

# New implementation of RTM access to L&B data: status and migration plans

Aleš Křenek<sup>1</sup>, František Dvořák<sup>1</sup>, Miloš Mulač<sup>1</sup>, Jiří Sitera<sup>1</sup>,  
Janusz Martyniak<sup>2</sup>, David Colling<sup>2</sup>

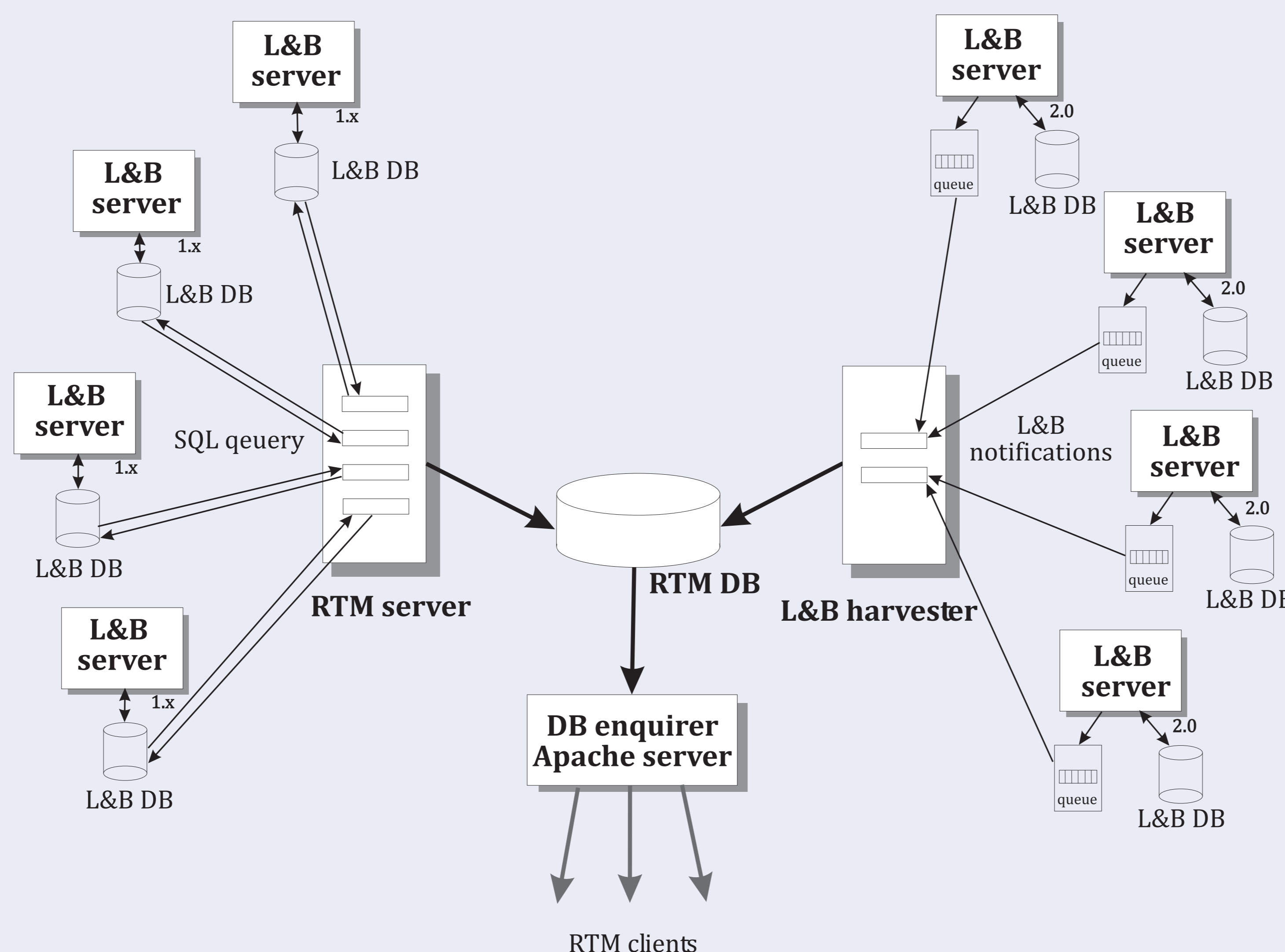
<sup>1</sup>CESNET, CZ, <sup>2</sup>Imperial College, UK

## Overview

Current implementation of RTM access to L&B data evolved gradually for several years, it reaches its scalability limits, and it also exhibits several known problems. Recently we decided to address these issue by reimplementing the L&B access module from scratch. We describe its design, introduce a working prototype, and discuss plans on smooth migration.

## Current Drawbacks

- Architecture of one thread per L&B **does not scale** well.
- Direct access to L&B database is not public interface, it yields **inaccurate results** and with next major L&B release (2.0) it would **not work** anymore.
- Queries are suboptimal, possibly generating unsustainable load on L&B and **blocking WMS/L&B** operation.
- Granting RTM the access to L&B databases is **unnecessary escalation of privilege** – the table-level grant give access to much more information than RTM really uses.



## New Architecture

Replacement RTM module, **L&B harvester**, was developed from scratch, based on **L&B notifications**.

- The module subscribes to **receive messages** on job state changes from all registered L&B's. The mechanism is reliable wrt. connectivity outage, and machine or software component crash. In the rare case of either side going down for longer than notification expiration (several hours), the harvester catches up with **bootstrap query**.

- The messages are processed and stored into **internal RTM database**.
- Data are picked there by other **unchanged** RTM components, and eg. used to draw maps of geographic job distribution as before. Therefore both implementations can co-exist during migration from L&B 1.9 to 2.0.

The new architecture addresses all the previous issues:

- One harvester thread accepts messages from multiple L&B's. Limiting factor is the **total number of incoming messages**, not number of connections.
- A message on a specific job is generated only on job state change. There are **no more massive queries** to L&B.
- Access to L&B uses its stable **public interface** only, not depending on changes in L&B internals.
- Authentication uses X509 identity of the harvester, leveraging **L&B authorization mechanisms**, including fine-grain control on access to job state fields (planned to be released in 2009).

## Performance

Performance measurements were focused on the new mechanism, starting with queued message at L&B server, and ending with preparation of SQL insert/update to RTM database. Tests were run between two quad-core machines at 2.5 GHz, sending 18309 messages (6 per job, 110 MB file size) in each thread.

threads	streams	time (s)	msgs/s	Mjobs/day
1	1	118	155	2.23
2	2	126	291	4.19
4	4	150	488	7.03
1	4	123	595	8.57

CPU load on both machines was low during the tests. The highest per-stream number (155 msgs/s) corresponds to network saturation (6 ms RTT). The last row shows that L&B harvester performs even better with traffic from multiple sources aggregated into single thread.

Altogether the results demonstrate that L&B message transport and the harvester are unlikely to become central bottleneck of the whole system.

## Main Achievements

- **Job state fully synchronized between L&B and RTM**
- **Reduced L&B load by information push**
- **AuthN/Z with X509 instead of DB password**
- **Promising performance**
- **Better scaling due to "1 thread : many L&B's" architecture**
- **Smooth migration from L&B 1.9 to L&B 2.0**

## References

- L&B home page and documentation <http://egee.cesnet.cz/en/JRA1/LB/>
- RTM home page <http://gridportal.hep.ph.ic.ac.uk/rtm/>



The EGEE project is building a Grid infrastructure for the scientific community. Grids are networks of computers spread across many sites but able to act together to provide a range of large scale facilities, from incredible processing power and mass storage to a platform for international collaboration.