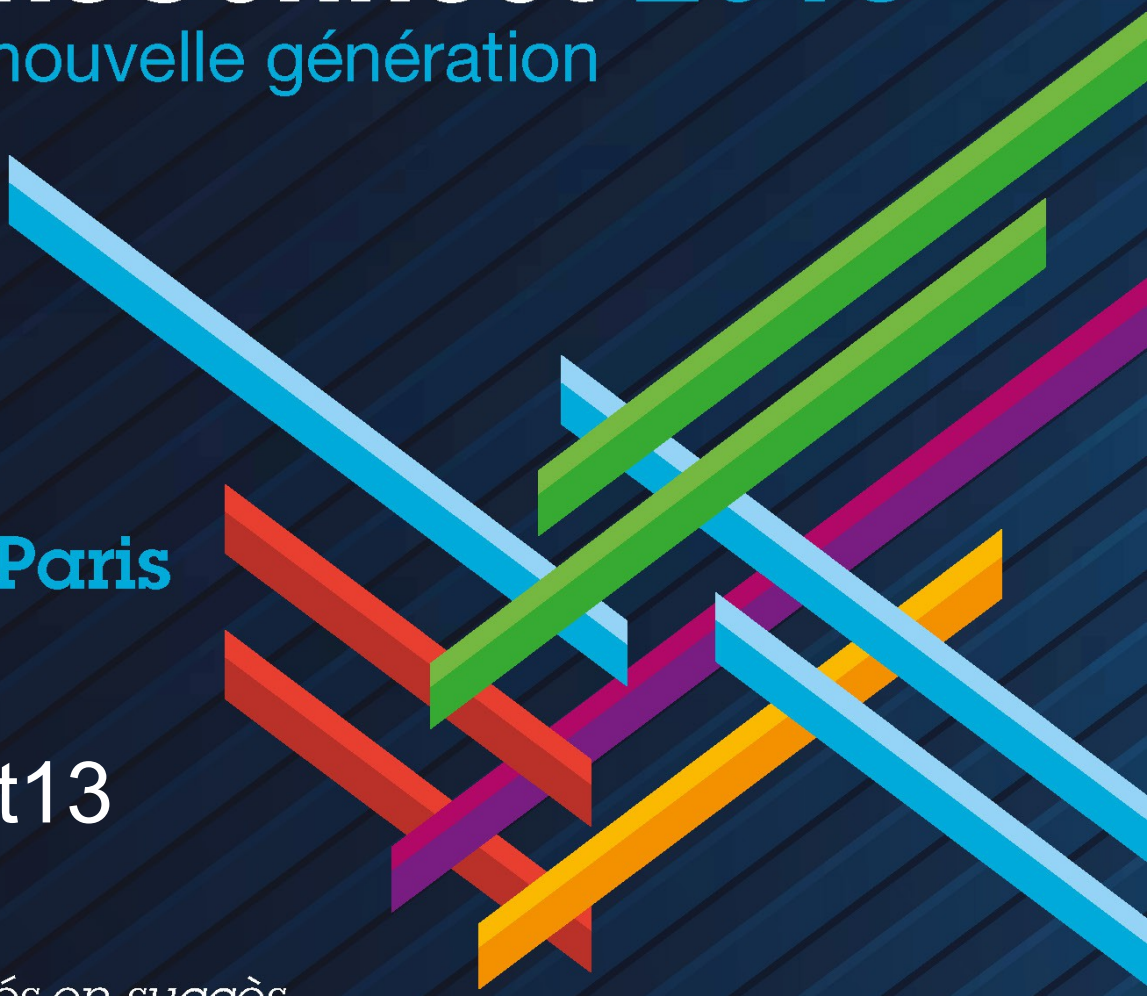
IBM SolutionsConnect 2013

L'IBM TechSoftware nouvelle génération



28, 29 et 30 août

IBM Client Center Paris

#solconnect13

Transformez vos opportunités en succès

IBM

WebSphere eXtreme Scale
WebSphere DataPower XC10 Appliance

les solutions de caching élastique

Hervé Grange
hgrange@fr.ibm.com
WebSphere Technical Sales and Solutions

Agenda

- Un cache élastique?
- WebSphere eXtreme Scale
- WebSphere DataPower XC10 Appliance
- Les usages
- References

Retail, Banking, Finance, Insurance, Telecom, Travel & Transportation …

Une explosion de la charge

Laptops, Ultrabooks, Tablets, Smartphones + Mobile Access  **3G, 4G, free wi-fi** = **Transaction Overload**

TV, Movie, Sports Personality mentions / endorses product + Social Media  **YouTube, twitter, Facebook** = **Transaction Overload**

E-mail, SMS, Pop-up, Click-thru Promotions, Web Crawlers + Targeted Advertising  **All of the above** = **Transaction Overload**

Le Cache élastique minimise l'impact d'une surcharge transactionnelle

Web Server Tier

IBM HTTP Server

App Server Tier

WebSphere Application Server

In Memory Data Grid

Improve Performance, Scalability & Availability

Highly Scalable Web Applications

Data-intensive Applications Extreme Performance

Back-end Systems Database Tier

Information Management

DB2 UDB

Pourquoi un cache mémoire en grille ?

Scalability issues with database/back end

- Adding extra hardware is not easy
- Licensing costs
- MIPS charges

Large volume of data

- Ability to handle increasing amount of data
- Handle data surges during product launches and live events

Fault tolerance and self-healing

- Need for automatic mechanisms to avert system failure affecting end-users
- Data integrity

Data redundancy and replication

- Maintain data reliability in case of failover

Qu’est-ce qu’un cache mémoire en grille (Datagrid)?

Distributed in-memory object cache

- Elastic, scalable, coherent in-memory cache
- Dynamically caches, partitions, replicates and manages application data and business logic across multiple servers

Capable of massive volumes of transactions

- Provides qualities of service such as transaction integrity, high availability, and predictable response times

Self-healing, allow scale-out / scale-in

- Automatic failure recovery
- on-the-fly addition / removal of memory capacity

Splits a given dataset into partitions

- Primary and Replica shards

Maturité des solutions

Forrester Research
MAKING LEADERS SUCCESSFUL EVERY DAY

*"IBM has a **strong strategy and execution plan** for WebSphere eXtreme Scale, as well as strong product features across the board."*

*"eXtreme Scale is especially strong in **runtime architecture**, distributed caching features, **performance** and **scalability** features, and support of **standards**."*

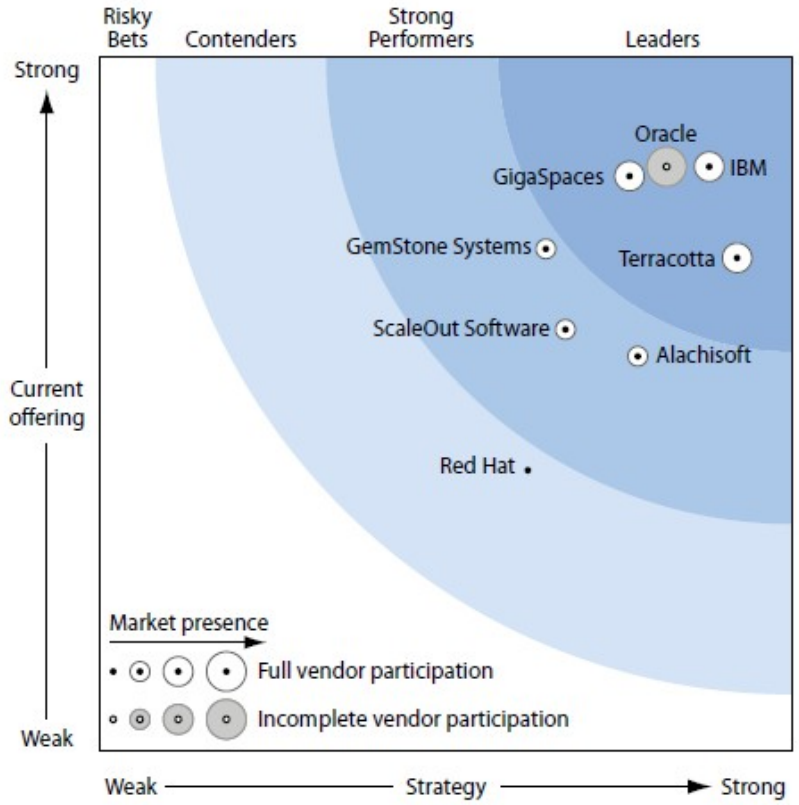
Gartner

"By 2014, at least 40% of large organizations will have deployed one or more in-memory data grids. Today, we estimate that less than 10% of large user organizations have IMDG products deployed in production."

