Grid & Virtualization Working Group

OGF23 gridvirt-wg

Erol Bozak, Chair
SAP, Development Architect

Wolfgang Reichert, Co-Chair
IBM, Senior Technical Staff Member

June 2008
Barcelona

OGF IPR Policies Apply

- "I acknowledge that participation in this meeting is subject to the OGF Intellectual Property Policy."
- Intellectual Property Notices Note Well: All statements related to the activities of the OGF and addressed to the OGF are subject to all provisions of Appendix B of GFD-C.1, which grants to the OGF and its participants certain licenses and rights in such statements. Such statements include verbal statements in OGF meetings, as well as written and electronic communications made at any time or place, which are addressed to:
  - the OGF plenary session,
  - any OGF working group or portion thereof,
  - the OGF Board of Directors, the GFSG, or any member thereof on behalf of the OGF,
  - the ADCOM, or any member thereof on behalf of the ADCOM,
  - any OGF mailing list, including any group list, or any other list functioning under OGF auspices,
  - the OGF Editor or the document authoring and review process
- Statements made outside of a OGF meeting, mailing list or other function, that are clearly not intended to be input to an OGF activity, group or function, are not subject to these provisions.
- Excerpt from Appendix B of GFD-C.1: "Where the OGF knows of rights, or claimed rights, the OGF secretariat shall attempt to obtain from the claimant of such rights, a written assurance that upon approval by the GFSG of the relevant OGF document(s), any party will be able to obtain the right to implement, use and distribute the technology or works when implementing, using or distributing technology based upon the specific specification(s) under openly specified, reasonable, non-discriminatory terms. The working group or research group proposing the use of the technology with respect to which the proprietary rights are claimed may assist the OGF secretariat in this effort. The results of this procedure shall not affect advancement of document, except that the GFSG may defer approval where a delay may facilitate the obtaining of such assurances. The results will, however, be recorded by the OGF Secretariat, and made available. The GFSG may also direct that a summary of the results be included in any GFD published containing the specification."

- OGF Intellectual Property Policies are adapted from the IETF Intellectual Property Policies that support the Internet Standards Process.

Agenda

5 min	**Recap of Previous Sessions**	Erol
30 min	**SAP Scenarios & Demo**	Erol
30 min	**Interface Requirements & Proposals**	Wolfgang
10 min	**Relation to RESERVOIR Project**	Wolfgang
15 min	**Next Steps & Discussion**	All

- Recap of Previous Sessions

Goals of the WG

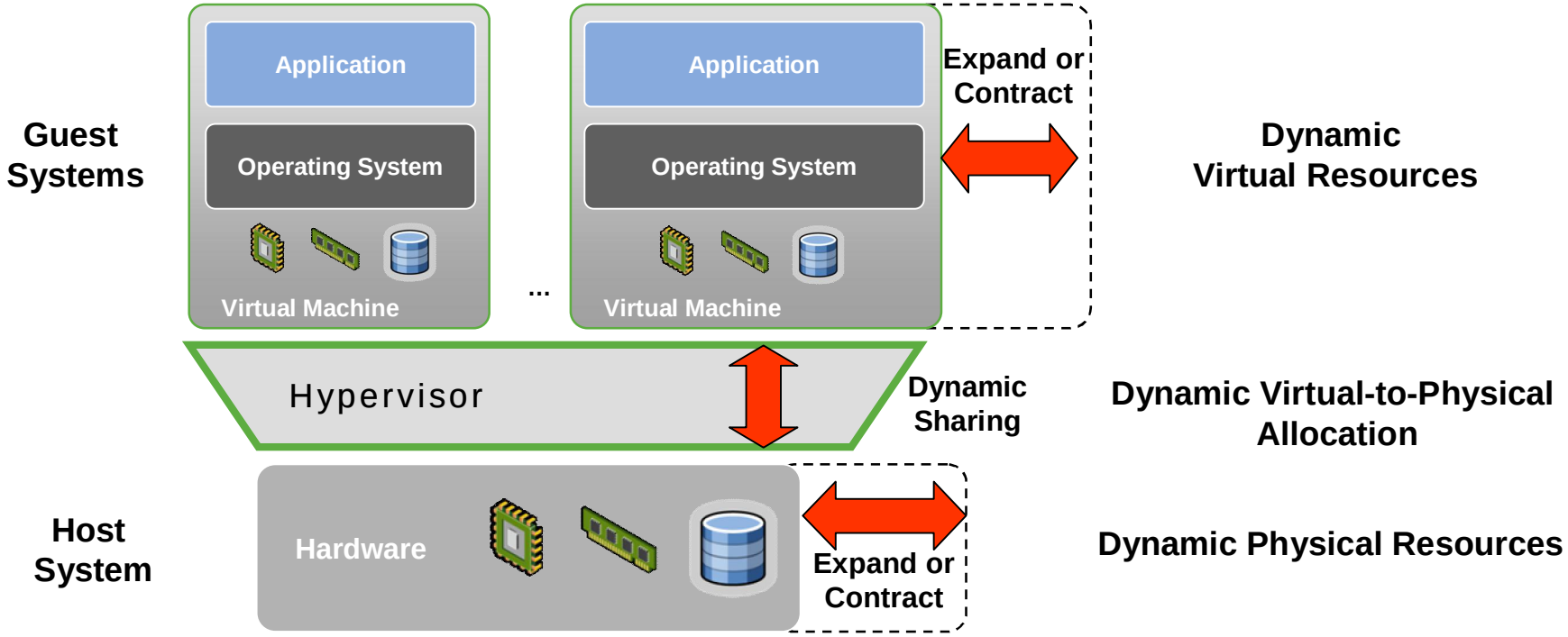
1. Verification that within existing Grid standards the specifications are neutral to virtualized systems and resources

2. Explore how virtualization technologies can be exploited to better support Grid use cases
   - Describe the use cases and scenarios

3. Define the requirements to the Grid architecture for integration with system virtualization platforms

4. Define profiles (i.e. usage recommendations)
   - Align with existing standards, e.g. DMTF - System Virtualization, Partitioning and Clustering (SVPC)

Recap: Virtualization Concept



- Virtualization decouples presentation of resources to consumers (applications) from actual resources through a virtualization layer (Hypervisor)
- Several virtual machines (VMs) may run on a single physical host
- Each VM has its own installed operating system and applications

Recap:
Key Capabilities of Virtualization

- **Creation of virtual systems on-demand**
  - Specify the environment the application / jobs needs to run
  - The environment of the allocating can be pre-configured an persisted as images that can be activated on creation (multiple times if necessary)
- **Dynamic resizing**
  - Change the configuration of virtual system
- **Isolation**
  - Applications / jobs can run isolated from each other
- **Snapshotting**
  - Suspending the virtual system and persisting the state which can be reactivated again
- **Migration**
  - Movement of virtual system among host systems (physical systems)

Recap: Use Cases Overview

Virtualization use cases

- Power saving
- Planned maintenance
- Changing capacity requirements
- Changing capacity offering/availability
- Stateful cloning
- Protecting long running jobs from system failures
- Reproducing situations
- Metering of job resource consumption
- Resource consumption enforcement
- Protection against malware
- Ensuring Security
- Avoiding conflicts
- Emulating an environment for legacy jobs

