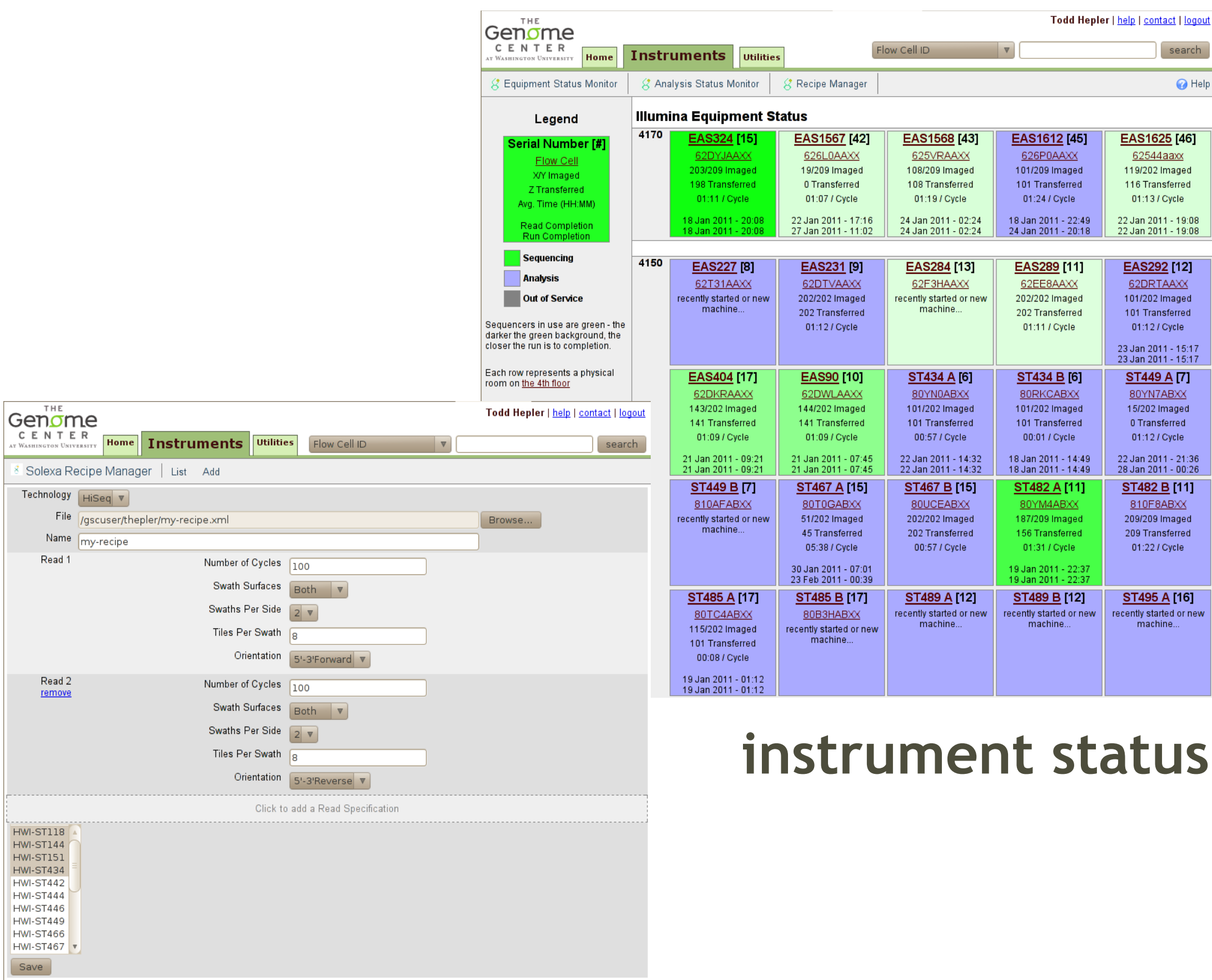


# Large-Scale Illumina Sequencing and Analysis

Todd Hepler, William Schroeder, Scott Abbott,  
Michael Kiwala, Indrani Das, Craig S. Pohl,  
David J. Dooling, Elaine R. Mardis,  
and Richard K. Wilson  
*The Genome Center at Washington University  
in St. Louis School of Medicine*

