NMDB:
Real-Time database for high resolution Neutron Monitor measurements

Grant Agreement Number 213007

Combination of Collaborative Project and Coordination and Support Action

Documentation of the database

C. T. Steigies, O. M. Rother

Kiel, Germany
30.12.2008

Document information

Project	
Project acronym:	NMDB
Project full title:	Real-Time database for high resolution Neutron Monitor measurements
Grant agreement no.:	213007
Funding scheme:	Combination of Collaborative Projects & Coordination and Support Actions
Project start date:	01.01.2008
Project duration:	24 months
Call topic:	INFRA-2007-1.2.1 Scientific Digital Repositories
Project web-site:	www.nmdb.eu
Document	
Deliverable number:	2.4
Deliverable title:	Database Documentation
Due date of deliverable:	month 12
Actual submission date:	30.12.08
Editors:	C. T. Steigies
Authors:	C. T. Steigies, O. M. Rother
Reviewers:	project consortium
Participating beneficiaries:	CAU, IZMIRAN, NKUA
Work Package no.:	2
Work Package title:	Architectural Design
Work Package leader:	CAU
Work Package participants:	CAU, IZMIRAN, NKUA,
Estimated person-months for deliverable:	1.5
Dissemination level:	PU
Nature:	report
Version:	30.12.2008
Draft/Final:	Final
No of pages (including cover):	14
Keywords:	database

Contents

Introduction

The lead beneficiary for the deliverables "2.1 Database prototype" and "2.2 Database documentation preliminary 1" was originally the work package leader of WP2, BIRA-IASB. Since BIRA experienced manpower problems and could not work on these deliverables, they agreed to withdraw from the project on May 15, 2008. The NMDB database is now hosted at CAU, and at the project meeting on September 13, 2008, it was formally decided by the consortium that CAU becomes the leader of WP2. After extensive discussions about the database form from May to July 2008, the format of the preliminary database was finalized and the prototype is online since July 8, 2008. The tables of the prototype database have been described on our website at `http://www.nmdb.eu/?q=node/59`.
During the midterm meeting in December, 2008 in Kiel, no changes to the database format were requested and no performance problems with the database, except a few network outages, were noted. Thus the database format as implemented by CAU in the summer of 2008, is the final format for NMDB. The layout of the different tables is described in this document.

1 Background

As documented in the report on BIRA's work on WP2, BIRA's subcontractor Sputnikweb has performed some work related to the NMDB database. The have studied the following subjects:

- Choice of the technology to be used
- Benchmarking the MySQL setup
- How to create a replication node
- Current database structure

BIRA has studied the two main open source database solutions, PostgreSQL and MySQL, and came to the conclusion that MySQL is most suitable for NMDB. The benchmarking tests with two servers have shown no speed problems when replicating small amounts of data. The database structure as defined by BIRA had to been changed significantly to be suitable for the needs of NMDB. The MySQL setup with two master nodes as tested by Sputnikweb is not suitable by NMDB, since NMDB has planned to use three or more database master servers. Moving from two to three or more master servers is not trivial, since MySQL5.0 does not support Multimaster Replication [1]. There are several strategies to achieve a similar setup, also MySQL5.1 which has been released in December 2008 has additional support for multi-master replication [2]. However, before any of these solutions can be used in a distributed setup with servers in Kiel, Athens, and Moscow, detailed tests still have to be performed to select the most suitable solution for NMDB.

2 Prototype database

A prototype database has been set up in Kiel using MySQL on a Debian GNU/Linux machine (`db01.nmdb.eu`). We are working to set up a mirror of the database in Kiel, which will be accessible as `db02.nmdb.eu`. Further mirrors will also be made available under the `nmdb.eu` domain.
A shell script has been developed to setup the tables that are needed when a new station joins NMDB. As of December 30th, 2008, we have tables for 20 stations in NMDB, see table 2.1.

short name	**full name**	**country**
AATB	Alma-Ata	Kazakhstan
APTY	Apatity	Russia
ATHN	Athens	Greece
ERV3	Erevan3	Armenia
ERV	Erevan	Armenia
ESOI	Tel Aviv	Israel
JUNG1	Jungfraujoch1	Switzerland
JUNG	Jungfraujoch	Switzerland
KERG	Kerguelen	France
KIEL	Kiel	Germany
LMKS	Lomnický štít	Slovakia
MCRL	Moscow test	Russia
MGDN	Magadan	Russia
MOSC	Moscow	Russia
NRLK	Norilsk	Russia
NVBK	Novosibirsk	Russia
OULU	Oulu	Finland
ROME	Rome	Italy
TERA	Terre Adelie	France
YKTK	Yakutsk	Russia

Table 2.1: List of stations submitting data to NMDB as of December 30, 2008.

