Online analytical processing with a cluster of databases

Author(s):
Röhm, Uwe

Publication Date:
2002

Permanent Link:

https://doi.org/10.3929/ethz-a-004448138 →

Diss. ETH No. 14 591

Online Analytical Processing with a Cluster of Databases

DISSERTATION

submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of

Doctor of Technical Sciences

presented by

UWE RÖHM

Diplom-Informatiker Univ., Universität Passau
citizen of Germany
born October 22, 1969

Prof. Dr. H.-J. Schek, examiner
Prof. Dr. M. Scholl, co-examiner

2002

Abstract

This dissertation investigates central architectural issues and performance aspects of clusters of databases for usage with online analytical processing. The objective is to develop a scalable infrastructure for on-line decision support systems which is capable of analyzing up-to-date data. This thesis pursues an approach with a *coordination middleware* on top of a *cluster of databases*. An important design principle is its component-oriented nature, so that the cluster can easily be extended with new hardware or software components. The challenge with such an infrastructure is to build it such that it provides scalability, performance, and correctness at the same time.

In the beginning, we concentrate on a query-only scenario and on possibilities to improve query performance by appropriate *query routing*. By replicating the data over all cluster nodes, several queries can be independently evaluated in parallel on different nodes. The task of the coordination middleware is to route each query to an appropriate cluster node. We develop a suite of query routing algorithms over encapsulated components. They approximate the contents of caches of the cluster nodes by means of recently evaluated queries. They use these approximations in order to route queries to the cluster nodes promising the fastest execution due to caching effects.

As a next step, we take updates into account, too. *Replication management* for a large cluster is still an open problem, especially as our intention is to provide scalability, correctness, and up-to-date guarantees at the same time. We develop a new method to replication management, which combines the principles of the open-nested layered transaction model with asynchronous replication management. It is founded in a formal framework, so that it can be proven to be correct. The resulting protocol behaves like synchronous replication to clients. They are guaranteed to access consistent, up-to-date data. At the same time, the protocol has the performance characteristics of lazy, asynchronous replication.

We then further refine this method and introduce a freshness-of-data-driven approach to replication management in a cluster of databases. The idea is to allow users to trade freshness of data for query performance. Clients submit their queries together with a *freshness limit* as new quality of service parameter. The result is *freshness-aware scheduling*. It makes use of the different degrees of freshness of the OLAP nodes in order to serve such queries which agree to access less fresh data sooner than queries asking for the latest data.

Finally, we are also interested in the performance characteristics of the presented algorithms. We have implemented the proposed routing and scheduling algorithms in a complete prototype of a coordination middleware. We report on the results of an extensive experimental evaluation using this prototype with the TPC-R benchmark for online analytical processing. It shows that cache approximation query routing can

significantly improve query performance as compared to non cache-aware routers, and that freshness-aware scheduling effectively allows users to trade freshness of data for faster query response time. Another nice result is that the coordination middleware does not become a bottleneck, even if providing clients access to up-to-date data in a cluster of databases with 128 nodes.

In summary, this thesis presents new approaches to query routing, replication management, and multi-version scheduling for online analytical processing in a cluster of databases. The approaches are founded on a formal theoretical background, implemented in a full-fledged prototype, and proven to be practicable by means of an extensive experimental evaluation.

Zusammenfassung

Diese Dissertation untersucht zentrale Aspekte von *Datenbankclustern* und ihrer Leistungsfähigkeit bei der Verwendung für Online Analytical Processing. Das Ziel ist eine skalierbare Infrastruktur für online Decision Support Systeme, die insbesondere Benutzern die Analyse aktueller Daten erlaubt. Die Arbeit verfolgt dabei einen Ansatz, der auf einer *Koordinationsmiddleware* basiert. Ein wichtiges Prinzip bei der Entwicklung des Systems ist Komponentenorientierung, so dass ein Cluster mittels neuer Hardware- und Softwarekomponenten einfach erweiterbar ist. Die Herausforderung ist dabei, ein System zu entwickeln, das sowohl Skalierbarkeit und Leistungsfähigkeit, als auch transaktionelle Garantien in sich vereint.

