Contents

Overview ........................................................................................................................................... 3
In-Memory analytics what is it? ........................................................................................................ 4
Who is it for? .................................................................................................................................... 5
What are the external factors driving In-Memory Analytics? ......................................................... 6
What are the internal drivers for in-memory analytics? ................................................................. 7
What benefits are sought? .................................................................................................................. 9
What organizational factors are required? ....................................................................................... 10
What technology has to underpin an in-memory initiative? ........................................................ 11
Data security must be of paramount concern ............................................................................. 12
Summary ......................................................................................................................................... 13

Yellowfin, and the Yellowfin logo are trademarks or registered trademarks of Yellowfin International Pty Ltd. All other names are trademarks or registered trademarks of their respective companies.

While every attempt has been made to ensure that the information in this document is accurate and complete, some typographical errors or technical inaccuracies may exist. Yellowfin does not accept responsibility for any kind of loss resulting from the use of information contained in this document.

The information contained in this document is subject to change without notice. This text contains proprietary information, which is protected by copyright. All rights are reserved. No part of this document may be photocopied, reproduced, stored in a retrieval system, transmitted in any form or by any means, or translated into another language without the prior written consent of Yellowfin International Pty Ltd.

The incorporation of the product attributes discussed in these materials into any release or upgrade of any Yellowfin software product – as well as the timing of any such release or upgrade – is at the sole discretion of Yellowfin.

This document version published April 2010 Copyright © 20010 Yellowfin International Pty Ltd.
Overview

Reporting and analytics has changed dramatically in recent years. Starting with static “green bar” reports in the mid-to-late ’70s, information could be abstracted from mainframe systems and, often manually, transferred to spreadsheets where data could be aggregated and analyzed. Data warehousing was the buzz of the ’80s, and while this did enable heterogeneous data sources to be centralized, projects were often grossly over budget and far below expectations. As technologies have matured and the advent of services-based architectures has become more prominent, data warehousing reinvented itself and emerged as what is now recognized as Business Intelligence.

However, the recent advent of in-memory analysis means that BI expectations have changed forever. Dealing with overly complex software designed for a handful of power users involving long deployment cycles and low project success rates is no longer acceptable. Today, smart companies are striving to spread fact-based decision making throughout the organization, but they know they can’t do it with expensive, hard-to-use tools that require extensive IT hand holding. The pace of business now demands fast access to information and easy analysis; if the tools aren’t fast and easy, Business Intelligence will continue to have modest impact, primarily with experts who have no alternative but to wait for an answer to a slow query.

In-memory analysis and reporting delivers rapid access to visually rich, interactive dashboards that anyone can build quickly and modify easily. It enables users to see and know their businesses in new ways and interactively explore data without limits. Now, users can engage their data with blazing speed, resulting in more informed, proactive decisions. For IT departments, it means which far less time spent on query analysis, cube building, aggregate table design, and other time-consuming performance-tuning tasks.

The in-memory approach is not new. The desire to take advantage of the speed of RAM has been with us for some time. Only recently, however, has the promise become a practical reality thanks to the mainstream adoption of 64-bit architectures that enable larger addressable memory space and the rapid decline in memory prices. Because of this rapidly changing infrastructure landscape, it is now realistic to analyze very large data sets entirely in-memory.

The benefits of in-memory analytics include:

- **Dramatic performance improvements.** Users are querying and interacting with data in-memory which is significantly faster than accessing data from disk.
- **Cost effective alternative to data warehouses.** This is especially beneficial for midsize companies that may lack the expertise and resources to build a data warehouse. The in-memory approach provides the ability to analyze very large data sets, but is much simpler to set up and administer. Consequently, IT is not burdened with time consuming performance tuning tasks typically required by data warehouses.
- **Discover new insights.** Business users now have self-service access to the right information coupled with rapid query execution to deliver new levels of insight required to optimize business performance, without the IT ‘bottleneck’.
In-Memory analytics what is it?

