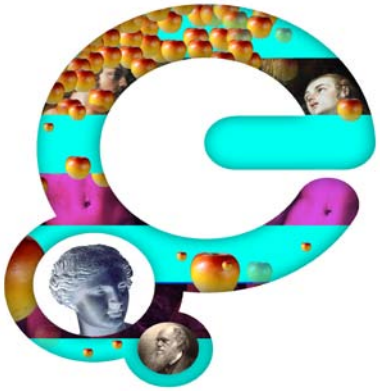


Europeana Prototype Architecture

Click to edit Master subtitle style

Bram van der Werf

Europeana v1.0 kick-off meetings

Overview

- What is the Europeana Prototype?
- Purpose of this presentation
- Evolution from Maquette to v0.3 and beyond
- Deployment diagram of current production environment
- Breakdown of components
  - Dashboard management interface
  - Image server
  - SOLR search server
  - Portal front-end application
  - Master/Slave replication setup

What is the Europeana Prototype?

- The Europeana prototype is a proof-of-concept environment to:
  - Validate candidate technologies for the v1.0 release with regards to searching, aligning semantics in a multidimensional content space, cross-lingual and cross-cultural information retrieval
  - Get hands on experience with ingesting and mapping a plethora of: metadata formats, languages, controlled vocabularies, etc.
  - Start building a pan-european network of content aggregators and cultural heritage partners
  - Validate strategies to guarantee future scalibility, maintainability, agility and extensibility of the Europeana business model.

Purpose

- Communicate high-level deployment architecture of the Europeana Prototype that is currently in production
- Give insight in how the current '*ingestion workflow*' feeds into the production environment.
- But not:
  - A review of how the current relates to EDLnet D2.5 'Europeana Outline Functional Specification', especially 3.3 *Functional Model Overview* and 3.1 *Logical data model.*
  - An in depth overview of technologies used.
  - A review of what did and did not work during the prototype development

Prototype evolution

- Maquette (January 2008)
- Blue design prototype v0.0 ‘Styx’ (June 2008)
  - 500.000 records searchable
  - Release under password protection to network partners
- Salter-Baxter design prototype v0.1 (November 2008)
  - 4.5 million records searchable
  - Official release in Brussel by EU Commission president Barosso
- New production environment release v0.2 (Christmas 2008)
  - Scaled to handle 8000 concurrent users
- MyEuropeana release v0.3 (March 2009)
  - Re-enabled MyEuropeana login user-space environment

