



# Reprocessing of Heavy Ion and Muon Data

Alexandre Vaniachine  
ATLAS Software & Computing Workshop  
April 4-8, 2011  
CERN, Geneva, Switzerland

# Responsibilities since the Previous S&C Workshop

- R. Walker and K. De provided overall coordination of GDP activities, plus covered numerous day-do-day operational issues such as
  - R. Walker migrated GDP crons to the new SLC5 machine (including job definition)
  - K. De debugged various GDP production issues when R. Walker and A. Vaniachine were on vacations
- A. Vaniachine served as a primary contact for the reprocessing and was granted rights to the job definition code (developed by P. Nevski)
  - 7K lines in a dozen of scripts
- W. Deng joined GDP team with responsibilities in
  - automating Sites Validation
  - eliminating duplicate files in production datasets
  - assuring event store integrity by eliminating corrupted events
- M. Borodin continued his responsibilities in database releases for reprocessing
- Further help is expected from D. Golubkov arriving at CERN tonight
- P. Nevski is back at CERN catching up on GDP developments while he was away



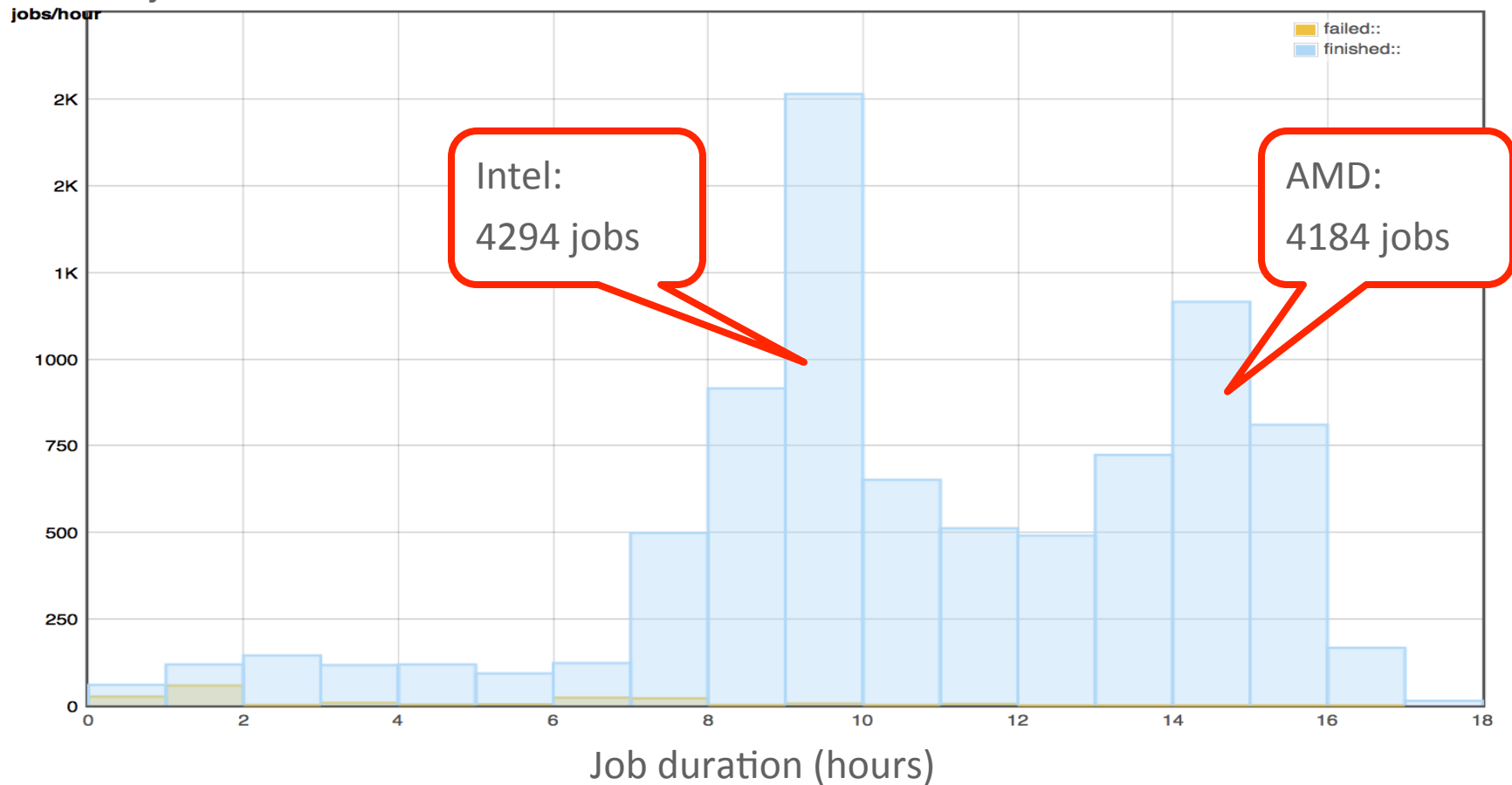
# Preparations for Reprocessing of Heavy Ion Data