As the name suggests, the key difference between conventional BI tools and in-memory products is that the former query data on disk while the latter query data in random access memory (RAM). When a user runs a query against a typical data warehouse, the query normally goes to a database that reads the information from multiple tables stored on a server’s hard disk. With a server-based in-memory database, all information is initially loaded into memory. Users then query and interact with the data loaded into the machine’s memory. Accessing data in-memory mean it is literally “turbo charged” as opposed to accessing that same data from disk. This is the real advantage of in-memory analysis.

In-memory BI may sound like caching, a common approach to speeding query performance, but in-memory databases does not suffer from the same limitations. Caches are typically subsets of data, stored on and retrieved from disk (though some may load into RAM). The key difference is that the cached data is usually predefined and very specific, often to an individual query; but with an in-memory database, the data available for analysis is potentially as large as an entire data mart.

Yellowfin’s in-memory database is an extension to its existing ROLAP architecture and is designed specifically to take advantage of the immense amount of addressable memory now available with the newest 64-bit operating systems. In-memory technology uses the multi-gigabytes of memory space available in 64-bit servers as for its data store. In-memory analysis is designed to improve the overall performance of a BI system as perceived by users, especially affecting complex queries that take a long time to process in the database or when accessing a very large database where all queries are burdened by the database size.

With its in-memory database, Yellowfin allows data to be analyzed at both an aggregate and a detailed level without the time-consuming and costly step of developing ETL processes and data warehouses or building multidimensional OLAP cubes. Since data is kept in-memory, the response time of any calculation is lightning fast, even on extremely large data sets analyzed by multiple concurrent users.

This kind of immediate, interactive analysis is particularly important when people are trying to uncover unknown patterns or discover new opportunities. With conventional reporting tools, users typically have a specific question, such as “What are my sales for this quarter.” With in-memory visualization tools, the question can be less specific and exploratory, akin to “show me the data and the patterns in the data.

In-memory analysis, however, needs to be taken in context with a total BI strategy that allows users to access data from sources that provide the most effective ROI for that data. An in-memory strategy needs to be cognizant and include additional data sources such as data warehouses and analytical appliances that ensure the responsiveness and performance to meet a users needs based on data volumes and complexity.
Who is it for?

In-memory BI is ideal for setting up departmental BI applications, and for meeting the BI needs of small to medium sized businesses as it requires very little up-front effort, and no ETL. Populated quickly from any database source, users can seamlessly use in-memory databases and associated meta-data layers as a source for many reports, dashboards, and analysis. Users can analyze the full depth and breadth of their data stores be they transactional systems or data warehouses since no limits are set in data scalability, analytical scope or drill-through capability.

By supporting rapid departmental deployments, in-memory analysis frees up reliance on IT departments that can be too slow in responding to the organization’s deluge of reporting requirements. A centralized data warehouse, that has to accommodate an organization’s diverse requirements, can translate into longer wait times to produce the reports required. Demand for in-memory is also strong with SMB’s that either lack the resources or expertise to build their own data warehouse; it offers an ideal alternative because users can analyze vast quantities of data in-memory and are often easier and faster to use than relational data marts.

Another main use for the in-memory approach is to facilitate a more exploratory, visual analysis. This approach is a stark contrast to many traditional query and OLAP products; instead of starting with a blank screen to build a query or a report, users start with a view of all the data. Held in memory, this data is then filtered down to just the information users are looking for with easy-to-use data selection, sliders, radio boxes and check boxes.

What this means is that with Yellowfin you can rapidly deploy to your end users as the number and complexity of the steps required for delivering a reporting project is significantly reduced (minimizing your project costs and risks) without the need to compromise on the speed and efficiency of your reporting and analysis.

(see figure below)
What are the external factors driving In-Memory Analytics?

There are a number of external factors driving the adoption of in-memory analysis technologies. These include:

**Availability of 64-bit computing**

Addressable RAM (Random Access Memory) is certainly a key ingredient in the success of in-memory analytics - the more memory available for analysis the greater the flexibility and performance of the system. As a result the major driver for in-memory analysis has been the innovations in hardware and associated operating systems – specifically the adoption of 64-bit operating systems. These 64-bit OSes makes in-memory BI enterprise-scalable, by providing Terabytes of space available for data storage and subsequent analysis.

