As your application’s data demands grow, maintaining scalability and performance is an increasing challenge. With Terracotta BigMemory, you can manage terabytes of data in-memory, cutting processing time from minutes to seconds, or less. This software solution for data-related performance and scalability problems can easily be plugged into your data-intensive applications today. Use BigMemory to give your applications instant access to terabytes of enterprise in-memory data storage for high performance at any scale; and transform your business by creating new opportunities with this volume of data. BigMemory provides a full set of enterprise data management capabilities, including high availability, durability, consistency and transaction support, and monitoring.

**Why in-memory?**

Applications need speed and scale in today’s hyper-fast, need-it-now world. Competitive pressures and escalating customer expectations have put a premium on absolute application performance and throughput. Successful businesses must scale flexibly and rapidly, which can stress traditional multi-tier application architectures.

If your applications rely on data stored in a central resource, such as a disk-based database or other type of server, you know that indirect data access soon becomes a bottleneck that slows application performance dramatically as scale increases. Alternatively, keeping data directly in the server’s inexpensive RAM, where your application runs, offers the best performance and maximum value.
Why big?

Data volumes continue to grow as applications become more connected and customer expectations increase. In fact, according to industry analysts, data volumes are expected to increase 10-fold every five years, quickly outpacing the capabilities of existing technologies and even Moore’s law. Since both business and customer value are often associated with the need for massive amounts of data, enterprises are always looking for a solution.

Additionally, there’s a growing need to dynamically process large volumes of data in real time to derive even more value and use data in new ways. Traditional data storage technologies just aren’t built to meet these requirements. Be sure to read the “BigMemory everywhere” section to see how companies are using BigMemory to transform what’s possible.

Why BigMemory?

BigMemory allows you to store data where it’s used: in memory. With BigMemory, you can easily store and manage your in-memory data in a re-usable, standard way that simply plugs into your application. As a result, you can harness the value of massive amounts of data, with real-time processing, while keeping all of it in your server’s RAM for maximum performance and scale. The performance improvements that result will continue to scale as your application’s data set and user base grow. There’s no simpler way to get predictable and fast access to large volumes of data.

BigMemory can ready your business for the increasing demands of a digital era by delivering capabilities such as its fast, low-latency data access, predictability, scalability, high availability, durability, consistency, monitoring and management. It is simple to get started and it works with your existing hardware and developer skill set. In addition, BigMemory supports the full consistency spectrum for data reliability and transaction support with maximum performance, including a high-availability configuration (no single point of failure), to meet enterprise requirements.

Snap-in performance and scale

For the absolute best enterprise application performance, you need to store data where it’s used and needed: in memory. Accessing data in your server’s RAM is hundreds of times faster than disk or network-based access. Since inexpensive servers with hundreds of gigabytes of RAM are increasingly abundant, it makes sense to use as much of it as you can. With BigMemory, you can access and manage all of your enterprise data in memory to achieve the scale your application requires.

Using BigMemory requires no code changes. Thanks to the Ehcache interface (the de facto get/put API), only a few lines of configuration are required. Your application will see benefits immediately and into the future as BigMemory allows you to scale up, scale out and scale to the cloud. Capacity is virtually unlimited as you can affordably buy servers with large amounts of RAM and move even terabytes of data into memory. BigMemory allows you to utilize your data in ways that weren’t possible before.

BigMemory customers are seeing an immediate speed-up in application response times as well as increased throughput and transaction rates. See first-hand how our customers are using BigMemory to transform their business, view Terracotta customer videos at http://www.softwareag.com/corporate/rc/.
Benefits of a tiered-storage architecture

BigMemory gives applications instant access to a large memory footprint and, at the same time, solves the inherent challenges associated with traditional data management approaches such as using Java® heap for in-process memory handling and distributed disk-based storage for handling larger volumes with high availability.

BigMemory solves three key problems commonly seen in applications with high data demands:

1. **Database-related delays**: BigMemory keeps application data in memory, eliminating costly database, web server or other data access delays.

2. **Unpredictable latencies**: BigMemory keeps Java heap (or similar) sizes small to eliminate latency problems introduced by garbage collection processes over which developers have no control.

3. **Complicated deployment and management**: Simply plug BigMemory into your application and keep hundreds of gigabytes or more of data cached in memory without the need to distribute that data across multiple instances.

