

# VERIFYING THE INTEGRITY AND RELIABILITY OF DATA-INTENSIVE WEB-SERVICES.

A Case Study

# MICROSOFTTRANSLATOR.COM

en.wikipedia.org/wiki/Portland\_Oregon - Translator - Windows Internet Explorer

http://www.microsofttranslator.com/bv.aspx?from=&to=zh-CHS&a=http%3A%2Fen.wikipedia.org%2Fwiki%2FPortland%2c\_Oregon

Microsoft® | Translator Privacy | Legal  
Translator Help  
Microsoft® is not responsible for the content below  
Powered by Microsoft® Translator

Translate URL: http://en.wikipedia.org/wiki/Portland\_Oregon  
English (Auto-Detected) → Chinese Simplified  
Views

Translated: 100%  
Mouse over text to see original  
Go to original page

新功能 登录/创建帐户

文章 讨论 阅读 编辑 查看历史记录 Search

## 波特兰,

Portland is a city located in the Northwestern United States, near the confluence of the Willamette and Columbia rivers in the state of Oregon. 45° 31'12"N 122° 40'55"W

**硅酸盐**是俄勒冈的位于美国西北, 拉米和哥伦比亚 Oregon 州河的交汇处附近的一个城市。2009年7月的它有一个人口估计的 582,130, [5]使其 30 人口最多的。硅酸盐是俄勒冈州的人口最多的城市和西北太平洋, 在第三次人口最多城市 西雅图、华盛顿和温哥华, 加拿大不列颠哥伦比亚省, 加拿大之后。约有 220 万人居住在硅酸盐城域网 (MSA) 23 人口最多的 2006 年 7 月美国。[6]

硅酸盐在 1851 年注册成立, Multnomah 县 县城。[7] 这座城市延伸轻微到 华盛顿县 西部和 克拉克县 南。它被受一位 市长 和四个其他专员为首的委员会基础的政府。

波特兰市已被称为世界最“环保”或“绿色”城市之一。[8] 强烈的土地利用规划的注意到城市与区域 [9] 和投资在 轻铁, 地铁, 一个独特的地方政府的支持。硅酸盐闻名的 microbreweries 和 microdistilleries, 以及其咖啡热情大号。它也是 跟踪外套 NBA 团队的家。

硅酸盐位于 西海岸海洋 气候 区域标记的温暖、干燥的夏季和雨季但温带冬天。这种气候非常适合于种植玫瑰, 和一个多世纪以来的硅酸盐一直被称为“玫瑰之城”[10][11] 与许多 玫瑰花园 —— 最突出 国际玫瑰园试验。

### 内容 [隐藏]

- 1 历史
  - 1.1 昵称
- 2 地理
  - 2.1 地形
  - 2.2 气候
- 3 城市景观
  - 3.1 公园和花园
- 4 文化与当代生活

