

Hohe Skalierbarkeit und Performance, Einfachheit und niedrige Total Cost of Ownership

Innovative Technologien in der **Business Intelligence Welt**

Wussten Sie, dass Microsoft einer der führenden Anbieter für Business Intelligence und Data Warehousing ist? In Unternehmen nutzen „Information Worker“ Microsoft Office, die IT verlässt sich auf Windows Server, die Kinder zuhause toben sich an einer Xbox aus und möglicherweise sind Sie stolzer Besitzer eines Windows Smartphones. Aber Business Intelligence?

In Gartners Quadranten für Business Intelligence Software liegt Microsoft im Jahr 2011 vorn in der Spitzengruppe (Bild 1). Mit der BI- und Datenmanagement-Plattform SQL Server 2008 R2 in Verbindung mit Office 2010 wurde eine starke Position erreicht. Die nächste SQL Server-Generation „Codename Denali“ – derzeit in einer öffentlichen Betaphase und ab Anfang 2012 allgemein verfügbar – bringt mit innovativen Analysetechnologien weiteren Schwung in den Microsoft BI Stack.

Der „BI Burger“

Das Thema Datenanalyse hat eine lange Tradition bei Microsoft. Bereits 1998 wurde mit dem SQL Server 7.0 neben der relationalen Datenbankmaschine zum ersten Mal eine analytische Engine ausgeliefert – die „Analysis Services“ (SSAS). Die Technologie wurde vom

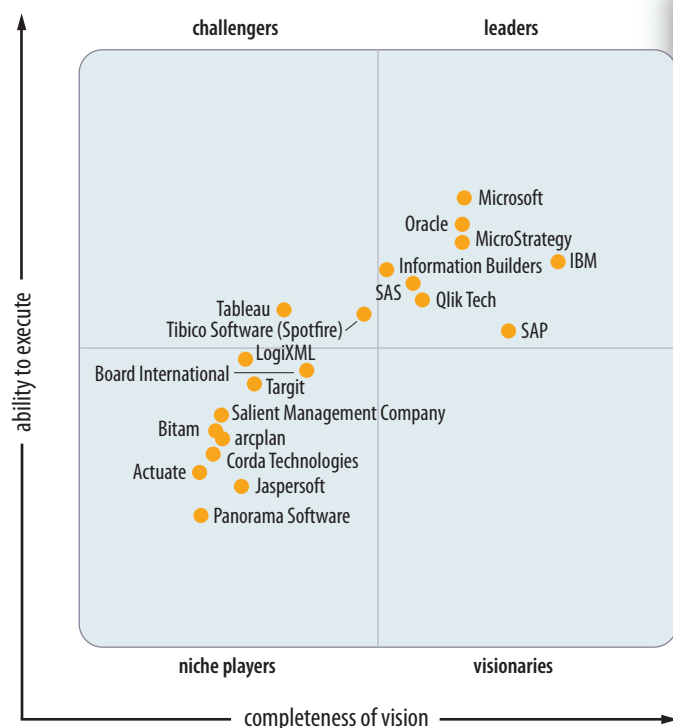


Bild 1: Gartners magischer Quadrant für BI 2011



israelischen Spezialanbieter Panorama übernommen und weiterentwickelt.

Die Analysis Services sind heute führend für Online Analytical Processing (OLAP) und Data Mining. Ein populäres Einsatzbeispiel ist ein 12 Terabyte Cube komprimierter Analysedaten, mit dem bei Yahoo eine geschäftskritische BI-Lösung betrieben wird. Als SSAS-Abfragesprache wird der von Microsoft entwickelte Quasi-Industriestandard MDX (Multidimensional Expressions) genutzt, mit dem sich beispielsweise auch SAP BW-Datenbanken abfragen lassen und der von allen wichtigen BI-Frontends genutzt wird.

Die Analysis Services sind aber nur ein Baustein im Business Intelligence-

Angebot des Unternehmens. Das Besondere: Im Gegensatz zu reinen BI-Anbietern finden Sie das Thema Datenanalyse nicht in einem einzelnen, hoch spezialisierten Produkt, welches nur von einem Power User angewendet werden kann. BI an jedem Arbeitsplatz heißt die Devise und daher erfolgt der Zugriff auf die von der IT zentral gemanagten Informationen vor allem über die gewohnten Office- und SharePoint-Anwendungen in Echtzeit. Unternehmen können also Software verwenden, die sie bereits im Einsatz haben und mit deren Administration die IT gut vertraut ist.

