Application Performance and Scalability with WebSphere z/OS V5

Sabine Holl **sabine_holl@at.ibm.com**
Alex Louwe Kooijmans **nl53347@nl.ibm.com**
Kevin J. Senior **kev_senior@it.ibm.com**

© 2003 IBM Corporation

Agenda

- From the viewpoint of the WebSphere system administrator . . .
- WebSphere z/OS System performance
- How EJB 2.0 improves performance and Enhanced Dynacache
- Performance testing
  - Why
  - How
  - When
- Case History
- References to some application design Best Practice

© 2002 IBM Corporation

WebSphere z/OS system performance

- WAS z/OS performance is very competitive
- WAS z/OS system code is rarely the source of performance problems
- “The system” is typically not a significant part of the path
- WAS z/OS performance has improved significantly release to release
- z/OS etc tuning is very important
- Read InfoCenter section on Performance or the new pdf manual:
  - WebSphere z/OS V5.0.1 Performance, Tuning and Monitoring SA22-7963
- There is value in the z/OS platform integration
- V5.0 Performance measurement has just started...

© 2002 IBM Corporation

Beware the Benchmark Trap!

- Benchmarks do not adequately reflect operational resource requirements
  - Distributed servers run 10% or less on average in production
  - Benchmarks do not consider availability, security, development/test …
  - Operational footprint usually much larger than benchmarks indicate

- zSeries was designed for the real world
  - Raw benchmark performance will be as good or better than other platforms
  - Price/performance component will not reflect the values we have discussed
  - A TCO study is the only way to make an equal comparison

Use caution in making decisions on a single data point: focus on sustained throughput and scale under real-world conditions

© 2002 IBM Corporation

WebSphere for z/OS V5 Performance Summary

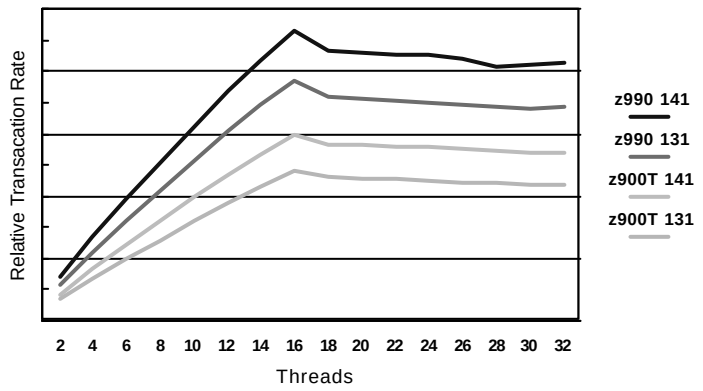
- WebSphere for z/OS performance is slightly improved in V5.0
  - Trade2 performance improved by 10%
  - Cost of managing transactions is much improved
  - MP scalability is excellent
- V5.0 uses less real storage than V4.01.
  - Trade2 can run up to a 4-way in 384m real storage
  - Savings is mainly in the amount of common storage used
- V5.0 performance work is still in-progress