## 波特兰

—— 城市 ——

### 波特兰市



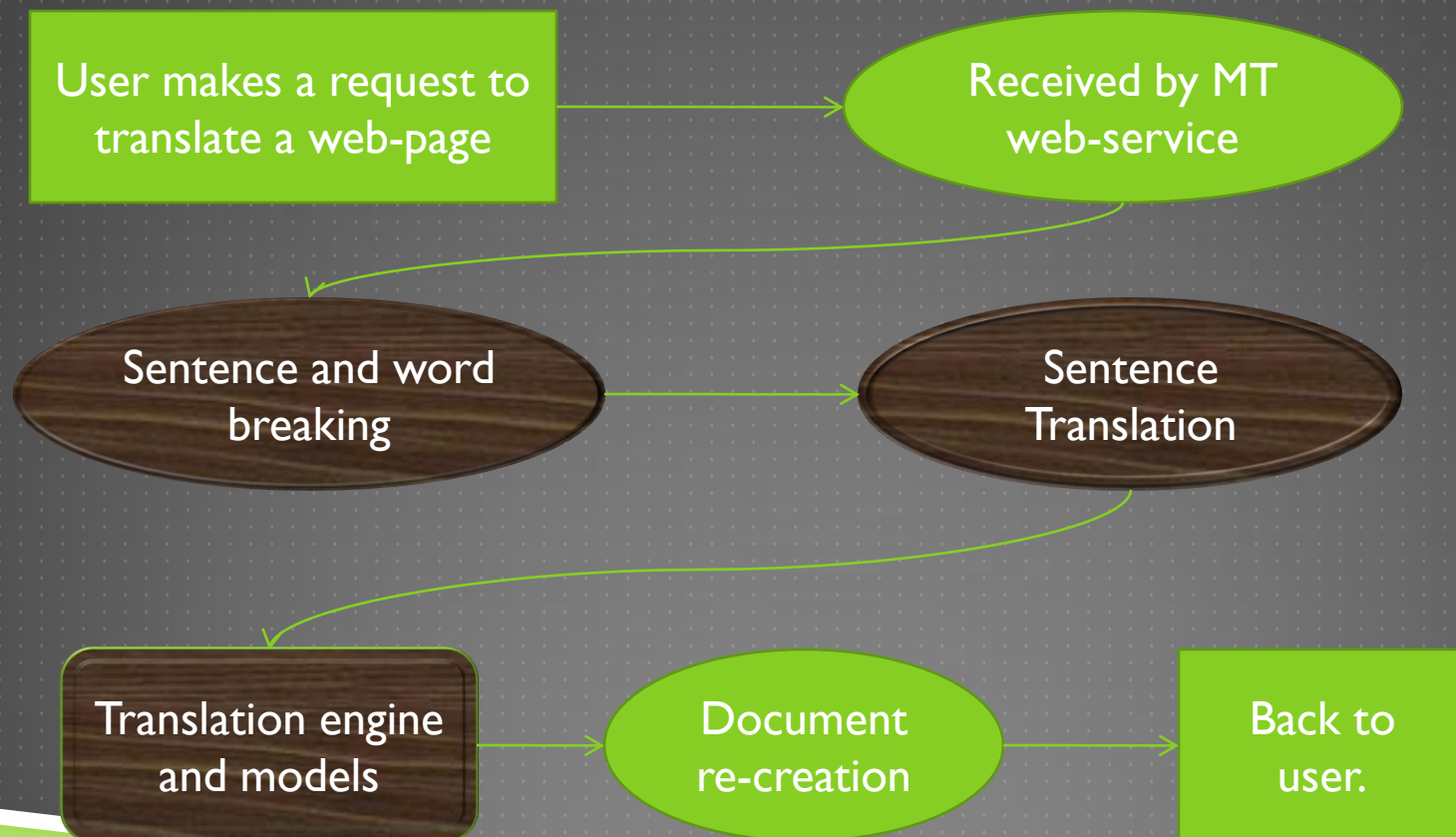
# REAL WORLD USAGES



In Welsh: "I am not in the office at the moment. Send any work to be translated."



# THE MSR MACHINE TRANSLATION (MT) SYSTEM



# COMMON APPROACH TO TESTING WEB-SERVICES

- ▶ Verify individual components first.
- ▶ Verify components in integration.
- ▶ Stress and performance testing.
- ▶ Mean-time-to-failure.
- ▶ Identify maximum sustainable traffic.
- ▶ Capacity planning.

# DATA-INTENSIVE WEB-SERVICES

- ▶ Obvious performance impact.
- ▶ A different set of failure points.
- ▶ Bottlenecks.
- ▶ Data integrity and availability.
- ▶ Growth patterns.
- ▶ Upgrade/versioning.

# THE DATA-INTENSIVE COMPONENTS

- ▶ Translation models hosted by a sub-set of the MT web-services.
- ▶ Data size: multiple GBs.
- ▶ Loaded at run-time, and memory-mapped files.
- ▶ Updated for every scrum cycle.

