



Business Intelligence – HP Integrity server reference configurations for Microsoft SQL Server 2008

Introduction	3
Microsoft Windows on HP Integrity servers	3
Performance challenges	3
Benefits of using the reference configurations	4
Multi-tier business intelligence architecture	5
ETL tier – Extraction, Transformation and Load (SQL Server Integration Services)	5
Data Warehouse tier – ROLAP (Relational Online Analytical Processing) – SQL Server	6
Data Mart tier – ROLAP (Relational Online Analytical Processing) – SQL Server	6
OLAP cube tier – (MOLAP – Multi-dimensional Online Analytical processing) – SQL Server Analysis Services	6
End users – Portals, data access and analysis tools	6
Why are they called “logical tiers” ?	7
Conformed data warehouse and data marts	8
How to determine the correct hardware configurations to support your business intelligence workload	9
HP BI reference configuration solutions for SQL Server	11
Data warehouse/data mart tier – queries accessing a ROLAP model – reference configurations	11
How to determine which data warehouse/data mart reference configuration is best for you	14
ROLAP – Low throughput (42 queries per hour) – reference configurations	15
ROLAP – Medium throughput (168 queries per hour) – reference configurations	16
ROLAP – High throughput (505 queries per hour) – reference configurations	17
ETL tier (SQL Server Integration Services) – reference configurations	18
Load/insert rates	18
Real-time updates to data warehouse (ROLAP)	20
Index builds – bulk load vs. row-at-a-time inserts	20
Complex business logic	20
Table look-ups	21
Do you need to add hardware to support SQL queries that perform table “look-up” operations during the ETL batch window?	21
Using the HP BI SQL Server sizing tool to estimate table look-up overhead	22
Effect of data compression on ETL processing in SQL Server 2008	22
OLAP tier (SQL Server Analysis Services) – reference configurations	22
Which OLAP reference configuration will support your query workload and meet your user service level agreements?	22
Example – to determine which OLAP reference configuration is best for your cube size and workload ..	24
Consolidation architecture (3-tier)	26
Next steps	26
Summary	27

Appendix – SQL Server 2008 setup.....	28
Logical drive configuration	28
Location of SQL Server binaries	29
System databases location	29
Setup transaction log.....	29
Setup tempdb	29
Memory allocation for OS.....	29
For more information.....	30

Introduction

Microsoft Windows on HP Integrity servers

IT has undergone a huge transformation in recent times. Businesses today are no longer looking at IT to support daily business operations – rather, they're looking at IT to run the business. Like many IT managers, you're probably feeling the pressure to improve operations, cut costs, and respond to a level of unprecedented change – all at the same time. You may be looking to consolidate or enhance your system's scalability so that you can do more with less. You might be looking for more processing power to better manage memory and I/O intensive applications or to fuel the increasing demand for business intelligence capabilities. Or, maybe you're feeling constrained by expensive proprietary offerings and want a standards-based operating environment that's just as reliable but more cost-effective and flexible. HP Integrity servers running Microsoft® Windows® Server and Microsoft SQL Server can help you meet these demands with a unique combination of mainframe-level reliability, scalability, and performance – and all the benefits of a standards-based solution. With Microsoft Windows Server and SQL Server on the HP Integrity line of servers, you have the reliability and availability necessary for your mission critical needs as well as scalability for your most demanding business intelligence queries.

HP Integrity servers are mission-critical. Integrity is not just a brand – it is a way of doing business through a tailored approach that encompasses services, technologies, and partnerships to consistently deliver predictable results in an uncertain world.

Integrity servers are trusted servers for scalable, mission-critical operations. From blades and entry-class servers all the way to Superdome servers, they provide a broad range of customer choice that is backed by world-class HP services and support.

Integrity's intelligently virtualized, standards-based environment can scale dynamically and provide continuous availability. For a mission-critical infrastructure flexible enough to adapt to new business needs, the answer is simple: HP Integrity – most trusted by customers.

Target audience: The target audience for this document consists of IT planners, architects, DBAs, CIO, CTO and business intelligence (BI) users with an interest in options for their BI applications and in the factors that affect those options.

Performance challenges

IT planners and architects have a wide range of technology options for their business intelligence (BI) applications. Each hardware platform decision requires them to research, and then balance, the tradeoffs between price, performance, long-term scalability, manageability, and vendor reputation. When rapid deployment is a key factor, the choice often spans dedicated BI hardware, rigidly prescribed single-vendor configurations, and building custom solutions from well-understood technologies. Such a range in choices often means complex, time-consuming, decision cycles that drain staff time in order to define and manage the benchmarks or proof-of-concepts. All of this takes away from the basic function of IT, which is to deliver information quickly to business users so that they can make sound, timely decisions.

In order to determine the appropriate hardware configuration to support an “assumed” workload, HP has developed several reference configurations which may be found in the matrices later in this document. The most significant system components in each reference configuration are:

- Server model number (for example, HP Integrity rx8640, etc.)
- Number of sockets and cores per server (for example, 1 CPU in a socket may contain multiple cores)
- Memory per server

- Fibre Channel HBAs per server to maintain I/O throughput requirements for a given query workload (*)
- Storage array model number and number of disk drives required to support each respective query per hour rate (*)

(*) NOTE: HP's reference configurations assume that most customers prefer HP StorageWorks Enterprise Virtual Array (EVA) or XP Disk Array SAN technology due to fibre channel's ease of management and high availability features in a very large database (VLDB) business intelligence environment. Should a customer prefer HP StorageWorks Modular Smart Array (MSA) (direct attached) storage, the SQL Server 2005 sizing tool will be able to determine an appropriate MSA configuration to meet the I/O throughput requirements based upon the assumed queries per hour (QPH) rate, query complexity, mix and database size.

Should a customer want a more customized configuration, which more precisely matches their unique load/insert rate, query complexity, query per hour rate, workload mix and database size, then the customer can run the HP Business Intelligence Sizer for Microsoft SQL Server 2005, which will model the unique requirements and generate a proposed hardware configuration. (See the *For more information* section for a link to the sizer.)

Your HP sales representative will be able to generate a current Bill of Materials (BOM), with the most recent part numbers and pricing based upon each reference configuration or sizing tool output report.

When you look at the hardware costs for a data warehouse, you will find that the bulk of the costs are related to storage. For example, it is not likely that a 1 TB data warehouse will perform well on only five 300 GB disks (RAID5). Therefore, it is important to insure that there are enough disks, controllers and Host Bus Adaptors (HBAs) configured to meet each customer's unique service level agreements. By comparison, the server/processor costs are only a small percentage of a data warehouse. That is why many customers are making sure that their server selections are based upon supporting an ever increasing query volume, query complexity, data base growth, increased sophistication of user queries and availability. HP Integrity servers provide that extra level of comfort by providing superior scalability, Reliability, Availability, and Serviceability (RAS) features. These features are becoming more important as business intelligence applications become more mission critical and is even more important should consolidation be considered because more users are supported in a consolidated/virtualized environment.

Benefits of using the reference configurations

IT staff time and resources are precious (limited and costly). HP BI reference configurations and sizing tools enable IT departments to do more with less, more quickly, and at lower risk. Some immediate benefits for deploying data warehouse and BI solutions based on the BI reference configurations include:

- Simplified product selection and sizing process – by starting with a reference configuration, which may be implemented as is or used as a starting point to configure the final system. By either using the reference configuration or adding to it, line of business (LOB) and IT can more quickly decide and purchase the optimal configuration for the application.
- Simplified, rapid deployment with predictable performance using balanced configurations
- Off-the-shelf components to enable re-purposing and lower inventory risk
- Proven, award-winning support, with access to performance and tuning services

In summary, HP reference configurations for SQL Server 2008 Data Warehousing provide optimal BI performance and enable customers to avoid common design pitfalls. Preliminary tests (in July 2008) show that SQL Server 2008 pre-release code, has performance improvements that were not available

