In-Memory Analytics with EXASOL and KNIME //

Dr. Marcus Dill

Analytics 2020

The volume and complexity of data today and in the future poses great challenges for IT systems. At the same time, real-time analyses are expected to be better, tools are expected to be more intuitive, and data quality is expected to be higher. Static reporting is out, and modern companies offer their users self-service business intelligence and support them with smart algorithms. Conventional IT architectures and classic technologies can't cope with such demands.

Analytics in the 21st century requires top performance in data processing and evaluation, analysis and forecasting techniques that users in business departments can handle, and modern front ends where correlations can be visualized comprehensibly and interactively. No single software provider today can offer top-quality tools in all these areas, and least of all at an acceptable price. If you don't want to compromise in terms of quality and costs, you choose best-of-breed technologies. Fortunately, integrating such technologies isn't the hurdle it used to be, as the solution developed by mayato for in-memory analytics with EXASOL and KNIME shows.

EXASOL and KNIME – each one a champion, and together a world-class duo

KNIME Analytics Platform offers a mature integrated open source package for modern data analysis and data mining: All the functions needed meet the requirements of major analytics projects are available, from comprehensive data import and data preparation features to the sophisticated modeling and subsequent export of the results. Its easy-to-use graphical user interface with logic similar to the established commercial suites puts KNIME Analytics Platform ahead of open-source programming environments such as R or Python. Furthermore, KNIME's high functionality is comparable to that of commercial suites – at much lower procurement and operating costs. KNIME also has numerous connectors, for example, to Hadoop, and is easy to integrate with modern BI and big data architectures thanks to a range of extensibility features.
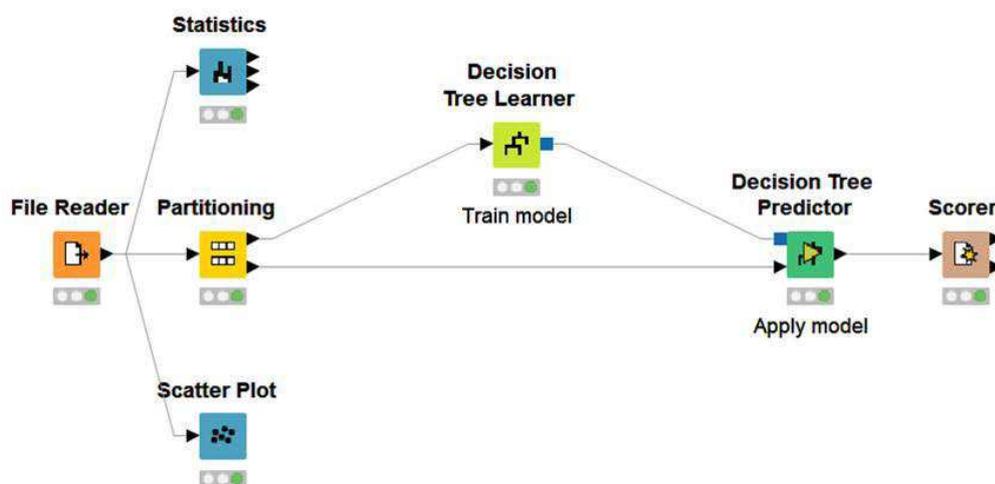
EXASOL is the undisputed technology leader in analytical databases. It is regularly rated best in benchmark tests and thoroughly deserves its title as the fastest database in the world. The high performance of data processing and queries is achieved through in-memory technology and impressive

parallelization capabilities, even on very large multi-node installations. EXASOL's position ahead of the competition is mainly due to the fact that it has exceptionally intelligent techniques to constantly optimize the use of the working memory – that is, the decision about what data is stored in-memory when and on what computer – and to adapt the data and its use. This means the analytical EXASOL database is not only always ultra-speedy, but it also works with practically no manual intervention. The fastest database in the world is therefore also the lowest maintenance.

EXASOL enables the integration of, for instance, R, Python, and Java within the parallel runtime architecture, making complex analyses and forecasts possible with hitherto unknown speed. Even programming languages such as R, Python, or JAVA have to be classified as expert-only technologies because of their lack of full-fledged integrated user interfaces, which are virtually impossible to use without programming knowledge. KNIME Analytics Platform, in contrast, has an intuitive user interface. Users don't need to know how to code to create data mining models and complex analyses. And the results are visualized as charts in an appealing and easy-to-interpret way.

