



WebSphere software

IBM WebSphere Extended Deployment: Providing enhanced infrastructure capabilities for SOA environments.

Contents
2 Introduction
3 Infrastructure virtualization: Fundamental to WebSphere Extended Deployment
4 WebSphere Extended Deployment: Advanced capabilities for SOA
5 Extending the value proposition: The IBM WebSphere business grid
7 Understanding the technical details of the WebSphere Extended Deployment business grid
11 High-performance and high- availability business-grid applications
13 Bringing it all together: Policy- based workload management
14 Conclusion
15 For more information

Introduction

Scale-out and distributed computing continues to gain prominence across the IT infrastructure landscape. Architectural approaches built on this kind of foundation offer inherent advantages, including the potential for price-performance economic benefits and greater flexibility. However, management, availability and throughput concerns can quickly outweigh these associated benefits. As a result, a balanced approach across infrastructure resources is often required. For most organizations, a mixture of both scale-up and scale-out computing resources is the preferred architectural approach and is often dictated by the types of workloads that need to be supported.

IBM WebSphere® Extended Deployment is an infrastructure product that operates equally well across distributed (scale-out) and centralized (scale-up) resources, and offers a consistent WebSphere application experience and approach across a number of heterogeneous resource types.

Java™ 2 Platform, Enterprise Edition (J2EE) standards have provided the framework for an application infrastructure that simplifies application development so that it is possible to deploy applications that take advantage of system services without binding the applications directly to the underlying service implementations. Now, application and infrastructure logic can be kept separate, enabling application developers to focus on building applications rather than building infrastructure code. IBM WebSphere Application Server Network Deployment provides the infrastructure for deploying applications while WebSphere Extended Deployment adds advanced capabilities such as resource virtualization, high-performance computing and intelligent, policy-based workload management.

Through these advanced capabilities, WebSphere Extended Deployment can provide:

- *Increased throughput, reduced response times and linear scaling for existing applications*
- *Improved responsiveness in a service oriented architecture (SOA) environment*
- *The ability to support new, innovative applications and workload types*

Infrastructure virtualization: Fundamental to WebSphere Extended Deployment

In October 2004, WebSphere Extended Deployment introduced the concept of infrastructure virtualization to the WebSphere platform. Why is this important? First, by separating the underlying infrastructure from the applications that run on it, workloads can be dynamically placed and migrated across a set, or pool, of resources. There's no longer a tight binding or one-to-one relationship between an individual machine (or set of machines) and an application. This loose coupling enables WebSphere Extended Deployment to intelligently manage and shift workloads around according to business policy. High-priority applications can be allocated the majority of resources; lower-priority applications are either designated to run later or moved to less-capable resources. These operations are all seamless to the user, but they require sophisticated job scheduling and workload-management capabilities like those built into WebSphere Extended Deployment.

Further, this notion of infrastructure virtualization enables the inclusion of a wide range of platforms, including larger machines such as mainframes, symmetric multiprocessor systems and distributed resources within the resource pool. This virtualization is commonly referred to as a *grid*. Indeed, the underlying principle of grid computing is the virtualization of both workload and information resources. These resources are typically homogeneous (blade servers or clusters), but they can be and often are heterogeneous.

Support for heterogeneous resources might not seem like an important point, but as virtualized infrastructures continue to gain traction, the ability to include a wide range of resource types within a grid (or pool) of resources and have your infrastructure software use and manage them in a seamless fashion, provides flexibility and choice. Using WebSphere Extended Deployment, application developers and system administrators can view your infrastructure resources as a single, consistent entity. This infrastructure virtualization also provides a foundation for your organization to help increase infrastructure value through the advanced capabilities integrated in WebSphere Extended Deployment.

WebSphere Extended Deployment: Advanced capabilities for SOA

SOA is a style or framework that lets you build, deploy and integrate services for IT resources, applications and business-process flows. It facilitates integration, enables you to modularize applications and provides a coherent view of a business process as a set of coordinated services. You can reuse these services to more rapidly build new applications, as well as integrate applications and business processes across your enterprise and with partners, suppliers and customers more seamlessly.