- The initial schedule for bulk HI reprocessing has been shifted by a month to assure software and conditions data readiness
  - Iterative Sites Validations cycles and several Slice Tests conducted by GDP assured the software release and conditions data quality required for HI reprocessing (A. Olszewski):
    - For the first time the reprocessing was done without the option `---ignoreerrors=True`
- In time for the HI reprocessing several new developments were delivered that were proposed at the previous S&C Workshop
  - Improvements in bamboo brokerage and multi-cloud capabilities
  - Improvements in Conditions DB Release packaging for reprocessing
  - Improvements in monitoring required for GDP operations during reprocessing
  - Prototype MetaMiner services required for mining the log files for errors and other metadata required for GDP operations during reprocessing



# Demonstrated Value of New Developments

- New monitoring capabilities (delivered by V. Fine) revealed unexpected shape of the job duration distribution



- MetaMiner prototype (by W. Deng) used the log files to determine that the shape is due to an unexpected software performance dependence on the CPU type

# Data Redistribution in Heavy Ion Reprocessing

- Several iterations were necessary to optimize the initial data distribution  
(A. Klimentov, A. Di Girolamo, C. Serfon, R. Walker et al.)

Tier1	BNL	FZK	IN2P3-CC	INFN-T1	NDGF-T1	PIC	RAL	SARA	TAIWAN	TRIUMF
TB	51.4	15.4	14.5	37.8	4.7	25.5	4.5	79.1	31.2	20.4

- Final data selection for reprocessing by the HI community has 0.25 PB in 44 runs
- For HI derived data (0.35 PB) CREM2 approved the pp data distribution scenario
  - In brief: two primary replicas ATLAS-wide for DESD and NTUP



# Day-to-day Monitoring Table: HI Reprocessing

- To update the table R. Walker updated the cron created by P. Nevski
- All HI reco jobs were defined with the help and guidance from P. Nevski
- Start delays are due to scheduled downtimes
- Tails discussed on Slide 11