Eine leicht verständliche Visualisierung des Microsoft BI Stacks ist der so genannte „BI Burger“ (Bild 2). Die Basis

bildet der SQL Server, der in einer klassischen BI-Infrastruktur die Funktionen für die Bewirtschaftung von Data Warehouses zur Verfügung stellt, zum Beispiel ETL für das Laden der Daten und ein leistungsfähiges RDBMS für die effektive Speicherung (inklusive Komprimierung oder Partitionierung). Im SQL Server sind aber auch die Analysefunktionen vorhanden und außerdem die Dienste für das Enterprise Reporting. Dies ist alles integriert in einer Lizenz.

Der Anwender kann seine individuellen Analysen mit Excel durchführen. Der Zugriff auf Analysis Services-Daten passiert über die gewohnten Pivot-Tabellen oder Formeln. Geht es um das Vorbereiten von Standardre-

ports, dann kommt als Berichtswerkzeug der Report Builder ins Spiel. Mit officartigem Look and Feel ausgestattet erlaubt er Anwendern das pixelgenaue Design von Berichten.

In der Mitte des BI Burgers hält der SharePoint Server die Dinge zusammen. Er stellt Funktionen zur Verfügung, die in jeder BI-Umgebung benötigt werden, wie etwa Kollaboration, Workflow, Versionierung und vieles andere mehr. Darüber hinaus steuert der SharePoint Server BI-Web-Portalfunktionen bei: Excel-Mappen können als Bericht veröffentlicht und im Browser betrachtet werden. Interaktive Tools erlauben eine Analyse direkt im Portal.

Der Reporting- und Analyse-Stack ist vollständig, aber das Thema Analyse entwickelt sich ständig weiter. Agiles BI und Self Service werden immer wichtiger. Und hier kommen neue Technologien ins Spiel.

Die PowerPivot In Memory-Revolution

Die „In Memory-Revolution“ begann bei Microsoft erstaunlicher Weise nicht im Backend, sondern bei des Controllers „liebsten Kind“: Excel. PowerPivot für Excel ist ein kostenloses Add In, mit dem sich Analysen großer Datenmengen auf jedem handelsüblichen Notebook oder Desktop PC durchführen lassen. Der Fachanwender bekommt seine persönliche Analyse-Engine, die es ihm nicht nur ermöglicht große Datensets zu verarbeiten – hunderte Millionen Datenzeilen sind möglich – über eine Vielzahl von Konnektoren können Daten aus unterschiedlichsten Quellen bezogen und miteinander kombiniert werden (Bild 3).

Damit löst PowerPivot „SVERWEIS“ ab – einer beliebten Excel-Funktion mit der sich Daten verknüpfen lassen, allerdings nur umständlich. Das AddIn basiert auf der Microsoft VertiPaq-Technologie. VertiPaq ist eine spaltenorientierte In Memory-Datenbank. Die Vorteile der Spaltenorientierung beschreibt Dr. Martin in dem Artikel „Auf dem Vormarsch“, ebenfalls diesem Heft, bereits ausführlich. Nur so viel: Durch die VertiPaq-Komprimierung kann man in einem 2 Gigabyte Adressraum (das ist die in Excel mögliche Speicher-



Bild 2: Der Microsoft „BI Burger“

menge für PowerPivot) bis zu circa 100 Gigabyte Daten unterbringen.

DAX – Data Analysis Expressions

Möchte der Anwender über die Standardfunktionen von Excel hinaus typische BI-Analysen durchführen, wie Zeitbetrachtungen oder Navigation in hierarchischen Strukturen, dann steht ihm dafür eine neue Ausdruckssprache

zur Verfügung: DAX – Data Analysis Expressions. DAX hat ähnliche Ziele wie MDX ist aber viel einfacher in der Anwendung. Ein Excel-Benutzer, der mit Formeln zurechtkommt, wird sich in DAX schnell einarbeiten können.

Was auf dem Client beginnt geht auf dem Server weiter. Mit PowerPivot für SharePoint lassen sich Excel PowerPivot-Mappen auf einem SharePoint veröffentlichen und mit den Kollegen teilen. Das Daten-Refresh kann dann von SharePoint zeitgeplant vorgenommen werden.