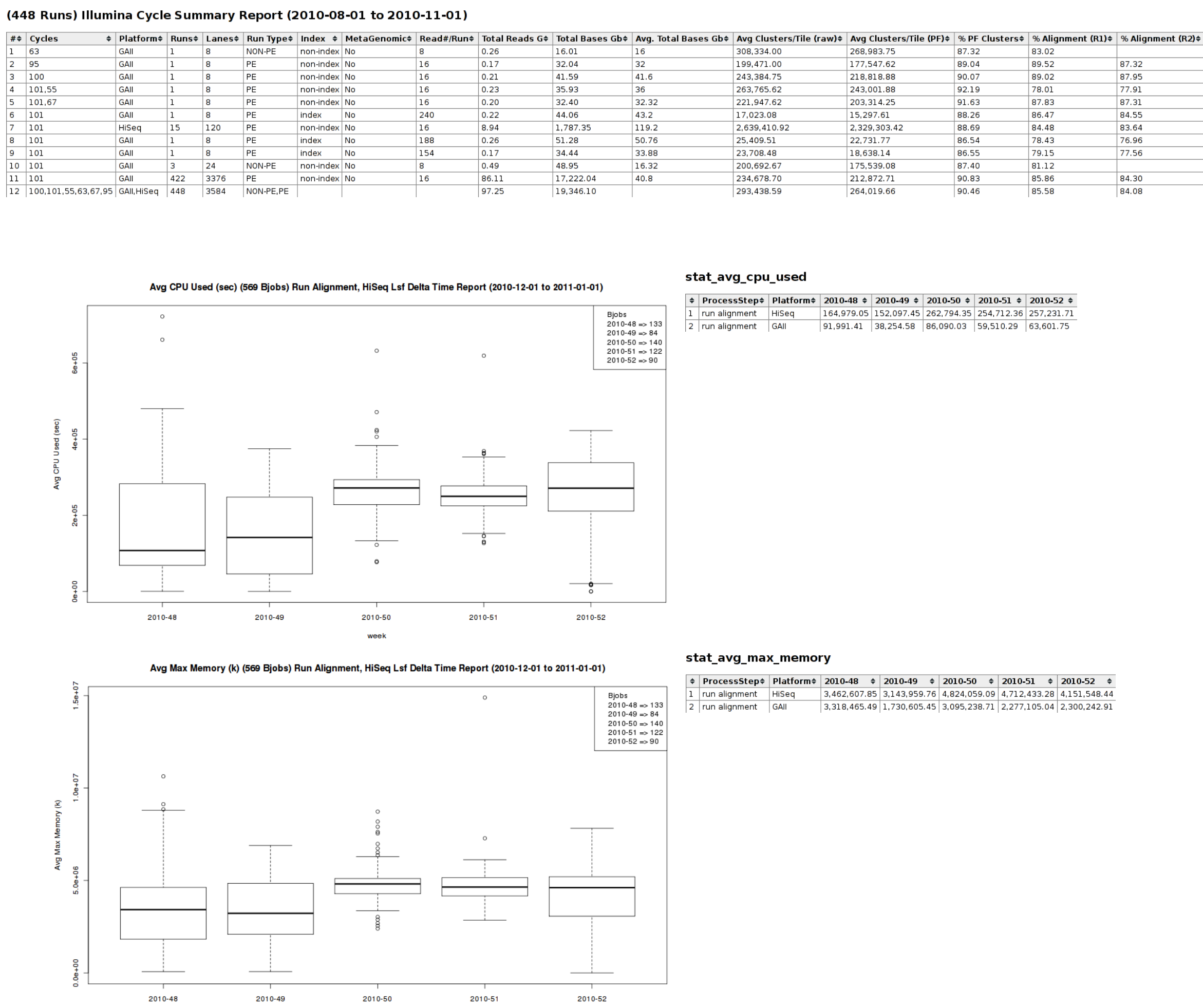
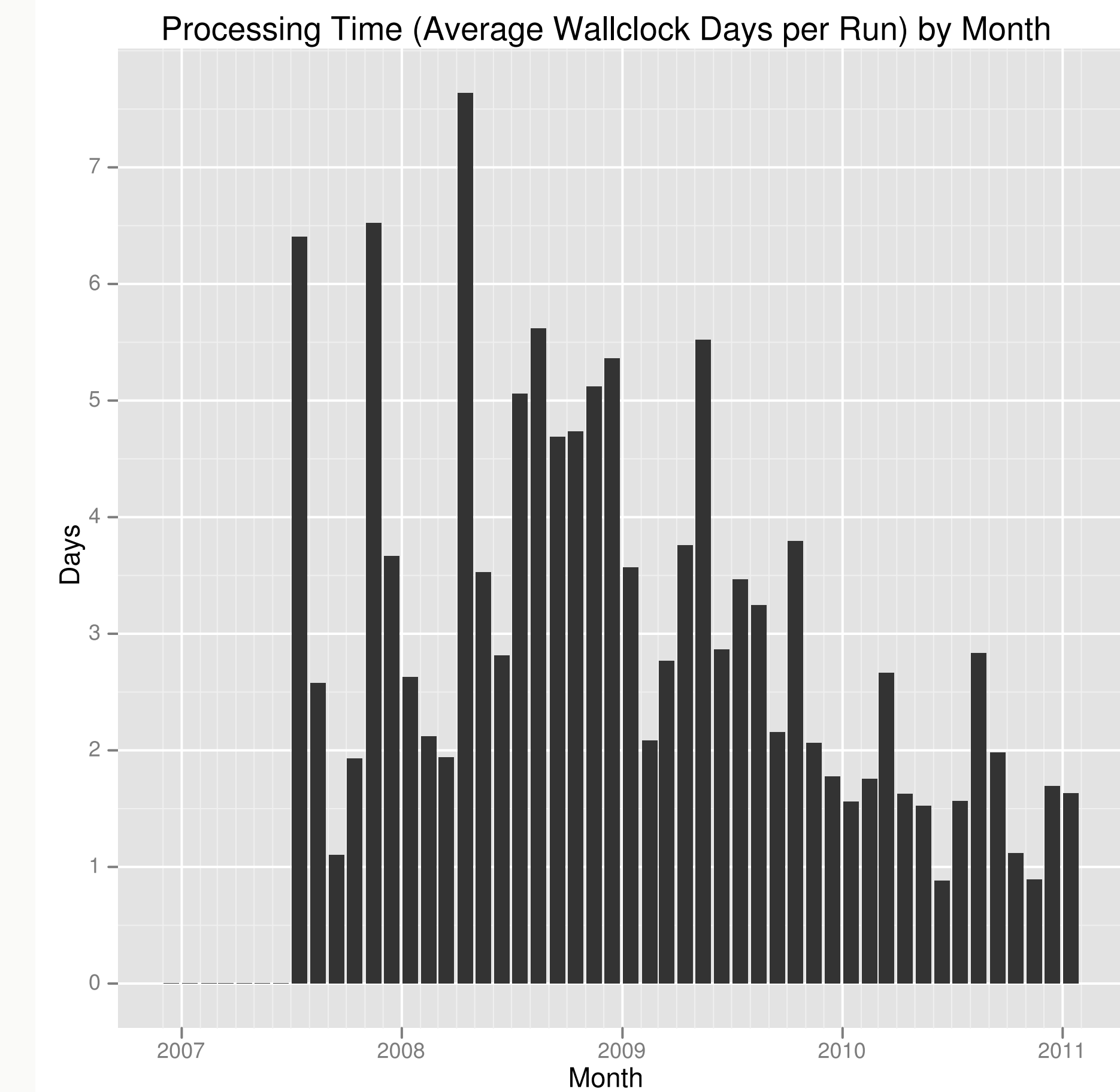
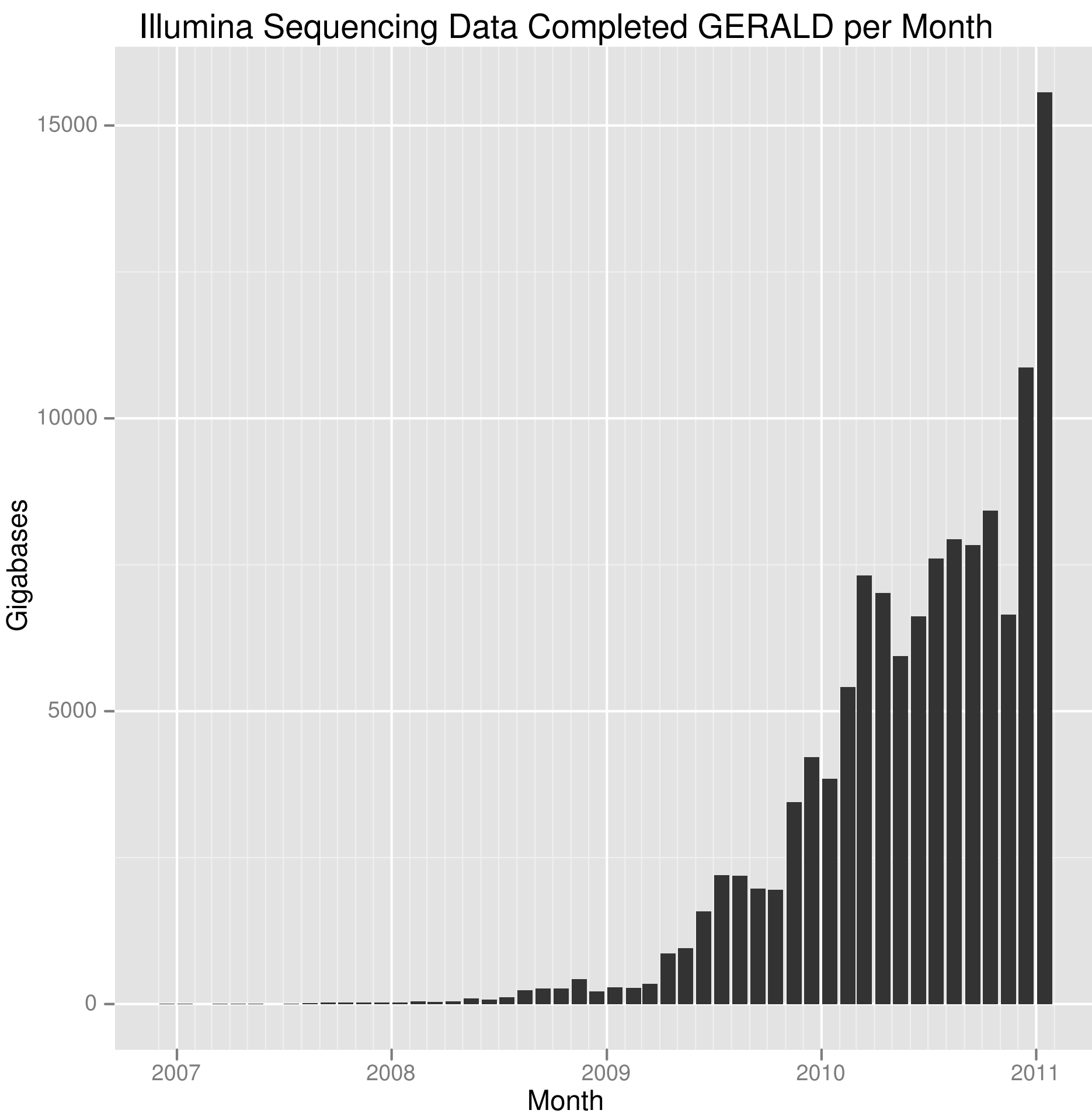
The Genome Center at Washington University has a large-scale, high-throughput Illumina sequencing and primary analysis pipeline that has required the development of customized software to maximize throughput, increase quality control and minimize costs. In order to facilitate this system, we have developed a custom disk allocation system for managing data types and processing requirements, an automated interface to the SCS/HCS instrument software, a system for parallelizing the primary analysis and CASAVA secondary analysis by lane or sub lane (index), and a systematic way for generating quality control statistics. Analyzed run data and metrics are stored in a data-warehouse database and reported back to technicians and project managers through a number of custom views, including web status monitoring and automated reports. This analysis pipeline continues to evolve in response to challenges of scale due to new sequencing instruments, vendor code releases, system bottlenecks, and maintenance outages.

