Distributed Computing With Data

John Chambers

Mark Hansen

David James

Duncan Temple Lang

May 18, 1998

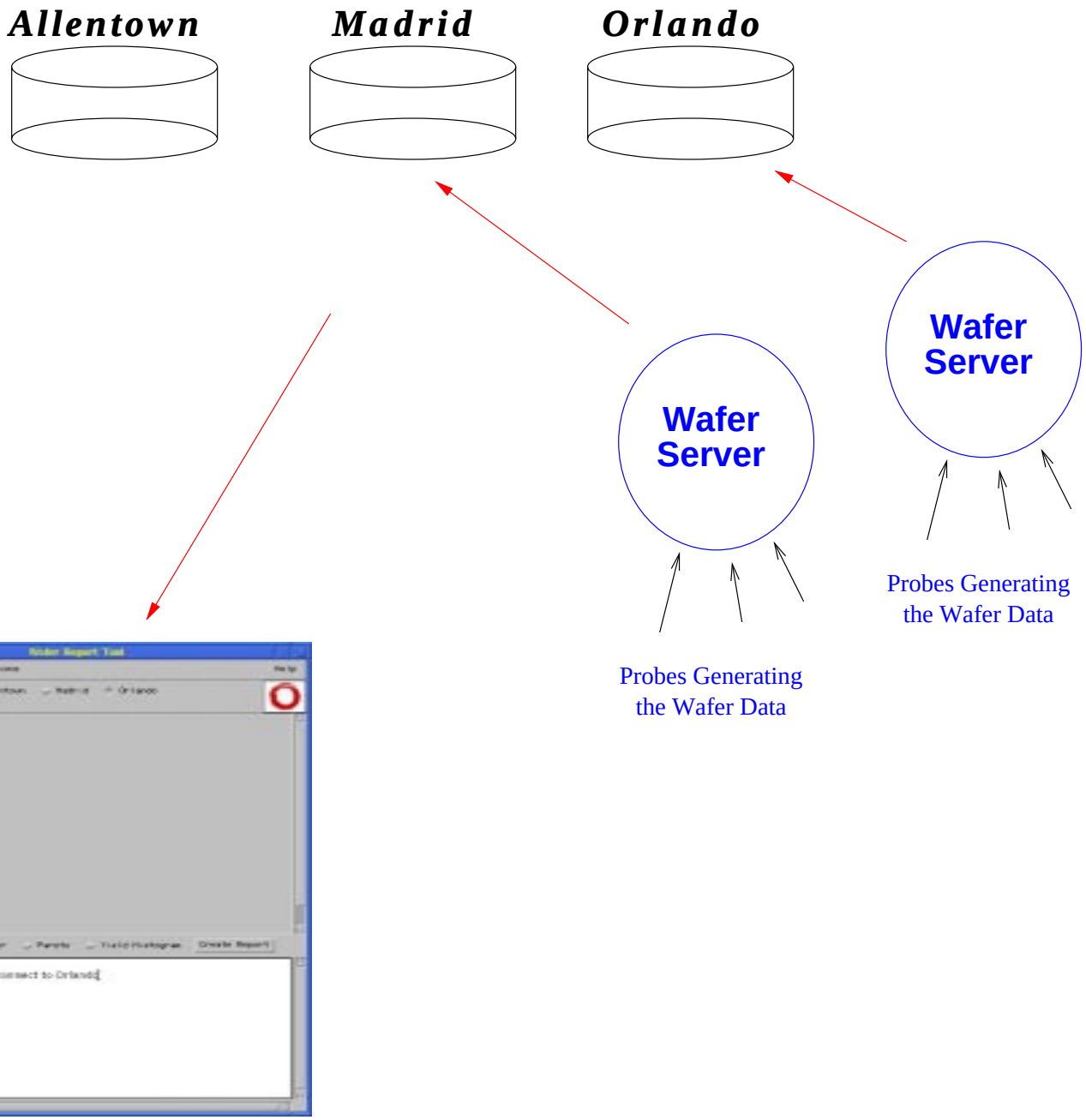
`cm.bell-labs.com/stat/duncan/DistComp/`

