

# Glossary

**Compute cluster:** A collection of network-connected computers that work together to improve performance and increase processing speed of tasks

**Compute cluster administrator:** The system administrator appointed by the customer's company/institution and responsible for installing, maintaining, and allocating computing resources to Partek Flow and the compute cluster. The Partek Flow administrator and compute cluster administrator are distinct roles usually held by two different individuals.

**Compute node:** A single computer that is part of a larger computer cluster. It is used to run jobs and can have a different amount of internal computing resources than the other computers within the same cluster.

**Computing resource:** The physical components required to do computational work including, but not limited to, the number of CPUs, RAM, and available hard disk storage.

**Central processing units (CPU) and cores:** A core is the basic computation unit of the CPU. At the hardware level, a CPU consists of one or more cores where each core is a processing unit. Increasing cores enables more jobs to run simultaneously and significantly increases work completed. At the software level, there is no distinction between CPUs and cores, therefore, the total number of cores is reported as the number of CPUs. This is the case with Partek Flow as all references to CPU counts are the total number of CPU cores in the computer.

**Gene Specific Analysis (GSA):** Partek developed method to identify the statistical model that is best for a specific gene among all the selected models, then using that model to calculate pvalue and fold change.

**Head node:** The controlling computer in a compute cluster. The job scheduler and Partek Flow server are both installed on the head node and accessible to users. In most cases, jobs are not processed on the head node, but on the compute nodes.

**Jobs:** Computational work requested by users of a computing cluster. Each job requires computing resources to perform properly.

**Job queue:** A list of computational work requested by all users of a computer cluster and managed by the job scheduler. This includes all non-Flow work. There can be several job queues on a computer cluster with each having unique names and resource limits.

**Job scheduler:** Responsible for managing jobs requested by users of a shared computing resource. Management is accomplished by prioritizing, running, and tracking the status of jobs based on available computational resources and per-user computing resource limits. Schedulers are commonly employed in computing clusters. Examples of job schedulers include SGE, Torque, and Slurm.

**Job submission script:** The computer code that specifies the resources required for a specific job, along with the directions for how to execute the job. When starting a remote worker, the job submission script is used to specify to the job scheduler that all available resources on a particular computer will be used along with the command to start the remote worker. Typically anyone using a cluster must manually write this code to interact with the cluster. Partek Flow performs this for users and allows them to interact with the cluster without writing code.

**Linux user account:** The account under which all work requested by a user of a computer cluster run. The Partek Flow server and workers all run under a single Linux user named flow and all work done by all Partek Flow users is done using this single Linux account.

**Partek Flow administrator:** A special Partek Flow user account that is used to configure the Partek Flow server and manage user accounts. This account has full permissions to complete any action inside Partek Flow such as assigning user and project permissions and deleting user data.

**Partek Flow internal worker:** A Partek Flow worker that runs on the same computer running the Partek Flow server. There can be only one internal worker. This worker should be used only for single-server Partek Flow installations and disabled on computer clusters in order to keep the head node responsive.

**Partek Flow queue:** A list of Partek Flow jobs running or waiting to run. This list is usually sorted by priority or submission time. Partek Flow workers process the work in the Partek Flow queue.

**Partek Flow remote worker:** A Partek Flow worker that runs on a different computer than the Partek Flow server. There can be zero or more remote workers, which are managed by the Partek Flow server. Remote workers can be started using the cluster job scheduler. It is assumed that the Partek Flow remote worker can use all resources provided by a compute node, thus only one remote worker per compute node should be running.

**Partek Flow server:** The program that schedules Partek Flow-related computational work and allows users to view and analyze Partek Flow-generated results. This program runs on the cluster head node and sends Partek Flow user requested computational work to Partek Flow workers.

**Partek Flow user account:** User accounts internal to Partek Flow. There is no equivalency between Partek Flow user accounts and Linux user accounts.

**Partek Flow worker:** A program installed on a single computer within a computer cluster and receives job requests from the Partek Flow server. The program will determine if the computer has the needed resources to complete the requested job, and based on the answer, deny or accept the job and report when completed.

**Pipeline:** A series of tasks used to process and analyze genomic data.

**Server:** Any program that users interact with in order to accomplish a specific task or to receive data.

**Task graph:** The combination of circular data nodes and rectangular task nodes displayed under the Analyses tab.



Your Rating:  Results:  38 rates