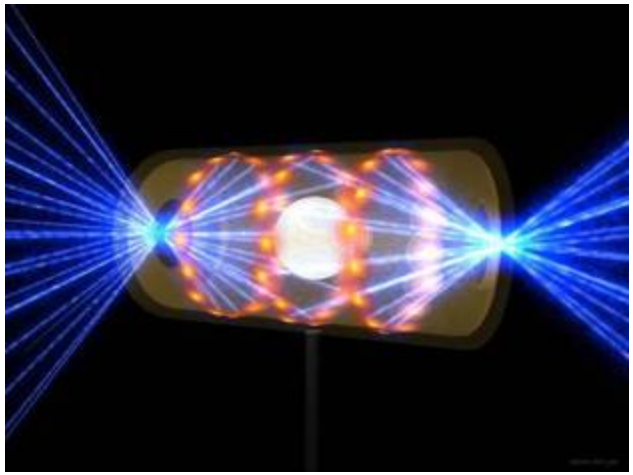
Processing Scientific Data: Three Unique Case Studies

Ronald Mak

Presented August 9, 2010
IBM Almaden Research Center



Presentation Agenda

- Case Study 1
  The Collaborative Information Portal (CIP)
  Mars Exploration Rovers (MER) Mission
  NASA Ames Research Center

- Case Study 2
  The Systems Health Information Portal (SHIP)
  NASA Ames Research Center

- Case Study 3
  Shot Data Management
  The National Ignition Facility (NIF)
  Lawrence Livermore National Laboratory

- Important Lessons Learned

Case Study 1:
The Collaborative Information Portal (CIP)

Case Study 1: The Collaborative Information Portal

The **Collaborative Information Portal (CIP)** was a key ground-based application used by the NASA's Mars Exploration Rovers (MER) mission.

Mission scientists, engineers, and researchers at JPL and around the world used CIP to access mission data in a secure and organized manner over the Internet.

Mission Control Center, JPL
Mars Exploration Rovers Mission

- My role: Senior Scientist Research Institute for Advanced Computer Science (RIACS)
  - Architect and lead developer of the CIP middleware, 2002-2003
  - Mission support at the NASA Ames Research Center and JPL, 2004

Mars Exploration Rover Mission

- Twin robot geologists search for liquid water on Mars in the past.

- **Launched:** June 10 and July 7, 2003
- **Landed:** January 3 and 24, 2004
- **Duration:** 90 days

  But after over six and a half Earth years, both are still operating (although one is has been stuck in a sand pile for several months).

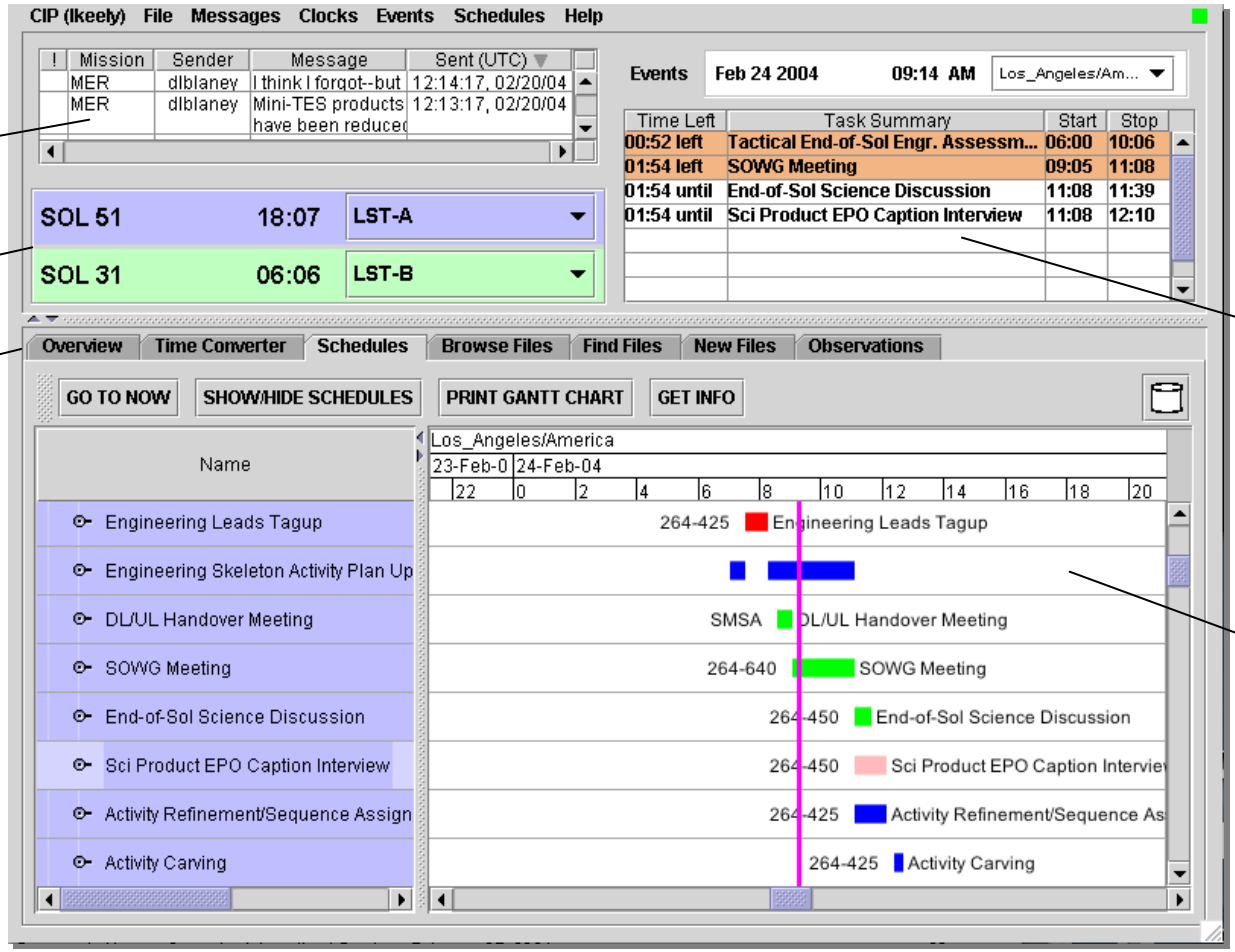
- **Mission Center:**
  Jet Propulsion Laboratory
  Pasadena, CA

MER Mission Requirements

- Time management
- Data management
- Personnel management

Mission Geology Lab, JPL

The Collaborative Information Portal (CIP)



Broadcast messages

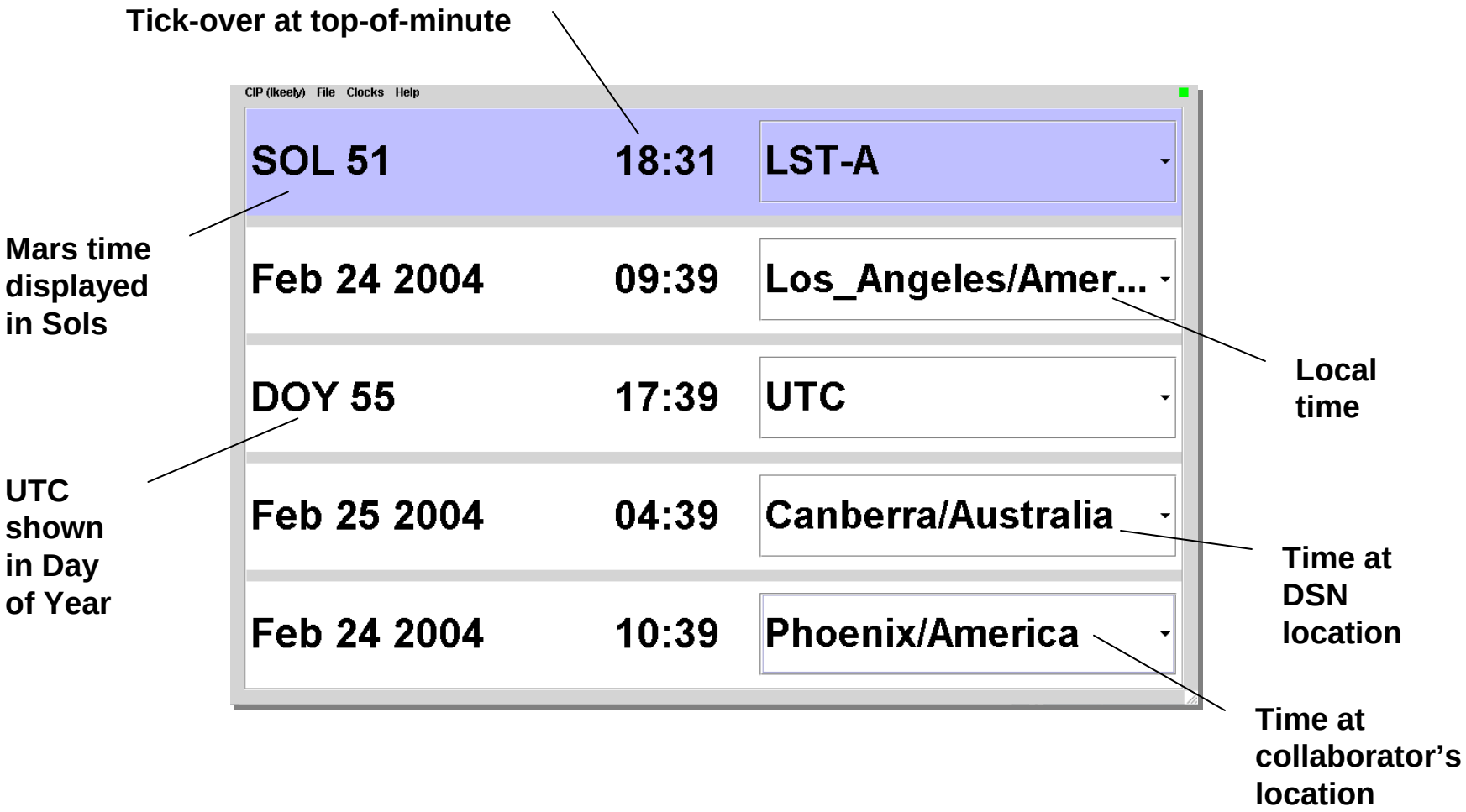
Clocks

Tool tabs

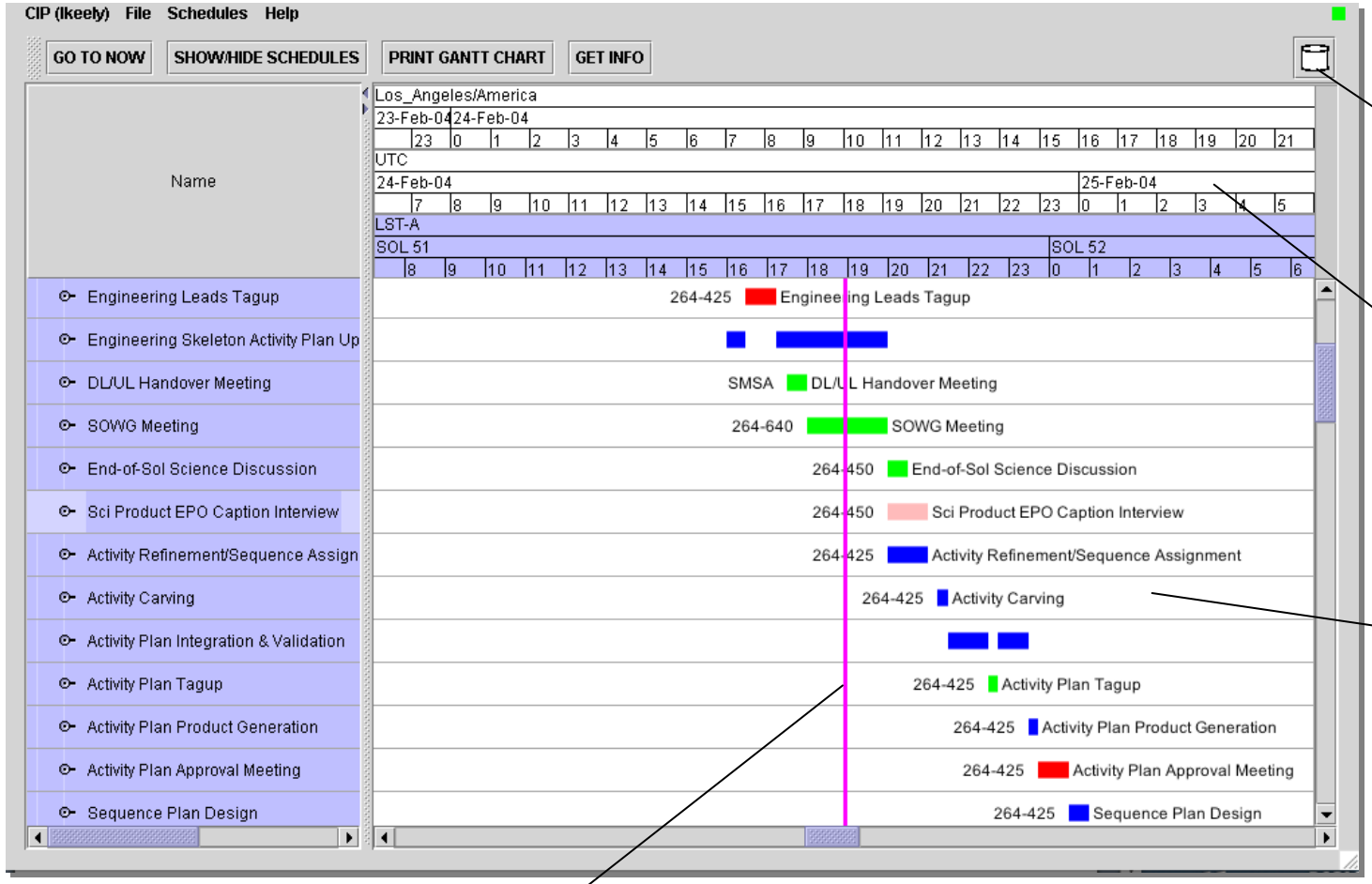
Event horizon

Schedule viewer

Clocks


Tick-over at top-of-minute
CIP (lkeely) File Clocks Help
SOL 51
18:31
LST-A
Feb 24 2004
09:39
Los_Angeles/Amer...
DOY 55
17:39
UTC
Feb 25 2004
04:39
Canberra/Australia
Feb 24 2004
10:39
Phoenix/America
Mars time displayed in Sols
UTC shown in Day of Year
Local time
Time at DSN location
Time at collaborator's location

