

Massive Beschleunigung von Datenanalysen

Riesige Datenmengen aus unzähligen Quellen und in unterschiedlicher Granularität zeitnah zu analysieren ist eine der Kernkompetenzen von IMS Health, dem führendem Marktforschungsunternehmen für die Pharma- und Gesundheitsindustrie. Der Anbieter entschied sich vor drei Jahren für ein hochskalierbares Datenbankmanagement-System, das die Performance zur Erstellung der Analysen einzelner Datenbestände um bis zu Faktor 80 steigern konnte.

Die Herausforderung

Zur Erschließung, Aufbereitung und analytischer Nutzung neuer Datenquellen wurde auf lokaler Ebene ein leistungsstarkes System benötigt, um große Datenmengen in ein Data Warehouse zu laden, zeitnah zu analysieren und ad hoc in entsprechenden Strukturen auszugeben. Darüber hinaus musste das eingesetzte System hochskalierbar sein, denn es kamen neue Datenlieferanten hinzu, was einherging mit ständig steigenden Datenvolumina und eine Integration verschiedener Datenquellen erforderte. Um diese Aufgaben professionell zu meistern, setzte IMS Health auf ein breites Produktportfolio mit verschiedenen Hardware, Datenbanken und Client-Server-Applikationen. Den speziellen, lokalen Bedürfnissen Rechnung tragend ergab sich die Notwendigkeit, ein alternatives System zu finden, auf dem Prototypen für neue, innovative Projekte und Produkte entwickelt werden konnten, ohne jedoch die Kosten massiv in die Höhe zu treiben.

Die Lösung

Als IMS Health vor drei Jahren die In-Memory-Cluster-Datenbanktechnologie von der EXASOL AG aus Nürnberg testete, erhielt die deutsche Firma die

Möglichkeit, die Leistung ihres Produkts EXASolution mit einem Proof-of-Concept unter Beweis zu stellen. Die Lösung brachte eine um bis zu Faktor 80 höhere Performance als die vorhandenen Client-Server-Systeme. Die Einführung der Lösung verlief reibungslos und nach etwa sechs Wochen stand das System für die erste produktive Nutzung bereit. Der Nürnberger Anbieter führte zunächst einen dreiwöchigen Stresstest auf einem Datenvolumen durch, das mit den Anforderungen der Marktforscher vergleichbar war, und setzte dann bei IMS Health den Cluster mit der Datenbanklösung auf. Auf Wunsch von IMS Health implementierte der Datenbankanbieter zwei wichtige Features: Es wurde eine zusätzliche Funktion im ETL- (Extraction Transformation Loading) Bereich integriert, die es ermöglicht, Dateien im FBV-Format in die Datenbank performant zu laden. Darüber hinaus wollte IMS Health in der Lage sein, eigene Client-Applikationen zu programmieren und über native Schnittstellen auf die Datenbank zuzugreifen. Die Bereitstellung eines Client-SDK (Software Development Kit) machte diese Anforderung möglich.

Das Data-Warehouse-System bestand zunächst aus fünf Servern mit Intel Xeon-CPU's, von denen einer als Stand-by für die Ausfallsicherheit sorgte. Die Prozesse der Befüllung konnte der Projektleiter bei IMS Health mit einem dreiköpfigen Team nach einer Schulung von EXASOL binnen sechs Wochen selbst durchführen. Das Team definierte die Strukturen für die Daten, konfigurierte das Sicherheitskonzept und stellte die Integration in die IT-Infrastruktur des Rechenzentrums sicher. Sehr wichtig war das Aufsetzen der Validierungsprozesse, das heißt die Implementierung der Regeln, nach denen die Daten validiert bzw. aufbereitet werden, bevor sie aus dem System extrahiert und für die Analysen und weiterführenden Systeme bereitgestellt werden. Zuletzt musste das Team noch die verschiedenen Ab-

fragen, Reports, Extrakte und Datenauschnitte programmieren, die in der Produktion verwendet werden. Bestimmte Prozesse und Analysen verlangen den Aufruf von kompletten Datenbeständen in der Größe von mehreren hundert Gigabyte, was sehr hohe Ansprüche an die Leistungsfähigkeit des Systems stellt. Deshalb musste das neue System die Nutzer in die Lage versetzen, die Performance Ihrer Prozesse und Programme signifikant zu erhöhen und dabei die Kosten niedrig zu halten.

Vorteile auf einen Blick:

- Leistungsstarkes Datenmanagement für vielschichtige Ad-hoc Abfragen
- Echtzeit-Auswertungen auf unbegrenzten Datenmengen
- Skalierbarkeit durch Hinzunahme neuer Server bei steigenden Anforderungen oder größerer Datenmenge
- Reduzierte Betriebskosten durch einfache Integration und Administration

Kontakt

EXASOL

EXASOL AG
Olga Mendscheritski
Neumeyerstr. 48
90411 Nürnberg
Tel.: 0911/23991267
Email: olga.Mendscheritski@exasol.com