For maximum value and flexibility, BigMemory becomes part of a tiered-storage architecture designed to keep the most important data where it’s needed most: in the local memory of your application. For high availability, consistency and scalability, all of BigMemory’s in-memory data is available on demand from an external server array and an on-disk backing store. Any portion of your server array can go offline at any time with no application downtime and no loss of data. BigMemory is the only data management solution that offers access to terabytes of in-memory data with this level of performance, availability and operational flexibility. Furthermore, with BigMemory, you’ll have enough headroom available to meet your growing data needs on the smallest hardware footprint possible.

Figure 1: Tiered storage performance
Maximize memory use with BigMemory

BigMemory offers the best of both worlds: reliable, super-fast access times at any scale plus enterprise-level scalability and manageability. Furthermore, it is often possible to achieve significant cost savings by maximizing server density. BigMemory helps you achieve the most from the memory resources available locally on each server to avoid the costs that go along with scaling out across additional servers when it may not be needed. With BigMemory, you can scale across the minimum number of servers necessary when you require it, while avoiding the complexity and cost involved when you don’t.

BigMemory maximizes the memory available to your application within each node in your application cluster. As a result, a large multi-terabyte in-memory data solution is instantly available to you at a fraction of the number of nodes. In customer deployments, we typically see the number of servers consolidate by a factor of four or more, thereby making fewer servers work more effectively by maximizing memory use.

Predictability and low latency

The top-most tier represents the most frequently-used data that BigMemory keeps within the native application cache (i.e., Java heap), allowing for read/write latencies less than one microsecond. The tier immediately below the heap represents BigMemory’s in-memory storage that’s a bit further away from the heap and hidden from the garbage collector, thereby eliminating unpredictable pauses due to internal housekeeping processes. Distributed data caches many terabytes in size can be stored across an in-memory data grid and accessed in less than 100 microseconds with no garbage collection penalties.

Why predictability and low latency are important

Many measure the performance of an enterprise application in terms of overall throughput achieved. This may be expressed as the number of transactions per second the application can process, etc. While throughput is an important measurement, it can be misleading because it’s just an average of responses over time. For instance, systems that can process thousands of requests per second (or more) may have quite a few responses with up to a full second of latency, or greater (see Figure 2). Even though a majority of the requests—or transactions—are processed with low latency, a few large latency outliers may be enough to violate your user Service Level Agreements (SLAs) or cause unpredictable problems later in the process chain.

![System Throughput](image.png)

Figure 2: Response Time Outliers
In a Java application, for instance, most latency outliers are due to long garbage collection pauses associated with a large Java heap. To achieve the best predictability in terms of absolute performance and latency (not just the average), you need to control and eliminate the unpredictable nature of the garbage collector. BigMemory does this by keeping all of your large data sets in memory, but outside of the Java heap. This eliminates both the performance bottlenecks of disk or network-based data storage, while eliminating the garbage collector pauses associated with otherwise large Java heaps.

Because BigMemory lets you store increasingly large amounts of data in memory while keeping the Java heap small, application performance will continue to scale while latencies stay consistently small and predictable. As can be seen in the benchmarks below, BigMemory provides your application the best of all worlds when it comes to sustainable throughput, low latency and high scalability. This design is capable of delivering the appropriate balance of these three aspects of performance, including data consistency and correctness, according to the specific business needs of your application.

“BigMemory has exceeded our expectations for stability and performance.”
— Director of Enterprise Architecture, Top online travel service
Enterprise data management

Keeping all of your enterprise data in memory is a big step towards achieving the highest performance and scalability. However, this alone isn’t enough. An enterprise grade in-memory solution requires a full suite of data management capabilities, such as scalability, high availability, data consistency, monitoring and multi-data center support.

Scalability

We’ve examined how BigMemory allows you to scale up and utilize all the inexpensive RAM available in today’s servers. With a server array, BigMemory also scales out across multiple servers for unlimited scalability and high availability.

The server array is an independently scalable set of storage servers that runs on commodity hardware. This array delivers enterprise-grade data management to BigMemory in the application tier. Each server in the array has an in-memory store and a disk-backed permanent store. Similar to RAID, this array is configured into groups of servers to form mirrored stripes. The data in the server array is partitioned across the existing stripes. Over time, more stripes may be added as needed to increase the total addressable cache size and I/O throughput.