© 2002 IBM Corporation

Java Primitive performance

Multi-threaded Java Performance

z990 vs z900T



Total 85% increase in Java throughput from hardware and software enhancements

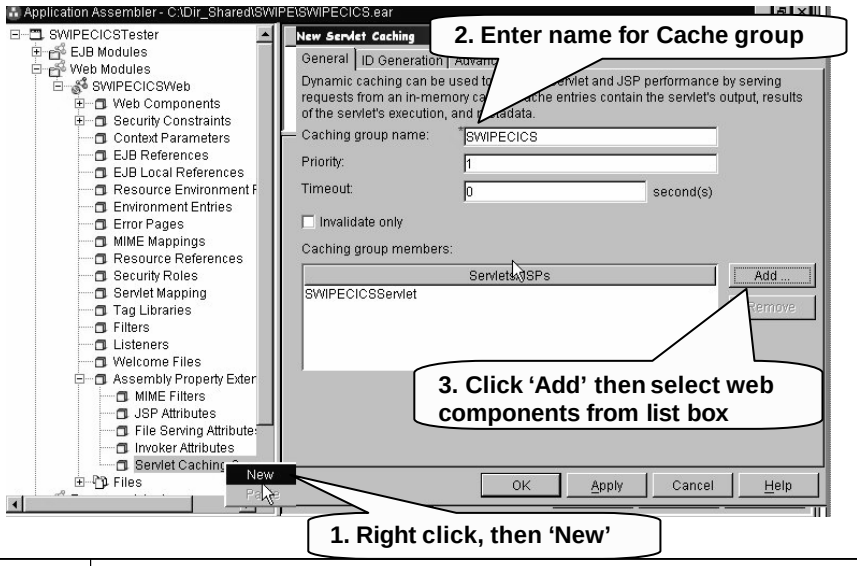
© 2002 IBM Corporation

EJB 2.0 Performance and Dynacache

- EJB 2.0
  - Local Interfaces
  - CMP caching
  - Access Intents
  - Application Profiling (setting CMP cache properties in deployment descriptor)
- WebSphere V5 Dynacache enhancements
  - Edge caching using Akamai ESI
  - Disk overflow of cached objects
  - More flexible cache policy deployment descriptor
- WebServices Dynacache