Schedule Viewer



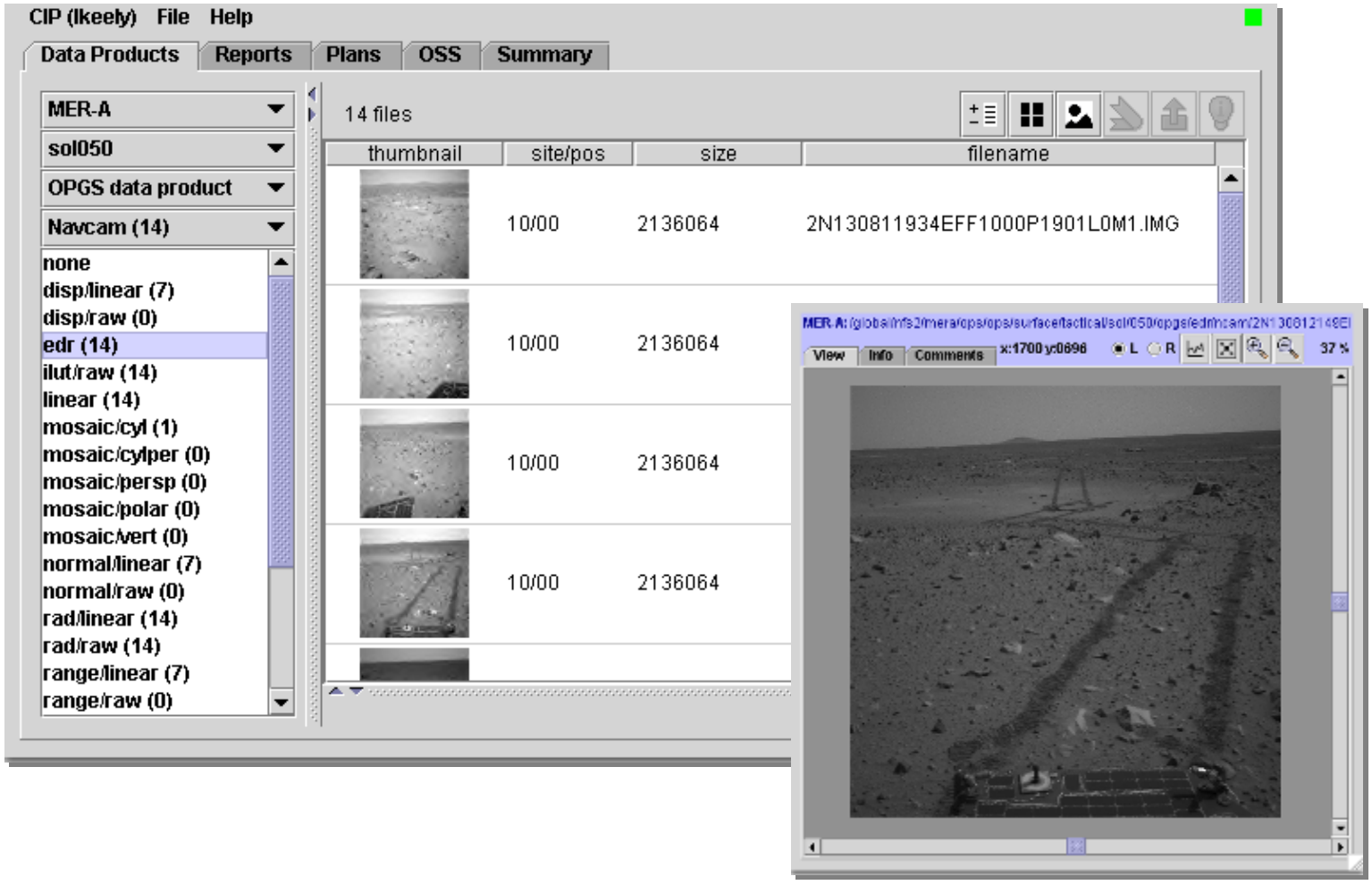
Schedule update indicator

Multiple time scales

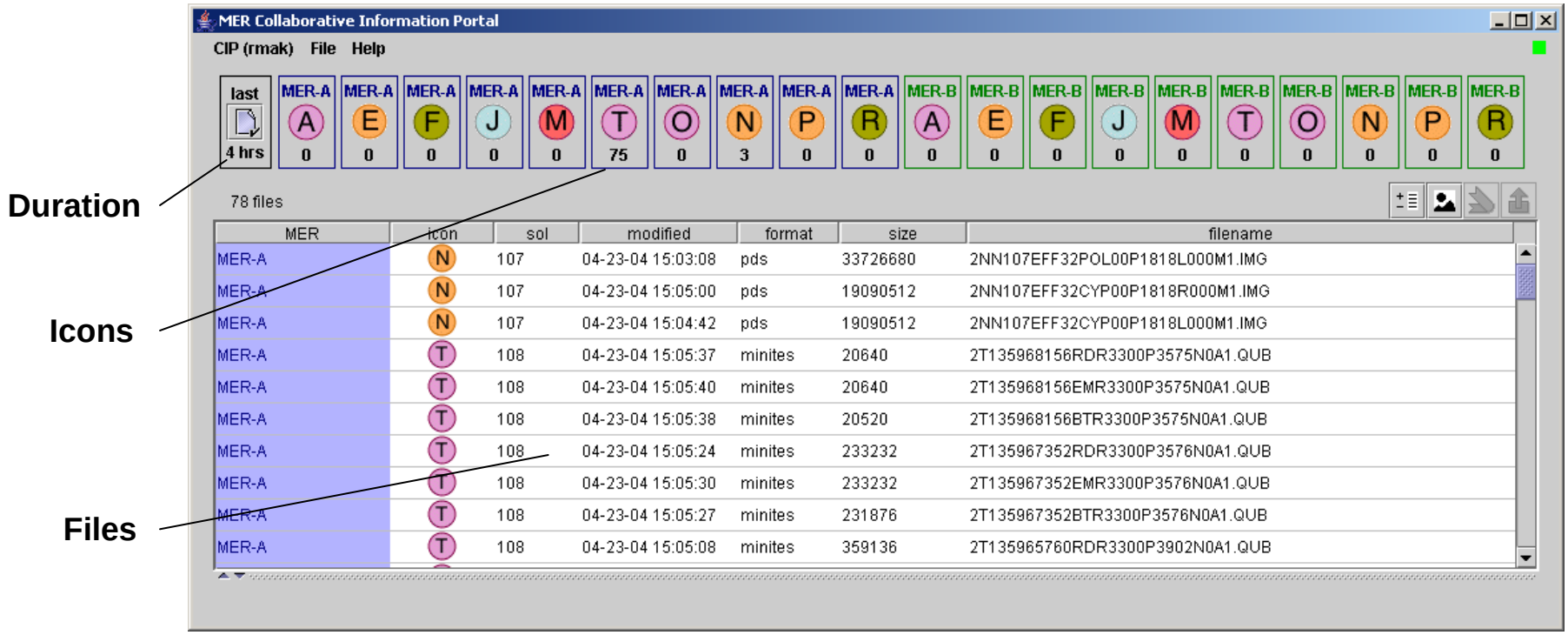
Unified schedules

“Now” bar

Data Product Navigator



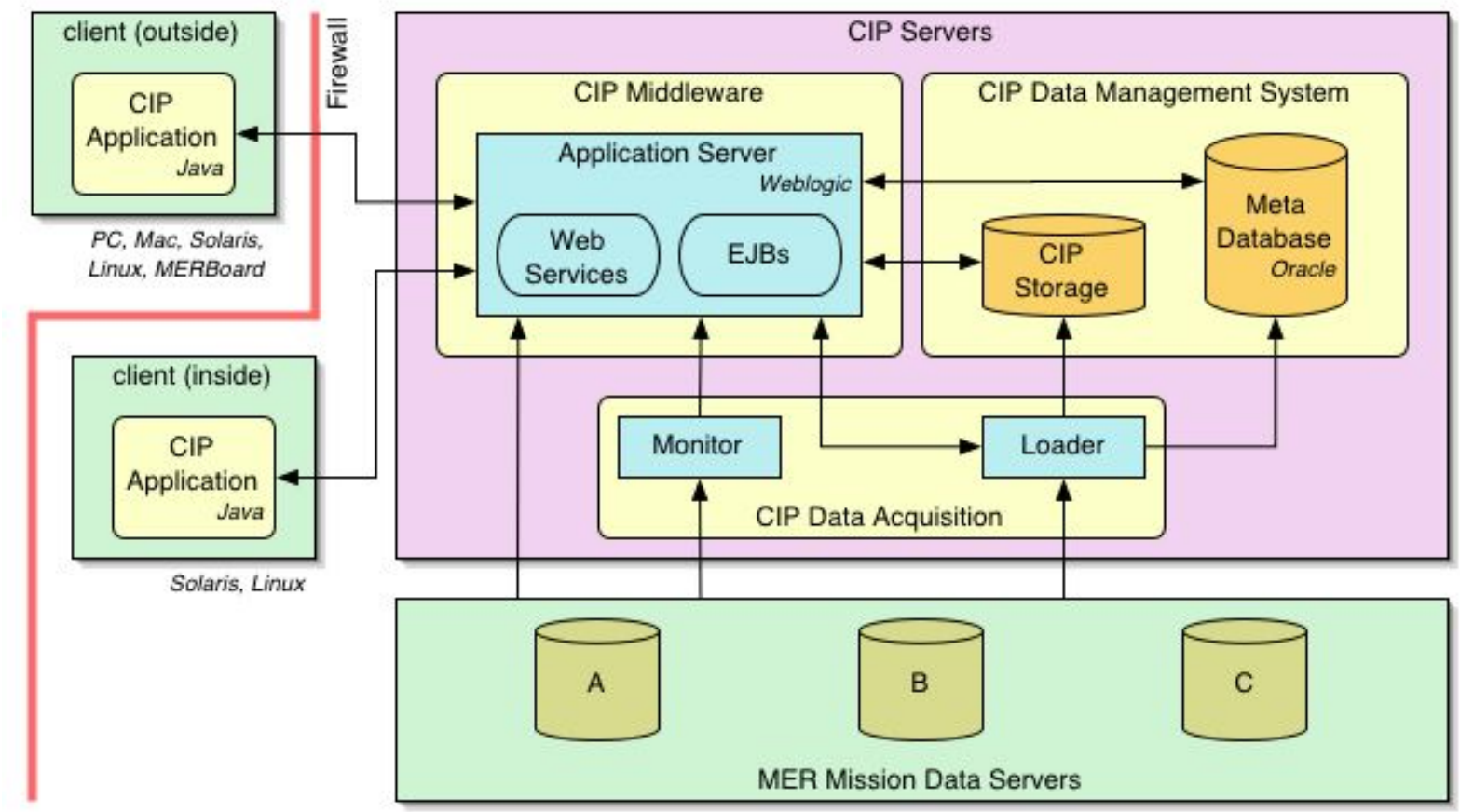
New File Notification



Three–Tiered Enterprise Architecture

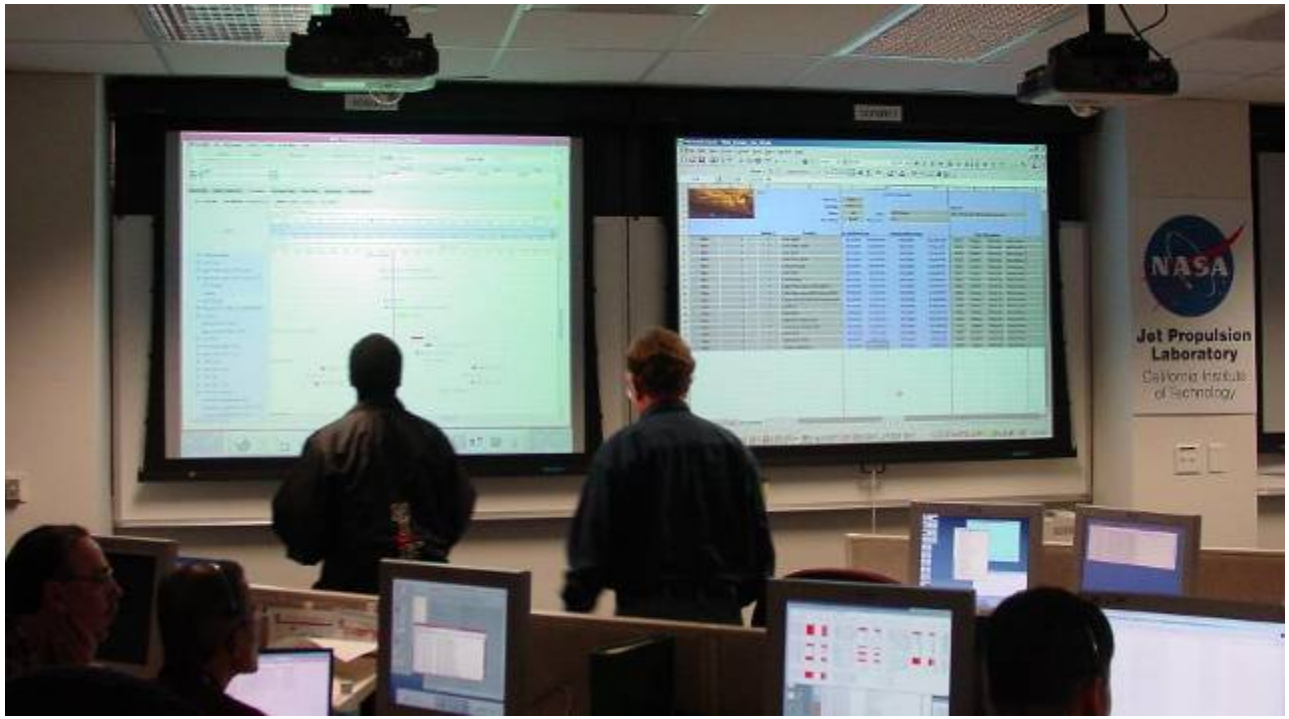
- Client
  - Java application (Swing)
- Middleware Server
  - Web services, Enterprise JavaBeans (EJB), Java Message Service (JMS)
  - Service-oriented architecture (SOA)
- Data Repository
  - Mission file servers (Unix)
  - File monitor (Java application)
  - Data loader (Java application)
  - Database (Oracle)

Architecture Overview



CIP Middleware Services

- User management
- Metadata
- Schedules
- Mars and Earth time
- File and directory
- Message

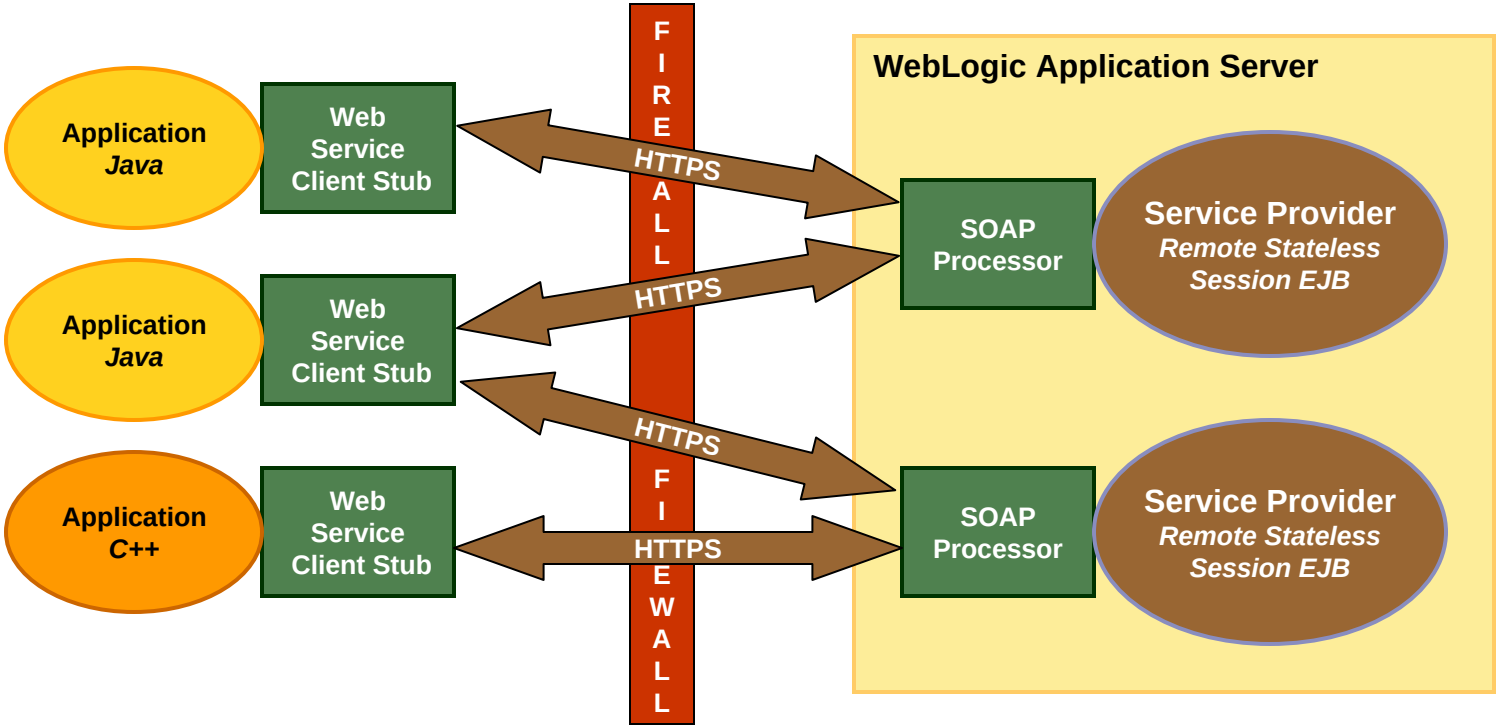


Mission Control Center, JPL

Middleware Technologies

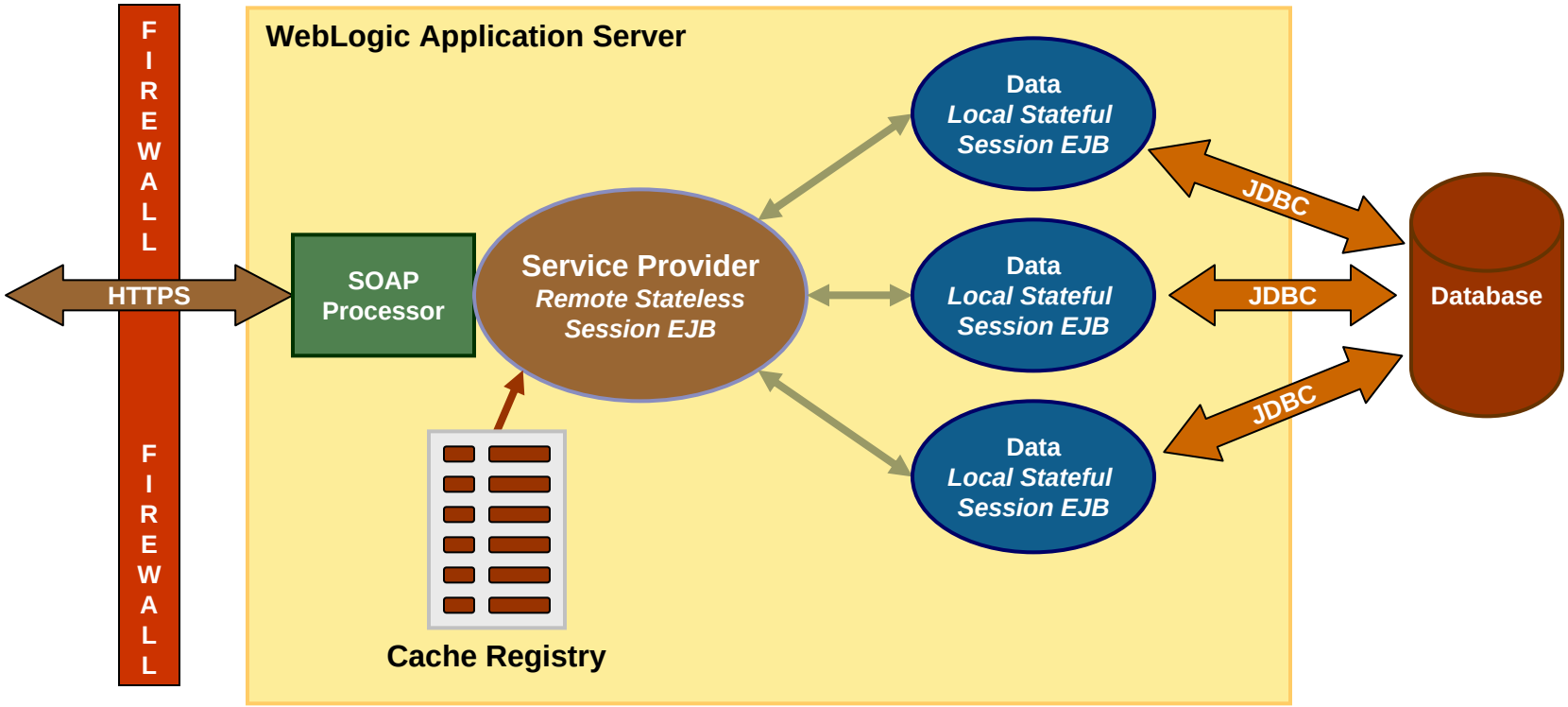
- Enterprise JavaBeans (EJBs) to achieve reliability, scalabilty, security, platform independence, and standards.
  - Stateless session beans
  - Stateful session beans
  - Message–driven beans
- Web services to expose the remote methods of the Service Provider EJBs to the client applications.
- Java Message Service (JMS) for synchronous and asynchronous messaging.
- BEA WebLogic application server.