Wir betrachten dazu zunächst den einfachen Fall mit rein lesenden Zugriffen ohne Datenänderungen und konzentrieren uns auf die Leistungssteigerung mittels geeigneter *Query Routing* Verfahren. Indem man Daten über alle Clusterknoten repliziert, wird es möglich, mehrere Analyseanfragen unabhängig voneinander und parallel auf verschiedenen Knoten auszuführen. Die Frage ist allerdings, welches der am besten geeignete Knoten für die Ausführung einer gegebenen Anfrage ist. Dazu stellen wir eine Reihe von neuen Query Routing Algorithmen für den Einsatz über eingekapselten Komponenten vor. Diese Verfahren approximieren den Inhalt der Datenbankpuffer in den Clusterknoten aufgrund der zuletzt dort ausgeführten Anfragen. Sie verwenden diese Approximationen, um Anfragen zu solchen Knoten zu schicken, die eine besonders schnelle Ausführung durch bereits gepufferte Daten erwarten lassen.

Als nächstes betrachten wir zusätzlich auch Datenänderungen. Replikationsverwaltung für einen grossen Cluster ist nach wie vor ein offenes Problem, insbesondere, da unser Ziel ein Verfahren ist, dass gleichzeitig Skalierbarkeit, Korrektheit und Zugriff auf die zuletzt geänderten Daten garantiert. Wir entwickeln dazu ein neues *Replikationsverfahren*, dass in sich Prinzipien eines offen-geschachtelten Transaktionsmodels mit asynchroner Replikationskontrolle vereint. Die Grundlage dazu bildet ein formales Gerüst, das auch einen Korrektheitsbeweis ermöglicht. Das neu entwickelte Replikationsverfahren garantiert allen Clients den Zugriff auf konsistente und aktuelle Daten. Es verhält sich dahingehend wie synchrone Replikation. Gleichzeitig hat das Verfahren aber die Leistungscharakteristiken von asynchroner Replikation.

In einem weiteren Schritt verfeinern wir dieses Replikationsverfahren dahingehend, dass zudem der *Frischegrad der Daten* variabel wird. Die Idee ist, Benutzern zu erlauben, Datenfrische gegen schnellere Antwortzeiten einzutauschen. Zu diesem Zweck führen wir eine Mindestfrische als neuen Quality-of-Service Parameter für Anfragen ein. Dies ermöglicht „Frische-basiertes" Scheduling, das die verschiedenen Frischegrade der Clusterknoten verwendet, um Anfragen, die mit weniger frischen Daten zufrieden sind, früher zu bearbeiten als solche, die nach aktuellen Daten fragen.

Abschliessend sind wir auch an den Leistungscharateristiken der neu entwickelten Verfahren interessiert. Dazu haben wir die vorgestellten Routing und Scheduling Algorithmen prototypisch implementiert. Wir präsentieren die Resultate einer umfangreichen Evaluierung dieses Prototypen mit dem TPC-R Benchmark für Online Analytical Processing. Es zeigt sich, dass approximatives Query Routing im Vergleich zu einfacheren Verfahren die Antwortzeiten deutlich beschleunigen kann. Weiterhin ermöglicht Frische-basiertes Scheduling Benutzern in der Tat, effektiv Datenfrische gegen Anfragegeschwindigkeit einzutauschen. Ein weiteres interessantes Ergebnis ist, dass die koordinierende Middleware keinen Flaschenhals bildet, selbst dann nicht, wenn alle Clients in einem Cluster mit 128 Knoten nach aktuellen Daten fragen.

Fassen wir noch einmal kurz zusammen: In dieser Dissertation werden neue Verfahren zum Query Routing, zur Replikationskontrolle und zu Mehrversionen-Scheduling für Online Analytical Processing in einem Datenbankcluster vorgestellt. Die Verfahren sind sowohl auf einer formalen Grundlage aufgebaut, als auch vollständig in einem lauffähigen Prototypen implementiert. Mit umfangreichen experimentellen Evaluation wurde abschliessend ihre praktische Verwendbarkeit gezeigt.