

Fedora 3: A Smooth Migration

Michael Durbin



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

The Scenario

- New versions of software promise exciting new capabilities and improvements.
- They also present new risks, and the upgrade process can be a source of apprehension and frustration.
- Fedora 3.x offers a great many improvements over 2.x
 - Content Model Architecture
 - Improved (fixed) resource index (over 2.2.3)
 - Performance improvements
- But the transition/upgrade is not without challenges



The Scenario: Our Repository

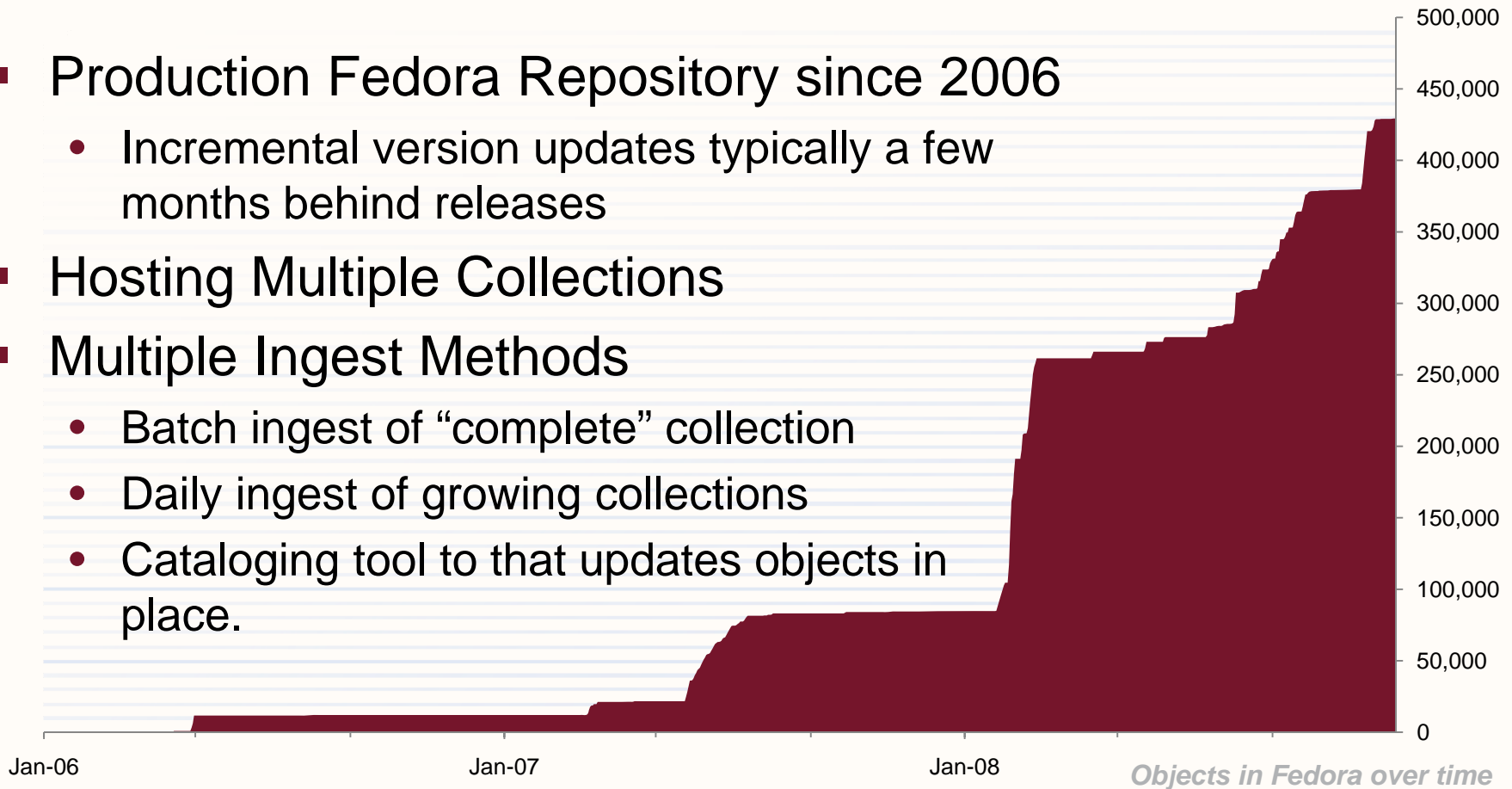
■ Production Fedora Repository since 2006