May 14, 2010

The Forrester Wave™: Elastic Caching Platforms, Q2 2010

by Mike Gualtieri and John R. Rymer
for Application Development & Delivery Professionals



Etude de Forrester TEI sur WebSphere eXtreme Scale

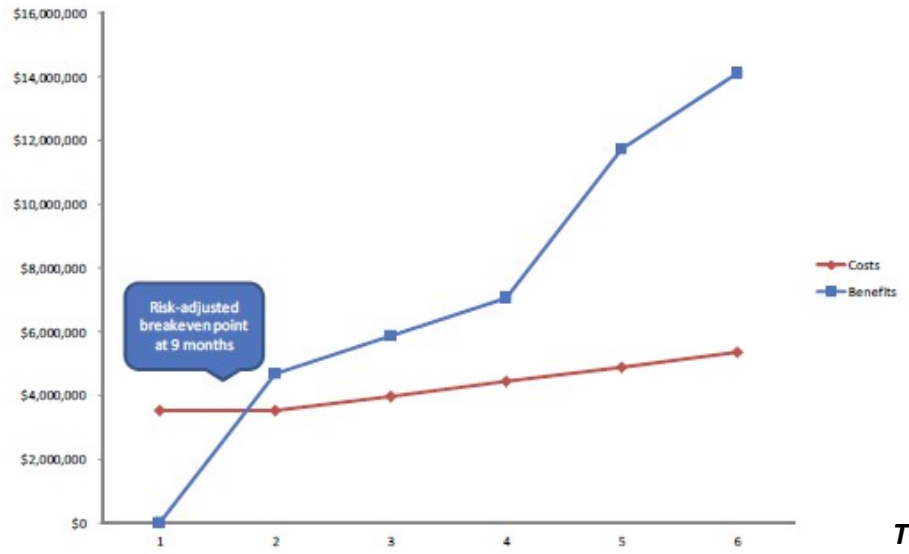
A Forrester Total Economic Impact™ Study Prepared For IBM

Total Economic Impact ™ Of IBM WebSphere eXtreme Scale February 2012

The financial analysis found that the organization experienced:

- ROI of 123%
- Payback Period of 9 months
- Net Present Value of $5.9 million

Three-Year Risk Adjusted Analysis



Source: Forrester Research, Inc.

- **Benefits.** The organization Forrester interviewed experienced the following benefits:
  - **Reduction in hardware and software costs.** This benefit represents the hardware and software savings associated with eliminating the need to expand to additional databases.
  - **Annual ongoing staffing costs**. This represents the savings from the ongoing maintenance of additional databases.
  - **Incremental gross revenue (Not quantified)** This benefit represents the incremental revenue associated with the mitigation in user drop-off when users experience slow response resulting from a large surge in traffic during live events and product launches

The following is taken from a commissioned study conducted by Forrester Consulting on behalf of IBM."

Le Cache élastique minimise l'impact d'une surcharge transactionnelle

Web Server Tier

WebSphere. software

IBM HTTP Server

App Server Tier

Application Server

In memory data grid

① 

DataPower XC10 for simple data oriented scenarios.

② WebSphere. software

eXtreme Scale for maximum flexibility and sophisticated application oriented scenarios.

Back-end Systems Database Tier

Information Management

DB2 UDB

Innovative Elastic Caching Solutions

"Data Oriented"

DataPower XC10 Appliance

- Drop-in cache solution optimized and hardened for data oriented scenarios
- High density, low footprint improves datacenter efficiency

Session Management

Extension for DynaCache

Side cache

Worldwide cache

Database buffer

Business Event Processing

Petabyte analytics

Extreme Transaction Processing

WebSphere. software

eXtreme Scale

- Ultimate flexibility across a broad range of caching scenarios
- In-memory capabilities for application oriented scenarios

"Application Oriented"

Elastic caching for linear scalability
High availability data replication
Simplified management, monitoring and administration

XC10

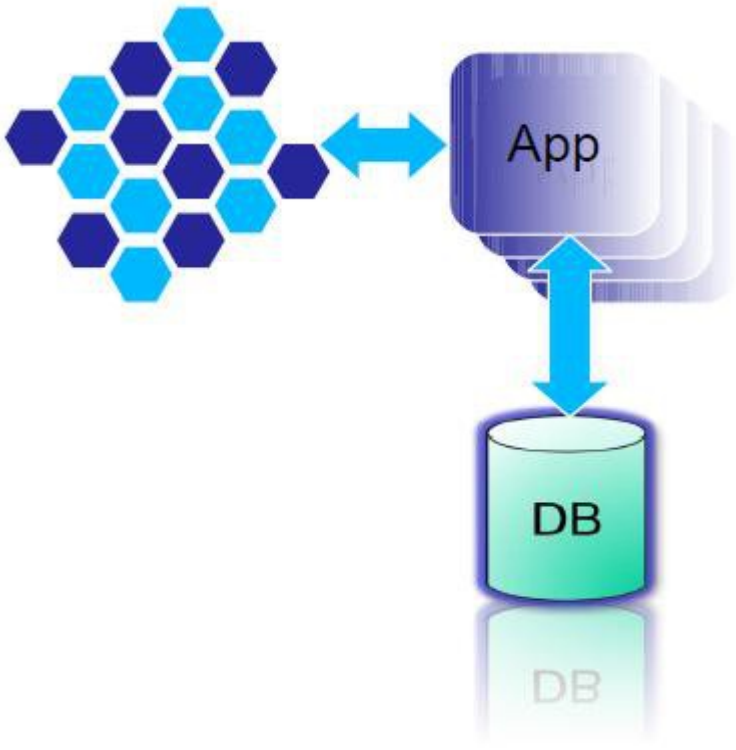
WebSphere software

eXtreme Scale

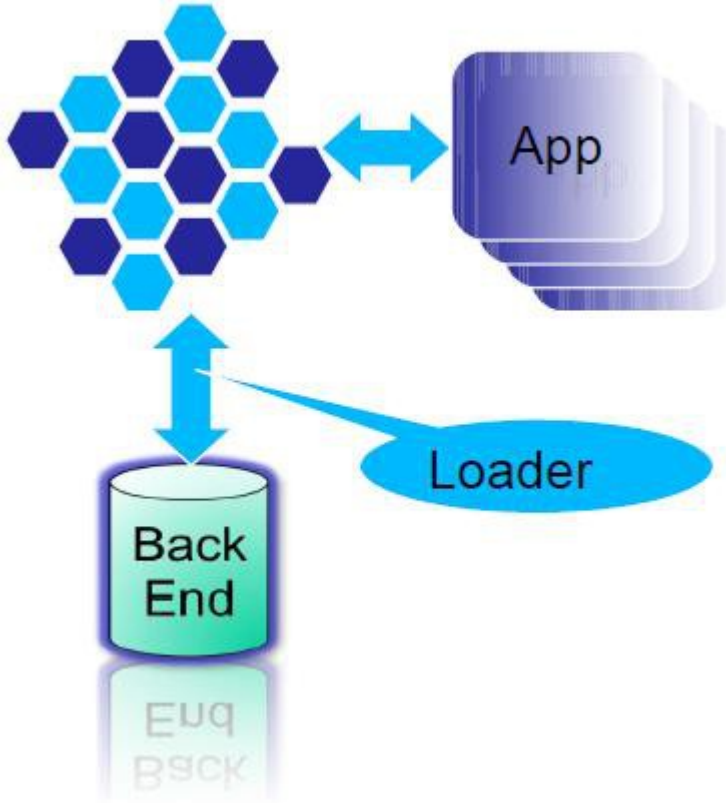
WebSphere software

eXtreme Scale

- Side Cache



Inline backend cache



Client cache within/without WebSphere Application, .Net, java or non-java (REST)

The XC10 appliance provides a complete, purpose-built, easy-to-use solution for common distributed caching scenarios

- Allows businesses to leverage the value of existing infrastructure investments
- Provides “drop-in” use for **Http Session** Management, Extended **DynaCache** service, and Simple Grid / ESB Caching scenarios, requiring little or no code changes to existing applications
- Offers the ability to quickly and easily increase **cache capacity and throughput as needs grow**
- Includes a flexible and simple user management interface for **monitoring and administration**
- Provides a large **240 GB** cache with near-linear scalability
- Reduces risk of data loss via **automatic replication**, delivering **high availability** and fault tolerance for higher user satisfaction and faster task completion

Elaborer un cache élastique via une appliance

WebSphere DataPower XC10

New!

