

dimensio
informatics

**Technische Hintergrundinformationen
zum Produkt**

DIMENSIO

Ultra-High-Speed für komplexe Datenbankanfragen

Extremer Performancegewinn
durch Indexierung und Gruppierung von Datenbankinhalten
auf der Grundlage inhaltlicher Datenzusammenhänge

| | |
|---|-----------|
| 1. EINFÜHRUNG..... | 1 |
| Ein Beispiel..... | 2 |
| Fazit | 2 |
| 2. DIMENSIO | 3 |
| Funktionsweise – wie geht was? | 3 |
| Unterschiede – anders als die Anderen?..... | 9 |
| Architektur – kennen Sie Lego®? | 10 |
| 3. INTEGRATION | 11 |
| Integrationspunkt 1 - die Anwendung..... | 12 |
| Integrationspunkt 2 - das Netzwerk | 13 |
| Integrationspunkt 3 - die Datenbank..... | 13 |
| 4. ANWENDUNGSGEBIETE | 14 |
| Business Intelligence - Data Warehouse Anwendungen..... | 14 |
| Volltextsuche | 15 |
| Telekommunikation | 15 |
| Bilddatenauswertung | 15 |
| Post-Processing, Materialnachweis, Garantiefallbearbeitung..... | 15 |
| Sequenzanalyse in Gendatenbanken | 16 |
| 5. GESCHWINDIGKEITSVORTEILE | 16 |
| 6. ALLEINSTELLUNGSMERKMALE | 17 |
| 7. KUNDENVORTEILE | 17 |
| 8. DIMENSIO INFORMATICS GMBH – DIE PERFORMER | 18 |
| Rückblick – wie alles begann..... | 18 |

1. Einführung

Datenbanksysteme speichern in der Regel große Mengen von Datensätzen, die nach eindeutig identifizierenden Merkmalen geordnet sind und danach durchsucht werden können.

Solche Merkmale sind z.B. Personalkennziffern, Vertragsnummern oder andere eindeutige Identifikatoren, die als Primärschlüssel bezeichnet werden und üblicherweise eine geordnete Speicherung der Datensätze gemäß dem Ordnungskriterium des Identifikators (z.B. auf- oder absteigend) erlauben. Primärschlüssel spannen damit einen linear geordneten, eindimensionalen Suchraum über den gespeicherten Datensätzen auf, in dem mit verschiedenen Verfahren nach Datensätzen gesucht werden kann.

Gängige Verfahren zum schnelleren Auffinden von Datensätzen unterteilen den eindimensionalen Suchraum in Intervalle und ordnen diese entweder sequentiell oder baumartig an, um einen Suchvorgang zu beschleunigen. Solche Verfahren bezeichnet man als Datenbankindexe oder – wenn der Kontext zu Datenbanken eindeutig ist - auch einfach als Indexe.

Die sich aufdrängende Analogie zu Inhaltsverzeichnissen und Indexen in Büchern ist in diesem Fall durchaus verständnisfördernd, da Inhaltsverzeichnisse einerseits Text-Intervalle in Form von Kapiteln und Abschnitten darstellen und andererseits deren Anordnung bzw. Reihenfolge im Buch repräsentieren. Im Gegensatz zur inhaltlich orientierten, also semantischen Intervallbildung in Büchern, teilen Datenbankindexe den Suchraum jedoch lediglich nach formal-mathematischen Kriterien auf (z.B. gleich große Intervalle, Metriken, etc.) und berücksichtigen die Semantik der Datensätze nicht.

Dennoch ist diese Art der Suchoptimierung für klassische Datenbank Anwendungen, wie Personalverwaltung, Lagerhaltung, etc. völlig ausreichend, weshalb die gängigen Datenbanksysteme für diese Art der Indexierung hoch optimiert und damit extrem performant sind.

Mit dem Aufkommen sogenannter Nicht-Standard-Anwendungen über Datenbanken gegen Ende der 1970er-Jahre trat das Problem mehrdimensionaler Suchräume auf, da für diese Art der Daten oftmals ein linearer Wertebereich für den Primärschlüssel nicht mehr ausreichte. Man denke z.B. an Bilddaten, bei denen die Bildpunkte innerhalb des Bildes durch einen X/Y-Koordinatenwert, also durch ein Wertepaar und nicht mehr nur durch einen Einzelwert, eindeutig zu identifizieren sind.

Auch für diese Art der Daten wurden Indexe erforscht und erzeugt: Die sog. mehrdimensionalen Datenbankindexe. Grundsätzlich folgen diese Indexe auch dem Prinzip der Intervallbildung, nun jedoch für mehrere Dimensionen, und der möglichst geschickten Anordnung dieser Intervalle in Datenstrukturen, z. B. in Mehrwegbäumen oder in mehrdimensionalen Gitterstrukturen.

Der Vorteil des schnellen Datenzugriffs ist durch diese Vorgehensart durchaus gegeben. Allerdings sorgt das auch hier beibehaltene Kriterium der formal-mathematischen Bildung von Intervallen dafür, dass diese Verfahren keinen Bezug zu den inhaltlichen Beziehungen zwischen den Datensätzen herstellen.

In den letzten 10 bis 15 Jahren hat sich das Anwendungsspektrum von Datenbanken allerdings wiederum erweitert und Anwendungen über mehrdimensionalen Suchräumen beziehen sich immer stärker auf die inhaltlichen, d.h. semantischen Beziehungen zwischen den einzelnen Suchraumdimensionen. Lediglich ein Beispiel hierfür sind die sog. Data-Warehouse-Anwendungen, bei denen Datensätze aus unterschiedlichen Datenbanken zusammengeführt werden, um unternehmenspolitische Zusammenhänge zwischen den Daten zu erkunden bzw. strategische Informationen daraus zu gewinnen.

Ein Beispiel

Ein einfaches Alltagsbeispiel soll das Problem der semantischen Beziehungen zwischen Suchraumdimensionen charakterisieren: In jedem Kochbuch befinden sich Rezepte für unterschiedliche Arten von Speisen, z.B. Mehlspeisen, Suppen, Speisen mit Fleisch, etc. und jedem Rezept ist eine Liste mit Zutaten beigefügt.

Leidet nun eine Person unter Zöliakie (einer Unverträglichkeit gegenüber Getreideeiweiß und, damit meist einhergehend, auch gegenüber Milchzucker), ist diese darauf angewiesen, jedes Rezept einzeln bezüglich der unverträglichen Inhaltsstoffe zu begutachten, denn die übliche Rezeptklassifizierung ist in diesem Falle untauglich.

Befinden sich die Rezepte jedoch in einer Datenbank, deren mehrdimensionaler Suchraum durch die Zutaten der Rezepte gebildet wird, kann bei einer inhaltsbezogenen Intervallbildung die gängige Klassifizierung ebenso erzeugt werden, wie die für bestimmte Krankheitsbilder wünschenswerte.

In heutigen Datenbanksystemen werden derartige Anwendungen meist durch zusätzliche Indexstrukturen bedient. Für jede Dimension des Suchraumes wird ein sog. Sekundärindex angelegt. Dieser bildet keine Primärschlüsselintervalle, sondern gruppiert die Primärschlüssel von Datensätzen entsprechend dem Auftreten bestimmter Werte in der Dimension des jeweiligen Suchraums. Die Analogie zu Buchindexen, bei denen zu bestimmten Schlagworten die Seitenzahlen deren Auftretens aufgeführt sind, ist auch hier hilfreich.

Diese Vorgehensart erzeugt jedoch eine z.T. große Anzahl zusätzlicher Datenstrukturen (pro Dimension ein kompletter zusätzlicher Index), welche den Primärschlüsselindex überlagern. Wie leicht vorstellbar ist, beschleunigt dieses Verfahren zwar das Auffinden von Datensätzen, das Ändern oder Einfügen neuer Daten wird im Extremfall jedoch stark behindert, da jeder Sekundärindex separat aktualisiert werden muss.

Fazit

Wünschenswert ist also ein Index, der die inhaltlichen Abhängigkeiten von Dimensionen eines mehrdimensionalen Suchraumes in einer einzigen Datenstruktur kompakt und effizient organisieren kann. Ein solcher Index kann als Ersatz der vielen Sekundär- und mehrdimensionalen Spezialindexe eingesetzt werden.