Virtualization capabilities

Live migration

Dynamic resizing

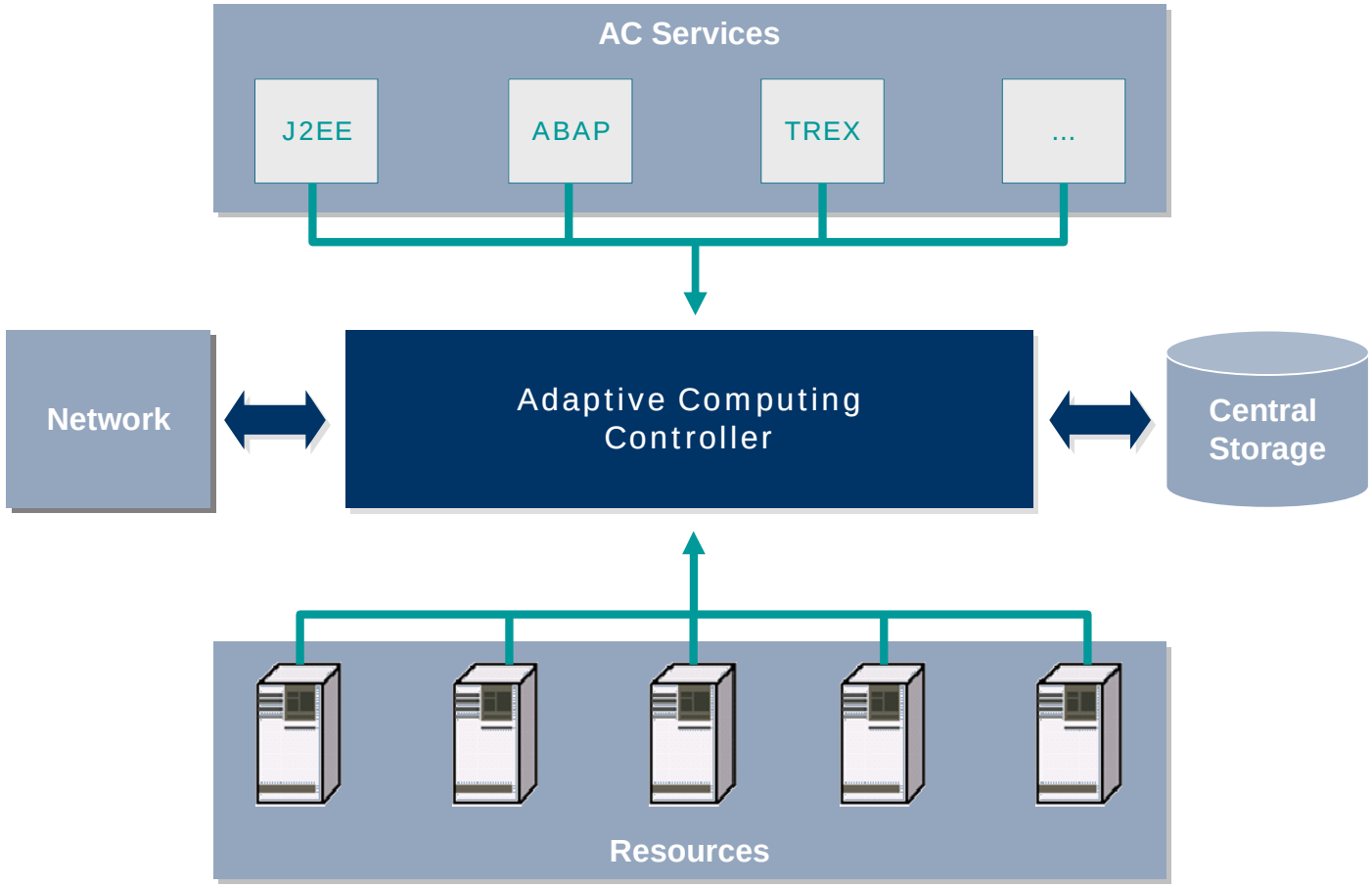
Snapshotting

Isolation

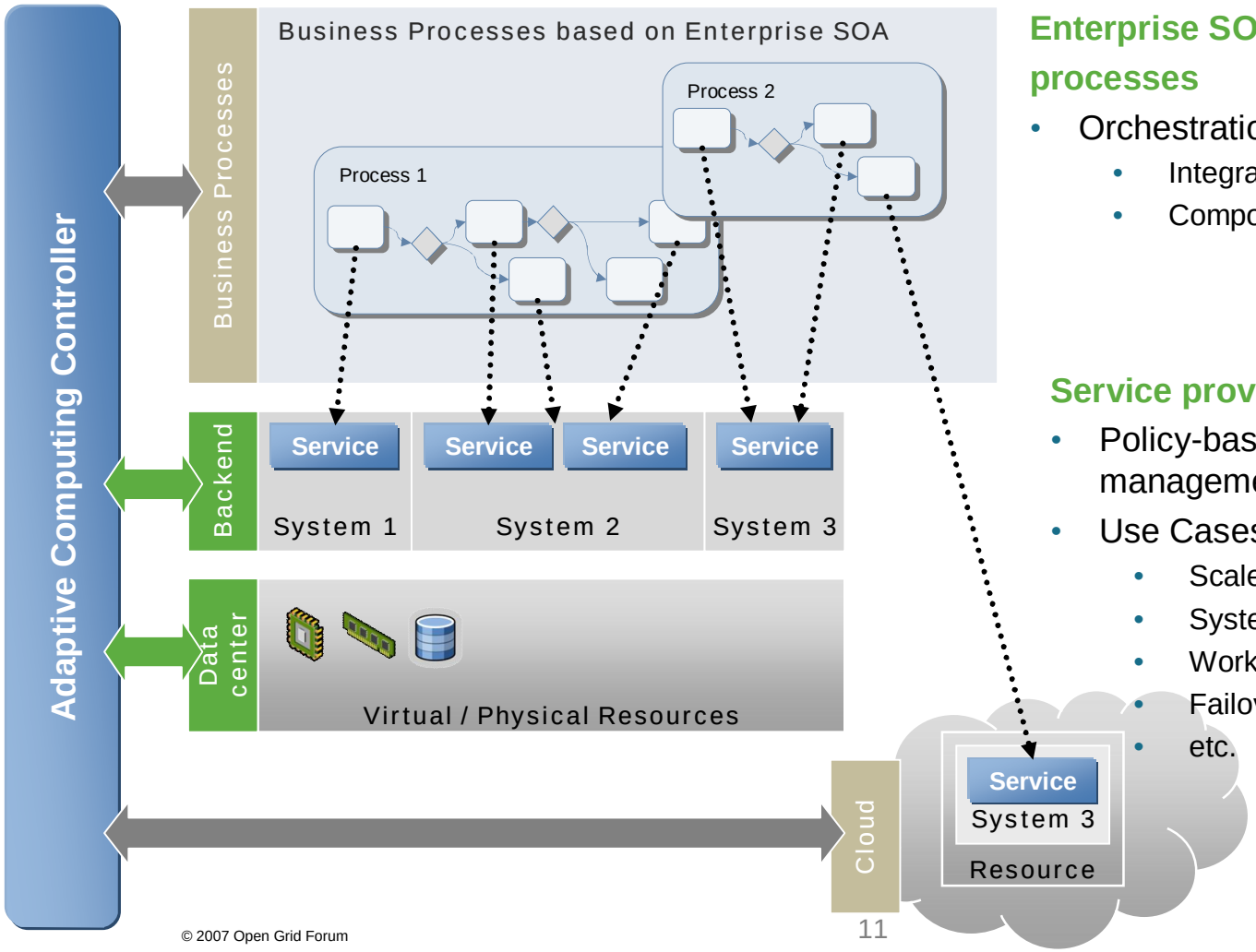
Provisioning

- Adaptive Computing @ SAP
- Demo

Adaptive Computing @ SAP



Enterprise SOA Resource Management



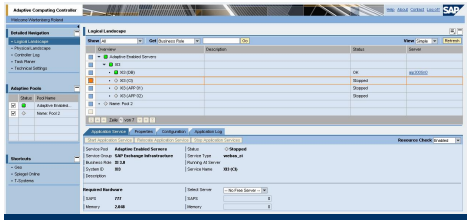
Enterprise SOA based business processes

- Orchestration at two scenarios:
  - Integration scenario
  - Composition scenario

Service providers hosted at systems

- Policy-based resource / capacity management
- Use Cases:
  - Scale-out
  - System provisioning
  - Workload balancing
  - Failover
  - etc.