date	ALL	CA	DE	ES	FR	IT	ND	NL	TW	UK	US
11-03-15	13752	1084	1455	1446	1381	2436	418	1548	38	1	3945
11-03-16	50366	2928	7470	4719	6279	4997	4592	3365	2206	926	12884
11-03-17	55238	2669	13640	3802	5621	3961	3882	3816	4418	5655	7774
11-03-18	55108	2889	12863	4159	4035	2284	4017	5403	4271	4998	10189
11-03-19	42547	732	6830	4260	2504	4007	3727	4153	4210	1825	10299
11-03-20	42160	1357	3960	5005	787	3767	3550	3032	3356	3311	14035
11-03-21	30036	3235	3873	1631	22	4386	3157	2993	2999	5434	2306
11-03-22	14013	3409	0	0	3	3456	226	972	4410	1537	0
11-03-23	12611	3250	0	0	0	3067	1	2739	3554	0	0
11-03-24	8286	1750	0	0	0	3247	0	3251	38	0	0
11-03-25	6517	2	0	0	0	3423	1	3084	7	0	0
11-03-26	660	0	0	0	0	567	0	93	0	0	0
11-03-27	35	0	0	0	0	33	0	1	1	0	0
11-03-28	18	0	0	0	0	0	0	0	18	0	0
<b>total done</b>	331347	23305	50091	25022	20632	39631	23571	34450	29526	23687	61432
<b>total jobs</b>	331953	23305	50091	25022	20632	39910	23575	34761	29526	23687	61444
<b>% done</b>	99.8	100.0	100.0	100.0	100.0	99.3	100.0	99.1	100.0	100.0	100.0
<b>aborted</b>	606	0	0	0	0	279	4	311	0	0	12
<b>% aborted</b>	0.2	0.0	0.0	0.0	0.0	0.7	0.0	0.9	0.0	0.0	0.0
<b>jobs left</b>	0	0	0	0	0	0	0	0	0	0	0



# Job Definition

- Worked smoothly mostly because of the efforts invested since I took the responsibility in December



date	ALL	CA	DE	ES	FR	IT	ND	NL	TW	UK	US
11-03-15	13752	1084	1455	1446	1381	2436	418	1548	38	1	3945
11-03-16	50366	2928	7470	4719	6279	4997	4592	3365	2206	926	12884
11-03-17	55238	2669	13640	3802	5621	3961	4592	3816	4418	5655	7774
11-03-18	55108	2889	12863	4159	5621	4007	4017	5403	4271	4998	10189
11-03-19	42547	732	6830	4260	4035	3727	4153	4210	4998	10189	10299
11-03-20	42160	1357	3960	5005	2504	3767	3550	3032	3356	1825	14035
11-03-21	12154	913	2142	1172	787	1561	944	1034	829	3311	2199
total done	271325	12572	48360	24563	20618	23013	21130	22351	19328	18065	61325
total jobs	331953	23305	50091	25022	20632	39910	23575	34761	29526	23687	61444
% done	81.7	53.9	96.5	98.2	99.9	57.7	89.6	64.3	65.5	76.3	99.8
aborted	0	0	0	0	0	0	0	0	0	0	0
% aborted	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
jobs left	60628	10733	1731	459	14	16897	2445	12410	10198	5622	119
% left	18.3	46.1	3.5	1.8	0.1	42.3	35.7	34.5	23.7	23.7	0.2
running	11378	1178	1731	459	14	2010	2445	1284	845	1493	119
days left	-	5	0	0	0	5	0	3	3	2	0



# ADCoS Provides Valuable Support

- Several bugs that were found during large-scale Sites Validation and Slice Test iterations performed by GDP have not been fixed, which led to event losses from the jobs that failed or produced corrupted outputs
  - Such lost events could be recovered with dedicated bug-fix software release
- In addition to the known bugs, during the main HI reprocessing campaign we encountered rare (previously unknown) bugs that result in event losses
  - To recover events, these bugs also have to be fixed in the software release for recovery
- Such rare bugs that slipped through thorough validation are hard to find, but since these bugs lead to event losses, we need to find them as soon as possible
  - That is why GDP appreciates much when shifters go beyond the (shallow) problem reporting and perform deep debugging down to the source of a job failure or when they sift through a multitude of trivial problems to report genuine reproducible bugs
    - Improvements in bug reporting documentation would be valuable
- To encourage such extra effort, I suggest to award extra OTP credits to shifters that submitted high quality savannah reports:
  - Jim Siegrist
  - Pavol Strizenec, et al...