Deployment diagram breakdown

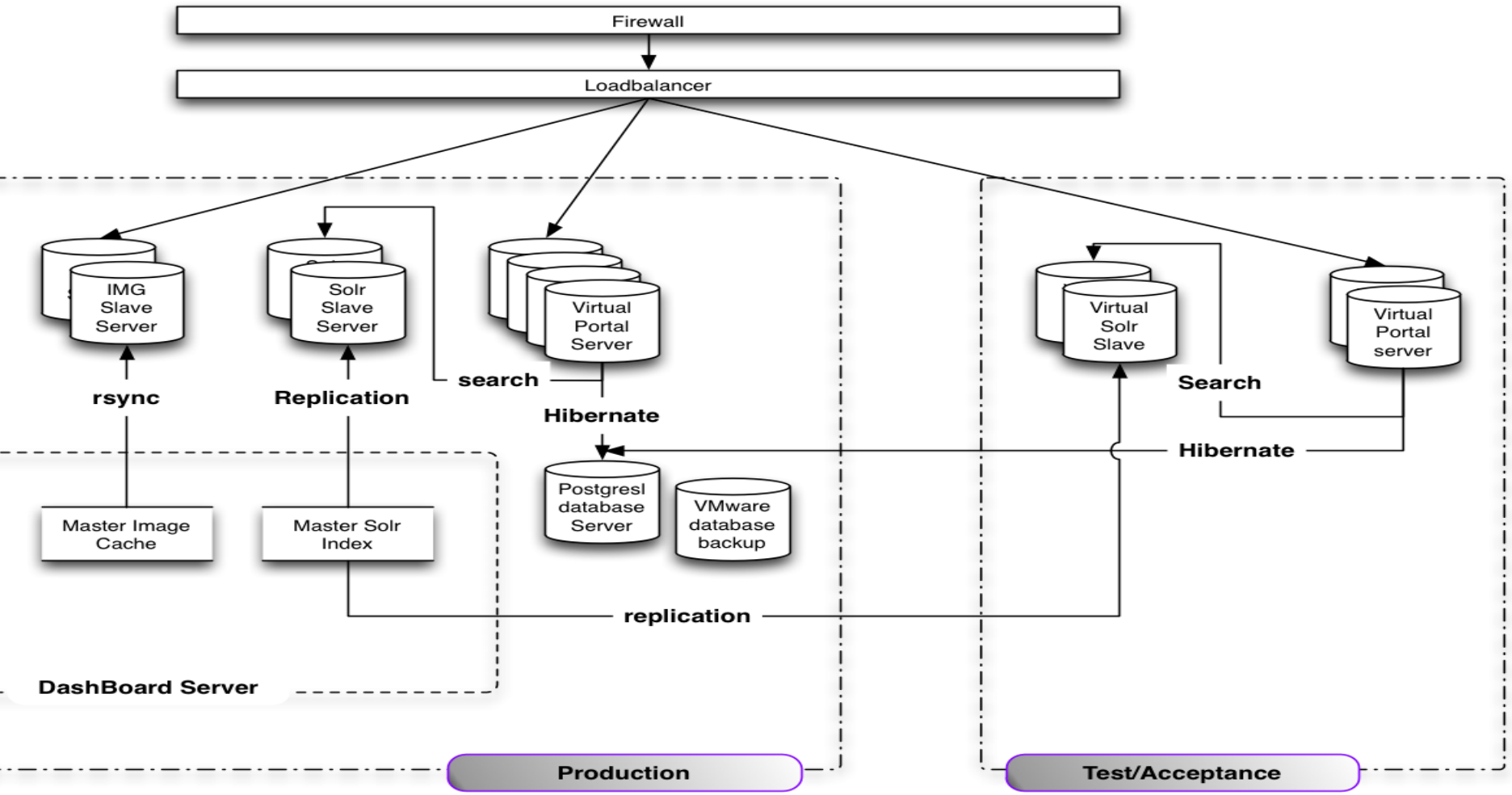


Image Server (1)