Der von der *dimensio informatics GmbH* angebotene Datenbankindex **DIMENSIO** realisiert eine solche komplexe, semantisch klassifizierende Datenstruktur. **DIMENSIO** unterscheidet sich von gewöhnlichen Datenbankindexen durch seine semantische Klassifikationskomponente.

2. dimensio

DIMENSIO ist ein mehrdimensionaler semantischer Index für Datenbanken mit hochdimensionalen Datenbeständen und besteht aus einer Klassifikations- und einer Verwaltungskomponente. Das Verfahren zur Klassifikation der Datenbankinhalte verwendet Künstliche Neuronale Netze zur parameterlosen Klassifikation und ist in der Lage, die Datensätze in einer Datenbank gemäß deren inhaltlichem Zusammenhang zu gruppieren. Anschließend wird diese Gruppierung in eine hoch effiziente Verwaltungsstruktur überführt, die einem R-Baum ähnelt (sog. V-Baum).

Funktionsweise – wie geht was?

DIMENSIO wirkt gegenüber anderen mehrdimensionalen Indexverfahren am effizientesten, wenn Anfragen zu beschleunigen sind, die lange Prädikatlisten enthalten, d.h. die eine größere Anzahl von Wert- oder Attributvergleichen in der SQL-WHERE-Klausel aufweisen. Ein Beispiel aus dem Anwendungsbereich Rezepte wäre eine Anfrage über folgender Relation:

Rezepte (Milch, Ei, Mehl, Butter, Margarine, Öl, Schwein, Rind, Lamm, Bohnen, Erbsen, Möhren, Spargel, Kartoffeln, Salz, Pfeffer, Paprika, Zucker);

bei der alle Attribute der Einfachheit halber als boolesche Werte modelliert sein sollen.

```
select *  
from   Rezepte  
where  Mehl = false   and  
       Milch = false  and  
       Rind = true    and  
       Erbsen = true  and  
       Möhren = true  and  
       Spargel = true;
```

wäre eine Anfrage an diese Rezeptdatenbank nach einem Rinderbraten mit Leipziger Allerlei unter Berücksichtigung der Unverträglichkeit gegenüber Getreideeiweiß (Mehl = false) und Milchzucker (Milch = false).

Um anhand dieses Beispiels zu verstehen, wie **DIMENSIO** arbeitet, versuchen wir im Folgenden nachzuvollziehen, wie die einzelnen Arbeitsabläufe aussehen.

Lernen

Im ersten Schritt werden vom Anwender (z.B. vom Datenbankadministrator) die für die Indexierung relevanten Attribute angegeben. In unserem Fall sind das alle Attribute der Relation **Rezepte**. Damit ist der mehrdimensionale Suchraum, in dem jedes Attribut mit seinem (tatsächlich in der Datenbank enthaltenen, nicht dem theoretisch möglichen) Wertebereich eine Dimension aufspannt und in dem die Datensätze später gefunden werden sollen, ausreichend definiert.

Die Übertragung der Anwenderangaben in das vom Index verwendete interne Format geschieht durch die *dimensio informatics GmbH* oder einen akkreditierten Systempartner (Anpassung des Query-Processors – vgl. Abschnitt „Architektur“), wird jedoch mit minimalem Kostenaufwand realisiert.

Im zweiten Schritt liest der Index den Datenbankinhalt blockweise ein und verwendet die in der Datenbank enthaltenen Datensätze als Beispiele für das Lernverfahren. Dabei wird nach folgendem Schema gelernt:

Jeder Datensatz markiert durch seine individuelle Wertekombination in den zu indexierenden Attributen einen Punkt im mehrdimensionalen Suchraum (Abb. 1). Im Index werden diese Punkte zunächst durch Neuronen als Zentren potenzieller Punktwolken vermerkt, d.h. es wird angenommen, dass noch weitere Datensätze existieren, deren Punkte im Suchraum sich in der Nähe der jeweiligen Neuronen befinden werden (Abb. 2).

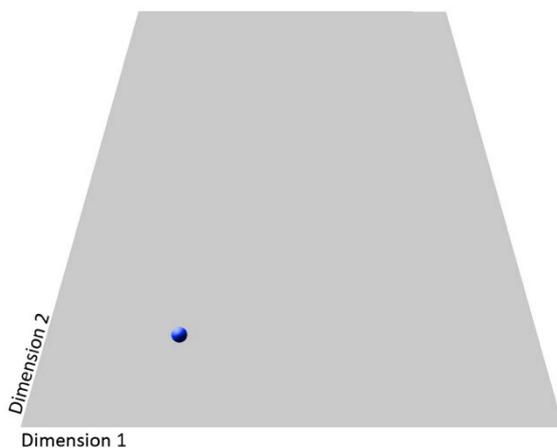


Abbildung 1

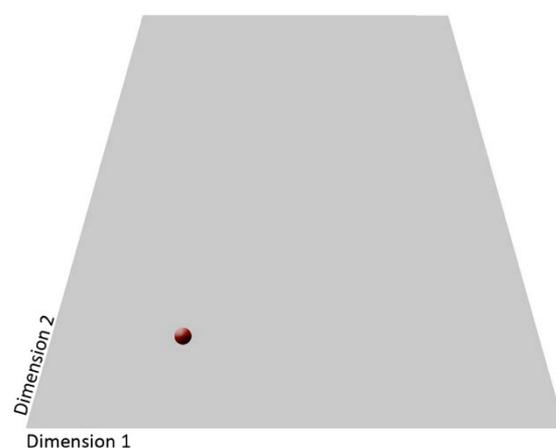


Abbildung 2

Kommen nun Datensätze hinzu, die ähnliche Inhalte haben, wie die bereits gelernten und liegen deren Punkte tatsächlich in der Nähe bereits bestehender Clusterzentren (Abb. 3), beeinflussen sie diese derart, dass sich die Neuronen – und damit die Clusterzentren - im Suchraum leicht verschieben (Abb. 4).

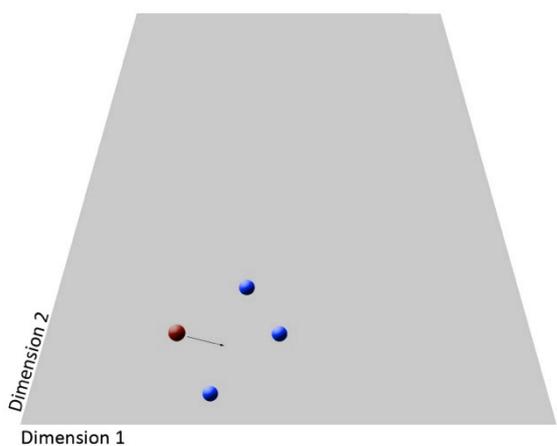


Abbildung 3

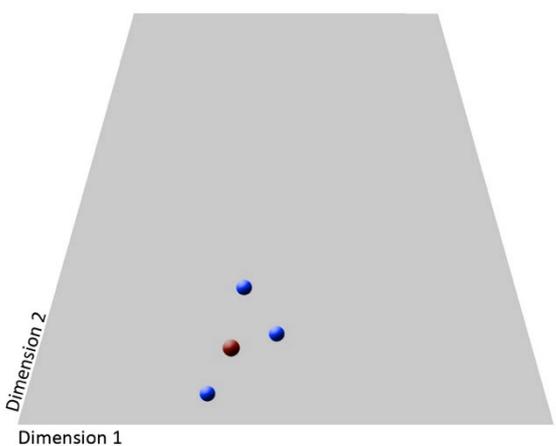


Abbildung 4

D.h. mit der Zeit entstehen um die Neuronen herum Punktwolken, deren Punkte Datensätze mit ähnlichen Inhalten repräsentieren (Abb.5). Dabei sind *ähnliche Inhalte* gleichzusetzen

mit einer Kombination aus Werten der zu indizierenden Attribute, die auf ihrer jeweiligen Wertebereichsskala nah beieinander liegen.