Middleware Architecture



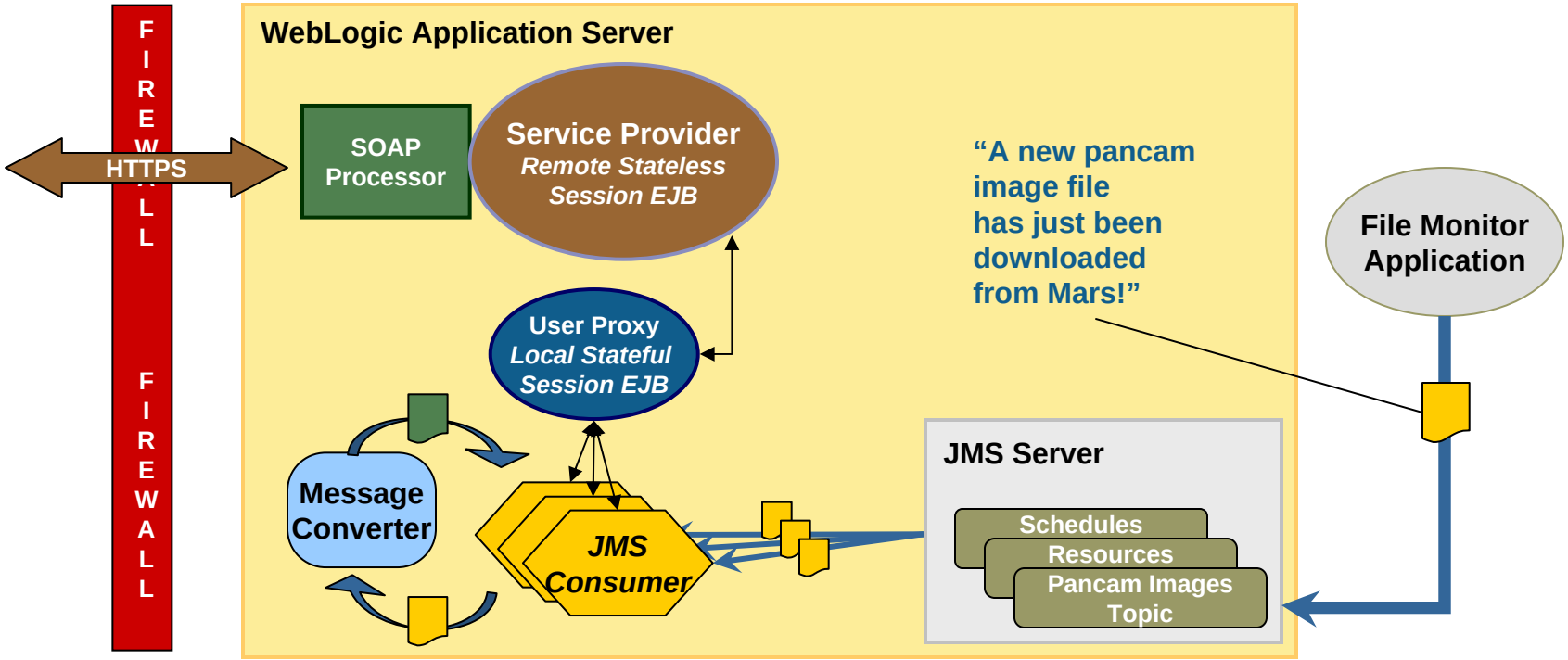
- Web Services expose the remote methods of the Service Provider beans.
- HTTPS encrypts the transmissions and gets them through the firewall.
- Stateless session beans are the Service Providers.

Caching Data



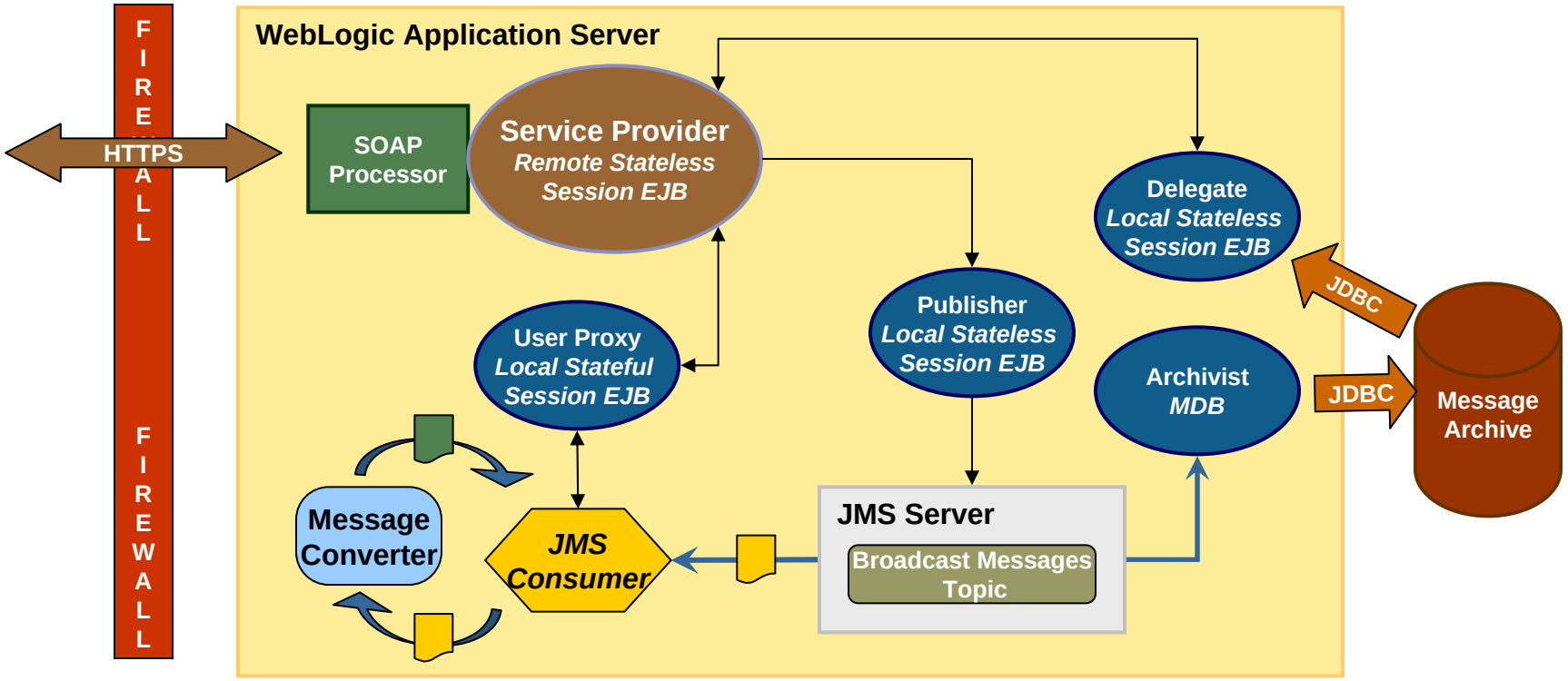
- Some middleware services cache data to improve performance.
- Stateful session beans cache data in memory.
- The cache beans access the backend data store.

Event Notification



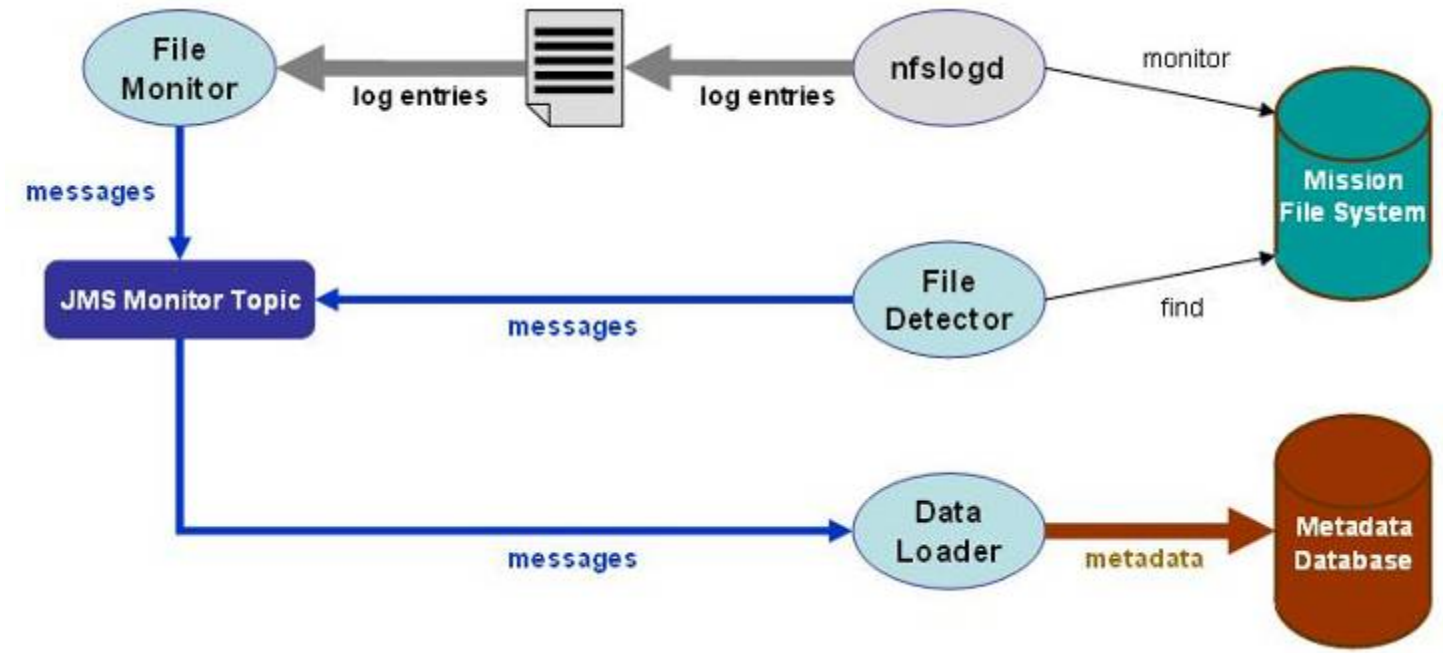
- Publish-subscribe model with topics
- Clients poll for their messages via web services.
- The consumer objects retrieve the messages.
- User proxies look up JMS consumer objects that represent the active subscriptions.
- Messages are converted for return via web services.

Broadcast Message Service



- The Publisher bean publishes messages to the Broadcast Messages topic.
- Message–driven bean (MDB) archives broadcast messages to persistent storage for later retrieval.
- The Delegate bean retrieves old messages for browsing.

Metadata Generation



- File changes are logged by nfslogd system program and the File Monitor, or found by the File Detector.
  - A message is sent to the JMS Monitor Topic for each file change.
- The Data Loader generates metadata about the file.
  - Retrieves messages from the monitor topic.

How Reliable Has CIP Been?

- The middleware code remained unchanged since the code freeze in early November 2003.
- CIP became operational in mid–December 2003, two weeks before the first rover landed.
- CIP has run continuously for as long as 153 Earth days without interruption.
  - As of the end of the day on December 31, 2005, the mission had lasted 17,904 Earth hours.
  - During that entire period, CIP has been down less than 10 hours, which is > 99.9% uptime.
- CIP was finally shut down this summer after 6½ years of failure-free service.

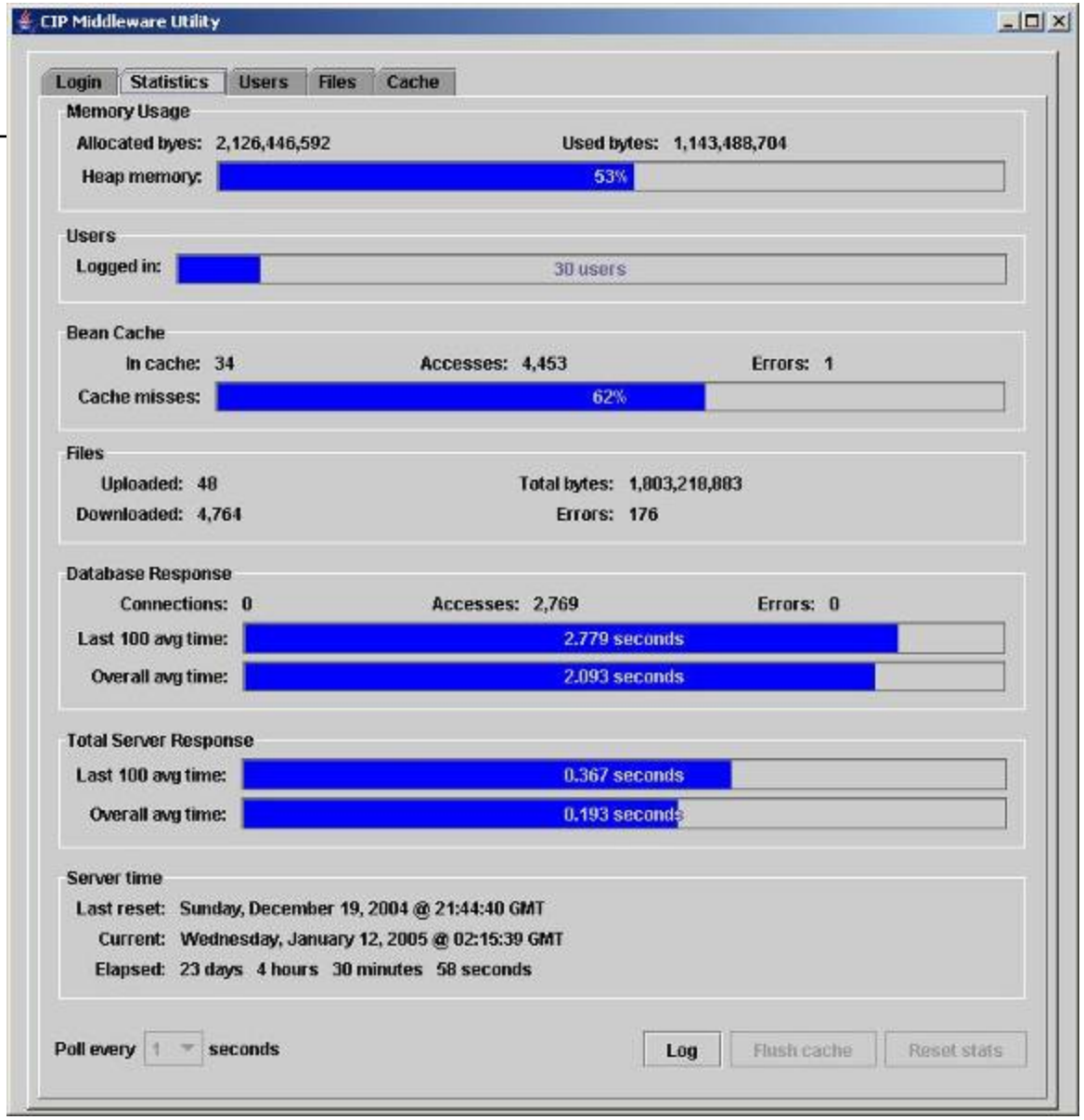
Example Log Entries: Streamer Service

```
2004-12-20 20:23:33,435 INFO : mjane: Streamer.putDataFile(/opt/bea/user_projects/cip/
conf/preferences/m.preferences)
2004-12-20 20:23:33,439 DEBUG: Begin upload of file '/opt/bea/user_projects/cip/conf/
preferences/sthompso.preferences'
2004-12-20 20:23:39,140 DEBUG: Completed upload of file '/opt/bea/user_projects/cip/conf/
preferences/sthompso.preferences': 35659 bytes
'/opt/bea/user_projects/cip/conf/global.properties': 13453 bytes
2004-12-20 20:28:57,516 INFO : jdoe: Streamer.getPreferences(user)
2004-12-20 20:29:29,721 INFO : jdoe: Streamer.getFileInfo()
2004-12-20 20:29:30,784 INFO : jdoe: Streamer.getFileInfo(/oss/merb/ops/ops/surface/
tactical/sol)
2004-12-21 19:17:43,320 INFO : jqpublic: Streamer.getDataFile(/global/nfs2/merb/ops/ops/
surface/tactical/sol/120/opgs/edr/jpeg/1P138831013ETH2809P2845L2M1.JPG)
2004-12-21 19:17:43,324 DEBUG: Begin download of file '/global/nfs2/merb/ops/ops/surface/
tactical/sol/120/opgs/edr/jpeg/1P138831013ETH2809P2845L2M1JPG'
2004-12-21 19:17:44,584 DEBUG: Completed download of file '/global/nfs2/merb/ops/ops/
surface/tactical/sol/120/opgs/edr/jpeg/1P138831013ETH2809P2845L2M1.JPG': 1876 bytes
```

- Timestamp, user name, method name
- Information about the uploaded or downloaded file