How does Adaptive Computing Work?



Adaptive Computing Controller

Computing Resources

Software installed on Central Storage

Computing Resource

Computing Resource C1

Central Instance

Computing Resource

Computing Resource

...

3 Relocate

1 Prepare Environment

2 Start

1) Stop Central Instance on C1
2) Unprepare on C1
3) Prepare on C2
4) Start Central Instance on C2

Database

Central Instance

Dialog Instance

...

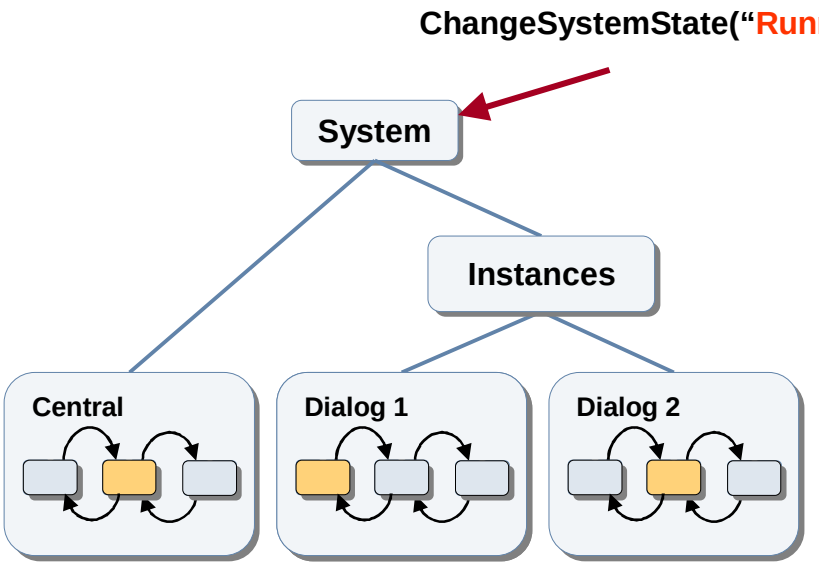
...

ABAP System (SID S01)

J2EE (SID S0n)

Central Storage

Execution Plan Generation



1) System Structure

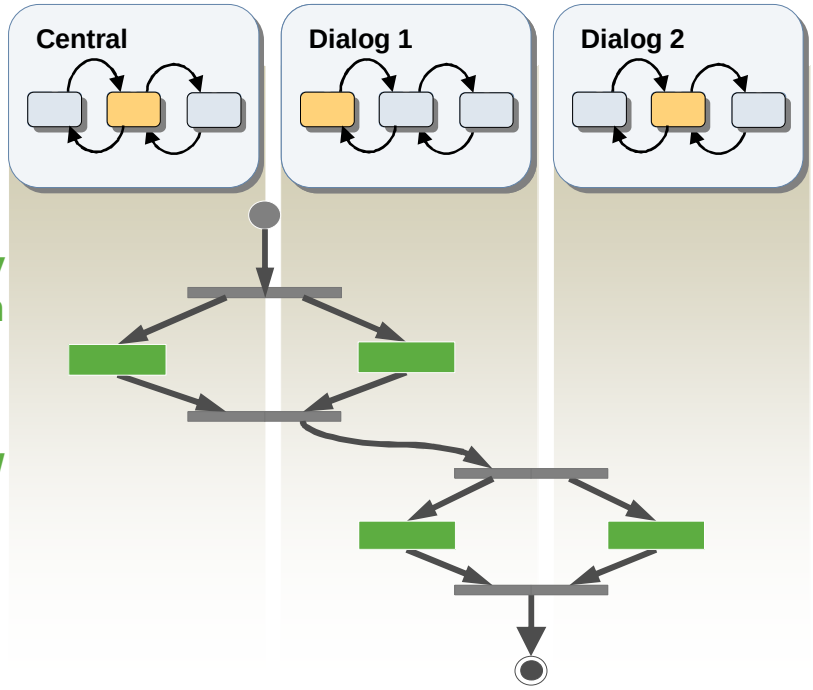
- Hierarchical
- Leaf node represent executable entities (e.g. SAP Instance)
- Leaf nodes have lifecycle states (stopped, started , prepared etc.)
- Intermediate nodes represent aggregation

2) Execution Plan

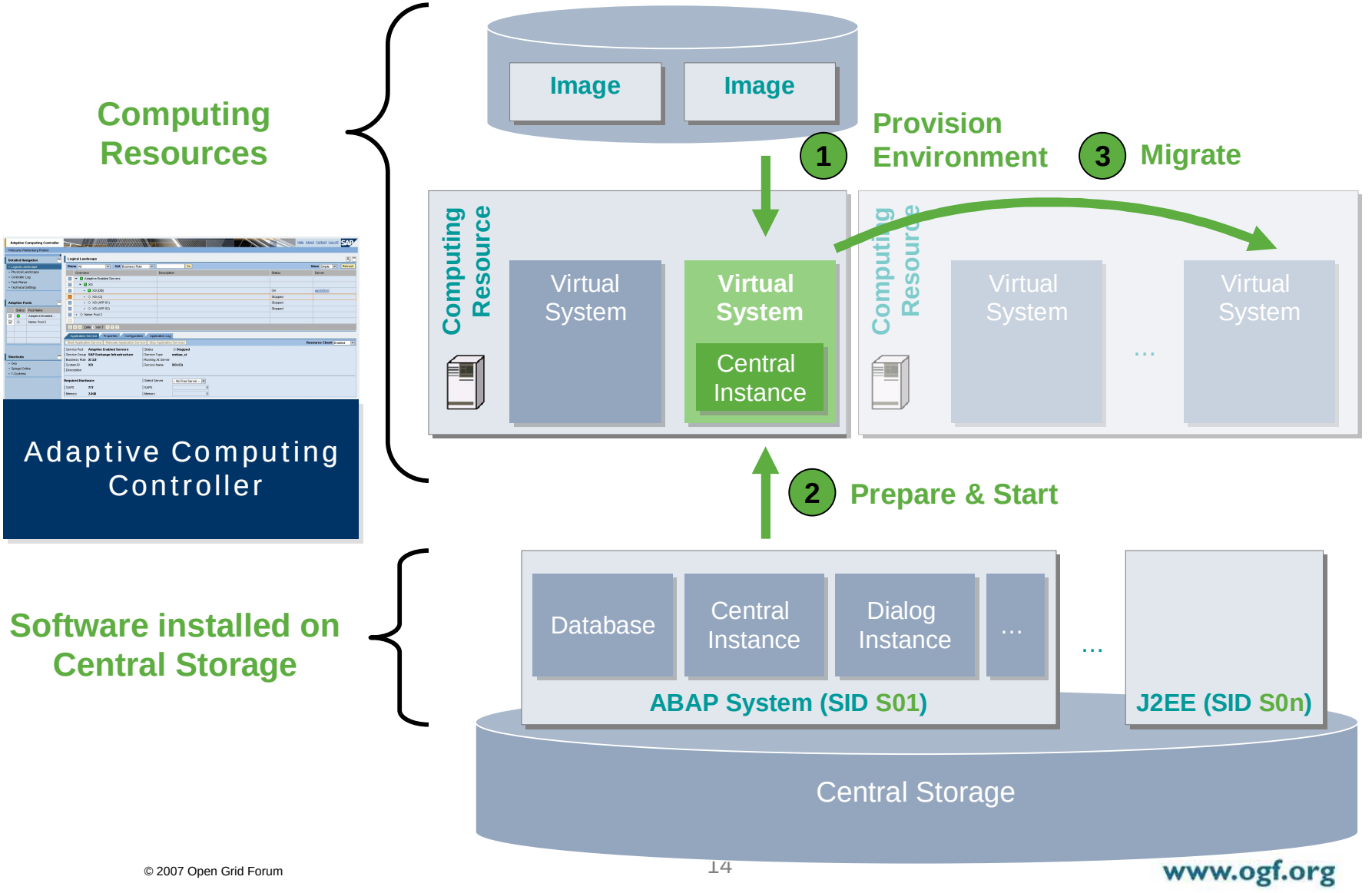
1. Action sequence per node and dependency calculation
2. Execution plan generation (i.e. workflow)