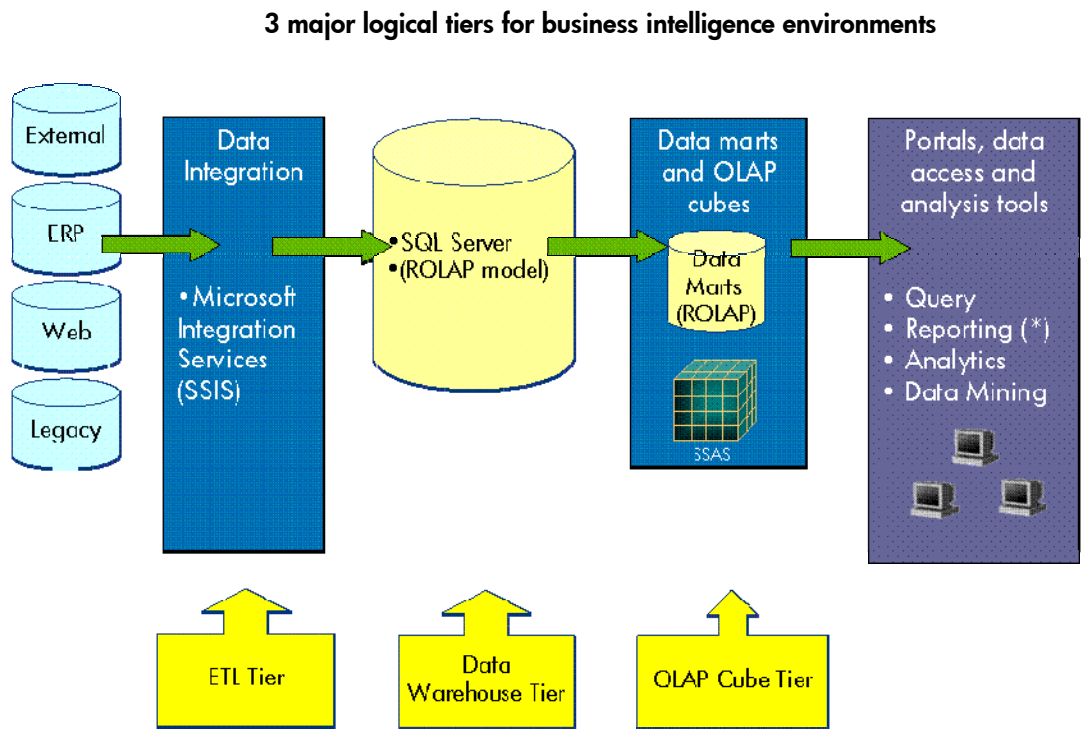
in SQL Server 2005. Therefore, when SQL Server 2008 is “generally available,” HP plans to re-run these reference configurations to take into account the new features.

Multi-tier business intelligence architecture

Reference configurations are available for all three tiers in a business intelligence environment:

- ETL Tier (ETL – Extraction, Transformation and Load – Microsoft SQL Server Integration Services (SSIS))
- Data Warehouse Tier (Relational Online Analytical Processing tier (ROLAP) – relational data warehouse or data mart)
- OLAP Cube Tier – Queries against the OLAP cube (OLAP – Online Analytical Processing – SQL Server Analysis Services)

Figure 1. Three-tier BI environment



ETL tier – Extraction, Transformation and Load (SQL Server Integration Services)

The ETL tier supports three main data migration functions:

- Extraction – This is typically the process of reading data from an online transaction processing (OLTP) database or Operational Data Store (ODS), which typically contains transactional data.
- Transformation – This process converts the extracted, operational data from its previous, normalized, OLTP structure into a form that uses a fact table(s) with conformed dimensions. The ETL software does this by remapping columns, using rules, look-up tables and sometimes combines the data with external sources (such as demographic and spatial data) to generate easier to use, meaningful information. Meta data is also, commonly, maintained for documentation purposes and to make sure that all users of the data warehouse have a consistent view of the information and how it may have been manipulated or derived in the multi-dimensional data warehouse structure.

- Load – the process of loading, inserting or updating the transformed data into the target data warehouse or data mart. The majority of time during the load is typically spent loading or inserting new data into the fact table(s). Occasionally, dimension tables are updated, but, the volume of changes is usually very low. Index maintenance may also be a time-consuming process depending upon the size of dimension tables and fact table partitions.

Data Warehouse tier – ROLAP (Relational Online Analytical Processing) – SQL Server

Transaction processing systems traditionally contain real-time operational data to support the business's day-to-day operations. The data on these servers is usually short lived. In addition, OLTP data structures are usually stored in 3rd normal form, which is not optimal for servicing business intelligence queries.

In order for management to make good strategic decisions, this operational data is sometimes stored in an Operational Data Store or, more commonly, directly loaded into a data warehouse or data mart from the transaction processing servers.

The data warehouse is a collection of data designed to support the management decision-making process. Traditionally, the ETL tier manipulates and transforms operational, demographic and other data to be stored in a multi-dimensional, star or snowflake scheme in the warehouse. These database structures, when coupled with meta-data and conformed dimensions, present a coherent picture of business conditions over time.

Data Mart tier – ROLAP (Relational Online Analytical Processing) – SQL Server

Data marts and data warehouses are similar in that they both store historical operational data in a multi-dimensional, star or snowflake scheme. The primary difference is that data marts tend to address the business needs of a smaller niche of business users. Some data marts may be very large, so the mart's vertical business niche is generally a more defining trait of a data mart, rather than size, when compared to a data warehouse.

Configuring ROLAP data marts is the same as sizing ROLAP data warehouses. Query volume, complexity and mix, coupled with the data warehouse size are the most significant factors in determining data mart resource consumption. Therefore, the same reference configurations may be used for ROLAP data marts.

OLAP cube tier – (MOLAP – Multi-dimensional Online Analytical processing) – SQL Server Analysis Services

OLAP cubes are very similar to data marts. However, rather than being implemented by using flat SQL tables in a multi-dimensional, star or snowflake scheme, the OLAP cube is an n-dimensional structure, which is, ideally, stored in memory

OLAP cubes generally give end users significantly faster response times because the data is summarized, aggregated and performs best when the frequently accessed cells of the cube are memory resident. However, this more rapid response time is time "paid for" by consuming hardware resources during cube build and aggregate creation time.

SQL Server Analysis Services also has data mining and analytical enhancements which make it easier for the business analyst to use to make decisions. Some of these operations include: slice, dice, drill-up, drill-down, roll-up, pivot, etc.) OLAP cubes also allow for hierarchies, which are more difficult to describe using traditional, ROLAP, SQL tables.

End users – Portals, data access and analysis tools

Business analysts typically have access to graphical, user friendly tools to perform analytics, data mining, or simply to generate reports and execute queries. Executive dashboards are also commonplace to provide flexible visual access to the information in the data warehouse, data mart or OLAP cube.

Examples of end user tools are: Microsoft PerformancePoint Server, Office Excel, SQL Server Reporting Services, etc.

Why are they called “logical tiers” ?

Many organizations have many data warehouse and data marts that have evolved over time. Each one of these environments, commonly, has a physical server for their ETL tier, ROLAP tier (data mart or warehouse) and/or OLAP cube(s) tier.

In any given department, it is likely, that ETL batch jobs load data into a database with a multi-dimensional (ROLAP) schema. Then, after the load, the ROLAP (data mart or warehouse) is used to build (optional) “front-end” OLAP cubes after the data is loaded into the ROLAP database.

Examples of queries accessing data in a multi-tier architecture:

End user query tools can typically extract information directly from the ROLAP data warehouse/mart or from OLAP cubes by pointing their query tool to these different multi-dimensional database structures. Hence, some customers do not bother implementing OLAP cubes and access information directly from the ROLAP data warehouse or data mart. Still other customers opt for a HOLAP model which is a “hybrid” environment, which allows users to drill through OLAP cubes and access ROLAP model data if the low level data is not stored in the cube.

From a performance point of view, in the real world, these physical server “tiers” typically have different peak time windows:

- Evening (run ETL and load/insert data into the ROLAP data warehouse/mart) – ETL tier and ROLAP tier are busy.
- Later that evening (build (optional) OLAP cubes) – OLAP cube and ROLAP tier are busy.
- During the day, users query the ROLAP data warehouse/mart or, optional, OLAP cube.

As you can see, the ROLAP data warehouse is the busiest server tier because it is active during the ETL window, cube build window and during the day when users query the database. It should be noted that it is, generally, a good idea to front end the ROLAP data warehouse with an OLAP cube or data mart. These front end tiers will alleviate the users’ query workload on the data warehouse during the day. Another benefit of OLAP cubes is that they provide significantly better response time because the cube should be memory resident and the aggregates managed by the cube dramatically improve performance.

By looking at the above data flow, you can see that it is common for one or more of these tiers (ETL, data warehouse/mart, OLAP cube) to be idle during different parts of the day. So, in order to reduce licensing, electrical, floor space, management and other costs, many companies are opting to consolidate these “logical” tiers onto a single HP Integrity, “scale-up”, server.

