



Portable Database Images (.PDI)

Ein Datenformat zur Extraktion von Wissen aus Datenbanken: Technologie, Potenzial und Vision

Dirk Owerfeld

Panoratio Database Images GmbH

<http://www.panoratio.com>

info@panoratio.com

Abstract

Portable Database Images stellen eine patentierte, RAM-basierende Technologie dar, die es erlaubt, die analytische Essenz (das „Wissen“) aus sehr großen Datenbanken zu einen „Portable Database Image“ von nur wenigen Megabyte Größe zu kondensieren – eine Art JPEG für die Statistik von Daten. Damit wird das Wissen aus großen Datenbanken spielend leicht handhabbar und portabel - äquivalent zu dem bekannten pdf-Format für Dokumente.

1 These: Paradigmenwechsel

Systematisch gesammelte Daten eines Unternehmens oder einer Institution werden mehr und mehr als das Wissen des Unternehmens interpretiert. Dieses rein maschinell ermittelbare Wissen steht jedoch niemals alleine, sondern wird ergänzt durch das jeweilige Fachwissen der Mitarbeiter.

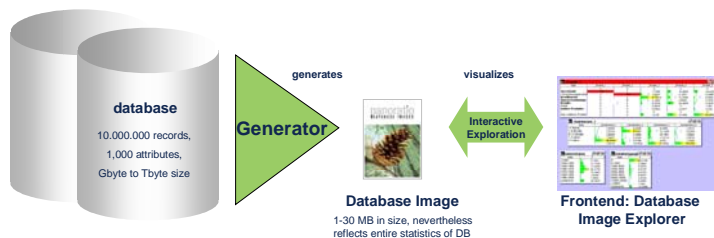
These: Genau so selbstverständlich wie heute nahezu alle Entscheider und Manager selbst E-Mails schreiben und lesen – ohne Zwischenschritte über ihr Sekretariat, eben weil es so einfach geworden ist - so werden diese in Zukunft auch selbst und direkt Antworten auf Fragen aus den Datenbeständen des Unternehmens finden, intuitive Vermutungen verifizieren und durch sanfte Vermischung von menschlichen Fachkenntnissen mit maschinelltem Wissen unerwartete Sachverhalte entdecken und Entwicklungstrends frühzeitig erkennen.

Technische Voraussetzungen für einen Paradigmenwechsel sind also vor allem die Fähigkeit, möglichst viele Parameter eines komplexen Sachverhalts (auch Variablen oder Dimensionen genannt) gleichzeitig berücksichtigen zu können, die Geschwindigkeit als auch die Handlichkeit (Portabilität). Diese können vor allem durch eine Verlagerung der analytischen Intelligenz von langsamen, festplattenbasierenden Systemen zu RAM-basierenden Systemen realisiert werden.

2 Portable Database Images

“Analytical Database Images” –

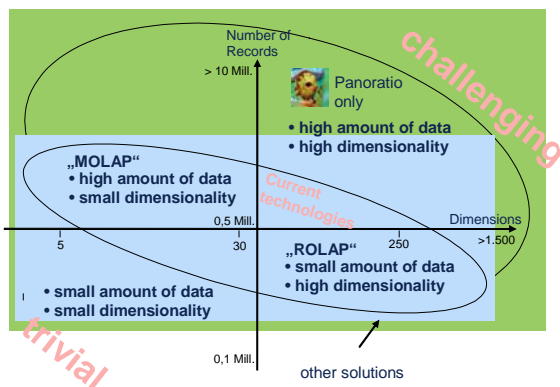
99% of the information relevant for analytics condensed to 1% of the size



Die Database Image Technologie ist ein völlig neuartiger RAM-basierender Ansatz: sie erzeugt (mittels des *Generators*) aus einer Datenbank durch patentierte Verfahren aus

dem Bereich AI/Machine Learning ein so genanntes Portable Database Image, d.h. ein statistisches Abbild der Datenbank, auf welchem dann ohne Verbindung zur Original-DB nahezu alle Analysen durchgeführt werden können (durch den *Database Image Explorer*).

Das Ergebnis der Generation eines Images stellt jedoch keine einfache Kompression im klassischen Sinne dar, sondern der Generator erzeugt ein

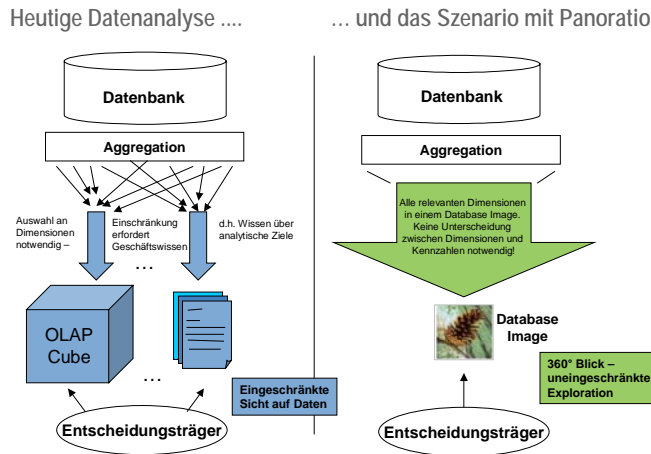


Database Image durch unüberwachtes, automatisches Lernen eines auf kombinierten, bayesianischen Netzen basierenden statistischen Modells der Datenbank. Ein solches Database Image ist in der Regel nur wenige MB groß und überschreitet selbst bei sehr großen Originaldaten (Millionen von Records, TeraByte Größe) niemals

die Größe des Hauptspeichers eines Standard-PCs, kann also jederzeit und an jedem Ort eingesetzt werden.

Damit verschiebt die PDI-Technologie die technischen Limitationen bisheriger Ansätze weit nach oben und vereinheitlicht mehrere klassische Technologien.

3 Konsequenzen



Der Quantensprung der Leistungsfähigkeit erlaubt völlig neue Anwendungen sowie Organisationsformen: statt in aufwändigen Iterationen zwischen IT-Experte und Anwender nur einen limitierten Blick auf die Daten zu erhalten, können in einem Database Image alle relevanten Dimensionen einer Domäne abgebildet werden. Denn erst wenn Anwender selbst auf allen Daten ihrer Domäne arbeiten können und sich menschliches Wissen sanft mit maschinellem Wissen vermischt, entfaltet sich der volle Nutzen der Daten!

4 Einsatzfelder und Referenzen

- **Business Intelligence (Analytical Customer Relationship Management)**
Karstadt Quelle Information Services, Direkt Anlage Bank, Baustoffhandel, ...
- **Plant Intelligence (Analyse technischer Daten und Zusammenhänge)**
Siemens Power Generation: Optimierung von Gasturbinen, Halbleiterfertigung, ...
- **Medizinische Daten**
Krankenhaus München-Schwabing: Klinik-Benchmarking (ICD-Codes), Siemens Health Services (Extraktion medizinischen Wissens aus Patientendaten).