The momentum behind SOA has been significant across the industry, but to use this approach effectively, you need to have the corresponding flexibility in your underlying infrastructure. WebSphere Extended Deployment helps create an On Demand Business infrastructure that spans both transactional and long-running workloads. This integrated environment dynamically determines how to optimally allocate application-infrastructure resources based on customer-defined business goals, providing the necessary flexibility and helping you realize the full benefits of SOA.

Extending the value proposition: The IBM WebSphere business grid

Although the core virtualization foundation enables the sophisticated workload and autonomic management features of WebSphere Extended Deployment, it also provides a platform for infrastructure optimization. The central concept here is one of helping to drive up infrastructure value and lower total cost of ownership (TCO) by increasing the resource usage across the WebSphere Extended Deployment virtualized pool of resources. Simply put, to drive up infrastructure usage, you either consolidate workload on fewer resources or you increase the amount of work your existing infrastructure is doing.

WebSphere Extended Deployment is a leader at providing optimized control and implementation for J2EE online transaction processing (OLTP) applications across a transactional grid of resources. By combining this core capability with the virtualization foundation and intelligent policy-based workload management, you can effectively consolidate OLTP workloads on fewer infrastructure resources. This consolidation has two important benefits. First, it can help lower TCO. Second, it can free your infrastructure to support new types of applications.

But WebSphere Extended Deployment provides more. With Version 6.0, WebSphere Extended Deployment supports new types of workloads, such as Java batch and computationally intensive workloads. This capability turns WebSphere Extended Deployment from a transactional grid into a business grid. Batch applications and workloads, such as records processing, are commonplace in many back-office computing environments. Further, as the overall cost of computing resources continues to decrease, many organizations are increasingly deploying and realizing business value from high-performance and computationally intensive applications. Examples of these types of applications include portfolio optimization and risk analysis in financial services, drug discovery in healthcare, product design and simulation in industrial and manufacturing firms, and generally any kind of data mining and analysis across multiple industries.

The ability to support new types of workloads is an important trend in WebSphere Extended Deployment because it enables you to consolidate workloads across a single, virtualized pool of infrastructure resources and to help realize increased infrastructure value. Typically, OLTP, computationally intensive and batch workloads have been run across separate, dedicated infrastructures. These workloads can now be combined into a single virtualized grid, taking advantage of the autonomic capabilities of WebSphere Extended Deployment to schedule work on idle computer resources (see Figure 1). Combine these capabilities with WebSphere Extended Deployment support for high-performance computing, strong capabilities and autonomies, and all of the core features of the WebSphere platform, and you have a powerful infrastructure value proposition.

