



Was wir vorhersagen, soll auch eintreffen!



In-Memory-Analytics mit Exasol und KNIME //

Dr. Marcus Dill



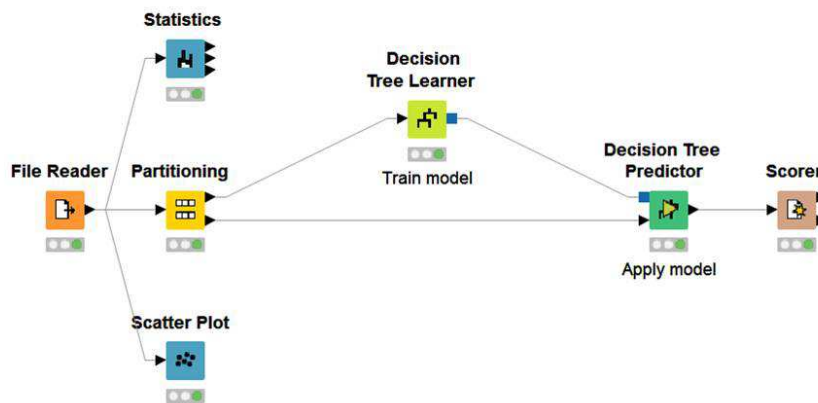
Analytics 2020

Volumen und Komplexität heutiger und zukünftiger Daten fordern IT-Systeme. Gleichzeitig steigt der Anspruch an Echtzeit-Analysen, intuitive Bedienbarkeit der verwendeten Tools und die Qualität der Daten. Statisches Reporting ist out, moderne Unternehmen bieten ihren Anwendern Self-Service BI und die Unterstützung durch intelligente Algorithmen. Herkömmliche IT-Architekturen und klassische Technologien sind durch all dies überfordert.

Analytics im 21. Jahrhundert erfordert höchste Performance bei Verarbeitung und Auswertung der Daten, für Fachanwender bedienbare Analyse- und Prognoseverfahren und moderne Frontends zur verständlichen, interaktiven Visualisierung von Zusammenhängen. Kein Softwareanbieter kann heute in allen diesen Bereichen Werkzeuge auf höchstem Niveau bieten, schon gar nicht zu einem vertretbaren Preis. Wer keine Kompromisse hinsichtlich Qualität und Kosten machen will, der setzt auf Best-of-Breed-Technologien. Dass deren Integration heute kein Problem mehr darstellt, zeigt die von mayato entwickelte Lösung für In-Memory-Analytics mit Exasol und KNIME.

EXASOL und KNIME – jeder ein Champion, gemeinsam ein Weltklasse-Duo

KNIME bietet mit seiner Analytics Plattform ein ausgereiftes Open-Source-Komplettpaket für moderne Datenanalyse und Data Mining: von umfangreichen Möglichkeiten zum Datenimport und der Datenvorverarbeitung bis hin zur ausgefeilten Modellierung und anschließendem Export der Ergebnisse stehen alle Funktionalitäten zur Verfügung, die auch den Anforderungen umfassender Analytics-Projekte genügen. Open-Source-Programmierungsumgebungen wie R oder Python hat KNIME eine leicht bedienbare grafische Nutzeroberfläche voraus, deren Logik sich an die der bekannten kommerziellen Suites von SAS oder SPSS anlehnt. Im Vergleich zu diesen wiederum besticht KNIME durch eine vergleichbar hohe Funktionalität – zu deutlich geringeren Anschaffungs- und Betriebskosten. Darüber hinaus verfügt KNIME über zahlreiche Konnektoren, z.B. zu Hadoop, und kann dank zahlreicher Erweiterungsmöglichkeiten leicht in moderne BI- und Big-Data-Architekturen eingebunden werden.





Exasol ist klarer Technologieführer für analytische Datenbanken. Regelmäßig beweist sie sich in entsprechenden Benchmarks als Testsieger und führt zurecht den Titel der schnellsten Datenbank der Welt. Die hohe Performance von Datenverarbeitung und Abfragen wird grundsätzlich durch In-Memory-Technologie und ein hohes Maß an Parallelisierbarkeit auf auch sehr großen Mehrknotenarchitekturen erzielt. Dabei ist Exasol der Konkurrenz vor allem dadurch voraus, dass außergewöhnlich intelligente Verfahren die Nutzung des Arbeitsspeichers, d.h. die Entscheidung, welche Daten wann auf welchem Rechner „in memory“ gehalten werden, ständig optimiert und den jeweiligen Daten und deren Verwendung anpasst. So ist Exasol nicht nur immer ultraschnell, sondern es ist dies auch praktisch ohne manuelle Eingriffe. Die schnellste Datenbank der Welt ist also gleichzeitig auch die wartungsärmste.