# Lost Data to be Recovered

- Absence of the option `---ignoreerrors=True` proved to be valuable
  - Several true errors were revealed, which could affect physics results if ignored
- 1)  $\sim 0.2\%$  level - misconfigured propagator: bug #79658
  - Only in two special HI runs with solenoid field off
    - Lost data will be recovered during the recovery step
- 2)  $\sim 4 \times 10^{-5}$  level - events from the last lumiblock lost: bug #78999
  - Has to be dealt with outside of the GDP operations
- Plus various “usual” reasons at the  $\sim 3 \times 10^{-6}$  level
  - No success after 25 attempts
    - <http://panda.cern.ch/server/pandamon/query?job=1213133868>
  - Corrupted DESDM file
    - DESDM files were not checked for data corruption: bug #80128
  - Possibly corrupted files in HI reprocessing: bug #80241
  - Other reasons still under investigation



# Finishing HI Reprocessing

## Preparations for Muon and 2011 Data Reprocessing

- Last week the reconstruction step for all 0.25 PB of HI data was completed
  - As requested, the reconstruction was completed in two weeks
- Subsequent HI data merging steps (submitted by J. Strandberg) are about to finish
  - Hampered by the (known) irreproducible software errors
    - Some merging jobs took 19 attempts to finish
- In the meantime the software fix was provided
  - Hopefully, there should be no such problems in muon reprocessing
- As soon as FR and DE finished with the HI reprocessing tasks, they got the next round of jobs from the muon reprocessing tasks (J. Strandberg)
  - Processing of one large run to validate the software release candidate for physics
- Upon these results a new software release for muon reprocessing was built yesterday
  - J. Strandberg launched the first task to test if everything is running like it should
    - If the test is successful, the bulk muon reprocessing campaign will be launched tonight
- The next week is the target date for the bulk start of the reprocessing of the 2011 data
  - GDP preparations are under way



# Advancements in Data Redistribution

- As in the HI case, the muon data distribution across the Tier-1 sites is uneven:

Tier1	BNL	FZK	IN2P3-CC	INFN	NDGF	PIC	RAL	SARA	TAIWAN	TRIUMF	Total
TB	99.7	28.3	30.5	9.3	6.6	-	13.2	24.9	10.2	8.3	230
Nb datasets	24	4	9	6	3	-	5	8	4	3	66

- Data Preparation coordinators advised that the data distribution should match the data processing capacities available at the Tier-1 sites
  - Other improvements in data distribution are under way to minimize the mismatch
- The optimal solution must take into account multiple factors:
  - Select T1 sites must finish early to be available for next reprocessing preparations
  - Scheduled downtimes at T1 sites
  - Redistribution of the derived data, etc.
- Data redistribution between sites is a non-trivial problem in linear programming
  - For the solution of such problems L. Kantorovich was awarded the Nobel Prize
    - GDP may engage this proven approach to achieve an optimal data redistribution
    - Yet another R&D topic for discussion at the S&C Workshop



# Conclusions and Summary

- Thanks to thorough preparations, reprocessing of 0.25 PB of HI data is on schedule
  - The reconstruction was completed in two weeks
  - Merging tasks are 99.97% done
  - Distribution of the outputs is ongoing following the “green light” from the HI community
  - The data recovery phase is under preparation
- Preparations for reprocessing of the 0.21 PB of muon stream data are completed
  - When Tier-1s finished with HI reprocessing tasks, they got the next round of jobs from the muon reprocessing tasks to validate the first candidate software release for physics
  - Following validation, the new software release was built yesterday
- Bulk submission for muon reprocessing is expected tonight
- In muon reprocessing (as with all pp data) there is one reco job per one input file
  - In case of HI data we have several reco jobs processing the same input file
    - Because of that the rate of data recall from tape was reduced by a factor of 3 in HI reprocessing
- Sites should expect higher rates of data recall from tape this week
- For the next week we plan the 2011 data reprocessing campaign
- **Summary: the GDP has been busy since the previous S&C Workshop**