Consolidating multiple servers within a single department and/or consolidating tiers from multiple departments are becoming commonplace. This type of hardware consolidation may result in considerable savings for the business. This document assumes that ETL, ROLAP and OLAP workload windows are not competing for system resources even if they all functionally execute on the same physical server. Of course, virtualization software, nPars (hard partitions), Soft NUMA (nonuniform memory access – soft partitions), WSRM (Microsoft Windows Server Resource Manager), Resource Governor in SQL Server 2008, etc. may be used to manage competing workloads on a single server. But, as of June 2008, the BI ATC (Business Intelligence Advanced Technology Center) has not developed a sizing model for competing functional tiers on a single server. Further discussion of consolidated architectures and designs are beyond the scope of this document, but, customers may want to consult with HP directly when considering consolidated architectures.

One final thought...from a business point of view, it should be noted that additional advantages may be obtained by implementing a consolidated, conformed data warehouse environment. A conformed

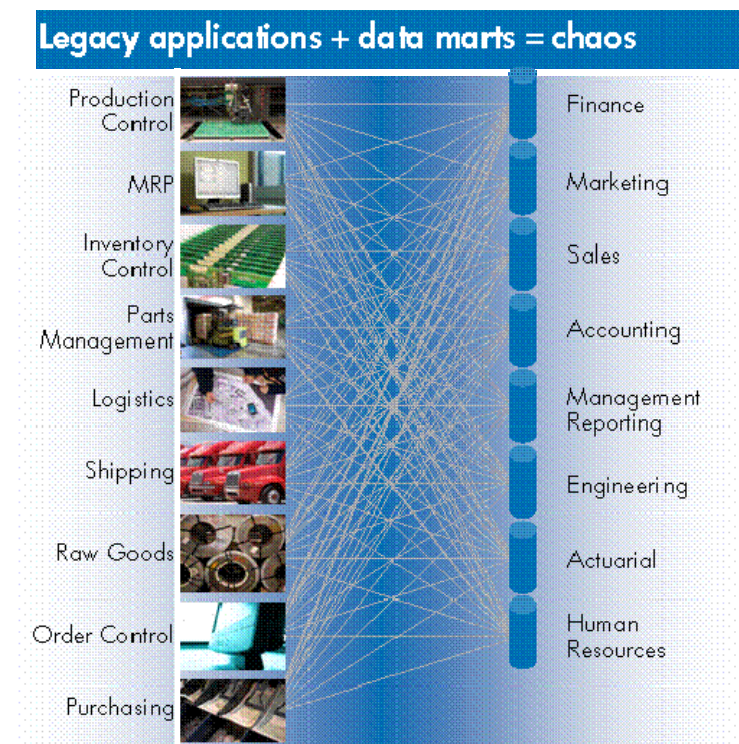
data warehouse uses “conformed dimensions.” This means that users throughout your organization agree upon dimensional terminology. This is frequently coordinated by using metadata or master data management techniques. The advantage of using conformed dimensions is that the business analysts and management will obtain a “single version of the truth” because people and the data warehouse, data marts and OLAP cubes are all using consistent terminology.

Conformed data warehouse and data marts

Historically, different departments typically developed their own business intelligence environment based upon their department’s own business needs. This evolutionary design tended to result in each department’s ETL tool deriving or transforming a data element’s meaning in such a way that the same data may mean different things to different departments depending upon the data mart accessed.

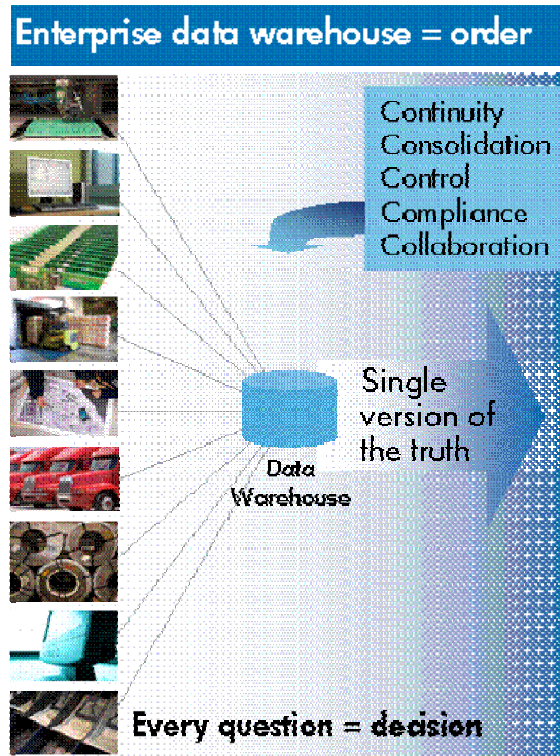
If upper management wants to access information across these data marts, it is likely that they will obtain inconsistent results because of the different use of terminology and data transformation at the departmental level.

Figure 2. The chaos caused by legacy application and data marts



A conformed data warehouse resolves these problems by using conformed dimensions. Developing a conformed data warehouse is more of a political challenge within a company than a technical challenge. Users and management must agree to use certain terminology and transformation should be standardized for commonly used functions. These corporate “standards” are then maintained in metadata and/or by using Master Data Management techniques.

Figure 3. Applying company standards to metadata or data management



Bear in mind that an organization still has the option of saving money by consolidating hardware as a separate project. Migrating to a conformed data warehouse environment may be implemented as a separate initiative, separate and distinct from hardware consolidation.

How to determine the correct hardware configurations to support your business intelligence workload

Historically, some people used to determine configurations by using a, so called, "sizing dashboard," or they simply configured a server based upon the budget a customer may have had available. These techniques will, most likely, result in a mis-configured server, which will not meet user service level agreements. HP advocates the following methods as being the most practical approaches to sizing servers to support a given BI workload:

- **Reference configurations.** This approach is addressed as the main subject of this document. Using reference configurations is fast, easy and provides a company with configurations which balance CPU processing power with I/O bandwidth capabilities. The reference configurations tend to have more than enough physical storage space to store the database because I/O throughput tended to be the limiting factor based upon the query workloads assumed in each scenario.
- **Sizing tools.** While reference configurations are useful in determining a "reasonable" configuration to support an "assumed" workload with an associated database size. It is very likely that a customer's "real world" workload will not match the workload that was assumed for the reference configurations. In this case, HP has provided an SQL Server BI sizing tool, which may be used to generate balanced configurations for workloads that a customer defines to the tool.

Bear in mind that the HP BI SQL Server sizing tool is not a “reference configuration selector.” It takes workload specifications and uses “atomic modeling techniques” to generate a unique configuration to support the workload defined to the tool. Sizing tools which do not use modeling technology simply “cross reference” into a matrix of reference configurations and guess which reference configuration might do the job. But, HP’s sizing model creates a unique configuration based upon each unique customer’s workload requirements and database size.

Note: The current HP SQL Server BI Sizing Tool can model the ETL tier (Integration Services), the data warehouse/data mart tier (ROLAP model), and queries against OLAP cubes (SQL Server Analysis Services).

HP Business Intelligence Sizer for Microsoft SQL Server 2005: (**Note:** you have to log on to ActiveAnswers, which has free registration)
<http://h20331.www2.hp.com/activeanswers/Secure/351481-0-0-0-121.html>

- **Benchmarks.** Years ago, companies used to perform benchmarks to find out if a given hardware configuration will support a given workload. The problem with benchmarks is that the benchmarked workload rarely, if ever, matches what real users will do. In addition, benchmarks tend to be very expensive because they are very people-, time-, and resource-intensive for all parties involved. HP has practical alternatives to custom benchmarks. Hardware sizing is now performed by using balanced reference configurations or by using the HP BI SQL Server sizing tool which models most business intelligence workloads. In addition, others have tried to use industry standard benchmarks (such as, TPC-H) to perform performance comparisons. Since companies do not run TPC-H queries in their production business intelligence environment, this type of comparison may not be meaningful. Therefore, pilots/prototypes are sometimes performed, which more directly engage business users.
- **Pilot/prototypes.** Unlike a benchmark, which tests unrealistic workloads, a pilot/prototype is the best way to prove that a given hardware configuration will support a “real world” workload providing users with their expected response times. Pilots/prototypes are more accurate because they involve the end user. Real data is loaded into the data warehouse/data mart, and end users execute real queries which they would expect to perform in their day-to-day operations. The hardware used for the pilot/prototype is selected by using one of the reference configurations or, better yet, by using a sizing model of what the customer thinks their workload will look like. Unlike OLTP applications, business intelligence is interesting because user queries will change as they gain experience and realize the power of the data warehouse in making business decisions. This is another reason why a pilot/prototype is superior to a benchmark. Since real users are involved with the process, HP can gain a thorough understanding of the workload that users will place on the system and determine future hardware requirements.
- **Capacity planning.** This brings us to the capacity planning phase. Capacity planning techniques may be used to predict hardware requirements based upon production workloads or make projections based upon a customer’s pilot/prototype system.