Interpretiert man unsere Abbildungen der Art, dass *Dimension 1* für die x-Koordinate und *Dimension 2* für die y-Koordinate steht, so befinden sich die Datensätze mit den Wertkombinationen (mittlerer x-Wert, kleiner y-Wert) in der Abbildung 5 etwa mittig unten, die Datensätze mit den Kombinationen (großer x-Wert, mittlerer y-Wert) rechts in der Mitte der Fläche und die Datensätze mit den Kombinationen (kleiner x-Wert, großer y-Wert) links oben. Gleichzeitig erkennt man, dass die das Zentrum eines Clusters repräsentierenden Neuronen in etwa in die Mitte der Cluster gerückt sind.

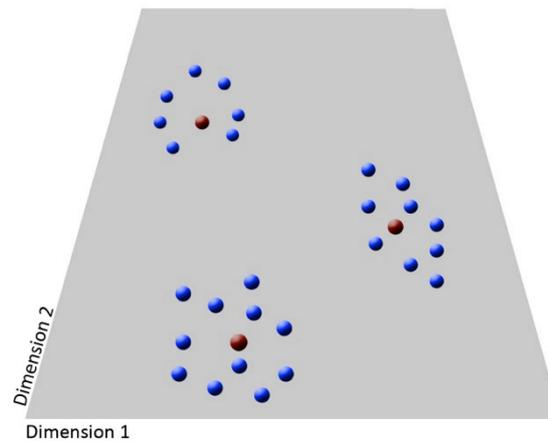


Abbildung 5

Da unser Lernverfahren hierarchisch verfeinernd arbeitet, und wir bei einer baumartigen Verfeinerungshierarchie die gesamte Datenbank als Wurzelknoten auffassen können, stellt die Abbildung 5 also die erste Knotenebene nach dem Wurzelknoten dar. Dementsprechend können sich große Cluster oder Cluster in denen sich Gruppen von weiteren Ähnlichkeiten erkennen lassen, auch weiter aufspalten und eine Anzahl von Untergruppierungen bilden (Abb. 6), was zu weiteren Knotenebenen führt.

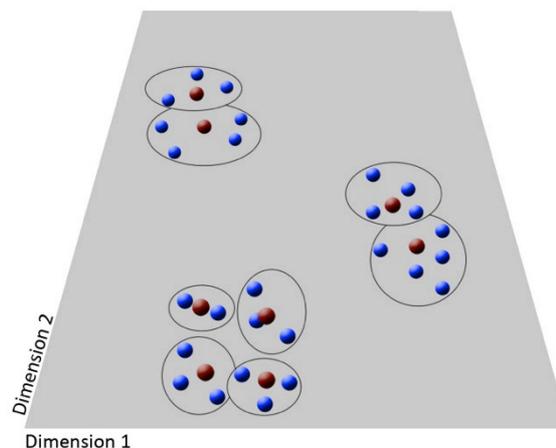
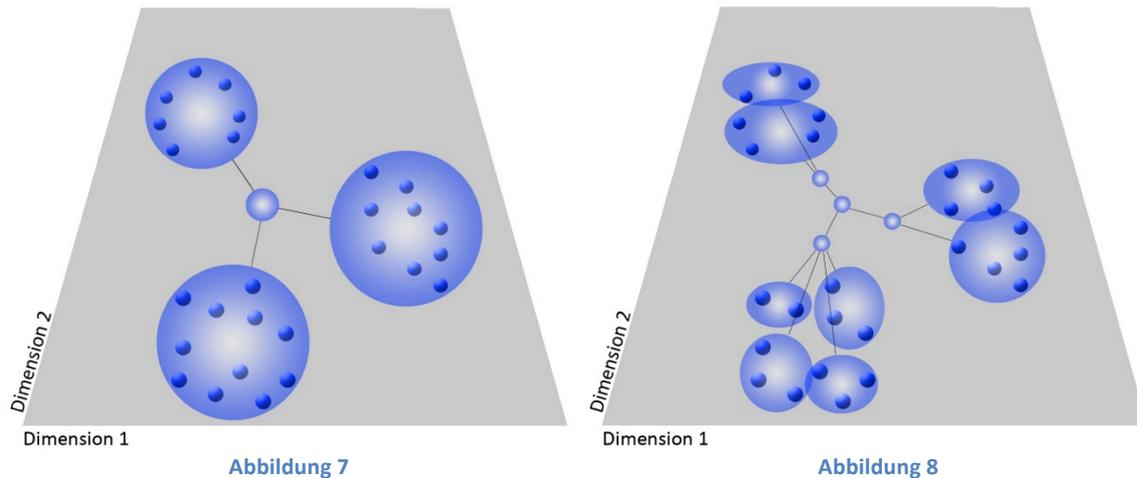


Abbildung 6

Im Extremfall könnte die Verfeinerung natürlich so weit gehen, dass pro Cluster nur noch ein einziger Datensatz enthalten ist – was allerdings in den seltensten Fällen sinnvoll sein wird, da in der Regel nach Wertebereichen und/oder Datensatzmengen angefragt wird. Deshalb

stoppt das Lernverfahren nach anwenderseitig vorgegebenen Bedingungen (z.B. minimale Anzahl von Datensätzen pro Cluster oder maximale Tiefe des Baumes).

Insgesamt entsteht durch dieses Lernverfahren ein hierarchisch verfeinerndes Abbild der Ähnlichkeiten aller Datensätze in der Datenbank, entsprechend der Kombinatorik ihrer Werte in den zur Indexbildung angegebenen Attributen (Abb. 7, mit einem eingefügten fiktiven Wurzelknoten und Abb. 8, in der die Knoten der ersten Ebene aus Abb. 7 verkleinert dargestellt sind).



Als Analogie zu der entstandenen Hierarchie kann die *Taxonomie* genommen werden. Sie gruppiert Objekte nach ähnlichen Eigenschaften und ordnet den Gruppen Begriffe zu (z.B. in der Biologie die Taxonomie des Tierreichs oder die Taxonomie der Pflanzenwelt). Unser Verfahren gruppiert ebenfalls Objekte (nämlich Datensätze) ordnet ihnen aber keine Begriffe zu, da es sich um ein vom Menschen unüberwachtes Lernverfahren handelt und die Benennung von Objekten oder Objektgruppen einen sprachlichen, vom Menschen gesteuerten Vorgang darstellt, der vom Erkennen eines Objektes oder einer Objektgruppierung völlig unabhängig ist.

Zum besseren Verständnis des Verfahrensergebnisses kann man die entstandene Struktur (Abb. 8) ordnen und durch einen Radialbaum darstellen, bei dem sich die Wurzel des Baumes in der Mitte befindet und – je nach Anzahl der Dimensionen, die zur Darstellung des Baumes verwendet werden – die Knoten kreis- oder kugelförmig darum herum angeordnet werden. Abbildung 9 zeigt einen solchen Radialbaum in dreidimensionaler Darstellung für eine reale Testdatenbank, die von **DIMENSIO** indiziert wurde.

Die Knoten des Baumes stellen die durch das Lernverfahren gebildeten Cluster dar, die Baumkanten repräsentieren die Verfeinerungshierarchie. Eine Einfärbung wurde hierbei bereits für die Knoten der ersten Ebene durchgeführt und an die nachfolgenden Kindknoten weitergegeben. Damit gibt die Knotenfärbung in diesem speziellen Fall Hinweise auf die Zugehörigkeit zu bestimmten Teilbäumen.

Bezogen auf unser Beispiel mit der Rezeptdatenbank könnte man den grün eingefärbten Teilbaum des Radialbaumes etwa dahingehend interpretieren, dass diese Knoten Rezepte mit Rindfleisch enthalten, während die blauen Knoten Schweinefleisch- und die margenta-farbenen Lammfleischrezepte enthalten.

Und weil von Ebene zu Ebene die Kombinatorik der Attribute spezieller wird, können wir davon ausgehen, dass unsere Anfrage nach einem Rinderbraten mit Leipziger Allerlei, verträglich für Zöliakiekranken, d.h. mit der Bedingung, dass bestimmte Attributwerte (Mehl und Milch) nicht (!) vorkommen dürfen, den Suchpfad im Radialbaum rasch eingrenzt und wir den entsprechenden Cluster sehr schnell finden.

Zu beachten ist allerdings, dass das Verfahren in der Praxis natürlich keine grafische Darstellung erzeugt. Sie wurde hier ausschließlich zu Erläuterungszwecken in diesem Papier erstellt.

