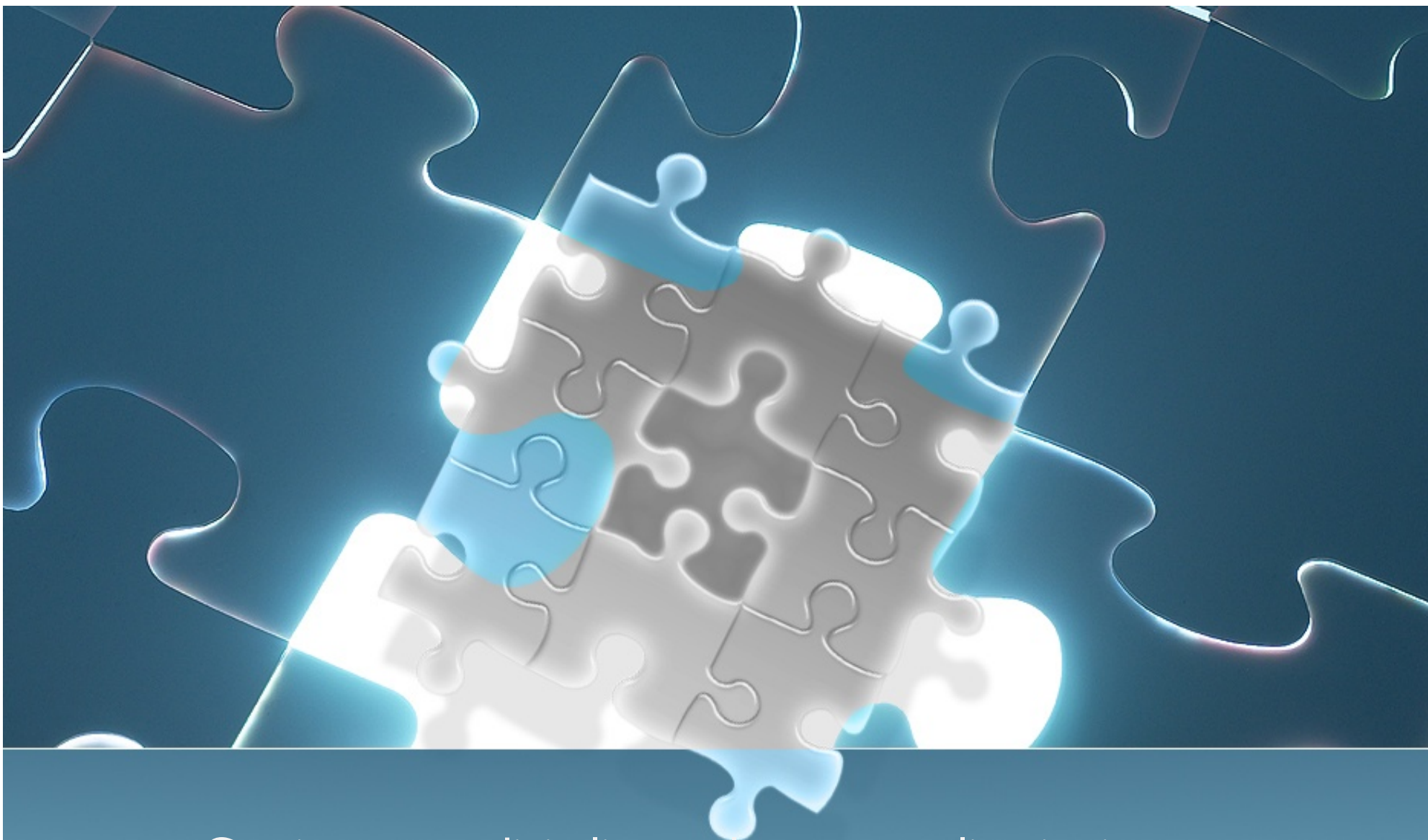




Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali

Costruire un **puzzle**... da dove cominciare?