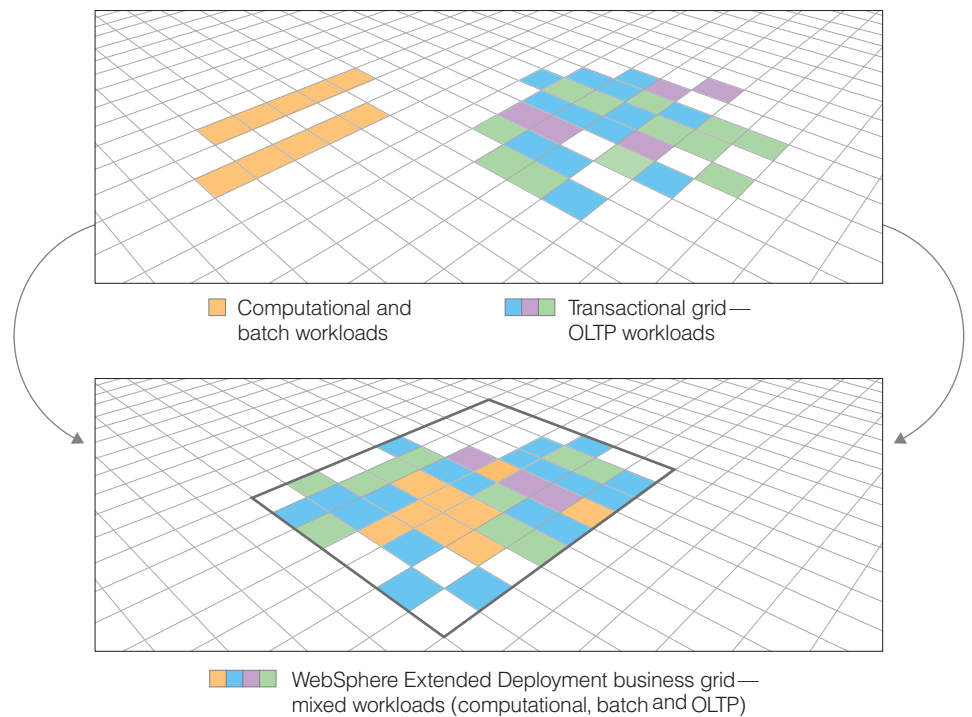


Figure 1. WebSphere Extended Deployment business grid

However, to realize the business-grid benefits of WebSphere Extended Deployment, it's important to identify application candidates across organizations within a company. Infrastructure virtualization and horizontal integration require a common infrastructure which can host multiple line-of-business (LOB) applications. Business units and departments have to share infrastructure resources, which is the whole point of infrastructure virtualization and a prime motivation behind SOA. Although this might seem like a challenging proposition, the benefits can be significant. Companies that have deployed cross-company virtualized infrastructures like WebSphere Extended Deployment have realized benefits in economies of scale and overall lower operational costs. And with WebSphere Extended Deployment policy-based workload management, it's easy to set priorities across multiple applications so that each business unit realizes its fair share of infrastructure resources. This capability helps mitigate reluctance from LOB organizations to share and use a common infrastructure.

Understanding the technical details of the WebSphere Extended Deployment business grid

This section covers some of the technical aspects of the WebSphere Extended Deployment business-grid capability.

Building business-grid applications

Business-grid applications are essentially just like other WebSphere applications. They have the same access to WebSphere resources, are packaged the same way as other WebSphere applications, and are deployed just as any other type of application would be deployed. Business-grid applications also benefit from the rich qualities of service inherent in the WebSphere Extended Deployment operating environment. Two types of business-grid application patterns are supported: Java batch applications and computationally intensive applications.

- *Batch applications*

Use the batch-programming model when you need to perform record processing with restart capabilities. For instance, when generating customer statements, a batch application might read records of the customer database and process each record to create a statement for each customer. If the batch application ends prematurely, you want the job to restart where it left off. The batch-processing container provides mechanisms to commit the transactions at an interval specified by policy. Batch applications take advantage of this capability by extending and implementing a set of classes and interfaces related to the batch-programming model.

The batch-programming model is a synthesis of mainframe batch applications within Enterprise JavaBeans (EJB) programming. It is geared toward record processing. Each step of the batch denotes input and output streams and can be queried by the container to obtain a logical cursor from the application. The cursor is opaque to the container and can be sent back to the batch application on restart, enabling the application to reposition itself for continuation. The application and the batch container cooperate to help ensure that correct restart information is captured at regular intervals. A global transaction is used to track the checkpoint information to help ensure that it's consistent with the application data.

You can submit batch applications using a command-line interface (CLI) or using a programmatic interface from an application.

- *Computationally intensive applications*

Computationally intensive applications are those types of applications that are typically run in parallel or high-performance computing environments. They are built by extending and implementing a set of classes and interfaces related to the grid application-programming model. These types of jobs essentially run as methods of a stateless session bean that extend a business grid interface. The programming model is lightweight, with a small number of additional interfaces to implement related to job life-cycle management.

You can also submit computational jobs from the CLI or by using a programmatic interface from an application.

Managing business-grid applications

Both batch and computationally intensive jobs obtain primary implementation parameters through XML job control language (JCL) (xJCL), and a set of properties is delivered to each job as specified in the xJCL. The intent of this operation is to allow specification of a small amount of job- and step-specific data through xJCL. If the input data set is large, then a staging approach should be used and a reference to the staged data passed as a property parameter to the job. The reference to the data might be in the form of a URL, database key or queue name.