Exasol erlaubt die Integration von R und Python innerhalb der parallelen Laufzeitarchitektur und damit komplexe Analysen und Prognosen mit nie gekannter Geschwindigkeit. Sowohl R als auch Python müssen aber mangels vollwertiger integrierter Benutzeroberflächen als Expertentechnologien gelten, die ohne Programmierkenntnisse kaum einsetzbar sind. Im Gegensatz hierzu bietet KNIME ein intuitives User Interface. Anwender können also auch ohne Coding Data-Mining-Modelle und komplexe Auswertungen erstellen. Sie erhalten die Ergebnisse in ansprechender Weise grafisch aufbereitet.

KNIME kann bereits out of the box bei der komplexen Datenanalyse auf Daten innerhalb von Exasol zugreifen und auf diese Weise entwickelte Modelle mit weiteren Daten beispielsweise für Prognosen zur Anwendung bringen. Typische Anwendungsfälle hierfür sind:

- Bestandskunden können anhand ihres Verhaltens und weiterer Daten in unterschiedliche Segmente eingeteilt und anschließend mit gezielteren, kostengünstigeren und erfolgreicherem Marketingmaßnahmen angesprochen werden.
- Bei der Neukundenakquise können Modelle zur Prognose von Kaufwahrscheinlichkeiten erstellt werden. Auch in diesem Fall sind fokussiertere Aktionen möglich.
- Verbundkäufe können analysiert und in Form von Produktvorschlägen in der Kundeninteraktion genutzt werden.
- Betrugsfälle im Kreditwesen können auf der Basis von historischen Verträgen auf typische Charakteristika hin untersucht werden, was im Anschluss auch bei neuen Geschäften Aussagen über Betrugswahrscheinlichkeiten erlaubt.
- Typische Konstellationen innerhalb von Maschinendaten lassen sich mit späteren Ausfällen, Verbrauchssteigerungen etc. korrelieren. Dieses Wissen kann im Anschluss als Frühwarnsystem dienen.
- Analog lassen sich in großen Mengen von Netzwerkdaten verdächtige Muster identifizieren, die auf Eindringlinge, Schadsoftware oder schlicht technische Probleme einzelner IT-Komponenten hinweisen können. Solchen Hinweisen kann gezielt nachgegangen und größere Schäden vermieden werden.

Der Fantasie, wie mit solchen und ähnlichen Ansätzen Geld verdient, Kosten gesenkt und Risiken gemindert werden können, sind keine Grenzen gesetzt.

mayato wird praktisch täglich mit neuen Fragestellungen für Analysen und Prognosen konfrontiert. Immer wieder erbringt mayato den Nachweis für das Potential von Analytics in den unterschiedlichsten Business Cases.



Financial
Analytics

Customer
Analytics

Industry
Analytics

Security
Analytics

mayato In-Memory-Analytics Project Accelerator (mayato ConKNEX)

Damit die Vorteile von KNIME in Kombination mit Exasol voll zum Tragen kommen, bedarf es in der Praxis jedoch noch einiger kleiner Erweiterungen in der konkreten Systemlandschaft, die mayato als Paket *In-Memory Analytics Project Accelerator (mayato ConKNEX)* gebündelt zur Verfügung stellt. **mayato ConKNEX** beinhaltet die folgenden Funktionen:

- KNIME kann auf Exasol im Standard nur über ODBC zugreifen. Damit werden die Vorteile der parallelen In-Memory-Architektur der Exasol-Datenbank nur teilweise genutzt. Mit der Runtime Integration von **mayato ConKNEX** können KNIME-Modelle direkt in die Datenbank deployed werden und dort innerhalb der Exasol-Plattform laufen. Die schnellste Datenbank der Welt dient den intelligenten Verfahren aus KNIME als zusätzlicher Turbo.
- KNIME verfügt über eine grafische Modellierungsumgebung, die Entwicklung und Ablauf von integrierten Analyseprozessen ermöglicht. Zugriffe auf Datenbanken, die innerhalb dieser Oberfläche definiert werden, sind auch in Verbindung mit Exasol möglich. Damit einige Besonderheiten von Exasol besser genutzt werden können, beinhaltet **mayato ConKNEX** zusätzliche Komponenten zum Zugang zu Exasol. Das Lesen und Schreiben von Daten über diese Komponenten gewinnt damit an zusätzlicher Performance.
- Weiterhin nutzt **mayato ConKNEX** spezifische Metadaten von Exasol, die über die Standardfunktionen in KNIME nicht speziell verfügbar sind.

mayato ConKNEX erlaubt somit die signifikante Reduktion von Entwicklungsaufwänden und Modell-Laufzeiten. Datenbank- und Data-Warehouse-Prozesse in Exasol einerseits und Analyseprozesse in KNIME andererseits lassen sich so miteinander integrieren, dass Daten direkt nach ihrem Eintreffen veredelt bzw. für Auswertungen verwendet werden können.

Analytics auf allerhöchstem Niveau!





mayato ConKNEX bringt Ihre analytischen Modelle direkt zu Ihren Daten

mayato ConKNEX ermöglicht es Ihnen zusätzlich, den in einem externen Analyse-Tool generierten Code effizient in Exasol über von mayato entwickelte UDF (User Defined Functions) auszuführen. Somit erstellen Sie komplexe Modelle in einer intuitiven Umgebung, die Sie direkt in Ihrer Datenbank ausführen. Sie sparen sich damit das erneute Einlesen Ihrer Daten für Analysezwecke.

Intuitive Benutzeroberflächen von KNIME ermöglichen eine benutzerfreundliche interaktive Datenanalyse. Durch die Kombination von Modulen für die Modellierung, Datenvorverarbeitung, Analyse und Visualisierung können komplexe Analysen ohne umfangreiches IT-Wissen durchgeführt und übersichtlich visualisiert werden. Klick- sowie Drag-und-Drop-Prinzipien vereinfachen die Erstellung und Durchführung von Analysen im Daily Business enorm.

Die von uns gewählte Architektur ist speziell für die Verarbeitung großer Datenmengen ausgelegt. Die Exploration der Modelle direkt auf der in-Memory-Datenbank von EXASOL bietet dabei eine erheblich verkürzte Antwortzeit, sodass Business Intelligence und Analytics-Anwendungen schneller Ihre unternehmerischen Entscheidungen unterstützen können.

Unsere Erweiterung kann nahtlos in Ihre bestehende Infrastruktur integriert werden. Mit ihr gewinnen vor allem mittelständische Unternehmen die richtige Ausgangslage, um

- › hoch performante Analytics-Grundlagen für die unternehmensspezifischen analytischen Fragestellungen zu schaffen
- › Algorithmen direkt auf der Datenbank auszuführen
- › kompaktes Data Mining auf großen Datenmengen mit maschinellem Lernen zu verbinden
- › die bestehende Infrastruktur um analytische Modelle zu erweitern

Als Resultat können sämtliche Algorithmen massiv parallel direkt im Cluster arbeiten. Diese Vorgehensweise führt zu einer schnellen Interaktion mit operativen Daten und ermöglicht Ihnen vielfältige Auswertungen, welche Sie aufgrund der hohen Performance vollständig auf der Datenbank umsetzen können.

Zusammengefasst ermöglicht Ihnen unsere Analytics-Erweiterung eine

- ✓ enorme Performance-Steigerung durch die jetzt zur Verfügung stehenden Möglichkeit der Datenanalyse in-Memory
- ✓ kostengünstige Alternative zu komplexen Data Warehouse Analysen mit meist hohem Administrationsaufwand
- ✓ Visualisierung und Analyse Ihrer Daten und somit neue Einsichten
- ✓ Self-Service-Zugang zu den relevanten Informationen, auch für Fachabteilungen ohne IT-KnowHow



Kontaktieren Sie uns //

Sie möchten mehr erfahren? Sie haben Fragen zu unserem Angebot?

mayato GmbH
Am Borsigturm 9
13507 Berlin

info@mayato.com

+49 / 30 4174 4270 0