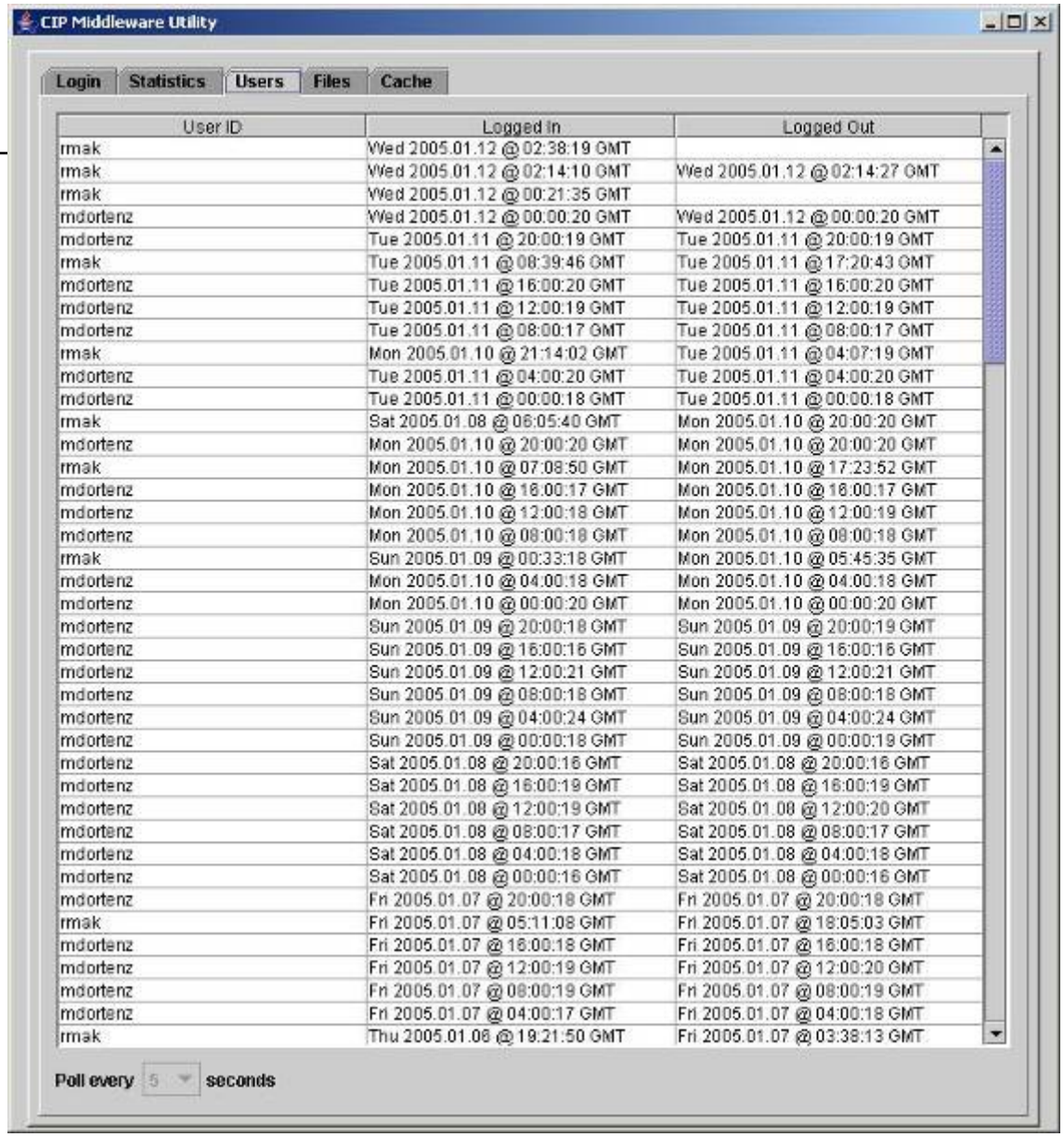
Middleware Monitor Utility

- Server statistics



Middleware Monitor Utility

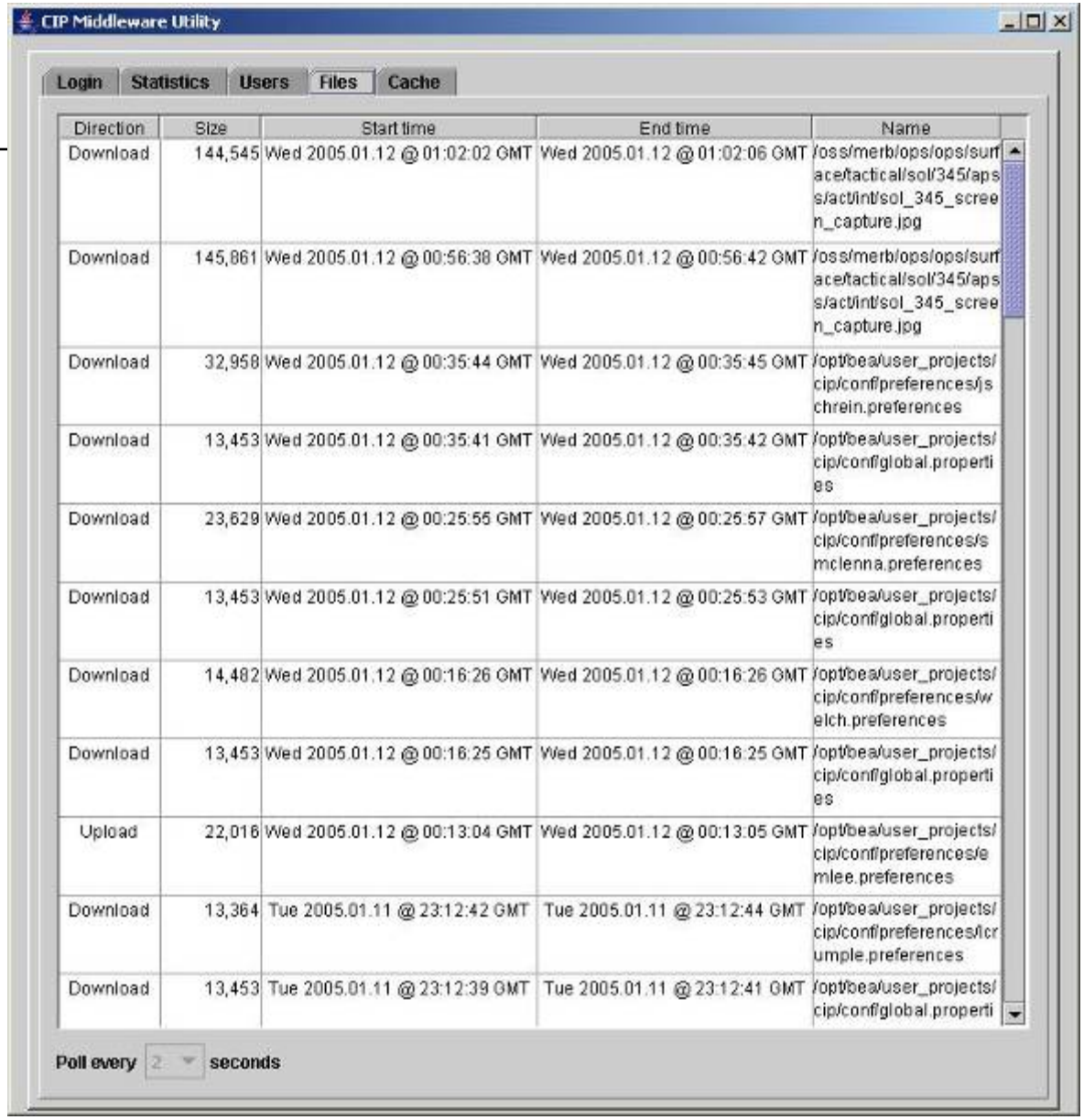
- Users



User ID	Logged In	Logged Out
rmak	Wed 2005.01.12 @ 02:38:19 GMT	
rmak	Wed 2005.01.12 @ 02:14:10 GMT	Wed 2005.01.12 @ 02:14:27 GMT
rmak	Wed 2005.01.12 @ 00:21:35 GMT	
mdortenz	Wed 2005.01.12 @ 00:00:20 GMT	Wed 2005.01.12 @ 00:00:20 GMT
mdortenz	Tue 2005.01.11 @ 20:00:19 GMT	Tue 2005.01.11 @ 20:00:19 GMT
rmak	Tue 2005.01.11 @ 08:39:46 GMT	Tue 2005.01.11 @ 17:20:43 GMT
mdortenz	Tue 2005.01.11 @ 16:00:20 GMT	Tue 2005.01.11 @ 16:00:20 GMT
mdortenz	Tue 2005.01.11 @ 12:00:19 GMT	Tue 2005.01.11 @ 12:00:19 GMT
mdortenz	Tue 2005.01.11 @ 08:00:17 GMT	Tue 2005.01.11 @ 08:00:17 GMT
rmak	Mon 2005.01.10 @ 21:14:02 GMT	Tue 2005.01.11 @ 04:07:19 GMT
mdortenz	Tue 2005.01.11 @ 04:00:20 GMT	Tue 2005.01.11 @ 04:00:20 GMT
mdortenz	Tue 2005.01.11 @ 00:00:18 GMT	Tue 2005.01.11 @ 00:00:18 GMT
rmak	Sat 2005.01.08 @ 06:05:40 GMT	Mon 2005.01.10 @ 20:00:20 GMT
mdortenz	Mon 2005.01.10 @ 20:00:20 GMT	Mon 2005.01.10 @ 20:00:20 GMT
rmak	Mon 2005.01.10 @ 07:08:50 GMT	Mon 2005.01.10 @ 17:23:52 GMT
mdortenz	Mon 2005.01.10 @ 16:00:17 GMT	Mon 2005.01.10 @ 16:00:17 GMT
mdortenz	Mon 2005.01.10 @ 12:00:18 GMT	Mon 2005.01.10 @ 12:00:19 GMT
mdortenz	Mon 2005.01.10 @ 08:00:18 GMT	Mon 2005.01.10 @ 08:00:18 GMT
rmak	Sun 2005.01.09 @ 00:33:18 GMT	Mon 2005.01.10 @ 05:45:35 GMT
mdortenz	Mon 2005.01.10 @ 04:00:18 GMT	Mon 2005.01.10 @ 04:00:18 GMT
mdortenz	Mon 2005.01.10 @ 00:00:20 GMT	Mon 2005.01.10 @ 00:00:20 GMT
mdortenz	Sun 2005.01.09 @ 20:00:18 GMT	Sun 2005.01.09 @ 20:00:19 GMT
mdortenz	Sun 2005.01.09 @ 16:00:16 GMT	Sun 2005.01.09 @ 16:00:16 GMT
mdortenz	Sun 2005.01.09 @ 12:00:21 GMT	Sun 2005.01.09 @ 12:00:21 GMT
mdortenz	Sun 2005.01.09 @ 08:00:18 GMT	Sun 2005.01.09 @ 08:00:18 GMT
mdortenz	Sun 2005.01.09 @ 04:00:24 GMT	Sun 2005.01.09 @ 04:00:24 GMT
mdortenz	Sun 2005.01.09 @ 00:00:18 GMT	Sun 2005.01.09 @ 00:00:19 GMT
mdortenz	Sat 2005.01.08 @ 20:00:16 GMT	Sat 2005.01.08 @ 20:00:16 GMT
mdortenz	Sat 2005.01.08 @ 16:00:19 GMT	Sat 2005.01.08 @ 16:00:19 GMT
mdortenz	Sat 2005.01.08 @ 12:00:19 GMT	Sat 2005.01.08 @ 12:00:20 GMT
mdortenz	Sat 2005.01.08 @ 08:00:17 GMT	Sat 2005.01.08 @ 08:00:17 GMT
mdortenz	Sat 2005.01.08 @ 04:00:18 GMT	Sat 2005.01.08 @ 04:00:18 GMT
mdortenz	Sat 2005.01.08 @ 00:00:16 GMT	Sat 2005.01.08 @ 00:00:16 GMT
mdortenz	Fri 2005.01.07 @ 20:00:18 GMT	Fri 2005.01.07 @ 20:00:18 GMT
rmak	Fri 2005.01.07 @ 05:11:08 GMT	Fri 2005.01.07 @ 18:05:03 GMT
mdortenz	Fri 2005.01.07 @ 16:00:18 GMT	Fri 2005.01.07 @ 16:00:18 GMT
mdortenz	Fri 2005.01.07 @ 12:00:19 GMT	Fri 2005.01.07 @ 12:00:20 GMT
mdortenz	Fri 2005.01.07 @ 08:00:19 GMT	Fri 2005.01.07 @ 08:00:19 GMT
mdortenz	Fri 2005.01.07 @ 04:00:17 GMT	Fri 2005.01.07 @ 04:00:18 GMT
rmak	Thu 2005.01.06 @ 19:21:50 GMT	Fri 2005.01.07 @ 03:38:13 GMT

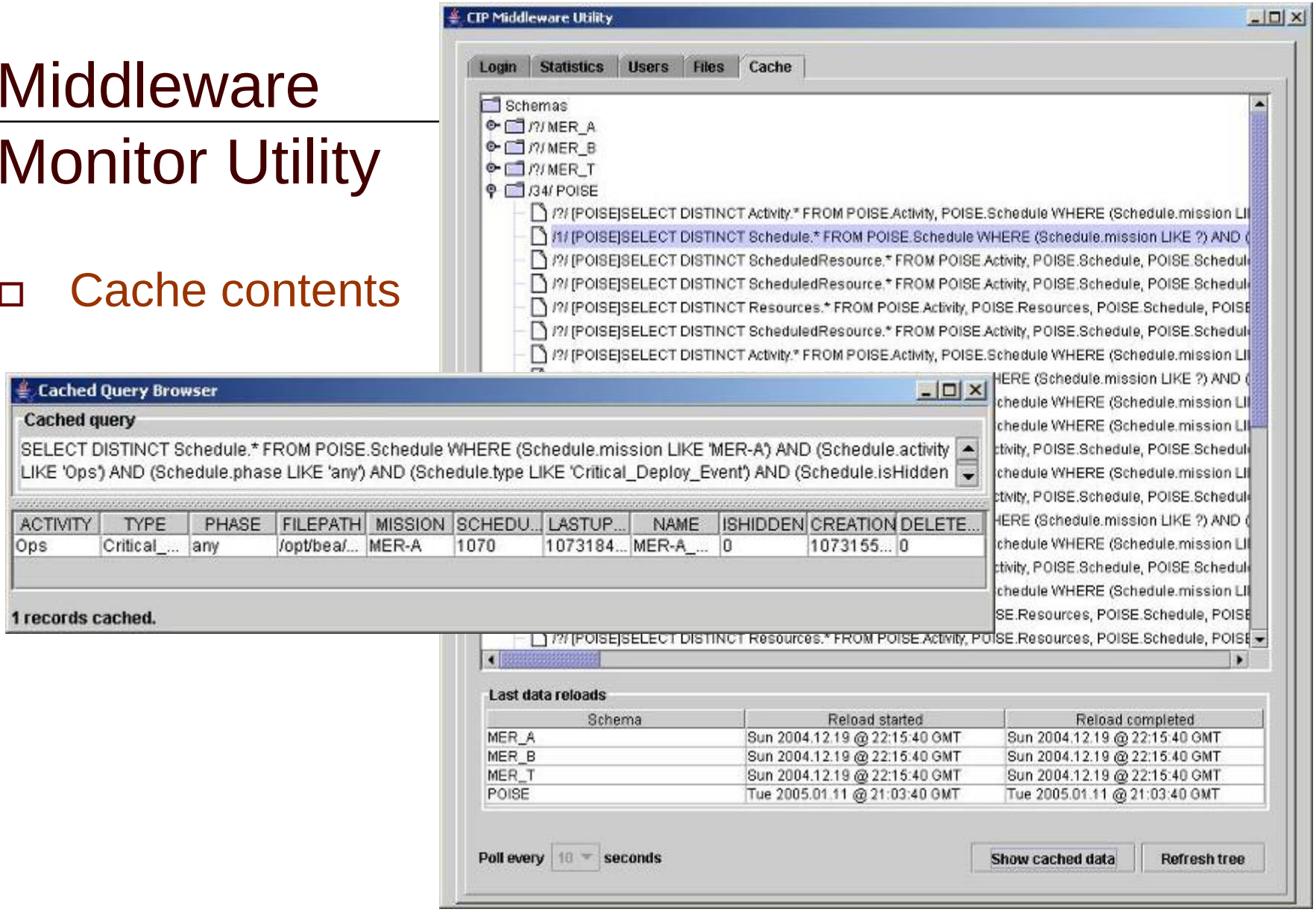
Middleware Monitor Utility

- Files uploaded or downloaded



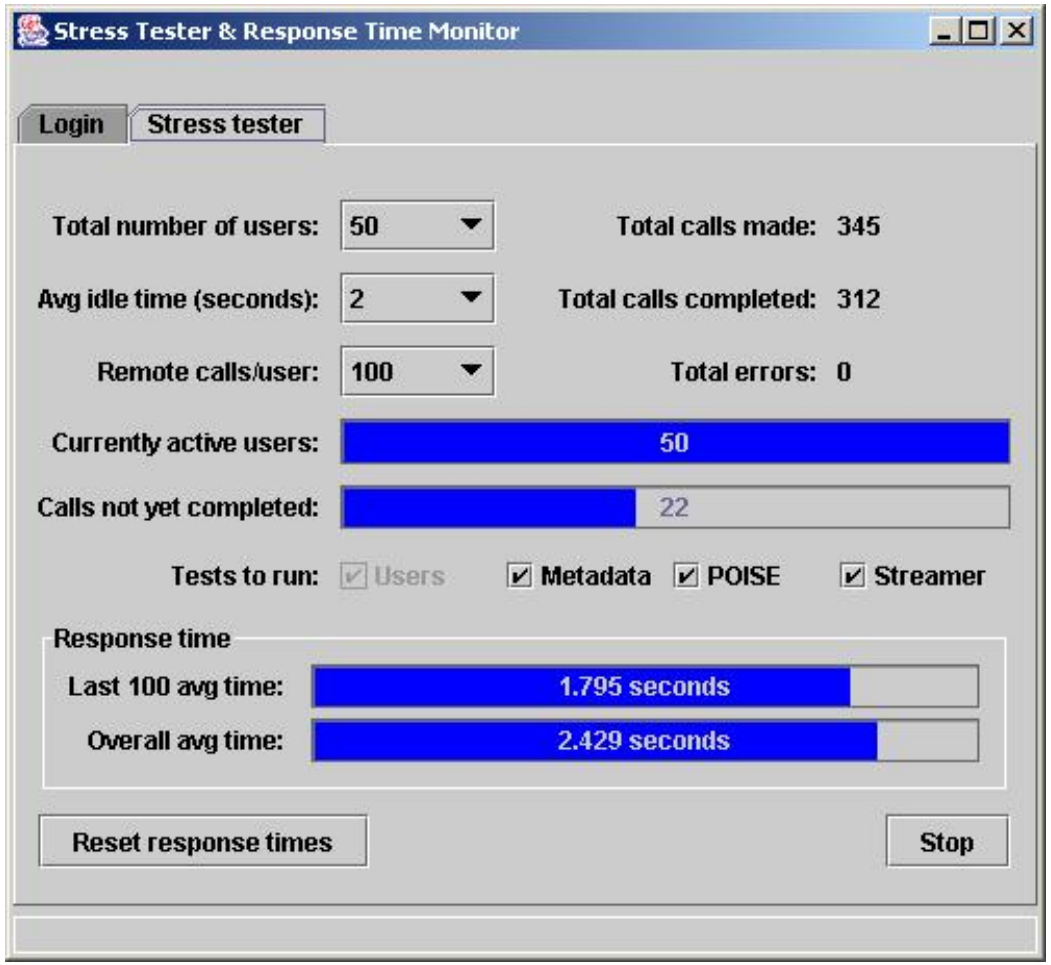
Middleware Monitor Utility

- Cache contents



Middleware Stress Tester

- Simulate multiple users performing client operations simultaneously.
- Can the server handle a heavy load?
  - What is its behavior when it starts to fail?