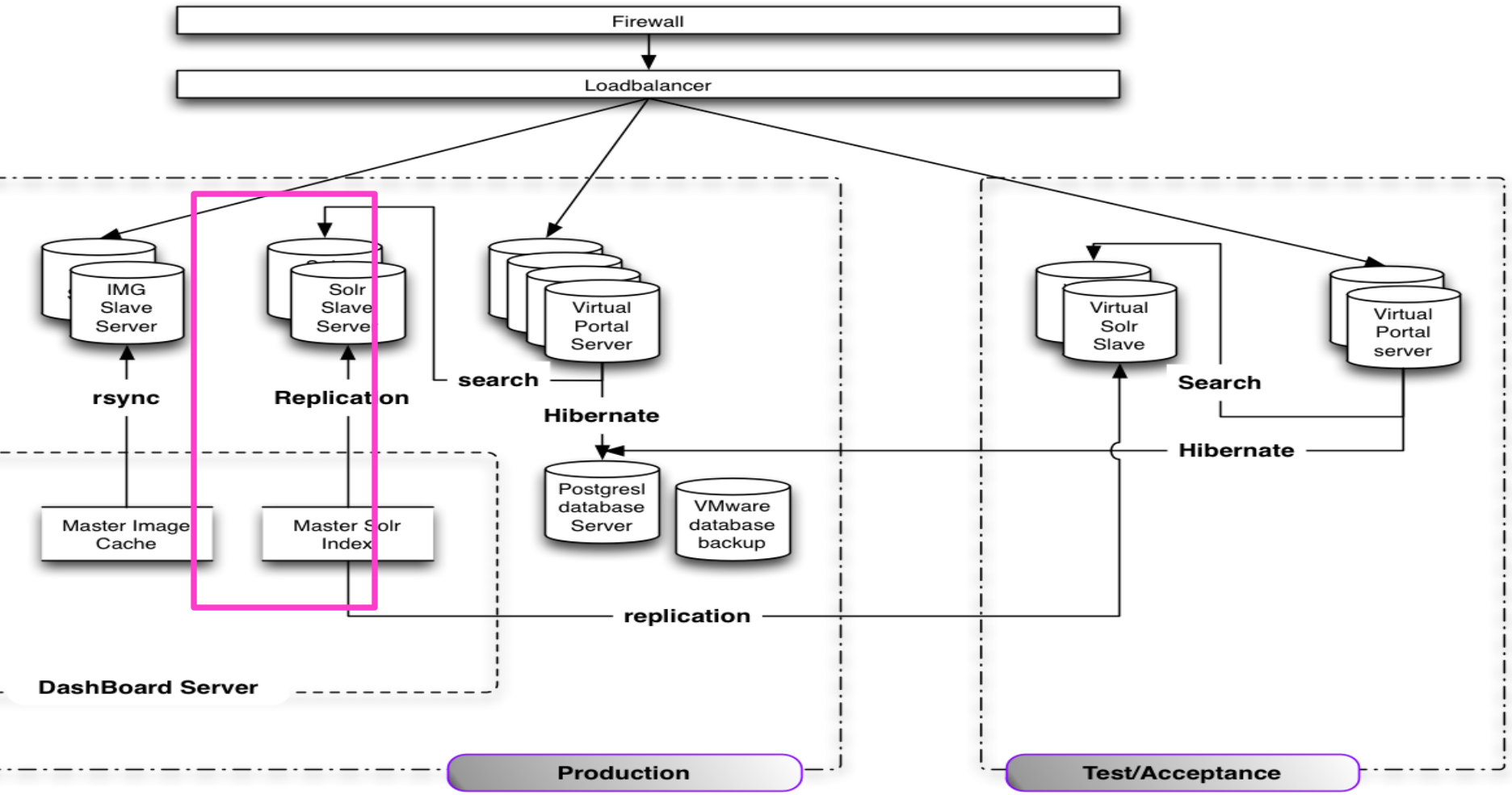