Dependency Diagram

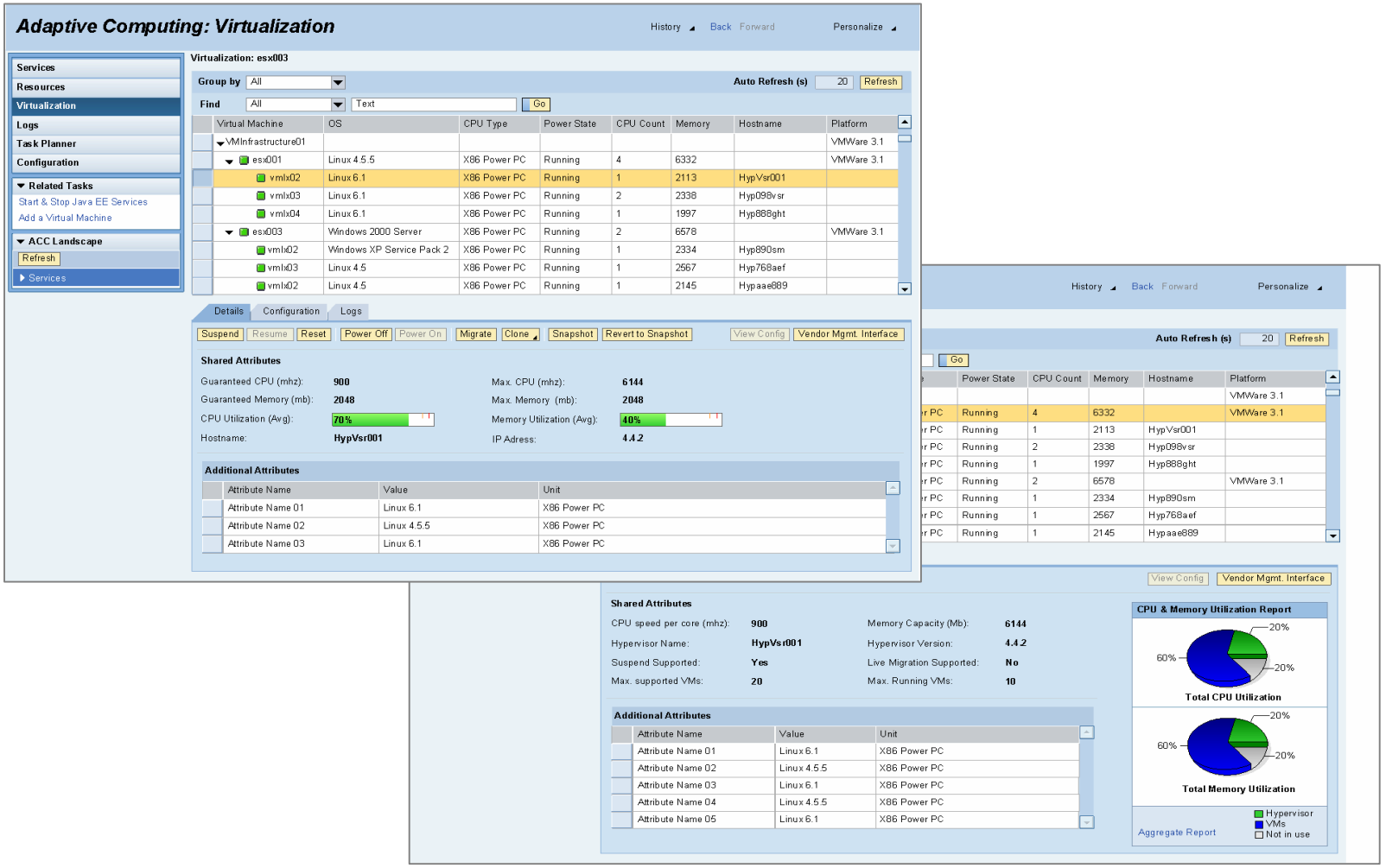
Workflow



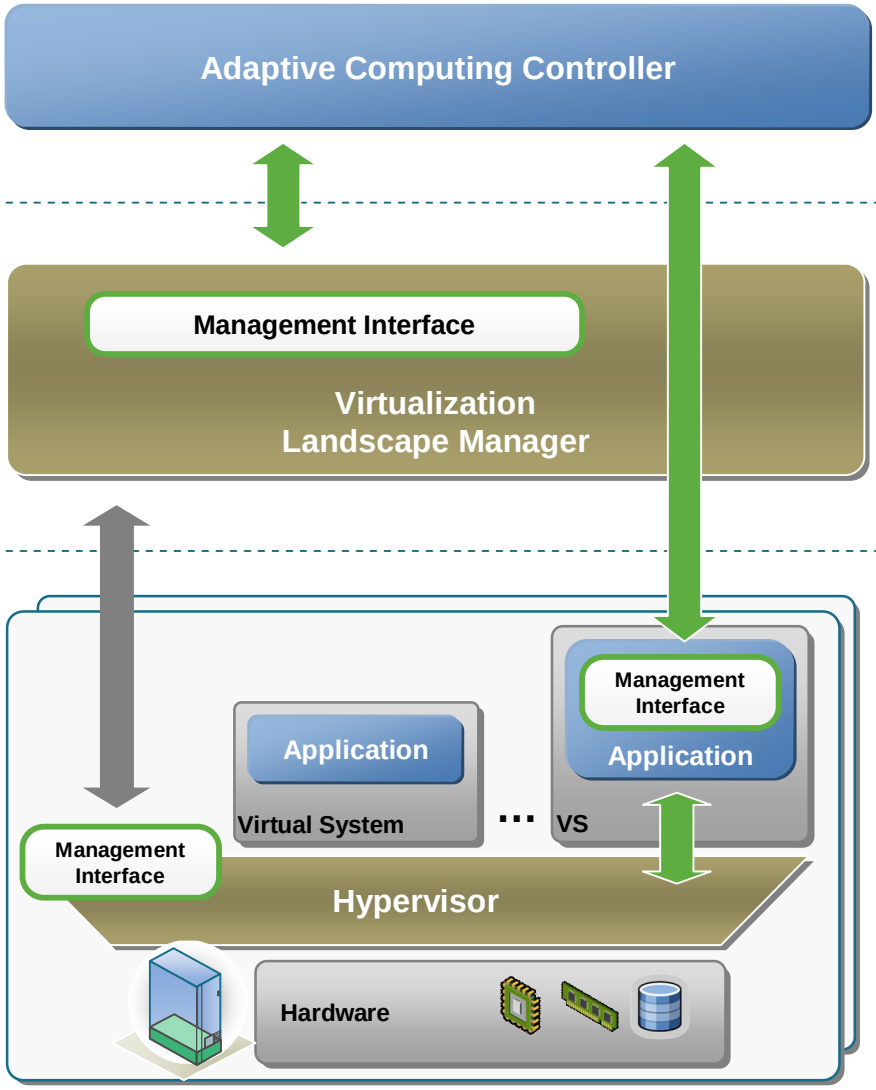
How does Adaptive Computing Work?



