



Using HEP Computing Tools, Grid and Supercomputers for Genome Sequencing Studies

K. De, A. Klimentov, T. Maeno, R. Mashinistov, A. Nedoluzhko, A. Novikov, S. Panitkin, A. Poyda, E. Ryabinkin, F. Sharko, I. Tertychyy, A. Teslyuk, T. Wenaus

- Recent advances of Next Generation Genome Sequencing technology led to increasing streams of sequencing data that need to be processed, analyzed and made available for bioinformaticians worldwide.
- Sophisticated computing software WMS and efficient usage of the supercomputers can greatly enhance this process. PanDA - Production and Distributed Analysis Workload Management System has been developed to address ATLAS experiment at LHC data processing and analysis challenges and it was able to cope with increasing LHC luminosity, ATLAS data taking rate and processing and analysis challenges.
- The success of the PanDA has drawn attention from other compute intensive sciences beyond HEP (and Grid) such as bioinformatics.
- We have adapted the biology SW pipeline PALEOMIX to run it on a distributed computing environment powered by PanDA well known in HEP.

Using HEP Computing Tools, Grid and Supercomputers for Genome Sequencing Studies

- We dramatically decreased the total walltime because of jobs (re)submission automation and brokering within PanDA, what was earlier demonstrated for the ATLAS applications on the Grid.
- Using software tools developed initially for HEP and Grid can reduce payload execution time for Mammoths DNA samples from weeks to days.