The iterative process involved with capacity planning insures that users' service level agreements are maintained and are supported by the appropriate hardware configurations.

HP BI reference configuration solutions for SQL Server

To reduce the risk and deployment time for BI solutions, HP has created a comprehensive set of BI reference configurations. The HP BI reference configurations draw on extensive lab, customer benchmark, and industry standard benchmark investments as well as global experience supporting thousands of data warehouse and business intelligence customers. The configurations enable IT planners and architects to start from a "best-fit" reference configuration that reflects their companies' raw data sizes and query workload characteristics. Each configuration can then be optimized based on their specific workload and requirements. An exact list of parts is best obtained and verified with your local HP sales representative or authorized reseller to insure all components and licensing costs are accounted for.

Each configuration reflects "real-world" BI customer environments and workloads. HP BI reference configurations are modular and can easily scale up by adding CPUs, memory or storage on an as-needed basis. Most importantly, the HP reference configurations were optimized for BI performance by avoiding bottlenecks and by maintaining balanced I/O for their respective workloads. Achieving balanced system I/O throughput specifically designed to satisfy the requirements of BI workloads was a key objective in developing the reference configurations because it maximizes system BI performance while optimizing return on investment (ROI) through proper utilization of all the hardware components.

HP reference configurations deliver on the promise of open, flexible, IT solutions based on standardized technologies. Because HP strongly supports customer needs for flexibility and re-use options, HP BI reference configurations combine architectural choice with predictable performance and standards-based technology.

This document addresses reference configurations for the three main business intelligence logical tiers:

- Data warehouse/data mart tier (ROLAP model)
- ETL tier (SQL Server Integration Services)
- OLAP tier – Queries against OLAP cubes (SQL Server Analysis Services)

Data warehouse/data mart tier – queries accessing a ROLAP model – reference configurations

For convenience, these configurations are arranged into size groups, each scoped for a limited range of raw data sizes. The current Microsoft SQL Server reference configurations are available for 500GB, 1TB, 2TB, 4TB, 7TB, 10TB and 20TB raw data size configurations.

The database size (RAID5 was assumed) is commonly larger than raw data sizes, in some cases as much as 5x larger than the raw data size. This "expansion factor" can vary widely depending on customer requirements, aggregates, indexes, and summaries, causing the instance size to expand. In addition, each configuration contains space which was allocated for TempDB and log files (RAID1).

As stated above, each configuration defines a matched set of servers and storage, which balances I/O throughput across all components to yield consistently high BI performance. I/O throughput is recognized as one of the best indicators of potential BI query performance. While the HP Integrity server reference configurations were developed with HP StorageWorks Enterprise Virtual Array (EVA) or XP Disk Array storage in mind, it is possible to run the HP Business Intelligence Sizer for Microsoft SQL Server 2005 to determine the correct number of HP StorageWorks Modular Smart Arrays (MSA) to support any given user workload requirement. While the use of the MSA will reduce costs significantly, the Integrity reference configurations opted to use HP StorageWorks EVA or XP storage

area networks (SANs) because they provide customers with an easier-to-manage subsystem that has higher availability than MSA storage.

- [EVA4100/6100/8100](#) series – The HP StorageWorks 4100/6100/8100 series of Enterprise Virtual Arrays (EVAs) continues to serve customers in the mid-range to enterprise-size market place with leading high performance, high capacity and high availability "virtual" array storage solutions. Not only do these solutions reduce IT costs and complexity, they save time, space and costs as compared to traditionally architected storage, and they are supported by a powerfully simple suite of management software, making it easy for users to achieve high levels of productivity.

Within each raw data size group, HP Integrity server configurations are provided for "low throughput" (42 queries per hour), "medium throughput" (168 queries per hour) and "high throughput" (505 queries per hour) workloads. Instead of quoting the number of users which a configuration may support, HP recognizes that it is not the number of users which drives hardware utilization levels, it is the "query per hour" workload, which users generate, that affects system performance.

For example, the SQL Server "medium throughput" Integrity reference configurations were designed to support 168 queries per hour. As you may suspect, 168 queries per hour may be generated by 100 users, 25 users or 1,000 users. Hence the actual workload placed on a server is a function of each "typical" user's average "think time" between queries submitted to the server.

Note: HP tends to avoid the term "simultaneous users" because "simultaneous" may mean different things to different people. HP uses "queries per hour" to more meaningfully define traffic rates.

Another significant influencing factor, which needs to be considered in conjunction with the number of queries per hour, is query complexity (that is, simple, medium, complex and very complex). A good "rule of thumb", or guideline, is that a medium query is 1.5 times the cost of a simple query, a complex query is 5 times the cost of a simple query, and a very complex query is 25 times the cost of a simple query.

Last, but not least, you will notice that each classification of user executes a different percentage of simple, medium and complex queries.

The assumptions below (Figure 4) are reasonable assumptions for most customers. However, it is highly recommended that each customer interview its user base to gain insight regarding the expected workload. HP has provided the [HP Business Intelligence Sizer for Microsoft SQL Server 2005](#) tool in order for customers to customize their own configurations to meet their user requirements.

Figure 4 is a summary of the assumptions used for the Integrity reference configurations:

Figure 4. Assumptions used in the Integrity Reference Configurations

Input Fields	Total # Users	Simple	Medium	Complex	Very Complex
Normal users	21 thru 252	75%	24%	1%	0%
Advanced users	3 - 36	50%	48%	2%	0%
Power users	1 - 12	5%	50%	44%	1%
Realistic response times (secs)		60	600	1,200	3,600
Average think time (secs)		1,800	3,600	5,400	7,200
Estimated number of I/Os per query		Varies depending upon DB size (17K – 342K)	Varies depending upon DB size (21M – 417M)	Varies depending upon DB size (56M – 1,112B)	Varies depending upon DB size (700M – 14B)

It is important to note that reference configurations and sizing models are very useful tools for estimating hardware requirements based upon a pre-defined workload for new business intelligence applications where the expected workload is being defined and estimated by discussions between the IT department and the user base. However, if a customer has an existing data warehouse, data mart or pilot/prototype, HP strongly recommends using actual Perfmon data to determine the most accurate configuration required to support any given workload. Using real world performance data (Perfmon or HP Performance Manager software) is the most desirable approach when performing upgrades, consolidation or ports, as opposed to using any sizing tool or reference configuration.

The HP Integrity reference configurations for Microsoft SQL Server 2005 were cross checked against real-world observation of customer workloads. However, the reference configurations were primarily based upon over 100 benchmarks and 1,000 performance graphs/metrics to build and calibrate the sizing model. The actual queries used to derive the Integrity reference configurations were scaled up and down based upon an industry standard 1TB business intelligence benchmark. So, if a customer suspects that its workload contains fewer or more I/Os per query, the customer can calibrate the reference configurations by using the [HP Business Intelligence Sizer for Microsoft SQL Server 2005](#) or by using real Perfmon or HP Performance Manager data, if available.

It should be noted that the medium, complex and very complex queries that were profiled for the reference configuration benchmarks in the tables below (low, medium or high throughput), tended to be VERY I/O bound. It is possible that customers can improve performance by using materialized views, summary tables, more indexes, or other means to reduce the hardware required to support their user population.

How to determine which data warehouse/data mart reference configuration is best for you

1. Estimate the number of queries per hour that your user population will generate. This will tell you if you want to use the low, medium or high throughput tables below. Don't forget that real people do not submit queries back-to-back with zero think time between queries. Users will tend to:
 - Think about the question they want an answer to and formulate a theory
 - Perform a simple query to determine the correct dimensions they may want to use
 - Submit the query
 - Take time to analyze the results
 - Formulate a new theory and submit another query
 - Don't forget that real people also take time to talk to co-workers, get coffee, take lunch, go to meetings, and perform other tasks. This all increases "think time" between queries, which in turn, lowers the query rate per hour.
 - It is also likely that many "normal users" and some "advanced users" will probably execute "canned" queries, which were optimized for performance by using materialized views, summary tables, indexes, or were simply displays of a Microsoft Office Word file, PDF, or Microsoft Office Excel spreadsheet that was generated during off-hours.
2. After you estimate your query rate per hour, cross reference to the appropriate column in the tables below based upon database size. This column will summarize the system's critical components.
3. If your workload does not align with one of the 21 reference configurations below, HP advises that you run the [HP Business Intelligence Sizer for Microsoft SQL Server 2005](#) to obtain a customized configuration to support your specific workload.