- Quantità e varietà di tecnologie in gioco: **pezzi sul tavolo**
- Lavoro di ricerca nella comunità open source: **selezione**
- Adattamento alle esigenze specifiche: **incastro**





Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali

Introduzione

- descrizione progetto

Architettura

- hardware
- filesystem
- applicazioni

Infrastruttura di rete

- caratteristiche

Conclusioni



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali: introduzione

- L'Ufficio IT e la Biblioteca del Ceris CNR partecipano al progetto di digitalizzazione **BESS** - Biblioteca elettronica di scienze sociali ed economiche del Piemonte, finanziato dalla Compagnia di San Paolo, (2011-2012).
- A cura del Ceris CNR la gestione dei prodotti della digitalizzazione: il **repository** e il **front-end**



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali: introduzione

7000 volumi da digitalizzare in 2 anni

- Formato dei file nativi:
pdf;
tiff in alta risoluzione;
txt dell'ocr;
- **14 terabyte** di spazio disco previsto





Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali: introduzione

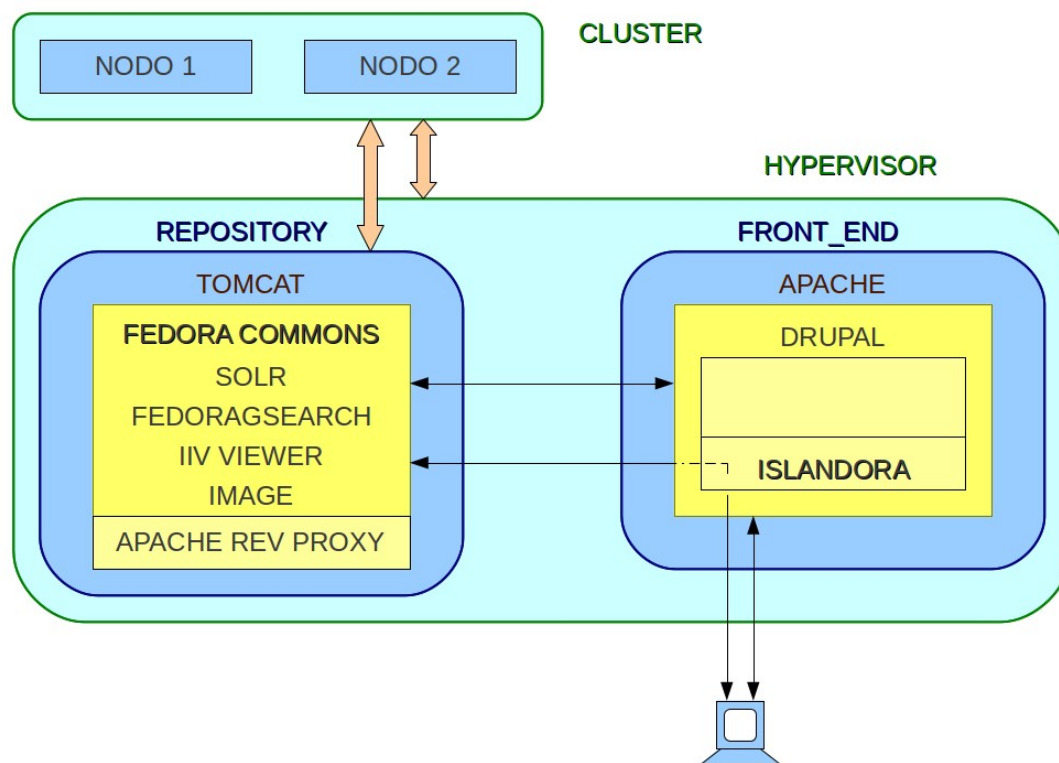
Tipologia dei volumi

- non più protetti da copyright (fino al 1930)
- copyright degli Enti appartenenti al gruppo Bess
- volumi protetti da copyright:
 - Richiesta scritta agli editori e/o agli autori di poter rendere disponibile la pubblicazione in forma digitale
 - Policy di accesso alle opere protette da copyright



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali: architettura

Schema architettura complessiva realizzata





Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali:

