Monitoring system for OpenPBS environment

V.Kolosov, E.Lublev, S.Makarychev
ITEP, Moscow

OpenPBS batch system is widely used in HEP community. Standard tools from OpenPBS package allow to check the current status of the system. This information is useful, but it is not sufficient for resource accounting and planning. As a solution of this problem we developed the monitoring system which parse the logfiles from OpenPBS and store the information into SQL database (PostgreSQL). This allows to analyze the data in many different ways using SQL queries. The system was used in ITEP during last two years for monitoring of the batch farm.

1. Introduction

The applications in High Energy Physics require a lot of intensive calculations. The computing farms with the batch systems are common way to do that and they a used very wide at this time in the HEP community. One of the such batch system is OpenPBS. It is rather reliable, fast, and scalable enough to be used in the large computing centers.

Standard tools of OpenPBS give us a possibility to control the jobs in the system, but they did not allow to monitor the statistics for a long term in a convenient way. It is possible to view the log files of OpenPBS to analyze some faults, but it is a real when you try to extract an integrated statistics for a long period grouped by users, by working group or by executing nodes. The simplest solution of this problem is the using of the SQL database for the OpenPBS log storage. We describe below one of the possible implementation which was developed in ITEP in 2001.

2. Monitoring system scheme

In the OpenPBS we have the interaction between the users, the OpenPBS server including scheduler and the sets of the working nodes. Physical users can have several accounts or usernames and can be a members of different working groups. Each working group can also have several system groups (analogs of the user account). All these system elements should be implemented in the database structure.

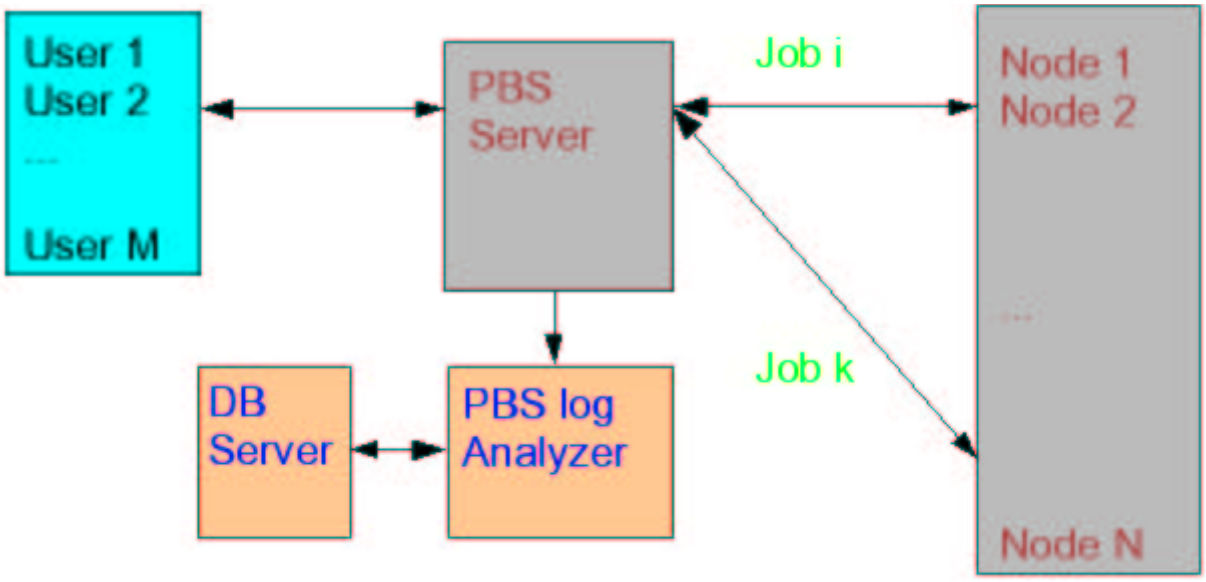


Fig 1. Scheme of the OpenPBS monitoring system.