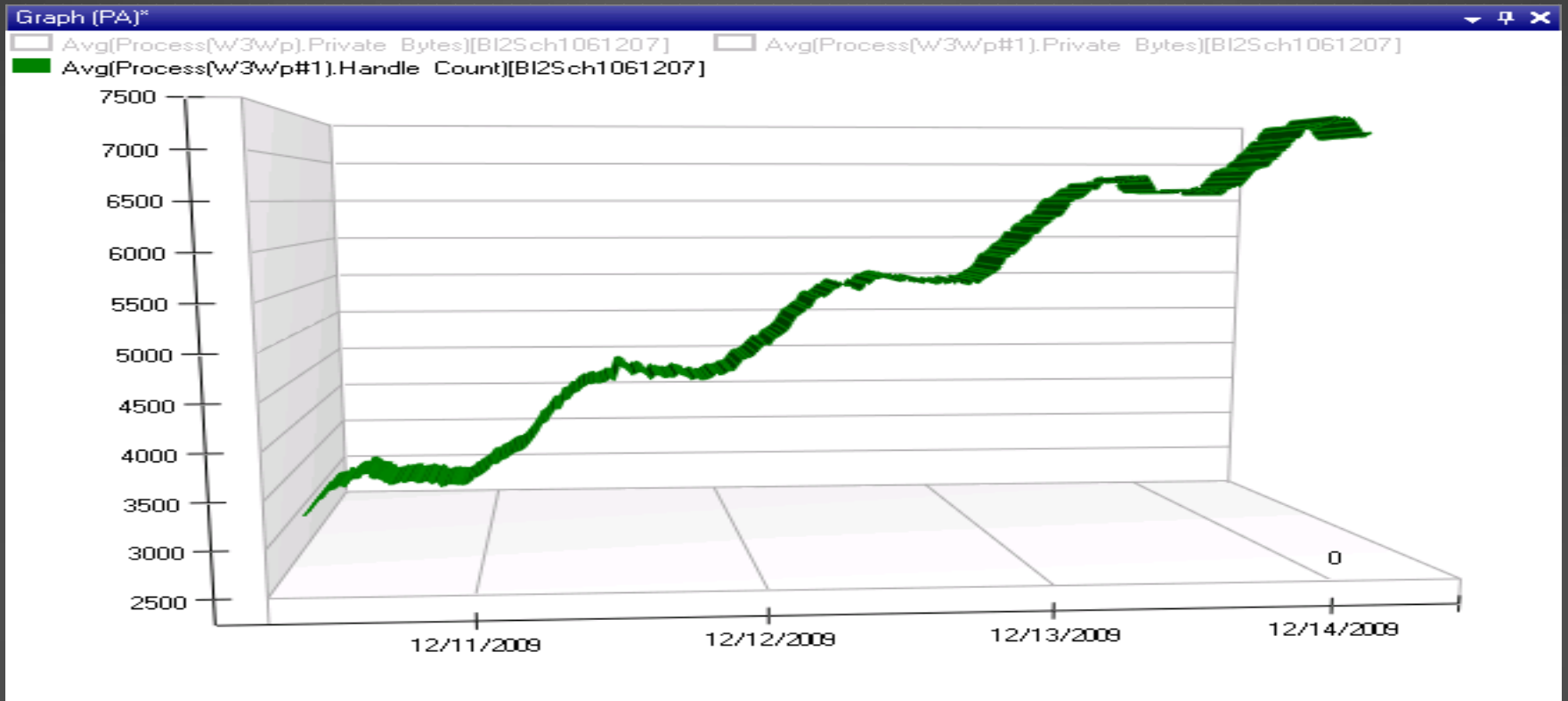


# MEMORY-MAPPED FILES

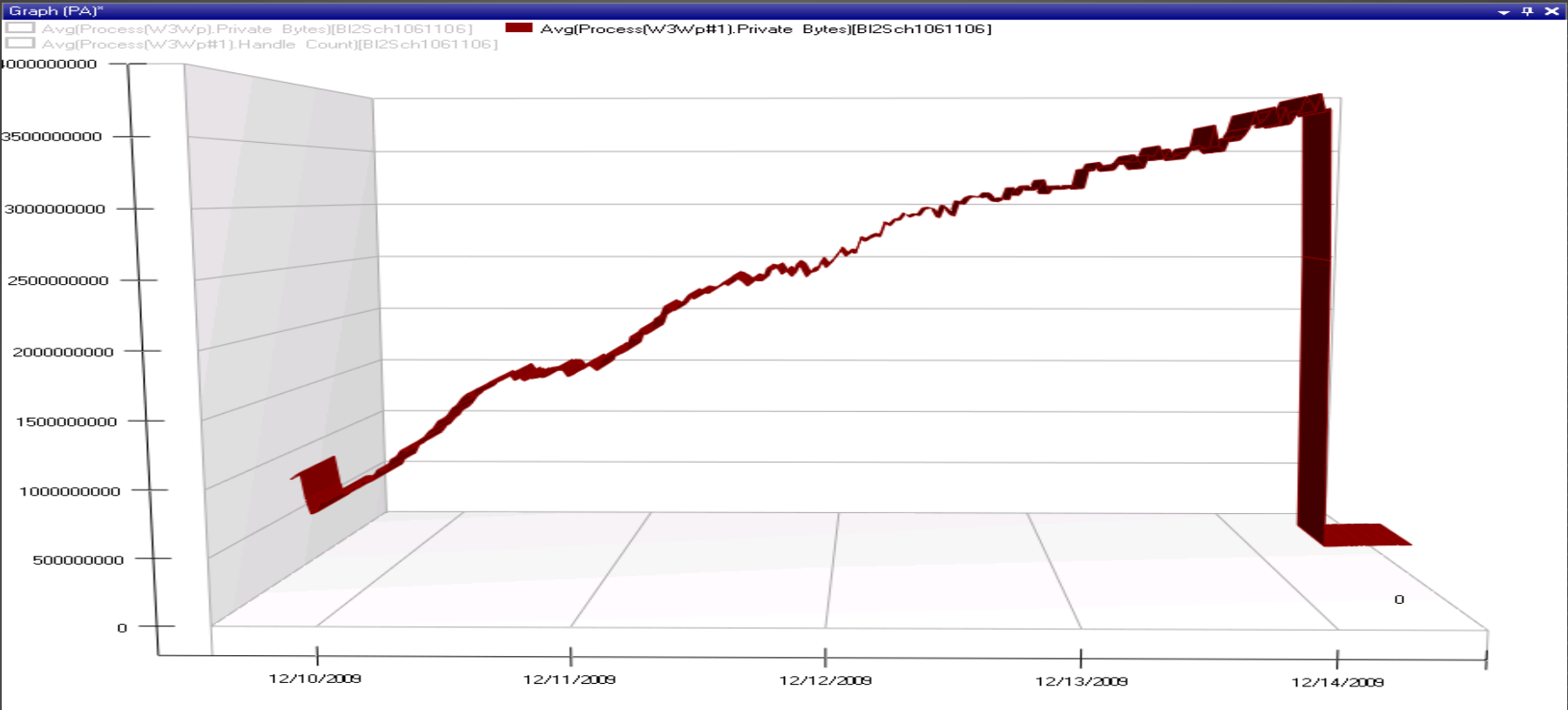
- ▶ A useful solution to increase performance.
- ▶ Works well for most scenarios.
- ▶ More not always better: increased contention/error rate for large data sizes.
- ▶ Run tests to measure optimal size.

	With 2 instances of memory mapped file	With 4 instances of memory mapped file
Throughput, in words per second	5,000	8,000
Percentage of requests failing	1.6%	13.5%

