



Mit Transbase® Hypercube Data Warehouse Anwendungen effizient betreiben

Transbase® Hypercube ist eine Transbase®-Option, die die innovative **Hypercube-Technologie** für komplexe analytische Anwendungen (OLAP) mit den übrigen Transbase®-Features verbindet.

Ausgezeichnete Technologie



Dieser relationale OLAP-Ansatz (ROLAP) gewährleistet die entscheidende Skalierbarkeit für große Datenbestände und kombiniert relationale Datenbankalgorithmen mit den Möglichkeiten der multidimensionalen Analyse.

Transbase® Hypercube ist mit allen Varianten von Transbase® einsetzbar.

Transbase® Hypercube deckt damit die hohen Anforderungen an alle Systeme ab, bei denen Analysen und Auswertungen nach verschiedenen Kriterien im Vordergrund stehen. Dazu zählen alle Anwendungen aus dem Umfeld Reporting, Business Intelligence und Business Performance Management.

Die Hypercube-Technologie

Die Hypercube-Technologie beinhaltet den multidimensionalen Hypercube-Index und die dazu nötigen Verfahren zur effizienten Manipulation und Auswertung der gespeicherten Daten. Mit der nahtlosen Integration in den Datenbanken wird diese Technologie einfach, robust und effizient anwendbar. Dadurch ergeben sich signifikante **Produktivitätsvorteile** sowohl bei der Erstellung, Wartung und Weiterentwicklung als auch im Betrieb der entsprechenden Anwendungen. Durch die ausgezeichnete Performance werden in vielen Fällen Ad-Hoc-Auswertungen überhaupt erst realistisch durchführbar.

Der Hypercube-Index vereint zwei Indexierungstechnologien zu einem effizienten und einfach einzusetzenden Werkzeug.

- Multidimensionales Clustering und
- Hierarchisches Clustering.



Die Kombination dieser Techniken wird daher auch als **Multidimensionales Hierarchisches Clustering** (MHC) bezeichnet. Es unterstützt in besonderer Weise die Filterung und Analyse von mehrdimensional und hierarchisch strukturierten Datenbeständen, wobei lineare und hierarchische Merkmale in beliebigen Kombinationen auftreten können.

Definition einer Hierarchie

```
create table dates
( cdate date not null,
  date_id integer not null,
  year integer not null,
  month integer not null,
  day integer not null,
  surrogate sdate compound
( year siblings 20,
  month siblings 12,
  day siblings 31,
  date_id siblings 1
) ),
primary key (date_id)
);
```

Definition einer Hypercube-Tabelle

```
create table turnover
( outlet_id integer not null
  references outlets,
  date_id integer not null
  references dates,
  product_id integer not null
  references products,
  turnover numeric(10,2),
  surrogate so for outlet_id,
  surrogate sd for date_id,
  surrogate sp for product_id,
  primary hckey( so, sd, sp )
);
```

Multidimensionales Clustering

Multidimensionales Clustering unterstützt die Verarbeitung von mehrdimensional, d.h. nach verschiedenen unabhängigen Merkmalen gegliederten Datenbeständen. Mit dem Hypercube-Index als physische Datenstruktur werden Datensätze als mehrdimensionale Punkte im Raum geschickt organisiert und nach ihrer "räumlichen" Position gespeichert.

Das bedeutet, dass sich Datensätze, die in räumlicher Nähe liegen, in wenigen Festplatten-Blöcken sammeln und damit durch wenige Leseoperationen vom Sekundärspeicher geladen werden. Dies beschleunigt insbesondere die typischen Abfragen, die mehrdimensionale Bereiche als Treffermenge haben, da bei großen Datenbeständen die Effizienz der Auswertung fast ausschließlich durch den Zugriff auf den Sekundärspeicher bestimmt wird.

Hierarchisches Clustering

Bei OLAP-Anwendungen sind die einzelnen Dimensionen oft hierarchisch organisiert. Die Hierarchien sind dabei im wesentlichen durch die möglichen Auswertungsebenen gegeben. Das **hierarchische Clustering** ermöglicht nicht nur die effiziente Filterung nach den Einheiten auf beliebigen Hierarchieebenen, sondern vor allem auch die effiziente Gruppierung und Auswertung des Datenbestandes nach diesen Einheiten.

Transbase® Hypercube verwendet dazu eigene numerische Codierungen, sogenannte **Surrogate**. Im Prinzip ähnelt dies der Vergabe von Postleitzahlen, die die Hierarchie abbilden. Diese Codierung wird dann für die geclusterte Speicherung der Daten im Hypercube-Index verwendet, der dabei die Surrogate der einzelnen Dimensionen für die multidimensionale Indexierung verschränkt. Die Vergabe und ggf. Aktualisierung dieser Surrogate wird dabei automatisch von Transbase® Hypercube vorgenommen.

Vom Entwickler werden lediglich die Hierarchieebenen für die Dimensionen und die entsprechenden Fremdschlüsselbeziehungen definiert. Durch die hierarchische Codierung mittels der **Surrogat-Technik** ergeben die Einschränkungen auf Hierarchieelemente jeweils zusammenhängende Intervalle im Surrogatbereich.