## Instrument Management

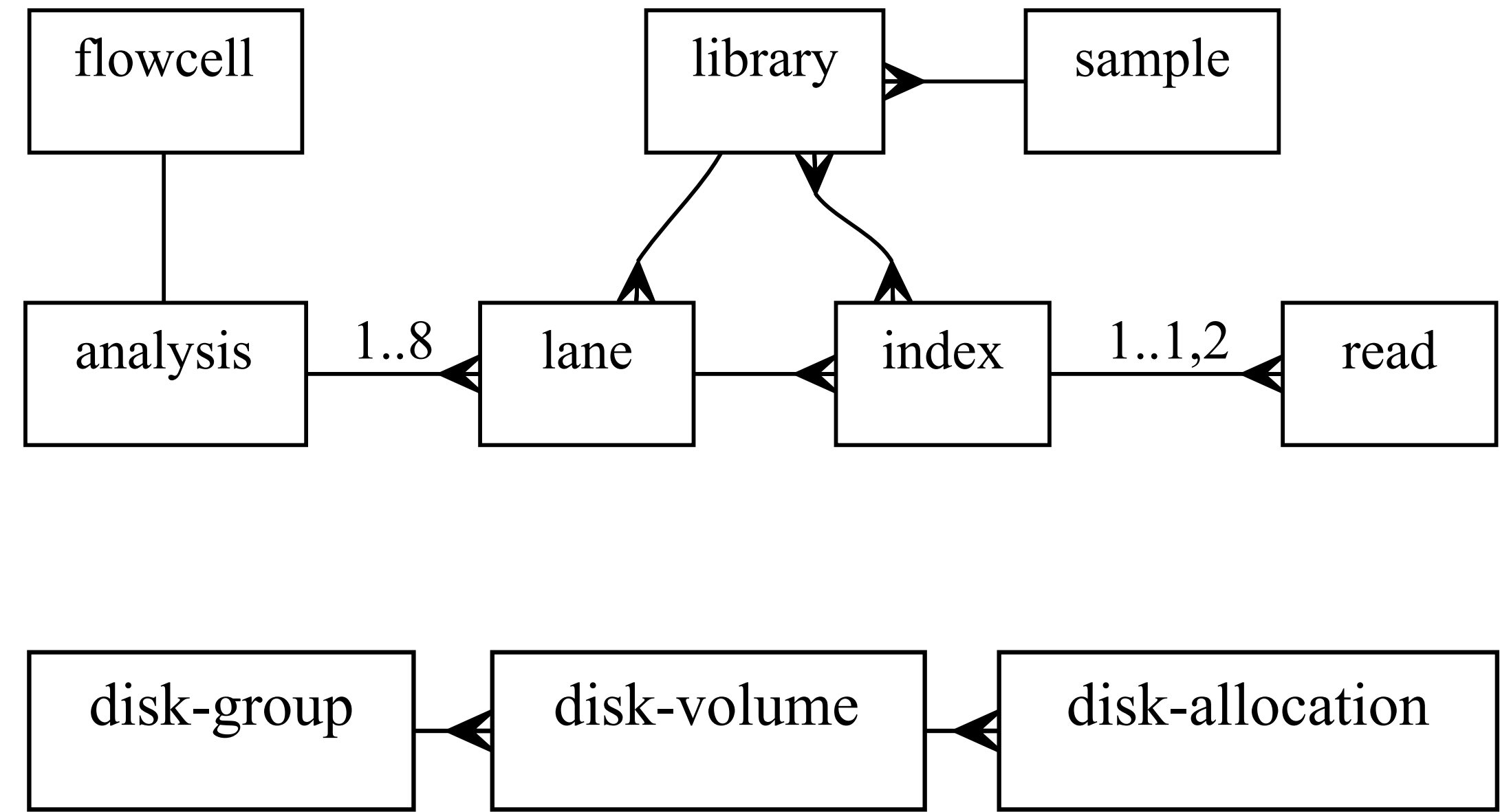


add recipes and assign to instruments

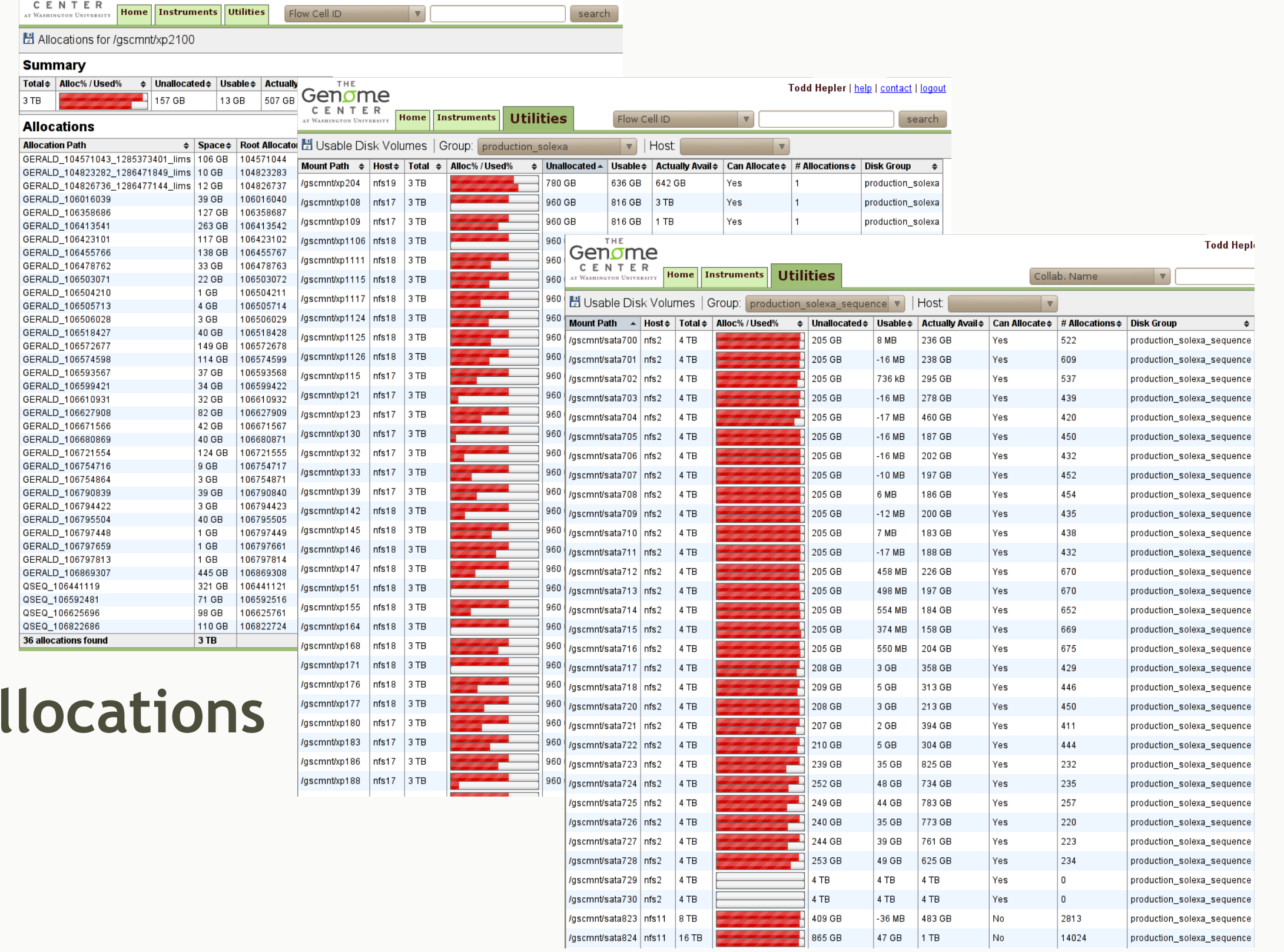