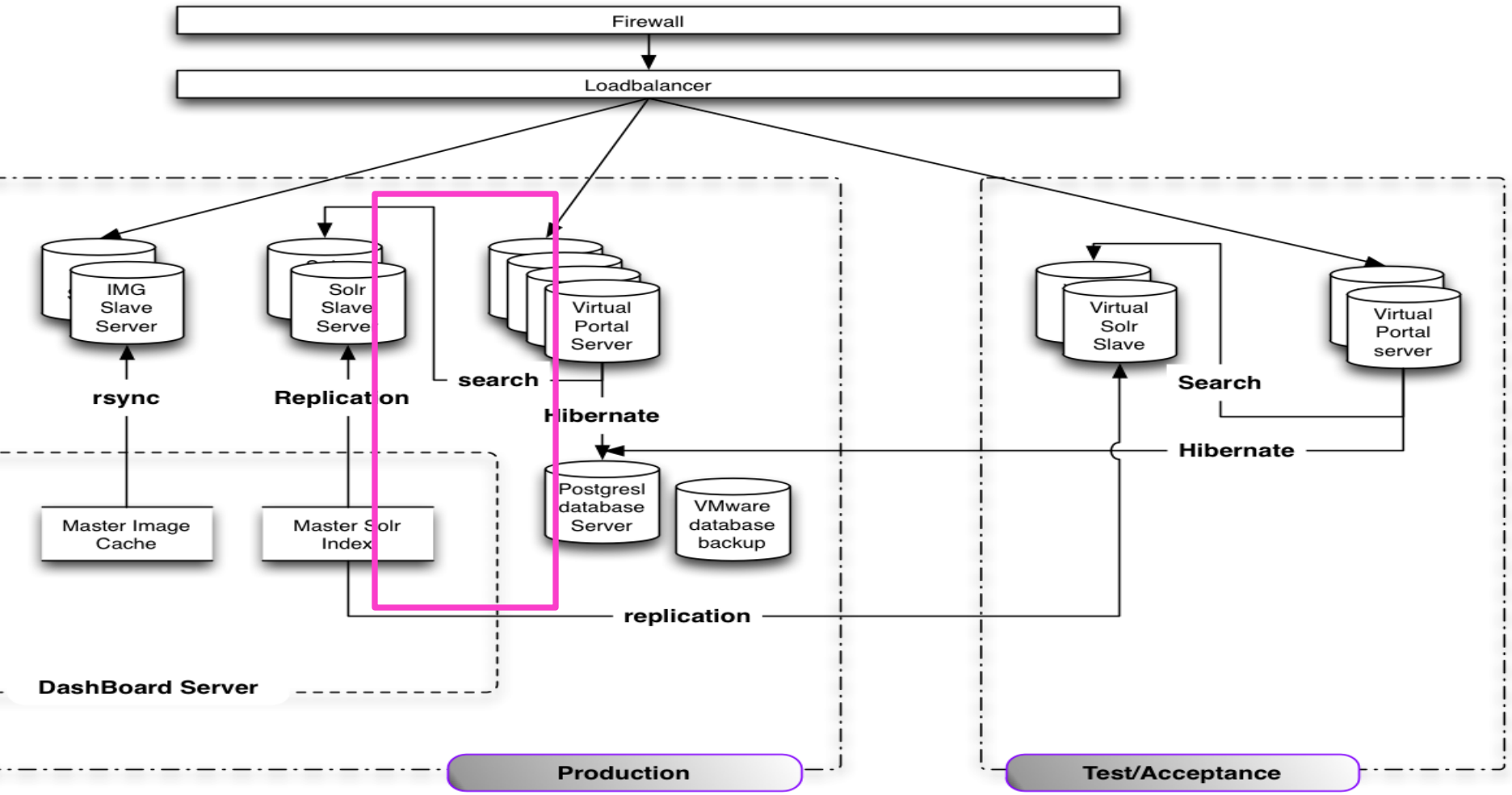