High availability: Guaranteed uptime and data access

To ensure maximum uptime and reliability, BigMemory runs in a high-availability configuration with no single point of failure. All in-memory data writes from the application layer to the server array are internally transactional and guaranteed. Any application server or server array node may be restarted or fail with no data loss.

The data management features of the server array provide a central authority that enables a number of run-time optimizations not available to other in-memory solutions. For example, transactions may be batched, folded and reordered at run time to increase throughput. Latency is minimized, because no cross-node acknowledgements are required. For high availability, each node in the server array is transactionally mirrored. Should a server node in a stripe be restarted or fail, one of the mirrors will automatically take its place, ensuring maximum uptime and data reliability.

The server array architecture allows new stripes to be added without rehashing all of the existing stripes. As a result, new stripes can be brought online instantly. Deploying BigMemory on a server array offers a number of capabilities that allow instant-on server onboarding:

- A bulk-loading mechanism that warms up new servers before adding them to the array, protecting the application from the runtime computational overhead and latency of in-memory data loading.
- A kick-start function makes new server array topology configurations instantly available to the application cluster. This means that when new servers are ready for service, the application can make use of their extra capacity immediately.
All of the high-availability features of BigMemory can scale geographically, across multiple data centers. This not only supports applications with highly distributed architectures, but also provides full continuity and disaster recovery in case you lose access to an entire data center of servers. BigMemory offers 100 percent reliability and high availability built into its architecture at all levels of scale with unmatched performance.

**The consistency spectrum**

Across the enterprise, there are typically requirements to support data access along a spectrum of consistency guarantees. This spectrum ranges from purely asynchronous operations suitable for read-only access, to fully transactional access to business-critical data at the other end of the spectrum. Because the level of consistency affects throughput and latency and is dependent on the business rules of the application, BigMemory offers configurable consistency guarantees to different data sets in the same application.

At one end of the spectrum, BigMemory allows fully asynchronous access to cached data. This yields the highest throughput and lowest latency, but the lowest consistency guarantees. In the middle of the spectrum, BigMemory enforces synchronous access to cached data. This yields a balance between fast access to data while reading and a coherent, stable view of the data as it changes. At the top end of the consistency spectrum, BigMemory enforces fully transactional, XA-compliant data access.

BigMemory ships with a default consistency setting that offers a balance of high consistency and high performance, but is easily configurable to suit the specific requirements of the application. No other in-memory data management solution supports the full range of consistency spectrum all on the same architecture and deployment topology and within the same application using the same API. BigMemory is a single solution that delivers predictable and cost-effective performance at all levels of scale and consistency.

**IoT and digital business in real time**

With all your data in memory available for real-time analysis, you can accelerate business decisions and satisfy customer requests much more quickly when compared to disk or network-based retrieval. Not only does this give you a potential competitive advantage and increased customer value, it allows you to harness value in your data that may have been inaccessible before, especially where the opportunity window for taking action on gained insights are brief.

Data mining for valuable, marketable, data relationships and associations is often performed in off-line batch processes. With BigMemory’s in-memory data management, this processing can be performed while customers use your application. As a result, your application can personalize the user experience in real time, adding value to the user, and providing new opportunities to market to your customers’ needs as they arise. For example, in a recent customer deployment, BigMemory reduced risk analysis calculation processing time from minutes to seconds, enabling them to act on their risk data during the customer transaction rather than after the fact.

BigMemory’s support for in-memory data also makes it possible to instantly stream large amounts of data to mobile devices, without waiting for disk-based retrieval or impacting other users on the system.

**In-memory performance for the Internet of Things (IoT)**

In-memory data management is a vital component when building agile solutions ready to meet the challenges presented by an ever more connected world. IoT applications have to scale and be able to deal with streams of data from millions of connected devices to service potentially millions of concurrent user requests. BigMemory makes it possible to handle the data velocities demanded by IoT use cases in order to enrich, analyze and act in real-time.
BigMemory is the first-choice platform for enterprises that require distributed in-memory data management to support IoT solutions demanding extremely low, predictable latency at any scale:

- Capture real-time data flows of any type of data, to and from any device
- Accelerate time to insight: gather, sort, analyze and act on data faster than your competition
- Hundreds of terabytes of heterogeneous data can be maintained in memory, with guaranteed latency of low milliseconds
- Data can be distributed across multiple commodity servers for redundancy

---

**Customer example: Leveraging the IoT to deliver new levels of service**

Our customer, a global provider of printing and content management solutions, wants to implement an innovative service offering preventative and corrective maintenance for its portfolio of connected devices in use at customer locations.