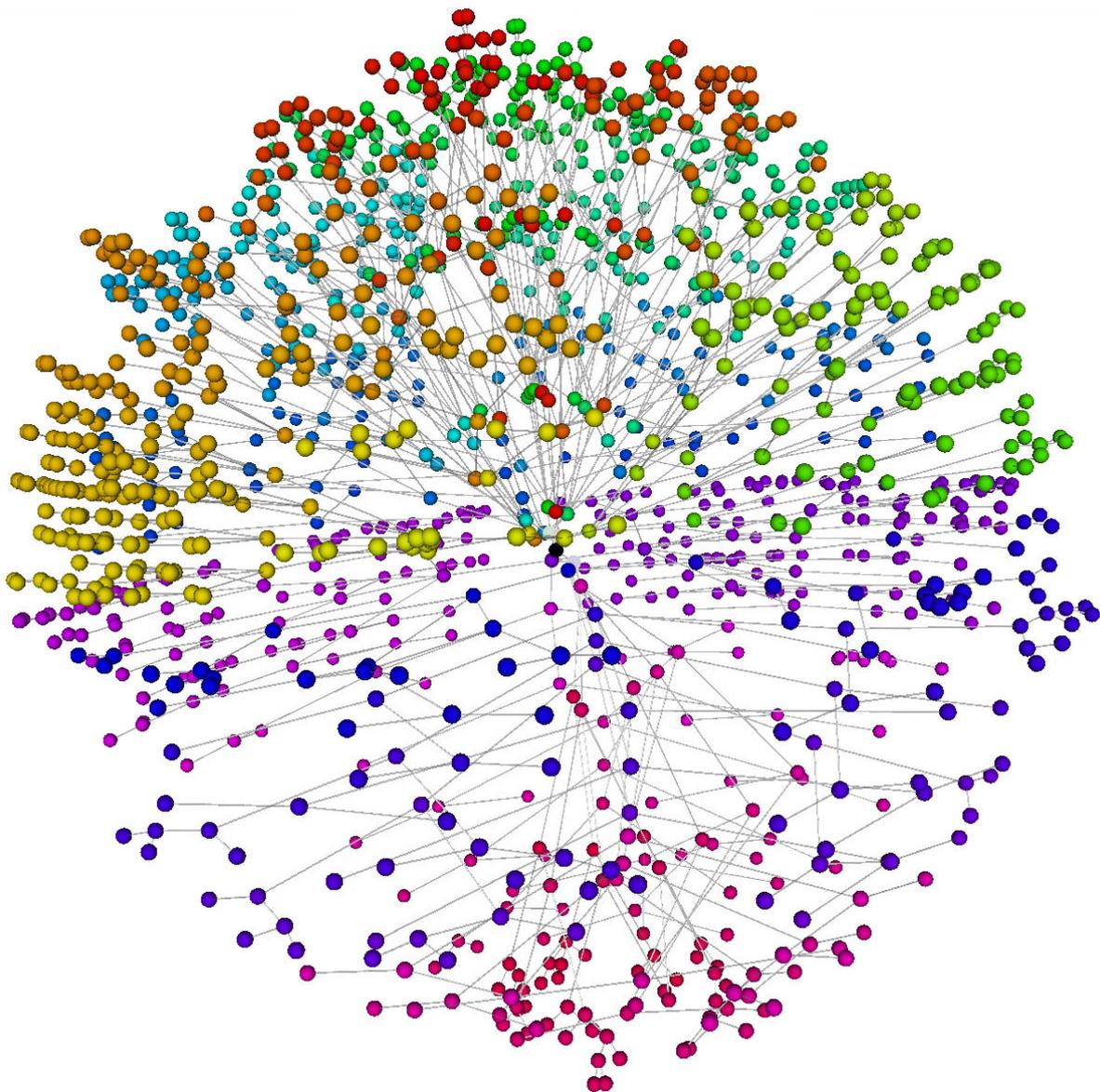


Abbildung 9

Auswerten

Wie bereits erwähnt, ist die Darstellung des Lernergebnisses in den Abbildungen 7 bis 9 lediglich eine Hilfestellung zum Verständnis des Verfahrensergebnisses. Außerdem ist besonders in Abbildung 9 leicht zu erkennen, dass der entstandene Baum eine nicht besonders

homogene Knotenverteilung im Raum aufweist und unbalanciert ist. Zur Auswertung als Indexstruktur ist dieses Ergebnis also ohne Umformungen in geeignetere Datenstrukturen nicht verwendbar.

Deshalb geschieht nach dem Ende des Lernens eine Umsetzung des Lernergebnisses in einen sog. Verwaltungsbaum, dessen Knotenstruktur geeignet ist, die Elemente der Cluster aufzunehmen. Da die Cluster mehrdimensionale Räume abgrenzen, die sich anhand der Punktwolken der Datensätze gebildet haben, und R-Bäume eine in der Datenbankwelt gut eingeführte Datenstruktur für die Indizierung räumlicher Strukturen darstellen, ist der Verwaltungsbaum eine spezielle Ausprägung des R-Baumes (Abb.10 und 11).

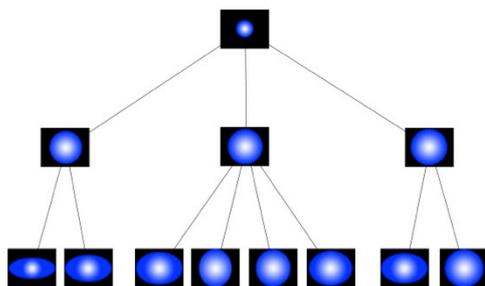


Abbildung 10

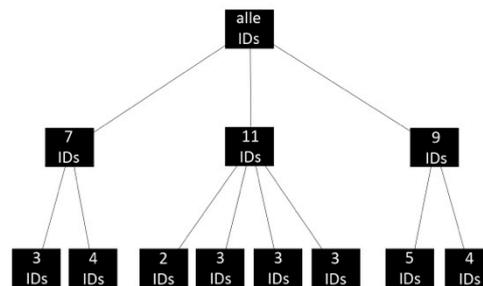


Abbildung 11

Ein Auswertungsvorgang kann nun wie folgt dargestellt werden (zur Integration von **DIMENSIO** in eine bestehende IT-Landschaft beim Kunden siehe Kapitel 4; an dieser Stelle wird lediglich auf die Beteiligung der **DIMENSIO**-Architekturkomponenten Client und Server eingegangen):

Im normalen Anfrageverlauf ohne **DIMENSIO** gehen wir hier davon aus, dass eine Datenbank Anwendung ihre SQL-Anfragen direkt an das Datenbanksystem schickt. In diesen Anfragen befinden sich Prädikatlisten, die sich u.a. auf die vom Index gelernten inhaltlichen Beziehungen zwischen Datensätzen beziehen und solche, in denen das nicht der Fall ist. Enthalten die Anfragen Prädikate, welche nicht gelernt wurden, bleibt es genau bei dem Ablauf der direkten Anfrageübergabe an das Datenbanksystem.

Enthalten die Anfragen jedoch Prädikate aus dem gelernten Indexbereich, dann werden die Anfragen nicht direkt an das Datenbanksystem geleitet sondern zunächst an den **DIMENSIO**-Client (der, je nach Integrationsweise – vgl. Kapitel 4 – an unterschiedlichen Stellen angesprochen werden kann). Dieser gibt in jedem Fall die Anfrage zur Indexauswertung an den **DIMENSIO**-Server weiter.

Im Server werden die Prädikate der Prädikatliste dahingehend ausgewertet, dass die darin angegebenen Attributwerte bestimmen, in welchen Clustern – bzw. in welchen Baumknoten des Index-Verwaltungsbaumes - nach den Datensätzen gesucht werden muss. Weil sich in den Baumknoten jedoch keine Datensätze sondern lediglich deren Repräsentanten (Primärschlüssel, vgl. Abb. 11) befinden, sucht der Server also nicht direkt die Datensätze aus der Datenbank sondern ersetzt die Prädikatliste der ursprünglichen Anfrage durch eine Menge

von Primärschlüsseln. Nämlich genau der Menge, die der Menge der ursprünglich gesuchten Datensätze entspricht. Diese veränderte Anfrage wird anschließend an das Datenbanksystem weitergeleitet und dort ausgeführt.

