

# Speed, Scalability, and Flexibility All at Once with New High Performance Analytics

by Andrew Ross, SAP AG

SAP NetWeaver 2004s<sup>1</sup> Business Intelligence (BI) is faster than ever before, accelerated by a massively boosted capability for high performance analytics (HPA) humming behind the familiar BI frontends.

Before HPA, most approaches to accessing BI data confronted IT staff with a maintenance challenge. Administrators had to study user behavior to target frequently asked queries, then build summary tables and database indexes for those queries. This was skilled work and hence a major cost driver. Response times were improved only for the targeted queries, so good judgment was required to satisfy users. To keep this work within bounds, access to the data was often restricted by means of predefined reports for selected users. Now, HPA enables companies to relax such restrictions.

Running HPA, users can call up aggregated sales data for their company's operations, slice it and dice it, drill down anywhere for details, and expect exact responses in seconds or less every time, no matter how unusual their request —

even with billions of records filling terabyte volumes.

For the busy executive, the key benefits of SAP NetWeaver BI with HPA are:

- It's fast. Average query response times are 10 to 100 times faster than traditional approaches.
- It's flexible. HPA's on-the-fly aggregation greatly reduces administrative overhead, so there's no technical need to restrict user freedom.
- It's inexpensive. State-of-the-art hardware can be scaled exactly to your requirements, and falling hardware prices ensure that HPA's total cost of ownership (TCO) advantage over other tuning approaches will only grow in the future.

## Orders-of-Magnitude Improvements

To realize the boosted BI analytics capability, SAP developed a new 64-bit HPA engine. The new engine is based on an existing SAP search engine, but features a number of radical innovations at the technical level.

The results are scorching. With the new HPA engine, you can expect speeds that are not just incremental improvements, but orders of magnitude

better than what you could get from most other analytics software.<sup>2</sup>

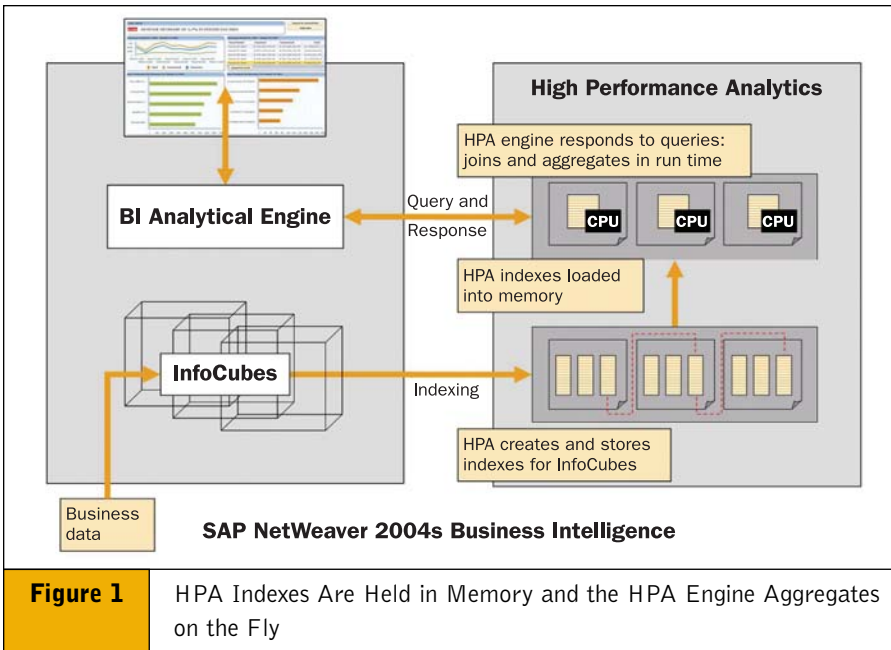
## Aggregation on the Fly

The HPA engine can aggregate key figures on the fly, during query runtime. It uses powerful and dedicated hardware resources to do all the work in memory, with no disk accesses to slow it down (see **Figure 1** on the next page). The engine takes the query entered by the user, computes a plan for answering it, joins the relevant column indexes to create a join path from each view attribute to the BI fact table, performs the required aggregations, and merges the results for return to the user.