In addition to common system structure we should include the tables describing the OpenPBS logfiles. As a result, after the normalization of the SQL tables we will have the following database structure.

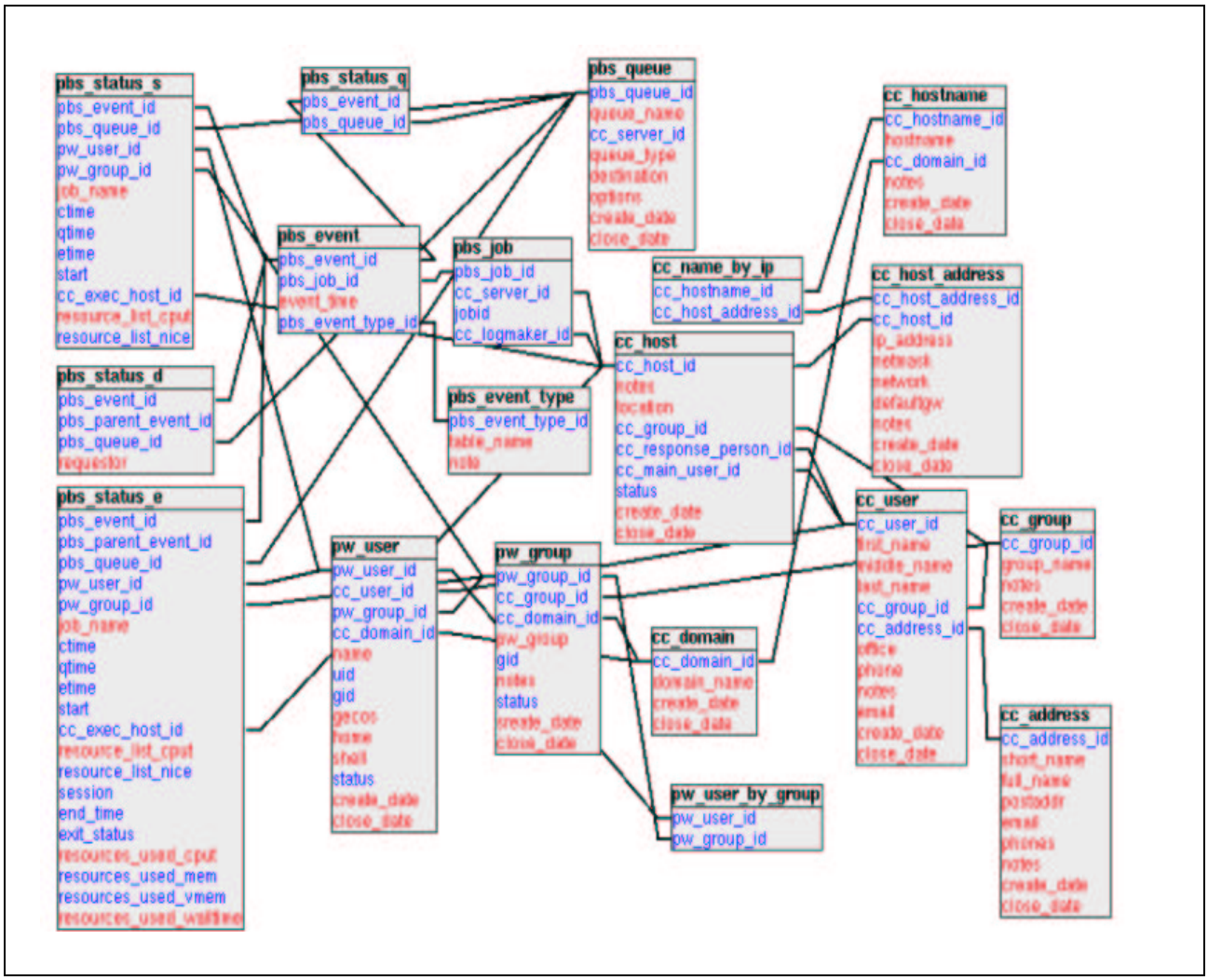


Fig 2. The scheme of the SQL database for the OpenPBS monitoring system.

