Grid Computing

Business Service Grid

- SOA service aggregation and brokering by integrating web services and grid

July 28, 2004

Yih-Shin Tan
Senior Technical Staff Member
SWG System House, IBM Corp
Raleigh, North Carolina, USA
Agenda

- Customer expectations
- Service Domain overview
- LSCE BioGrid exploitation
- Product outlook
- References
On Demand Business

An on demand business is an enterprise whose business processes—integrated end-to-end across the company and with key partners, suppliers and customers—can respond with speed to any customer demand, market opportunity or external threat.
Start with The Bigger Picture—Business and IT

Where you start depends on YOUR organization’s priorities.

Increasing flexibility and simplification are key—business models, processes, infrastructure, plus financing and delivery
What Businesses Need From an On Demand Infrastructure

**Business Flexibility**

Integration

Business flexibility through integration of people, processes and information within and beyond the enterprise

**IT Simplification**

Infrastructure Management

IT simplification through automation and virtualization, enables access to and creates a consolidated, logical view of resources

- Availability
- Security
- Optimization
- Provisioning
- Infrastructure Orchestration
- Business Service Management
- Resource Virtualization of Servers, Storage, Distributed Systems/Grid and the Network

Built on open industry standards
What Customers are Telling Us

“I want to protect the investments I’ve already made in my IT infrastructure and skills base”

“I want my IT resources to adjust on the fly to the demands of business critical applications”

“I want to scale easily to accommodate transaction volume growth”

“I want a real-time view of what’s going on in my entire application environment”

“I want to get the most out of my infrastructure”

“I want to ensure high availability and reliability to accommodate peaks in transaction volumes”

“I want to ensure consistent and predictable performance for business critical applications”

“I want to reduce the human intensity of monitoring and managing my IT environment”
Grid and Web Services: Convergence?

The definition of WSRF means that Grid and Web communities can move forward on a common base.
Service Domain Motivation – simplicity and optimization of SOA

(Web) Admin clients → Service Domain application

(Web) user clients → Service Domain application → other clients

Service Domain application → Main Service Domain

Main Service Domain → Secondary service domain

Platform specific implementations → shared services → ... → shared services → J2EE, EJB, etc.
Architecture of a single Service Domain Node

1. **create**
   - Private Registry

2. **deploy**
   - Operational

3. **prepare**
   - Policy

4. **register**
   - Provided services

5. **subscribe**
   - Incoming req

6. **Service Domain Aggregation Engine**

- A collection of services represented as a single WSDL service to broker requestors to suppliers.

- Types of services:
  - Type 1
  - Type 2
  - Type 3
  - Type 4
  - Type 5
Topology of Service Domains

Clients

Main Service Domain

Aggregated ports/WSDL

Secondary domain North

Secondary domain South

Multiple secondary service images

service definitions and policy rules

single logical service image

Local service deployment

service-1 portfolio analysis

service-2 video

. . .

service-n Life Science

. . .

2ndry-x

2ndry-y

2ndry-z

. . .

multiple secondary service images

Aggregated ports/WSDL
Main Features of Service Domains

- **Web Services Aggregation Model**
  - One virtual port mapped to nested many service instances, multiple virtual ports in a domain, WSDL based, SOAP, SOAP-RPC, WSIF, .NET bindings

- **Policy driven service selection, brokering, etc.**
  - Rules: user and supplier service level definitions, filtering, discovery, selection, access, recovery, event handling, service level mapping, business relationship

- **Flexible node topology, recursive and self-discovery**
  - Join, merge, split, or leave domain nodes; plug or unplug services; dynamically

- **Autonomic and fail-over processing**
  - Built-in fail-over on exceptions, event logs and handlers to interoperate with global PD, monitoring, self-healing, and automation components

- **Workload Distribution**
  - General, Service features, relationship, and QoS based selection factors

- **Easy Customer exploitation**
  - End-to-end plug-in points defined around the service visualization-virtualization pattern