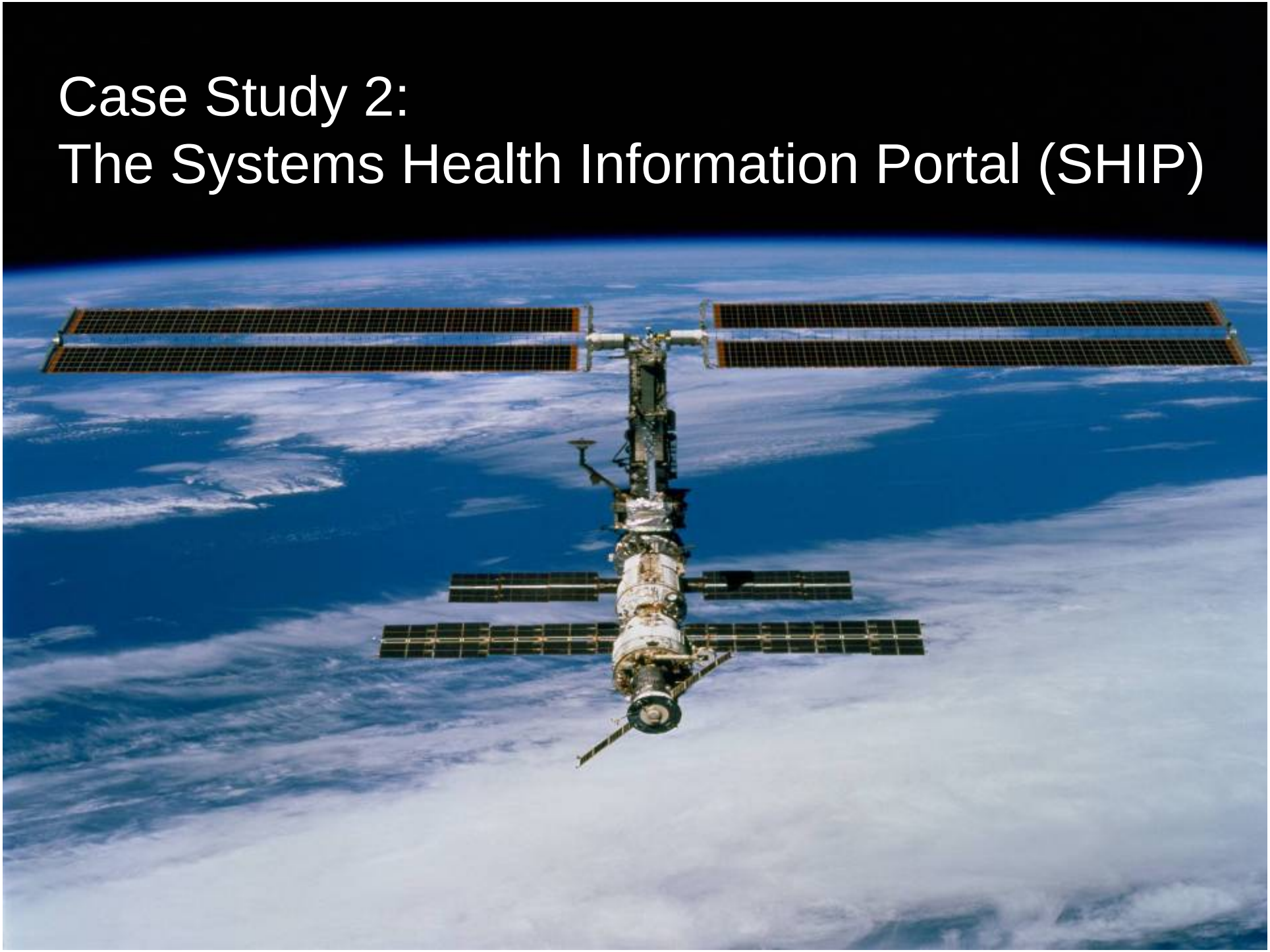


CIP Summary

- Provided situational awareness for MER mission personnel
  - Download images, data, schedules, current time
  - Collaboration
  - Alerts
- Service-oriented architecture (SOA)
  - Java Swing and C++ desktop client applications
  - J2EE technologies
  - Web services
  - Mission file servers
  - Metadatabase
  - BEA WebLogic, Oracle RDBMS
  - Open source software

Wind-sculpted sand patterns on Mars

Case Study 2:
The Systems Health Information Portal (SHIP)



Case Study 2: The Systems Health Information Portal

The **Systems Health Information Portal (SHIP)** constantly monitors and analyzes sensor data gathered from a manned NASA space vehicle.

If there is a fault with the vehicle, SHIP quickly analyzes the situation and recommends corrective actions for the astronauts on board, even if contact with ground control is lost.

Life on board the International Space Station

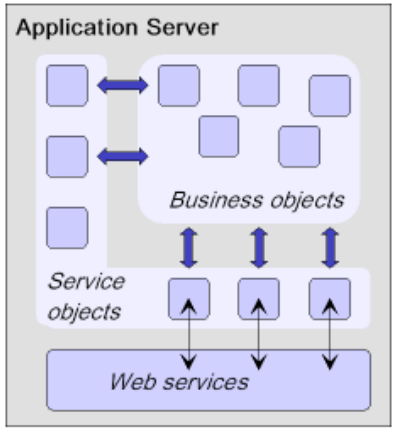
- My role: Project Scientist
  University of California at Santa Cruz
  - Architect, lead developer, and systems integrator, 2004-2005

SHIP Requirements

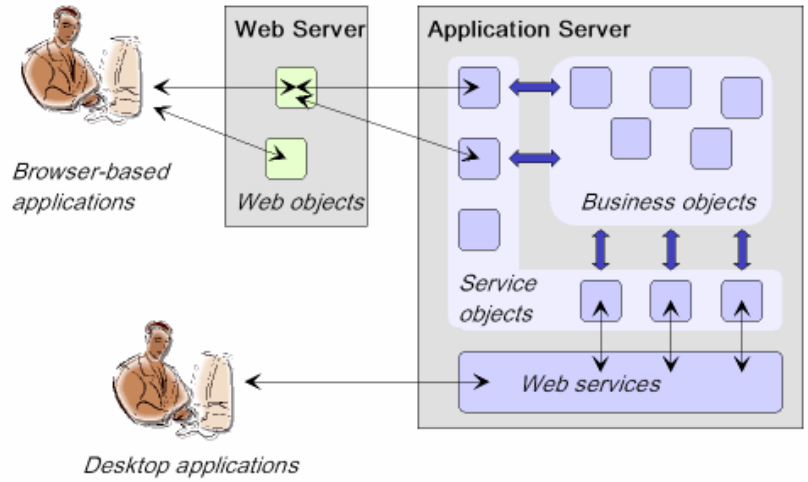
- Access to disparate data sources
  - Databases, files, live instrument streams, web services, web pages, programs
- Multiple data formats
  - Relational tables
  - XML, text, and binary files
  - Proprietary and legacy data formats
- Generate and manage reports
  - Fault analyses
  - Prognostications
  - Procedure manuals
- Test bed
  - International Space Station (ISS)

Rendezvous in low Earth orbit (LEO) of a manned space capsule with the main rocket engines in preparation for a trip to the Moon or to Mars.