Image Server (2)

- ESE-element '*europeana:object*' uri's are preprocessed using ImageMagick and stored in 4096 folders (=163) as SHA1 hashed strings for very fast retrieval
- Simple cache servlet for retrieval of the object as http-requests from the users' browser
- Replication:
  - Thumbnails are cached on the Dashboard server that function as the 'Image Cache' Master
  - Master 'Image Cache' is synced with production slaves every hour via rsync
- Statistics: served over 75 million images in first 3 months of 2009

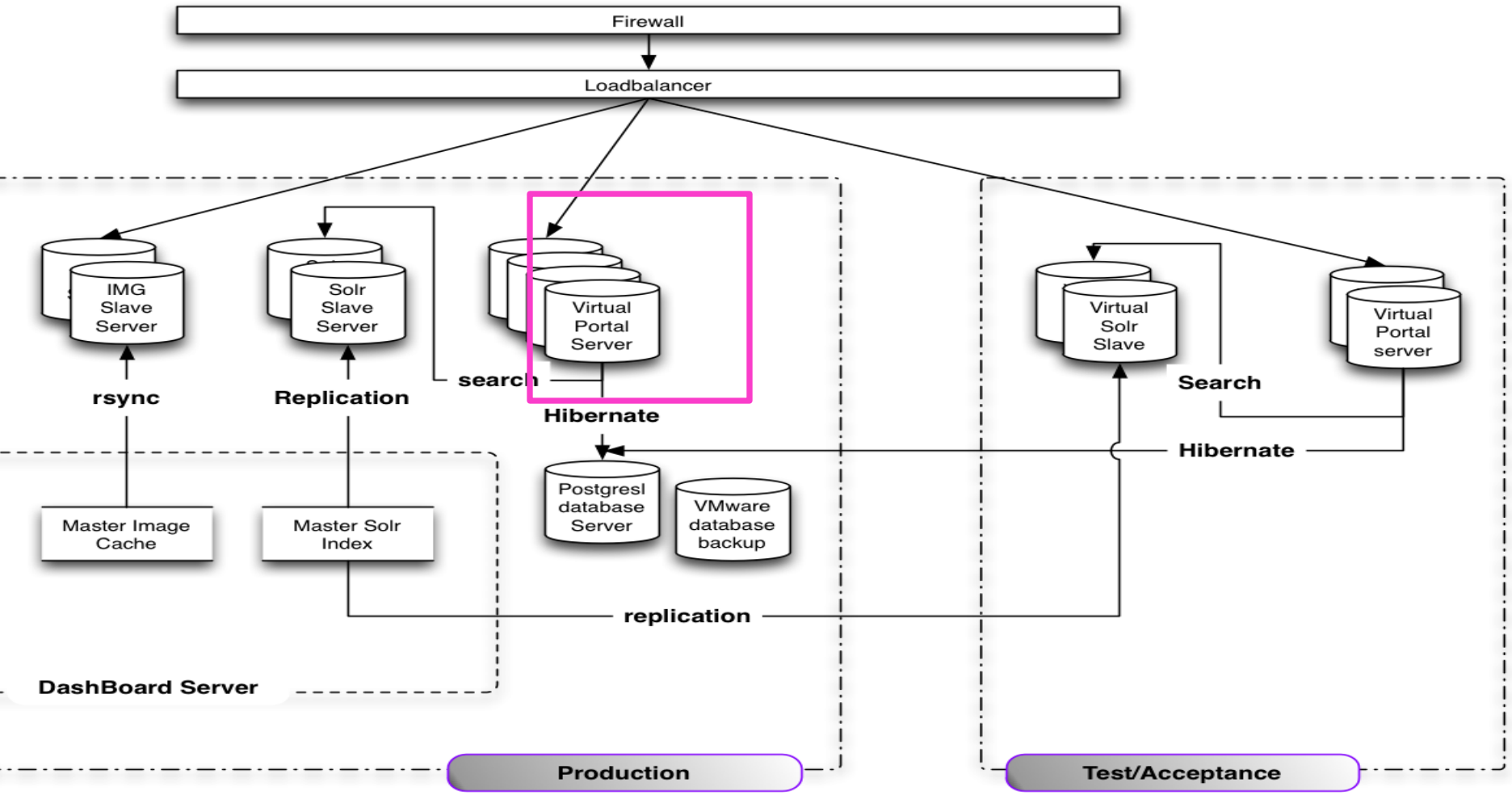
SOLR Search Server (1)



SOLR Search Server (2)

- Apache SOLR v1.3 http://lucene.apache.org/solr/
- REST-based interface for querying and updating the index
- Build in facetting, related item search, spellchecking 'did you mean', replication, result-caching, multiple output formats
- Fast and scalable indexing. Know to scale into billions of records
- Multithreaded indexing. Current single threaded indexing speed ca. 1 million records per hour.
- v1.4 additions: carrot2 clustering algorithms, autosuggestion, faster numerical range queries (TrieRange Queries) for temporal and spatial querying