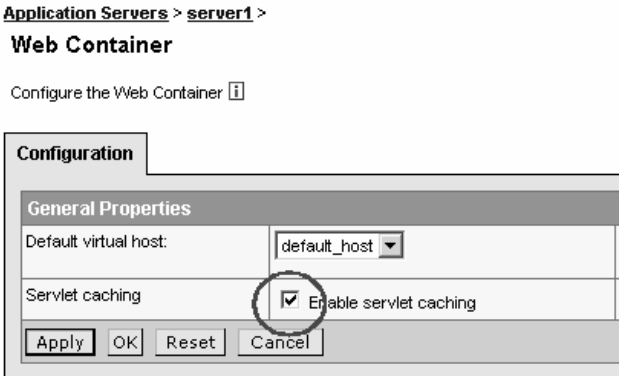
© 2002 IBM Corporation

Using the AAT to set up Dynacache policy



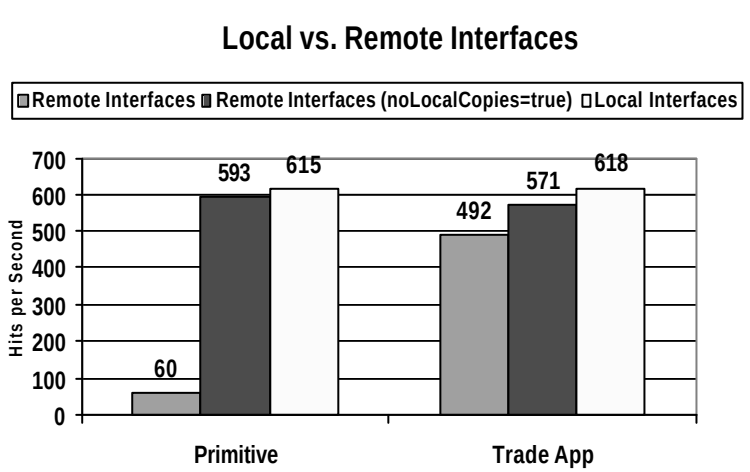
© 2002 IBM Corporation

Enable Dynacache in server



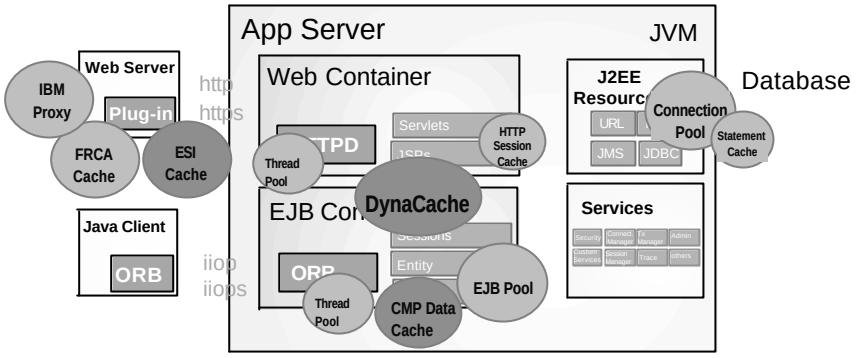
© 2002 IBM Corporation

EJB 2.0 Local vs. Remote Interfaces



© 2002 IBM Corporation

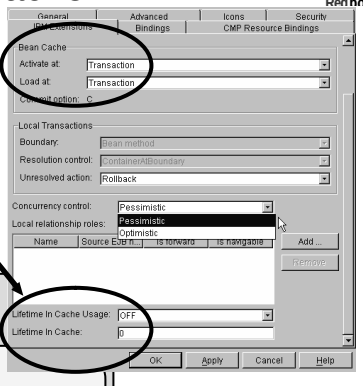
WebSphere Caching Overview

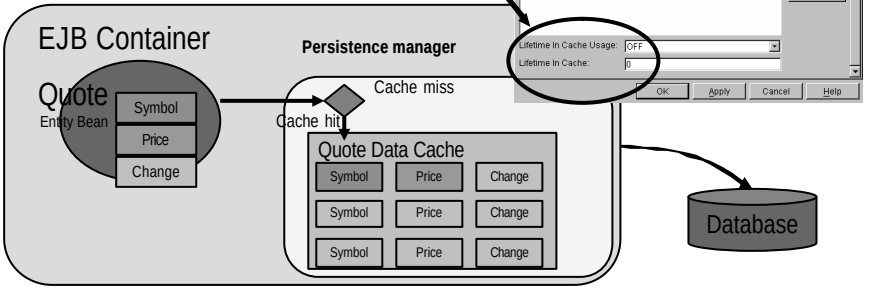


© 2002 IBM Corporation

EJB Data Caching – Lifetime In Cache

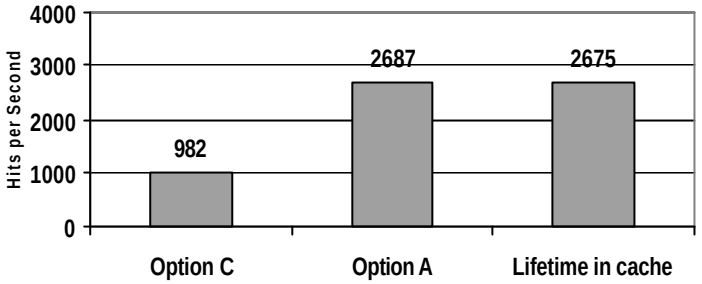
- CMP Entity Beans can be cached
- Time-to-Live in cache is specified in AAT
- Must have one of the following access intents
  - Pessimistic Read
  - Optimistic Read
  - Optimistic Update