For the monitoring system we keep in the mind the possibility to operate under the heavy load in parallel for few PBS farms. As a result of the search for the solid open source solution for this task we decide to use PostgreSQL database. Using this database and DBI interface module for perl the PBS log parser was written (for server_priv/accounting/). DBI allows us to be independent on choice for SQL database for log storage. One significant feature of this parser is the possibility to use the transaction mechanism for writing the logs from several PBS cluster into one database.

Some stress tests were done and it was shown that it is possible provide the monitoring up to tenths thousand jobs per day. Some serious bugs were fixed during two years of using this monitoring system but there were no significant crashes with data corruption. Some architectural improvements allow us to improve this rate up to few millions events using the code separation for log parser and log writer (they will be implemented as separate daemons).

The log writer is an essential part of the monitoring system but tools for presentation are required. The second version of the interface for WWW was written in the summer of 2003. It is combine the parser for standard PBS tools like “qstat” for presentation of the current state of the cluster and the interface for the database with PBS logs for presentation of the long term statistics .

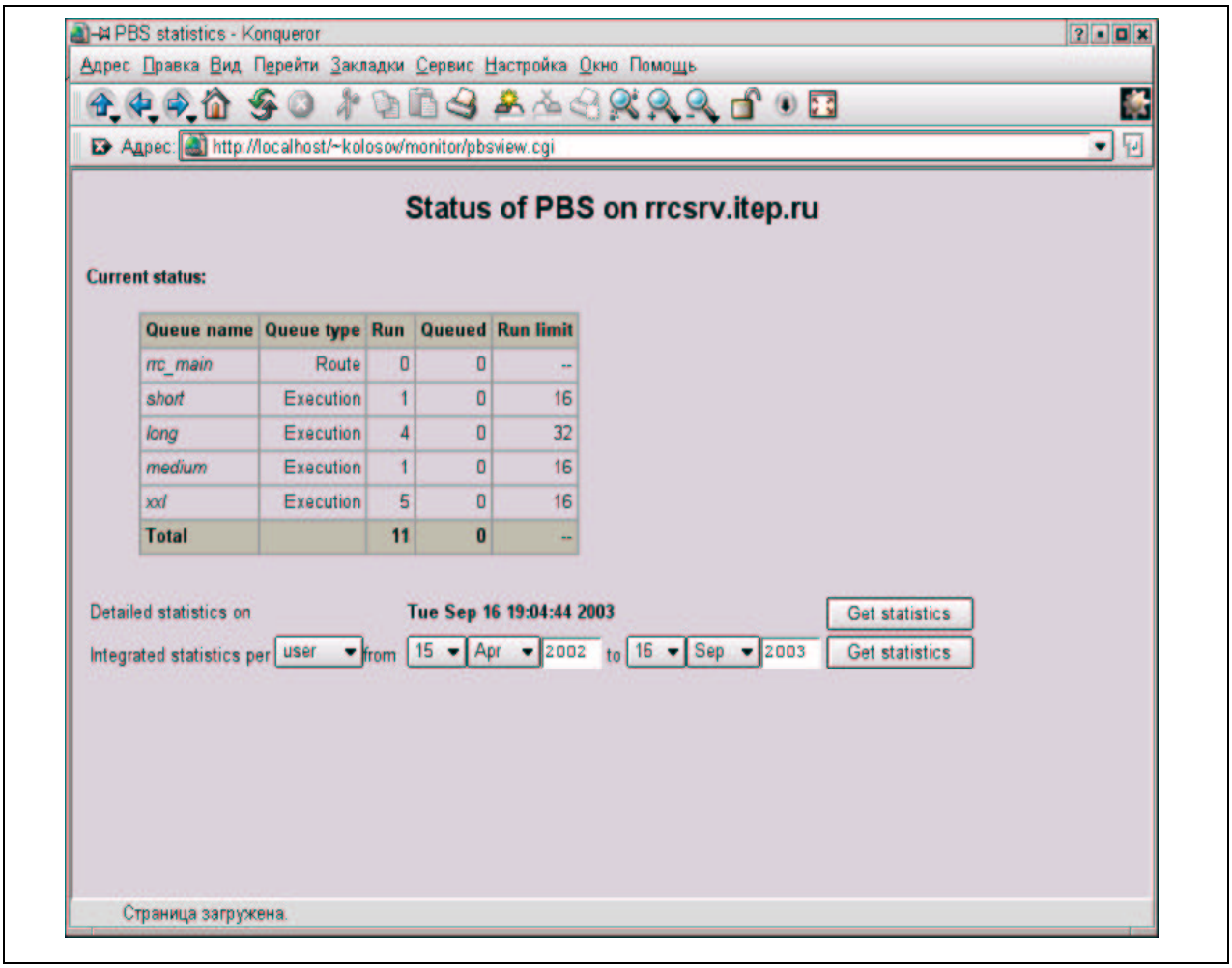
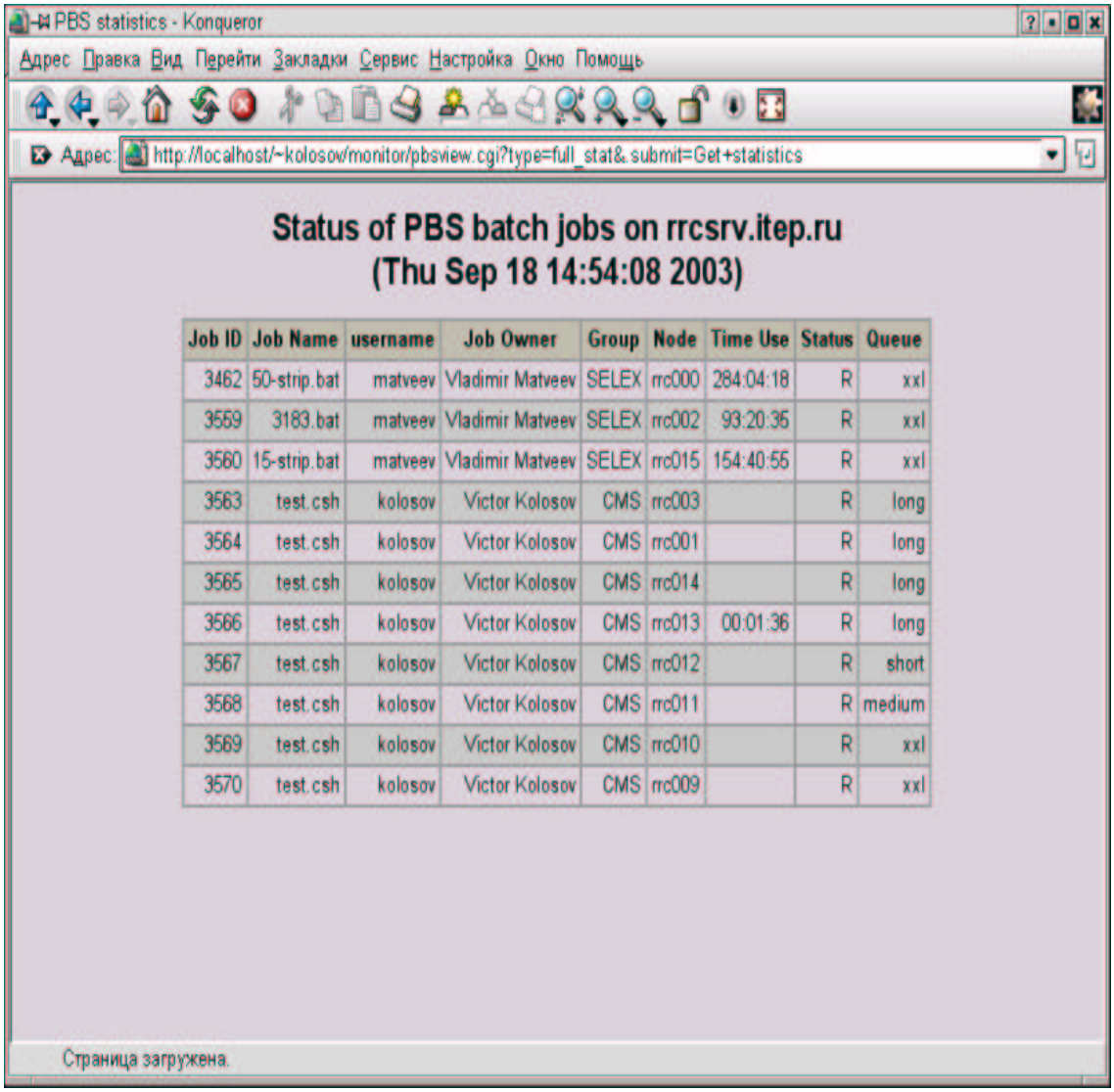