This ability to aggregate during runtime is the decisive benefit of HPA. All the chores of prebuilding indexes and aggregates and realigning them regularly fall away. The realignment slots in nightly or weekly load windows are freed up, and administrators who previously did these chores can spend their time more productively. Users benefit from predictable response times, which encourage them to slice and dice their data in new ways and gain new insights. Freed from the old bottlenecks, companies can empower more users.

<sup>1</sup> SAP NetWeaver 2004s is the mySAP Business Suite edition of SAP NetWeaver 2004. It is a minor release that delivers on specific needs of the mySAP and xApps solutions delivered by SAP in 2005. An implementation of SAP NetWeaver 2004s is recommended only to customers requiring it for those solutions. However, some customers may wish to implement SAP NetWeaver 2004s in order to benefit from specific enhancements, such as high performance analytics.

<sup>2</sup> My team first demonstrated these performance improvements in late 2003, and later work has increasingly confirmed our expectations. For a high-level overview, see "Business Intelligence Best (and Worst!) Practices" by Peter Graf in the January-March 2005 issue of *SAP Insider* ([www.SAPinsider.com](http://www.SAPinsider.com)).



## Analytics Demo Wows Audience

At SAPHIRE 2005 in Boston, SAP Executive Board Member Shai Agassi showed off the new high performance analytics (HPA) capability delivered as part of SAP NetWeaver 2004s Business Intelligence (BI). From a BI analytic dashboard (sample screen shown in **Figure 2**), Agassi launched complex queries against a billion data records and got results back in seconds. He then showed how to compose new dashboards in minutes from a set of services with the help of a drag-and-drop tool called Visual Composer.<sup>3</sup>



**Figure 2** Sample BI Analytic Application Screen

For the demonstration, the HPA engine ran under 64-bit Linux on eight blades, each with dual Intel Xeon processors, mounted above a server running SAP NetWeaver 2004s BI, and a filer, all in a standalone cabinet.

<sup>3</sup> For more information on this drag-and-drop tool, see "Visual Composer — A Model-Driven Development Tool for Enterprise Portal iViews" by Karl Kessler in the October-December 2004 issue of *SAP Insider* ([www.SAPinsider.com](http://www.SAPinsider.com)).

## Vertical Decomposition of Tables

The HPA engine decomposes table data vertically, into columns that are stored separately (see **Figure 3**). This makes for more efficient use of memory space than row-based storage, since the engine needs to load only the data for relevant attributes or characteristics into memory.

This is a good idea for analytics, where most users want to see only a selection of data. In a conventional database, all the data in the table is loaded together, in complete rows, whereas the new engine touches only relevant data columns. The engine can also sort the columns individually to bring specific entries to the top. The column indexes are written to memory and cached as flat files. Efficiency is improved because both the memory footprint of the data and the input-output flows are smaller.

## Data Compression

Data for HPA is compressed using integer coding and dictionary lookup. Integers represent the text or other values in table cells, and the dictionaries are used to replace integers by their values during post-processing. In particular, each record in a table has a document ID, and each value of a characteristic or an attribute in a record has a value ID. An index for an attribute is simply an ordered list of value IDs paired with sets of IDs for the records containing the corresponding values (see **Figure 4**).

To compress the data, a variety of methods are employed, some of them highly innovative and covered by recent patent applications. Integer compression greatly reduces the average volumes of processed and cached data. This allows for more efficient numerical processing and smart caching strategies, which reduce the data volumes and flows by an average factor of 10. Hence all query processing can be performed in main memory.

## Horizontal Partitioning of Indexes

The HPA engine can partition large tables horizontally for parallel

processing on multiple machines in distributed landscapes (see **Figure 5**). This enables it to handle huge data volumes yet stay within the limits of installed memory. Formerly, large data volumes were kept on disk. Now the volumes can be split over multiple hosts by a round-robin procedure to build up parts of equal size, so that they can be processed fast and in parallel. A logical index server distributes join operations over partial indexes and merges partial results, all so smoothly that the index looks the same as ever to BI.