architettura

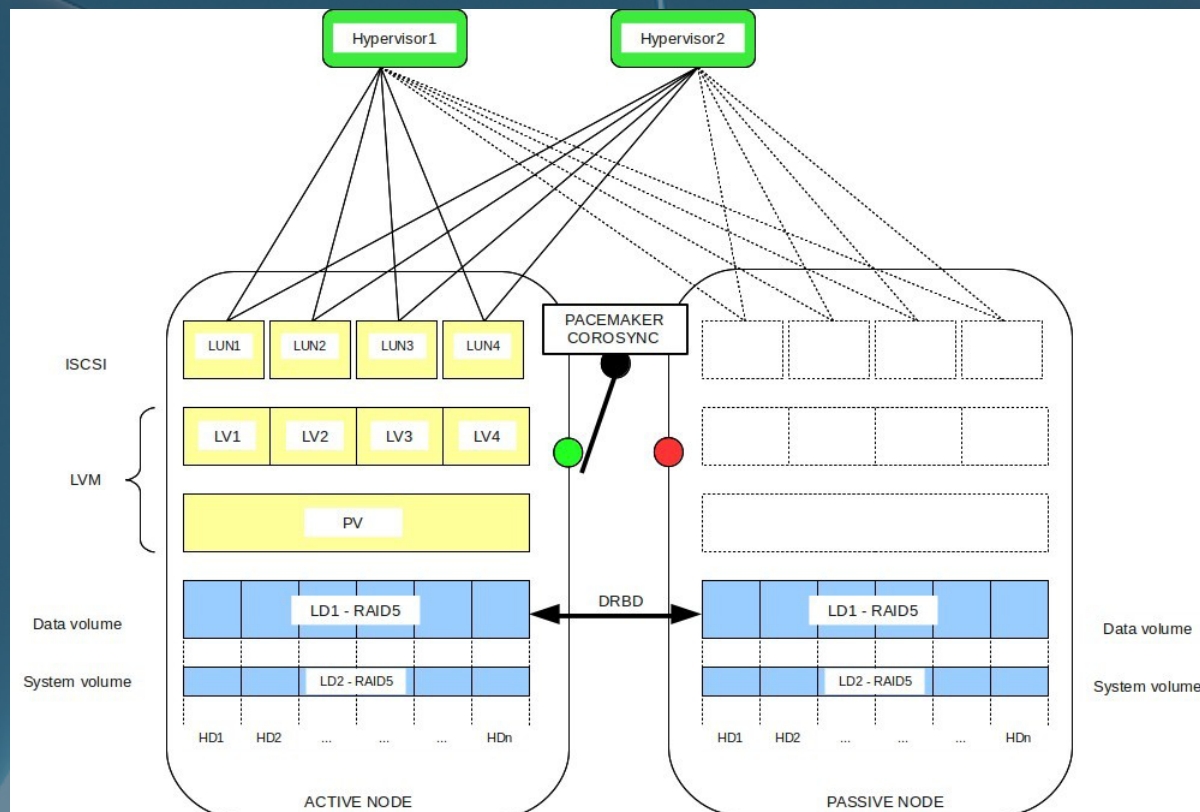
Hardware

- 1 **cluster** a due nodi attivo/passivo realizzato tramite software open-source
- 2 **Hipervisor** ridondati e connessi al cluster



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali: architettura

Schema a blocchi del cluster e hipervisor





Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali:

architettura

File system

- **Macchine virtuali** (VM) file system locale sull' hypervisor
- **Backup VM**, copia delle VM sul cluster via iSCSI
- **Dati**, file system sul cluster tramite iSCSI e Akubra data store per Blob (Binary large objects)



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali: architettura

Applicazioni

- Soluzione a **due server**
- Separazione applicazioni **Java/PHP**
- **Tomcat** esterno
 - Adore-djatoka, fedora, fedoragsearch, fop, iiv, imagemanip, saxon, solr



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali:

architettura

applicazioni

Repository: **Fedora Commons**

- Gestione e conservazione oggetti digitali
- Supporta elevate quantità di dati
- Oggetti composti da datastream
- Descrizione semantica delle relazioni tra gli oggetti
- Architettura a modelli espandibile e personalizzabile
- Harvesting OAI-PMH
- Operazioni di ingesting programmabili



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali: architettura

Repository : Fedora Commons

screenshot dei **datastreams** relativi ad una **pubblicazione**

Repository

Search: *

fgramsci:TO024-00001

Properties

Label: Sisifo 1

Created: 2011-10-11T13:23:49.388Z

Modified: 2011-10-11T13:23:49.388Z

Owner:

State: Active (A) ▼

Commit Changes

Export Object

View Object XML

Purge Object

Datastreams

ID	Label	MIME Type
DC	Dublin Core Record for this object	text/xml
RELS-EXT	RDF Statements about this object	application/rdf+xml
TN	Thumbnail	image/jpeg
INDEX	Index	text/plain
PDF	Reference PDF	application/pdf

Add Datastream

Refresh List



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali: architettura

Repository : Fedora Commons

screenshot dei **datastreams** relativi ad una singola pagina della pubblicazione

