read
PatRec:works with single	3/98	3/98	3/98	
connect tracks to MVD	3/98	4/98	8/98	
PatRec:works with	4/98	4/98	4/98	
PatRec:upgrade (Kalman filter?)	4/98	4/98		simple pat rec OK
combine muTR & muID	?/98		7/98	
PatRec:performance w/o			***9/98***	
Cal: connect to database	*5/98*	*5/98*	*9/98*	
final check of code for	*5/98*	*5/98*	*8/98*	
Geom:verify muTR as	6/98	6/98	*6/98*	
Geom:verify muID as	?/98	3/98	3/98	
DetResp:tune muTR response	?/98	6/98	6/98	
DetResp:muTR in STAF	*6/98*	*6/98*	*8/98*	
simulate DCM zero	1/99	1/99	1/99	
Cal:sw to calculate & store	1/99	1/99	1/99	
begin work on event	?	4/98	4/98	
Cal:cathode strip pos calib	2/99?	2/99	2/99	
MUID:discriminant	*3/98?*		*9/98*	
MUID: road finder		*4/98*	*9/98*	

Analysis development

- Physics working groups underway
  - each participates in MDC
- DST contents defined
  - will reduce size after MDC
  - working groups make mini/micro DSTs
  - separate streams planned
- physical structure of DST, mini-, micro-DST to be finalized after MDC
  - manpower limitation
- Tag database for event selection under development for MDC
- Calibration strategy being developed (see talk of Axel Drees)
  - Subsystem calibration software started

Questions to the committee

- Is PHENIX on track for day-1 readiness?
  - is performance OK for this stage?
  - are priorities right?
  - will we meet our requirements?
- Do we have right mix of own vs. externally written software?
- Is the plan right?
  - what is missing or overemphasized?
  - is it feasible?
    - is there enough time & manpower?

Manpower

- Reconstruction - all volunteers
  - J. Mitchell specified/coordinated tasks
    - Central has contribution from 22 people
    - includes 5 students
    - people working 1/2 - 2/3 of their time
  - Muon: 6 people (1 student) ~1/2 time
- Core Offline
  - 2.5 FTE
    - 1.5 FTE required to maintain software
    - offer for postdoc in the works
    - trying to recruit more collaboration help
      - T. Ichihara (RIKEN) - benchmarking, data export
    - M. Purschke where ONCS overlaps
- Simulation
  - 3.5 FTE (with 1 student 1/2 time)
  - P. Chand (BARC) currently banned!
- Analysis - remains to be seen

Manpower concerns

- Core offline seriously understaffed
  - skills required deter casual contributions
  - need long-term collaboration commitments
  - missing production manager,
    - GC & microDST implementation,
    - Data model development,
    - ROOT (& other) tools developer
- Simulations team very small
  - focus is shifting
  - geometry database implementation looms
  - incredible effort to achieve milestones
  - Q/A is ramping up
- Reconstruction
  - contribution from collaboration good
  - manpower tradeoff between optimization & analysis development
- Analysis
  - need to switch framework
  - organize infrastructure vs. individual's code

Cost to Complete

- Core offline
  - spent so far: $622K
    - $450K for manpower
  - remaining:
    - $69K in FY98
    - $180K in FY99

- Simulations
  - spent so far: $160K
    - $70K for manpower
  - remaining:
    - $50K in FY99
      - 1/2 postdoc + expenses
      - $15K for computer equipment