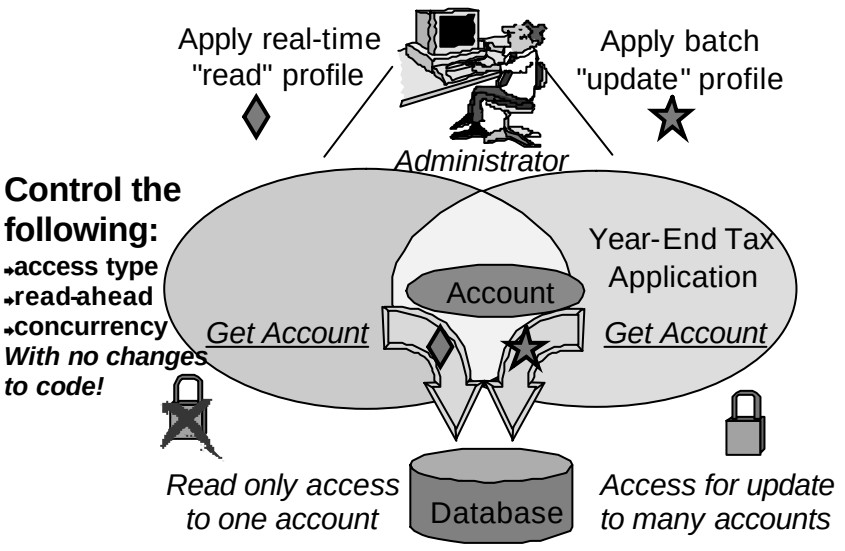
© 2002 IBM Corporation

EJB 2.0 Caching Options



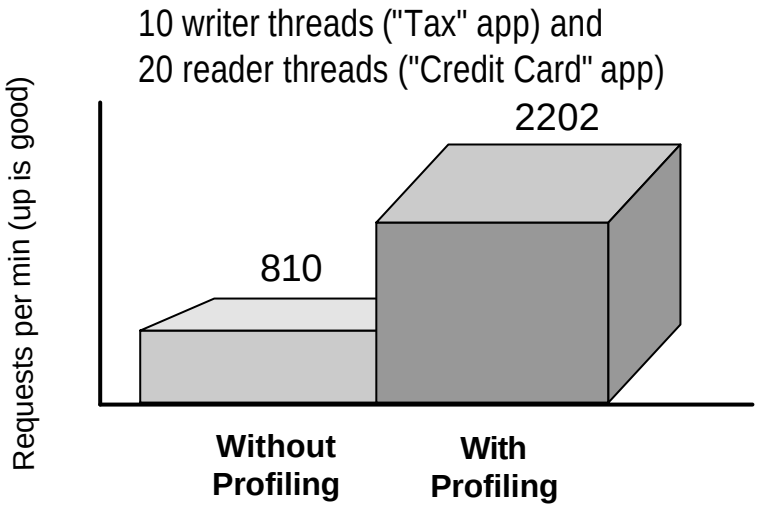
© 2002 IBM Corporation

Application Profiling Optimization



© 2002 IBM Corporation

Some Performance Evidence



© 2002 IBM Corporation

Best Practice: Transactions

- The choice of the transaction level determines the path length and scalability of the application
- Transactions = synchronous processing
- The more complex the UOW gets:
  - The more synchronization has to take place
  - The more dependencies between datasources and objects exist
  - The more resources are consumed and the longer the path length gets
- Transaction level influences the way locking is performed in the specific DB
- So, use transactions only if the application design requires it
  - The default should be "TX_NOT_SUPPORTED" in the deployment descriptor
  - Evaluate the usage of transactions on a case-by-case base

© 2002 IBM Corporation

Best Practice: Transactions (2)

- When using transactions, keep them as short as possible and decouple as soon as possible
  - To avoid long-running locks
- Start transactions on the server, close to the datasource, and not on remote clients
  - Transactions over the network run much, much, longer
- Avoid distributed transactions involving multiple physical systems
  - Overhead to synchronize over the network increases heavily
  - Backing out a transaction becomes very complex
  - Locking can really become a big problem
- Try to decouple units of work, especially in a multi-server environment:
  - Asynchronous communication
  - Messaging, JMS
  - Requires compensation transactions

© 2002 IBM Corporation

Performance Test Objectives

- 6 primitive questions relate to WebSphere Performance
  - **Who**
  - **What**
  - **Where**
  - **When**
  - **Why**
  - **How**
- Case study

© 2002 IBM Corporation

Who Should Be Involved?