Job-status information can be queried through the management interfaces. If the job needs to store large amounts of output data for access by the submitter, then the submitter should supply a reference as a parameter indicating where the output should be sent.

You can submit, query and cancel business-grid jobs. A job submission returns a job ID that you can use to manage the job. Job management is possible through programmatic interfaces or through the administrative console (see Figure 2).

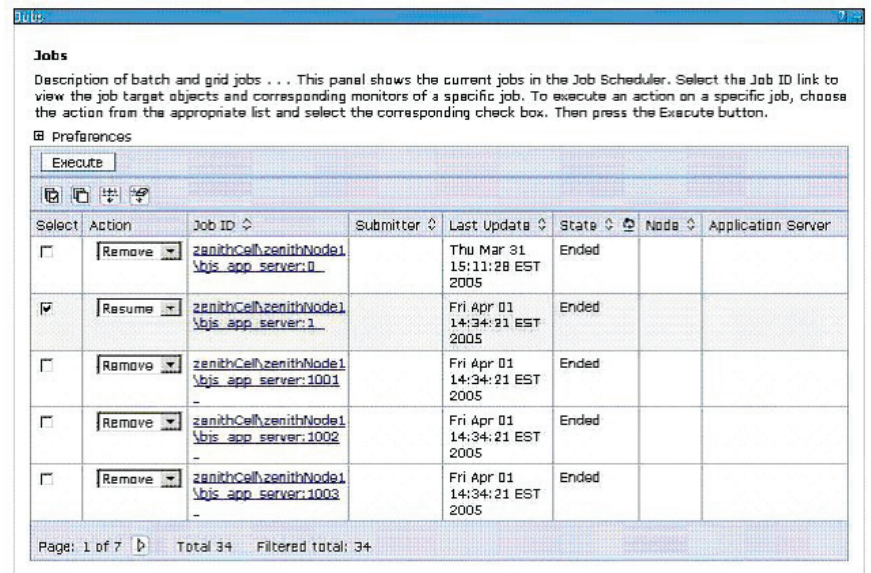


Figure 2. Job-management interface

Job management is a highly secure mechanism. Only system operators with the correct privileges and the job owners themselves can manage jobs. Administrative-console panels provide information on which jobs are currently running, which are queued to run and which jobs have been completed. Also, visualization panels graphically depict overall job statistics, throughputs and resource usage, such as processor usage and memory consumption.

Scheduling business-grid applications

The WebSphere Extended Deployment business grid uses a policy-based job scheduler. Just as WebSphere Extended Deployment uses service policy to determine how to queue transactional work and how many instances of which applications to keep active, service policy is also used for job scheduling.

Each business-grid application can be assigned a service policy that describes its overall importance. The job scheduler uses the application's importance to determine when and where it should be dispatched relative to other work active in the system. It also works closely with WebSphere Extended Deployment application placement to help ensure that adequate job instances exist when and where needed.

The job scheduler is highly available. If the node or process hosting the scheduler fails, the scheduler is recovered in another process and possibly on a different node.

High-performance and high-availability business-grid applications

Although the architecture is designed from the ground up to support throughput and performance, WebSphere Extended Deployment offers built-in features that enable high-performance computing and high-availability capabilities to be integrated within the infrastructure virtualization foundation. These capabilities are available to any WebSphere Extended Deployment application, and therefore offer more value when you use WebSphere Extended Deployment to support computationally intensive and batch applications. And because they are built using Java and paired with exceptional tools, you have the potential to realize new levels of programmer productivity for these types of applications.