Demo



Levels of Integration



- **Adaptive Computing Controller provides „higher level“ capabilities for managing SAP systems and resources**
- Manages physical as well as virtual resources

- **Platform specific „Virtualization Landscape Manager“ (VLM) manages distributed sets of virtual systems and resources in a landscape**
- Proprietary interfaces exposed at this level !

- **Platform specific interface exposed at the „Virtualization Engine“ (VE), e.g. Hypervisor**
- Standardization is focused on this level
- Initiative at the DMTF: SVPC (Server Virtualization, Partitioning & Clustering)
- Application awareness needed !

Capabilities

- Operations
  - Types:
    - Create (SVPC: Define) (see provisioning)
    - Destroy
    - Power on (SVPC: Activate) / Power off (SVPC: Deactivate) / Reset
    - Shutdown, Reboot
    - Migrate
    - Clone
  - Retrieving “operation logs” of executed and/or executing operations
  - Asynchronous / synchronous processing of operations
    - Tracking progress
    - Cancelable operations
- Provisioning of Virtual Systems
  - Templates based virtual system creation
  - OVF (Open Virtual machine Format) based provisioning
- Configuration
  - Configuration exclusively for resizing
  - 1. Activation of predefined size configuration (e.g. Small, Medium, Large, XL)
  - 2. Customization via standard metrics

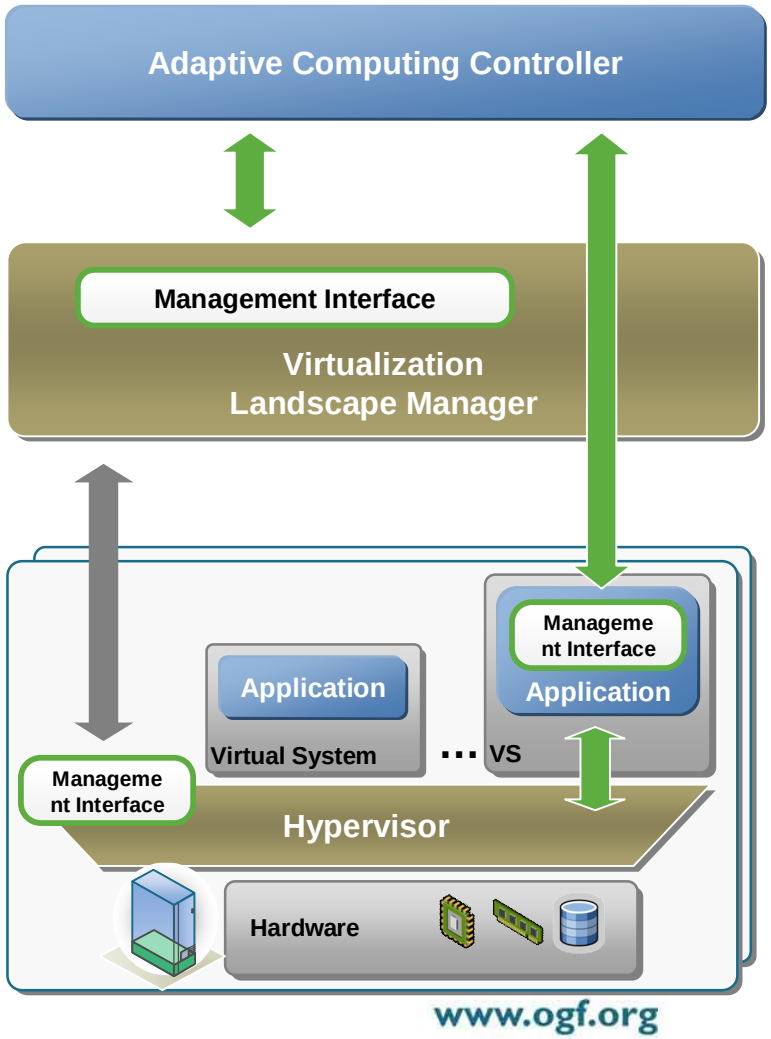
Capabilities

- Inventory / Discovery / Monitoring
  - Inventory & discovery of templates and virtual systems and hosts
  - Monitoring of virtual systems and hosts
  - Selective monitoring (through attribute specification) for all managed entities in the virtualization landscape, e.g. virtual systems, host systems, operations (tasks) etc.
- Snapshot
  - Creation of snapshots and reverting virtual systems from its snapshots
- Other
  - Retrieving alerts from the virtualization platform
  - Alerts may be raised for errors & exceeding thresholds (i.e. policy violation)
  - Alerts have to have a reference to the alert details / situation