# MEMORY LEAKS

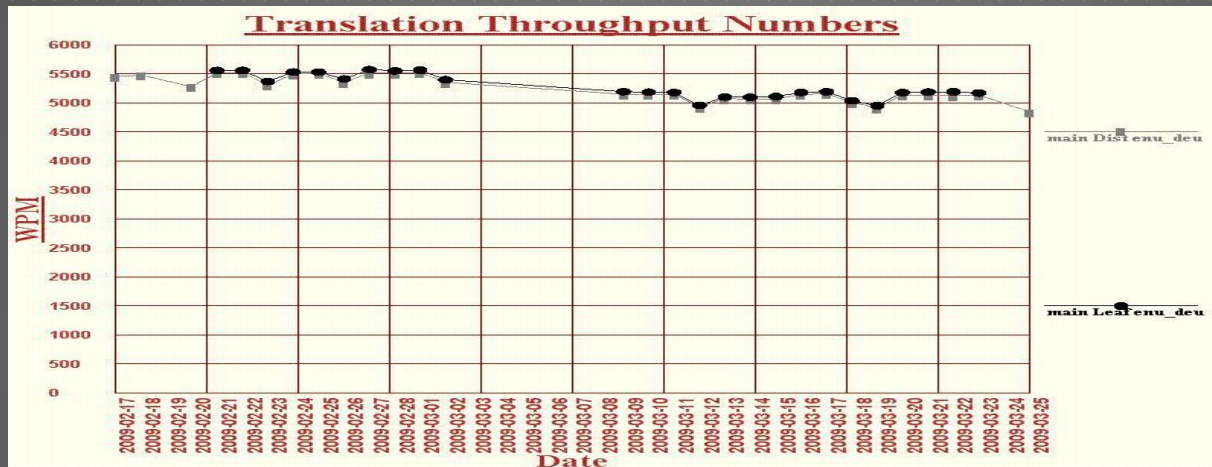


# SIDE-EFFECT OF HANDLE COUNT INCREASE



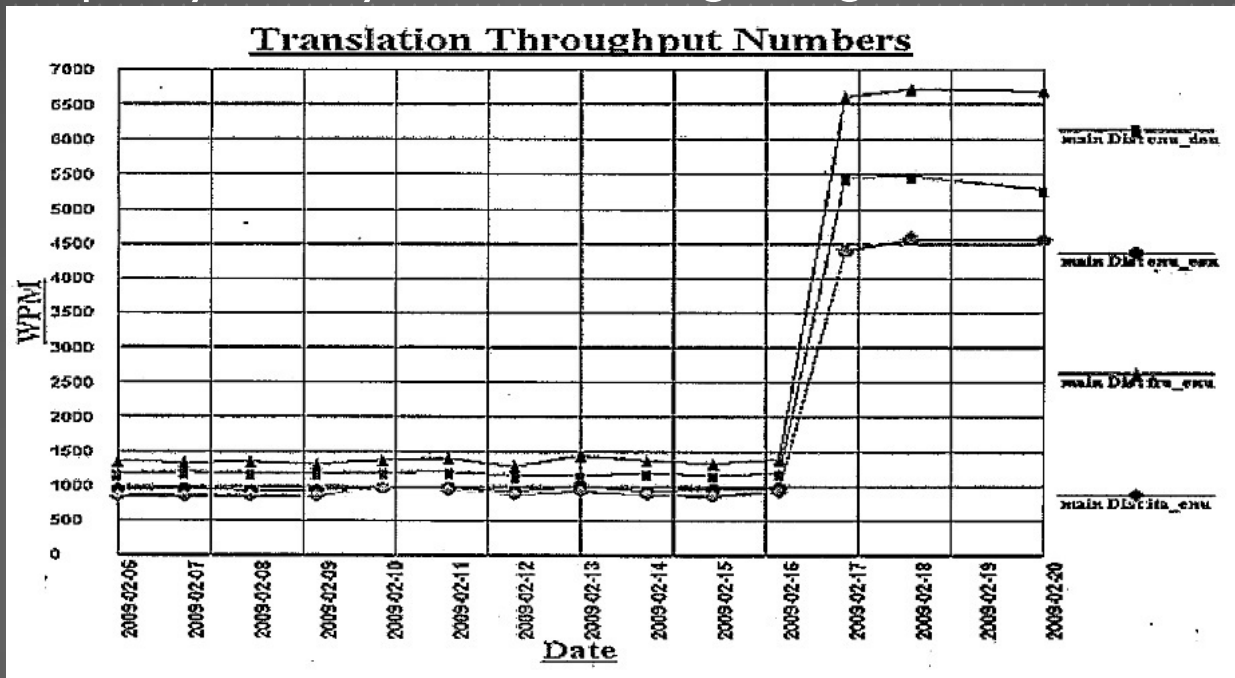
# OVERHEAD OF INDIVIDUAL COMPONENTS

- ▶ In a distributed system, it is important to minimise communication or processing overhead.
- ▶ Identify primary contributor to a bottleneck.
- ▶ Tests run regularly measuring throughput at different layers.