Spalten-Power in den Analysis Services

Die Revolution geht weiter! In der zukünftigen Version der SSAS ist eine spaltenorientierte Engine enthalten, welche die vorhandenen multidimensionalen Verfahren ergänzt (klassisches MOLAP oder ROLAP). Dieser „tabular mode“ basiert auf einer Weiterentwicklung der VertiPaq-Engine. Gleichzeitig wurde DAX zu einer Definitionssprache für die BI-Modellierung weiter entwickelt. Hier geht es nicht mehr um „persönliches BI“, wie in PowerPivot, sondern um deutlich größere Analysemodelle und Datenmengen. Die Modelle werden nicht mehr in Excel gebaut, sondern in Visual Studio, dem traditionellen Microsoft Entwicklerwerkzeug. PowerPivot-Modelle können bei Bedarf importiert werden.

Der „tabular mode“ ist nicht nur aus Performancegründen interessant. Die



BI an jedem Arbeitsplatz heißt die Devise und daher erfolgt der Zugriff auf die von der IT zentral gemanagten Informationen vor allem über die gewohnten Office- und SharePoint-Anwendungen in Echtzeit.

WEB-TIPP: www.microsoft.de/bi

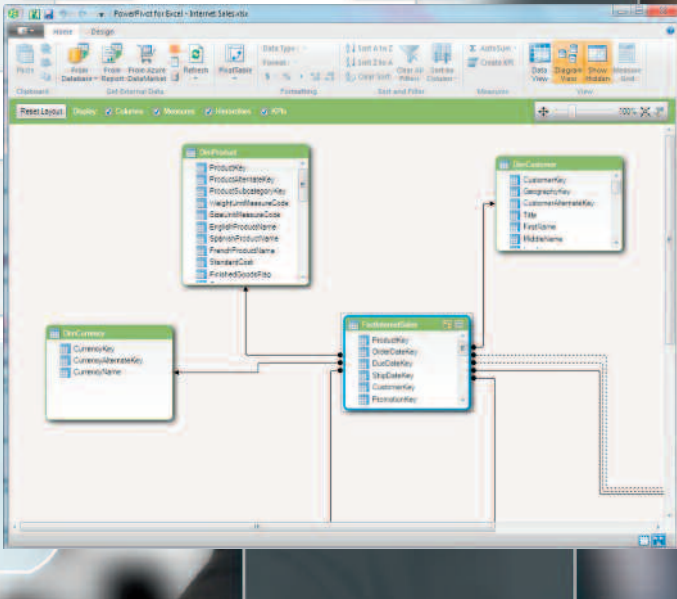


Bild 3: PowerPivot in Excel

derzeit aktuellen Analysis Services im multidimensionalen Modus liefern auch heute schon eine Geschwindigkeit, die für die meisten B-Anwendungen mehr als ausreicht. Ein wichtiges Argument für die spaltenorientierte Speicherung und Modellierung liegt in der Einfachheit der Anwendung, Entwicklungszeiten können reduziert werden. Es sind weniger Spezialkenntnisse erforderlich.

Für sehr große oder sehr komplexe Projekte wird weiterhin der Einsatz der multidimensionalen SSAS-Engine empfohlen. In Projekten lassen sich die beiden Modi auch parallel einsetzen. Für die Clients ist der Speicher-Modus vollkommen transparent. Für die Zukunft

der Analyse gilt aber: Spaltenorientierung und In Memory sind Trumpf!

Ein Booster für das Data Warehouse

Auch in der relationalen Welt tut sich etwas: Mit den „Column Store Indizes“ halten spaltenorientierte Techniken Einzug in die Datenbankmaschine. Ohne Spezialwerkzeuge lassen sich damit analytische SQL Abfragen um Größenordnungen beschleunigen. Mit einem simplen T-SQL-Kommando `CREATE COLUMNSTORE INDEX` oder das Management-GUI werden die Daten einer Basistabelle in das Spaltenformat gebracht. Der Begriff „Index“ ist

etwas irreführend, da es sich nicht um einen üblichen Index handelt, sondern um das Speichern im hoch komprimierten VertiPaq-Format.