Datenauswertung

Ein spezieller Algorithmus sorgt für die optimale Ausnutzung der in den Surrogaten enthaltenen Hierarchieinformation:



- Zunächst werden die Restriktionen der Dimensionen ausgewertet. Dadurch entstehen Intervalle auf den Dimensionscodierungen, die auf der Faktentabelle über den Hypercube-Index besonders performant evaluiert werden.
- Durch die Hierarchie-Kodierungen in der Faktentabelle kann bereits eine Gruppierung und Zusammenfassung (z.B. Summierung) von Daten vorgenommen werden, ohne dass dazu die einzelnen Datensätze mit den Dimensionen kombiniert werden müssen.
- Der Aufwand für die Kombination der wenigen verbleibenden Gruppen mit den Dimensionen (der sog. „Rückjoin“) ist daher nahezu vernachlässigbar.

Durch diese Query-Abarbeitung beschleunigt sich die Ausführung solcher Anfragen um Größenordnungen im Vergleich zu herkömmlichen Datenbanksystemen.

Vorteile mit Transbase® Hypercube

Transbase® zeichnet sich durch eine Vielzahl von Eigenschaften aus, die bei relationalen Datenbanken Stand der Technik sind, bei spezialisierten OLAP-Lösungen aber nicht unbedingt in gleicher Weise vorhanden sind. Gerade bei den genannten Anwendungen sind sie aber wegen der großen Datenmengen von entscheidender Bedeutung.

Dazu zählen:

- Performanz und Skalierbarkeit
- Backup und Recovery
- strikter Transaktionsschutz
- Datenschutz durch Verschlüsselung

Nur durch die vollständige und nahtlose Integration der Hypercube-Technologie als echte Index-Technologie in den Datenbankkern ergeben sich die entscheidenden Vorteile in allen Bereichen des Application-Lifecycle.

Auswertungseffizienz

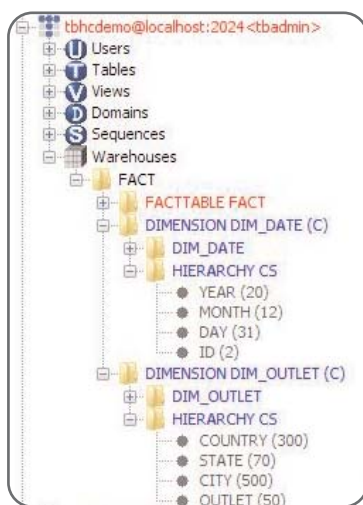
Die Auswertungseffizienz ergibt sich durch die ausgezeichneten Clustering-Eigenschaften des Hypercube-Index sowie durch die Einbeziehung der komplexen, aber effizienten Auswertungsalgorithmen in den Datenbankkern. Dabei ist vor allem die Integration der Hypercube-Algorithmen in den normalen Query-Optimierungs-Prozess zu nennen.



Datenhaltung

Durch die automatische Pflege der Hierarchieinformation (Surrogate) in den Datenbeständen entfällt die Notwendigkeit einer speziellen Aufbereitung der Daten bei der Übernahme, was den ganzen Prozess erheblich erleichtert. Außerdem tragen folgende Features dazu bei, dass die Datenübernahme und Auswertung in einfacher und effizienter Weise erfolgen kann.

TBJ-Tree



- Stored-Procedures und User-Defined-Functions ermöglichen komplexe Transformationen und Auswertungen während des ETL-Prozesses.
- Der Transbase® OCI-Reader und der Transbase® JDBC-Reader ermöglichen den direkten Zugriff auf Oracle- bzw. JDBC-fähige Datenbanken und machen so den Umweg über Export und Import überflüssig.
- Verbreitete ETL-Tools können über Standard-Schnittstellen wie ODBC, OLEDB, ADO.NET und JDBC angeschlossen werden.
- Spezielle Algorithmen beschleunigen das Einfügen großer Datenmengen in Transbase® Hypercube.
- Mittels analytischer Funktionen können komplexe Auswertungen direkt in SQL formuliert werden (z.B. RANK, PARTITION OVER). Dadurch werden Reporting-Applikationen entlastet, während das Datenbanksystem seine Optimierungsmöglichkeiten voll ausschöpfen kann.

Anwendungsentwicklung

Die Effizienz der Anwendungsentwicklung ergibt sich durch:

- übersichtliche Beschreibung der Hierarchien in der Definition des Datenbankschemas,
- standardkonforme SQL-Abfragen ohne spezielle Konstrukte, die die Semantik der Abfragen verschleiern,
- den automatischen Einsatz des Hypercube-Index durch den Query-Optimierer,
- die Unterstützung von Standardschnittstellen.

Für die Modellierung von Datenbanken mit Transbase® Hypercube eignet sich besonders der mit Transbase bereitgestellte **graphische Datenbank-Editor TBJ**. Er unterstützt den Entwurf und die Analyse von Datenbank-Schemata durch eine formularorientierte Darstellung und Definitionsmöglichkeit von Datenbankobjekten. Dabei werden insbesondere auch Hierarchien und MHC-Schemata übersichtlich dargestellt.



Anwendungsbeispiele

Transbase® Hypercube wurde z.B. erfolgreich in der Marktforschung zur Aufbereitung der Verkaufszahlen von Musik, Büchern und und Spielen (media control GfK International GmbH), sowie von Mobilfunkanbietern (e-plus, O2, T-Mobil) zur geographischen Analyse von Funkmessdaten eingesetzt.

Dies zeigt exemplarisch die Eignung von Transbase® Hypercube als effizientes Werkzeug für Statistikdatenbanken und geographische Informationssysteme sowie viele andere Bereiche mit vergleichbaren Anforderungen.

Kontakt

Transaction Software GmbH
Willy-Brandt-Allee 2
81829 München

Tel.: +49 89 / 627 09 - 0

Fax: +49 89 / 627 09 - 11

info@transaction.de
www.transaction.de
www.transbase.de