

Link: <https://www.computerwoche.de/a/quo-vadis-data-warehousing,2508971>

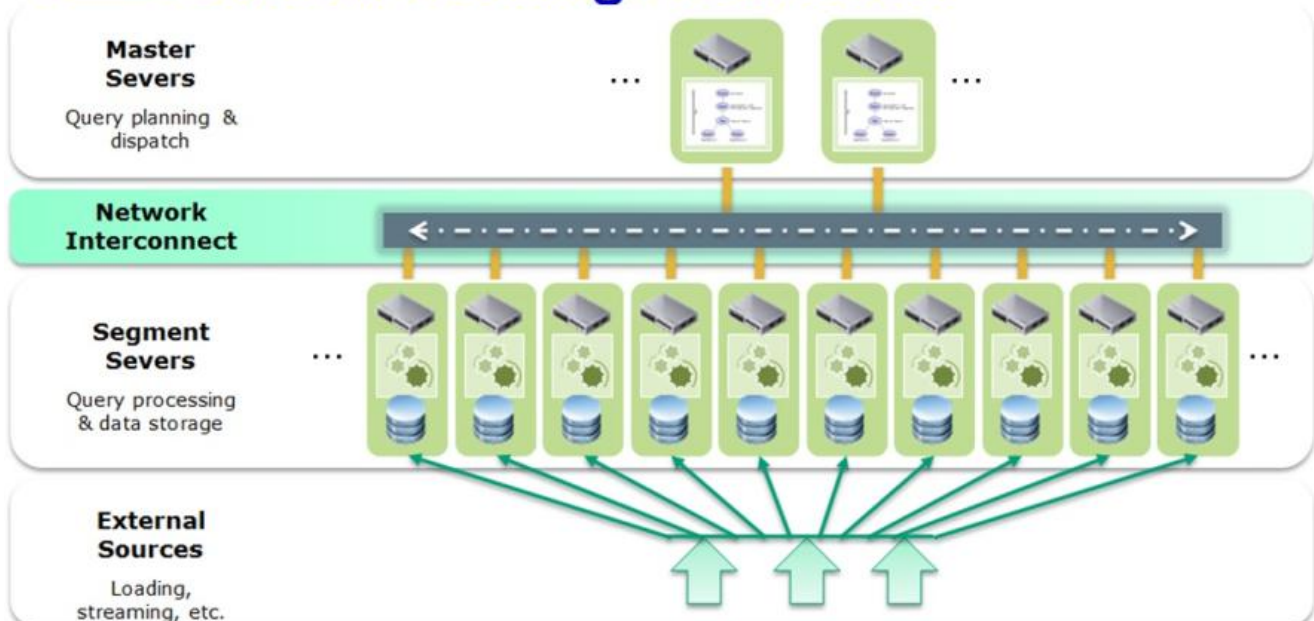
Massive Parallelverarbeitung bei Big Data

Quo vadis, Data Warehousing?

Datum: 10.04.2012
 Autor(en): Klaus Manhart

Big Data, unstrukturierte Daten und immer ausgefeiltere BI-Analysen in immer kürzeren Zeiten lassen herkömmliche Data Warehouses alt aussehen. Massive Parallelverarbeitung könnte die zukunftsweisende Technologie für das Data Warehousing sein.

Shared Nothing Architektur



MPP-Architekturen: In der Shared-Nothing-Architektur erhält jeder Knoten einen Teil der Gesamtdaten, die er mit seinen eigenen Ressourcen bearbeitet.

Foto: Data Mart

Data-Warehouse-Systeme (DWH) sind in vielen Unternehmen in die Jahre gekommen. Technische und ökonomische Schwächen haben Performance-Probleme, höhere Betriebskosten und immer aufwendigere Entwicklungsprozesse zur Folge. Spätestens mit der Verarbeitung großer Massendaten - Stichwort Big Data - und dem damit einhergehenden Zuwachs an Analysebedarf sind herkömmliche DWHs überfordert.

Viele BI- und Datenbankexperten sehen im Massive Parallel Processing (MPP) die zukunftsweisende Technologie für das Data Warehousing, heißt es in einer Analyse des BI-Spezialisten Data Mart Consulting. Grundsätzlich unterscheidet man bei MPP-Datenbanken zwischen zwei Architekturen: Shared Nothing und Shared Everything.

- In der Shared-Nothing-Architektur erhält jeder Knoten einen Teil der Gesamtdaten, die er mit seinen eigenen Ressourcen bearbeitet. Ein "Master-Server" organisiert sowohl bei ETL-Prozessen als auch bei allen abfrageorientierten Prozessen die optimale Verteilung der Daten bzw. der Last auf die Knoten. Diese Architektur ist ausschließlich für Data-Warehouse-Anwendungen geeignet und erzeugt hier sehr gute Performance.