**Growing data volumes**

"It's not news to IT execs that electronic data is proliferating at a dizzying rate. IDC has estimated that that by 2011, the amount of electronic data created and stored will grow 10 times the 180 exabytes that existed in 2006, reflecting a compound annual growth rate of almost 60%".

Many enterprises are being stymied by a massive proliferation of data in their databases and applications. Growing volumes of transaction data are being digitally captured and stored, along with unstructured forms of data files such as email, video, and graphics. At the same time tighter regulations put the onus on organizations to maintain this data and keep it available for years to come. Much of this data still resides on legacy systems, which are costly to operate and maintain.

To rapidly access this data on an ad-hoc basis for analysis purposes without the need to build complex data warehouses is another driver for in-memory analysis. With an in-memory database it is possible to connect to a legacy data store, populate a database, conduct your analysis and then discard the in-memory database once your analysis is complete. In this scenario if the organization needs to develop a full data mart for access the cost of development, storage etc would be significant but in-memory analysis makes access and analysis of large data volumes possible at incredibly low costs.

**Demand for real time analysis**

Achieving real-time Business Intelligence presents many challenges but one of the main hurdles to overcome is slow query performance due to limitations of traditional BI infrastructure. Today, many organizations have a data warehouse that stores a version of their transactional information and makes it available for queries and analysis. Data warehouses are updated with periodic snapshots of operational data. This process is commonly referred to as extract, transform and load (ETL), and can take anywhere from a few hours to several days to complete. As a result, the data in the warehouse can be at least a day or two out of date—or even a week or a month behind the source systems.

Whilst this latency is generally acceptable for trend analysis and forecasting, traditional data warehouses simply can't keep pace with today's BI requirements for fast and accurate data. They were not designed to deliver complex analytics on terabytes of data quickly and, as the volume of data used in organizations grows exponentially, extracting information becomes more time-consuming and complex. This is where in-memory databases that support incremental loads shine. The ability to bulk load data in near real time provides end users with instant access to their data for real time reporting.
What are the internal drivers for in-memory analytics?

There are three main internal drivers for the development of in-memory analytics within organizations. These include:

**Speed of deployment**

With the growth of data volumes and proliferation of database driven applications organizations are finding it harder to manage their BI projects for all levels of the organization. There are many cases where users simply want access to their specific transactional systems for reporting and analysis purposes without the need for a data warehouse deployment.

In-memory analytics doesn’t so much make it easier to consume analytical applications rather than make it easier to build them; essentially it removes the need to build a complicated disk-based performance layer, such as relational-based aggregates and multidimensional cubes. Business units won’t need IT to build these complicated disk-based performance layers if they can use analytical applications running on top of an in-memory performance layer. It is difficult for IT to build data marts for all user requests, and most users are unwilling to use BI without it.

Slow query performance will stunt end-user adoption; therefore in-memory analytics will drive wider BI adoption, as it will be much easier for users to get the performance they demand for all analytical applications. There will be no need to wait for the IT bottleneck to break as fewer resources will be required to build summary and aggregate tables to be used for analysis.

**Traditional BI project delivery times**

![Graph showing traditional BI project delivery times]

- **Average implementation time:**
  - 17 months in total; 5 months to deploy the 1st usable analytic application

- **Mean annual expenditure on BI software:**
  - $1.1 million for companies with >1,000 employees

- **Project Success Rate:**
  - 31% success rate; at best

- **Meeting needs - right data to the right person:**
  - only 36% are confident that reports and dashboards deliver the right data to the right person at the right time.

*IDC Business Intelligence Survey*

**Speed of analysis**

Interactive data visualization has been adopted as the common front-end to analytical applications, driven by the ubiquity of rich internet applications. These tools, with their attractive displays, are being more widely adopted by those users who aren’t accustomed to the grid style of analysis and reporting offered by relational databases and spreadsheets. By definition, interacting and exploring data using visualization tools requires only the most intuitive of tasks, such as clicking on a pie wedge, or circling the dots on a scatter plot. However, for this type of visualization to be adopted the speed of analysis has to be fast. End users expect instantaneous results from their queries.