The solution will take real-time status and sensor data feeds from hundreds of thousands of individual devices and compare, contrast and combine these with in-house data to service those devices.

With BigMemory capturing the high volume of incoming device data and providing the high-speed platform for data access, the company can build new applications to monitor machines in real time and proactively address service issues, thus offering an enhanced service for managing equipment installed on premise. In this way, this company is able to bring to market a ground-breaking new level of service that strengthens its competitive position.

**BENEFITS**

- Higher value service contracts and enhanced customer loyalty
- More efficient use of company resources to remotely track and manage devices
- Able to set aggressive target for growth of their managed services business (from 17% to 50% over the next 5 years)

---

**Disruptive business models require a Digital Business Platform**

The need for innovation has never been greater. To remain competitive, successful organizations are implementing digital transformation strategies in order to build the new solutions needed to meet and exceed the growing expectations of customers.

Software AG’s Digital Business Platform is an integrated, flexible suite of technologies and tools to manage the digital transformation in order to increase operational efficiency, improve customer experiences and create new business models. It offers best-in-class IoT platform services and edge services for the cloud, hybrid and on-premise to help enterprises build highly innovative IoT frameworks. Platform services include device connectivity and management, in-memory storage, streaming and predictive analytics with connection to machine learning, enterprise and cloud integration, API management, process design, application enablement and data analysis/visualization capabilities.
Customer example: Groundbreaking innovation in the insurance industry

This company saw an opportunity to use internet-connected technologies to transform the auto insurance industry. Its devices are installed in over 15 million vehicles capturing behavioral, contextual and driving data, which is then analyzed and added to one of the world’s largest driving telemetry databases with more than 380 billion kilometers of data and analysis of more than 265,000 accidents.

The challenges of integration, process management, real-time analytics and operational intelligence all require super-fast access to a mix of high-velocity data and contextual data-at-rest.

BigMemory provides the in-memory data fabric needed to deliver the combination of robustness and performance required to transform huge volumes of driving data into new telematics solutions. Its customers include insurance companies, car rental and fleet management organizations, motor manufacturers and governmental authorities, all of whom use the company’s solutions respectively to price risk, manage and pay claims, and maximize their customer relationships.

Using BigMemory and Software AG’s Digital Business Platform to leverage big data and the IoT, this organization is able to meet new customer demands and bring innovative new products to market.

BENEFITS

- Building a new, scalable telematics monitoring platform
- Able to create completely new markets
- Analyze driver behavior to offer personalized insurance policies and real-time crash detection
- Help consumers who demonstrate lower risk to save money
- Lower liability exposure for insurers
BigMemory and webMethods

BigMemory is shipped with webMethods, the leading solution for integrating systems and business processes, within and between connected enterprises. WebMethods reduces effort, time-to-market and ongoing operational costs. Together, webMethods and BigMemory provides the foundation of Software AG’s Digital Business Platform, acting as the glue that connects its capabilities, as well as delivering the ability to perform at any scale with the agility needed to build applications ready for constant change.

As an example, a large international retailer ran into problems when attempting to apply its existing infrastructure to support new online business opportunities. The solution had to manage an online catalog for 100,000 products and provide all 500+ branches with the ability to choose which items to stock and assign their own pricing.

The need to render item availability and pricing per shop online in real time (from a total set of more than 40 million potential combinations) proved too much for existing systems. Response times from the incumbent mainframe-centered architecture were poor, too costly to maintain, and could not scale to cope with peak hour demands combined with the need to support large file processing. Data latency was seen to cause significant loss of business.

The customer utilized the native integration between BigMemory and webMethods Integration Server to quickly implement a solution that pulled data from the mainframe and stored it in memory for access by its online applications, massively accelerating the large amount of static data lookups required. BigMemory delivers the scalability, availability and performance needed for real-time data access by the online search and recommendation engines.