.

What's new :

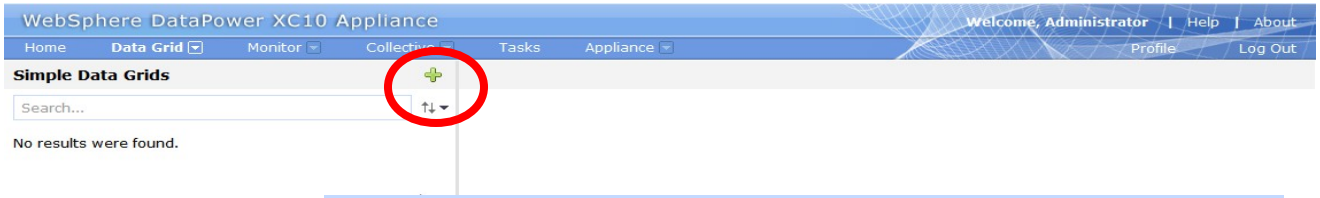
- **Multi-data center support** allows customers to host data on XC10s in multiple locations with data kept in synch through multi-master replication
- **Support for elastic caching for WAS Liberty** brings scalability, fault tolerance and high availability
- **Dynamic cache replacement** for WebSphere Portal
- **Monitoring enhancements** ease administration and improve serviceability
  - Support for Simple Network Management Protocol (SNMP) traps for event notification
  - Capability to query grid contents
- **Spring Cache 3.1 Support**
- **.Net client delivery**
- **Near-cache invalidation**

Improved TCO, increased operational efficiency and productivity, and better response time

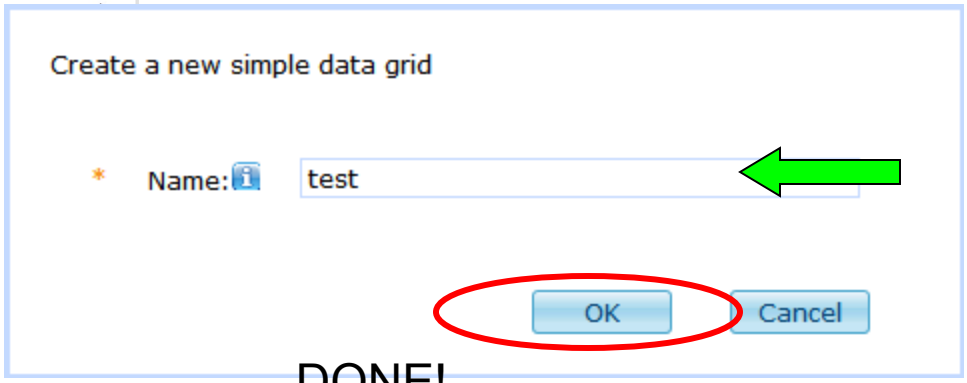
Créer simplement une Grille de cache sur le XC10
3 sélections

1.

2.



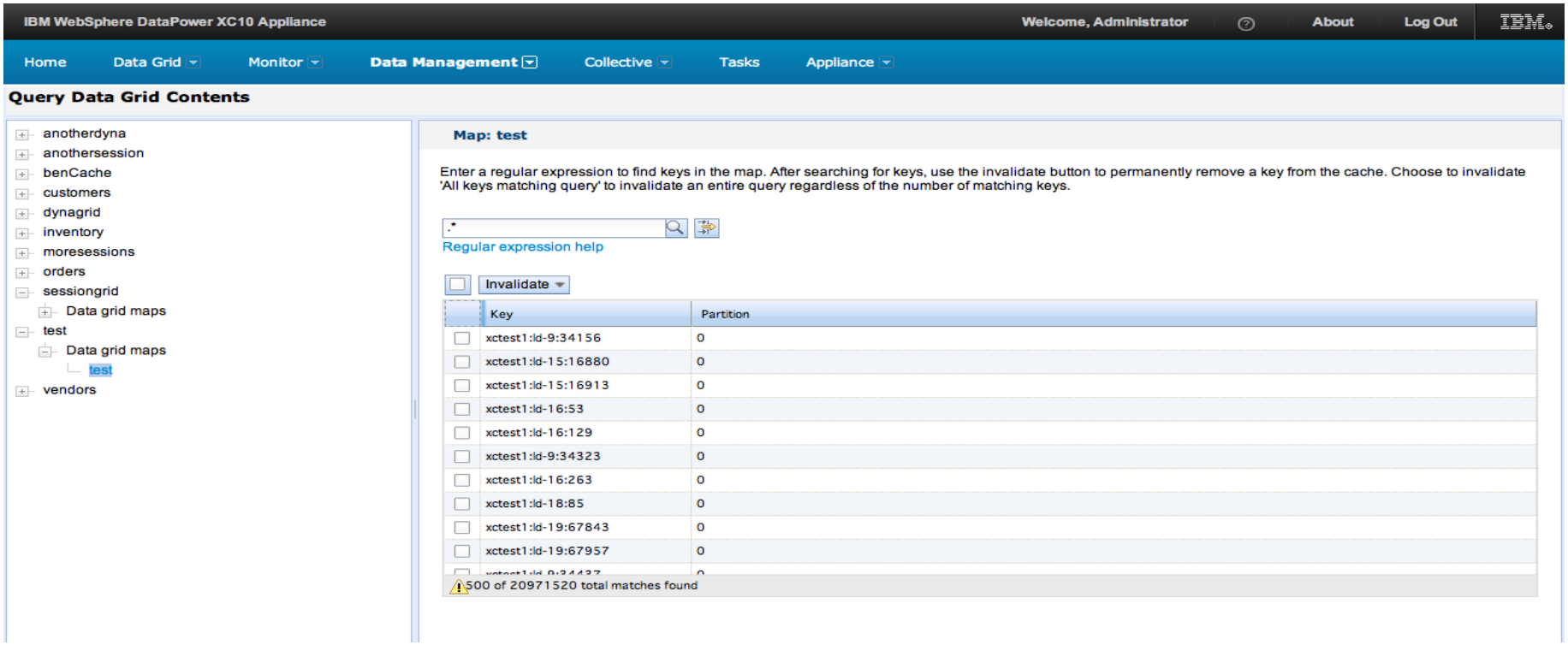
3.