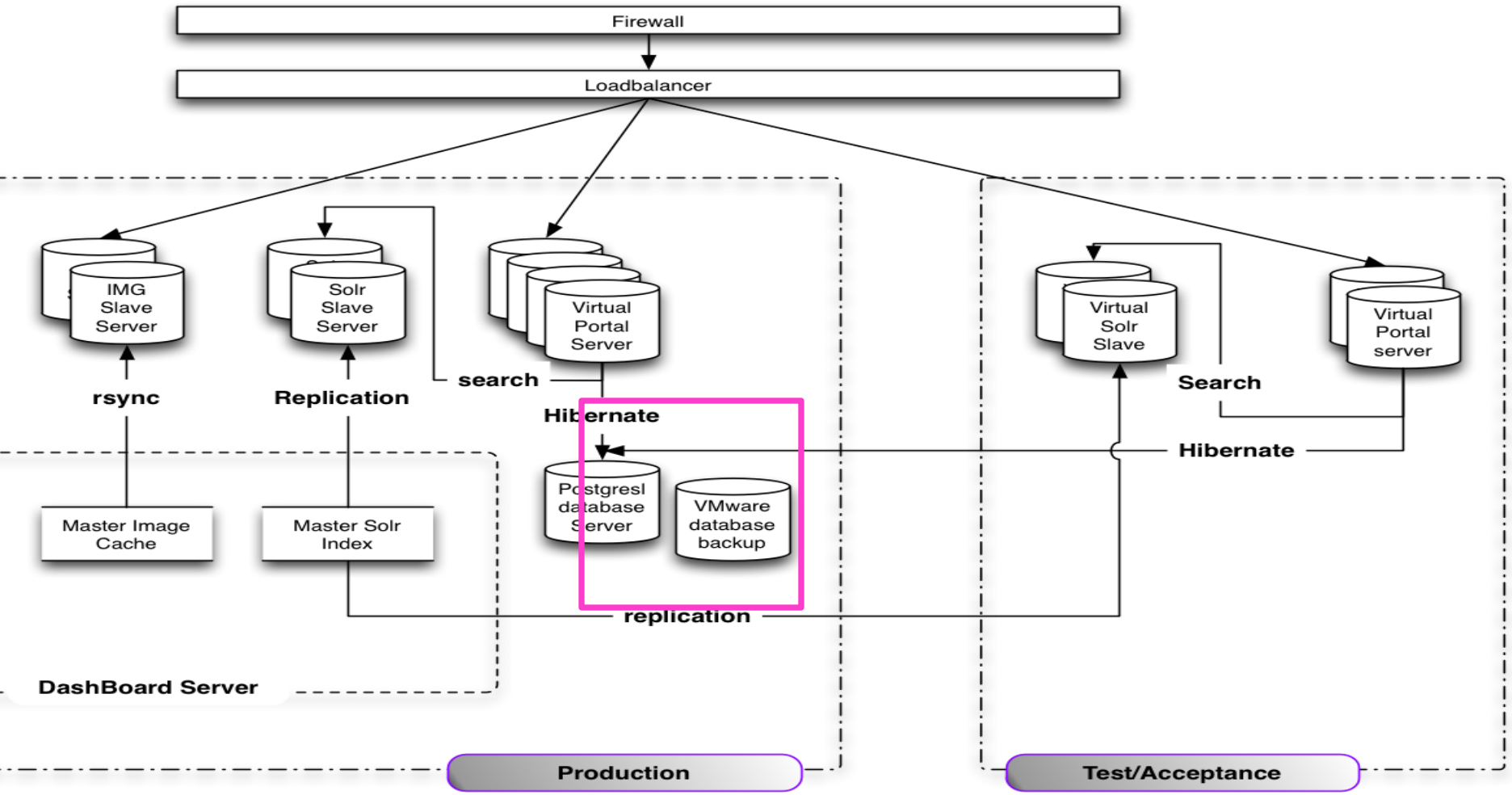
Portal front-end application (1)



Portal front-end application (2)

- Is the main Europeana Graphical User Interface:
  - Dynamically renders the pages
  - Reformulate queries of SOLR
  - Handles user authentification and localisation
- Statistics: served over 13 million pages and processed over 70 million http-request in the first 3 months of 2009