Betrachten wir noch einmal die *where*-Klausel unserer Beispielanfrage,

```
select *  
from Rezepte  
where Mehl = false and  
      Milch = false and  
      Rind = true and  
      Erbsen = true and  
      Möhren = true and  
      Spargel = true,
```

dann sehen wir, dass mit den ersten beiden Prädikaten ein im Kontext von Rezepten sehr starkes Ausschlusskriterium gegeben ist, denn die weitaus meiste Anzahl von Rezepten enthält gerade diese Bestandteile – zumindest Mehl. Daher dürfen wir davon ausgehen, dass bereits im ersten Schritt der Baumsuche weit über die Hälfte der in der ersten Ebene vorhandenen Knoten als Einstiegsmöglichkeit in eine weitere Teilbaumsuche ausgeschlossen werden können und der Index einen Cluster, in dem die gesuchte Prädikatkombination zutrifft, rasch finden wird.

Aber selbst wenn kein derart starkes Ausschlusskriterium in der Abfrage enthalten ist – was der Regelfall sein wird – findet der Index den gesuchten Cluster schneller als andere Indizes, denn die spezielle Kombinatorik der gesuchten Attributwerte leitet quasi direkt und ohne Umwege zum Suchergebnis. Eine Auswertung von Prädikatlisten über das tatsächliche Durchsuchen von Attributen oder auch von indizierten Attributen entfällt.

Zusammenfassend kann man sagen: Weil das Klassifikationsverfahren von **DIMENSIO** die gesamte Kombinatorik der angegebenen Attribute (bezogen auf die in der Datenbank tatsächlich enthaltenen Datensätze) gelernt hat, weiß **DIMENSIO** was Sie anfragen, bevor Sie es tun.

Unterschiede – anders als die Anderen?

Verfahrenstechnisch: Der hohe Geschwindigkeitsvorteil von **DIMENSIO** gegenüber anderen Verfahren wird durch die extreme Vereinfachung der Anfrage realisiert, welche die Arbeit des Datenbanksystems trotz langer Prädikatliste auf einen einfachen Primärschlüsselzugriff reduziert. **DIMENSIO** befreit also das Datenbanksystem von der Auswertung langer Prädikatlisten durch eine Auswertung der gelernten Clusterung! Übrig bleibt für dieses lediglich die konkrete Suche der Datensätze nach deren Primärschlüsseln – und genau dafür ist jedes Datenbanksystem hervorragend optimiert.

Systemtechnisch: **DIMENSIO** ist ein vom Datenverwaltungssystem und vom Aufbau der IT-Landschaft unabhängiges, eigenständiges Produkt! Es ist nicht integraler Bestandteil einer Applikation, kann jedoch in eine solche via API integriert werden. Es ist nicht integraler Bestandteil eines Datenbanksystems, kann aber als Fremdsystem mittels PlugIn-Technik in ein solches integriert werden. Da **DIMENSIO** lediglich auf die tatsächlichen Inhalte von Daten-

satzattributen abzielt, ist es nicht einmal notwendig, dass sich die Datensätze in einer -. Wie auch immer modellierten - Datenbank befinden.

Technologisch: *DIMENSIO* benötigt keine In-Memory-Technik, kann aber mit Systemen, die In-Memory-Technik verwenden, sehr gut zusammenarbeiten. Gemeinsam erzielen beide Systeme noch einmal einen deutlichen Geschwindigkeitsgewinn, denn die zugehörigen Datensätze der im Index gruppierten Primärschlüssel liegen nicht auf einer Festplatte sondern werden in komprimierter evtl. auch spaltenbasierter Form im Hauptspeicher vorgehalten. Es entfallen also auch noch die letzten Plattenzugriffe.

DIMENSIO verwendet selbst keine spezielle Form der Primärorganisation (z.B. spaltenorientiert), kann aber mit Systemen zusammenarbeiten, die so organisiert sind. Gleiches gilt für Kompressionstechniken. *DIMENSIO* verwendet selbst keine Datenkompression, weil der Index keine Datensätze vorhält (ausgenommen im Cache), kann allerdings mit Systemen zusammenarbeiten, die Datenkompression verwenden. Und selbstverständlich gelten beide Aussagen auch in Kombination miteinander.

Zusammengefasst: Wegen seiner absoluten Unabhängigkeit ist *DIMENSIO* auch hervorragend geeignet, minimal-invasiv in bestehende IT-Landschaften (Rechenzentren, Clouds, etc.) integriert zu werden! Selbst in Situationen, bei denen Applikationen und Datenverwaltungssysteme Kunden gehören, die nicht einmal einen minimalen Eingriff zulassen – aus welchen Gründen immer. Hier greift die Lösung der *DIMENSIO-Appliance* (mehr dazu im Abschnitt 3). Außerdem ist der Index wegen seiner losen Kopplung zum Datenverwaltungssystem völlig unabhängig von dessen Ausprägung. M.a.W. *DIMENSIO* funktioniert auch ohne DBMS, also zum Beispiel mit sog. Flat Files.

Architektur – kennen Sie Lego®?

DIMENSIO ist ein Client/Server-System. Der Client nimmt dabei lediglich die Anfragen entgegen und leitet diese an den Server weiter. Damit wird eine große Unabhängigkeit von der Einsatzumgebung erreicht, die es ermöglicht, *DIMENSIO* als stabiles System in unterschiedlichste IT-Landschaften zu integrieren. Im Folgenden gehen wir auf die Architektur des Servers ein.

Dabei wurde auf eine durchgehende Baukastenarchitektur Wert gelegt. Diese ermöglicht es, mittels standardisierter Funktionseinheiten, eine extreme Leistungssteigerung zu realisieren. Aufbauend auf einer Trennung in *Modulebene* und *Basisebene* (Abb. 12) wird die eigentliche Funktion von der benötigten Infrastruktur entkoppelt. Die Modulebene ermöglicht das Zusammenstellen von gewünschten funktionalen Anforderungen und Leistungswerten.

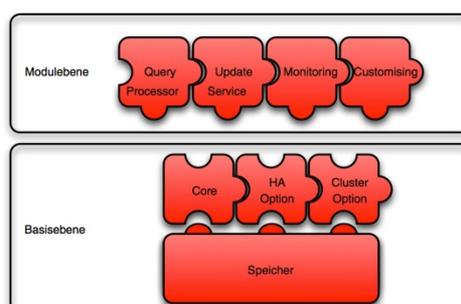


Abbildung 12

So ist der *Query-Processor* zunächst die für den Kunden am wichtigsten erscheinende Komponente, denn sie ist spezialisiert auf seinen Anwendungsbereich (z.B.: BI, Volltextsuche oder Post-Processing) und ist angepasst auf die zu lernenden Attribute (vgl. Abschnitt „Funktionsweise“). Beabsichtigt ein Kunde **DIMENSIO** gleichzeitig für mehrere Anwendungsgebiete einzusetzen, kann diese Komponente daher auch mehrfach im System enthalten sein.

Mindestens ebenso wichtig ist jedoch der *Update-Service*, denn er sorgt dafür, dass **DIMENSIO** nicht nur ein Index für *Information Retrieval* Zwecke ist – also lediglich auswertenden Charakter besitzt – sondern auch im realen Datenbankbetrieb mit Update- und Delete-Funktionalität sicher verwendet werden kann. Dabei sorgt ein abgestimmtes Verfahren zwischen Lern- und Verwaltungskomponente für die laufende Aktualität des Index und für ein ausgeglichenes Suchverhalten bei Unbalancen im Baum.

Für Techniker und Administratoren stellt sicher die *Monitoring-Komponente* eine zentrale Instanz dar, denn mit ihr lassen sich die Lastverteilung im System und die erzielten Leistungssteigerungen überwachen.

Und schließlich stellt die *Customising-Komponente* die aktive Schnittstelle zwischen Administrator, Systempfleger und **DIMENSIO** dar. Sie ermöglicht es, das Lernverhalten zu justieren, die Datenbank-Konnektivität zu administrieren und die Datenmodellierung anzupassen.

Gegenüber der Modulebene realisiert die Basisebene die eigentliche Systemumgebung, auf der die konfigurierten Module der Modulebene ausgeführt werden. Hierbei erfolgt eine aktive Lastverteilung zwischen den verfügbaren Komponenten.