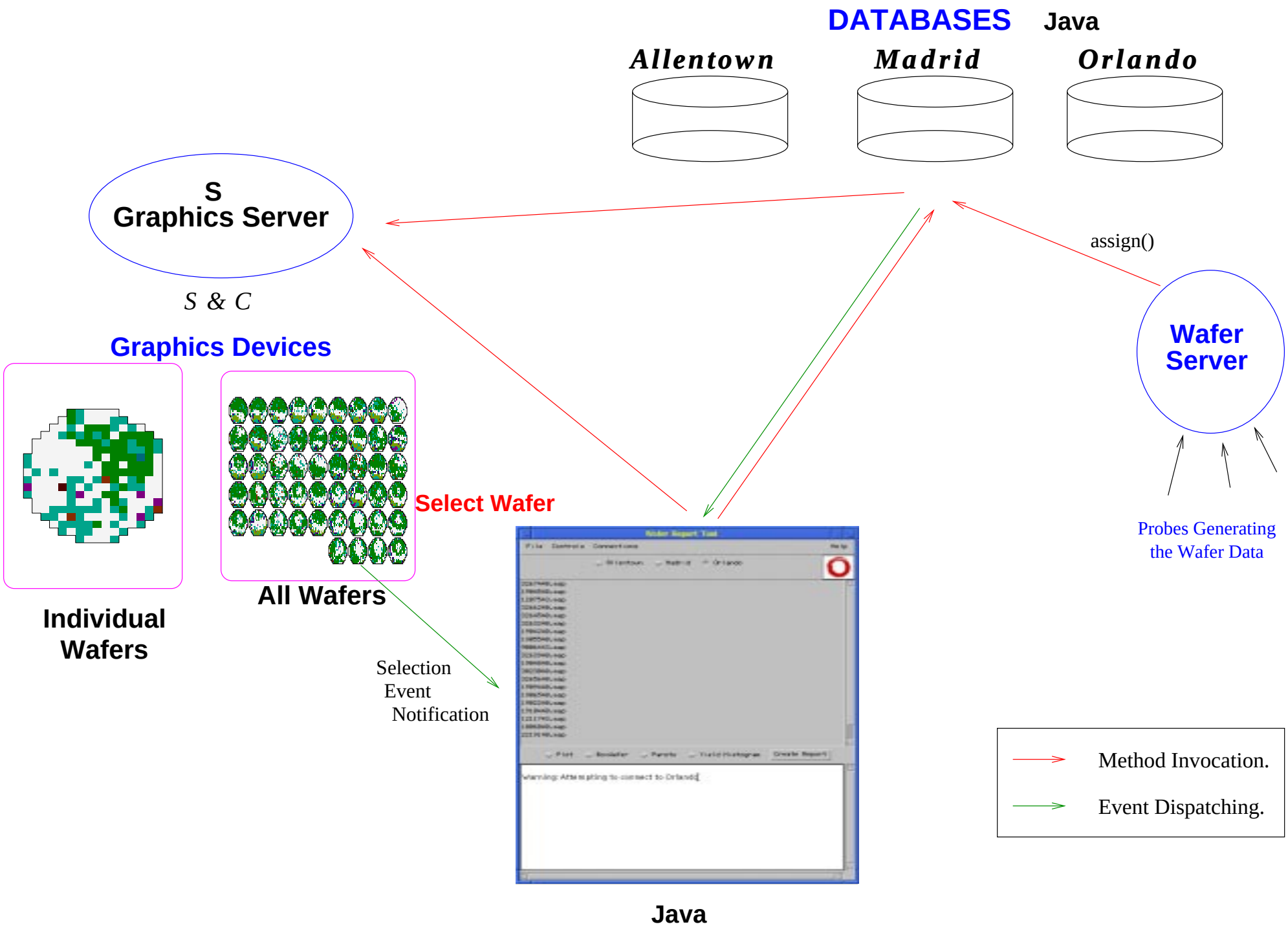


Outline

- Scenarios involving distributed computing.
- Requirements for a distributed environment.
- CORBA
- Next Generation Statistical Services.