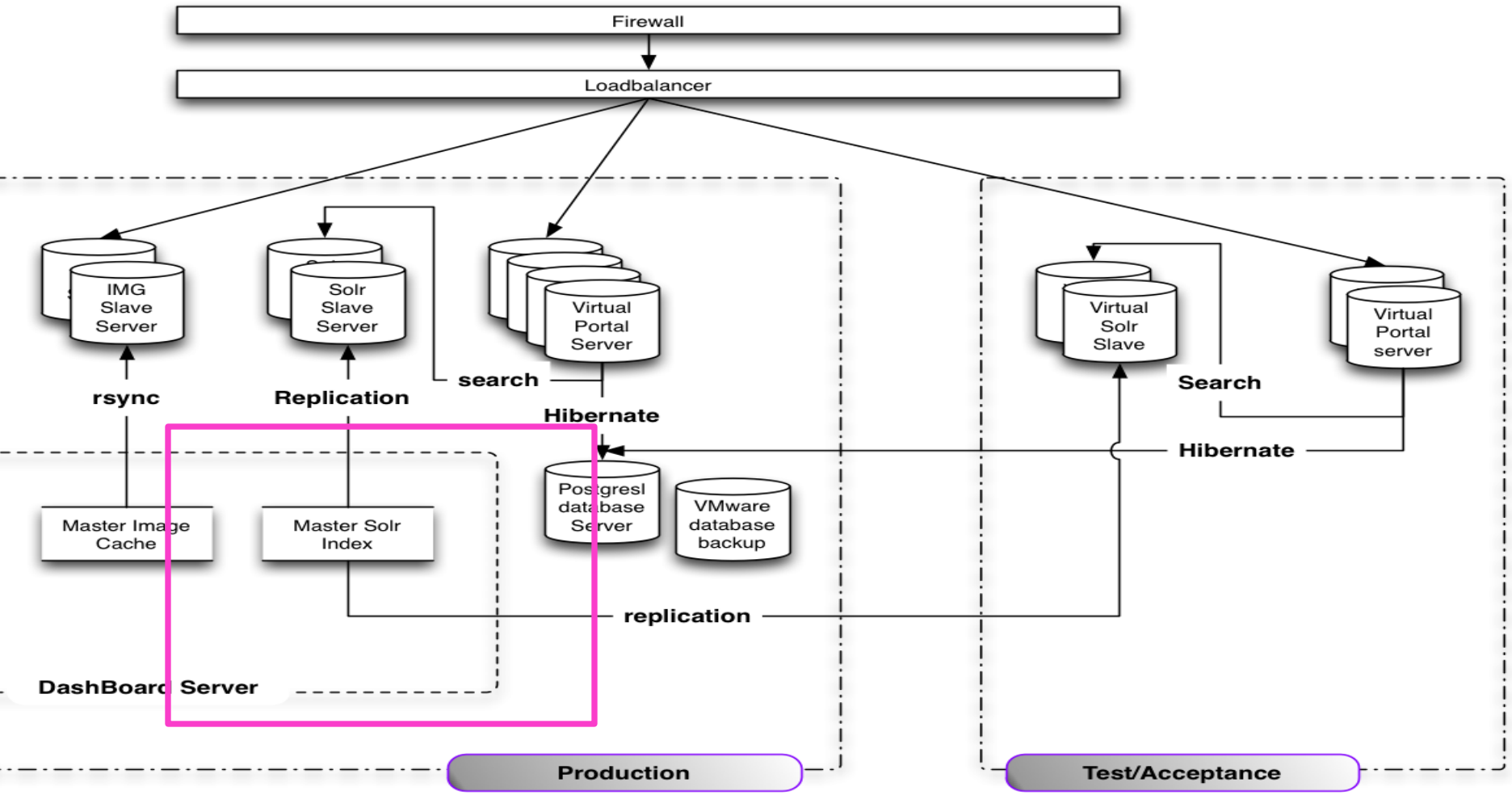
Database Server (1)



Database Server (2)

- Postgresql 8.3 database server
- Contains:
  - Europeana records, Europeana Objects, Users, Saved searches/tags/items, interface translations, static pages, content for partner page (i.e. partners and contributors), etc.
- Main usage:
  - Datastore for ingested material.
  - Uses europeana object table for caching process
  - Uses european records table for indexing process
  - Dynamically load interface translations and translated static pages (v0.3.1)
  - User validation and management

Dashboard Administration Interface (1)



Dashboard Administration Interface (2)

- Relative new comer to the Europeana prototype space
- Runs on the Dashboard Server
- Multi-user and multi-role adminstrative interface to:
  - Edit user roles
  - Collection management: importing, indexing, caching
  - Edit interface and staticpage translations
  - Manage partner and contributers list
  - Provide content related statistics
- Future extensions:
  - Ingestion wizard?
    - Upload, analyse, map, normalise, test content in sandbox, etc.

Questions?

Bram van der Werf
Technical Director **Europeana**
bram.vanderwerf@kb.nl