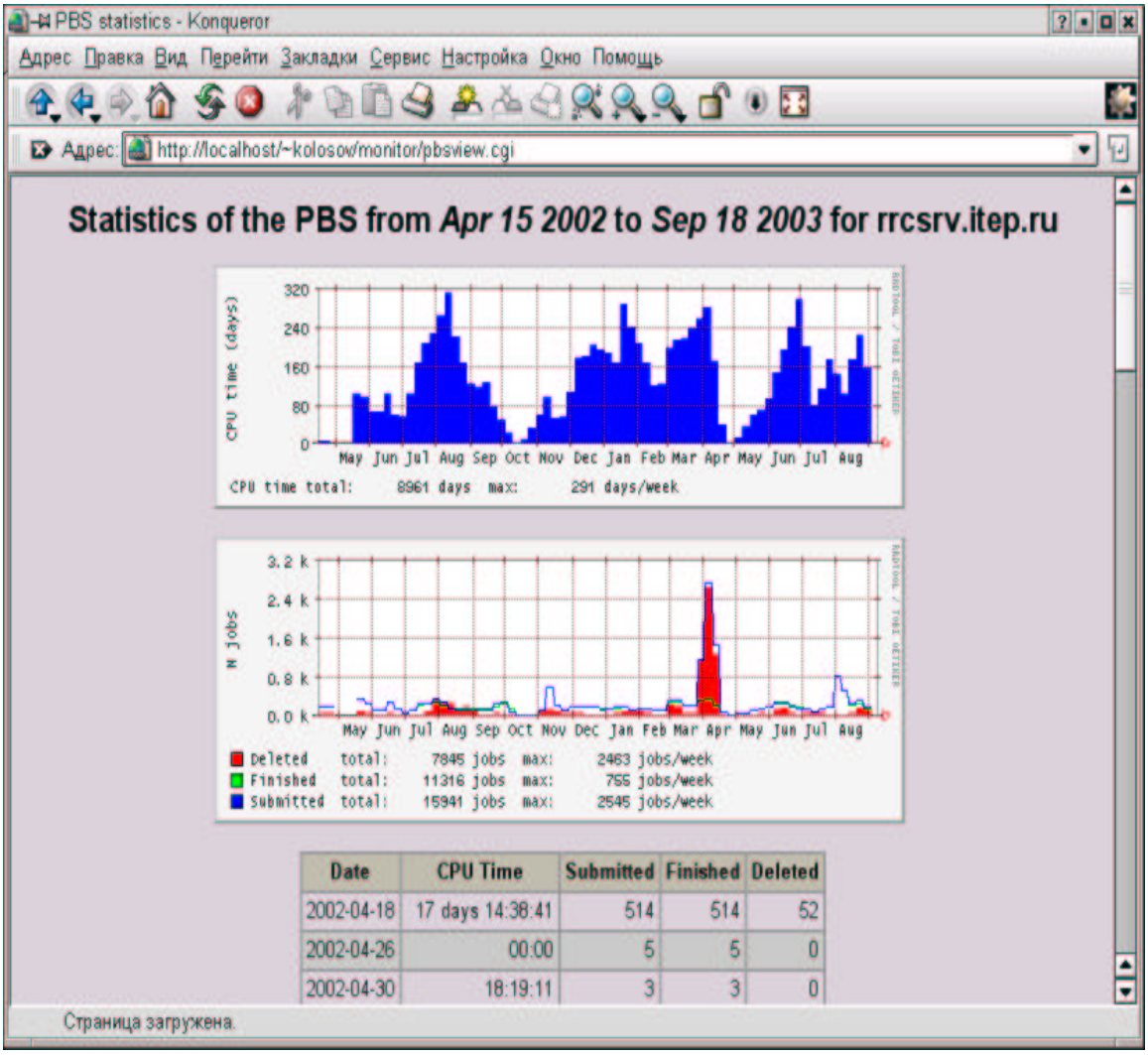
Fig 3. The main page for Web interface.