- **Everyone**
  - Application Level
    - Architect
    - Developer
    - Experienced user
  - System Level
    - System Administrator
    - DBA
    - WebSphere Administrator
    - Network administrator
  - Managers/Executives

© 2002 IBM Corporation

Why – The Goal

- Increase Customer satisfaction by minimizing response time and maximizing throughput.

© 2002 IBM Corporation

When Do I Think Performance?

- Conflicting Advice
  - Anonymous -

    "Adding good performance to an application after coding is like adding lightness to an airplane after it is built".
  - Java Performance Tuning by Jack Shirazi –

    "At the code writing stage, your emphasis should not be on optimizing...."
- Solution (?)
  - Architecturally think about performance from the beginning
  - Programmatically get the application working then performance tune
  - . . . But leave enough time ! !

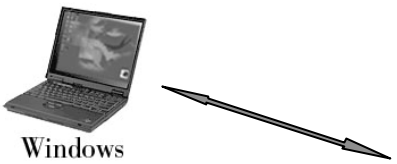
© 2002 IBM Corporation

When – Performance/Stress Testing

- 3 Things to be aware of :
  - Scale of Production Environment in Test
    - Is there enough power in the target environment ?
  - Scale of Production Workload in Test
    - Does the test tool have enough power to drive the workload?
    - Limitations of PC-based load test tools
    - Can they generate enough load?
    - Do you know how to configure and manage a cluster to generate load?
    - Have you bought enough licenses?
  - Data creation / validation
    - De-personalization required?
    - How much test data needed to support 15 minute test run?

© 2002 IBM Corporation

WebSphere Studio Workload Simulator (WSWS)



Windows-based "Controller"

Test Case Capture
Test Case Scripting
Test Execution and Monitoring
Test Analysis

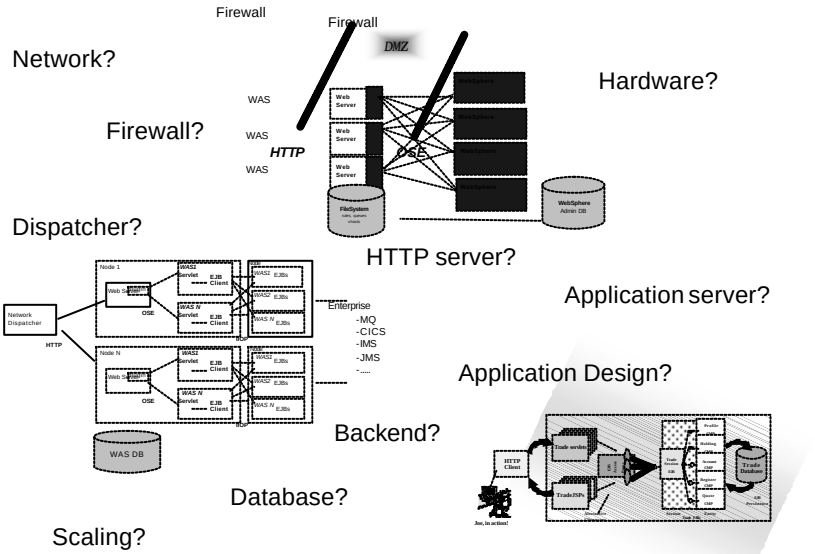
z/Series-based "Engine"

Load generation

All test functions are accessible through the Windows-based controller

© 2002 IBM Corporation

Where Is the Problem?