Repository

Search: *

fgramsci: T0024-00001-0001

Properties

Label: Sisifo 1 - page 0001

Created: 2011-10-11T13:23:54.483Z

Modified: 2011-10-11T13:23:54.483Z

Owner:

State: Active (A)

Commit Changes

Export Object

View Object XML

Purge Object

Datastreams

ID	Label	MIME Type
DC	Dublin Core Record for this object	text/xml
RELS-EXT	RDF Statements about this object	application/rdf+xml
TN	Thumbnail	image/jpeg
JP2	Compressed JP2	image/jp2
OCR	Scanned text	text/plain
TIFF	Reference HD image	image/tiff

Add Datastream

Refresh List



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali:

architettura

Applicazioni

Front-end: **Drupal**

- CMS, aperto, AAI

Framework : **Islandora**

- modelli e componenti che collegano gli oggetti del repository al front-end
- tutto nel repository

Piattaforma di ricerca: **Sorl**

- motore di indicizzazione e gestore delle ricerche



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali: architettura

Drupal, visualizzazione volumi: singola pagina

The screenshot displays the BESS Digital Library (demo) interface. The browser address bar shows the URL: `demo.bess-piemonte.it/fedora/repository/fgramsci:TO024-00001/-/Sisifo 1`. The page title is "BESS Digital Library (demo)". The navigation menu includes "BESS Digital Library". The search bar is labeled "Search" and contains the text "Title". The main content area is titled "Bess Digital Library" and features a "Read" button. Below the button, there is a thumbnail of a book cover titled "Sisifo" and a list of metadata: "idee ricerche programmi dell'Istituto Gramsci piemontese gennaio 1984 6.Q>2j PERCHE SISIFO di Mario Dogliani". The right sidebar contains a list of text: "er questa pubblicazione, che non vuole essere una rivista, ma, più semplicemente, una raccolta di brevi testi che indicano i temi sui quali stiamo lavorando, abbiamo scelto un nome mitico che, prima di essere banalizzato nel ritratto moralistico della sofferenza di una pena, era simbolo del lavoro: non di un lavoro trasfigurato, felice e pacificato, ma del lavoro così com'è, con la sua durezza e la sua «alienazione» (la terra che Sisifo lavora è quella degli eroi; i cui frutti non possono essere raggiunti - da Tantalo - per saziarsi). Questo titolo ci sembra contenere un messaggio chiaro: che concepiamo il nostro lavoro come uno sforzo continuativo, che sa di non avere altro senso che quello di aggiungere un granello infinitesimo, e forse". The footer of the page shows the URL: `http://demo.bess-piemonte.it/fedora/repository/fgramsci:TO024-00001/-/Sisifo 1#tabset-tab-2`.



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali:

architettura

Drupal, ricerca full text e filtri

BESS BIBLIOTECA ELETTRONICA DI SCIENZE SOCIALI ED ECONOMICHE DEL PIEMONTE

BESS Digital Library (demo)

Home > Chiamparino (x)

Search Results

(1 - 20 of 23)

Navigation

- Bess Digital Library

Search

FullText

Chiamparino

search

Filter

Description

- dell'istituto (23) + -
- gramsci (23) + -
- idee (23) + -
- piemontese (23) + -
- programmi (23) + -
- ricerche (23) + -
- 1988 (9) + -
- aprile (7) + -
- dicembre (7) + -
- 1986 (4) + -

Creator

- gramsci (23) + -
- istituto (23) + -
- piemontese (23) + -

Subject

- cultura (23) + -
- industria (23) + -
- innovazione (23) + -
- lavoro (23) + -
- piemonte (23) + -
- ricerca (23) + -

Sisifo 12 - page 0052

idee, ricerche, programmi dell'Istituto Gramsci piemontese. Dicembre 1987

... anni novanta di Sergio **Chiamparino** Organismi direttivi Comitato scientifico: Aldo Agosti, Arnaldo...

... Buran, Giampiero Carpo, Federico Cereja, Sergio **Chiamparino**, Claudio Dellavalle, Alfio Mastropaolo...

Sisifo 19 - page 0047

idee, ricerche, programmi dell'Istituto Gramsci piemontese. Aprile 1990

... "in negativo?"» Introduzione di S. **Chiamparino** 1 ° dicembre «Che cosa sono in grado di dire sulla...