DATABASES


Allentown
Madrid
Orlando
Wafer Server
Wafer Server
Probes Generating the Wafer Data
Probes Generating the Wafer Data


DATABASES
Java
Allentown
Madrid
Orlando
S
Graphics Server
S & C
assign()
Wafer
Server
Graphics Devices
Select Wafer
Individual
Wafers
All Wafers
Selection
Event
Notification
Probes Generating
the Wafer Data
Method Invocation.
Event Dispatching.
Java

Wafer Yield Reporting

- Local databases store wafer information by lot ID
- Device asynchronously sending data from processing to appropriate database
- End User on Web:

  Browse contents of different databases.

  Text summaries of one or more lots.

  Identify wafer of interest graphically - highlights in Report interface list.

  Plots of different statistics
  yield, composite maps, wafers, failure types.

Implementation Details

- New wafers announced to Report GUI and Graphics Server.
- Graphical selection events in Graphics Server.
- Data transferred directly between Databases and Graphics server.
- Computations for histogram - yields - done in specialized database.
  Communicated as vector of numbers.
- Avoid generating all static plots.

Examples

- Statistician

  Using functionality from different packages & languages.

  User manages communication.

  Link Fortran, C/C++, Java, Perl, scripts.

- Developer

  Distribute new methods.

  Offer cycles on machine.

  Link to user's environment.
  Multiple language bindings.

- Computing Resources and Performance

  Specialized hardware.

  Multi-tasking on multiple machines - "Parallel" Computing.

High Level Solutions

- Adding features to existing languages helps in the short term, but

  Expensive testing,

  Conflicts of design,

  Can never keep up-to-date.

- Avoid highly specialized individual applications with own formats.
  Use existing “standard” languages.

Is there a better way?

- Integrate access to features in existing languages.
- Need general communication between components.

Distributed Computing Environment Architecture

- Modular & Persistent components/servers/objects.
- Inter-component communication.
- Multi-directional and loop communication.
- Multi-lingual communication with support for common languages and extensibility.
- Structured Programming -
  - strong type checking in the server's own language.
  - Exceptions.
- Synchronous and asynchronous invocation - blocking and non-blocking.
- Support for different interfaces - GUIs, etc.
- Additional High level services
  security, naming, licensing, . . .
- Transparent Web access.