Die Standardausführung der Basisebene enthält den Speicher zur Speicherung des Index und zur Pufferung von häufig angefragten Datensätzen sowie die *Core-Einheit*, welche zur Verarbeitung des Index benötigt wird.

Optional kann diese Core-Einheit durch eine *HA-Option* (HA = High Availability) ergänzt werden, die im Falle eines Ausfalls der Core-Einheit deren Aufgaben nahtlos übernimmt. Hierdurch kann ein ausfallsicherer Betrieb gewährleistet werden.

Ebenfalls optional ist die *Cluster-Option*, welche eine Lastverteilung auf weitere Core-Einheiten übernimmt und damit die Leistungsfähigkeit des Gesamtsystems **DIMENSIO** noch weiter erhöht. Durch diese Komponente wird **DIMENSIO** Cloud-fähig (vgl. Version **DIMENSIO Power Cloud** im nächsten Abschnitt).

3. Integration

DIMENSIO ist ein minimal-invasives Tuningprodukt, das – unabhängig vom Ort der Integration in eine bestehende IT-Landschaft – nur extrem geringen Aufwand verursacht. Dabei wird der Server (vgl. Abschnitt „Architektur“) in vier verschiedenen Versionen geliefert:

- **DIMENSIO Appliance:** Vorkonfiguriertes, optimiertes Komplettsystem.
- **DIMENSIO Virtual Appliance:** Nutzt bestehende Kunden-Ressourcen.
- **DIMENSIO Software Service:** Integrierbar in bestehende Software.
- **DIMENSIO Power Cloud:** Lösung für die IT-Cloud.

Um unterschiedlichen Kundenanforderungen entgegen zu kommen, gibt es technisch drei Möglichkeiten der Integration in eine bestehende IT-Infrastruktur bzw. Softwareanwendung.

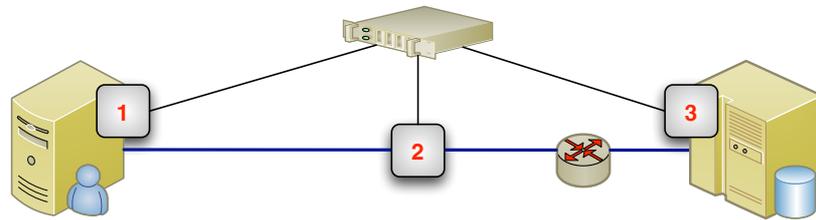


Abbildung 13

Abbildung 13 zeigt die möglichen Integrationspunkte (1-3), die im Folgenden erläutert werden.

Integrationspunkt 1 - die Anwendung

Dieser Ansatz kommt für Softwareanbieter (Softwarehäuser, ISVs) oder Unternehmen, die Zugriff auf den Sourcecode ihrer Anwendung haben, in Frage, aber auch für sog. Endkunden. Änderungen am Code sind minimal, da in jedem Fall lediglich der **DIMENSIO**-Client seitens der Anwendung angesprochen werden muss und für diese Art der Integration eine eigene Programmbibliothek (API) ausgeliefert wird. Damit ist eine problemlose Integration rasch und kostengünstig möglich. Dennoch sind die Anforderungen zwischen Softwarehäusern/ISVs und Endkunden oftmals unterschiedlich:

Für die Ausführung des **DIMENSIO**-Servers für Softwarehäuser und ISVs kommt zunächst der **DIMENSIO Software Service** in Frage, denn dieser kann vollständig und für Endkunden unsichtbar in eigene Anwendungen integriert werden. Softwarehäuser und/oder ISVs können auf diese Weise ihre hauseigenen Produkte extrem beschleunigen und quasi unter ihrem eigenen Namen von der Performanzsteigerung profitieren.

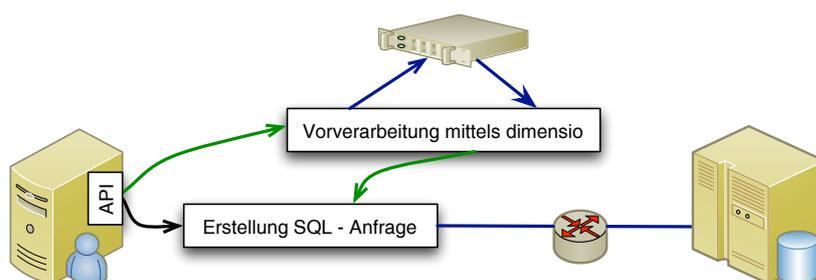


Abbildung 14

Wird zwar eine Anbindung aber keine Integration in die Anwendung gewünscht, was bei Endkundenprojekten vorkommen kann, kommen zunächst die **DIMENSIO Virtual Appliance** und eventuell auch die **DIMENSIO Appliance** als Lösung in Frage (Abb. 14).

Die **DIMENSIO Virtual Appliance** läuft auf vom Kunden bereitgestellter IT, d.h. der Kunde stellt eine virtuelle Maschine mit der notwendigen Leistungsfähigkeit für den **DIMENSIO**-Server zur Verfügung und bietet auch den geforderten Speicherbereich (storage capacity). Ist das nicht möglich oder nicht gewünscht, kann die **DIMENSIO Appliance** zum Einsatz kom-

men, denn diese wird von der *dimensio informatics GmbH* komplett vorkonfiguriert und in einem eigenen Gehäuse mit allen notwendigen Anschlüssen ausgestattet ausgeliefert.

In beiden letzteren Fällen muss in der Anwendung lediglich das Ansprechen des **DIMENSIO**-Client organisiert werden, was in der Regel mit wenigen Codezeilen getan ist.

Integrationspunkt 2 - das Netzwerk

Eine sehr elegante Möglichkeit der Integration in eine bestehende IT-Infrastruktur ist für IT-Dienstleister, wie z.B. Rechenzentren und/oder Cloudbetreiber, die Integration in das Netzwerk (Abb. 15). Diese Integrationsvariante erfordert überhaupt keine Änderungen an der bestehenden IT-Infrastruktur. Weder die Anwendung, noch die Datenbank werden angefasst oder geändert.

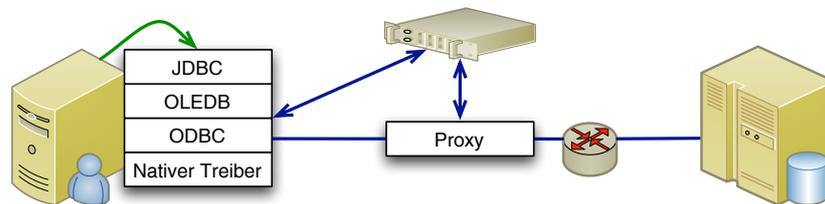


Abbildung 15

Bei dieser Lösung wird der **DIMENSIO**-Client in den Datenbanktreiber integriert. Der direkte Zugriff der Anwendung auf die Datenbank wird unterbunden, es wird lediglich eine Umleitung installiert, die sämtliche Anfragen an die Datenbank prüft. Ein Proxy filtert dann die Anfragen; von **DIMENSIO** indizierte Tabellen werden an den **DIMENSIO**-Server weitergeleitet. Der **DIMENSIO**-Server optimiert Anfragen an indizierte Tabellen, erweitert sie um Informationen zum schnelleren Datenbankzugriff und reicht diese Anfrage an die Datenbank weiter. Antworten, die von der Datenbank kommen, werden direkt an die anfragende Anwendung weitergereicht.

Zur Auslieferung kommen in diesem Fall die **DIMENSIO Appliance**, wenn es sich um normal strukturierte IT-Dienstleister – etwa Rechenzentren - handelt, oder die **DIMENSIO Power Cloud** für Cloudbetreiber. In jedem Fall handelt es sich um vorkonfigurierte, völlig selbstständig arbeitende Auslieferungen, die lediglich eine Integration in das Netzwerk des Kunden benötigen. Mit diesen Lösungen ist der Begriff „minimal-invasiv“ am deutlichsten dokumentiert.

Integrationspunkt 3 - die Datenbank

Eine schnelle und einfache Integrationsvariante für Anwender, die Zugriff auf das Datenbanksystem haben oder auch für Dienstleister, die Datenbanksysteme anpassen, ist die Anbindung des **DIMENSIO**-Servers direkt an das Datenbankmanagementsystem (Abb. 16).