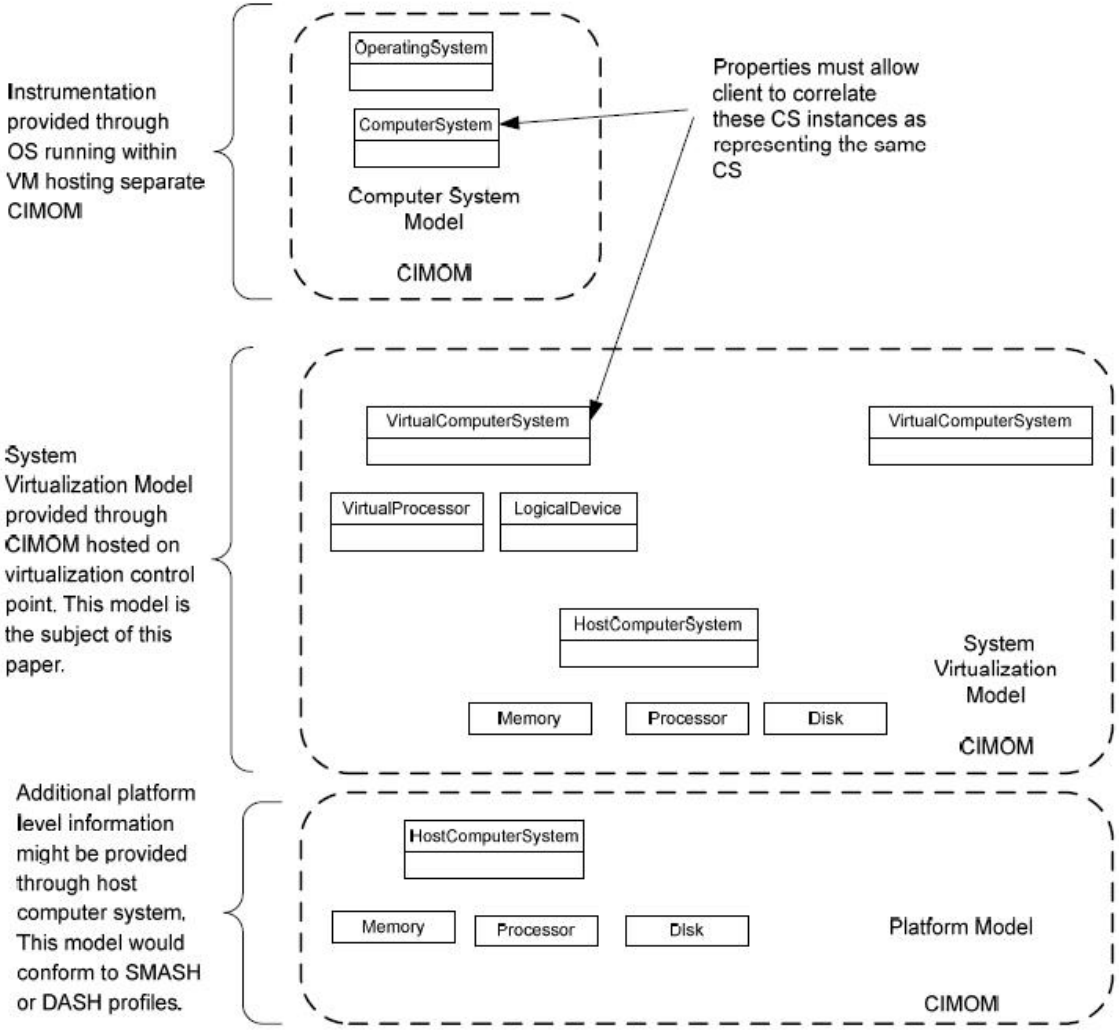
- Interface Requirements & Proposals

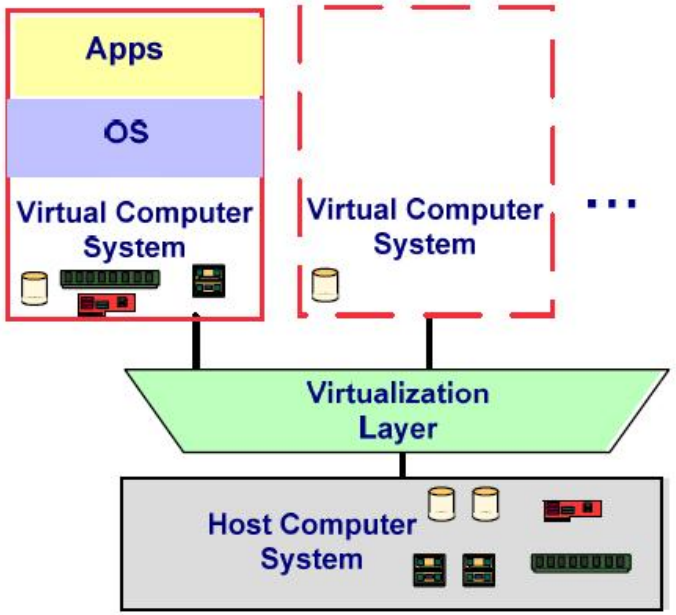
Management Stack

- Cloud / Grid
- Datacenter / Landscape (*heterogeneous environment*)
  - SAP ACC
  - …
- Cluster (*homogeneous environment*)
  - Virtual Center (VMware)
  - HMC (Hardware Management Console, IBM)
  - IBM Director
  - ZENworks Orchestrator (Novell)
  - …
- Virtual Systems running OS
- Servers / Hypervisors
  - IBM POWER Hypervisor, z/VM
  - VMware ESX
  - XEN
  - …