The most obvious advantages of in-memory technology, then, are the improved performance of analytical applications. Since data is being accessed directly from RAM, ad-hoc query results are returned in an efficient, highly responsive fashion that OLAP cubes would find difficulty matching.
Gartner Research confirms that not only can data be retrieved faster, but in-memory analytics solutions perform calculations and query results much faster than disk-based approaches. This allows users to slice and dice data and create robust reporting applications without the limitations associated with multidimensional cubes or aggregate tables. Real-time, ad-hoc query capabilities, then, can be extended to even high volume transaction industries such as financial services and retail.

**Technical challenges associated with traditional BI**

Traditional BI presents a number of challenges to IT that in-memory overcomes. These technical challenges include:

1. Reporting requirements have a large impact on I/O time and network resources when dealing with large volumes of data. In order to minimize this, a company needs optimized file and database access methods. In-memory overcomes this process by shifting all the data and workload into memory.

2. The aggregation of data to produce summaries at user query time, which is typically not feasible with the immediacy of response time that users want. Because of this, many data marts are designed to pre-calculate and pre-store the summaries. This can take huge amounts of processing and data movement. Optimized aggregation algorithms are needed to enhance performance. With in-memory, data aggregation is not required. The vast amounts of data that can be stored and analyzed means that an aggregation process is not required.

3. The amount of data for database loads means that ETL processes take many hours to complete. These loads involve index creation, which requires sorting. Because of the amount of data used in typical loads, the fastest possible sorting algorithms are necessary. With an in-memory database such as Yellowfin, bulk loaders mean that huge volumes of data can be loaded and compressed ready for reporting in minutes.

**Growing challenges for BI and the data warehouse**

<table>
<thead>
<tr>
<th>Today</th>
<th>IT Pain</th>
<th>Business Pain</th>
</tr>
</thead>
<tbody>
<tr>
<td>More Data</td>
<td>Higher System Costs</td>
<td>Impact to financial performance</td>
</tr>
<tr>
<td>Gb to Tb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>More end users</td>
<td>Slower Performance</td>
<td>Delayed decisions or responses</td>
</tr>
<tr>
<td>100’s to 1000’s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>More Complex queries</td>
<td>Missed reporting windows</td>
<td>Missed deadlines and penalties</td>
</tr>
<tr>
<td>Ad hoc reports and analysis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The technology supporting BI today must provide SPEED, SCALABILITY, FLEXIBILITY and AVAILABILITY at a low operating cost.
What benefits are sought?

Three critical benefits needs to be sought by organizations looking to adopt in-memory analysis technologies. These are:

Higher end user adoption leads to improved decision making

Perhaps the most significant benefit, albeit less recognized, of in-memory technology is that ease of use allows business analysts and line-of-business managers can build their own reports and dashboards with very little training or technical expertise. This encourages higher levels of user adoption due to the autonomy that users have to explore their own data. The value of data shifts away, then, from those that manage the data to the stakeholders who use and own the data, allowing them to make better decisions. Solutions such as Yellowfin enable users to see and know their business in new ways and interactively explore data without limits.

Delivering more BI with less infrastructure to more users

In-memory analysis offers the extraordinary benefit of eliminating a large amount of repetitive and expensive processing that would normally burden the database servers. By creating in-memory databases for analysis, the source database engine only needs to populate the in-memory database once instead of processing a query each time a report or dashboard is run.

The in-memory database can be scheduled to populate, either as a batch or incremental load, overnight during off-peak hours, thus freeing up database capacity during the peak interactive hours of the day. By offloading the database server with in-memory analysis, companies can deliver faster performance, deliver more reports per hour, and free up capacity on the source database servers for operational purposes.

Enabling BI self service and disposable analytic applications

Finally, in-memory analysis gives companies a very easy mechanism to carve out subsets of the enterprise BI environment for convenient departmental or workgroup usage. Analytic applications for different subject areas can be created for different workgroups by a centralized IT group, and those workgroups can operate autonomously using in-memory databases without impacting the data warehouse workload.