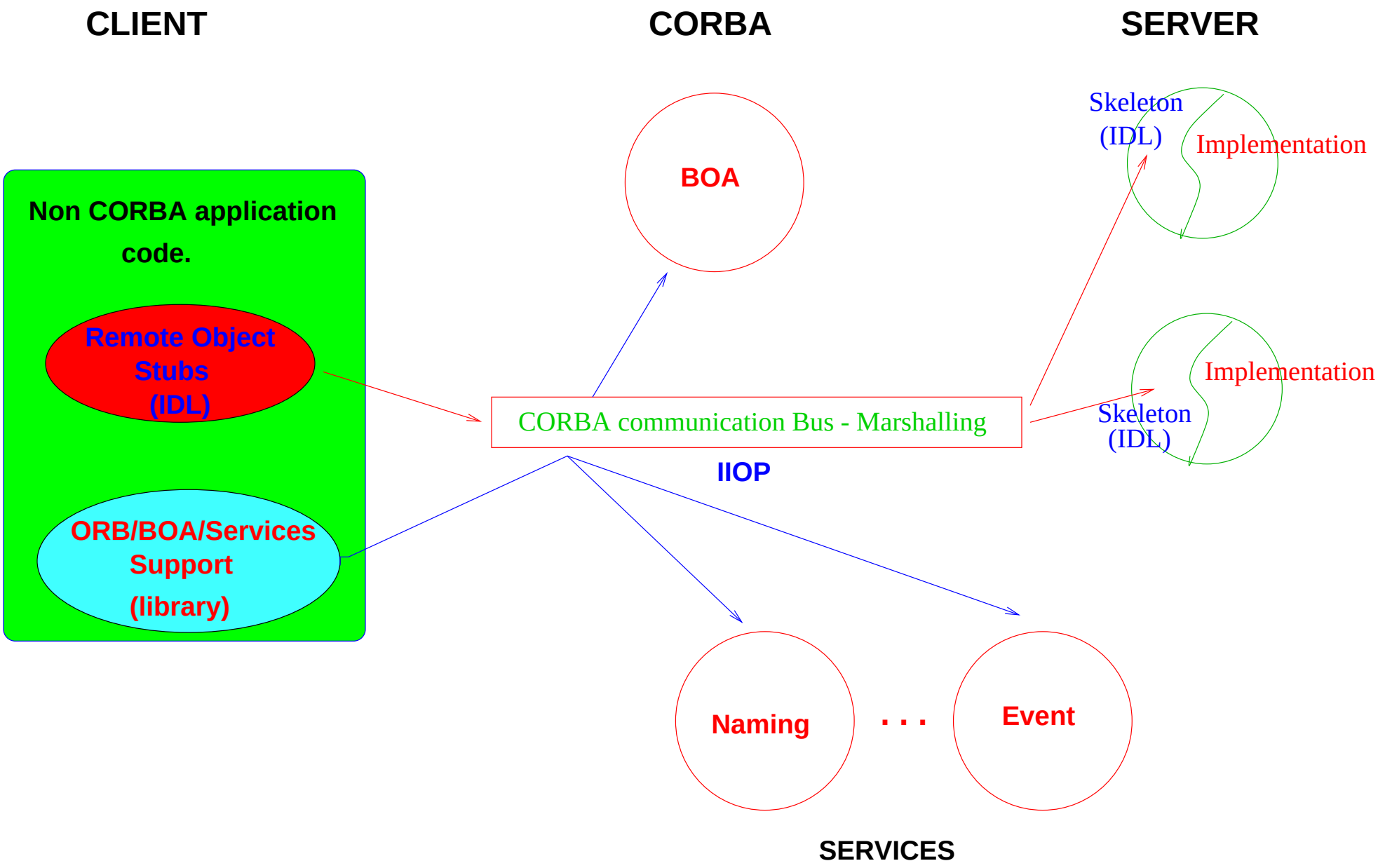
Details to Avoid

- Pairwise communication interface.
- Data format between machines.
- Fault tolerance.

Common Object Request Broker Architecture
CORBA

- Standard from Object Management Group (OMG).
- Implemented by numerous vendors commercial and non-commercial.
- Support for large number of operating systems.
- Multi-Domain support - not Statistics.
- Direct support from Java creators, Netscape,. . .
- Object Oriented.
- Separates implementation from description via IDL - Interface Definition Language.

CORBA Operation Details



Additional CORBA Services

- Dynamic Interface Repository.
- Naming Service - locate (and creating) named objects.
- Trading Service - locate objects satisfying some criteria.
- Event channels for asynchronous notification.
- Inter-ORB communication protocol
  different implementations communicate.
- Transaction Service.
- Concurrency Control.
- Security.
- Licensing Service.

Distributed Complexities

- Design of framework - properties & locations of objects.
- Data copying, consistency & synchronization.
- Task Management & Utilization of Resources.
- Debugging & Run-time Visualization.

Future Directions - Statistical Application

- Basic Classes
  - Develop primitive IDL Generic Data classes (e.g. vector, list, matrix, model, result,. . .) .
  - Share among different statistical systems.
- Objects as Servers
  - “Publish” data & methods from chosen language.
  - Tools to generate IDL

Future Directions - Statistical Application

- Dynamic Clients

  Discover remote objects and Interfaces.

  Provide automated high level methods to invoke remote methods.