# STAYING WITH IT.

- ▶ Run perf tests regularly.
- ▶ More quickly identify effect of rolling changes.

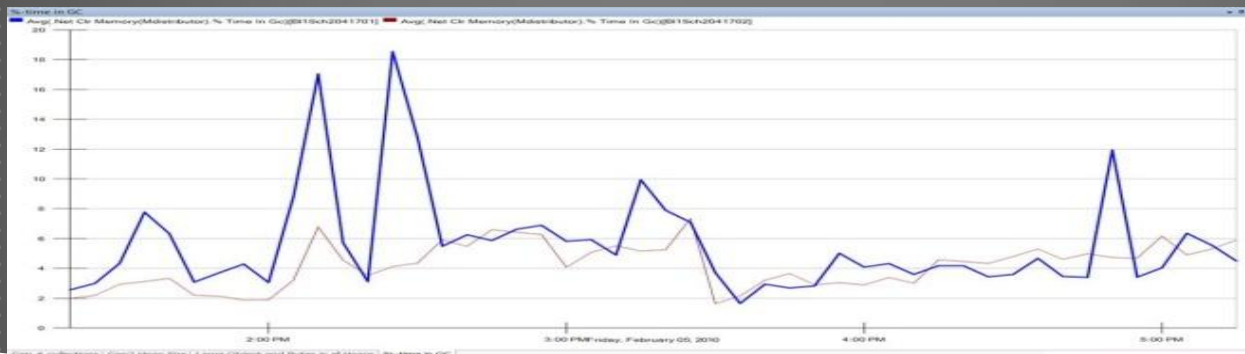
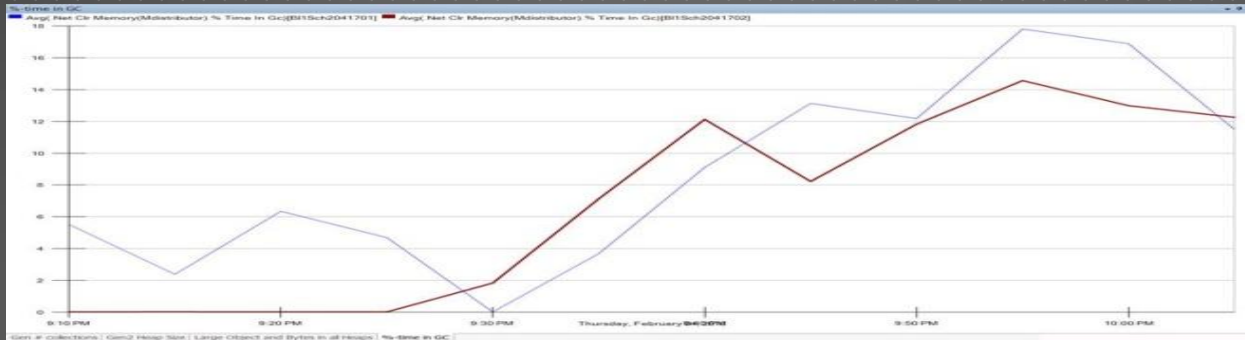


# IMPACT OF GARBAGE COLLECTION

- ▶ For data-intensive applications/services, any overhead is performance hit.
- ▶ Side-effect of managed cache: time in GC.
- ▶ Useful to measure: percentage time in GC, large object heap size,

# FLAVOURS OF GARBAGE COLLECTION

- ▶ Difference between Workstation and Server-GC:



# CONCLUSIONS

- ▶ Squeeze throughput to the utmost at each layer.
- ▶ Stay on top of version/build changes.
- ▶ **Keep It Simple.**