- **GUI Integration**
  - JSP/Servlets, JSP/Portlets, or other UI schemes hooks up easily to administrative and user invocation task beans
Customer Use Case

- **Life Sciences BioGrid**
  - A joint project between IBM and Taiwan government (MOEA)
  - Subsidy from National Research Program for Genomic Medicine
  - Solve real world life sciences R&D problems

- **Application of Service Domain technology**
  - Concrete SOA deliverables
  - Initial project milestones (6/2004)
    - Integration of Bioinformatics Tools (discussion today)
    - Grid enablement for Bioinformatics value-added database
Integration of bioinformatics tools
An Example of Infectious Disease Analysis (IDA)

- Perform Gene analysis to research cause of disease
- Bioinformatics Tools used
  - SRS – search protein sequence database
  - BLAST – perform sequence alignment
  - EMMA – calculate homology relationship among sequences
  - EPROTDIST – compute distance matrix from protein sequences
- Manual Method
  - Existing bioinformatics tools scattered on various systems and internet.
  - Various biotech information sources with different schema.
  - Complicated bioinformatics tools usage combination too hard to manage the biotech analysis.
  - User proceeds IDA step by step manually
Problems

- Existing bioinformatics tools scattered on various systems or internet, needs an integrated solution with standard specification.

- A biotech analysis process usually contains complicated bioinformatics tools usage combination (5+ steps) for each DNA/Protein sequence, the number of DNA/Protein sequences involved in analysis may be more than thousand, biologist proceed most of those steps transition manually. An automation mechanism is required.

- Many bioinformatics tools take long running time, or large volume of information generated, it is required to support the status/information management of biotech analysis process.

- Need an integrated user interface for biotech analysis status/information presentation with user’s preference.
IBM Life Sciences Solution

- **Integration**
  - Service oriented architecture (WebSphere Enterprise Service Bus - Service Domain)
  - Bioinformatics tools and biotech analysis processes are web services based
  - Front end integration (WebSphere Portal)
  - End to end integration (WebSphere Process Choreography)
  - Platform (WebSphere Application Server)

- **Automation**
  - Accelerate the biotech analysis process

- **Personalized biotech analysis management**
  - Based on user's preference to manage the biotech analysis processes

- **Existing product/technology utilization**
  - See list above
Prototyping IDA flow

Keywords: TFIIIA

PubMed

SRS

browsable reference list

pick seq

mutifasta

Emma

XML File

blast

1

1 or more sequences

more

submit multiple parallel jobs

Display result

Prophecy

.dnd

.dnd

Treeview

.eprodist

Display result

profy

Ex: tf3a-human

.aln

.phy

Profit

Profit
An IDA Flow Segment
Main Portal page: access to automated IDA flow, or manually to dispersed tools
IDA Portal page: keep this flow, current state is SRS finished
IDA Portal page: per flowchart, select SRS result sequence to prepare for BLAST
IDA Portal page: set runtime parameters and go BLAST
IDA Portal page: wait for BLAST to finish, leave to resume later or wait here
IDA Portal page: BLAST result back, continue the next step of the flow chart
**PALSdb Integration**

To maintain the state of the bio analysis flows, user submitted data is uploaded for processing, and the state of the bio analysis flows is maintained. SOAP messages are exchanged between the Application Server (process) and the Application Server (services), with File IO used for file transfer.

The Portal Server communicates with the Application Server (process) and the User Interface, with HTTP requests and responses for user management (Portlets).

The Shared File System stores data uploaded by users or generated by bio flow, and is accessed through File IO.

For querying and generating programs, the Grid1.tw.ibm.com and PLASdb Services are used.