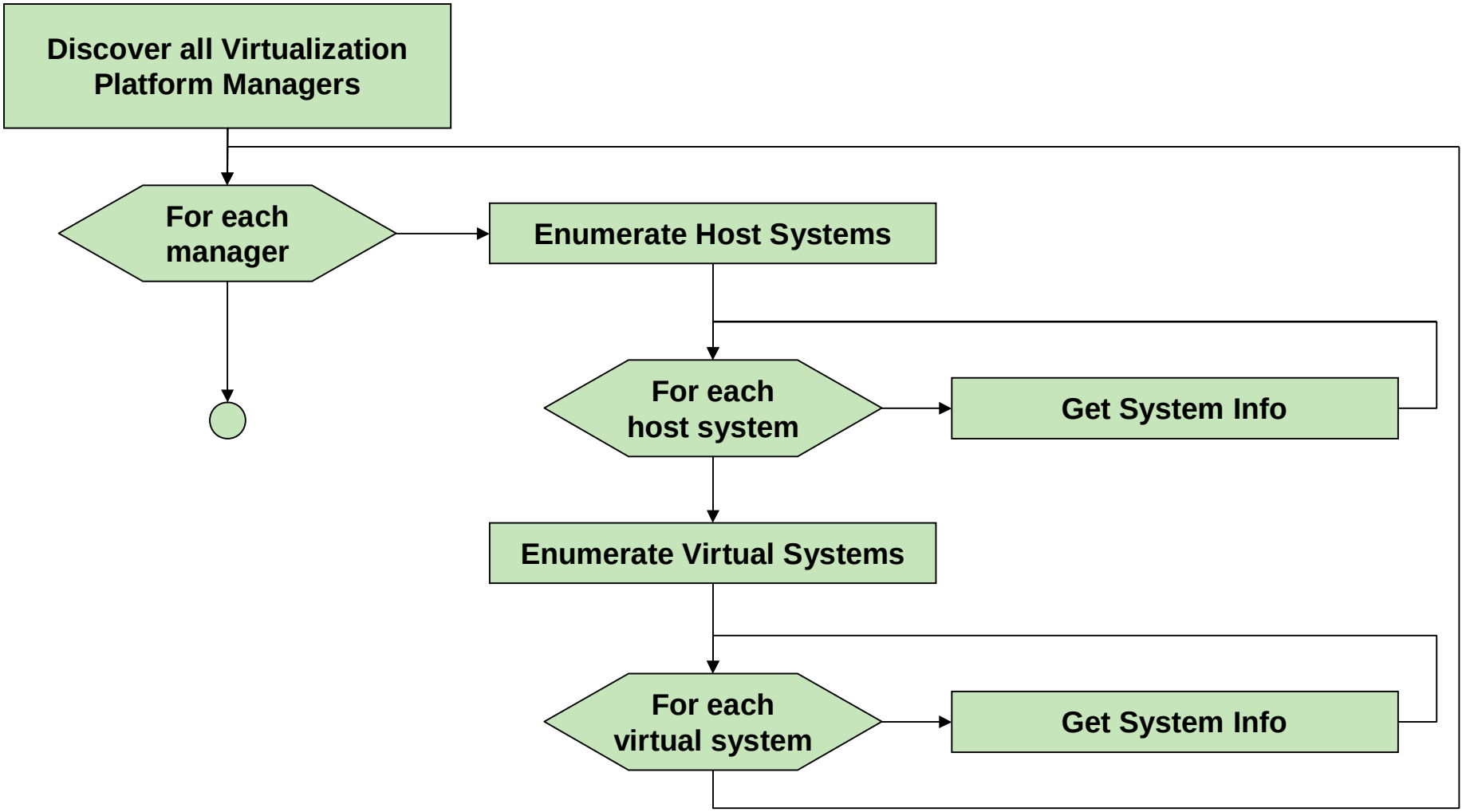


CIM System Virtualization Model

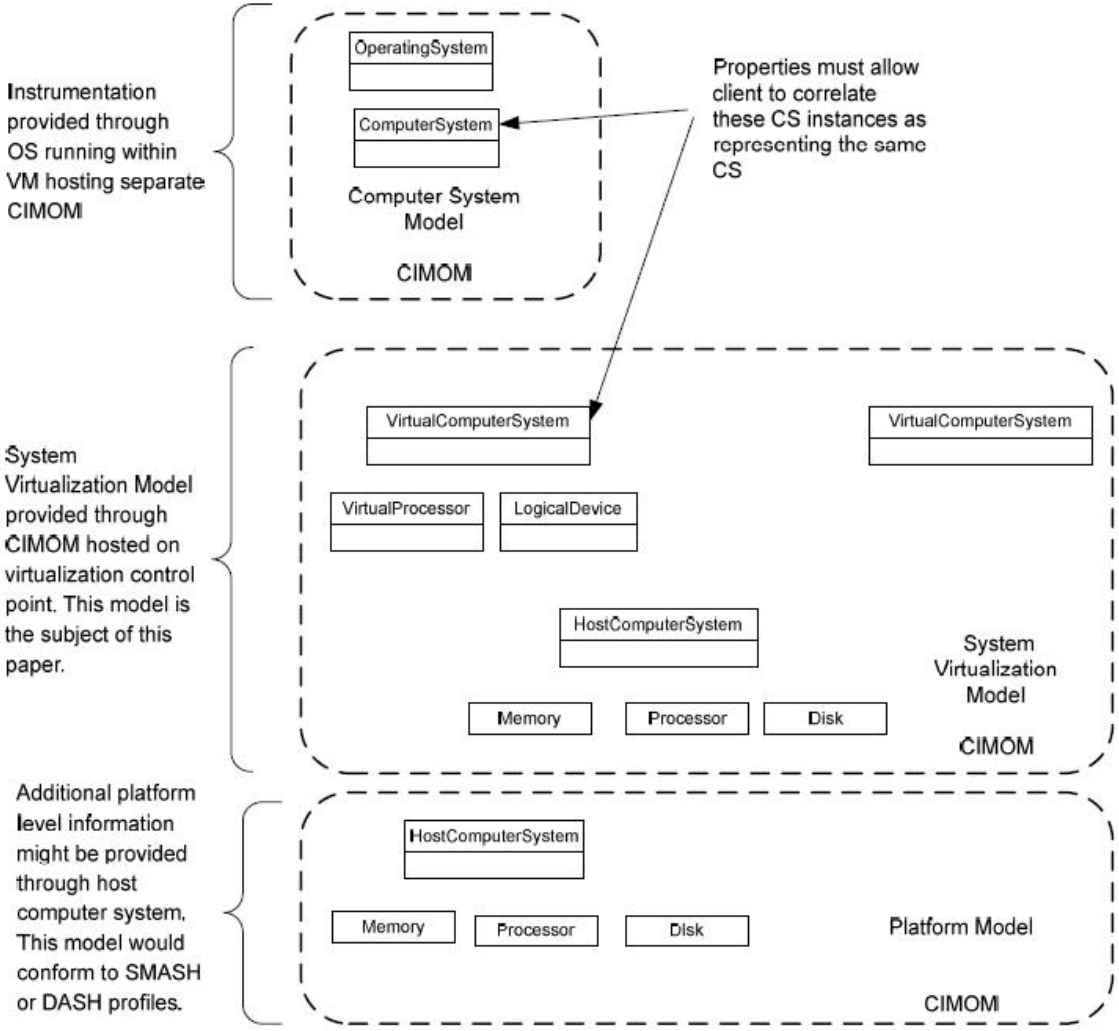


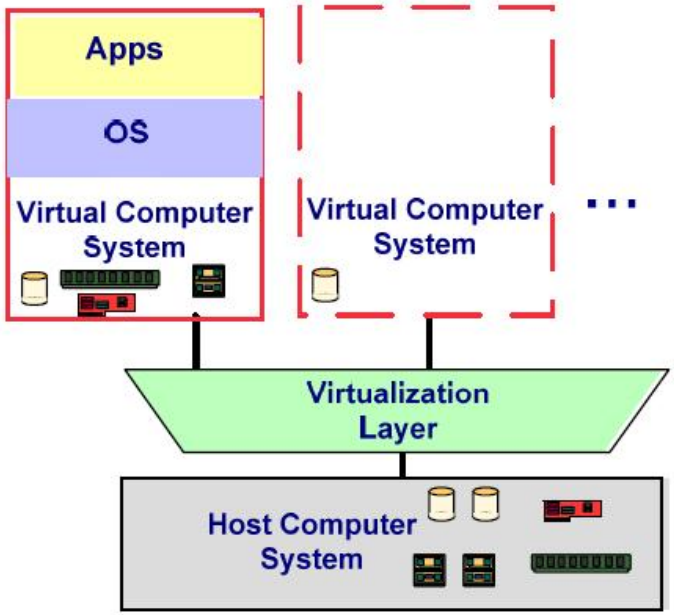


Topology Discovery



Virtualization Modeling Environment





Topology Discovery: CIM Approach

- CIM = distributed model
  - Manager needs to establish contact with each CIMOM
    - SLP broadcast to discover all CIMOMs
    - Query CIM_RegisteredProfile → "System Virtualization"
  - Enumerate instances

- Virtualization manager knows ComputerSystem data, but not OperatingSystem data
  - Need to contact CIMOMs in OS for application driven scenarios
    - Many (hundreds/thousand) CIM connections required

  - Associators indicate relation
    - CIM ComputerSystem ID is not hostname

➢ Higher level API could be more efficient than CIM

Topology Discovery…

GetAllHostSystems

IN

- With Filter: Hardware architecture (INTEL, POWER, …) *optional*
- With Filter: Virtualization type (XEN, VMware, z/VM, …) *optional*

OUT

- List of system identifiers

GetAllVirtualSystems

IN

- With Filter: State (Active, Defined, …) *optional*
- With Filter: RunningOperatingSystem (Linux, Windows, AIX …) *optional*

OUT

- List of system identifiers

GetAllVirtualSystemsOnHost

IN

- System identifier
- With Filter: State (Active, Defined, …) *optional*
- With Filter: RunningOperatingSystem (Linux, Windows, AIX …) *optional*

OUT

- List of system identifiers

Topology Discovery…

GetSystemInfo

IN

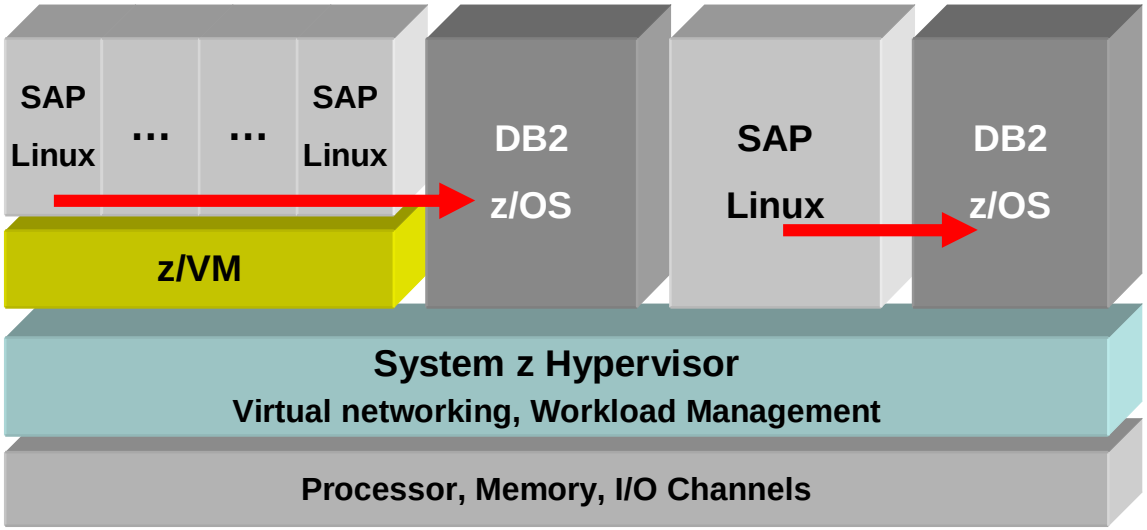
- List of system identifiers

OUT

- List of base properties for each system

Topology Discovery: Implications

- Virtualization is not limited to two layers, rather it is a tree
  - A virtual system might be a host system for other virtual systems
  - The tree is navigated starting with the (physical =) lowest level host system