This scalability enables HPA to run on advanced computing infrastructures, such as blade servers. HPA runs on 64-bit platforms, where memory address space limitations are finally a thing of the past.

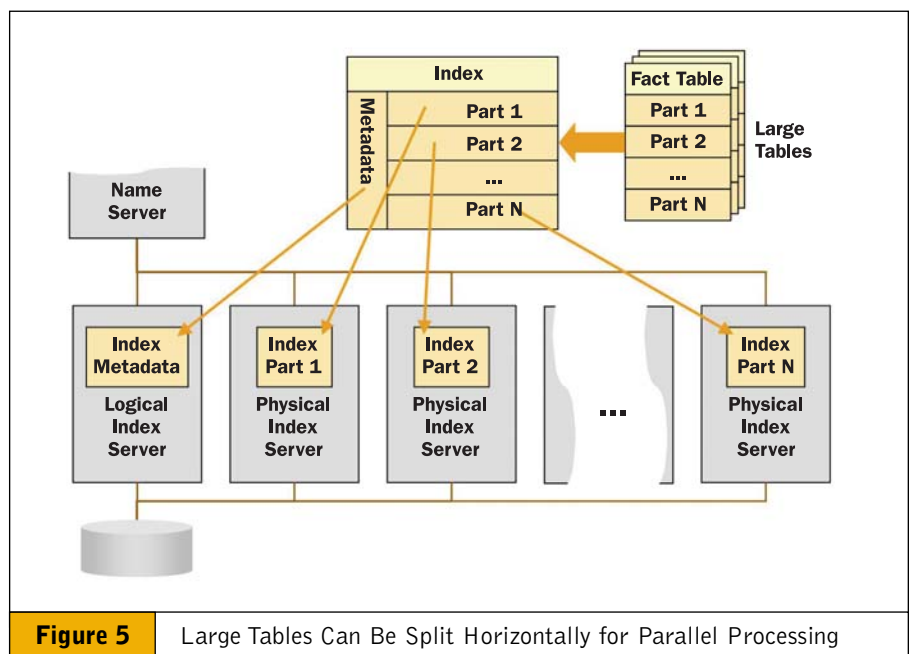
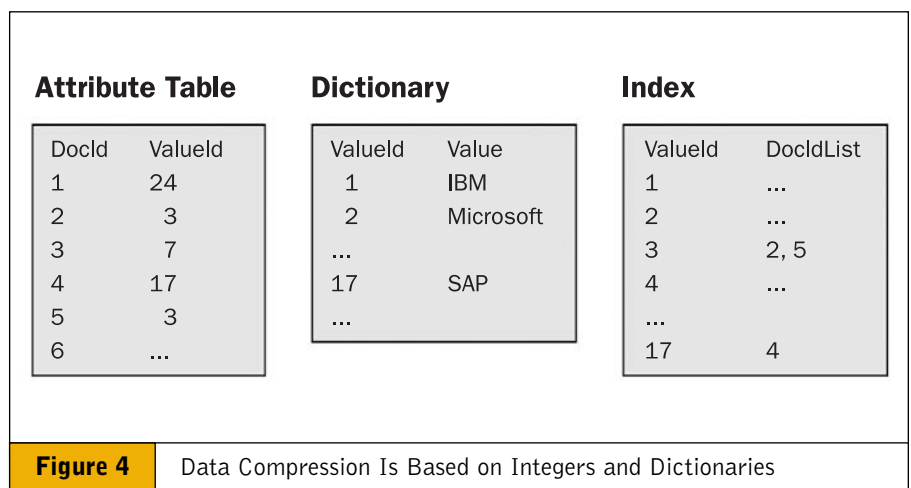
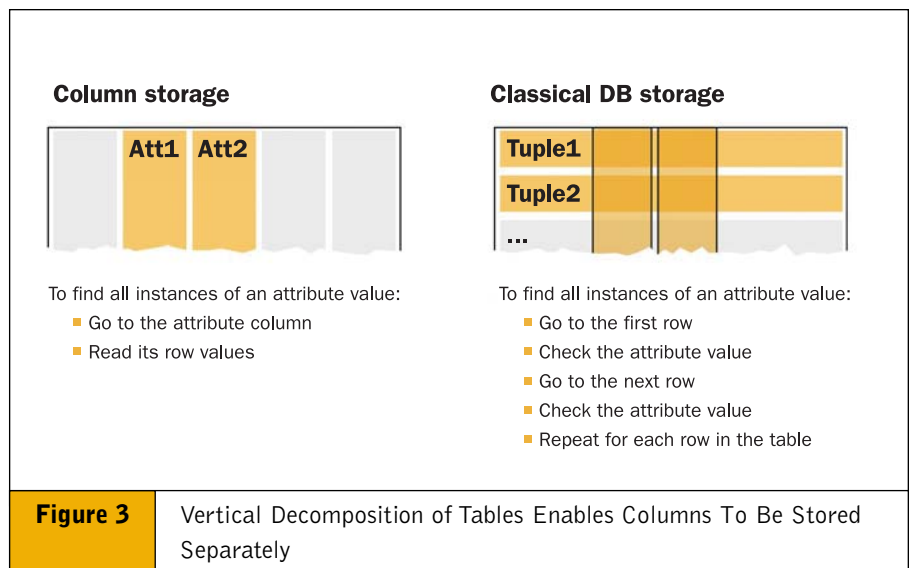
### Scalable Multiserver Architecture

