

EGU2020-13637

<https://doi.org/10.5194/egusphere-egu2020-13637>

EGU General Assembly 2020

© Author(s) 2020. This work is distributed under the Creative Commons Attribution 4.0 License.



## Performance analysis and optimization of a TByte-scale atmospheric observation database

**Clara Betancourt**, Sabine Schröder, Björn Hagemeier, and Martin Schultz

Jülich Supercomputing Centre, Jülich Research Centre, Jülich, Germany

The Tropospheric Ozone Assessment Report (TOAR) created one of the world's largest databases for near-surface air quality measurements. More than 150 users from 35 countries have accessed TOAR data via a graphical web interface (<https://join.fz-juelich.de>) or a REST API (<https://join.fz-juelich.de/services/rest/surfacedata/>) and downloaded station information and aggregated statistics of ozone and associated variables. All statistics are calculated online from the hourly data that are stored in the database to allow for maximum user flexibility (it is possible, for example, to specify the minimum data capture criterion that shall be used in the aggregation). Thus, it is of paramount importance to measure and, if necessary, optimize the performance of the database and of the web services, which are connected to it. In this work, two aspects of the TOAR database service infrastructure are investigated: Performance enhancements by database tuning and the implementation of flux-based ozone metrics, which – unlike the already existing concentration based metrics – require meteorological data and embedded modeling.

The TOAR database is a PostgreSQL V10 relational database hosted on a virtual machine, connected to the JOIN web server. In the current set-up the web services trigger SQL queries and the resulting raw data are transferred on demand to the JOIN server and processed locally to derive the requested statistical quantities. We tested the following measures to increase the database performance: optimal definition of indices, server-side programming in PL/pgSQL and PL/Python, on-line aggregation to avoid transfer of large data, and query enhancement by the explain-analyze tool of PostgreSQL. Through a combination of the above mentioned techniques, the performance of JOIN can be improved in a range of 20 - 70 %.

Flux-based ozone metrics are necessary for an accurate quantification of ozone damage on vegetation. In contrast to the already available concentration based metrics of ozone, they require the input of meteorological and soil data, as well as a consistent parametrization of vegetation growing seasons and the inclusion of a stomatal flux model. Embedding this model with the TOAR database will make a global assessment of stomatal ozone fluxes possible for the first time ever. This requires new query patterns, which need to merge several variables onto a consistent time axis, as well as more elaborate calculations, which are presently coded in FORTRAN.

The presentation will present the results from the performance tuning and discuss the pros and cons of various ways how the ozone flux calculations can be implemented.