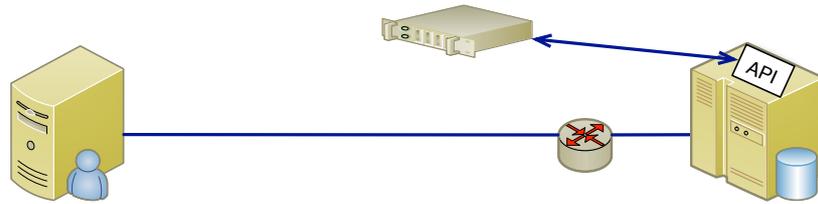


Abbildung 16

Der **DIMENSIO**-Client wird bei dieser Integrationsvariante als Plug-In dem DBMS hinzugefügt. Dazu muss das DBMS eine entsprechende Schnittstelle bereitstellen. Zur Zeit bieten nur Oracle und DB2 Schnittstellen an, die einen für **DIMENSIO** ausreichenden Funktionsumfang zur Verfügung stellen. Im Fall von Open-Source-Datenbanksystemen kann die Integration jedoch auch direkt im Code des Systems vorgenommen werden.

Erfolgt von Seiten der Anwendung ein Zugriff auf Tabellen die von **DIMENSIO** indiziert sind, leitet das DBMS die Anfrage an die Schnittstelle und damit an den **DIMENSIO**-Client weiter. Der **DIMENSIO**-Server liefert dann auf Anfrage des Client die Indizierungsinformationen via Plug-In direkt an das DBMS.

Diese Variante bietet naturgemäß die größten Chancen für eine Leistungssteigerung, da neben der Integration des **DIMENSIO**-Systems auch direkte Kooperationen zwischen Datenbankherstellern und der *dimensio informatics GmbH* denkbar sind.

4. Anwendungsgebiete

Wie bereits erwähnt, unterteilen gewöhnliche Datenbankindexverfahren den Suchraum in mathematisch oder auch technisch bestimmte Intervalle und beziehen die *inhaltlichen* Zusammenhänge von Datensätzen in den Indexierungs-Prozess nicht mit ein. Umgekehrt ist es in der Regel so, dass Programme aus der Künstlichen Intelligenz, die inhaltliche Auswertungen von Datenbeständen vornehmen, nicht die Fähigkeit aufweisen, ihre Analyseergebnisse anschließend als Zugriffsunterstützung für große Datenmengen zu verwenden. Durch die Integration von Techniken aus den beiden Bereichen, Künstliche Intelligenz und Datenbanken, mit **DIMENSIO** ist aber genau das möglich geworden.

Grundsätzlich ist der erfolgreiche Einsatz von **DIMENSIO** davon abhängig, ob und in welchem Maße die Klassifikationskomponente im Gegensatz zur mathematisch-technischen, inhaltsunabhängigen Intervallbildung einen zusätzlichen Nutzen für die Indexierung von Datenmengen erbringt. Das ist nachgewiesener Maßen zumindest bei komplexen Anfragen der Fall, bei denen viele Attributwerte zu korrelieren sind.

Business Intelligence - Data Warehouse Anwendungen

Data Warehouses werden verwendet, um Datenbestände, die für unterschiedliche Zwecke eingesetzt werden, an einem Ort zusammenzuführen, um sie dort nach übergeordneten Kriterien auszuwerten (physische Datenintegration). Die Auswertungen beziehen sich in der Regel auf unternehmenspolitische Zielsetzungen, zu treffende Managemententscheidungen oder zur Analyse von Geschäftsvorgängen über längere Zeiträume. **DIMENSIO** hat in diesem

Bereich Geschwindigkeitssteigerungen gegenüber Clusterlösungen mit In-Memory-Technik von bis zu einem Faktor von $10^3 = 1.000$ erbracht.

Volltextsuche

Volltextsuche in umfangreichen Dokumentenbeständen ist inzwischen eine Selbstverständlichkeit geworden, sowohl im privaten Bereich (z.B. Suche mittels Internetsuchmaschinen) als auch im geschäftlichen Umfeld (Dokumentenmanagementsysteme, DMS). Obwohl die Suchmaschinen sehr effizient und damit schnell arbeiten, konnte **DIMENSIO** zeigen, dass mit einem orthogonalen Suchansatz (keine Verschlagwortung sondern eine fensterbasierte Suche in einem 26^3 -dimensionalen Suchraum) ein Geschwindigkeitsgewinn um den Faktor 15 oder höher erzielt werden konnte.

Telekommunikation

Telekommunikationsunternehmen sind aufgrund des starken Wettbewerbsdrucks ständig gezwungen Ihre Netzinfrastruktur zu analysieren und zu optimieren. Dazu werden, unter anderem, technische Daten, wie sie bei jeder Nutzung von Handy-Diensten anfallen, egal ob Gespräche, SMS-Versand, Daten- und Web-Dienste, regelmäßig ausgewertet. Die Anzahl der auszuwertenden Datensätze liegt üblicherweise in der Größenordnung von 10^9 bis 10^{10} . Die Antwortzeiten bewegen sich im Bereich von 20 bis 60 Minuten. **DIMENSIO** reduziert diese Zeiten auf wenige Sekunden bis zu etwa zwei Minuten, abhängig von der Anzahl der Messdatensätze.

Bilddatenauswertung

Eine ganze Reihe von Anwendungen und Problemstellungen macht es heute notwendig Datenbanken daraufhin zu durchsuchen ob in ihnen ein gleiches oder ähnliches Bild bzw. Muster bereits vorhanden ist. Als Beispiel sei genannt die Erkennung von biometrischen Merkmalen im Bereich der Sicherheitstechniken. Aufnahmen die von einer Kamera gemacht werden müssen sofort mit tausenden oder gar Millionen von Datensätzen in einer Datenbank verglichen werden und das in Bruchteilen von Sekunden. **DIMENSIO** reduziert die Suchzeiten in diesem Anwendungsbereich um bis zu einem Faktor von $10^4 = 10.000$.

Post-Processing, Materialnachweis, Garantiefallbearbeitung

Materialnachweise und die Bearbeitung von Garantiefällen erfordern, dass Produkte, die bereits längere Zeit aus dem Produktionsprozess ausgeschieden sind, virtuell noch einmal aufgeschlüsselt werden müssen, um ihre Bestandteile erkennen zu können und um Zulieferer bzw. Herstellungsdaten zu identifizieren (Post-Processing). Solche Prozesse sind in der Regel zeitaufwändig und dauern zum Teil Stunden, da die normalen Produktionssysteme dafür nicht wieder einsetzbar sind. **DIMENSIO** erzielt in diesem Bereich Geschwindigkeitssteigerungen um den Faktor $10^3 = 1.000$ bei zum Teil deutlich reduziertem Aufwand an einzusetzender Hardware.

Sequenzanalyse in Gendatenbanken

Weltweit werden Protein- und Gensequenzen tausender verschiedener Organismen in Datenbanken bereitgehalten. Über einen Vergleich dieser Sequenzen kann unter anderem auf die Funktion menschlicher Gene rückgeschlossen werden. Gleichermaßen wird mit diesen Daten Grundlagenforschung und angewandte Entwicklung, beispielsweise von Krebsmedikamenten, betrieben. Schon einfache paarweise Alignments sind recht langsam. Multiples Alignments, d.h. die gleichzeitige Analyse mehrerer Sequenzen, dauern sehr viel länger. **dimensio** erzielt hier erhebliche Verbesserungen. **DIMENSIO** kann in diesem Bereich analog zur Volltextsuche eingesetzt werden und deutliche Geschwindigkeitssteigerungen möglich machen.

5. Geschwindigkeitsvorteile

Erste Performanz-Messungen mit einem universitären Prototyp und einem am Markt führenden Datenbanksystem (Quelle: Dissertation TU Chemnitz, O. Görlitz, 2004) hatten bereits gezeigt, dass bei einem 10-dimensionalen Suchraum gegenüber der Technik mit Sekundärindizes über **99%** und gegenüber einem systemspezifischen Composite Key Index immer noch etwa **85%** der sonst notwendigen I/O-Operationen eingespart werden¹.