- Incremental version updates typically a few months behind releases

■ Hosting Multiple Collections

■ Multiple Ingest Methods

- Batch ingest of “complete” collection
- Daily ingest of growing collections
- Cataloging tool to that updates objects in place.

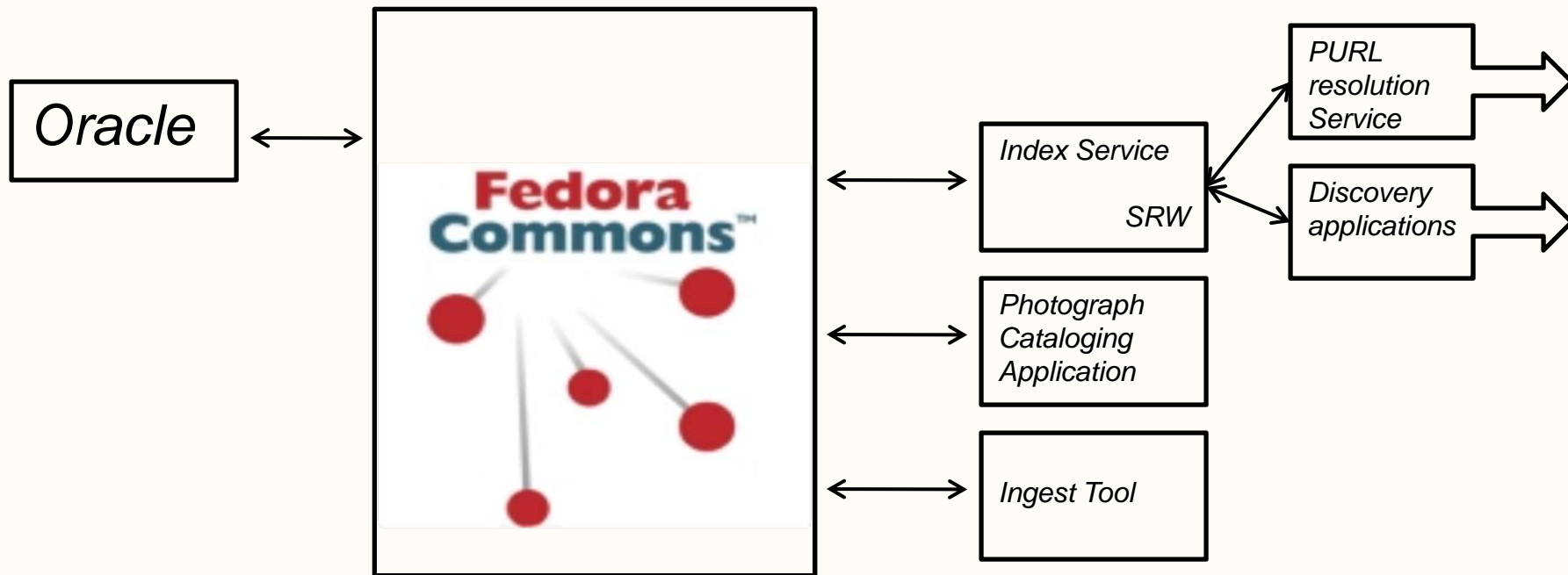


DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

The Scenario: Servers and Services

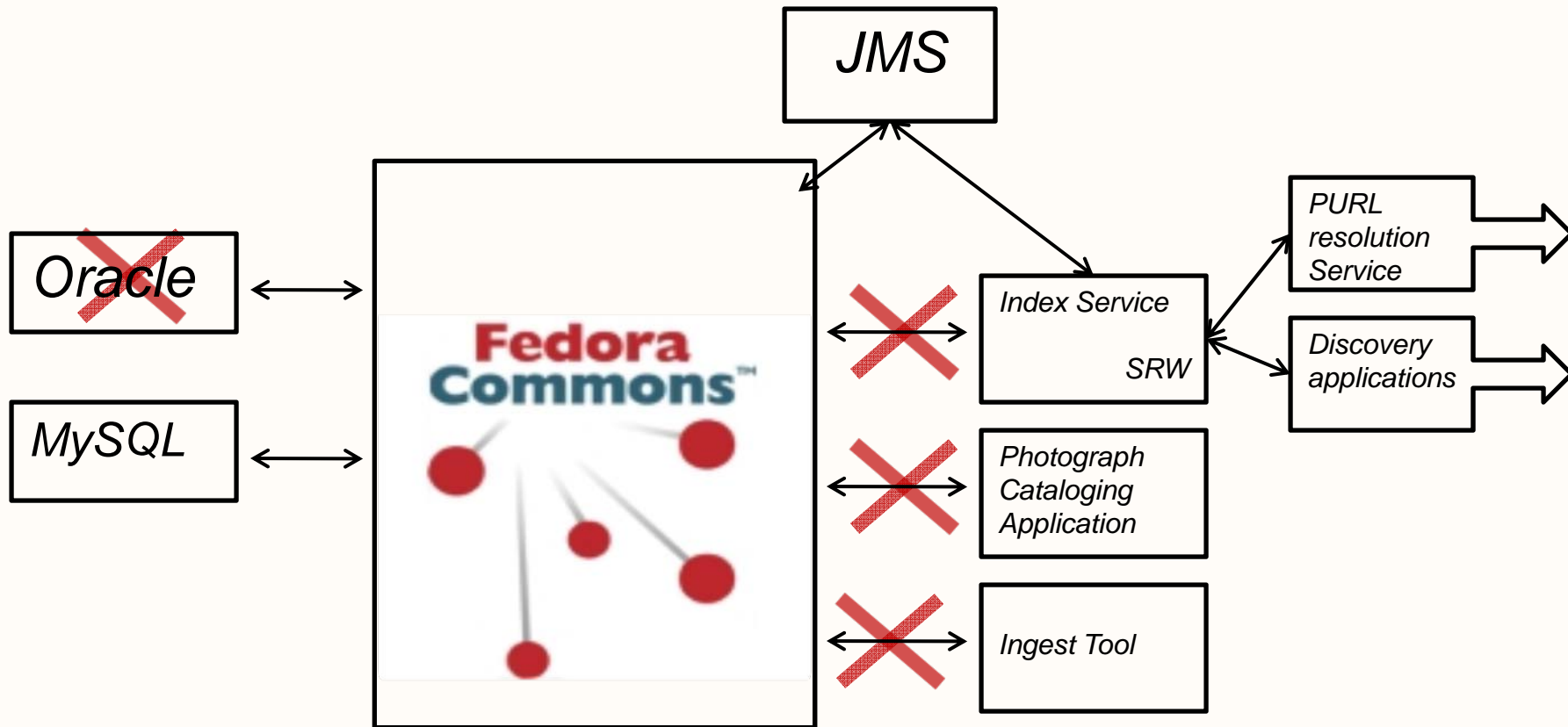


The Scenario: Migration Requirements

- Migrate to new backing database
 - Allow for a testing period and potential roll-back
 - No significant downtime
 - Minimal Ingest Suspension
 - Perform migration on a single server
-
- **No significant disruption in the user experience**



The Scenario: Servers and Services



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Migration Guide

- Helpful and complete migration guide:
 - <http://fedora-commons.org/confluence/display/FCR30/Upgrading+from+2.x>
- Overview
 - Install
 - Analyzer
 - Generator
 - Transformer
 - Rebuild
 - Ingest
 - Content model objects
 - Service Deployment objects