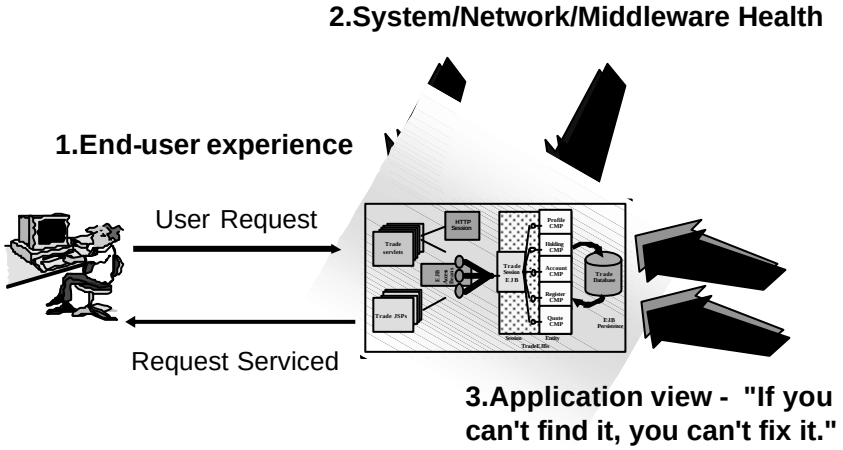
IBM Corporation

Where – Monitoring Principles

- If you don't know you have a performance problem, you can't fix it
- If you don't know where to look, you can't fix it
- If you can't find it, you can't fix it

© 2002 IBM Corporation

Where - Monitoring Dimensions



© 2002 IBM Corporation

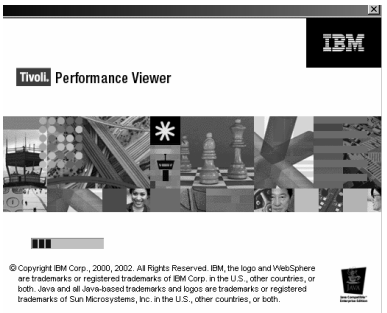
WebSphere z/OS Monitoring

- SMF 120
- **NEW !** Tivoli Performance Viewer
- WebSphere Studio Application Monitor
- JInsight
- . . .other ISV products

© 2002 IBM Corporation

How – Tivoli Performance Viewer

- Part of IBM WebSphere V5.0.1 Client Development Kit for z/OS Download from <WAS_HOME>/bin/setup.exe
- The Performance Viewer is a stand-alone performance monitor for WebSphere Application Server
- The Performance Viewer retrieves data from Application Servers using the EPM/PMI client.
- The Performance Viewer provides default views for all the EPM/PMI performance data



© 2002 IBM Corporation

Planning WebSphere z/OS Performance Testing

- Are you going to test or wait and see?
- What will be the cost of failure of your e-business in front of all your best customers? Does the testing cost exceed this?
- Good decision! So …..How many users? What transaction rate?
- How much time do you think it will take to perform a load test?
  - 1 day?
  - 1 week?
  - 1 month?
  - 1 year?
  - Experience suggests that a simple application requires several man/weeks and a complex application 3-6 man/months.

© 2002 IBM Corporation

Performance Analysis Example

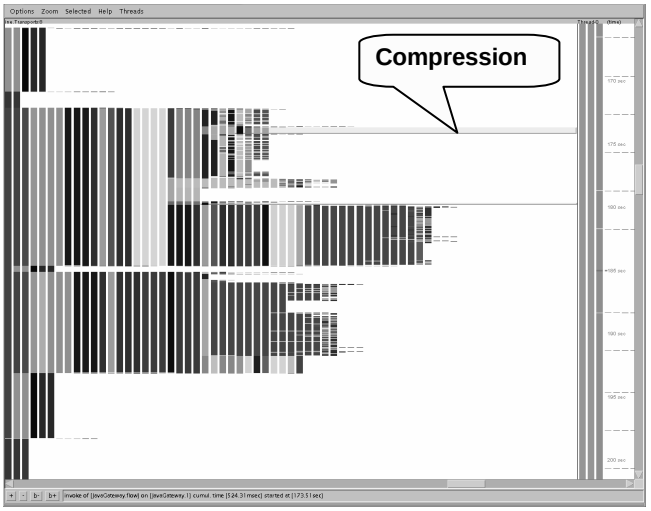
© 2002 IBM Corporation

Large UK Retailer

- Situation:
  - WebSphere application on zLinux (z900)
  - Migrating from 4 large NT servers
  - Mission Critical
  - 1600 Users
  - Poor performance halting migration and roll-out
- Action:
  - Jinsight trace taken within a few hours
  - 3 potential problems / improvements identified
    - Compression / Decompression of session data
    - Not re-using CTG/CICS connections
    - Opportunity to re-write CICS COMMAREA to java types mapping code, originally EAB/CCF