Sisifo 14 - page 0023

idee, ricerche, programmi dell'Istituto Gramsci piemontese. Luglio 1988



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali: infrastruttura di rete

Ruolo della rete

Indispensabile per il **funzionamento**:

- Cluster DRBD
- Applicazioni Java
- Server web

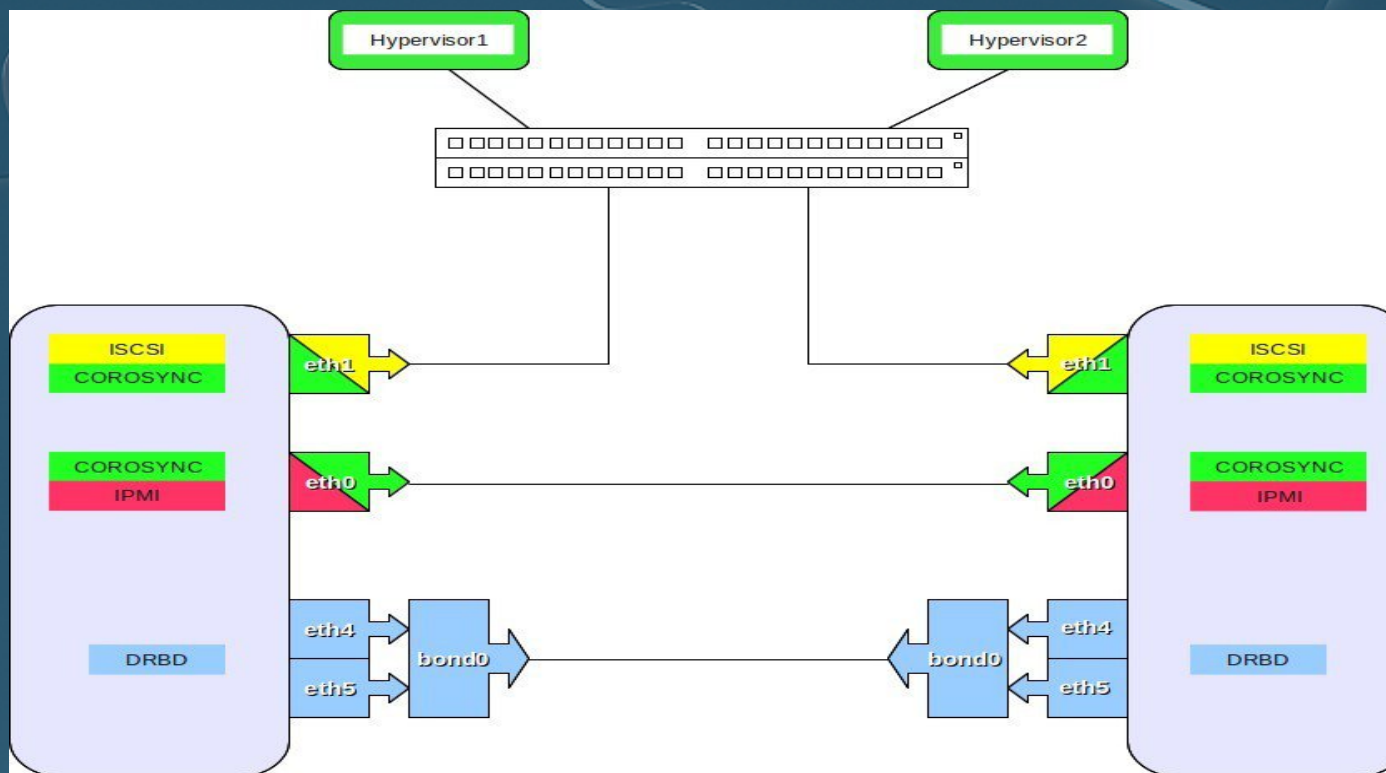
VLAN, IPv6

Sicurezza



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali: infrastruttura di rete

Connessioni di rete del cluster





Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali:

referimenti

Riferimenti

- Giancarlo Birello g.birello@ceris.cnr.it

Riferimenti bibliografici

- Rapporto tecnico [Ceris-CNR n.37](#)

Link

- <http://islandora.ca/>
- <https://wiki.duraspace.org/display/ISLANDORA/Islandora>
- <https://github.com/islandora>
- <http://fedora-commons.org/>
- <http://drupal.org/>
- <http://lucene.apache.org/solr/>
- <http://demo.bess-piemonte.it>
- <http://www.bess-piemonte.it>
- <http://www.to.cnr.it>



Gestire opere digitali con cluster e applicazioni open-source in ambienti virtuali:

conclusioni



Sviluppi futuri

