

**Development of scheduler algorithm for
torque pbs and tcl program language.
Its usage inside IHEP grid-infrastructure**

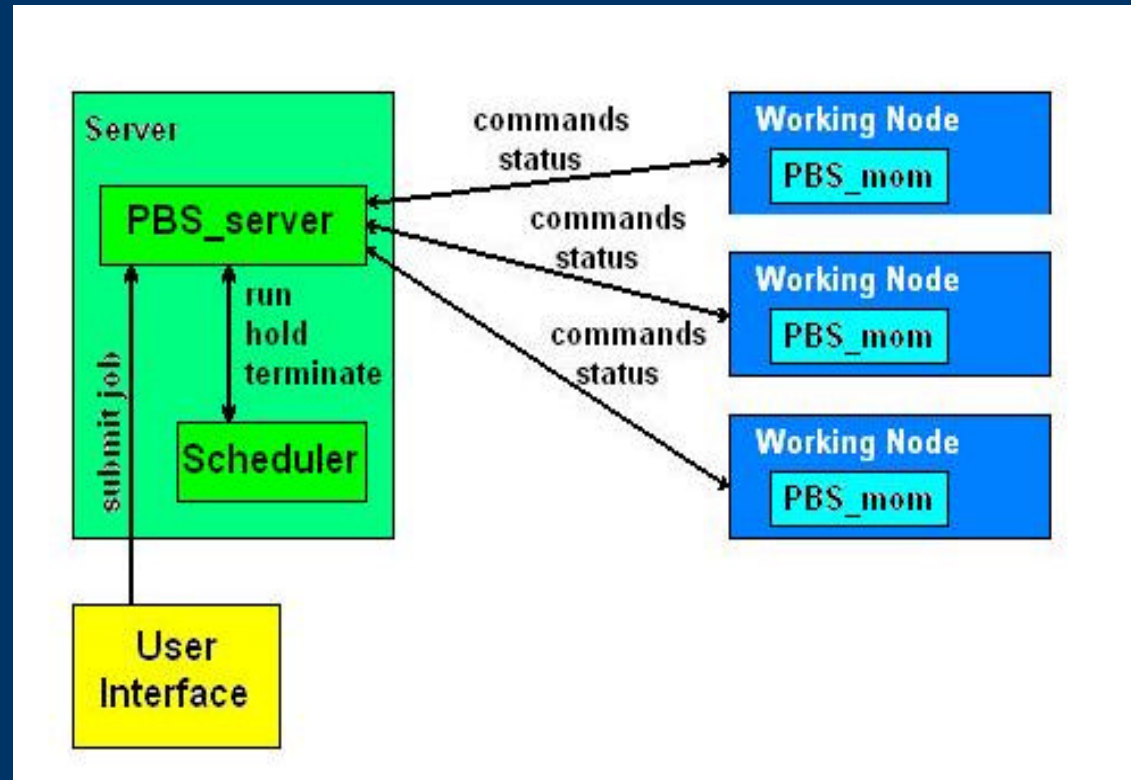
G.V. Volchkov

Institute for High Energy Physics

Protvino

General scheme of PBS

The Portable Batch System is a batch job and computer system resource management package. As such, it will accept batch jobs, a shell script and control attributes, preserve and protect the job until it is run, run the job, and deliver output back to the submitter.



Reasons for development of own scheduler

MAUI Lacks:

1. Absence of configured planning algorithms;
 2. Uses own planning algorithm with greater number of adjustment parameters;
 3. Absence of precise classification for tasks from the point of view of planning policy;
 4. Absence of dynamic CPU reservation;
 - 5 .Complex perception of influence of changeable parameters on a planning policy.
-
-

Variants of schedulers for PBS

Integrated in TORQUE Resource Manager

C based scheduler

BaSL scheduler

Tcl based scheduler

External scheduler

MAUI

Comparison of schedulers opportunities

Scheduler type	+	-
C based scheduler	Creation of own planning algorithm	? Dynamic planning script replacement impossibility
BaSL scheduler		? Impossibility to realize necessary planning algorithm by given means of language
Tcl based scheduler		? Standard algorithms described on C are complex for perception
		? Usage of additional programs to realize planning policy
		? Dynamic planning script replacement impossibility