ROLAP – Low throughput (42 queries per hour) – reference configurations

Table 1. Low Throughput configuration(s)

User data warehouse size	500GB	1TB	2TB	4TB	7TB	10TB	20TB
Architecture	SMP	SMP	SMP	SMP	SMP	SMP	SMP
HP Integrity Server models	rx3600 or BL870c	rx3600 or BL870c	rx7640	rx7640	rx7640	rx7640	rx8640
Nodes	1	1	1	1	1	1	1
Sockets/cores per node	1/2c	1/2c	2/4c	4/8c	6/12c	8/16c	16/32c
Memory per node	8GB	16GB	24GB	32GB	48GB	64GB	128GB
FC Cards per node DP = dual port	1x 4Gb/s	2x DP 4Gb/s	4x DP 4Gb/s	4x DP 4Gb/s	6x DP 4Gb/s	8x DP 4Gb/s	14x DP 4Gb/s
SAN Switch	SAN 4/16	SAN 4/16	SAN 4/32	SAN 4/32	SAN 4/32	SAN 4/64	SAN 4/64
Storage Array (number of disks)	EVA4100 (30)	EVA6100 (54)	2xEVA8100 (110)	2xEVA8100 (178)	3xEVA8100 (296)	4xEVA8100 (496)	1xXP12000 (488)
Raw Storage TB	2.1	3.8	7.9	12.8	21.3	35.7	35.6
Usable Storage TB	1.5	2.5	5	9	16	25	25
OS*	Windows	Windows	Windows	Windows	Windows	Windows	Windows
Estimated concurrent users**	25	25	25	25	25	25	25
Estimated queries per hour (QPH)	42	42	42	42	42	42	42

* Microsoft Windows Server 2003 for Itanium-based Systems

** 25 "Concurrent users" were used to generate 42 queries per hour (QPH). Actual customers may have hundreds of users who generate 42 queries per hour or 10 users who generate 42 queries per hour. Therefore, HP recommends that the QPH rate be used as the main guide vs. the number of concurrent users for each respective database size.

ROLAP – Medium throughput (168 queries per hour) – reference configurations

Table 2. Medium throughput configurations(s)

User data warehouse size	500GB	1TB	2TB	4TB	7TB	10TB	20TB
Architecture	SMP	SMP	SMP	SMP	SMP	SMP	SMP
HP Integrity Server models	rx3600 or BL870c	rx3600 or BL870c	rx7640	rx7640	rx8640	rx8640	HP Integrity Superdome
Nodes	1	1	1	1	1	1	1
Sockets/cores per node	1/2c	2/4c	4/8c	6/12c	8/16c	10/20c	20/40c
Memory per node	16GB	24GB	32GB	64GB	96GB	128GB	256GB
FC Cards per node DP = dual port	1x DP 4Gb/s	2x DP 4Gb/s	4x DP 4Gb/s	6x DP 4Gb/s	8x DP 4Gb/s	12x DP 4Gb/s	26x DP 4Gb/s
SAN Switch	SAN 4/16	SAN 4/16	SAN 4/32	SAN 4/32	SAN 4/64	SAN 4/64	3x SAN 4/64
Storage Array (number of disks)	EVA4100 (38)	EVA6100 (72)	2xEVA8100 (144)	3xEVA8100 (232)	4xEVA8100 (388)	6xEVA8100 (648)	2x XP12000 (696)
Raw Storage TB	2.7	5.1	10.3	16.7	27.7	46	51
Usable Storage TB	2	3	8	11	21.5	30	37
OS*	Windows	Windows	Windows	Windows	Windows	Windows	Windows
Estimated concurrent users**	100	100	100	100	100	100	100
Estimated queries per hour (QPH)	168	168	168	168	168	168	168

* Microsoft Windows Server 2003 for Itanium-based Systems

** 100 "Concurrent users" were used to generate 168 queries per hour (QPH). Actual customers may have a thousand+ users which generate 168 queries per hour or 20 users which generate 168 queries per hour. Therefore, HP recommends that the QPH rate be used as the main guide vs. the number of concurrent users for each respective database size.

ROLAP – High throughput (505 queries per hour) – reference configurations

Table 3. High throughput configurations

User data warehouse size	500GB	1TB	2TB	4TB	7TB	10TB	20TB
Architecture	SMP	SMP	SMP	SMP	SMP	SMP	SMP
HP Integrity Server models	rx3600 or BL870c	rx7640	rx7640	rx8640	rx8640	rx8640	HP Integrity Superdome
Nodes	1	1	1	1	1	1	1
Sockets/cores per node	2/4c	4/8c	6/12c	8/16c	12/24c	16/32c	28/56c
Memory per node	24GB	32GB	48GB	96GB	192GB	256GB	256GB
FC Cards per node DP = dual port	2x DP 4 Gb/s	4x DP 4 Gb/s	6x DP 4 Gb/s	14x DP 4 Gb/s	12x DP 4 Gb/s	20x DP 4 Gb/s	48x DP 4 Gb/s
SAN Switch	SAN 4/16	SAN 4/32	SAN 4/32	SAN 4/64	SAN 4/64	3x SAN 4/64	4x SAN 4/64
Storage Array (number of disks)	EVA6100 (62)	2xEVA8100 (112)	3xEVA8100 (228)	7xEVA8100 (396)	XP12000 (326)	2xXP12000 (552)	2xXP24000 (1330)
Raw Storage TB	4.4	8	18.4	28.5	23	40	98
Usable Storage TB	3	6	11	20	17	26	64
OS*	Windows	Windows	Windows	Windows	Windows	Windows	Windows
Estimated Concurrent Users**	300	300	300	300	300	300	300
Estimated Queries per hour	505	505	505	505	505	505	505

* Microsoft Windows Server 2003 for Itanium-based Systems

** 300 "Concurrent users" were used to generate 505 queries per hour (QPH). Actual customers may have thousands of users which generate 505 queries per hour or 50 users which generate 505 queries per hour. Therefore, HP recommends that the QPH rate be used as the main guide vs. the number of concurrent users for each respective database size.

ETL tier (SQL Server Integration Services) – reference configurations

After reviewing the data warehouse reference configurations for ROLAP queries, shown above, some readers may have concluded that the reference configurations for querying the data warehouse tier are primarily associated to MByte/second rates driven by the query per hour (QPH) rate and the mix of query complexity. This is true. In addition, HP tends to avoid the term “simultaneous users” because “simultaneous” may mean different things to different people.

However, when sizing the ETL tier you are not concerned with users or “queries per hour.” The SQL Server reference configurations for the ETL tier are most significantly influenced by:

- Load/insert rates
- Index builds
- Complex business logic
- Table look-ups

Load/insert rates

If not known, in order to estimate daily or weekly load/insert rates per batch window, it is possible to calculate backwards. For example:

- If a customer told HP that they believe they will have a 15TB data warehouse containing 5 years of data
- If they were to load data daily for 5 years, then they would be performing 1,825 loads within 5 years
- If one assumes an average of 600 bytes per row, then simple arithmetic will allow you to assume that an average of 15 million rows will be loaded per evening load if you have a 15TB data warehouse.

Note: In the real world, actual daily loads will have more or less than 15 million rows. So, the customers will have to decide if they want to size for peak (could be 2 – 5 times average) or use the average volume for initial estimates.

- It should also be noted that, in the real world, more data is loaded than simply the fact table. For example, dimension table updates are typically very small and have negligible impact on system performance.

People may be surprised that the load/insert rate generally does not put a heavy load on the server. For example, you will notice in the following chart, that you need the same number of processors to insert 500,000 or 15 million rows within one hour and build 10 indexes on 200 million row tables or partitions within 3 hours. Of course, CPU utilization on the 1 processor (dual-core) server will be higher when loading 15 million rows than loading 500,000 rows within the same time window, but, the workload still fits within a one processor server.

The chart below shows that a reference configuration to support loading data into a 15TB data warehouse using SSIS tools will be an HP ProLiant BL460c server; 1/4c; 2GB of memory, and the server running the SSIS packages will require an HP Integrity BL860c server; 1/2c; 4GB of memory. This relatively small server configuration will support the ETL tier for customers:

- Inserting less than 15 million rows into the fact table while maintaining ten indexes on the fact table within 1 hour (a large number of inserts should be executed within a single “begin/end transaction” statement).