## Reports



## Schema



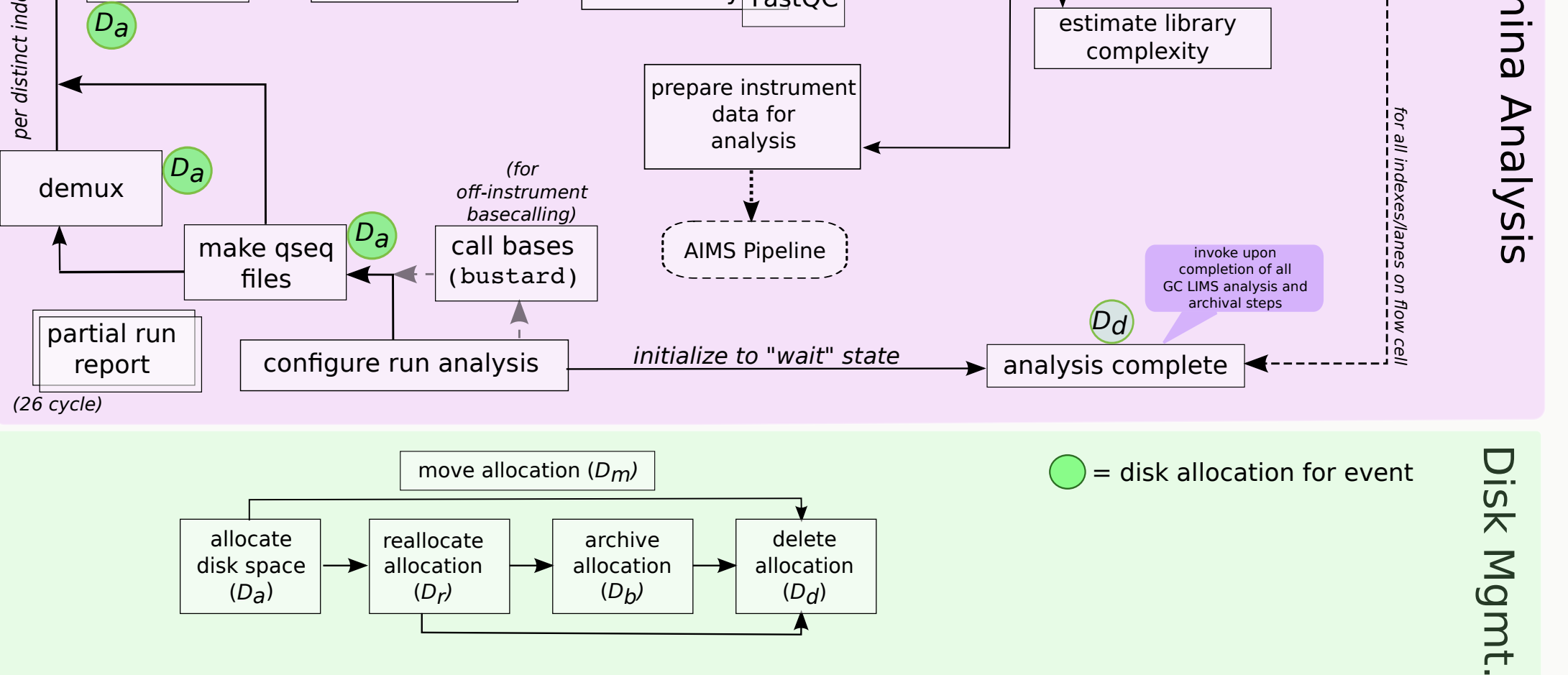
## Disk Management



allocations