TCL Scheduler components

initialization script - scheduler reads the initialization script once on start up and executes it. The given file is adjusting. It contains all variables and values which necessary for realization policy set of planning.;

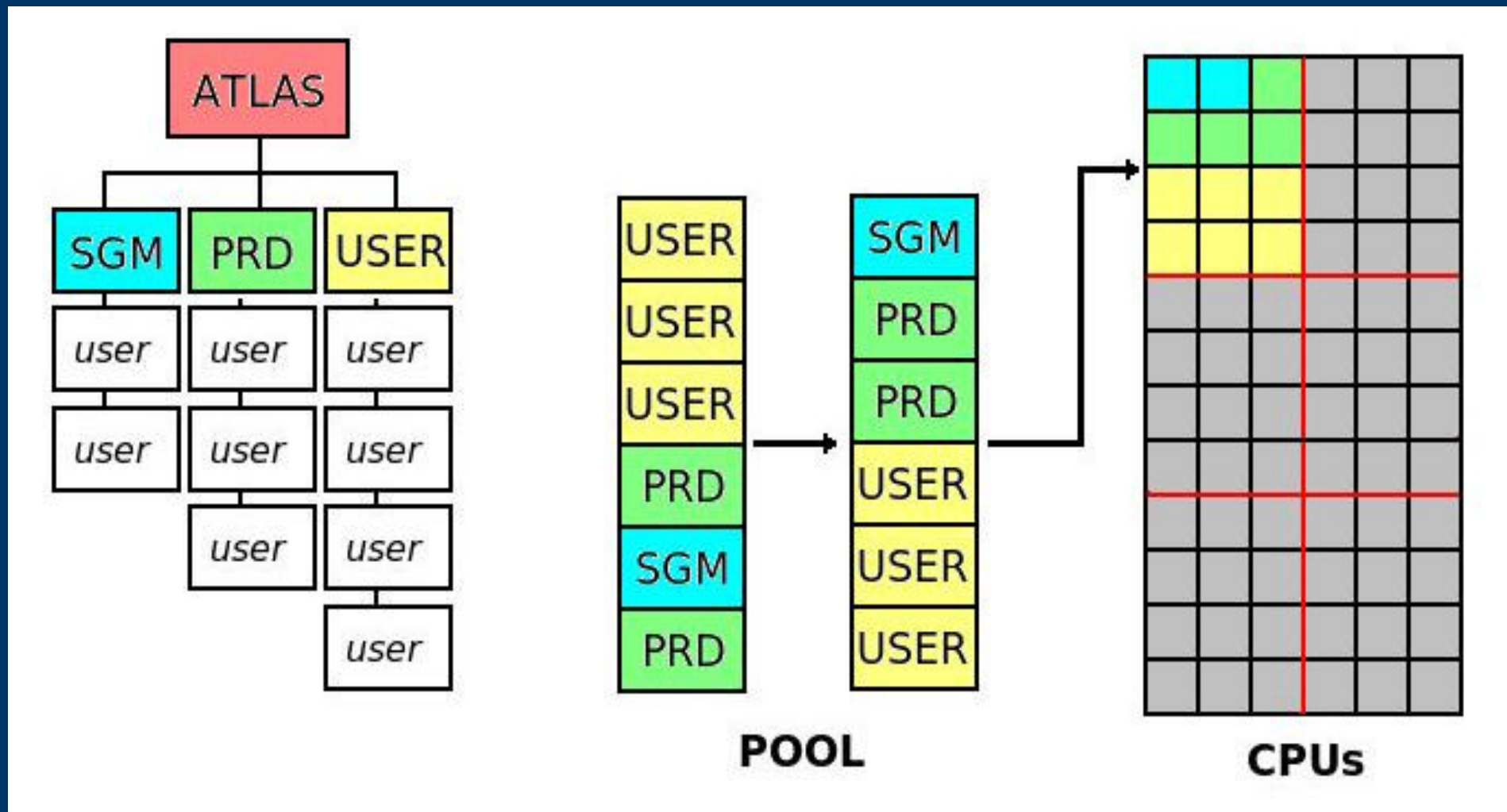
body script - this is the file that will be executed each time a “schedule” command is received from the Server. This file contains algorithm of planning task processing and sending THEM TO Server FOR execution;

log-file – contains all events about job execution for 24 hours.