- Building 10 indexes against a 200 million row dimension table partition in three hours

If you plan on performing unusually complex business logic, string manipulation or large sorts, which were not accounted for when developing the HP reference configuration, then additional hardware will be required, but, HP feels that this configuration will support daily loads for most customers who have a 15TB data warehouse (DW).

Note: If you would like to alter the ETL reference configuration workload assumptions listed below, please download and run the HP SQL Server BI Sizing Tool to obtain a configuration which will more closely support your environment.

Figure 5. SQL Server sizing table

Approximate DW size	15TB
# of rows in fact table (million)	22,500
Approximate DW size (GB)	14,870
# of loads (365/year * 5 years)	1,825
# of partitions (months)	60
# of rows to insert into fact table (*)	15,000,000
# of indexes maintained during insert	10
Time window to perform inserts	1 hour
Fact table has existing clustered index	yes
# of rows in fact table partition (1 month)	375,000,000
# of rows in table/partition to be indexed (usually largest dimension table or partition)	200,000,000
# of indexes to be built	10
Time window to build indexes	3 hours
Business function complexity & table lookups	medium
Architecture	2 tier
ETL Server (SSIS/Tools) – Sockets/cores – memory	ProLiant BL460c; 1/4c; 2GB
Server (Target DW/DM) – Sockets/cores – memory	Integrity BL860c; 1/2c; 4GB
FC cards per node	1 x 4Gb/sec
SAN switch	SAN 4/24
Storage Array to store DW and support ETL	2 * EVA8100 (266) (**)

(*) NOTE: The reference configurations assume row-at-a-time inserts, while maintaining indexes during the insert operations. Please see discussion below on *Index builds – bulk load vs. row-at-a-time inserts*.

(**) Note: The amount of storage required for the “target” data warehouse should be the higher of either the ETL reference configuration or the ROLAP reference configuration. You should not total the storage requirements for both tiers, since the ETL window typically does not interfere with users submitting ROLAP queries.

If your ETL batch window does not interfere with end user query or cube processing, you may want to consider consolidating the 1 processor (2 core); 4GB of memory into your ROLAP data warehouse. This consolidation of the ETL tier with the ROLAP tier will reduce network traffic between tiers, especially with regard to returning result sets for table lookups.

Real-time updates to data warehouse (ROLAP)

Some customers are starting to develop “real time” BI environments which insert/update the data warehouse or ODS (Operational Data Store) in real-time. Since real-time inserts/updates occur throughout the day instead of the more traditional approach of loading data during an evening or weekend batch window, the resources required to support real-time inserts/updates is minimal.

Bear in mind that it is desirable to build queuing mechanisms into this real-time insert/update architecture to support unexpected spikes in volume and to handle recovery.

Index builds – bulk load vs. row-at-a-time inserts

The ETL reference configurations assumed that the customer will be performing row-at-a-time inserts vs. bulk load operations. This is because some customers have indexes on the fact table and in HP test scenarios it was usually more efficient to perform row-at-a-time inserts while maintaining indexes instead of performing a bulk load and rebuilding indexes after the load on large fact tables. HP benchmarks placed multiple row inserts into “transaction mode” to speed up the ETL operation.

It is also a best practice to partition your fact table and large dimension tables for optimal performance. Some of the more significant advantages of partitioning are:

- Parallel I/O across multiple partitions will maximize performance.
- If fact table is partitioned by date, then the oldest partition may be dropped as it ages off of the data warehouse (that is, greater than 5 years old in the reference configurations).
- Instead of building indexes on the whole fact table, it is far more efficient to build indexes on the partition(s) which were affected during the load. So, only the most recent partition(s) need to have rebuilt indexes after a bulk load.

Note: As stated above, many times row-at-a-time inserts may be faster than a bulk load and index rebuild.

Complex business logic

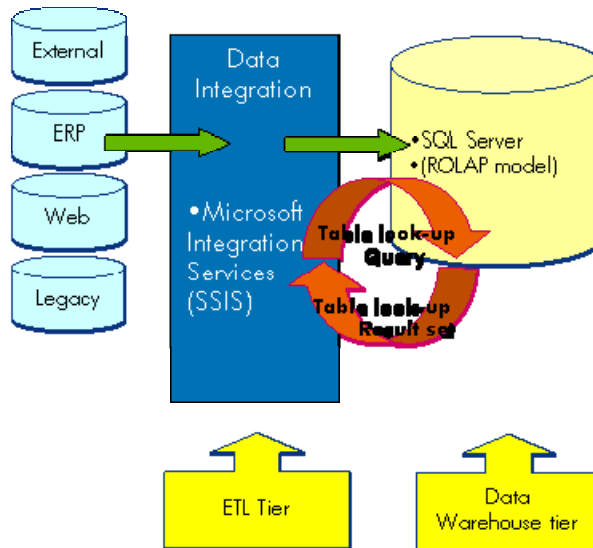
The ETL reference configurations assume that business logic is relatively simple (for example, remapping columns, etc.) However, the SQL Server BI sizing tool allows the user to:

- Specify a number of data conversions, aggregate calculations, “integer-string conversions,” and “string – text conversions”
- Estimate business logic complexity (simple, medium, complex and very complex)

Table look-ups

Table look-ups are generally SQL statements used to perform data cleansing and validation operations. They may be costly ETL operations depending upon the complexity of SQL queries executed. However, the bulk of the table look-up overhead is born by the data warehouse and typically not by the ETL tier, but, the ETL tier (Integration Services) will have to compare data elements, convert encoded data into text, etc. as part of its data cleansing and transformation functions.

Figure 6. Information flow in SQL Server ROLAP model



Note: In SQL Server 2008, SQL Server Integration Services (SSIS) will have a new feature which will allow SSIS to cache table look-up data sets within the ETL tier. These data sets are expected to be stored locally to SSIS as files and will be accessed to minimize overhead on the ROLAP data warehouse tier. Conversely, this local caching of data on the ETL tier will force the ETL tier to consume more CPU and I/O resources than is consumed on SQL Server 2005. Therefore, HP plans to produce new reference configurations and recalibrate the HP SQL Server BI Sizing Tool to take this feature into account after Microsoft officially releases SQL Server 2008.

Do you need to add hardware to support SQL queries that perform table “look-up” operations during the ETL batch window?

Reference configurations are usually beneficial before a customer has a data warehouse up and running. At this point in time, a customer’s understanding of table-look-up complexity is typically not known. Therefore, the reference configurations assume that the data being loaded is clean and that

any data cleansing, edit checks, remapping of columns, etc. will not consume a significant amount of extra CPU resources over and above load/insert operations.

It is important to note that table look-ups and other data cleansing operations usually execute during off hours, when business user queries are VERY low or non-existent. Therefore, it is fair to assume that the data warehouse tier, which normally handles end-user queries during the day, will have plenty of resources to execute table look-up operations to support typical ETL workloads.

Using the HP BI SQL Server sizing tool to estimate table look-up overhead

As stated above, HP believes that its reference configurations are representative of most “real world” workloads and that table look-up adjustments should not be necessary because business users rarely execute queries during the ETL execution window. Therefore, the data warehouse’s hardware resources are typically freed up in the evening, during SQL Server Integration Services execution.

However, If the customer believes that their table look-up queries are more complex or will place a heavier workload on the data warehouse than their end user population will do during the day, then the ETL, table look-up queries, should be modeled separately using the SQL Server BI Sizing tool (“Queries against a ROLAP model, data warehouse”).

In order to take into account simple, medium, complex or very complex table-lookup operations, the user of the sizing tool should treat table look-ups like any other query accessing a ROLAP data warehouse. From a performance point of view, if a query is initiated by an ETL tool (Integration Services) vs. a real end user (human being), the resources required to support the query is virtually the same.

In summary, it is important to bear in mind that hardware that supports an end-user workload during the day tends to be sufficiently large to support ETL table look-ups during evening hours or over the weekend. Since the ETL jobs and business user queries typically do not overlap, it is fair to size the data warehouse for the larger of the two workload requirements, which is usually “end user queries” accessing the data warehouse.

Effect of data compression on ETL processing in SQL Server 2008

SQL Server 2008 will have a new, optional, feature, which allows tables and partitions to use data compression. It is expected that compression will consume some additional CPU cycles during ETL load or insert processing. Therefore, the July 2008 reference configurations and July 2008 release of the HP SQL Server BI Sizing Tool will assume uncompressed data.