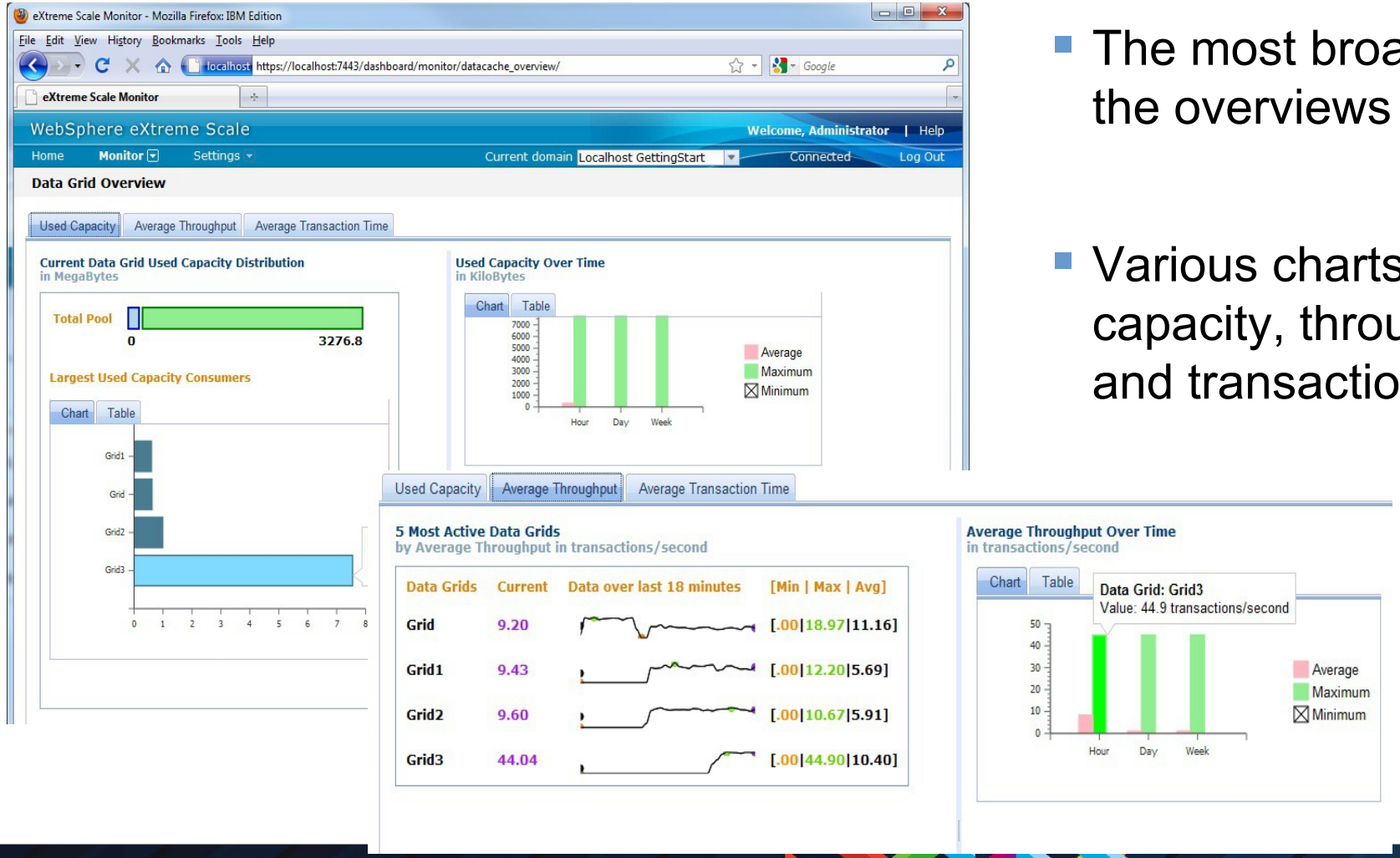
DONE!

Query Console – WXS & XC10

Allows monitoring, inspection, and invalidation of specific cache content



Monitoring console, Overview of all Data Grids



- The most broad of the overviews
- Various charts on capacity, throughput, and transaction time

The XC10 Appliance Virtual Image for Developers

http://www.ibm.com/developerworks/downloads/ws/wsdvad/

Get started using the WebSphere DataPower XC10 Appliance with just a download

The WebSphere® DataPower XC10 Appliance virtual image is an open virtual appliance (OVA) package that you can import into your virtualization environment. With this image, you can use the caching appliance software for creating and managing simple, session, and dynamic cache data grids.

Downloads

(Updated 07/22/2013)

XC10-Developer_2.5.0.1.zip

- For the application , need to download the cache client :

http://www-01.ibm.com/support/docview.wss?uid=swg24035358

WebSphere eXtreme Scale Client standalone	19 Jul 2013	Language Independent	3175429917	FC DD
WebSphere eXtreme Scale Client for WAS 7	19 Jul 2013	Language Independent	584367293	FC DD
WebSphere eXtreme Scale Client for WAS 8	19 Jul 2013	Language Independent	686793175	FC DD
WebSphere eXtreme Scale Client for .NET	19 Jul 2013	English	20892392	FC DD

WebSphere eXtreme Scale

WXS provides an in-memory data grid which, dynamically processes, partitions, replicates, and manages application data and business logic across hundreds of servers

- **Automatically** handles **replication** which can be either synchronous or asynchronous
- Handles **advanced placement** so that replicas can be placed in **different physical zones**
- **Fully elastic** in that servers can be added and removed a automatically redistributes data
- Allows clients using **different object representations** to **share data stored** in the data grid
- Provides automatic **integration with databases**
- Provides **HTTP Session Management**
- Flexible deployment model allowing significant **customization**
- Proven **multi-data center capabilities**
- Proven **low-latency access to data**

In Memory Data Grid for elastic scalability and next generation cloud environments

IBM WebSphere eXtreme Scale

- **A powerful, scalable, elastic in-memory grid** for your business-critical applications
- **Improves business agility** by allowing you to overcome traditional IT performance limitations
- **Scales elastically** at runtime to thousands of JVMs, limited only by available physical machines

What's new :

- **Support for elastic caching for WAS Liberty** scalability, fault tolerance and high availability
- **Dynamic cache replacement** for Portal content rendering improves throughput and performance.
- **Support for Install Manager** in WAS & Liberty environments reduces deployment time and costs
- **A Grid Query Console** monitoring, inspection, and invalidation of specific cache content
- **Resource Adapter for Java EE platforms**. Allows integrating the WXS transaction with an EE-managed transaction.
- **eXtreme Memory** support in WebSphere Application Server processes
- **WXS console** support on all 64-bit platforms and WebSphere
- **Spring 3.1 Cache** integration support.
- **Continous Query, Disk OverFlow, HPEL,**

Improved TCO, increased operational efficiency and productivity, and better response time

IBM® WebSphere® eXtreme Scale V8.6

Powerful In Memory Data Grid capabilities for the Next Generation of Cloud and Mobile Applications

WebSphere eXtreme Scale provides enhancements to support Enterprise-wide Data Grids

What's New

- Java and .NET applications can now interact natively with the same data in the same data grid, leading the way toward a true enterprise-wide data grid.
- A new REST Gateway provides simple access from other languages.
- WXS 8.6 delivers a faster, more compact serialization format called eXtreme Data Format (XDF), which is neutral to programming languages.
- A new transport mechanism, eXtreme IO (XIO) removes the dependency on the IBM ORB, enabling easier integration with existing environments.
- Built in pub/sub capabilities enable WXS 8.6 to update client "near caches" whenever data is updated, deleted, or invalidated on the server side.
- API enhancements enable continuous query of data that is inserted and updated in the grid.
- Enhanced Monitoring capabilities for better grid management and visualization

Features/Business Value

- Support for non-Java applications enables single, enterprise wide solution.
- Better integration with existing environments including IBM portfolio products WebSphere Message Broker and WebSphere DataPower XI52.