Multiple virtualization layers on IBM System z

Virtual System Settings and Capabilites

GetVirtualSystemSettings

IN

- System identifier

OUT

- List of settings

GetResourceAllocationSettings

IN

- System identifier

OUT

- List of resource allocation data
- Available resources in the pool

GetVirtualSystemManagementCapabilities

IN

- System identifier

OUT

- List of supported management services

ChangeVirtualSystemSettings

IN

- System identifier
- List of changed settings

OUT

- List of settings

ChangeResourceAllocationSettings

IN

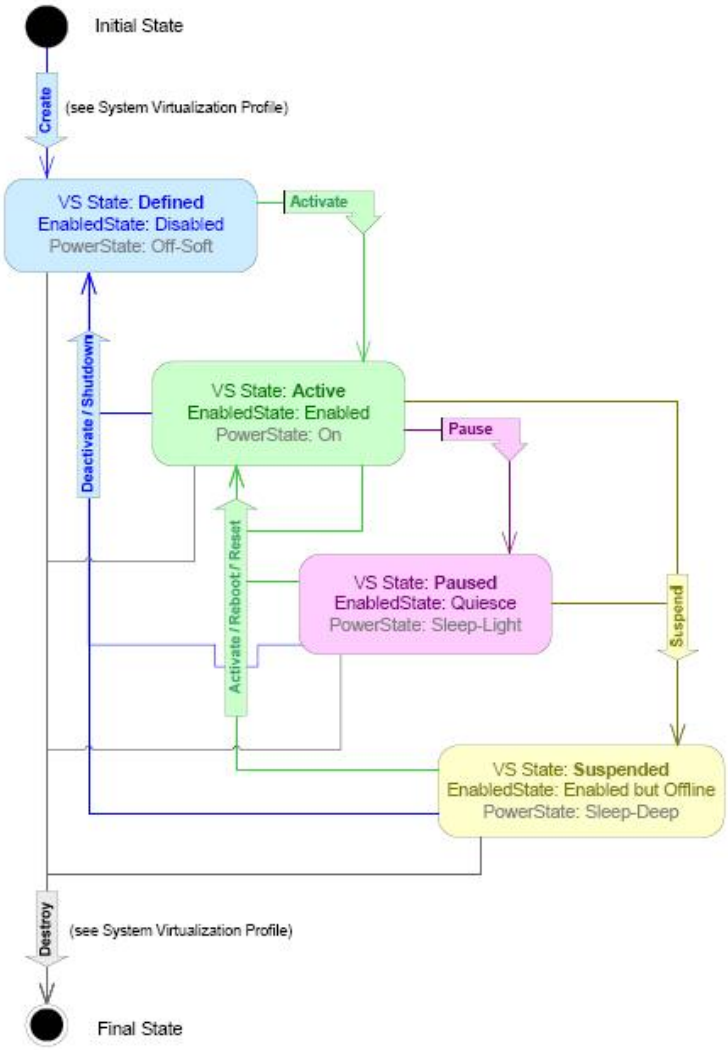
- System identifier
- List of changed resource allocation data

OUT

- List of resource allocation data

Virtual System Lifecycle Operations

- Virtual system state transitions
  - Create / Destroy
  - Activate / Deactivate
  - Pause / Suspend
  - Shutdown / Reboot / Reset
- Value-added operations
  - Migrate
  - Clone
  - Create/Apply/Destroy snapshot
- Helper services
  - Query progress of asynchronous operations
  - Cancel asynchronous operations
  - Retrieve operation logs



CIM virtual system state model

Snapshot / Live Migration

CreateSnapshot

IN

- Virtual System identifier
- Name
- Description

OUT

- Snapshot identifier

ApplySnapshot

IN

- Virtual System identifier
- Snapshot identifier

DestroySnapshot

IN

- Snapshot identifier

ListAllSnapshots

IN

- Virtual System identifier

OUT

- List of snapshot identifiers and attributes

MigrateVirtualSystem

IN

- Virtual System identifier
- Target host system identifier
- Mode (Test / Prepare / Execute)
- Properties [*optional*]

OUT

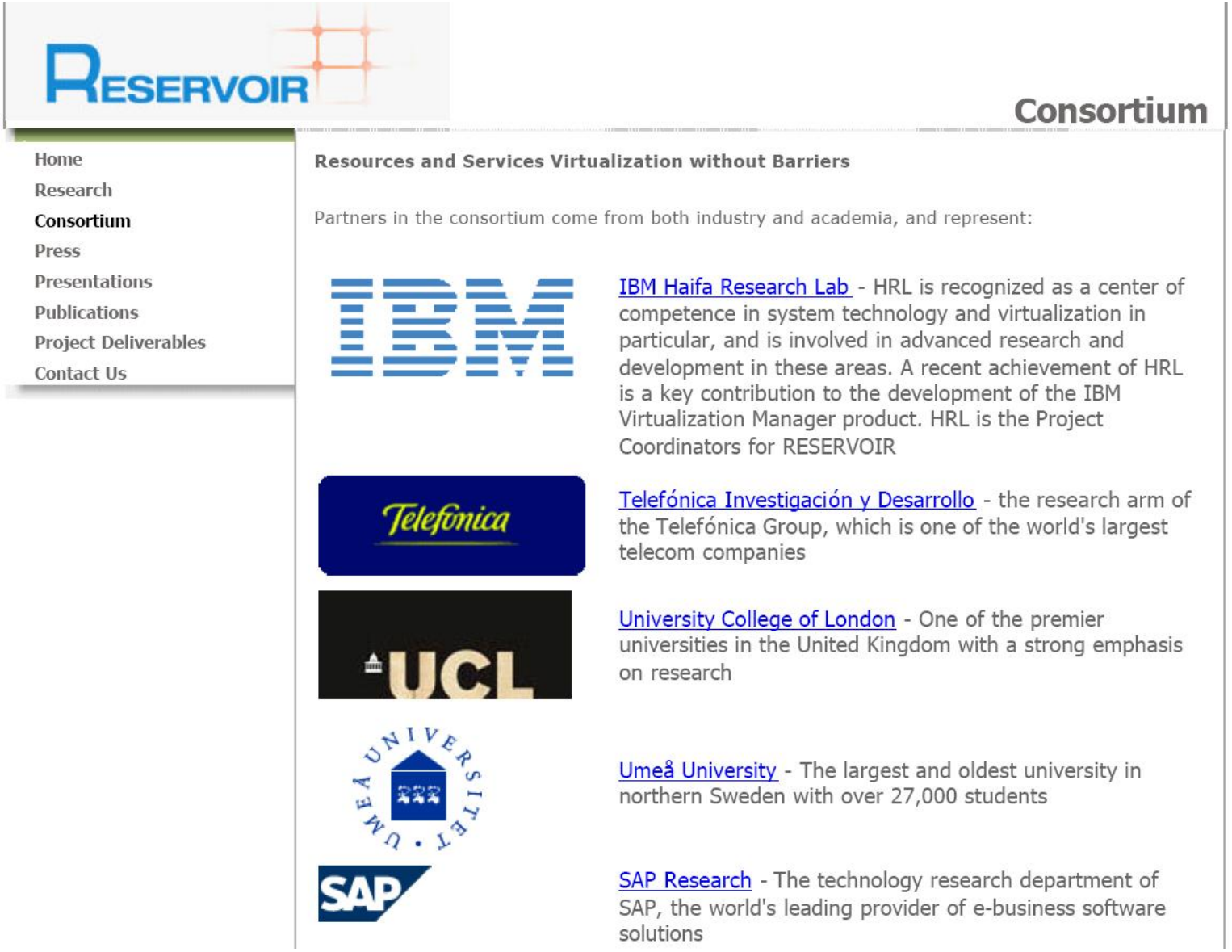
- Job identifier

- Relation to RESERVOIR project

RESERVOIR

http://www.reservoir-fp7.eu/

RESERVOIR: Consortium



RESERVOIR: SAP Use Cases and Requirements

- Creating a manifest of a service application
  - Images, contextualization scripts, DB content, all other configuration data
  - Means to contextualize and customize
- Provisioning a service application from a manifest
- Dynamic adjustment of resource allocation
  - Capacity planning
  - Automatic adaptive resource allocation / Self-optimization based on SLA and actual workload
- Elastic array of virtual execution environments
  - Dynamic scale-out by adding virtual servers to a service application
- Live migration
- . . .

- Next Steps & Discussion

Next Steps

1. Publish work group paper for public comments
   - Document type: “Informational”

2. Alignment with DMTF System Virtualization Working Group

3. Alignment with RESERVOIR