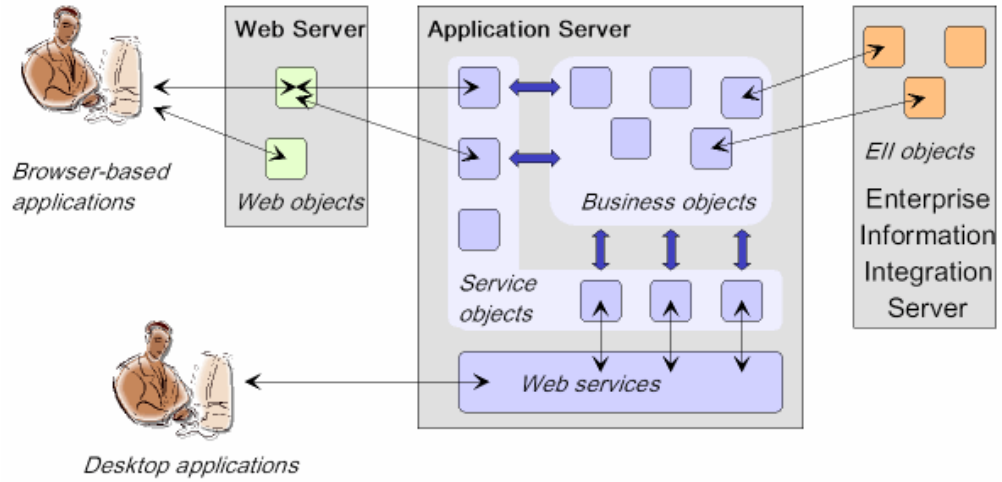
SHIP Architecture



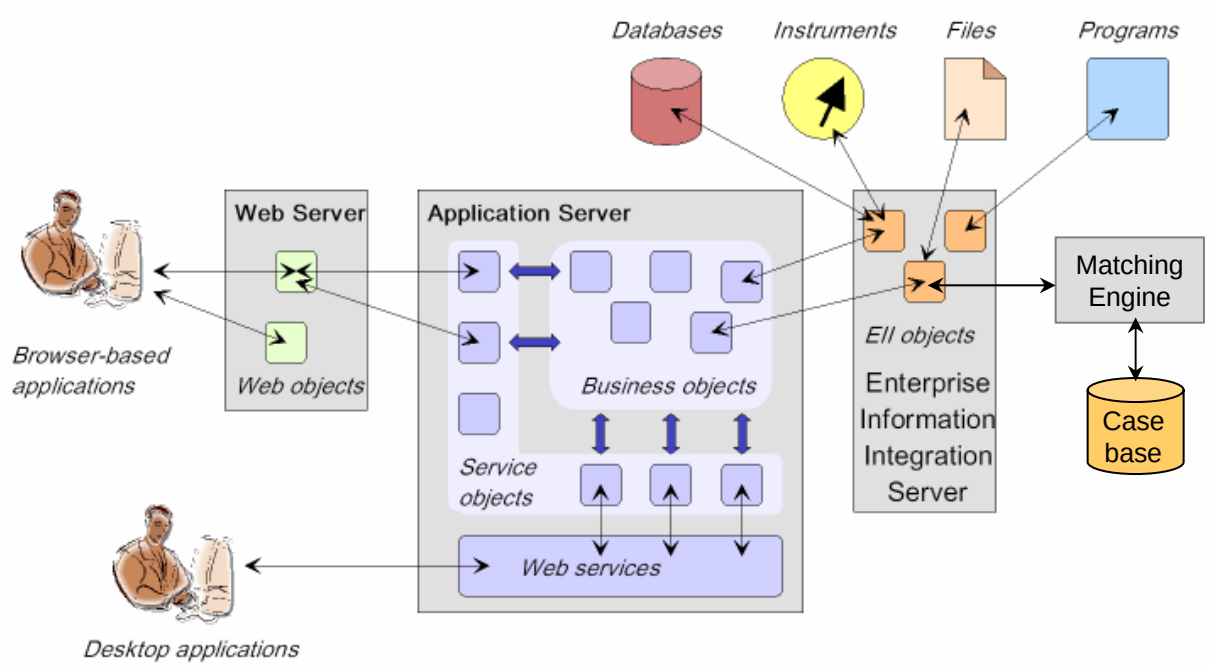
SHIP Architecture



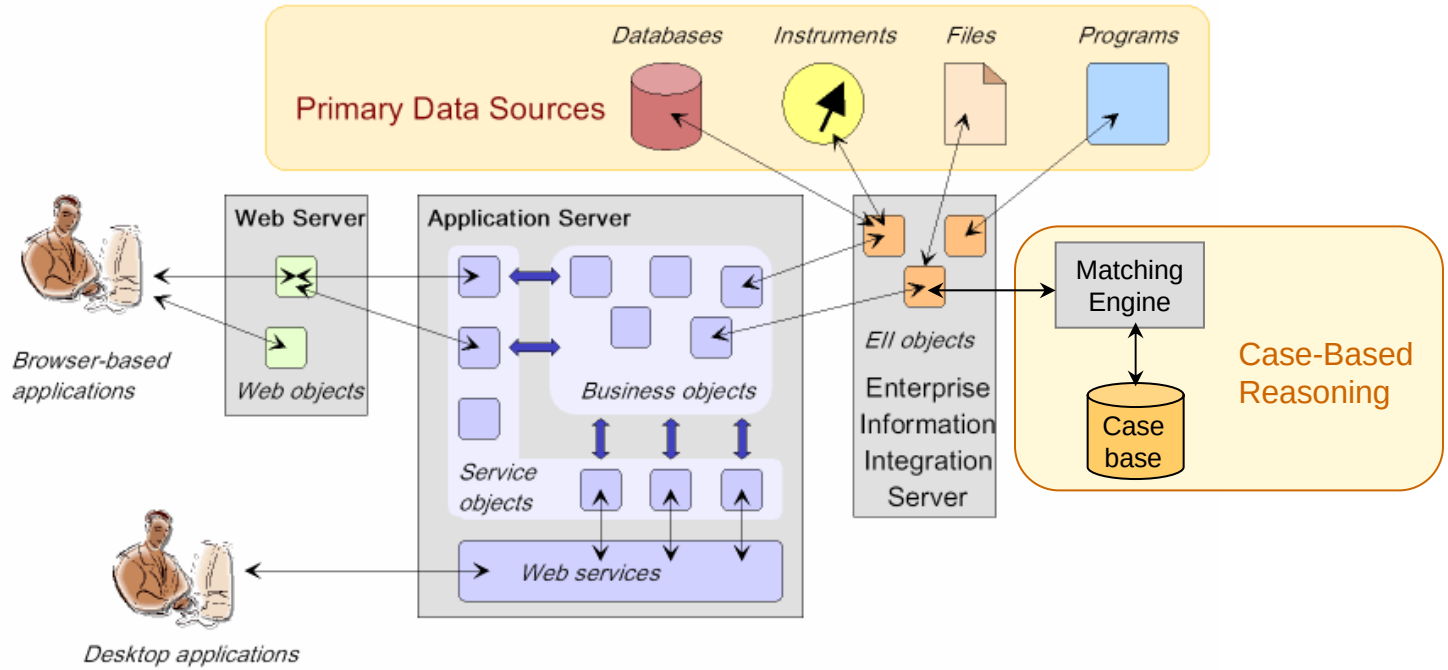
SHIP Architecture



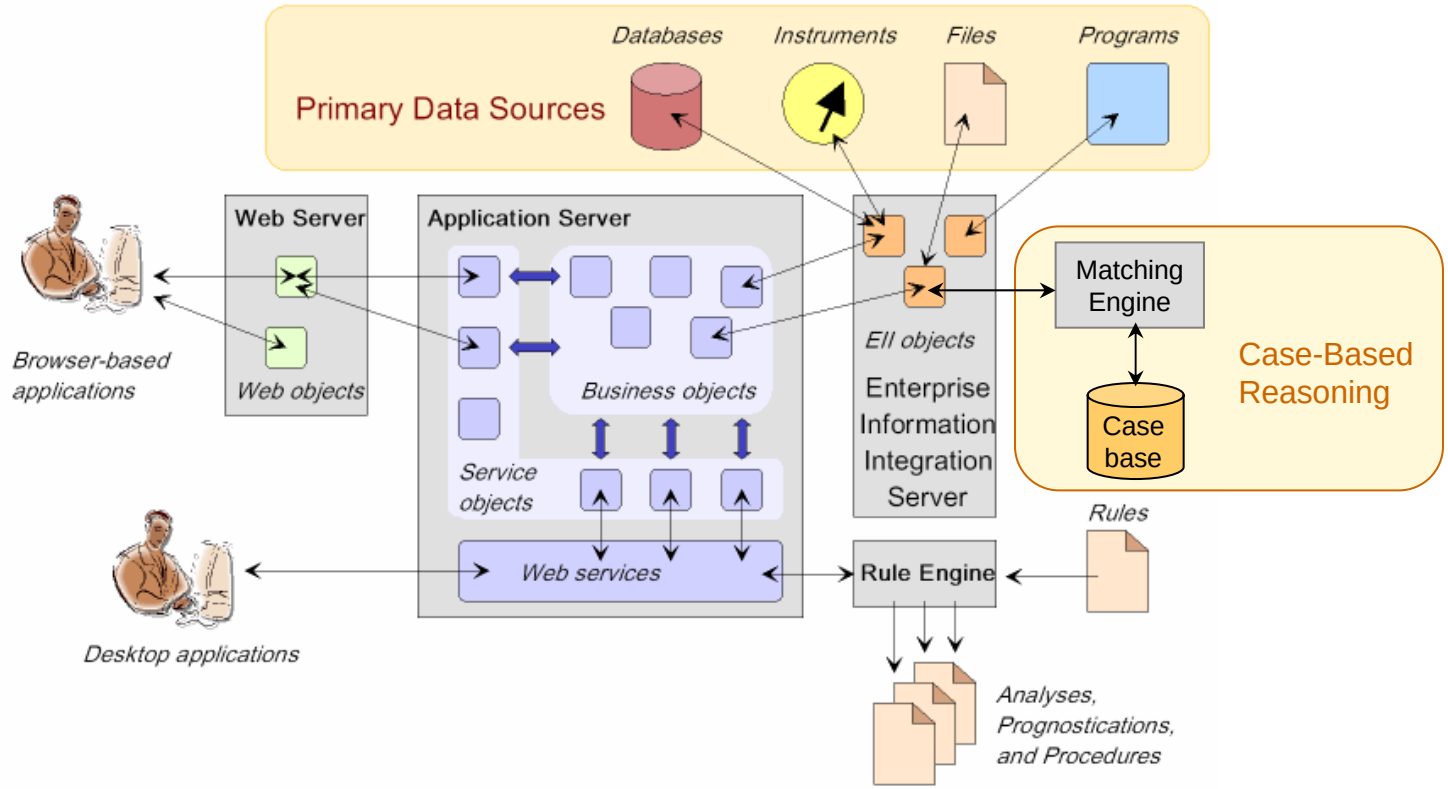
SHIP Architecture



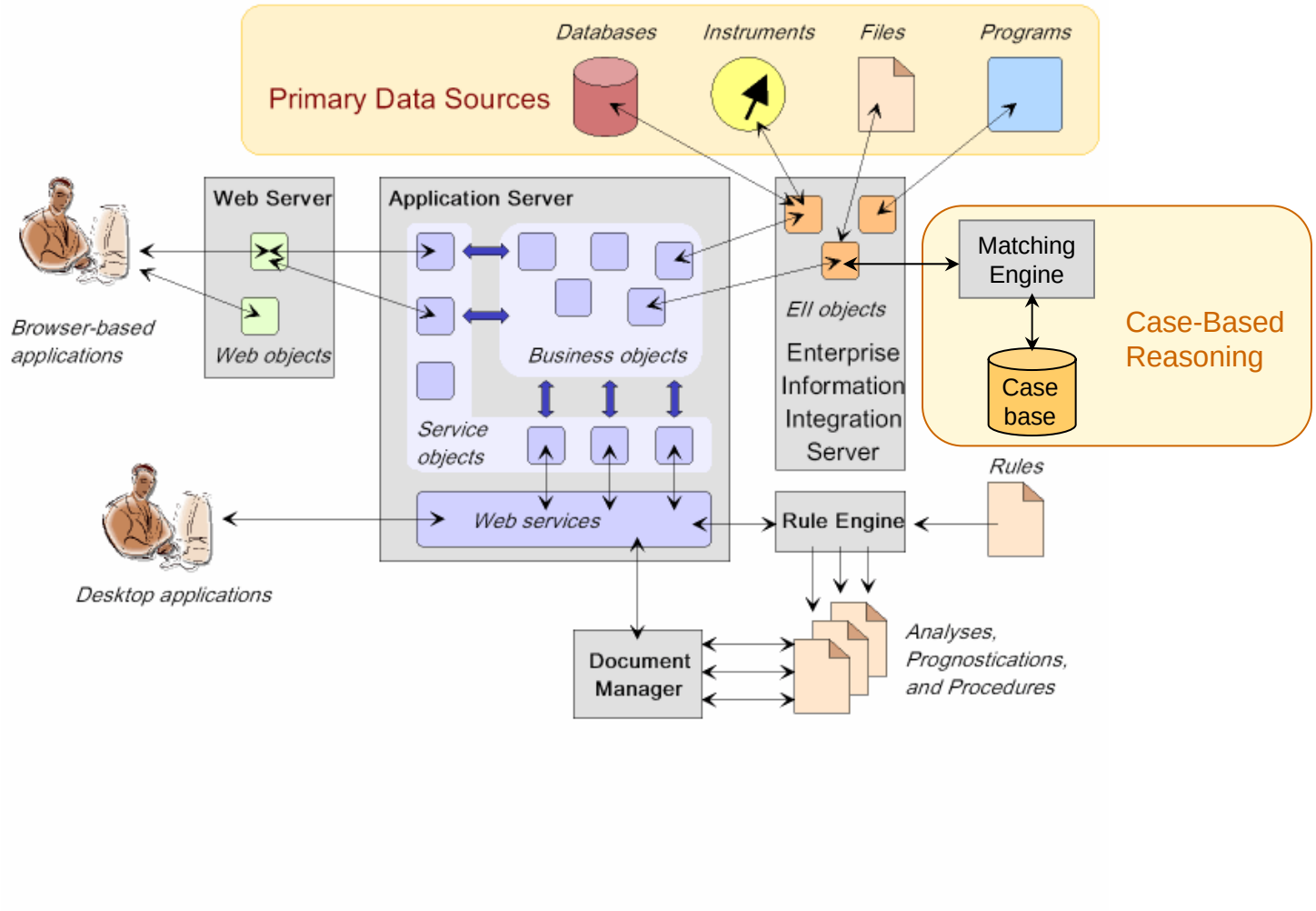
SHIP Architecture



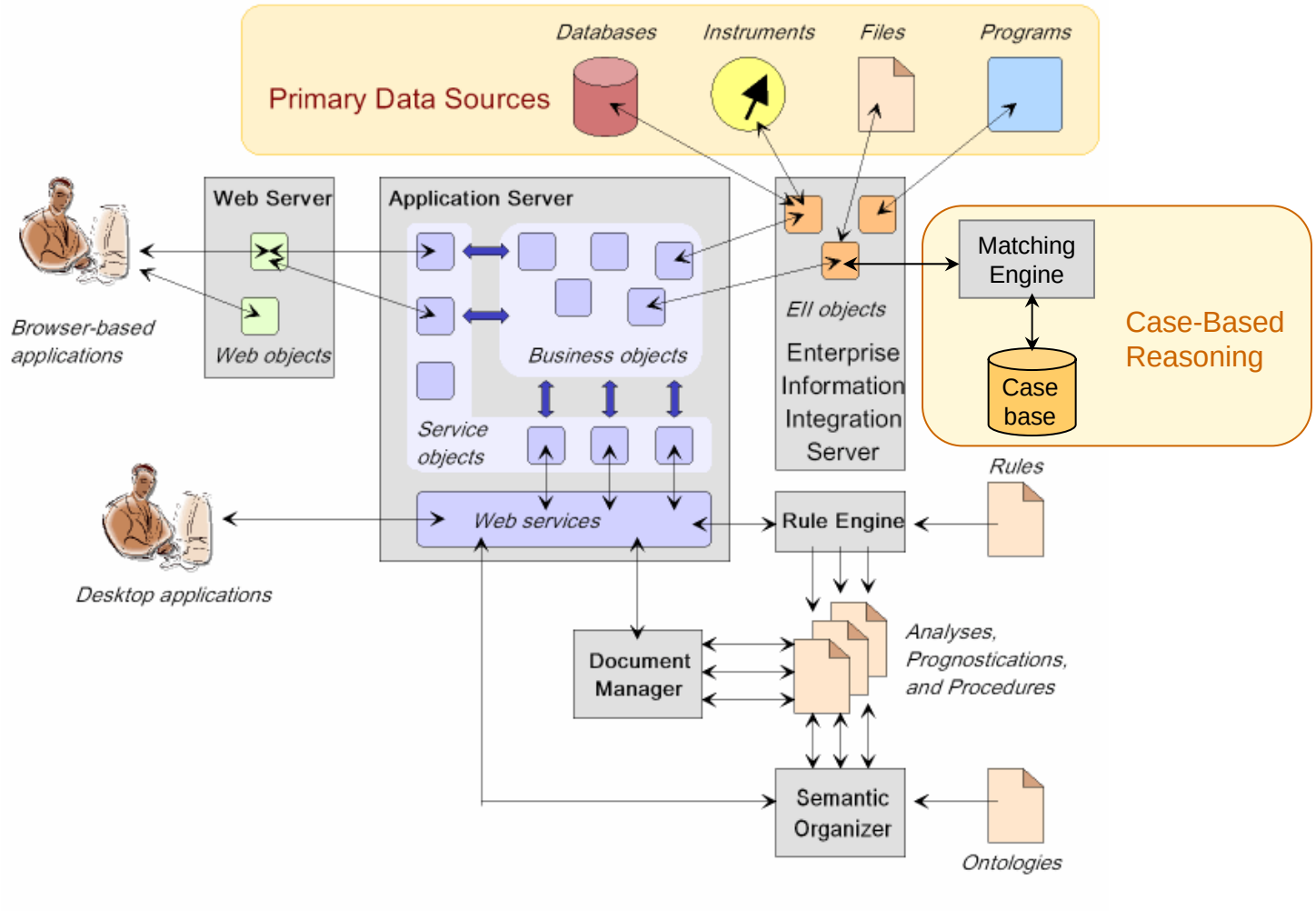
SHIP Architecture



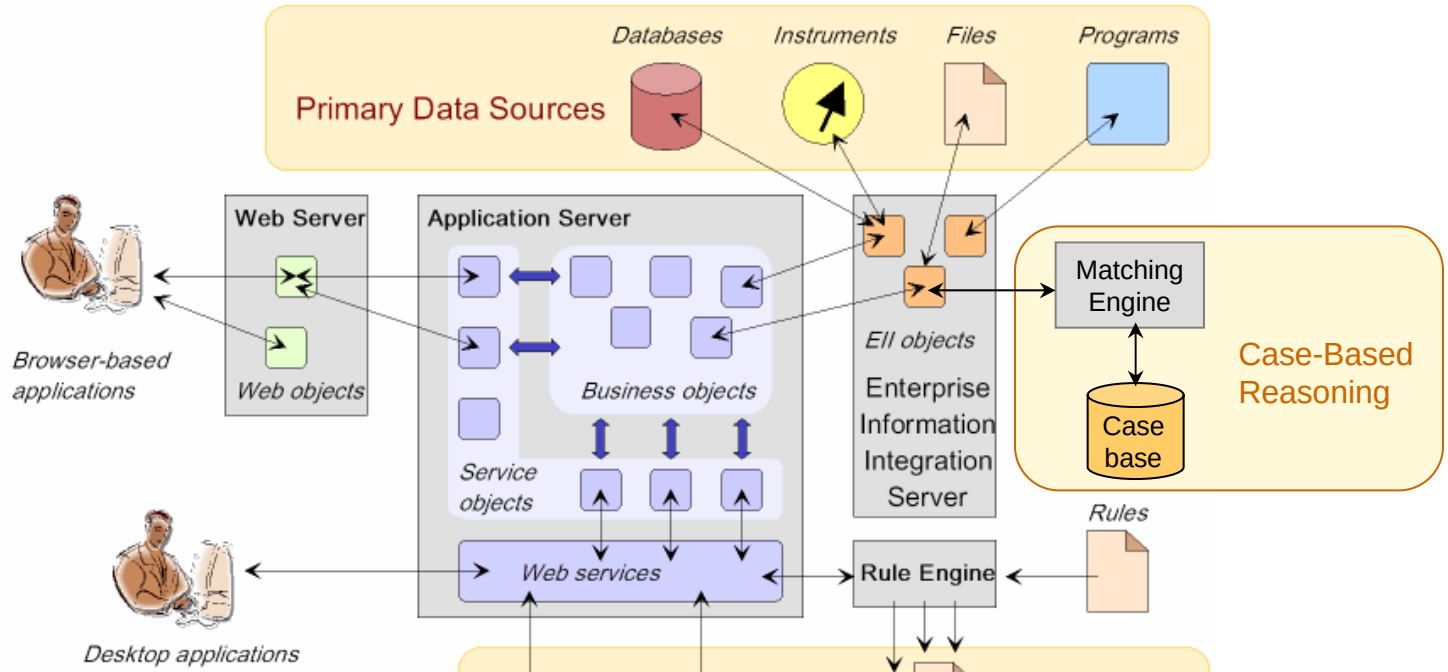
SHIP Architecture



SHIP Architecture

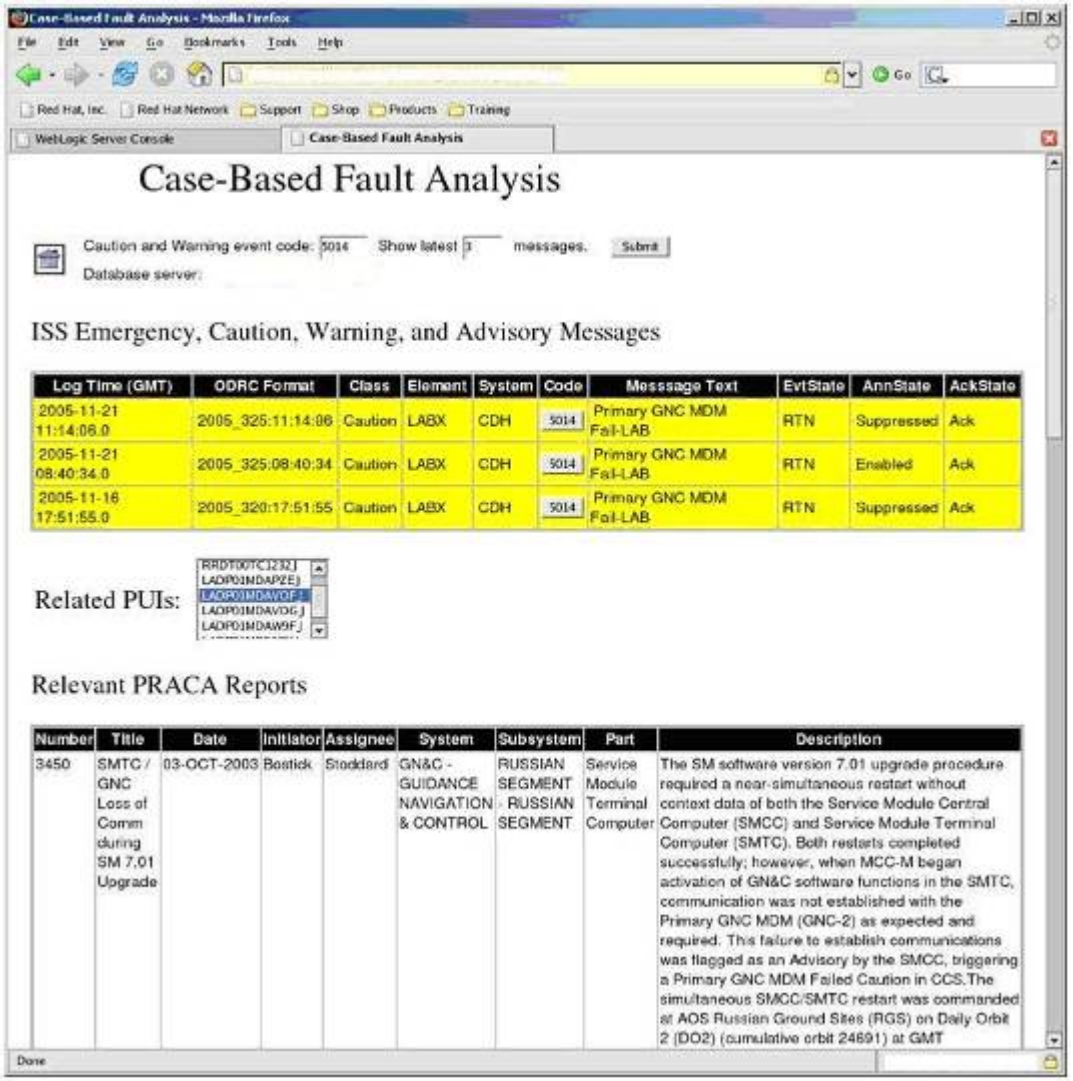


SHIP Architecture



Raw data
- ➔ Integrated information
  - ➔ Analysis
    - ➔ Knowledge
      - ➔ Action

SHIP Screen Shot



Case-Based Fault Analysis - Mozilla Firefox

Case-Based Fault Analysis

Caution and Warning event code: 5014 Show latest 3 messages. Submit

Database server:

ISS Emergency, Caution, Warning, and Advisory Messages

Log Time (GMT)	ODRC Format	Class	Element	System	Code	Messsage Text	EvtState	AnnState	AckState
2005-11-21 11:14:06.0	2005_325:11:14:06	Caution	LABX	CDH	5014	Primary GNC MDM Fail-LAB	RTN	Suppressed	Ack
2005-11-21 08:40:34.0	2005_325:08:40:34	Caution	LABX	CDH	5014	Primary GNC MDM Fail-LAB	RTN	Enabled	Ack
2005-11-16 17:51:55.0	2005_320:17:51:55	Caution	LABX	CDH	5014	Primary GNC MDM Fail-LAB	RTN	Suppressed	Ack

Related PUIs:

Relevant PRACA Reports

Number	Title	Date	Initiator	Assignee	System	Subsystem	Part	Description
3450	SMTC / GNC Loss of Comm during SM 7.01 Upgrade	03-OCT-2003	Bostick	Stoddard	GN&C - GUIDANCE NAVIGATION & CONTROL	RUSSIAN SEGMENT - RUSSIAN SEGMENT	Service Module Terminal Computer	The SM software version 7.01 upgrade procedure required a near-simultaneous restart without context data of both the Service Module Central Computer (SMCC) and Service Module Terminal Computer (SMTC). Both restarts completed successfully; however, when MCC-M began activation of GN&C software functions in the SMTC, communication was not established with the Primary GNC MDM (GNC-2) as expected and required. This failure to establish communications was flagged as an Advisory by the SMCC, triggering a Primary GNC MDM Failed Caution in CCS.The simultaneous SMCC/SMTC restart was commanded at AOS Russian Ground Sites (RGS) on Daily Orbit 2 (DO2) (cumulative orbit 24691) at GMT

- ISS
  International Space Station
- PUI
  Program Unique Identifier (part number)
- PRACA
  Problem Report and Corrective Action

SHIP Summary

- Problem Resolution and Corrective Action (PRACA) system for manned space vehicles
  - Disparate data sources: live sensor feeds, archived legacy data, XML files, Word documents, web pages, web services, databases, etc.
  - Case-based reasoning and rules-based diagnoses
  - Semantics-based knowledge management

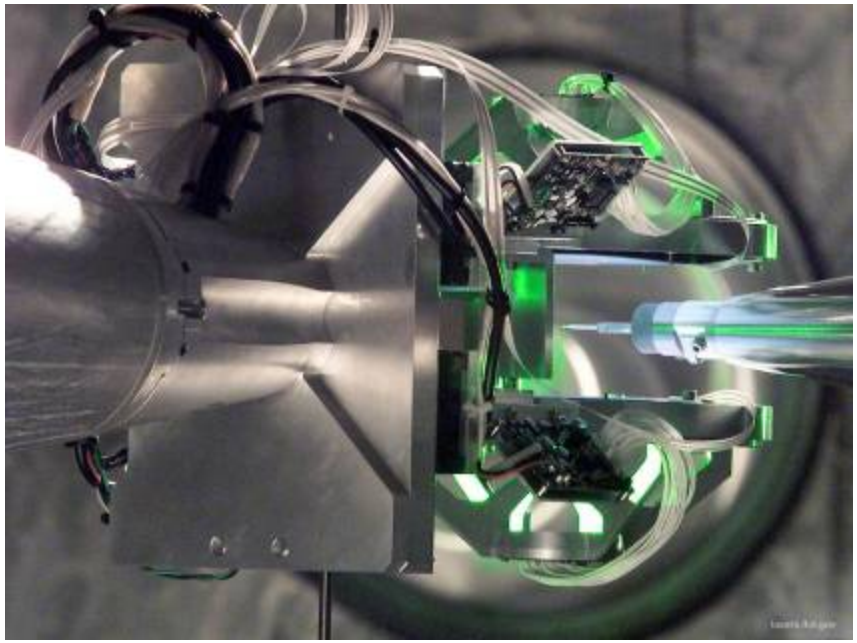
- Service-oriented architecture (SOA)
  - Web-based and desktop client applications
  - J2EE technologies
  - Web services
  - Enterprise information integration (EII)
  - Application integration
  - JavaServer Faces, BEA WebLogic, Composite Software
  - Open source software


Case Study 3:
Shot Data Management

Case Study 3: Shot Data Management

The National Ignition Facility (NIF) is a major laser-based fusion energy research project at the Lawrence Livermore National Laboratory.

Each simultaneous firing of its 192 powerful lasers at a BB-sized target generates gigabytes of data that **Shot Data Management** routes in near real time to the project scientists and researchers.



Target positioner

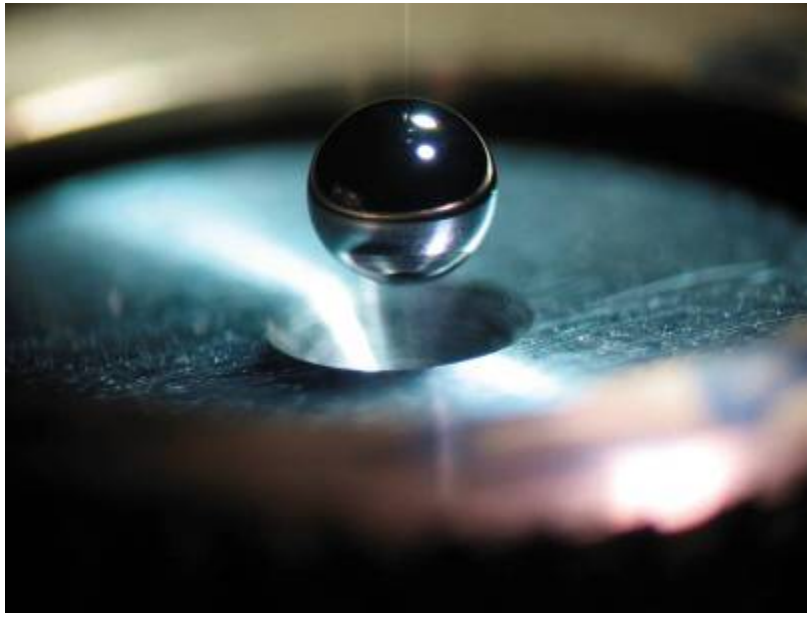
- My role: Enterprise Software Strategist
  - Architect, catalyst, lead developer, and systems integrator, 2006-2007
  - High security clearance (Level P)

Lawrence Livermore National Laboratory
Science in the National Interest

NIF
The National Ignition Facility

Shot Data Management Primary Services

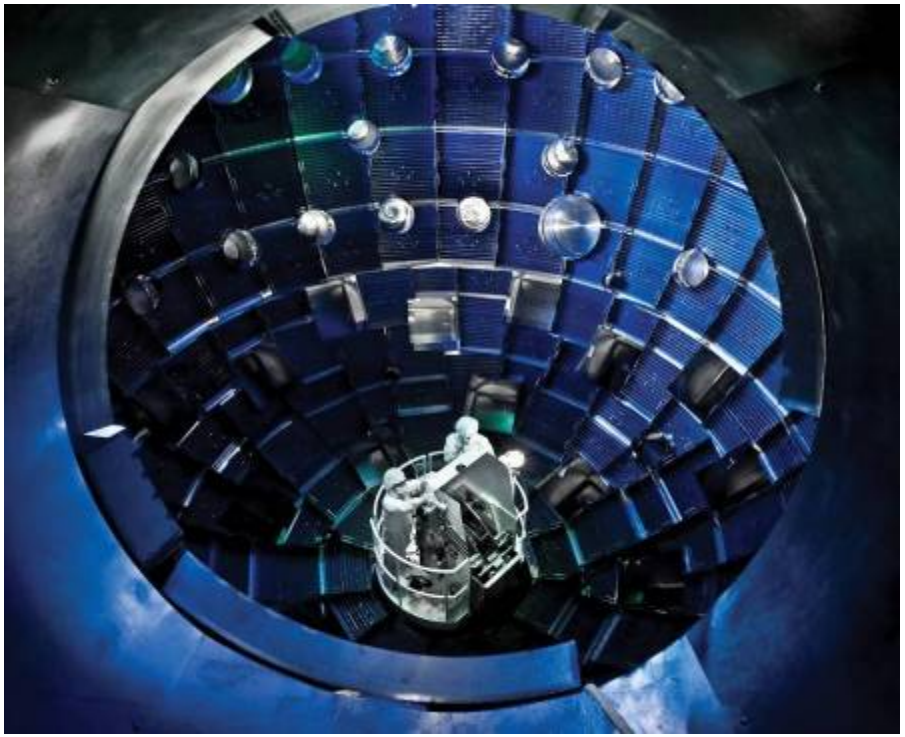
- Data provisioning
  - Integration of disparate and heterogeneous data sources
  - Secure access and downloads
  - Metadata-based queries
- Data management
  - Version management and revision control
  - Archive data from NIF's expected 30-year lifespan
- Data marts
  - Historical and trend analysis of specific subject areas
  - Support ad hoc data reporting



The target: A polished 2 mm capsule filled with cryogenic hydrogen fuel.

Requirements

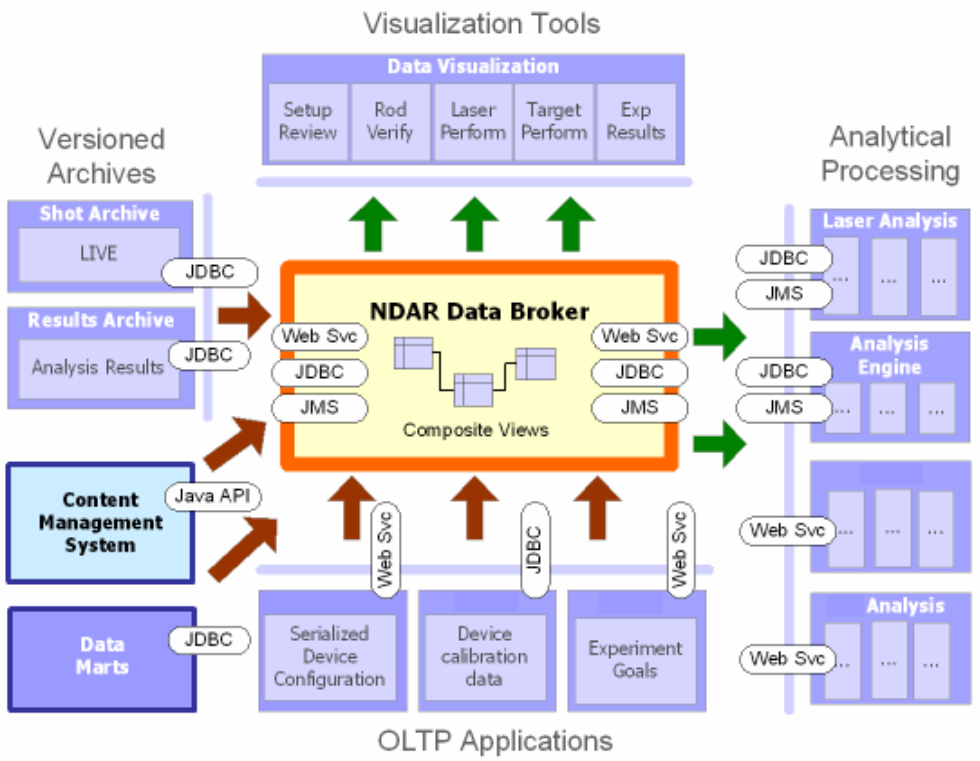
- Application insulation
- Eliminate duplicate data
- Near real time data access
- Reusable data services
- Decision support
- Quality control
- Hierarchical storage management
- Data security
- Integrate legacy data silos
- Workflow management



NIF target chamber

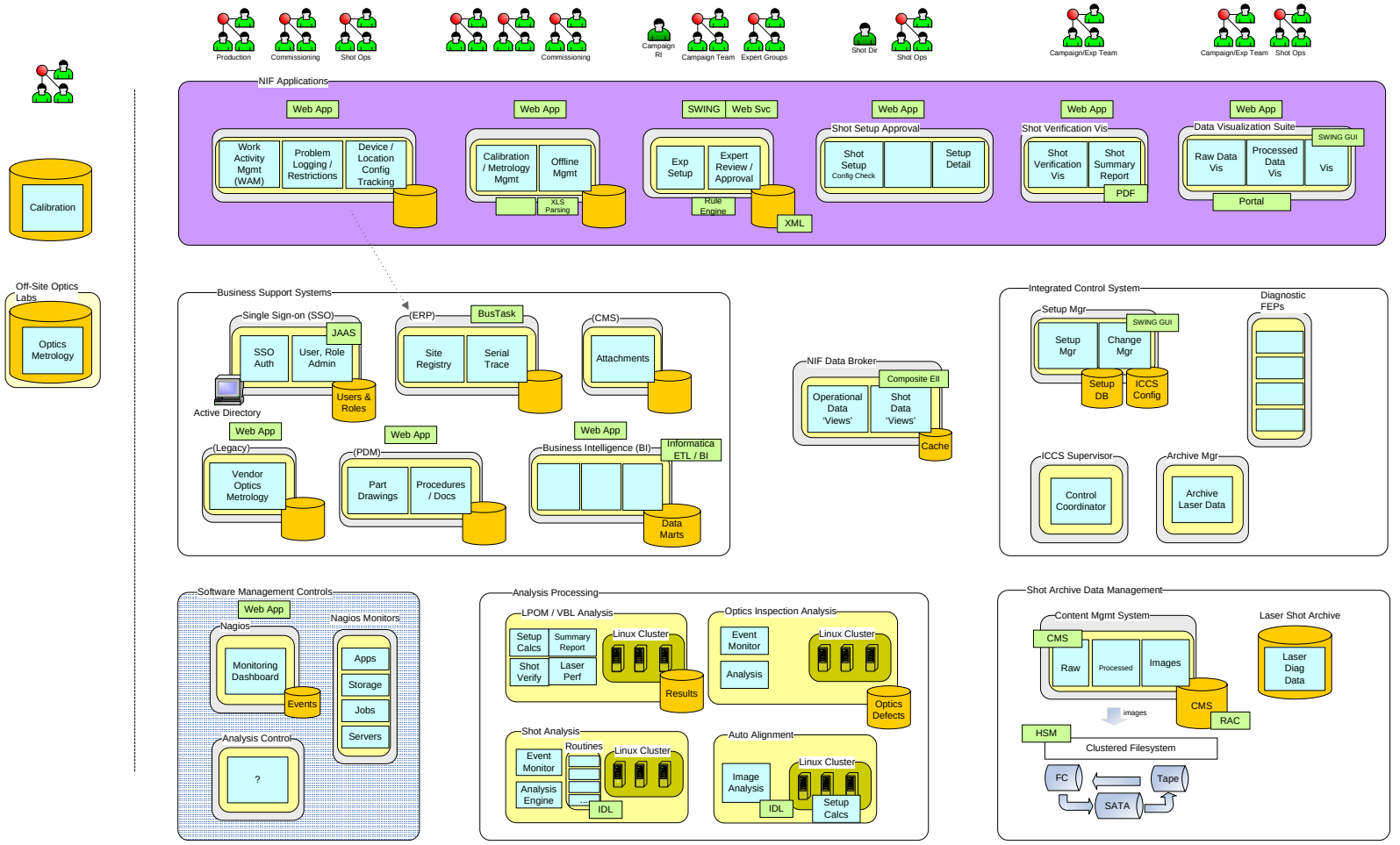
Architecture Overview

- Industry standards
  - Web services
  - XML data sources
  - JDBC and ODBC (database interfaces)
  - Java Message Service (JMS)
  - BPEL (Business Process Execution Language)
- Java programming language
- SOA (Service-Oriented Architecture)
- Linux server clusters

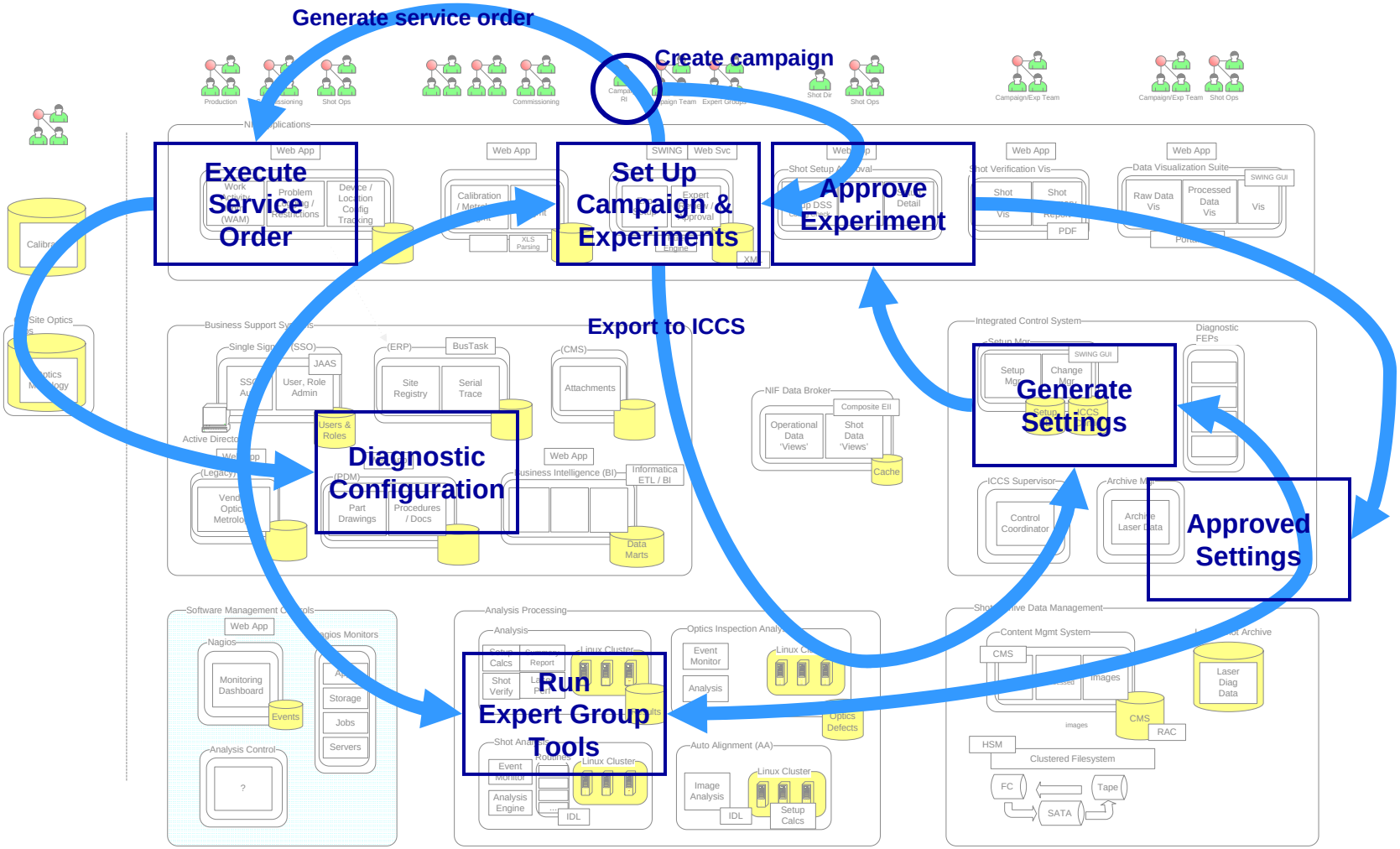


- Oracle RDBMS
- Oracle CMS (Content Management System)