Client Benefits

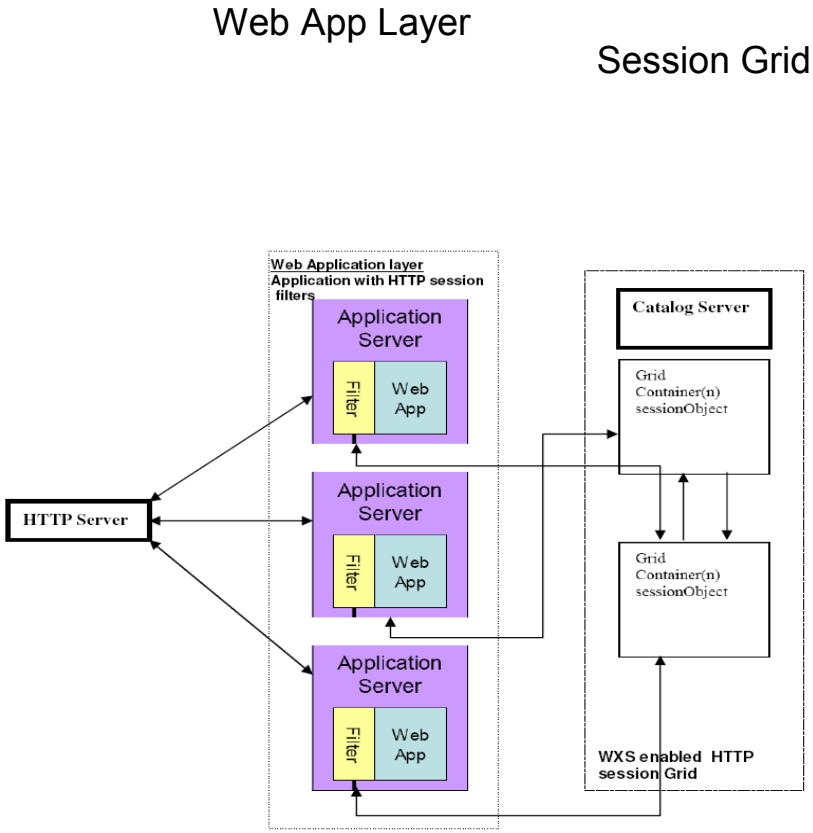
- Consistent and predictable performance for business-critical applications, Elastic scalability,High availability of data
- Reduced load on back-end systems

Common Elastic Caching Scenarios

HTTP Session Replication
Side Cache
Data Access Layer
Data Grid + Compute Grid

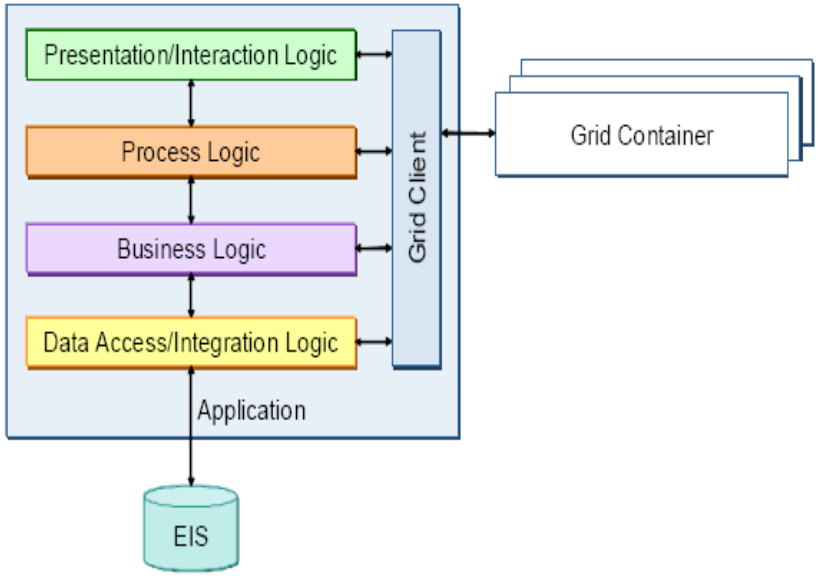
HTTP Session Replication

- Many enterprise applications today require HTTP session persistence.
- A grid of JVMs can be established with the sole purpose of storing HTTP session (or any java) objects.
- Isolating the application runtime from grid runtime, thereby, freeing up the JVM heap for application use.
- Provide linear scalability to accommodate growth (in # of sessions or size of session objects).
- Providing replication and management of session objects within the grid.
- Can even store session objects across datacenters.



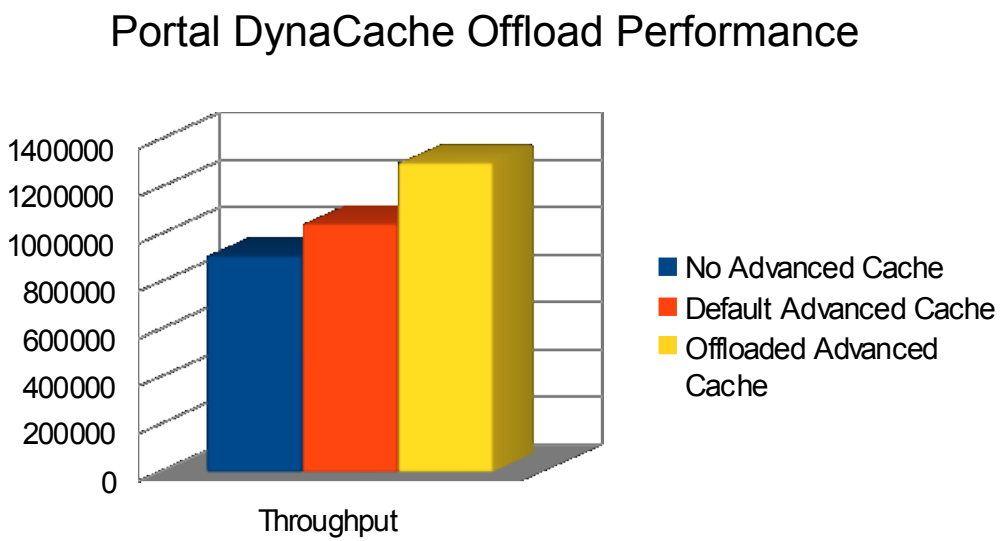
Side Cache /XS API/Dynamic Caching

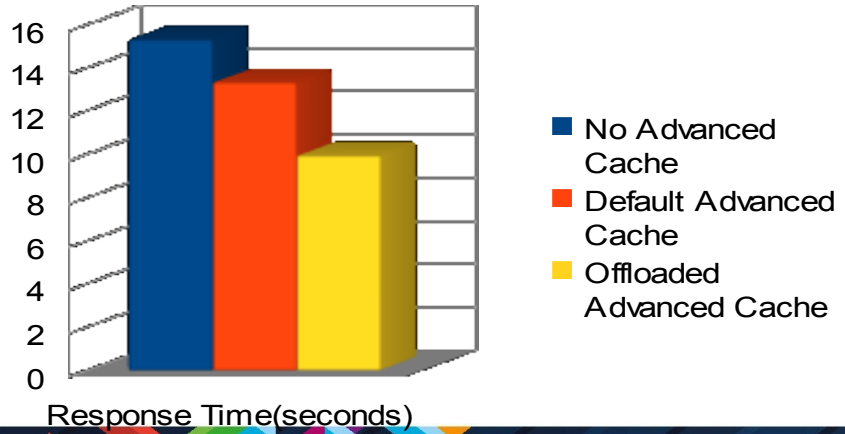
- Client first checks the grid before using the data access layer to connect to a back end data store.
- If an object is not returned from the grid (a cache “miss”), the client uses the data access layer as usual to retrieve the data.
- The result is put into the grid to enable faster access the next time.
- The back end remains the system of record, and usually only a small amount of the data is cached in the grid.
- An object is stored only once in the cache, even if multiple clients use it. Thus, more memory is available for caching, more data can be cached, which increases the cache hit rate.
- Improve performance and offload unnecessary workload on backend systems.



Portal Customer Experience – Steady State Comparison

- With Elastic Cache average throughput in our steady state/concurrent user scenario was consistently faster than with Default Advanced Cache
  - 30% throughput improvement in our scenario
- With Elastic Cache average steady state response-times are consistently faster than with Default Advanced Cache
  - 5.5 second improvement observed in our scenario
  - Customer results may vary
- Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. Actual performance in a user's environment may vary.