WebSphere Extended Deployment ObjectGrid feature

The WebSphere Extended Deployment ObjectGrid feature provides a customizable, pluggable object-caching framework that enables applications to share object data using a variety of consistency models. Object data can be application data that can be retrieved from a common data source such as a database or file system, or short-lived objects such as session data (for example, shopping-cart information). Objects can be stored in the grid and then accessed from multiple applications, helping to reduce the number of trips to the data source and avoid the cost of repeatedly re-creating objects. The ObjectGrid feature is especially valuable to computationally intensive applications that usually have a number of worker nodes that need to share data among them. Also, if one server fails in the ObjectGrid, other servers in the grid can cache the object and supply the application with little or no loss of service.

WebSphere partitioning facility

Application partitioning using the WebSphere partitioning facility, when used correctly, can help support significant performance improvements and better scalability. And because the WebSphere partitioning facility is tightly integrated with the high-availability manager, it also supports high availability. The partitioning facility functions by alleviating a bottleneck that can be caused by many servers requiring access to the same database. By partitioning applications, the WebSphere partitioning facility can improve data-access performance by up to 20 times. As a result, business-grid applications with significant data volumes, and input and output can use the WebSphere partitioning facility to help increase application performance.

WebSphere high-availability manager

The foundational WebSphere Application Server Network Deployment product contains significant support for high-availability services. The high-availability manager runs as a service within each WebSphere process (deployment manager, node agents or application servers) and monitors the health of WebSphere singleton services. In the event of a server failure, the high-availability manager can redirect any singleton service that was running on the failed server to a peer server. This feature is especially useful for computationally intensive applications where master nodes gather information from worker nodes, typically over a long time period.

WebSphere Extended Deployment is one of the few platforms that offers support for applications that are both computationally data intensive and require high availability in a single, integrated package.

Bringing it all together: Policy-based workload management

Given a mix of OLTP, batch and computationally intensive features, WebSphere Extended Deployment can intelligently assign resources to each workload type and its associated applications according to business policy. This capability enables workload consolidation across a virtualized pool of WebSphere resources. For example, a bank running OLTP applications during normal business hours can use the WebSphere Extended Deployment autonomic workload-management capability to run computationally intensive portfolio-analysis or risk-analysis applications on resources traditionally dedicated to the OLTP applications.

WebSphere Extended Deployment systems administrators use service policies to describe the relative importance and target response times of various applications. When service policies are deployed describing service goals, WebSphere Extended Deployment works to achieve the goals. As demand for various applications changes, WebSphere Extended Deployment manages response times by starting or stopping application instances as the load on each application changes. It also manages the usage of the available instances according to resource share needed to achieve policy goals. WebSphere Extended Deployment provides management tools to help visualize current loads and the effects of its policy enforcement.

Conclusion

With companies looking to unlock infrastructure value to realize new levels of business success, WebSphere Extended Deployment is the right product, at the right time and the right place. It is built from the ground up on a platform of virtualization, which enables it to seamlessly and unobtrusively manage applications according to business and IT objectives. Companies specify priorities and policy, and WebSphere Extended Deployment manages resources based on those specifications. WebSphere Extended Deployment also offers robust support for high-availability and high-performance computing to allow you to feel confident in deploying critical applications. And it provides infrastructure value by enabling you to deploy new types of applications and workload across a single, unified grid.

Further, as companies continue to implement SOA throughout the enterprise, many services, some very fine-grained, some larger-grained, some mobile, are going to require a responsive and scalable IT infrastructure. WebSphere Extended Deployment meets the challenge by providing a resilient, scalable, highly available and fault-tolerant infrastructure using autonomic capabilities to meet service-level agreements (SLAs) and business-performance needs. With these capabilities, you can optimize the resource utilization and management of your IT infrastructure, while enhancing the quality of service for your critical applications.

For more information

To learn more about IBM WebSphere Extended Deployment, contact your IBM representative or IBM Business Partner, or visit:

ibm.com/software/webservers/appserv/extend

To join the Global WebSphere Community, visit:

www.websphere.org



© Copyright IBM Corporation 2006

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
06-06
All Rights Reserved

IBM, the IBM logo and WebSphere are trademarks of International Business Machines Corporation in the United States, other countries or both.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries or both.

Other company, product and service names may be trademarks or service marks of others.