Various bioinformatics programs such as BLAST, EMBOSS, and others are integrated into the system.
Project status

- On track of project milestones
  - Bioinformatics Flows (shown earlier)
  - Bioinformatics value-added database

- Exploit/incorporate new IBM ODOE and Industry Standard Technologies, whenever available
  - IBM WebSphere XD
  - IBM Virtualization Engine
  - Globus Toolkit, IBM ETTK . . .

- Migrate timely with appropriate solution platforms
WebSphere Extended Deployment
Virtualization Engine
Emerging Technology Toolkit
Enterprise Service Bus/Web Services Gateway
...WebSphere Extended Deployment Version 5.1 Provides Advanced Features

Dynamic Operations
- Resource pooling and allocation in a WebSphere environment
- Application differentiation through operational policies
- Dynamic routing and workload management

Extended Manageability
- Flexible modes of runtime operations: manual, supervised and on demand
- Runtime operations tree map, charting and event visualization tools

High Performance Computing
- Application partitioning
- High availability services
- J2EE development of high-end OLTP applications
Virtualized WebSphere Environment

“I want my IT resources to adjust on the fly to the demands of business critical applications”

Scale on demand through dynamic allocation of WebSphere resources

Create pools of resources that can be shared among applications

Breaks the tie between application clusters and machines which can now be shared amongst applications, optimizing resource utilization and simplifying overall deployment

Scale for unpredictable application demand

Dynamically adjusts application resources as needed depending on demand

Scale beyond the defined application server pool with Tivoli Intelligent Orchestrator (option)

Dynamically expands and contracts resources by adding and removing machines into and out of resource pools when it is deemed necessary by monitoring logic
Virtualization Engine

Suite for Servers
- Tivoli Provisioning Manager for IBM @server®
- IBM Grid Toolbox
- Enterprise Workload Manager
- IBM Director Multiplatform
- Virtualization Engine Console

Storage
- IBM TotalStorage® Productivity Center
- IBM TotalStorage SAN Volume Controller
- IBM TotalStorage SAN File system
IBM Grid Toolbox V3 for Multiplatforms

**GT3.0 Core (with IBM Contributions)**
OGSi, Notification, Registry

**IBM substitutions / enhancements**
Grid Services
CMM (a.k.a. CRM), Policy, Repository, Container Mgmt
Management UI Admin,
Command-Line Tooling
OGSi Logging (using WAS facilities)

**GT3.0 Services**
Job Execution Services
Resource Management Services
Information Services
GT3-Security-Compatible

**Embedded Technologies**
WAS Bobcat 5.0.2
Database (Cloudscape)

Integration of GT3.x and IBM technologies into a Single Product
Available on AIX, and Linux on xSeries, pSeries, iSeries and zSeries
OS/400 available 6/25/04
Emerging Technology Toolkit

- **WS-MetadataExchange**: The WS-Metadata Exchange specification was recently published by IBM, BEA, Microsoft, and SAP and used in a WS-Policy scenario.

- **Web Services Distributed Management (WSDM 0.5)**: Oasis specifications define management of IT and Web service resources via Web service protocols.


- **Autonomic Manager Toolset (AMTS)**: The Thermostat demo has been converted to use the new WS-Resource Framework prototype implementation, rather than the Globus Toolkit 3.0 implementation of OGSI 1.0.

- **Common Base Event (CBE) Version 1.0.1**: common event format for logging, management, problem determination, autonomic computing, and e-business functions

- **Service Domain**: web service aggregation and brokering

- **Others . . .**
More Information

- Details about the Service Domain technology will be published in a special issue of IBM System Journal about Grid Computing in November, 2004

- LSCE Taipei BioGrid demo site and documentation
  - Contact Yih-Shin Tan, ystan@us.ibm.com

- Service Domain technologies in ETTK 1.2 from IBM alphaWorks
  - http://alphaworks.tech/ettk

- White papers on IBM developerworks, starting from

- China Research and Education Grid

- IBM Grid (VE) and Software Group (WebSphere XD) web sites
  - http://www.ibm.com/software
Questions?

Thank You