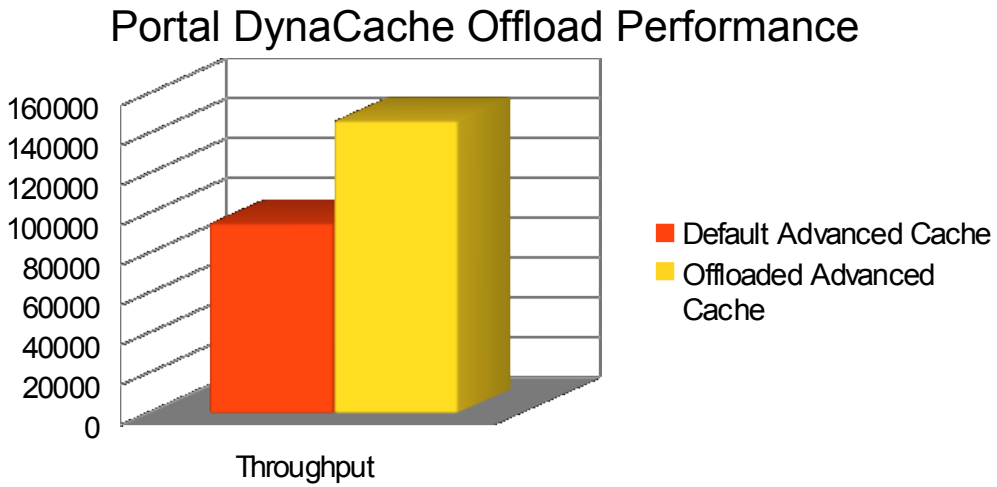


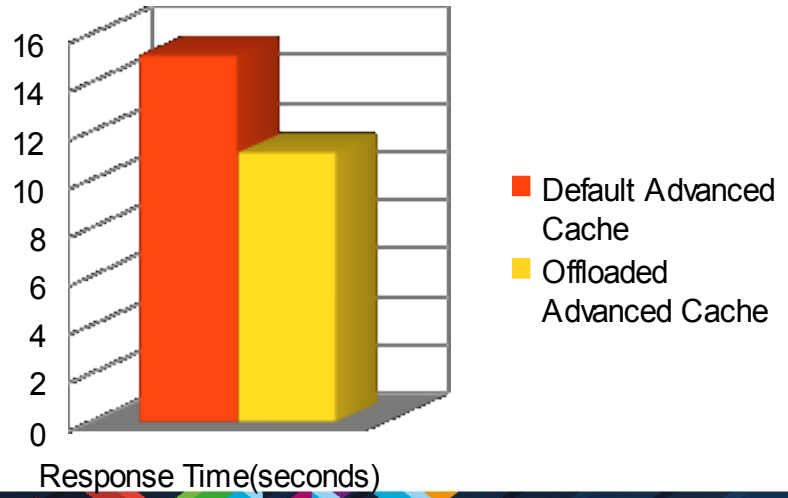


Reducing Portal warm-up time – Cold Start Results

- With Elastic Cache average throughput of a newly started server consistently faster than with Default Advanced Cache
  - 54% throughput improvement in our scenario
- With Elastic Cache average response-times are consistently faster than with Default Advanced Cache
  - 4 second improvement observed in our scenario
  - Customer results may vary
- With Elastic Cache response times improve faster due to faster cache hydration

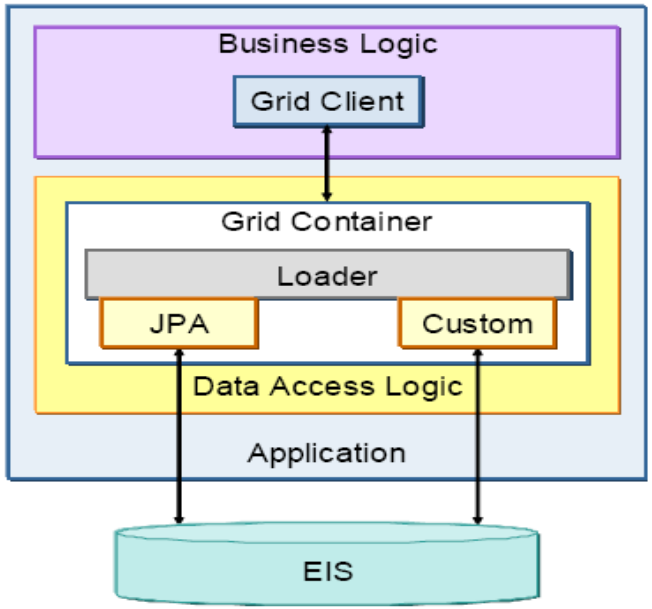
- Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. Actual performance in a user's environment may vary.



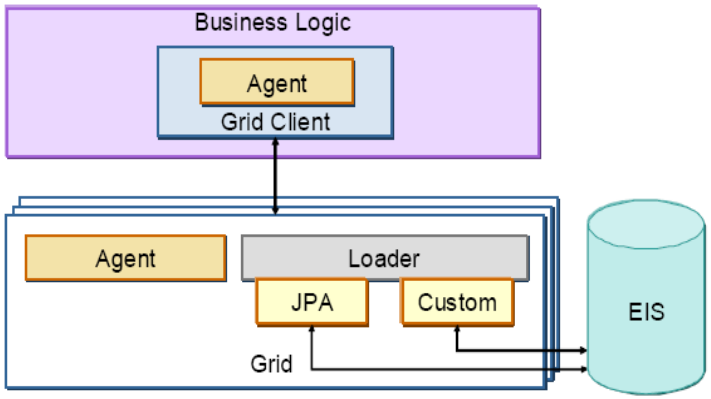


Data Access Layer – (WXS only)

- The grid can be used as a special data access layer where it is configured to use a loader to get data from the back-end system.
- The loader can be configured to asynchronously write back the changes to the back-end system
- Even when the back end is down, the application can write changes. They are buffered in the grid to be written back when the back-end system comes online again.
- Reduced dependency on back end data store



Data Grid + Compute Grid– (WXS only)



- The DataGrid scenario is the natural extension of the data access layer scenario.
- Grid is the primary system of record.
- The grid is used to update and query the data.
- Write-behind functionality allows for the grid to push changes to the data into the back-end data stores.
- Again, the grid scales linearly simply by adding new grid containers.
- Performance and efficiency benefits.
- Little to no load on back-end data stores

Objects in the Elastic Data Grid – Key, Value pairs

Key Value pair examples:

- integer key, string value
- (hex) integer key, compound value
- compound key, compound value

ObjectMap

Key	Value
1	Daniel
2	Jennifer
…	…

ObjectMap

Key	Value
A123B45C	John::Person Name: John Last name: Doe Birth date: 08.07.1999
D567E89F	Jane::Person Name: Jane Last name: Doe Birth date: 03.05.1979
…	…