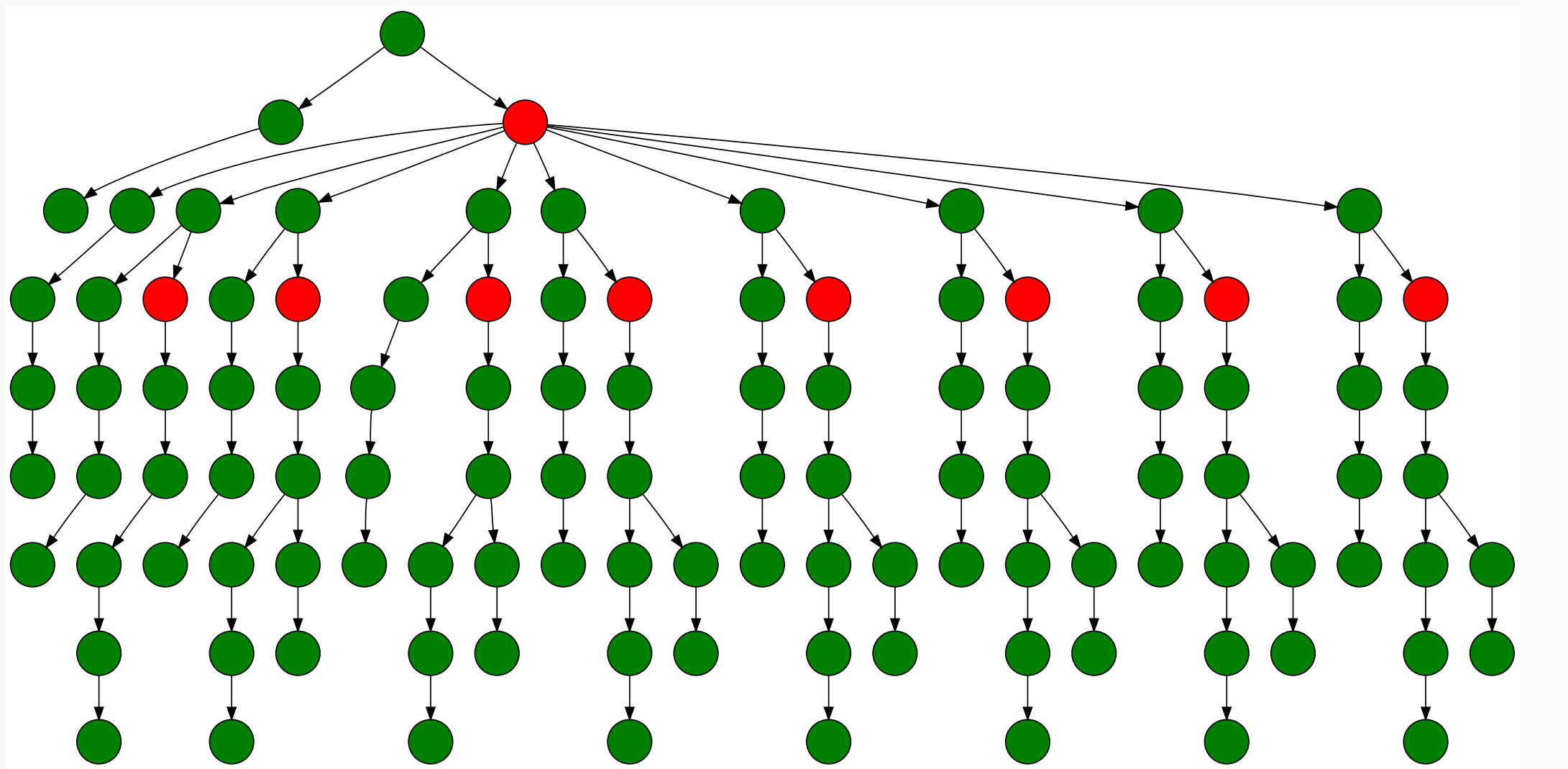
volumes by disk group

## Pipeline



1 Run - 8 lanes - non-indexed

● 1 core job ● 4 core job



## Future Work

- o Integrate CASAVA 1.8
- o Further analyze compute resources and optimize
- o Scale with increased number of HiSeq instruments