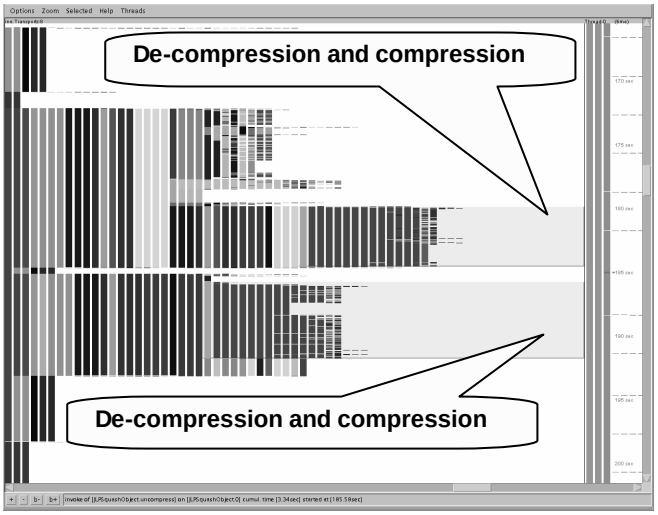
© 2002 IBM Corporation

1. Compression



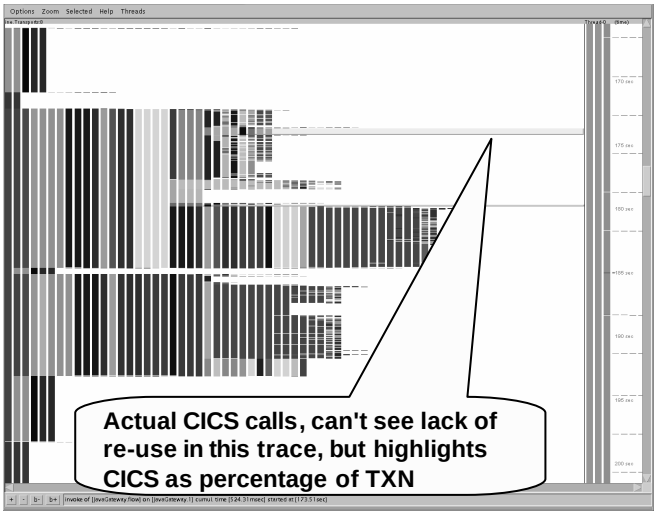
© 2002 IBM Corporation

2. De-compression



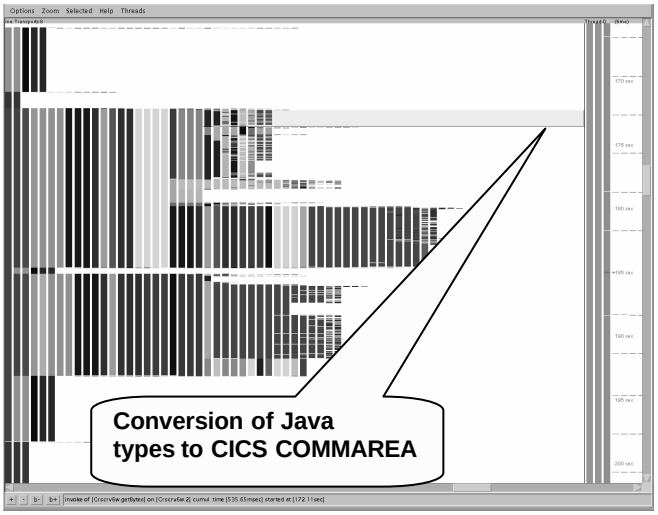
© 2002 IBM Corporation

3. CICS calls



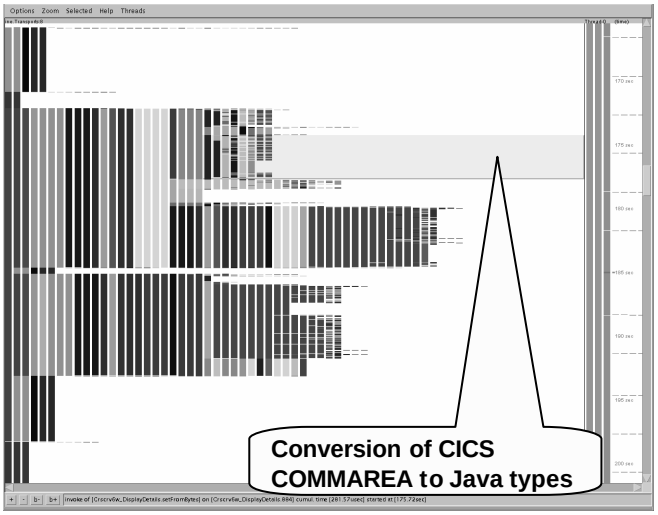
© 2002 IBM Corporation

4. Data mapping: ToBytes



© 2002 IBM Corporation

4. Data mapping: FromBytes



© 2002 IBM Corporation

Main Recommendations

- Remove session data compression
  - Originally an IBM recommendation to solve another problem
  - No longer needed as J2EE application server technology envolved
  - Removed in 2 hours
- Pool CTG Connections
  - Customer code already there, but not being used in this application

Other Recommendations

- Re-write mapping code
  - A large job to get right
  - Eclipse/WSAD tooling being updated to generate hyper-fast mapping code
  - Decided to see what first 2 recommendations provide

© 2002 IBM Corporation

Results

- **Results**
- Removal of compression code reduced CPU utilization but over 40%
- CTG pooling code in test, expected to reduce CPU by another 20%
- Migration and roll-out completed, all 1600 users running on z/Linux
- Peaking at 80 transactions per second
- 200 tx/sec expected during December trading peak
- Customer happy.
  Commented on z/Linux stability compared to previous platform.