In addition to this in-memory analysis lends itself particularly well to conducting ad hoc analysis within the organization. This is ad-hoc in the true sense of the word. This is the ability to rapidly source data and build an analytic application to meet a specific investigative need. For example conducting analysis on the key drivers of customer attrition. In this case the analysis is undertaken; then programs built to reduce customer attrition. On completion it is possible that analysis database used is no longer required and can be disposed of. In a traditional BI environment this thinking might result in huge push back but with the rapid ROI of in-memory analysis this type of BI is going to become more and more prevalent.
What organizational factors are required?

To achieve the benefits sought in an in-memory analysis project a key set of organizational attributes need to be in place.

It is vital that a company has already established a competency for Business Intelligence, or has a general analytics strategy in place. In-memory analysis is a sub set of a broader BI program. It should not be embarked upon in isolation. The general framework for the delivery of a successful BI project applies equally well to in-memory BI projects.

Gauge the costs and benefits of in-memory deployments

An in-memory deployment involves supporting another data set and structure. In-memory investments often deliver great value, but the benefit may be hard to quantify and articulate. Businesses often aren’t aware of tangible, cost-of-ownership benefits such as saving disk space and reducing administrative labor that would otherwise be required to speed queries.

Companies contemplating in-memory approaches would do well to raise the profile of these backroom tasks to better build the business case. Positioning in-memory analysis with the user segments and types of applications is critical in the cost/benefit analysis and overall BI tool strategy.

Understand the BI bottlenecks

Are users complaining about poor query response times? Do complex queries time out before finishing? Does poor performance prevent users from asking important business questions? If so, consider in-memory technology that delivers an integrated user experience. It is critical that the application does not constrain users but offers flexibility in terms of data access as well as report publishing and analysis.

Identify the business opportunity and consider what users do with the data

If your BI deployment is report-oriented and does not facilitate what-if analysis, interactive filtering, discovery of patterns and new opportunities, then adding an in-memory visualization may be beneficial. If users routinely dump data into spreadsheets for interactivity, it maybe a warning sign that your BI environment is too inflexible and not aligned with the business requirements. Conversely, there are times when all users need is a report or a single number. Don’t expect in-memory analytics to be the answer for everyone.

Understand your scalability and delivery

For broad deployments accessing large data volumes, ensure that your infrastructure and application supports 64-bit deployments of your preferred operating system, whether that’s Windows, Linux or Unix.

In addition broad, distributed/extranet deployments demand rich web-based interfaces and enterprise-class administrative features. There is little advantage in procuring an in-memory database if the visualization component does not scale to meet your deployment needs. In memory analysis should not be constrained to the desktop – it needs to be server based.
What technology has to underpin an in-memory initiative?

The success or failure of in-memory analysis does rest to some degree on the technology chosen to be the delivery platform. The fundamental requirement is that this platform is web-centric, beyond that there are some essential technology components that assist to deliver the business benefits sought. These are:

**Enterprise scalability and security**

All BI solutions must include enterprise administrative features, such as usage monitoring, single sign-on and change management; and this is just as true for in-memory solutions. It is therefore, critical that you choose solutions such as Yellowfin that can provide enterprise class infrastructure that enable you to scale your deployment as your users grow.

**Integration with your existing data warehouse and OLAP cubes**

While some vendors tout in-memory as a way of avoiding building a data warehouse, this option usually applies to smaller organizations that may only have a single source system. For larger companies that have multiple source systems, the data warehouse continues to be the ideal place to transform, model and cleanse the data for analysis.

Look for tools that are designed to integrate with and leverage existing BI environments. An in-memory solution that is tightly integrated into the visualization tool is critical. However, it is equally important that the visualization tool can also access your OLAP cubes and data warehouse tables without the need for an in-memory middle-layer. Without this option a purely stand-alone in-memory solution can lead to yet another version of the truth, adding complexity to your BI environment.