OLAP tier (SQL Server Analysis Services) – reference configurations

The use of SQL Server Analysis Services is optional and is not used in all business intelligence environments. However, as customers realize the benefits of OLAP cube technology, HP is finding that it is being implemented more frequently to provide users with significantly faster response times than that of traditional ROLAP data warehouses / data marts. OLAP cubes also provide data mining capabilities and may be front ended by PerformancePoint and/or Excel which provides users with executive dashboards and other advanced BI features.

Which OLAP reference configuration will support your query workload and meet your user service level agreements?

In order to determine which OLAP reference configuration will support an estimated number of user queries per hour in your environment, you first should determine which HP server model will support your cube size. (Note: When using a cell-based server, each cell board may be figured with 128GB of memory. This allows an Integrity Superdome server to support up to 2TB of physical memory).

One of Microsoft’s “best practices” is to insure that your most frequently accessed cells and aggregates are memory resident. Since this “hit” percentage is frequently not known in the early

phases of software development, HP's reference configurations assume that the entire cube will be memory resident in its OLAP cube reference configurations.

Ideally, a memory resident cube will not have any disk I/O once it is in memory. But, some queries may require extra computation, sorts and aggregation. Therefore, HP has added a reasonable number of disk drives to support these unforeseen I/Os.

If the customer decides not to keep the entire cube in memory and if non-memory resident cells and aggregates are accessed to execute the query, the customer should expect more disk I/O activity and increased CPU utilization levels. This will, obviously, have a counterbalancing effect on query throughput and expected response times, but HP hardware may be easily configured to support shifting workloads and may even be modeled by using capacity planning techniques.

Since each HP server has different memory limits, after you determine your cube size, the table below will tell you the minimum number of processors required to support your memory requirements.

- Superdome (2TB) server – Note: SQL Server's practical limit for a single OLAP cube is 1TB
- rx8640 (512 GB) server
- rx7640 (256 GB) server
- rx6600 (192 GB) server
- BL870c (92 GB) server
- BL860c (48 GB) server

Figure 7. Processors needed to support your memory requirements

Cube size	48 GB	92 GB	192 GB	256 GB	512 GB	1 TB
Up to 400 users / 3,800 QPH	Integrity BL860c; 2P/4C; 48GB					
Up to 1800 users / 17,000 QPH		Integrity BL870c; 4P/8C; 92GB				
Up to 1800 users / 17,000 QPH			Integrity rx6600; 4P/8C; 192GB			
Up to 4000 users / 38,000 QPH				Integrity rx7640; 8P/16C; 256GB		
Up to 4000 users / 38,000 QPH					Integrity rx8640; 8P/16C; 512GB	
Up to 8000 users / 75,000 QPH						Superdome; 16P/32C; 1TB
EVA (disks)	EVA4100 (14)	EVA4100 (26)	EVA4100 (26)	EVA6100 (44)	EVA6100 (44)	EVA6100 (86)
MSA (disks)	MSA1500 (16)	MSA1500 (32)	MSA1500 (32)	MSA1500 (54)	MSA1500 (54)	MSA1500 (105)

Based upon the table above, one can understand which HP Integrity servers can maintain a certain size cube in memory (which is a Microsoft "best practice"). The next step is to determine how many processors are required to support your workload [Query per Hour (QPH) rate]. By using the left

column in the reference configuration table above you can find the maximum expected query per hour (QPH) rates which its respective server is expected to support.

It is possible that your server selection may be based more on cube size (memory requirements) than the query volume that you need to support. So, if your QPH rate is significantly less than the proposed server capacity, you may want to consider consolidating the OLAP tier with another functional tier to make more effective use of your hardware investment. Please see the consolidation section later in this document if this is an option.

IMPORTANT: In order to obtain a more detailed OLAP cube configuration for your unique environment, HP suggests using the HP BI SQL Server Sizing Tool.

Example – to determine which OLAP reference configuration is best for your cube size and workload

As you can see from the table below, if your cube is estimated to be 40GB, then you can fit your cube in all HP Integrity servers.

By cross referencing with the left column of the table, you can estimate the query per hour (QPH) rates that each server is expected to support. Therefore, if you believe that you will have less than 3,800 QPH, then you should configure an Integrity BL860c server (2P/4C); 48GB of memory and 14 to 16 disks depending upon if you would like to use EVA or MSA storage.

Figure 8. Which OLAP reference configuration is best for you.

40 GB cube will fit in all servers

Cube size	48 GB	92 GB	192 GB	256 GB	512 GB	1 TB
Up to 400 users / 3,800 QPH	Integrity BL860c; 2P/4C; 48GB					
Up to 1,800 users / 17,000 QPH		Integrity BL870c; 4P/8C; 92GB				
Up to 1,800 users / 17,000 QPH			Integrity rx6600; 4P/8C; 192GB			
Up to 4000 users / 38,000 QPH				Integrity rx7640; 8P/16C; 256GB		
Up to 4000 users / 38,000 QPH					Integrity rx8640; 8P/16C; 512GB	
Up to 8000 users / 75,000 QPH						Superdome; 16P/32C; 1TB
EVA	EVA4100 (14)	EVA4100 (26)	EVA4100 (26)	EVA6100 (44)	EVA6100 (44)	EVA6100 (86)
MSA	MSA1500 (16)	MSA1500 (32)	MSA1500 (32)	MSA1500 (54)	MSA1500 (54)	MSA1500 (105)

But note that if your cube is estimated to be 225GB, then your cube requirement cannot be supported by a blade server. Nor will you be able to keep the cube memory resident in an Integrity rx6600 server because its maximum memory is 192GB. Therefore, the best reference configuration for a 225GB cube will be an Integrity rx7640, rx8640 or Superdome server.

Figure 9. Range of servers needed to fit your needs

225 GB cube will fit in rx7640 server or larger

Cube size	48 GB	92 GB	192 GB	256 GB	512 GB	1 TB
Up to 400 users / 3,800 QPH	Integrity BL860c; 2P/4C; 48GB					
Up to 1,800 users / 17,000 QPH		Integrity BL870c; 4P/8C; 92GB				
Up to 1,800 users / 17,000 QPH			Integrity rx6600; 4P/8C; 192GB			
Up to 4000 users / 38,000 QPH				Integrity rx7640; 8P/16C; 256GB		
Up to 4000 users / 38,000 QPH					Integrity rx8640; 8P/16C; 512GB	
Up to 8000 users / 75,000 QPH						Superdome; 16P/32C; 1TB
EVA	EVA4100 (14)	EVA4100 (26)	EVA4100 (26)	EVA6100 (44)	EVA6100 (44)	EVA6100 (86)
MSA	MSA1500 (16)	MSA1500 (32)	MSA1500 (32)	MSA1500 (54)	MSA1500 (54)	MSA1500 (105)

By cross referencing with the left column of the table, you can estimate the query per hour (QPH) rates that each server is expected to support. Therefore, if you believe that you will have less than 38,000 QPH, then you should configure an Integrity rx7640 (8P/16C) server; 256GB of memory and 44 to 54 disks depending upon if you would like to use EVA or MSA storage.

Note: For cost reasons, a customer may decide that 192GB of memory is close enough to 225GB, so they may opt to consider an rx6600 server instead of the rx7640 server. But, if they physically configure less memory than the portion of the cube they plan to frequently access, then extra CPU and disk resources should be configured to support the extra work that SQL Server Analysis Services will have to perform to bring the required cells into memory.

Note: If your QPH workload is significantly less than what your server is expected to support, then reducing memory and the server model number may be possible. However, response times will not be as good as when the cube will be memory resident. In addition, processor and disk utilization will be higher than expected.

Consolidation architecture (3-tier)

When designing a business intelligence environment, the three logical tiers (ETL, ROLAP and OLAP) may either reside on three separate physical servers or they may be consolidated onto a single, scale-up, server such as the Integrity server. The HP Integrity platform may scale to 64 processors/128 cores with multiple SQL Server instances and Windows operating systems. Since, the hardware required for loading data into the data warehouse and the resources required to query OLAP cubes is usually not significant, then customers may want to consider physically consolidating these logical tiers onto a single physical server.

The reference configurations were designed to support various, pre-defined workloads. However, there are many times when a customer may decide to save costs by consolidating one or more BI tiers:

- ETL tier (Microsoft SQL Server Integration Services (SSIS))
- Data Warehouse tier – Relational Online Analytic Processing (ROLAP) query workload
- OLAP cube tier – Queries accessing Microsoft Analysis Services (MOLAP/HOLAP) cubes

Cost savings may be observed by reducing licensing fees, simplifying server management because there are fewer servers to maintain, saving floor space, reducing electrical costs and more effectively utilizing hardware resources, because functionally isolated servers may be underutilized for long periods of time.