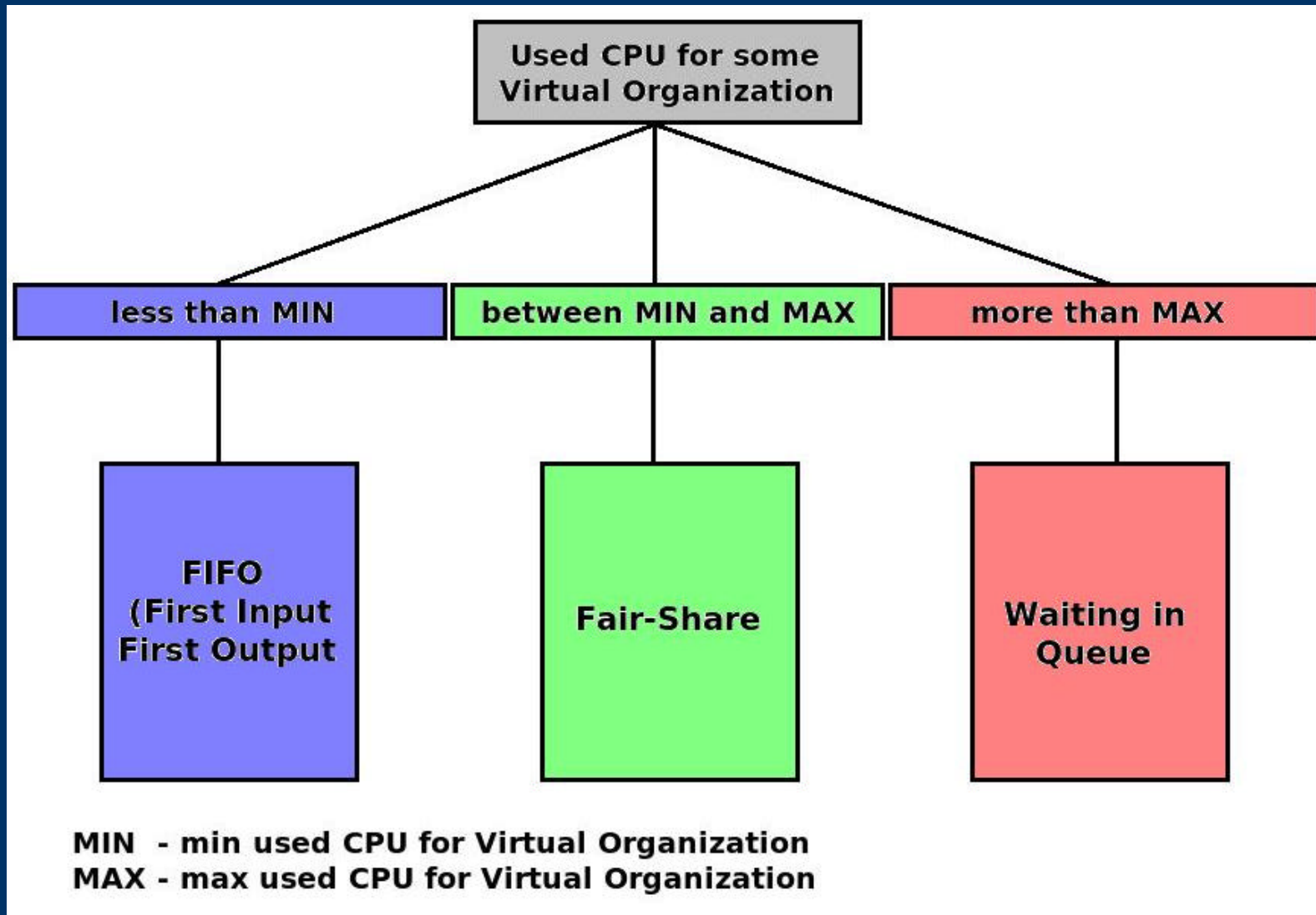
TCL Scheduler Policy

1. Dynamic CPU reservation for internal users
 2. Restriction of started tasks for:
 - users
 - groups
 - queues
 3. Opportunity of usage two types of planning for every Virtual Organization which dependence on system congestion:
 - FIFO (First Input First Output)?
 - Fairshare
 4. Opportunity of priority set up
 - between Virtual Organization
 - inside every Virtual Organization
 5. Log events for 24 hours and record them in file
-
-

Example of priority distribution inside Virtual Organization



Choice algorithm scheme



Tcl Scheduler in work

Server	Max	Tot	Que	Run	Hld	Wat	Trn	Ext	Status
ce0003.m45.ihep.	0	173	101	72	0	0	0	0	Active