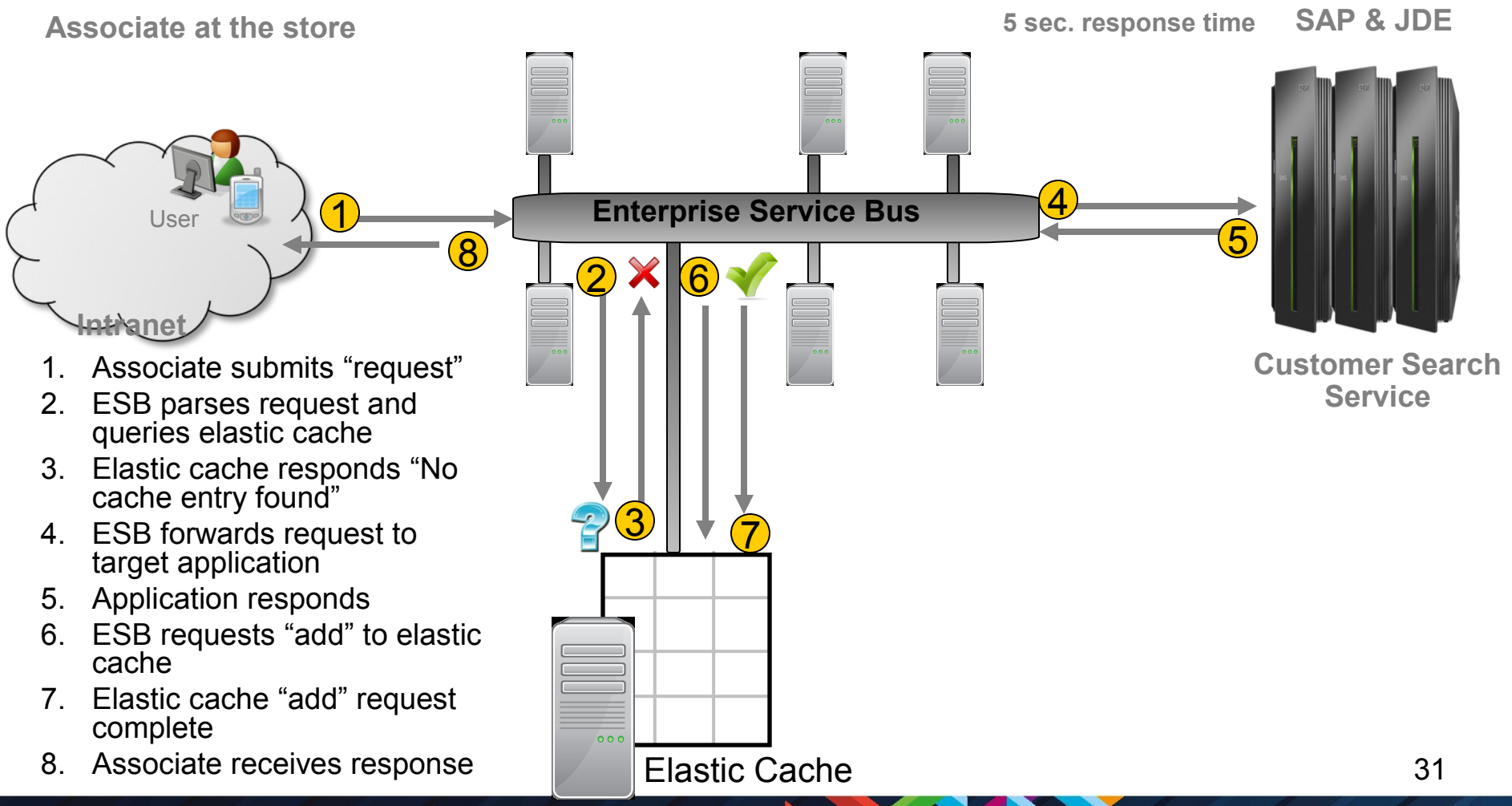
ObjectMap

Key	Value
John Key::Person Key Last name: Doe Birth date: 08.07.1999	John::Person Last name: Doe Birth date: 08.07.1999 Name: John Gender: Male
Jane Key::Person Key Last name: Doe Birth date: 03.05.1979	Jane::Person Last name: Doe Birth date: 03.05.1979 Name: Jane Gender: Female
…	…

XC10/XS REST Gateway

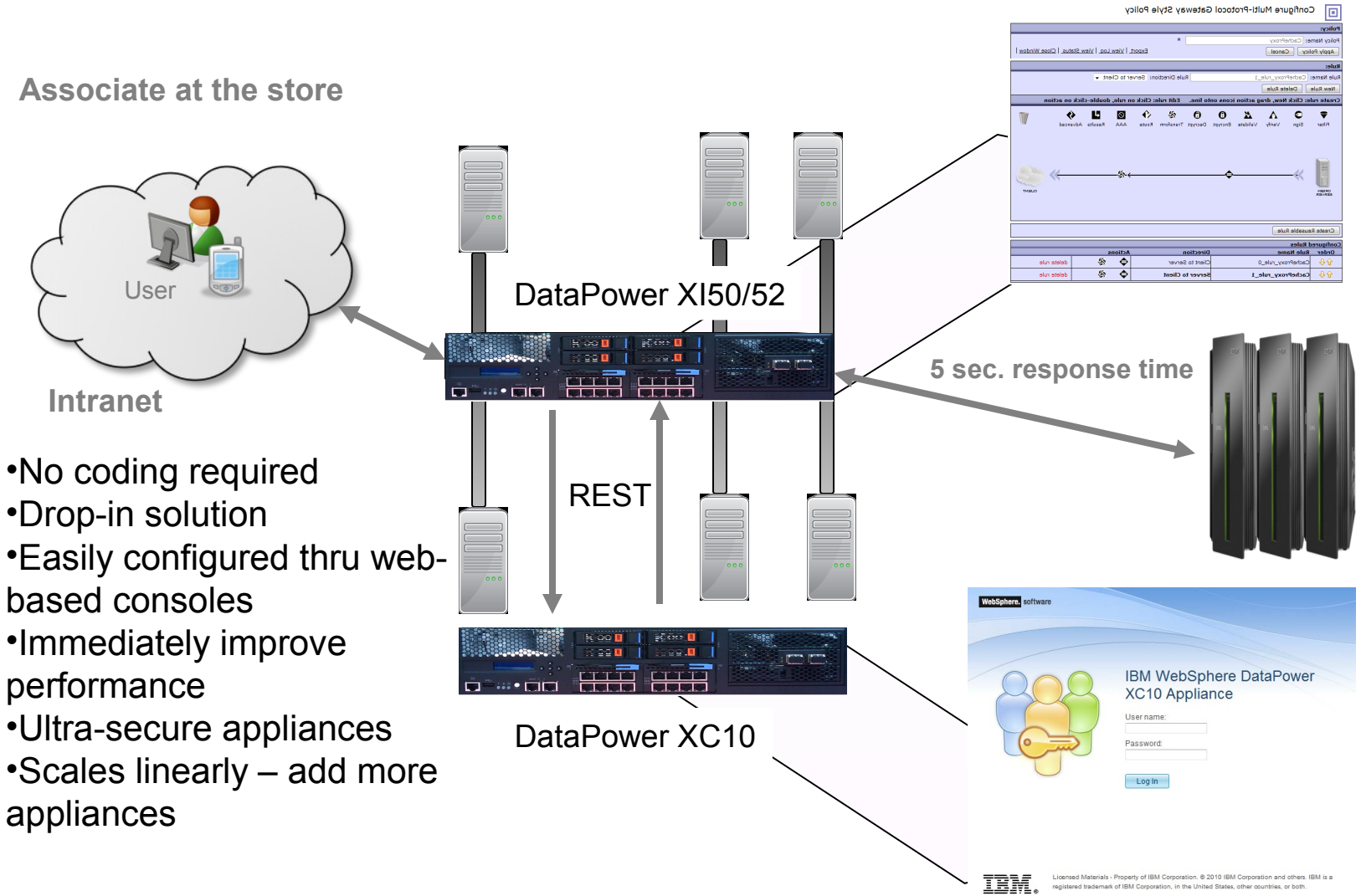
- Enables non-Java™ based clients access to simple data grids using a set of HTTP based operations. (e.g. .NET, php)
- Simple HTTP methods
  - HTTP POST method to insert/update data in the grid
  - HTTP GET method to get data from the grid
  - HTTP DELTE method to delete data from the grid.
- Supports creation of dynamic maps with 3 Time to Live (TTL) templates
  - None (no time to live expiration)
  - Last Access Time
  - Last Update Time
- non-Java™ clients and Java™ clients can access the same data grids

Customer Search Service with elastic cache as side cache for ESB
Request flow for a “cache miss”



1. Associate submits “request”
2. ESB parses request and queries elastic cache
3. Elastic cache responds “No cache entry found”
4. ESB forwards request to target application
5. Application responds
6. ESB requests “add” to elastic cache
7. Elastic cache “add” request complete
8. Associate receives response

WebSphere DataPower XC10 used as a Side Cache for IBM Datapower XI50/XI52



- No coding required
- Drop-in solution
- Easily configured thru web-based consoles
- Immediately improve performance
- Ultra-secure appliances
- Scales linearly – add more appliances

DataPower XI50/52 + XC10: Travel and Transportation

Online Reservations

100x
performance
improvement

Reservations System

- **Before:** 3-5 sec response time
- **After:** .01 -.05 sec response time
- Caching service requests
- Improved the average response time of the Global Distribution System requests for Fare Availability and Category Availability
- 52% caching rate
- 10 minute cache resulted in 40% reduction in load on the back-end systems
- Maintained high data integrity. Faster responses were also accurate
- POC in 3.5 hrs