Even straight out of the box, KNIME Analytics Platform can access data within EXASOL for complex data analysis and can take models developed in such a way and add more data, for example, for forecasts. Typical application examples here are:

- Existing customers can be divided into different segments based on their behavior and other data, and then addressed with more targeted, more cost-effective, and more successful marketing measures.
- For customer acquisition, models to forecast purchase probabilities can be created. In this case too, focused campaigns are possible.
- Companion sales can be analyzed and used to generate product suggestions in customer interaction.
- Cases of fraud in banking can be examined on the basis of historical contracts to identify typical characteristics, subsequently enabling statements to be made about fraud probability even for new business.
- Typical constellations in machine data can be correlated with subsequent downtimes, increases in consumption, and so on. This knowledge can then be used as an early warning system.
- Similarly, suspicious patterns can be identified in large quantities of network data which could indicate intruders, malware, or simply technical problems in individual IT components. Such indications can be investigated specifically and major damage can be avoided.

As a result of these diverse application scenarios, numerous opportunities emerge to increase sales, reduce costs, or minimize risks using such approaches.

mayato is confronted with new issues for analyses and forecasts practically every day. Again and again, we prove the potential of analytics in a wide variety of business cases.

mayato In-Memory Analytics Project Accelerator (mayato ConKNEX)

To fully exploit the benefits of KNIME in combination with EXASOL, practice has shown that some enhancements are necessary in the specific system landscape, which mayato provides as a complete package known as *In-Memory Analytics Project Accelerator* (mayato ConKNEX). mayato ConKNEX contains the following functions:

- In the standard edition, KNIME Analytics Platform can only access EXASOL via JDBC. This means that only some of the advantages of the parallel in-memory architecture of the EXASOL database can be used. With the runtime integration of mayato ConKNEX, KNIME models can be deployed directly in the database and can run there within the EXASOL platform. The fastest database in the world serves as an additional turbo for the intelligent features in KNIME Analytics Platform.
- KNIME Analytics Platform has a graphical modeling environment that enables the development and running of integrated analysis processes. Accesses to databases defined within this interface are also possible in connection with Exasol. So that some features of EXASOL can be better deployed, mayato ConKNEX includes additional components for access to EXASOL. Performance is thus increased for the reading and writing of data using these components.
- Furthermore, mayato ConKNEX uses specific EXASOL metadata that is not specifically available through the standard functions in KNIME.

mayato ConKNEX can therefore significantly reduce development effort and model runtimes. Both database and data warehouse processes in EXASOL and analysis processes in KNIME Analytics Platform can be integrated so that data can be refined or used for evaluations right after its arrival in the system.

Excellence in analytics!

mayato ConKNEX takes your analytical models straight to your data

In addition, mayato ConKNEX enables you to execute the code generated in an external analysis tool efficiently in the analytical EXASOL database with UDFs (user-defined functions) developed by mayato. You

can then create complex models in an intuitive environment, which you can execute right in your database – so you don't have to import your data again for analysis purposes.

KNIME Analytics Platform's intuitive user interfaces make data analysis interactive and user-friendly. Through the combination of modules for the modeling, data preparation, analysis, and visualization, complex analyses can be performed and clearly visualized without in-depth knowledge of IT. Click and drag-and-drop principles simplify the creation and execution of analyses in daily business tremendously.

The architecture we selected is specially designed for processing large quantities of data. Because the exploration of the models takes place right on the EXASOL in-memory database, the response time is much shorter, meaning that business intelligence and analytics applications can help you make the right business decisions faster.

Our enhancement can be seamlessly integrated with your existing infrastructure. With it, midsize businesses in particular are ideally placed to:

› Establish a high-performance basis to answer their company-specific analytical questions

› Execute algorithms directly in the database

› Combine compact data mining in large data quantities with machine learning

› Extend the existing infrastructure with analytical models

As a result, all algorithms can work massively, parallelized, and directly in the cluster. This procedure leads to fast interaction with operational data and empowers you to perform diverse evaluations, which you can execute completely on the database thanks to the high performance.

Our analytics enhancement enables you to:

✓ Increase performance enormously by using in-memory data analysis

✓ Benefit from an affordable alternative to complex data warehouse analyses with high administrative effort

✓ Visualize and analyze your data, giving you new insights

✓ Gain self-service access to the relevant information, even for business departments with no IT expertise

Contact us //

Want to know more? Do you have any questions about our products and services?

mayato GmbH
Am Borsigturm 9
13507 Berlin

info@mayato.com