© 2002 IBM Corporation

References

- Willy Chiu, Design for Scalability - An Update, IBM, 2001
  http://www7b.boulder.ibm.com/wsdd/library/techarticles/hvws/scalability.htm
- Harvey W. Gunther, WebSphere Application Server Development Best Practices for Performance and Scalability
  http:/ibm.com/software/webservers/appserv/ws_bestpractices.pdf
  "Reducing Garbage Collection Times and Sizing Memory" by Nagendra, Nagarajayya and J. Steven Mayer
  http://wireless.java.sun.com/midp/articles/garbage/
- "Performance Testing Protocol for WebSphere Application Server-based Applications"
  http:/www7b.software.ibm.com/wsdd/techjournal/0211_polozoff/polozoff.html
- Running WebSphere Benchmark Sample Trade3 with Web Performance Tool (WPT)
  http://www7b.software.ibm.com/dmdd/library/techarticle/0303lau/0303lau.html
- Joines, Wilenberg, Hygh, Performance Analysis for Java Web Sites,
  Addison-Wesley, 2002, ISBN 0-201-84454-0.
- WebSphere Developer Domainat: http://www7b.software.ibm.com/wsdd/
- Redbooks:

  SG24-6198-IBM WebSphere V5.0 Performance, Scalability, and High Availability
  (This Redbookis for Distributed WebSphere but contains a section on application design for performance.)

  SG24-6195-00 IBM WebSphere Application Server V5.0 System Management and Configuration

© 2002 IBM Corporation