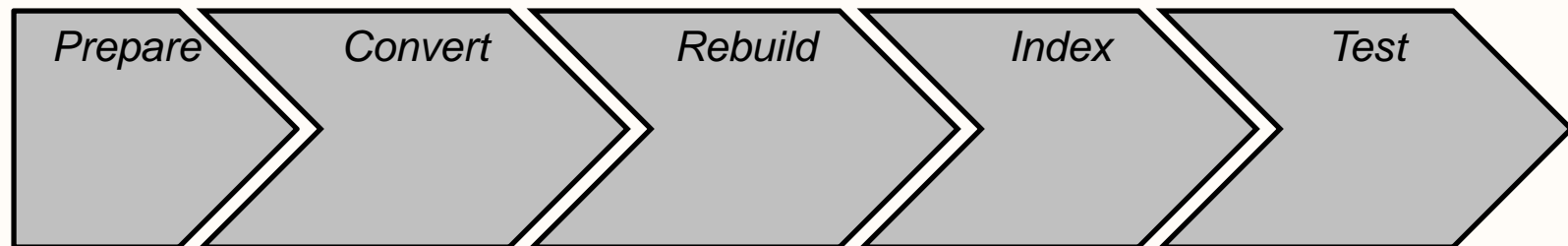


DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

The Migration Plan



- Some steps may take a lot of machine time
- Other steps may take a lot of human time
- Problems could appear at any time which could throw off schedule

One week should be sufficient



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Migration: Preparation

- Set up environment
 - user, database
 - Issued an ingest/update freeze
 - Deployed an XACML policy change to enforce it
 - Run fedora 3.1 installer
 - New tomcat instance on new ports
 - Points to new database
 - Update configuration
 - E-mail list, datastream and object directories, default pid namespace
 - Start fedora
-



Migration: Data conversion

- Analyzer

- Duplicate objects cause problems
- 3 hours for ~400K objects
- PIDs were divided into groups with identical signatures



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Migration: Analyzer - Content Models

- **7 intended content models resulted in 53!?!?**
 - Analyzer output can be used to audit content models
 - Migration process allows a unique, low cost opportunity to clean up objects
 - This can quickly become overwhelming
 - Setting the analyzer to ignore characteristics can simplify this process
- **Our Strategy**
 - Retain original logical models
 - Minimize required changes to external tools
 - Allow for easy future experimentation



Migration: Generator and Transformer

- Generator
 - Creates XSTL stylesheets to update FOXML objects
- Transformer
 - Fedora 3 was pointed to a COPY of the old object directory
 - Quick enough to merit a dry run
 - Modifies FOXML files in place



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services

Migration: Tools

- Ingest tool and Photo Cataloging application
 - Refactor to use new API
 - Content models
 - RELS-EXT methods
- Index service / SRU search
 - Refactor to use JMS
 - Update to handle new FOXML version
 - Update to reflect new CMA
 - multiple content models
 - Stored in RELS-EXT



Migration: Rebuild

- Rebuild the database
 - Less than 6 hours
- Rebuild the resource index
 - Over 11 hours
- Rebuild external Lucene search index directories
 - Over 20 hours
 - Also serves as a test and validation of repository data migration and index XSLT
 - Note: behavior or invalid or unauthorized requests has changed such that document() calls fail (as they should) when they wouldn't before



Migration: Switchover

- Stopped fedora 2 (downtime starts)
- Moved index directories
- Replaced web applications
 - fedora, SRW, index service, PURL resolver
- Refreshed environment settings (FEDORA_HOME)
- Started fedora 3
- Ingested content model objects
- Enjoyed success
- Started playing with new features



Thanks!

Detailed migration notes available online at:
<https://wiki.dlib.indiana.edu/confluence/x/J4A2BQ>

Questions?



DIGITAL LIBRARY PROGRAM

INDIANA UNIVERSITY

Libraries/University Information Technology Services