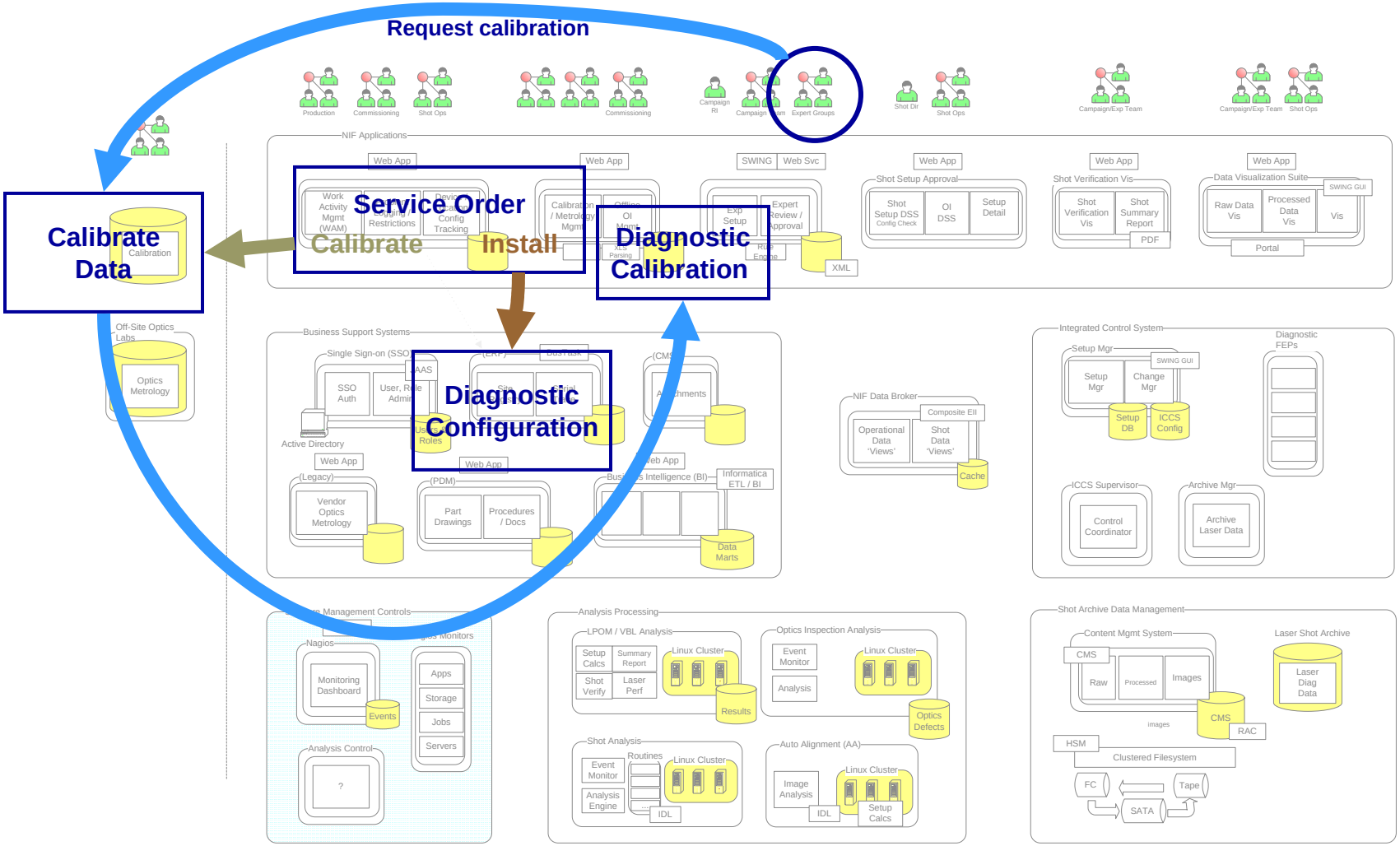
Shot Data Management Architecture



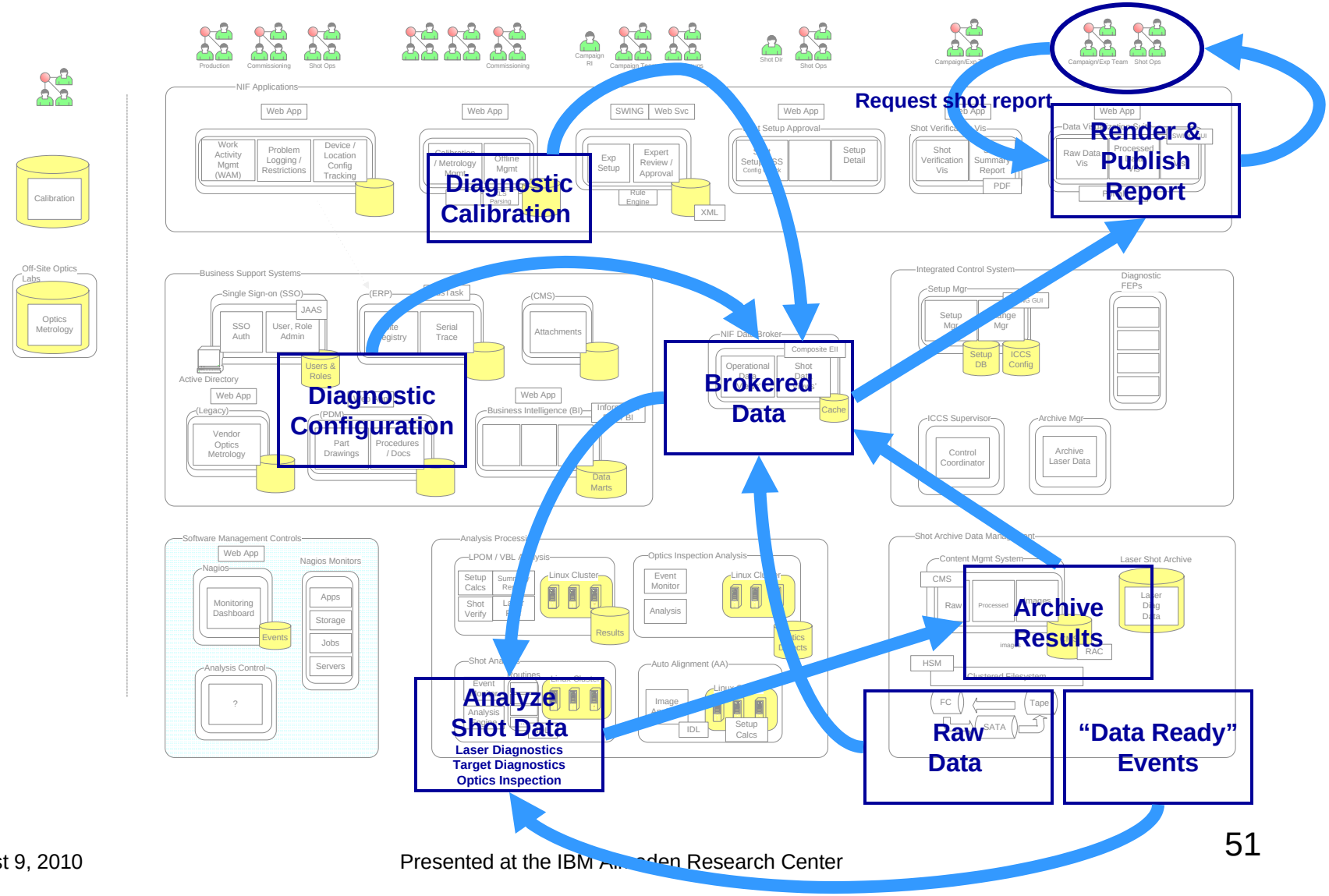
Workflow: Set Up Campaign & Approve Experiment



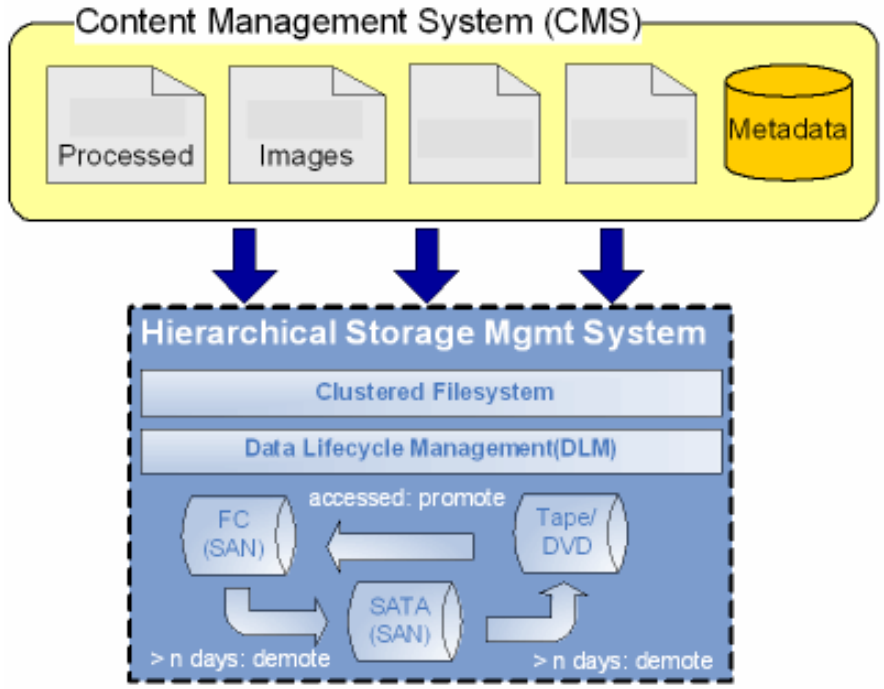
Workflow: Manage Configuration & Calibration



Workflow: Analyze & Visualize Shot Results

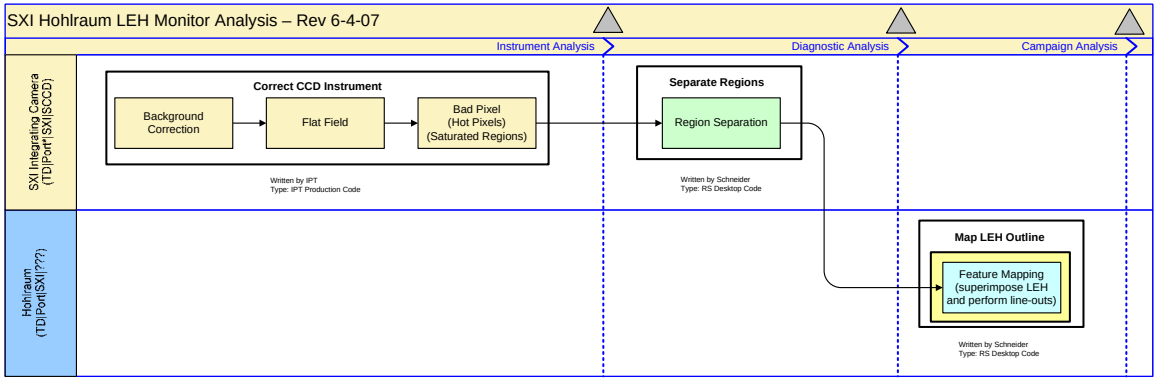


Data Archiving



- Hierarchical storage management system

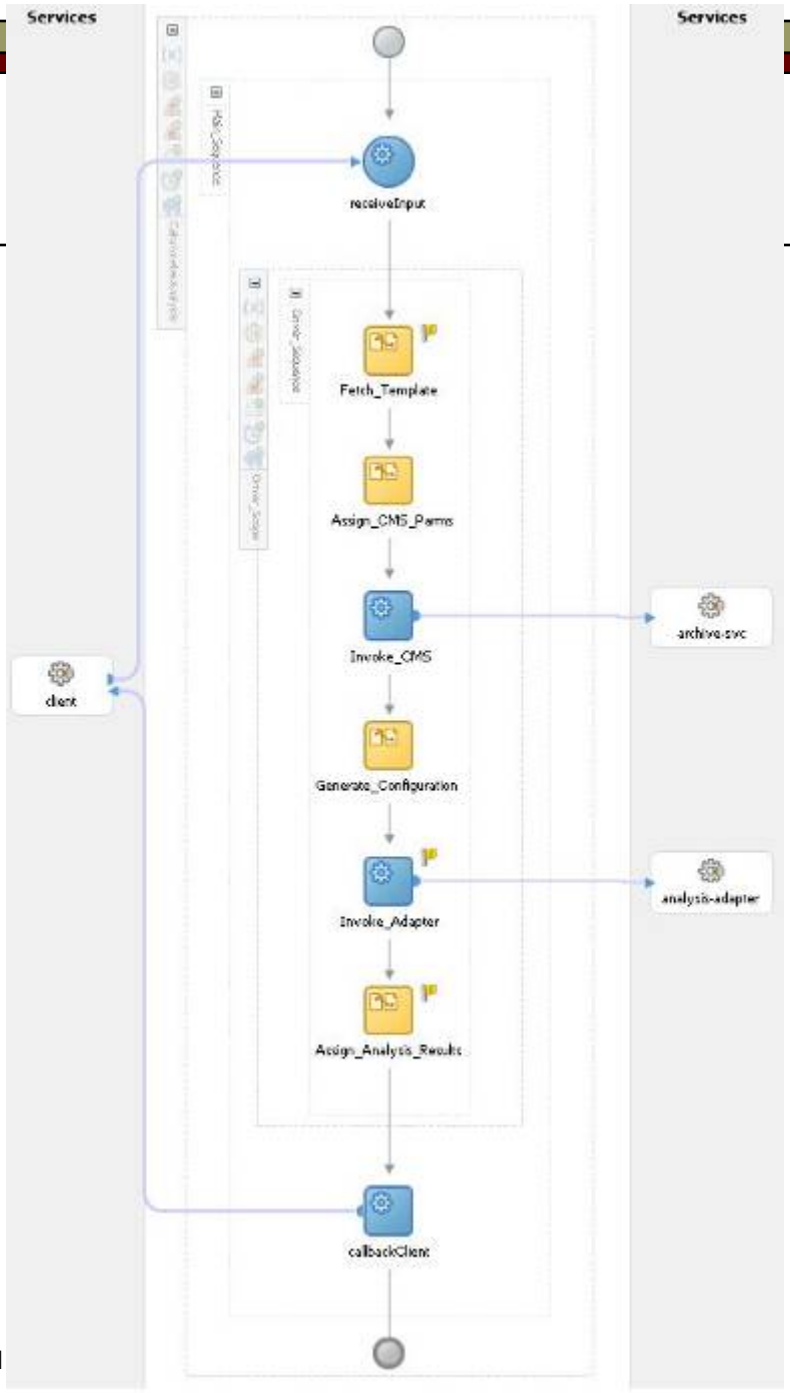
Visualizing Workflows



- Large display of many nodes in multiple “swim lanes”
  - Each node represents a unit of work to be performed.
  - Each swim lane represents a category of work.
  - Results flow from one node to the next based on rules.
- The display is maintained in real time.
  - Currently active nodes are highlighted.

Creating a Workflow

- Create workflow diagrams using an IDE.
- Compile workflow diagrams into BPEL scripts.
  - Business Process Execution Language (XML)



Shot Data Management Summary

- Large, complex distributed system based on SOA
  - Integration of legacy data silos
    - data warehouses
    - relational databases
    - content managers
  - Data archiving
    - HSM
    - 30 years of data
- Workflows
  - Shot approval and setup
  - Configuration and calibration
  - Results analysis and visualization

Installation of the target chamber

"This system saved 6 months off the software development schedule."

Important Lessons Learned

Sunset on Mars

Important Lessons

- Scientific data management systems share common features
  - Distributed and web-based
  - Integration of legacy subsystems
  - Integration of disparate and heterogeneous data
  - Near real-time data access with security
  - Data is unstructured
  - Data queries driven by metadata
  - Users collaborate to analyze data and write reports.

Important Lessons

- Keys to project success and high reliability
  - Well-understood requirements
  - Highly dedicated and competent developers
  - Good software engineering
    - Solid but flexible service-oriented architecture (SOA)
    - Industry standards and best practices
    - Unit testing
    - Monitoring tools
    - Create alternate paths around risky areas of code that you can “hot switch” at run time.

Important Lessons

- Keys to project success and high reliability *cont'd*
  - Strong project leadership
    - Realistic project schedule with clear milestones
    - Agile (iterative) development methodology
    - Maintain constant communication with users
    - The leader must actively drive the project from milestone to milestone
  - Obsessive stress testing
    - Discover the system limits (before the users do)
    - Learn the system's behavior under extreme stress.


Questions?

WEEKLY WORLD
NEWS®
February 17, 2004

DRUNK PILOTS START THEIR OWN AIRLINE

ALIEN SKULLS FOUND ON MARS!

INSIDE: Photos gov't doesn't want you to see!

TOP SECRET NASA REPORT: 'ANCIENT ASTRONAUTS WERE OUR ANCESTORS'

$1.99 US / $2.29 CANADA
07>
0 71658 51030 2

PLUS

SADDAM'S PRISON LOVER TELLS ALL!

The Retrieve-Reuse-Revise-Retain Cycle