Integrity server SMP configurations can be run as soft partitions (soft-NUMA) in HP Integrity systems or as hard partitions (nPar), which are electrically isolated partitions in a single SMP Integrity server. nPars, also, provide high security and allow for high availability configurations. Through the use of Instant Capacity on Demand (iCAP), customers can improve processor utilization and meet peak performance demands. Coupling this technology with Microsoft Windows System Resource Manager (WSRM) software, consolidation of multiple tiers becomes practical.

Next steps

Next steps involve putting HP BI reference configurations or the [HP Business Intelligence Sizer for Microsoft SQL Server 2005](#) to work for you!

Let HP help you get started. Contact your HP sales representative for more information on solutions based on BI reference configurations that can improve time to deployment for Microsoft SQL Server business intelligence applications. Your sales representative can also arrange for a total cost of ownership (TCO) assessment to compare performance of a BI configuration relative to your existing environment.

Summary

Companies are presented with a wide range of technology options for their business intelligence (BI) applications. Each hardware platform decision requires them to research, and then balance, the tradeoffs between price, performance, long-term scalability, manageability, and vendor reputation. To reduce the risk and deployment time for BI solutions, HP has created a comprehensive set of BI reference configurations. The HP BI reference configurations draw on extensive lab, customer benchmark, and industry standard benchmark investments as well as HP's global experience supporting thousands of data warehouse and business intelligence customers. The reference configurations presented in this white paper reduce the effort to a choice of already balanced and all-inclusive configurations, depending on the size of the database and the likely workload of the customer. Links are also provided to [HP Business Intelligence Sizer for Microsoft SQL Server 2005](#) available to assist customers in optimizing their choices.

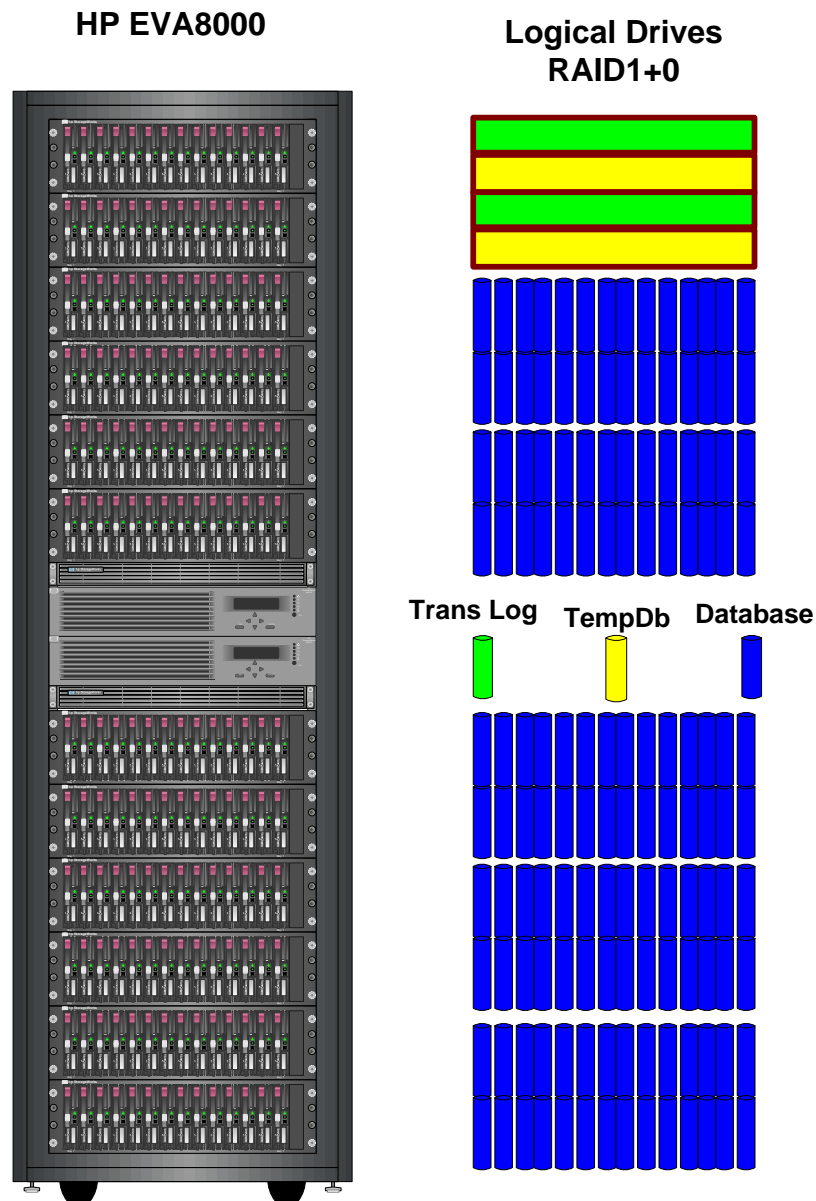
For generic HP business intelligence information, please go to: <http://www.hp.com/go/bi>

Appendix – SQL Server 2008 setup

Logical drive configuration

The external storage configurations listed below depict a best practice representation of logical drive layouts for EVA storage devices. The configurations use RAID 1+0 for the transaction log and tempdb and place the database files on RAID5 disks by default. These external storage configurations guidelines may not be optimal for all workloads and database designs; however, the guidelines provide a good starting point.

Figure A-1. External storage configuration and logical drives



Location of SQL Server binaries

The SQL Server 2008 binaries should be placed on a drive separate from all the data drives. The reference configurations are designed with the intent that the SQL Server binaries should be placed on the same drive as the OS. There is no performance benefit to separating the OS and SQL Server binaries on different drives.

System databases location

For reference configurations with a SAN, the system databases, such as master and msdb, should be located on one of the SAN drives allocated for data use. It is beneficial to locate the databases on a SAN drive, instead of on a disk attached directly to the server, for high availability reasons. In case you need to execute your disaster recovery plan and replace the entire server, it is easier to recover the system if the original system databases are located on a shared SAN drive, instead of on a local drive.

Setup transaction log

With an online transaction processing (OLTP) workload, good transaction log throughput is critical. However, during query processing, the transaction log is infrequently used, so its location is not critical to system performance. On the other hand, the transaction log will be used during the data load window and its optimal physical location will vary depending upon the recovery model used by the database administrator (DBA) during load processing. It is common to place the log files on separate drives from the data drives and use RAID1 for best write performance.

Setup tempdb

The tempdb configuration is extremely important for good SQL Server performance. It is recommended to set the recovery model of tempdb to **simple**. Allow for tempdb files to automatically grow as required and set the file growth increment to a reasonable size to avoid the tempdb database files from growing by too small a value. Pre-allocate space for all tempdb files by setting the file size to a value large enough to accommodate the typical workload in the environment. To maximize bandwidth, create as many tempdb files as processor cores and create them of equal size. Sometimes, better performance may be achieved by placing tempdb files on separate drives from the data drives. This is because disk thrashing may be reduced and tempdb performance is easier to monitor. On the other hand, there may be cost savings by striping tempdb across the data disks even though performance may not be optimal and even though it may be more difficult to monitor and tune the impact of tempdb versus database performance.

Memory allocation for OS

In a default SQL Server 2008 setup, SQL Server can use all memory available on the system. This can have a detrimental performance effect on the operating system itself. Leave 10% of the memory available for the operating system. For example, if your system has 128GB of RAM, set maximum memory available to SQL Server 2008 to 128GB less 10% or 115GB, leaving approximately 13GB to the operating system.

For more information

HP storage options, <http://www.hp.com/storage>

HP Integrity servers, <http://www.hp.com/go/integrity>

Microsoft SQL Server: Big Data, <http://www.microsoft.com/sql/bigdata/default.aspx>

HP Business Intelligence Services,
<http://h20219.www2.hp.com/services/cache/10912-0-0-225-121.aspx>

HP Business Intelligence Solutions,
<http://h71028.www7.hp.com/enterprise/cache/3905-0-0-225-121.html>

HP Business Intelligence Sizer for Microsoft SQL Server 2005 (requires, free, ActiveAnswers user ID and password),
<http://h71019.www7.hp.com/ActiveAnswers/cache/351481-0-0-0-121.html>

To help us improve our documents, please provide feedback at www.hp.com/solutions/feedback

© 2008 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Microsoft and Windows are U.S. registered trademarks of Microsoft Corporation. Intel and Itanium are trademarks of Intel Corporation in the U.S. and other countries. TPC-H is a trademark of the Transaction Performance Processing Council.

4AA2-2222ENW, September 2008

