Star Computing Requirements

Peter Jacobs, LBNL

Oct. 20, 1997

- RCF Review, July '97: *The experimental computing requirements need more study, especially those requirements that have profound effects on the RCF system architecture and on the size and scope of the remote facilities...unrecoverable architectural choices should not be made on the basis of unjustified and indefensible requirements.*
- Computing for RHIC is on unprecedented scale, capabilities will directly impact physics output

- Previous estimates (ROCOCO-2) are now out of date or seen to be inadequate, need updating
- Task Force of STAR collaborators convened by spokesman in August ’97 to undertake bottom–up review of STAR computing needs
- Preliminary Task Force report due Oct.1, ’97, final report due Nov. 1, ’97
- Status: behind schedule! Several drafts circulated, continual progress, but no clear estimates yet

StarCompReq Task Force: Issues

- Large computing facilities for RHIC should be *scalable*: only incremental cost to grow with increasing physics demands
- Distinguish between **character** and **scale** of STAR offline analysis:

  character: *relative* requirements of universal DST production, CPU-intensive and I/O-intensive μDST production and analysis

  scale: overall *magnitude* of CPU cycles, data volume, bandwidths required for each step

StarCompReq Task Force: Approach

- Clearly separate estimates of *character* and *scale*
- To estimate character, study in detail small number of analysis projects from each general STAR physics category
- Selected analyses should range from *typical* to *demanding* on computing resources
- For selected analyses: estimate in detail required number of events, data volume and cpu cycles for DST and μDST production
- Both data analysis and simulations
- Estimates based upon previous experience for similar analysis plus model calculations for RHIC

- Task Force contains experts in all areas of STAR Physics
- New ingredient in estimates for most of us is OODBM
- After *character* established, try to estimate *scale* by assessing fraction of STAR analysis effort engaged in each physics category with how many projects
- Rule of thumb: our *best* estimate will be accurate to no better than a factor of two

StarCompReq Task Force: Membership

Chair Peter Jacobs (LBNL)

Soft hadronic physics W. Llope (Rice), S. Margetis (Kent State)

Hyperon physics K. Wilson (WSU), P. Jones (Birmingham)

Leptons and D-mesons T. Ullrich (Yale)

High Pt, jets and photons T. LeCompte (ANL), W. Christie (BNL)

Event-by-event physics T. Trainor (UW), I. Sakrejda (LBNL)

Spin physics G. Eppley (Rice), H. Spinka (ANL)

Peripheral collision (two-photon) physics S. Klein (LBNL)

Particle correlations (HBT) S. Pandey (Wayne State), N. Xu (LBNL)

Ex Officio Lanny Ray (Texas), Torre Wenaus (BNL)

StarCompReq Task Force Report

1. Introduction
2. Physics Motivation
3. General Comments on Offline Analysis
4. Data Sets
5. DST Production
6. Data Volume
7. μDST Production and Analysis
8. Simulations and Corrections
9. Basis for CPU Estimates
10. Scaling to Total STAR Requirements
11. Non–RCF Facilities
12. Summary
13. Conclusions

StarCompReq Task Force: Status

- *This is work in progress: all tables are under development and many numbers are not yet defendable*
- Simulation estimates are particularly weak or nonexistent
- Detailed discussion of numbers in tables by committee is discouraged: they are shown only to indicate approach and status

STAR Physics Categories

Colliding System	Observables	Online Selection	Offline Event or Track Selection
Au-Au	Single Particle Spectra	global	μDST
	Hyperon Yields and Spectra	global	DST, μDST
	Correlations	global	μDST
	Event-by-Event	global	μDST
	Dileptons	global	μDST
	Peripheral Collisions	specific	minor
	High p_T π^0, γ, jet	specific	μDST
pp	All	specific	minor

Data Sets for Selected Analyses

Data sets required for selected “rare” processes that drive size of data volume

Preliminary

Physics Category	Observable	Particle	# central Au-Au events
Soft Hadronic Probes	dN/dy	$\bar{p}$	50K
	$< p_T >$	$\bar{p}$	20K
	dN/dydp_T	K^0_s	50K
		$\bar{p}$	200K
Hyperons	dN/dydp_T	Λ	1M
		Ξ^-	20M
	dN/dy	Ξ^-	2M
	total yield	Ω^-	20M
Correlations	R_{out}, R_{side}, R_{long} vs. y, k_T	$\pi\pi$, p_T>1 GeV/c	2M
		$K^\pm K^\pm$	2M
		pp	2M
	2-D (?)	$K^0_s K^0_s$	4M
	multiparticle	pions	2M
Event–by–Event			all events
High p_T	p_T> 10 GeV	γ	>10M (triggered)
	p_T> 20 GeV	Jets	>10M
Leptons and D–mesons	yields	$\phi \to e^+e^-$	10M
		$J/\psi \to e^+e^-$	10M
		D mesons	3.5M

Data sets for $\gamma\gamma$, pp and spin

Preliminary

Physics Topic (section)	Observable	# Events	# equivalent central Au-Au events
Peripheral Collisions	all, 1 year's running	20M	15K
pp and Spin	all, 1 year's running A_{LL}	60M 260M	1M 4M

μDST Production and Analysis

Data volume and CPU for selected "demanding" analyses
Pre–Preliminary (Note non-uniform metrics for CPU estimates)

Physics Category (section)	Observable	μDST Volume	CPU Time
Soft Hadronic Probes	dN/dp$_T(\bar{p})$	40 GB	negligible
Hyperons		2 GB	
Correlations		1 TB	260 HP735 days
Event–by–Event	all	5 TB	3 GFlop-years
High p$_T$	Jets, γs	few % of DST	few % of DST
Leptons and D–mesons	$\phi \rightarrow e^+e^-$	340 GB	$3 \cdot 10^7$ sec
	$J/\psi \rightarrow e^+e^-$	32 GB	$3 \cdot 10^7$ sec
	D mesons	450 GB	$3 \cdot 10^8$ sec
Peripheral Collisions	all	few $\cdot$ 10 GB	negligible
pp and Spin	all	150 GB	negligible

Simulation requirements, selected topics

Simulations are for correction of data only, not large scale study of MC event generators

Not even Pre–Preliminary

Scaling to Full Star Requirements

- Use current membership of Physics Working Groups to gauge eventual mix of analysis projects
- Estimate ∼100-150 simultaneous projects
- must develop various scenarios for mix of standard/demanding analyses as function of time
- time–dependent ingredients: STAR hardware configuration, detector understanding, run plan, configuration of RCF and other computing facilities, accelerator performance,...
- try to determine systematic error bar on final estimates

StarCompReq Task Force: Summary

- Work in progress, many gaps, no serious numbers yet
- time scale for first numbers that are “discussable” is 1-2 months
- This is a long term study: must progress and improve in close consultation with RCF as STAR and RHIC develop