Job id	Name	User	Time Use	S	Queue								
206897.ce0003	STDIN	alicesgm	03:53:35	R	alice	212945.ce0003	STDIN	cms026	0	Q	cms		
208551.ce0003	STDIN	alicesgm	03:11:10	R	alice	212944.ce0003	STDIN	cms026	0	Q	cms		
211250.ce0003	STDIN	atlas057	12:37:12	R	atlas	212946.ce0003	STDIN	cms026	0	Q	cms		
211795.ce0003	STDIN	alicesgm	07:51:26	R	alice	212947.ce0003	STDIN	cms051	0	Q	cms		
211801.ce0003	STDIN	alicesgm	09:15:22	R	alice	212949.ce0003	STDIN	cms026	0	Q	cms		
211803.ce0003	STDIN	alicesgm	09:18:50	R	alice	212950.ce0003	STDIN	cms026	0	Q	cms		
211808.ce0003	STDIN	alicesgm	07:53:08	R	alice	212951.ce0003	STDIN	cms051	0	Q	cms		
211815.ce0003	STDIN	alicesgm	08:08:34	R	alice	212953.ce0003	STDIN	cms051	0	Q	cms		
211817.ce0003	STDIN	alicesgm	07:57:33	R	alice	212954.ce0003	STDIN	cms026	0	Q	cms		
211879.ce0003	STDIN	alicesgm	07:15:25	R	alice	212955.ce0003	STDIN	cms051	0	Q	cms		
211881.ce0003	STDIN	alicesgm	07:19:49	R	alice	212956.ce0003	STDIN	cms026	0	Q	cms		
211883.ce0003	STDIN	alicesgm	07:09:53	R	alice	212958.ce0003	STDIN	cms026	0	Q	cms		
211888.ce0003	STDIN	alicesgm	07:19:46	R	alice	212959.ce0003	STDIN	cms051	0	Q	cms		
211892.ce0003	STDIN	alicesgm	07:01:16	R	alice	212960.ce0003	STDIN	cms026	0	Q	cms		
211899.ce0003	STDIN	alicesgm	07:13:31	R	alice	212962.ce0003	STDIN	cms026	0	Q	cms		
211910.ce0003	STDIN	alicesgm	06:05:13	R	alice	212963.ce0003	STDIN	cms051	0	Q	cms		
211915.ce0003	STDIN	alicesgm	06:14:29	R	alice	212964.ce0003	STDIN	cms051	0	Q	cms		
212018.ce0003	STDIN	alicesgm	06:30:19	R	alice	212965.ce0003	STDIN	cms051	0	Q	cms		
212023.ce0003	STDIN	alicesgm	06:23:38	R	alice	212966.ce0003	STDIN	cms026	0	Q	cms		
212026.ce0003	STDIN	alicesgm	06:18:51	R	alice	212967.ce0003	STDIN	cms051	0	Q	cms		
212031.ce0003	STDIN	alicesgm	06:22:02	R	alice	212968.ce0003	STDIN	cms026	0	Q	cms		
212131.ce0003	STDIN	alicesgm	04:36:14	R	alice	212970.ce0003	STDIN	cms051	0	Q	cms		
212135.ce0003	STDIN	alicesgm	05:15:04	R	alice	212971.ce0003	STDIN	cms026	0	Q	cms		
212137.ce0003	STDIN	alicesgm	04:41:37	R	alice	212972.ce0003	STDIN	cms051	0	Q	cms		
212139.ce0003	STDIN	alicesgm	04:23:52	R	alice	213102.ce0003	STDIN	cms051	0	Q	cms		
212140.ce0003	STDIN	alicesgm	05:19:40	R	alice	213111.ce0003	STDIN	cms026	0	Q	cms		