The use of scalable and distributed search technology enables investment in hardware and network resources to be continually optimized to reflect changing availability requirements and load levels (see **Figure 6** on the next page). In each landscape, a name server maintains a list of active services, switches to backups where necessary, and balances load over active services.

Although the first customers will implement HPA on preconfigured hardware that they simply plug into their existing landscape, future customers will be able to configure new hardware dynamically. If the capacity is available, HPA will then run on existing hardware and scale rapidly to suit changing load. In an adaptive landscape, HPA instances will simply be replicated as required, by cloning services, and form groups with master and backup servers and additional cohort servers for handling query load. Such groups can be optimized for both high availability and good load balancing, with the overall goal of requiring zero administration.

### InfoCubes as Join Graphs

A BI InfoCube is a star schema for representing structured data (see



**Figure 7).** A large fact table is surrounded by dimension tables (D) and sometimes also X and Y tables. The S tables spell out the values of the integer IDs used in the other tables.

HPA models and algorithms are tailored and optimized to work well with InfoCubes. The HPA metamodel is designed to represent a star schema logically as a join graph, where joins between the tables forming the star schema are predefined in the model and materialized at runtime by the new query engine. The HPA metamodel bridges the gap between InfoCubes, which contain structured data, and search engine technology, which was originally developed to work with unstructured data.

