



netarchive.dk

Distributed Installation

Lars Clausen
Netarkivet

Why distributed install?

- Additional resources
- Redundancy
- Geographical distribution
- Protection from attacks/sabotage
- Protection from systematic failure
- Partial upgrades

Applications

- ❑ Basic building block
- ❑ Each application runs in a separate JVM
- ❑ Settings files may be shared
- ❑ Each application must have own JMX ports, log files
 - Can be set on command line
- ❑

“Typical” arrangements

- ❑ Bitarchive separately
- ❑ Central server for harvestdefinition, arcrepository
- ❑ Index server with viewerproxy
- ❑ Harvestcontroller with sidekick

Locations

- ❑ Abstract term for copy of archive
- ❑ Each application belongs to a location
- ❑ Only really matters for bitarchive and bitarchivemonitor
- ❑ Batch jobs can be sent to either location
- ❑ Location-agnostic batch jobs can be sent to the faster location

Settings

- Location: `archive.arcrepository.*`,
`archive.bitarchive.thisLocation`,
`archive.bitarchive.thisCredentials`
- Viewerproxy: `viewerproxy.hostname`
- JMS: `common.jms.broker`
- RemoteFile: `common.remoteFile.*`
- Mail: `common.notifications.*`,
`common.mail`

Ports

- ❑ JMS Broker: 7676, 33700 to broker machine
- ❑ FTP/HTTP RemoteFile (multiple places)
- ❑ Web GUI (incoming)
- ❑ Mail (outgoing)
- ❑ JMX+RMI (to monitor machine)
- ❑ ViewerProxy (incoming)

JMX

- Used for monitoring only (so far)
- Each application has two JMX ports:
 - `<jmxport>` for JMX messages
 - `<jmxport>+100` for RMI
- Logging MBeans accessible for last 100 logs
-

Startup order and machine dependencies

- Few startup dependencies thanks to JMS
- HarvestController before SideKick
- ArcRepository on same machine as BitPreservation
- Bitarchives before BitarchiveMonitor
-

Shutdown and restart

- Must applications can be killed
 - Harvesters will not resume a job
 - Batch runs will be aborted and not return
- When restarting, clear the JMS queues
 - Deserialization errors
 - Old messages no longer pertinent

Using a proxy for the proxy

- ❑ Authentication proxy before viewerproxy
- ❑ Provides log of who saw what
- ❑ Legal requirement
- ❑ Simple