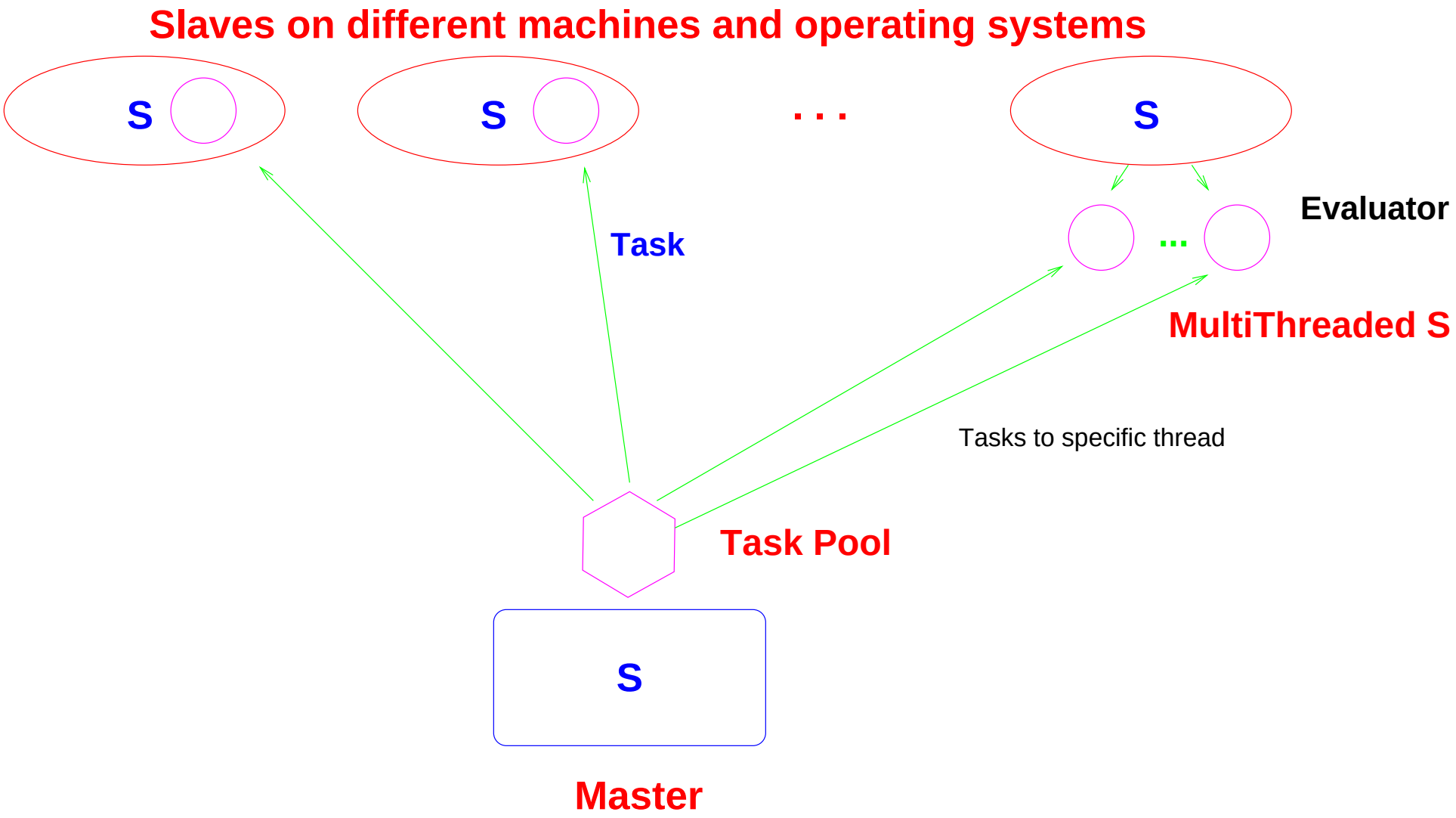
  Dynamically coerce local objects automatically via IDL descriptions recursive.

- High Performance Computing

  Automatic sub-task identification and synchronization.

  Scheduling & Fault Tolerance.

Distribution of S Tasks



Future Directions - Statistical Application

- Databases & SQL

  Expressions in statistical language (e.g. subsetting).

  Operations in Database (using SQL).

- Visualization Systems

  Tools from OpenGL, VRML, Java, etc.

  Linked independent components.

- GUI Development

  Interface in one language, actions other(s).

  Manage GUI objects remotely.

- Enhanced Thread Support
- Web-Based reporting
- On-line Dynamic Documentation
- Multi-User Systems

Summary

- CORBA provides foundation.
- Mechanism for transparently providing
  - Data & Service distribution and sharing.
  - High Performance Distributed Computing.
  - Integration of new & existing tools.
  - Web-based tools.
- Need support for common data types in different languages.
- Focus on new & high level problems.