212141.ce0003	STDIN	alicesgm	05:23:54	R	alice	213119.ce0003	STDIN	cms051	0	Q	cms		
212145.ce0003	STDIN	alicesgm	05:21:55	R	alice	213135.ce0003	STDIN	cms051	0	Q	cms		
212151.ce0003	STDIN	alicesgm	04:28:55	R	alice	213152.ce0003	STDIN	cms026	0	Q	cms		
212153.ce0003	STDIN	alicesgm	05:14:03	R	alice	213153.ce0003	STDIN	cms026	0	Q	cms		
212154.ce0003	STDIN	alicesgm	04:31:09	R	alice	213218.ce0003	STDIN	cms051	0	Q	cms		
212555.ce0003	STDIN	alicesgm	04:25:25	R	alice	213229.ce0003	STDIN	cms051	0	Q	cms		
212556.ce0003	STDIN	alicesgm	04:33:12	R	alice	213357.ce0003	STDIN	cms051	0	Q	cms		
212576.ce0003	STDIN	alicesgm	05:21:56	R	alice	213406.ce0003	STDIN	cms026	0	Q	cms		
212586.ce0003	STDIN	alicesgm	05:16:04	R	alice	213456.ce0003	STDIN	cms026	0	Q	cms		
212588.ce0003	STDIN	alicesgm	04:20:38	R	alice	213479.ce0003	STDIN	cms051	0	Q	cms		
212782.ce0003	STDIN	cms051	04:43:04	R	cms	213645.ce0003	STDIN	alicesgm	0	Q	alice		
212787.ce0003	STDIN	cms051	03:50:39	R	cms	213813.ce0003	STDIN	cms026	0	Q	cms		
212791.ce0003	STDIN	cms051	03:48:48	R	cms	213817.ce0003	STDIN	cms026	0	Q	cms		
212794.ce0003	STDIN	cms051	03:50:54	R	cms	213865.ce0003	STDIN	cms026	0	Q	cms		
212797.ce0003	STDIN	cms051	03:50:08	R	cms	213893.ce0003	STDIN	cms026	0	Q	cms		
212801.ce0003	STDIN	cms051	03:43:52	R	cms	213980.ce0003	STDIN	cms026	0	Q	cms		
212803.ce0003	STDIN	cms051	03:41:26	R	cms	213990.ce0003	STDIN	cms026	0	Q	cms		
212808.ce0003	STDIN	cms051	03:30:45	R	cms	214079.ce0003	STDIN	cms026	0	Q	cms		
212813.ce0003	STDIN	cms051	03:21:09	R	cms	214433.ce0003	STDIN	prdcms05	0	Q	cms		
212816.ce0003	STDIN	cms051	03:20:16	R	cms	214455.ce0003	STDIN	alicesgm	0	Q	alice		
212819.ce0003	STDIN	cms051	03:18:09	R	cms	214456.ce0003	STDIN	alicesgm	0	Q	alice		
						214458.ce0003	STDIN	alicesgm	0	Q	alice		
						214459.ce0003	STDIN	alicesgm	0	Q	alice		
						214529.ce0003	STDIN	prdatl03	0	R	atlas		
						214530.ce0003	STDIN	prdatl03	0	R	atlas		

Summary

Now scheduler works in test mode on IHEP farm.
After trial period mistakes will be revealed and scheduler development will be continued. Now bugs are not discovered and it is not time to do any conclusions

Thanks for attention