Figure 6: webMethods integrated within the solution architecture.

Improving the online experience reduced customer attrition, increased revenues and strengthened competitive position by providing:

- Fast, uniform data access across multiple platforms
- In-memory management and enrichment of consolidated data
- Product-store combinations available online with localized pricing

The solution also sets the stage for growth as more products and branches can be added with no impact on latency. Additionally, more applications can be easily integrated using webMethods to offload data into BigMemory, further improving performance in the retailer’s e-commerce environment.
BigMemory everywhere

Companies large and small have turned to BigMemory to accelerate business-critical, data-intensive applications and analyses. This is because BigMemory is broadly applicable, suitable for many types of applications, as well as enterprise level initiatives focused on opportunities available through digital transformation and IoT solutions.

In fact, you may be using BigMemory in common enterprise applications today without even knowing it. Have you booked a flight online? Some of our customers use BigMemory to speed up that highly transactional web-based process. Have you charged dinner on your credit card? BigMemory is currently being used for real-time fraud detection, scanning through hundreds of gigabytes of bank data in the second or two it takes to get an approval. Have you streamed video to your mobile device? BigMemory provides the scale needed to serve data to thousands of concurrent viewers without requiring a data center full of hardware.

Here are just a few more examples of how BigMemory is making a big difference for market leaders in a wide range of industries.

<table>
<thead>
<tr>
<th>FINANCIAL SERVICES</th>
<th></th>
</tr>
</thead>
</table>
| A global financial services firm processes trade reconciliations in a tight four-hour window—accomplishing what canned database reports couldn’t. | - Accelerate the processing of trade orders, credit card authorizations and other high-volume transactions  
- Speed up large-scale data analysis for risk management, asset management or real-time fraud detection  
- Provide fast, reliable access to aggregated account data on customer portals  |

<table>
<thead>
<tr>
<th>TELECOMMUNICATIONS</th>
<th></th>
</tr>
</thead>
</table>
| A major broadband carrier improved the performance of their billing system, boosting processing success rate from 80 to 99 percent. | - Speed up billing, subscriber provisioning and other high-volume transactions  
- Improve call center efficiency with faster access to account data  
- Expand subscriber self-service offerings with scalable online applications  
- Implement a scalable, ultra-fast solution for network management  |

<table>
<thead>
<tr>
<th>HIGH-TECH, INTERNET AND ONLINE GAMING</th>
<th></th>
</tr>
</thead>
</table>
| A leading cloud service provider has achieved 100 percent uptime for its online meeting service by storing session data in memory. | - Accelerate searches, purchases, ad placements and other online transactions  
- Handle spikes in demand and long-term traffic growth  
- Ensure stable response times at any scale  
- Boost service uptime  |

<table>
<thead>
<tr>
<th>ENTERTAINMENT AND MEDIA</th>
<th></th>
</tr>
</thead>
</table>
| A U.S. cable operator guarantees seamless TV viewing on the iPad® with real-time user authentication. | - Scale web services to millions of concurrent users/viewers  
- Display dynamic or aggregated data at lightning speed  
- Ensure a superior user experience  
- Reduce hardware costs  |

<table>
<thead>
<tr>
<th>ENTERTAINMENT AND MEDIA</th>
<th></th>
</tr>
</thead>
</table>
| Europe’s leading hotel portal ensures speedy online reservations—reducing database use by more than 50 percent at the same time. | - Increase throughput rates for booking, ticketing and other high-volume transactions  
- Instantly display such dynamic data as weather, delivery status, arrival time or traffic updates to enhance the value of customer portals  |

<table>
<thead>
<tr>
<th>ENTERTAINMENT AND MEDIA</th>
<th></th>
</tr>
</thead>
</table>
| A U.S. government agency can now meet internal SLAs for three applications in two data centers. | - Improve the performance and scalability of mission-critical applications  
- Support real-time data analysis and high-velocity data processing  
- Deliver web services scalable to millions of citizens  
- Comply with mandated SLA  |
Get started

Try BigMemory free for 90 days.

Register for a free trial at www.terracotta.org/downloads/bigmemory.

For more information, talk to your Software AG representative and visit www.softwareag.com/terracotta.