Improved reliability and scalability of reservation channels

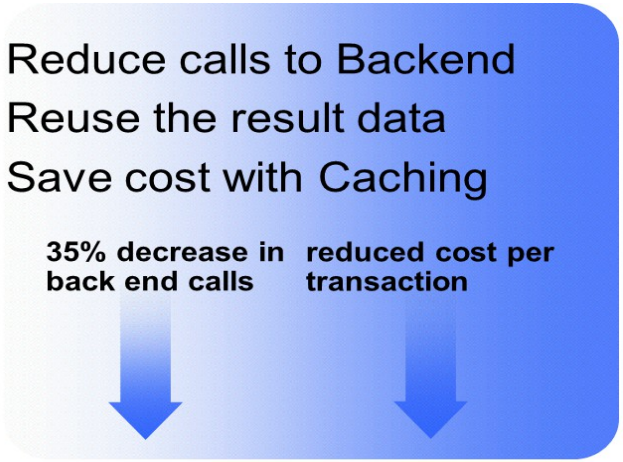
Reduced traffic to backend systems

Deliver high performance & consistent response times

Scale with simplicity and lower TCO

XI50/52 Integration as side cache

Wireless Telecommunication Company



Before:

Customer has multiple systems that call the rebateEligibilityCheck service to get the eligibility for an individual to receive rebates towards the purchase of a new device. Customer relies solely on information from the API call to a third-party service to respond to this request. There are up to 130,000 calls per hour to the third-party service for this particular API. Customer incurs a cost for incremental calls to the third-party service. The data that the third-party service provides is static for each 24 hour period, and there are multiple calls per individual in each 24 hour period.

After:

It reduced the number of back-end API calls by caching the responses for up to 24 hours.

Solution:

Using IBM's caching appliance solution (XC10), A web service call is made to XI50 to obtain up to date eligibility information. If the web service response is already in the side cache (XC10), then Datapower (XI50) returns that result. Otherwise, XI50 will make an API call to a back end system to retrieve the result, which will then be placed in the XC10 for future use and to the requesting service. Upon storing web service responses in the XC10, the XI50 sets the Time-to-Live (TTL) to 24 hours (or 86400 seconds). Also the TTL value could be changed in real-time, without having to start/stop anything, either on the XC10 or XI50.

Client Usage: Social Media Website

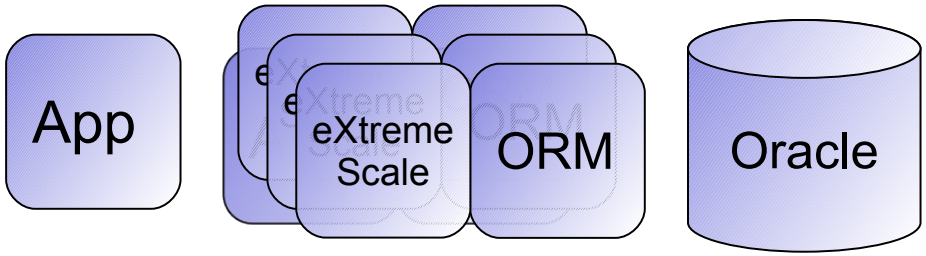
Entertainment

7 Billion
requests per day

10x
reduced response times

Fantasy Sports Web Infrastructure

- **Before:** 60ms response time against database
- **After:** Improved to 6ms response time
- 450k concurrent users
- 80k requests per second up to 1M in 2011
- 6 weeks from concept to production



Performance data is based on measurements and projections using benchmarks in a controlled environment. Actual performance in a user's environment may vary.

Client Usage: Investment Bank

Investment Banking

12 Million
orders per day

4x
increase in revenue

40x
number of transactions supported

Next-generation Order Management System

- **Before:** Oracle RAC based architecture unable to scale to necessary demands
- **After:** 300K transactions / day → 12M / day
- Revenue up 4X and growing … “all because of WebSphere eXtreme Scale”
- Response time drops to 2.5ms
- Moving to “22 x 7” operations (more than 9AM - 4PM)

Performance data is based on measurements and projections using benchmarks in a controlled environment. Actual performance in a user's environment may vary.

Client Usage: One of the largest retail US banks

Retail Banking & Investments

22 Million
online banking users

35x
reduced response times

$500k
reduced costs per month

20x
reduction in “lost sessions”

Next-generation Online Banking

- **Before:** 700ms to login with 2 backend calls
- **After:** 20ms to login with profile cache access
- $6M/year cost savings in MIPs reduction
- 700k logins per hour across 3 data centers
- 8Gb of data transfer per hour between DC’s
- 60 million page views/day
- 10 million logins per day
- 3 active data centers

Performance data is based on measurements and projections using benchmarks in a controlled environment. Actual performance in a user's environment may vary.

Resources

- Fully functional J2SE trial download
  http://www.ibm.com/developerworks/downloads/ws/wsdg/learn.html
- Wiki documentation
  http://www.ibm.com/developerworks/wikis/display/objectgrid/Getti
- User's Guide to WebSphere eXtreme Scale
  http://www.redbooks.ibm.com/abstracts/sg247683.html
- Data Sheet
  ftp://public.dhe.ibm.com/common/ssi/pm/sp/n/wsd14088usen/WS
- XC10 Web Site
  http://www-01.ibm.com/software/webservers/appserv/xc10/
- Getting Started Wiki for XC10
  http://www.ibm.com/developerworks/wikis/display/extremescale/2

Additional resources

Weekly video podcasts covering customers questions and forum posts on the IBM WebSphere eXtreme Scale product.

YouTube

http://www.youtube.com/user/ibmextremescale#

developerWorks.

WebSphere Extreme Transaction Processing for Developers Space will discuss various topics for developing and deploying XTP applications and will point out emerging trends, benefits, challenges, and features associated with it.

http://www.ibm.com/developerworks/spaces/xtp

Questions

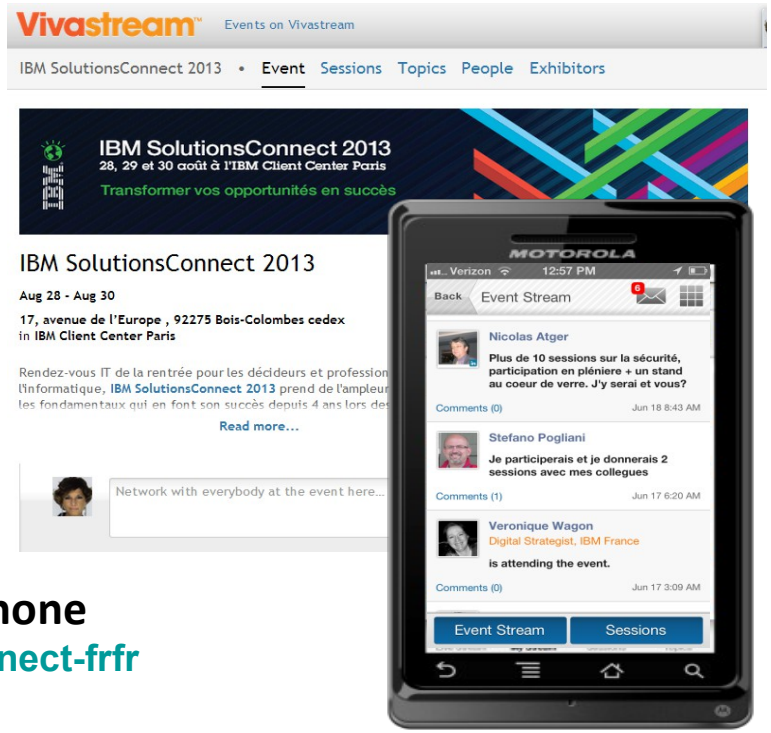
http://www.ibm.com/software/fr/websphere/

IBM SolutionsConnect 2013

L'IBM TechSoftware nouvelle génération

La plate-forme d'échange Vivastream™

- Développez votre réseau
- Découvrez les experts sur les sujets qui vous intéressent
- Echangez avec les speakers et les experts
- Regardez qui participent aux sessions pour lesquelles vous êtes inscrits
- Evaluez les sessions auxquelles participez



Inscrivez-vous sur le web ou avec votre smartphone

http://www.vivastream.com/events/ibmimt-solutionsconnect-frfr

Viva

Available on the App Store

ANDROID APP ON Google play

BlackBerry App World