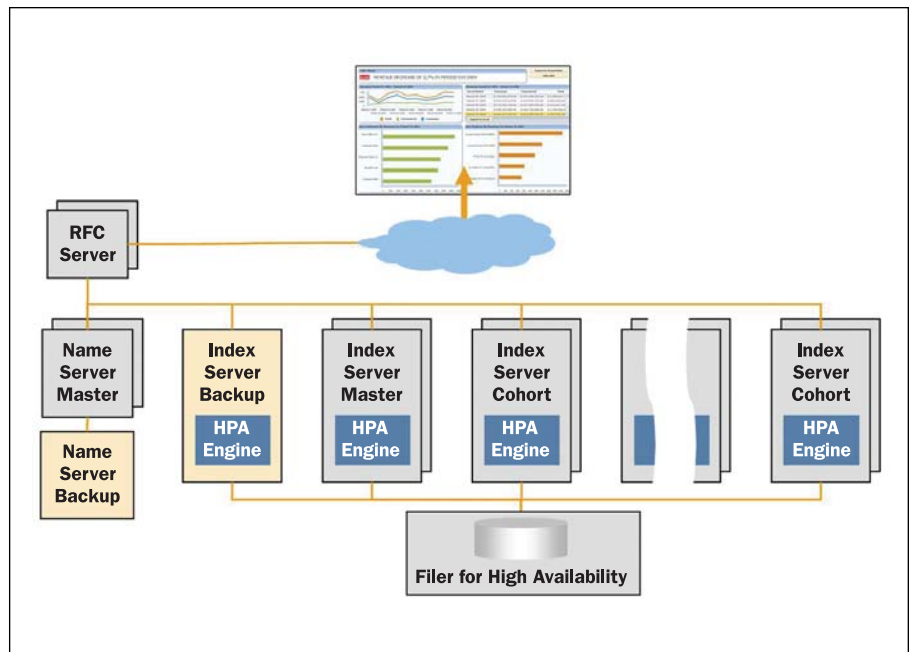
### Plug and Play

In summary, the technical benefits of SAP NetWeaver BI with the HPA capability are:

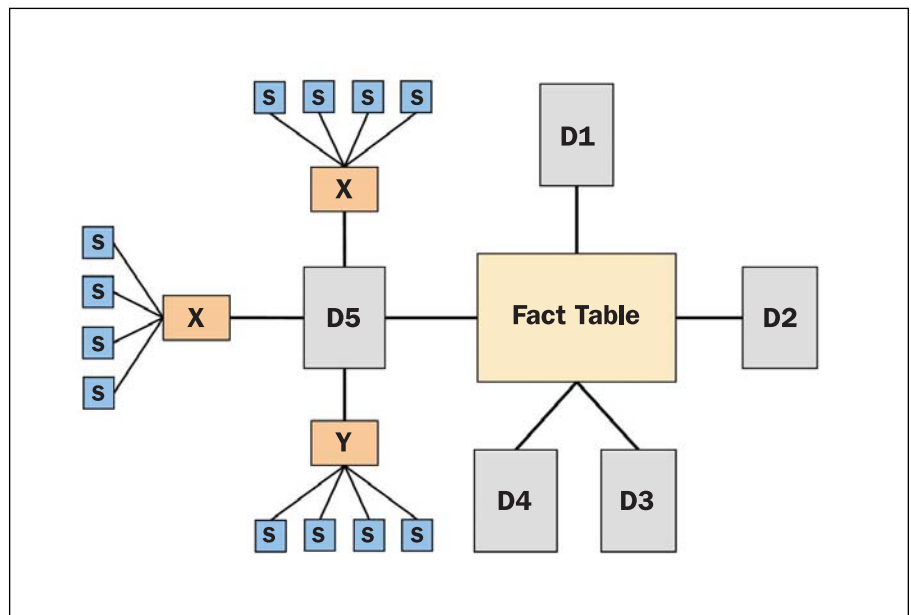
- *Speed:* The use of advanced compression and aggregation algorithms enables the new engine to achieve query response times that are on average about 10 times faster than previous approaches.
- *Flexibility:* The new approach greatly reduces administrative overhead. With HPA, there are no hunts for frequently asked queries, no pre-aggregations, and no realignment runs. So there is no technical need to restrict user freedom.
- *Ease of use:* To deploy HPA, customers can simply plug in the new hardware (available preconfigured in sizes from XS to L) and use it through their SAP NetWeaver 2004s BI frontends.<sup>4</sup>

For more information on HPA and other new or enhanced capabilities in

<sup>4</sup> During the Ramp-Up phase (six months starting October 2005), HPA will only be available on preconfigured hardware (as a set of blade servers that can be plugged into an existing rack or cabinet in the customer landscape). Unrestricted shipment begins in Q2 2006.



**Figure 6** The High Performance Server Architecture Is Highly Scalable



**Figure 7** The Standard BI InfoCube Is Represented as a Join Graph

SAP NetWeaver BI, just contact your local SAP sales office, and check the latest news at [www.sap.com/solutions/netweaver/businessintelligence/index.epx](http://www.sap.com/solutions/netweaver/businessintelligence/index.epx). ■

Andrew Ross is a developer in the SAP NetWeaver AS TREX team, which developed the HPA capability in collaboration with SAP

NetWeaver BI. He became part of the TREX team in 2003. He joined SAP in 1999, working at first in SAP Active Global Support. Earlier he was a computer science editor at a publisher in Heidelberg. He is British and holds four degrees in technical philosophy, three from Oxford and one from London.