Der Administrator kann es sich dabei sehr leicht machen: Best Practice ist das Auswählen aller Spalten der Basistabelle. Komplizierte Optimierungsüberlegungen entfallen. Der SQL Server-Abfrageoptimierer findet selbst heraus, welche Abfragen von der Verwendung des Spaltenindex profitieren. Spalten-Indizes profitieren übrigens besonders stark von den modernen Prozessorarchitekturen mit einer Vielzahl von Kernen, da die Abfragen stark parallelisiert ausgeführt werden können.

Parallel Data Warehouse für Big Data

Neben den spaltenorientierten Engines gibt es bei den Analysedatenbanken die Klasse der hoch parallel arbeitenden Systeme. In diesen werden Abfragen durch viele Rechenknoten und Prozessoren verteilt abgearbeitet. Die Microsoft-Variante dieser MPP-Technologie hört auf den passenden Namen Parallel Data Warehouse (PDW) (Bild 4). PDW ist eine Appliance, die gemeinsam von Microsoft und führenden Hardware-Herstellern entwickelt wurde. In Deutschland wird PDW im Rahmen der Microsoft Hewlett Packard Frontline-Partnerschaft von HP angeboten.

Rein physisch betrachtet besteht ein PDW aus zwei Racks zu je 42 Höheneinheiten, die prall mit Rack-Servern und -Speicher gefüllt sind. Spannend an der Lösung ist, dass vollständig auf den Einsatz speziell entwickelter Hardware verzichtet wird. Stattdessen kommt bewährte und natürlich hochwertige Commodity Hardware zum Einsatz, deren Komponenten fein aufeinander abgestimmt sind. Auf den Rechenknoten arbeiten normale SQL Server. Eine zusätzliche Softwareschicht kümmert sich um die optimale Verteilung der Daten und das Finden eines idealen parallelen Ausführungsplans für die Abfragen.

Im Hewlett Packard PDW bestehen die Knoten dieses Shared Nothing-Clusters aus Servern der Serie DL 360 G6 und für den Storage werden normale FS2000 SAN-Einheiten eingesetzt. Intern sind die Knoten durch Dual Infini-

o Hewlett Packard und Microsoft: Strategische Allianz

Ingo Schreiber, Microsoft GmbH, zu der Zusammenarbeit von Microsoft und Hewlett Packard: „Microsoft und HP haben Anfang 2010 eine umfassende weltweite „Infrastructure-to-Applications“-Vereinbarung angekündigt. Ziel ist es, gemeinsam Komplettlösungen für die IT Infrastruktur aus einer Hand anzubieten. Mit speziell aufeinander abgestimmter Hardware und Software, sowie zusätzlichen Produkten und Service-Angeboten, werden Kunden in die Lage versetzt, Zeit und Geld zu sparen, weil die gemeinsam entwickelten Komplett-Lösungen sich einfacher und schneller in die vorhandene IT integrieren lassen.“

HP und Microsoft betreiben ein gemeinsames Kompetenzzentrum in Böblingen. Dort kann Parallel Data Warehouse live erlebt werden und es ist möglich Proof of Concepts mit eigenen Daten durchzuführen. Interessenten wenden sich an Ralph Kemperdick (ralph.kemperdick@microsoft.com) oder Stefan Weber (stefan.weber@hp.com).