The setup script creates five tables per station, one table for hourly data (STATION_1h), one table for real-time high resolution data (STATION_ori), one table for revised high resolution data (STATION_rev), one table for environmental data (STATION_env), and one table for meta data (STATION_meta), with columns as defined in the following tables. The script also creates a user and password for the new station, which is then

sent to the operators of the station. These credentials are needed to insert data into the tables for this station, no other user has permissions to write data to these tables. In WP4 several user tools have been presented that submit data to these tables in NMDB using these credentials.
All times and dates recorded in NMDB are in GMT. One DATETIME format which is accepted by MySQL is `YYYY-MM-DD hh:mm:ss`. All counts are given as counts per second for the whole monitor. Where not noted otherwise, SI units are used in NMDB. All pressures are given in mbar. All temperatures are given in ℃. All velocities are given in m/s. STATION stands for one of the stations from table 2.1.

3 Hourly data

Hourly data is stored in the table STATION_1h. This table contains the best available hourly data, i.e. previous data is overwritten, in case errors in the data are corrected. The countrates stored in this table are hourly averages, the values represent counts per second for the whole detector.

column	format	comment
start_date_time	DATETIME	start date and time of the 3600 s count interval
uncorrected	FLOAT	uncorrected count rates, counts per second
corr_for_efficiency	FLOAT	count rates corrected for pressure and efficiency
corr_for_pressure	FLOAT	count rates corrected for pressure
pressure_mbar	FLOAT	atmospheric pressure in millibar
last_change	TIMESTAMP	automatically generated timestamp

4 Original data

High resolution data that is recorded in real-time is stored in the table STATION_ori. This table contains data in the highest available time resolution and data is stored as measured, no data is ever overwritten. The countrates in this table represent counts per second for the whole detector.

column	format	comment
start_date_time	DATETIME	start date and time of the data collection interval
length_time_interval_s	FLOAT	duration of the data collection interval in seconds
measured_uncorrected	FLOAT	uncorrected count rates, counts per second
measured_corr_for_efficiency	FLOAT	count rates corrected for pressure and efficiency
measured_corr_for_pressure	FLOAT	count rates corrected for pressure
measured_pressure_mbar	FLOAT	atmospheric pressure in millibar

5 Revised data

Revised high resolution data is stored in the table STATION_rev. The format of this table is equal to STATION_ori, but is only used if a data set in STATION_ori needs to be revised. Only pieces of revised data shall be submitted here. The table contains an automatically updated timestamp field "last_change", as well as a two digit integer "version" field.

column	**format**	**comment**
start_date_time	DATETIME	start date and time of the data collection interval
length_time_interval_s	FLOAT	duration of the data collection interval in seconds
revised_uncorrected	FLOAT	uncorrected count rates, counts per second
revised_corr_for_efficiency	FLOAT	count rates corrected for pressure and efficiency
revised_corr_for_pressure	FLOAT	count rates corrected for pressure
revised_pressure_mbar	FLOAT	atmospheric pressure in millibar
version	INTEGER	version number of the data
last_change	TIMESTAMP	date and time of the last change of the data

6 Environmental data

Optional environmental data such as temperature, wind speed and humidity is stored in the table STATION_env. The table also contains an automatically updated timestamp field "last_change".

column	format	comment
date_time	DATETIME	start date and time of the 3600 s count interval (This must be identical to the date and time in the _ori and _rev tables.)
measured_temperature_c	FLOAT	air temperature in degrees Celsius
measured_relative_humidity	FLOAT	relative humidity (%)
measured_wind_speed_m_s	FLOAT	wind speed in m/s
last_change	TIMESTAMP	timestamp of the entry

7 Meta data

Optional "human-generated" electronic log files of the station are stored in the table STATION_meta. This table also contains an automatically updated timestamp field "last_change".

column	format	comment
start_date_time	DATETIME	start of time period
end_date_time	DATETIME	end of time period
quality_flag	INTEGER	quality flag as per GLE standard
comment_environment_condition	TEXT	log entry
comment_revision	TEXT	revision of log entry
last_change	TIMESTAMP	timestamp of the entry

8 Station information

In addition to the station specific tables, NMDB contains one table with information from all stations (`NM_station_information`), describing the stations and their equipment. This table is filled in by NKUA based on the query they generated in deliverable 3.1 and which has been filled in by all participants. The query contains questions about the following topics: general information, station information, station electronics information, acquisition system, and local server information. For every change at a station, for example change of the station manager, change in the registration electronics, or change in the number of counter tubes, a new entry with the new information is created in this table. Since every entry also contains a start and an end date, the complete station history is recorded in this table. Details about the contents of this table are given in the report to deliverable 3.1 and will not be repeated here.

Bibliography

[1] Baron Schwartz, Peter Zaitsev, Vadim Tkachenko, Jeremy D. Zawodny, Arjen Lentz, and Derek J. Balling. *High Performance MySQL*. O'Reilly, 2nd edition, 2008.

[2] Multi-master and circular replication. http://dev.mysql.com/doc/refman/5.1/en/mysql-cluster-replication-multi-master.html.