Status of PBS batch jobs on rrcsrv.itep.ru (Thu Sep 18 14:54:08 2003)

Job ID	Job Name	username	Job Owner	Group	Node	Time Use	Status	Queue
3462	50-strip.bat	matveev	Vladimir Matveev	SELEX	rrc000	284:04:18	R	xxl
3559	3183.bat	matveev	Vladimir Matveev	SELEX	rrc002	93:20:35	R	xxl
3560	15-strip.bat	matveev	Vladimir Matveev	SELEX	rrc015	154:40:55	R	xxl
3563	test.csh	kolosov	Victor Kolosov	CMS	rrc003		R	long
3564	test.csh	kolosov	Victor Kolosov	CMS	rrc001		R	long
3565	test.csh	kolosov	Victor Kolosov	CMS	rrc014		R	long
3566	test.csh	kolosov	Victor Kolosov	CMS	rrc013	00:01:36	R	long
3567	test.csh	kolosov	Victor Kolosov	CMS	rrc012		R	short
3568	test.csh	kolosov	Victor Kolosov	CMS	rrc011		R	medium
3569	test.csh	kolosov	Victor Kolosov	CMS	rrc010		R	xxl
3570	test.csh	kolosov	Victor Kolosov	CMS	rrc009		R	xxl

Fig 4. Full statistics window for OpenPBS.



Date	CPU Time	Submitted	Finished	Deleted
2002-04-18	17 days 14:38:41	514	514	52
2002-04-26	00:00	5	5	0
2002-04-30	18:19:11	3	3	0

Fig 5. Integrated statistics for long period.

3. Conclusions

The first and the main result of our work is creation of the system which can operate at least two year without serious problem and which can allow significant database redesign without changing a lot of code. In addition we find a way for improvement the performance of the monitoring system. This solution can be used for the developing of the high performance monitoring system for common purpose because the existing SQL solution provides better log rate than the existing file based monitoring systems like MRTG or RRDtool.

4. Acknowledgments

This project was carried out with financial support from INTAS (grant 00-00440 INTAS CERN).

5. The Bibliography

[1] http://www.openpbs.org/
[2] ANSI X3.135-1992 (R1998), Information Systems – Database Languages SQL, and ISO/IEC 9075:1992, Information technology - Database Languages SQL
[3] Practical PostgreSQL, John C. Worsley, Joshua D. Drake, O'Reilly & Associates, Paperback, Bk&CD edition, Published January 2002