In den Jahren 2009 bis 2011 wurden dann mit dem inzwischen verbesserten Prototypen weitere Tests (PoCs) durchgeführt, von denen im Folgenden einige aufgelistet sind:

Im Rahmen einer landesweiten Kartierung von Sachsen ging es darum, wie sich **DIMENSIO** im Vergleich zu etablierten mehrdimensionalen Datenbank-Extensions (in diesem Falle der Spatial-Extension eines großen Datenbankherstellers) verhält. Dabei wurde über einem 50-Meterraster von Sachsen mit jeweils 20 Werten pro Rasterpunkt (= 20 Dimensionen) der PoC mit knapp **300 Mio. Datensätzen** durchgeführt. Zielsetzung war es, mindestens 1.000 Anfragen pro Sekunde zu erreichen. Die Laufzeiten pro Anfrage betragen dabei **ohne DIMENSIO** ca. **0,15** Sekunden, entsprechend 6 Anfragen pro Sekunde, und **mit DIMENSIO** rund $5 * 10^{-4}$ Sekunden, was etwa 2.000 Anfragen pro Sekunde bzw. einem Steigerungsfaktor von **300** entsprach!

Ein weiterer Test mit realen Industriedaten aus dem Bereich MES² (Manufacturing Execution System = Fertigungssteuerungssystem) erfolgte über einer Tabelle mit **6,6 Mio. Datensätzen** mit dem folgenden SQL-Statement:

```
select count(*) from (  
  select Attr-1, count(*) Attr-2 , NVL(Attr-3, 0),NVL(Attr-4, 0)
```

¹ Getestet wurde auf einer Datenbank (auf Sun II Ultra Sparc Server) mit 2,1 Mio. 10-dimensionalen, zufällig erzeugten Datensätzen. Der Wertebereich jeder Dimension umfasste das Intervall $[-10^4, 10^4]$. Aus der Klassifikationskomponente resultierten 10^5 hyperkugelförmige Cluster bzw. Gruppierungen mit einem Radius von 250. Um einen Vergleich mit den herkömmlichen Methoden mehrdimensionaler Indizierung zu bekommen, wurden über der Tabelle in klassischer Primärschlüsselorganisation zwei Arten von im System integrierten Sekundärschlüsselindizes angelegt: Ein sog. kaskadierter Schlüssel für 10 Dimensionen (Composite Key Index) und 10 einzelne Sekundärindizes. An diese Konfiguration wurden sieben Reihen zu je 100 Testanfragen mit wachsender Bereichsgröße (zwischen 100 und 3000 Antwortdatensätzen) gestellt, wobei jede Anfrage mindestens eine Gruppierung tangierte. Gemessen wurde die zeitrelevante Anzahl der I/O-Operationen (gerundet).

² Weitere Details hierzu unter <http://www.dimensio-informatics.com/html/di-ais.html>

```
from Relation
where Attr-5 = 0 and Attr-6 is not null
      and Datetime >= to_date('10.08.2009','dd.mm.yyyy')
      and Datetime < to_date('11.08.2009','dd.mm.yyyy')
      and Attr-7 is not null
group by Attr-1, NVL(Attr-3, 0), NVL(Attr-4, 0))
```

Dabei ergab sich im Durchschnitt über alle Testläufe ein Faktor für die Geschwindigkeitssteigerung von **$0,97 * 10^3$** – also knapp einem Faktor von **1.000**.

Weil die *dimensio informatics GmbH* u.a. IBM-Entwicklungspartner ist, hatte schließlich auch das IBM Innovation Center in Ehningen eingeladen, um den Prototypen von **DIMENSIO** im eigenen Hause auf einer aktuellen DB2-Installation zu testen. Zu verifizieren war der Geschwindigkeitsvorteil, der mit **DIMENSIO** gegenüber normalen DB2-Zugriffen entsteht³. Am Ende des Testtages lag die durchschnittliche Zeit für eine Anfrage in DB2 im Falle von ad-hoc-Anfragen mit **DIMENSIO** bei **3%** vom direkten DB2 Zugriff (100%) und bei **68%** für statisch gebundene Anfragen (d.h. ohne Prepare-Zeit).

6. Alleinstellungsmerkmale

- Hochdimensionaler Index für beliebig viele Dimensionen
- Intervalldefinition durch semantische Analyse der Daten und Clusterbildung mittels Künstlicher Neuronaler Netze
- Minimal-invasive Integration in bestehende IT-Landschaften (Integration in Applikation, Netzwerk oder DBMS)
- Funktioniert mit jeder Datenverwaltung (nicht nur Datenbanken, sondern auch kundenspezifische Speicherlösungen und/oder Flat Files)

7. Kundenvorteile

- bis zu 10.000 mal schnellere Datenauswertung gegenüber klassischen Lösungen
- der enorme Zeitgewinn ermöglicht optimierte Geschäftsprozesse
- einmalige Kostenersparnis, da im Vergleich zu Standard-Lösungen geringere Hardwareinvestitionen und geringere Investitionen in zusätzliche CPU-Lizenzen von Datenbanken notwendig sind
- laufende Kostenersparnis, beginnend bei Supportaufwand, da Hard- und Software weniger komplex sind, über Wartungsgebühren, bis hin zu geringeren Energiekosten, da weniger Hardware und weniger Rechenzeit benötigt wird
- neue Möglichkeiten der Datenauswertung die früher aufgrund der langen Wartezeiten keine Nutzerakzeptanz fanden

³ Die dafür aus dem BI-Bereich abstrahierte Testdatenbank enthielt eine Relation mit 998 Spalten (Wertebereich jeweils 0 oder 1, d.h. Eigenschaft vorhanden oder nicht) und ca. 12.800 Datensätze. Ausgeführt wurde der Test auf einem System x 3755 mit 4 x AMD Opteron 64 Dual Core 3,2 GHz und 12 GB RAM unter Ubuntu 10.04 (64-bit). Für DB2 kam die Version 9.7 FP2 zum Einsatz. Getestet wurden jeweils 50 Punktabfragen, die zwischen einem und 42 Tupeln zurücklieferten.

8. **dimensio informatics GmbH – die Performer**

Die *dimensio informatics GmbH* – the spirit of speed - ist ein unabhängiger, speziell auf Performanzprobleme fokussierter Technologieanbieter. Gemeinsam mit Partnern (ISVs, OEM-Partnern, DBMS-Herstellern, etc.) werden Produkte und IT-Dienstleistungen zur Analyse und Beschleunigung von Geschäftsprozessen angeboten. Performanceengpässe werden analysiert und, aufbauend auf den eigenen Produkten individuelle, minimal-invasive Lösungen konzipiert. Die Softwareprodukte der *dimensio informatics GmbH* nutzen hochmoderne Technologien. Sie lassen sich einfach und kostengünstig in bestehende IT-Infrastrukturen integrieren und arbeiten mit allen gängigen Businessanwendungen und allen Datenbankmanagementsystemen zusammen.

Rückblick – wie alles begann

Die Ursprünge von **DIMENSIO** stammen aus dem universitären Umfeld Ende der 1980er Jahre. Erste Versuche, den Index hardware-basiert zu realisieren, scheiterten. Mitte der 1990er Jahre erfolgte eine Wiederaufnahme des Themas als software-basierte Lösung.

1999 erhielt eine diesbezügliche Diplomarbeit den Universitätspreis der Technischen Universität Chemnitz und im Jahre 2000 wurde auf der CeBIT in Hannover ein erster Prototyp des inzwischen ICIX (Intelligent Cluster Index) genannten Verfahrens ausgestellt. In Fortführung der Arbeiten entstand im Laufe der Jahre ein stabiler universitärer Prototyp.

Im Juli 2010 konnte nach erfolgreicher Teilnahme des universitären Teams an verschiedenen Businessplanwettbewerben die Gründung der *dimensio informatics GmbH* als Ausgründung der Technischen Universität Chemnitz realisiert werden, die seit Oktober 2010 vom Technologie Gründer Fonds Sachsen (TGFS) finanziell abgesichert wird.

dimensio informatics GmbH
Brückenstr. 4
09111 Chemnitz
Deutschland

Telefon +49 371 26 20 19 0
Telefax +49 371 26 20 19 10
info@dimensio-informatics.com
www.dimensio-informatics.com

Lego® ist eingetragene Marke der LEGO-Group; dimensio informatics®, dimensio® und the spirit of speed® sind eingetragenen Marken der dimensio informatics GmbH