

Platform Data Affinity

Smart, data-aware scheduling for
Platform Symphony



Features

- Tag data with user-defined resource attributes
- Define expressions representing resource preferences
- Account for relative costs of data transfer
- Express data preferences at the application, session or task level
- Flexible client and server-side programming interfaces

Benefits

- Reduce task run-times by up to 80%
- Reduce network congestion
- Lower latency for time-critical applications
- Reduce cost by consolidating servers, storage and networks
- Integrate easily with popular data caching solutions

Overview

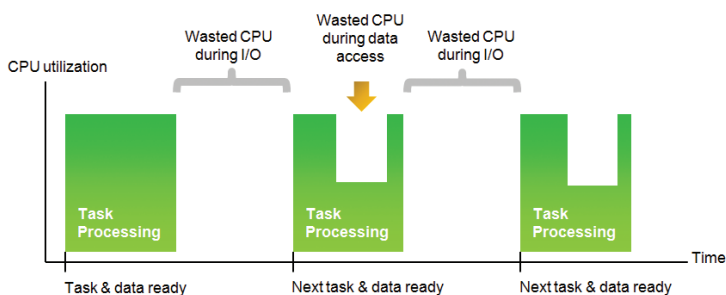
Platform Data Affinity takes data locality into account when scheduling service-oriented workloads on Platform Symphony grids often eliminating the need to load large (gigabyte or terabyte sized) datasets. Smarter data handling results in dramatic gains in application performance for data-intensive applications and more efficient use of shared compute, network and storage infrastructure.

By enabling a preferential association between a task and a service instance or host that has needed application data, overhead and latency associated with data movement are reduced. Users or application developers can define their own “tags” to label data in advance or at run-time depending on the application. Hosts having needed data are selected based on flexible user-defined expressions. By scheduling workloads intelligently based on data requirements, network traffic can be reduced dramatically, and application performance can be boosted by a factor of four or more.

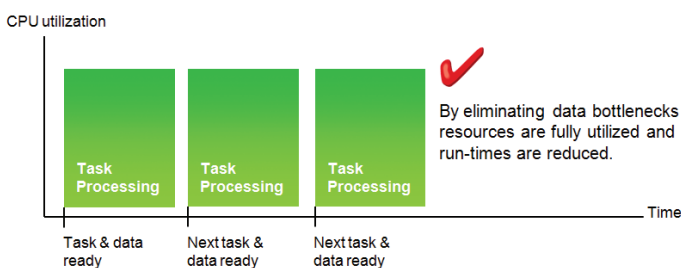
Flexible, location-aware data services

As processors continue to get faster, and compute farms scale to hundreds or even thousands of nodes, it becomes impossible to keep up with throughput requirements by simply scaling up network-attached storage (NAS) environments. Inevitably, applications face long-delays accessing needed data over congested networks. Platform Data Affinity solves the latency bottlenecks caused by moving data sets between hosts by taking into account data locality and placing workloads in an optimal and cost efficient way. Smarter task placement results in reduced network traffic between nodes, faster file access times, and the opportunity to consolidate previously separate layers of server, network and storage infrastructure to reduce capital and operational cost.

Without Platform Data Affinity



With Platform Data Affinity



Integrate with data caching solutions

While Platform Data Affinity does not require data caching or a parallel file system solutions to provide benefits, it is integrated with leading distributed cache systems including VMware's GemFire. Developers can also use a provided data-aware scheduling plug-in to optionally integrate their own in-house developed applications that maintain an awareness of data locations. This means that Platform Data Affinity can easily tie into existing optimization initiatives. As you distribute and cache data across the nodes in a grid, Platform Data Affinity intelligently places workloads to ensure optimal performance with minimal data-staging wait-times.

Product Capabilities

Easily tag application data

Platform Data Affinity provides an easy to use mechanism to label resource attributes such as large data sets available on specific service instances or host computers. These data attributes can be provided by the service, at the host-level, by a data-aware scheduling plug-in. This flexible mechanism for naming and classifying data provides application developers and administrators with flexibility including the ability to name and classify datasets in a way that suits their unique applications.

Expressing data preferences

Client applications can express preferences for one or more datasets in flexible ways. Expressions can include basic numeric operations and can take into account weighting factors and non-data related resource preferences such as memory to make the most efficient scheduling decision possible. Platform Data Affinity factors the cost of moving data with numeric values associated with each data resource. These values can be adjusted depending on factors like dataset size and network speed to help decide where workloads should be placed based on how cost-efficient it will be to move data to where it is not already available.

Data-aware scheduling plug-in

The data-aware scheduling plug-in for Platform Symphony provides another way to obtain the location and relative cost of obtaining data. Using this optional plug-in, Platform

Symphony can obtain resource data from a centralized source that can be customized to integrate with third party products or in-house applications.

Application or task level optimization

When making scheduling decisions, Platform Data Affinity takes into account multiple sources of information including any attribute data associated with Symphony services, host-level attribute data, and results obtained through an optional data-aware scheduling plug-in. If none of these values are available, calculations can still proceed with default value for a session. This flexibility enables Platform Symphony to make informed scheduling decisions even in environments where knowledge about data location and relative transfer costs are incomplete.

Client & server-side programming interfaces

Client and Service side programming interfaces enable data related resource requirements to be expressed at run-time. This is valuable because for many applications, details like data requirements and the relative sizes of datasets may not be known in advance. Using the client-side API, developers can specify attribute definitions and create their own resource preference expressions. They can then bind these resource preferences to a task, ensuring that the task will be dispatched preferentially to service instances having the best access to needed data.

Platform Data Affinity Supported Environments Operating Systems

- Windows® 2003, 2003R2 64-bit, 2008, 2008R2 64-bit, Vista®
- Windows 7, Windows HPC Server® 2008
- Red Hat Enterprise Linux® (RHEL) 4, 5, 6
- SUSE Linux Enterprise Server® 9,10
- IBM AIX® 5
- Oracle® Solaris® 64
- Other Linux distributions 2.6 or greater with glibc 2.3, 2.4

Development Environments

- C, C++, C# on Microsoft .NET, COM, VB6/VBA, Java 1.4.2,1.5,1.6, Sun JVM, BEA JRockit, VisualAge C++ (AIX), Intel C++ and various GNU compilers.
- Microsoft Visual Studio Professional 2008/2010, ECLIPSE 3.2.2,3.3.0

Platform Computing is the leader in cluster, grid and cloud management software - serving more than 2,000 of the world's most demanding organizations since 1992. Our workload and resource management solutions deliver IT responsiveness and lower costs for enterprise and HPC applications. Platform has strategic relationships with Cray, Dell™, HP, IBM®, Intel®, Microsoft®, Red Hat®, Fujitsu and SAS®. Visit www.platform.com.

World Headquarters

Platform Computing Corporation
3760 14th Avenue
Markham, Ontario
Canada L3R 3T7
Tel: +1 905 948 8448
Fax: +1 905 948 9975
Toll-free Tel: 1 877 528 3676
info@platform.com

Sales - Headquarters

Toll-free Tel: 1 877 710 4477
Tel: +1 905 948 8448

North America

New York: +1 212 888 6270
San Jose: +1 408 392 4900

Europe

Bramley: +44 (0) 1256 883756
London: +44 (0) 20 3206 1470
Paris: +33 (0) 1 41 10 09 20
Düsseldorf: +49 2102 61039 0

Asia-Pacific

Beijing: +86 10 82276000
Xi'an: +86 029 87607400
Tokyo: +81(0)3 6302 2901
Singapore: +65 6307 6590

**Platform
Computing**