Yellowfin takes a flexible approach whereby the system administrator can configure the server to perform processing either against the in-memory database, or alternatively, push processing down to the underlying data store. The decision on which approach is optimal for a given deployment will depend a lot on the query performance characteristics of the data store. For example, a traditional OLTP data store may benefit significantly from in-memory processing, whereas a query optimized analytic data store may provide performance similar to or better than in-memory processing. Combining this flexible architecture with the cost advantages of not using an OLAP server gives customers choice and a BI platform that can grow as their data and analysis requirements do.
Ensure real time data refresh

Because reporting data is potentially extracted from a source system or a data warehouse and then loaded into memory, data latency can be a concern. Front-line workers in a customer service center, for example, need near-real-time, highly granular (detailed) data. If an in-memory tool contains last week’s product inventory data, it’s probably not of use to customer service reps. Thus, the suitability of an in-memory tool and the success of the deployment may hinge on the degree to which the solution can automate scheduled incremental data loads.

Minimize administration overhead

In-memory analytic tools often introduce some of the same concerns that OLAP stores create: namely, they usually create another data source, with its own calculations and business definitions. This is where tools such as Yellowfin differ from other in-memory approaches: existing queries, reports and dashboards automatically take advantage of an in-memory database, seamless to users. Administrators are not adding calculations and business logic within another layer; they reside within the existing meta-data layer for reporting that is already built.

Web-based development and deployment.

Some in-memory tools are not nearly as web enabled as their conventional BI counterparts. This seems to reflect both technology immaturity and a tendency to be a niche deployment. However, for successful adoption with minimal administrative overhead web based development and deployment is critical. Both the visualization tool and in-memory database need to be server based deployments to ensure data access security and application upgrades can be easily managed. Solutions such as Yellowfin, provide a single web based platform for delivering your Business Intelligence needs. From connection through to design, modeling and visualization, your users work within a fully integrated browser application that encourages collaboration and an iterative approach to report development - leading to analytical applications that meet the needs of your end users.

Data security must be of paramount concern

In memory applications have the potential to expose significantly more data to end-users then ever before. This raises security issues regarding how data is accessed, where it is stored and who has access to that data.

In determining the best strategy for your in-memory deployment security needs to be foremost in your selection criteria. There are two aspects of security the location of your data. Where is it stored and is that storage secure? And secondly, who has access to that data store? The most secure location for your data is on a centralized server, whether hosted or internal. Not only is this more secure but it maintains basic controls regarding data governance.

To understand this consider a scenario where users are able to conduct complex queries by downloading up to 100 million rows of data to their desktop from many data sources, or data feeds from the web. Sure the information can then be sliced and diced into reports or users can create BI applications on their desktops and share them with colleagues. Sounds great in theory but fraught with danger in practice. With this level of data on a laptop it is free to leave your premises and get lost or stolen in the worst case or published without any form of governance at best.
In addition to centralized storage your in-memory analysis need to conform to data security measures as well. These means that data access profiles for your users need to be adhered to throughout your reporting process. Organizations spend an enormous amount of effort in securing their transactional applications and so it is critical that when it comes to the data the same level of security is present. This means that users only have access to the data they are authorized to access, and that this access is changed as the employee’s role changes.

**Summary**

In-memory technology is ideal for setting up departmental and SME BI applications with minimal up-front effort, no need for ETL technology, and no need for a decision support database. In-memory databases can be populated quickly and easily from any database source or even a spreadsheet. Once the data resides inside n-memory, it can act as a source of reusable definitions for many reports, and dashboards.

When selecting an in-memory solution consider one that operates seamlessly within an end-to-end BI platform where its usage is completely transparent to users and report developers. Look for technology that has been designed to avoid the excessive administrative burdens and can scale to enterprise levels in terms of user number, data security and data governance. Your in-memory analytical application needs to be more than a one trick pony. It needs to be able to support the multitude of BI needs that exist.

So what does in-memory analysis mean for today’s organization? As with most challenges in Business Intelligence, technology is only a part of the answer. Thereal value of in-memory BI is not only how fast it is, but more importantly, the decisions that can be enhanced, the tough business questions that can now be answered and the new opportunities that will be discovered.

---

**Find out more**

Contact Yellowfin at [www.yellowfin.bi](http://www.yellowfin.bi) and ask for our proven roadmap to assist you to successfully implement Yellowfin BI into your organisation.