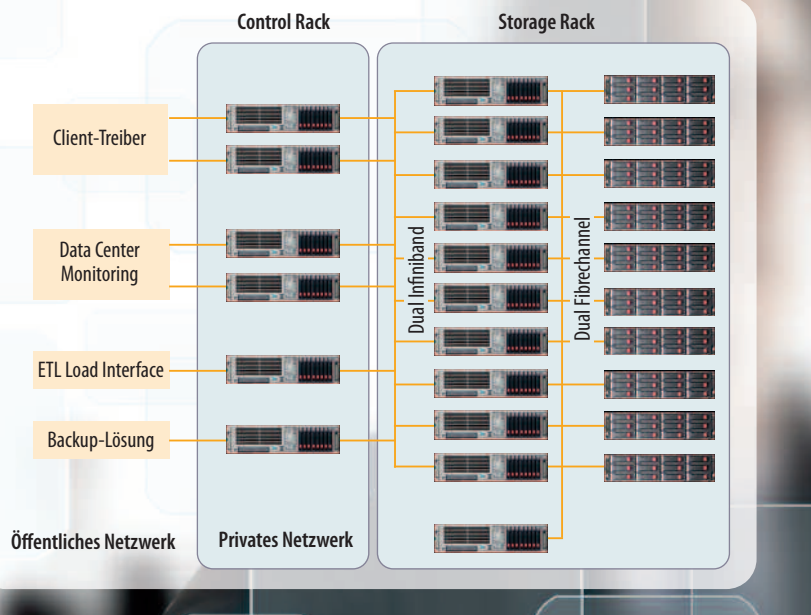


Bild 4: Parallel Data Warehouse Architektur

band miteinander verbunden. Die Startversion hat eine Kapazität von etwa 91 Terabytes und kann durch zusätzliche Daten-Racks auf bis zu 500 Terabytes ausgebaut werden.

Fast Track Data Warehouse: Eine Nummer kleiner

Nicht jeder benötigt gleich ein Data Warehouse, das sich in den Multi-Terabyte-Bereich skalieren lässt. Wenn man auf der Suche nach einer einfach zu implementierenden und dennoch sehr leistungsfähigen DWH-Plattform ist, dann könnte „Fast Track DW“ eine Möglichkeit sein.

Fast Track ist „Appliance Like“, aber eine klassische, SMP basierte Lösung. Fast Track adressiert, genau wie auch PDW, die im Artikel „Auf dem Vormarsch“ in diesem Heft von Dr. Martin beschriebenen Performanceprobleme großer Data Warehouses. DWH-Abfragen sind häufig „I/O bound“, während die CPUs noch reichlich Kapazität hätten, gehen Storage und Datenbusse in die Knie. Fast Track begegnet dem durch eine I/O-optimierte Konfiguration auf der Basis von Microsoft Referenzarchitekturen.

Es werden exakt dieselben Hardware-Komponenten eingesetzt, wie sie in den Rechenknoten der PDW-Appliances vorkommen. Fast Track ist bereits ab wenigen Terabyte Warehouse-Volumen interessant und skaliert von vier bis

circa 50 Terabyte. Weitere Informationen in den Webtipps.

BI Burger auf der Raststätte

In diesem Artikel wurden die verschiedenen Facetten der analytischen Datenbanken im Microsoft Technologie-Stack aufgezeigt: Von der „persönlichen“ Analysedatenbank PowerPivot für Excel, über die Analysis Services, den Spaltenindizes in der relationalen Datenbank bis hin zu den großen DWH-Appliances.

Denken Sie an diese Möglichkeiten, wenn Sie Datenanalysen in Ihrem Unternehmen vorbereiten wollen und denken Sie an den BI Burger, wenn Sie demnächst an der Raststätte einen Stopp einlegen.

GEORG URBAN, TECHNOLOGIEBERATER
BEI DER MICROSOFT GMBH

Info: Webtipps

- [1] Gartners BI Studie 2011 und weitere Informationen zu Microsoft BI: www.microsoft.de/bi
- [2] Microsoft Website zum Thema Data Warehousing und Appliances: www.microsoft.de/sql
- [3] Mehr über PowerPivot: www.powerpivot.com

Interview mit Thomas Strehlow

Datenana

Beim Mobilfunkkonzern E-Plus Gruppe wurde innerhalb von zwei Jahren die überalterte BI-Lösung durch ein modernes System auf der Basis des Microsoft Technologie-Stacks eingeführt. Thomas Strehlow, Geschäftsführer der ORAYLIS GmbH stand uns für ein kurzes Interview zur Verfügung.

? Herr Strehlow, was waren die Besonderheiten Ihres BI-Projekts bei E-Plus?

! **Thomas Strehlow:** Mobilfunk ist ein sehr dynamisches Geschäft. Ständig werden neue Produkte erfunden. Bei E-Plus können sich die Kunden beispielsweise „Mein Base“-Tarife sehr individuell aus einer Vielzahl von Optionen selbst zusammenstellen. Außerdem werden die Datenquellen ständig weiterentwickelt. Diese Agilität muss das BI-System mitmachen.

Glücklicherweise hatten wir es hier mit einem sehr reifen und emanzipierten Kunden zu tun. Das ist eine wichtige Voraussetzung für das Gelingen von BI-Projekten.