- Die Shared-Everything-Architektur ist dem seit langem bekannten Datenbankcluster sehr ähnlich. Die Knoten teilen sich alle verfügbaren Ressourcen und nutzen einen gemeinsamen Datenbestand, der in der Regel auf dem Storage liegt. In Anwendungsbeispielen hat dieses Konzept zwar Performancenachteile bei reinen Data-Warehouse-spezifischen Aufgaben gezeigt, ist aber im Gegensatz zur Shared-Nothing-Architektur auch für die OLTP-Anwendung ausgelegt und geeignet.

Wer beherrscht das Thema Big Data?



BIG DATA

Machen Sie mit und gewinnen Sie den Big-Data-Award!

Die COMPUTERWOCHE sucht die besten Big-Data-Projekte. Bewerben Sie sich als Anbieter von Produkten und Dienstleistungen zusammen mit Ihrem Referenzkunden beim COMPUTERWOCHE- Award "Big Data 2012". In Zusammenarbeit mit einer renommierten Jury kürt die Redaktion am 26. September in Offenbach am Main je drei Gewinner in den beiden Kategorien "Managing Data" und "Analyzing Data". Darüber hinaus vergeben wir in einer Sonderkategorie Publikumspreise für das beste Big-Data-Projekt sowie die innovativste Big-Data-Technologie.

Weitere Informationen unter: www.bigdata12.de¹

Was MPP bringt

Eine Migration der bestehenden Datenbank zu einer MPP-Datenbank wirkt sich laut dem Analysepapier vor allem positiv auf die Parameter Kosten, Performance, Komplexität und Funktionalität aus.

- **Kosten**
Bei einer MPP-Datenbank müssen keine Heerscharen an Beratern beschäftigt werden, die Hilfstabellen, Materialized Views, Bitmap-Indizes und SQL-Programme erstellen, um beispielsweise Ladeprozesse einige Minuten schneller zu machen. Auch die ständigen Bemühungen um eine Beschleunigung von Reporting und Analyse, damit die Antwortzeiten den Vorstellungen der Anwender entsprechen, sind damit hinfällig. Vielmehr reduziert sich durch die automatischen Optimierungsmodelle des Master-Servers die gesamte Datenbank-Administration bei vielen Anbietern auf einen Bruchteil.
- **Performance**
Sämtliche MPP-Anbieter gehen je nach Aufgabenstellung von einem Performancegewinn mit Faktor 10-120 aus. Das ergibt in jedem Fall genug Reserven für die mittlere Zukunft. Zusätzlich sind die neueren Systeme auch im laufenden Data-Warehouse-Betrieb umfangreich skalierbar.
- **Komplexität**
Auch die Komplexität des Datenbankschemas wird durch den Wegfall zum Beispiel von Hilfstabellen, Materialized Views und Bitmap-Indizes erheblich reduziert. Der größte Unterschied zum herkömmlichen DWH liegt sicher darin, dass durch den großen Performancegewinn auf die verdichtete Analyseschicht weitgehend verzichtet werden kann. Für den nächtlichen Ladeprozess, bei dem die Verdichtung der atomaren Analysedaten in der Regel einen nicht unerheblichen Teil der Gesamtzeit in Anspruch nimmt, ist dies somit ebenfalls von Vorteil.
- **Funktionalität**
MPP-Datenbanken bieten einige Data-Warehouse-Funktionalitäten, die man bisher lediglich von mehrdimensionalen OLAP-Datenbanken kannte. Zusätzliche Verdichtungsfunktionen "on the fly" ermöglichen in den meisten Fällen den Verzicht auf verdichtete Daten.

Links im Artikel:

¹ <http://www.bigdata12.de/>

IDG Business Media GmbH

Alle Rechte vorbehalten. Jegliche Vervielfältigung oder Weiterverbreitung in jedem Medium in Teilen oder als Ganzes bedarf der schriftlichen Zustimmung der IDG Business Media GmbH. dpa-Texte und Bilder sind urheberrechtlich geschützt und dürfen weder reproduziert noch wiederverwendet oder für gewerbliche Zwecke verwendet werden. Für den Fall, dass auf dieser Webseite unzutreffende Informationen veröffentlicht oder in Programmen oder Datenbanken Fehler enthalten sein sollten, kommt eine Haftung nur bei grober Fahrlässigkeit des Verlages oder seiner Mitarbeiter in Betracht. Die Redaktion übernimmt keine Haftung für unverlangt eingesandte Manuskripte, Fotos und Illustrationen. Für Inhalte externer Seiten, auf die von dieser Webseite aus gelinkt wird, übernimmt die IDG Business Media GmbH keine Verantwortung.