? Welche Funktionen deckt die neue BI-Lösung ab?

! **Thomas Strehlow:** Es geht um verschiedenste Aspekte des Wholesales, des Marketings und des Vertriebs. Das Datenmodell wurde in enger Zusammenarbeit mit dem E-Plus Controlling entwickelt. Wichtig ist vor allen Dingen die integrierte Sicht über unterschiedliche Themenbereiche hinweg. In

lyse bei E-Plus



der neuen Analysedatenbank werden Informationen einerseits hoch aggregiert vorgehalten, andererseits gibt es auch Unmengen von Detailinformationen: Von Verträgen bis hin zu jeder einzelnen Geschäftsaktivität. Das ermöglicht ein blitzschnelles Ausliefern von Kennzahlen über die aggregierten Werte und gleichzeitig ein Drilldown bis hin zur feinst möglichen Granularität.

Q *Wieso hat sich Ihr Kunde E-Plus für das Microsoft Tool Set entschieden?*

! **Thomas Strehlow:** Zum einen passen die Microsoft Tools sehr gut in die bestehende IT-Landschaft. Zum anderen wird in diesem Projekt ausgesprochen hohe Performance benötigt, um die Akzeptanz bei den Usern herzustellen. Die Anwender benötigen Online Analysefunktionen, um Ihre Fragestellungen selbst beantworten zu können. OLAP ist aber nur dann sinnvoll, wenn die Anwender keine Wartezeiten spüren. Die Analysis Services sind ein sehr ausgereifter und flexibler Analyseserver. Ein Anwender berichtete uns, dass er sich vorkommt, als wäre er vom Armaturenbrett eines VW Käfers in das Cockpit eines Düsenjets gewechselt.

Q *Über welche Frontend-Tools werden die Informationen an die Fachanwender verteilt?*

! **Thomas Strehlow:** Die Standardberichte werden von den eingebauten Reporting Services des SQL Servers ausgeliefert. Diese bringen alle notwendigen Funktionen bereits mit. Zusätzlich verwenden die OLAP-Nutzer den ihnen vertrauten Executive Viewer von IBM. Die Analysis Services lassen sich sehr gut mit unterschiedlichen Frontends „verheiraten“.

Q *Welche Vorteile bringt dem Kunden der Einsatz der Fast Track Appliance in einem Analyse-Projekt?*

„Mit den SQL Server
Analysis Services
können wir gnadenlose
Performance erzielen.

Die wenigsten Abfragen
dauern länger als
zwei Sekunden.“

Thomas Strehlow
Geschäftsführer bei der
ORAYLIS GmbH

! **Thomas Strehlow:** Die Daten des BI-Systems stammen aus verschiedenen Quellsystemen. Für das Zusammenführen und die Bereinigung wird ein klassisches Data Warehouse eingesetzt. Fast Track vereinfacht den Einsatz der Server. Die Administratoren müssen sich nicht mehr um die Konfiguration und das Tuning kümmern.

Q *Bitte geben Sie uns ein paar Informationen zum Sizing der Lösung.*

! **Thomas Strehlow:** Zurzeit befinden sich 10 Milliarden Faktendaten im System. Es werden Informationen von 20 Millionen Telko-Kunden vorgehalten. Die Analysen werden von 600 Anwen-

dern genutzt. Der Analyseserver muss pro Tag 2,5 Millionen Abfragen beantworten.

Q *Sagen Sie zum Abschluss noch ein paar Worte zu Ihrer Firma!*

! **Thomas Strehlow:** ORAYLIS gibt es seit dem Jahr 1999 und wir haben uns auf das Thema BI für große Unternehmen und dem Mittelstand spezialisiert. Wir sind Microsoft Gold-Partner für BI und Data Plattform.

Herr Strehlow, vielen Dank für das Gespräch!

Fakten zur Plattform

Das Data Warehouse und die Analysis Services laufen auf drei Hewlett Packard Fast Track-Maschinen:

- DL585 G6 (je 4 * Sixcore CPUs, 64 GB RAM)
 - DWH-Datenvolumen derzeit: 2 TB
 - Drei Analysis Services Cubes, mit je bis zu 300 GB Größe
- Zur Beschleunigung der Analysen wird eine Fusion-io-Karte eingesetzt.

