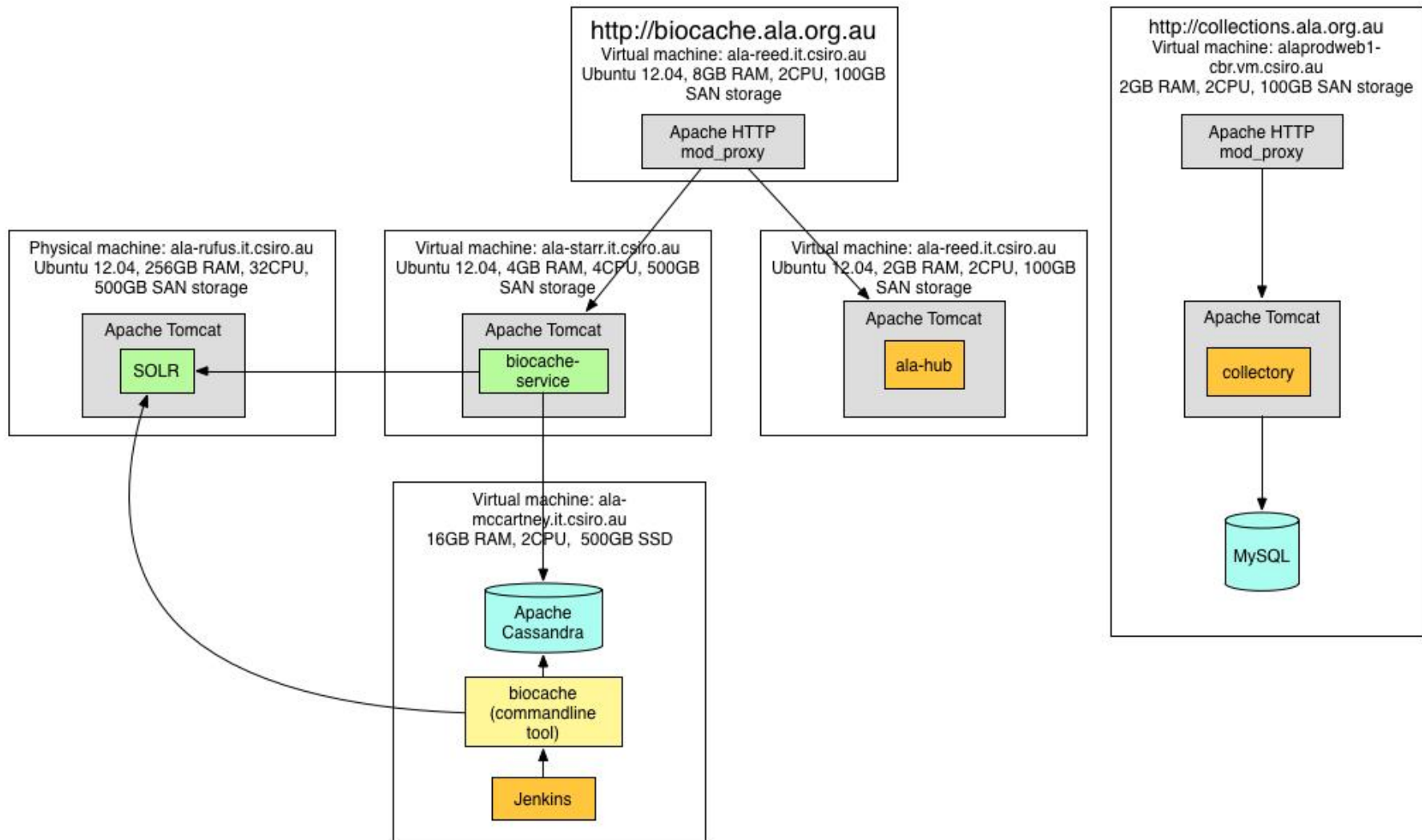# **Operation**

Paris workshop - May 2015
Dave Martin

Agenda

- Server requirements
- Data management
  - ALA
  - GBIF
- Database backups, index regeneration
- Resourcing

http://biocache.ala.org.au
Virtual machine: ala-reed.it.csiro.au
Ubuntu 12.04, 8GB RAM, 2CPU, 100GB
SAN storage

Apache HTTP
mod_proxy

http://collections.ala.org.au
Virtual machine: alaprodweb1-
cbr.vm.csiro.au
2GB RAM, 2CPU, 100GB SAN storage

Apache HTTP
mod_proxy

Physical machine: ala-rufus.it.csiro.au
Ubuntu 12.04, 256GB RAM, 32CPU,
500GB SAN storage

Apache Tomcat

SOLR

Virtual machine: ala-starr.it.csiro.au
Ubuntu 12.04, 4GB RAM, 4CPU, 500GB
SAN storage

Apache Tomcat

biocache-
service

Virtual machine: ala-reed.it.csiro.au
Ubuntu 12.04, 2GB RAM, 2CPU, 100GB
SAN storage

Apache Tomcat

ala-hub

Apache Tomcat

collectory

MySQL

Virtual machine: ala-
mccartney.it.csiro.au
16GB RAM, 2CPU, 500GB SSD

Apache
Cassandra

biocache
(commandline
tool)

Jenkins

Suggested specifications

- ## SOLR index server
  - 32GB+ instance, 640GB storage
- ## Cassandra server
  - 16GB+ instance, SSD is recommended
  - 4CPU, 8 or 16
- ## Collectory, Biocache services, Hub web app
  - 3 x 4GB instance, 32GB storage
  - 2CPU, Biocache service 4CPU or better

Data management

- ALA
  - Upload DwCA
  - URL endpoints
  - IPT Integration - create data provider
  - Automation - Jenkins
- GBIF
  - Uses GBIF web services

Database backups

- Collectory
  - mysqldump
- Cassandra
  - nodetool snapshot
- Both tools can be ran against live DB

Index regeneration

- biocache commandline tool
  - "`biocache bulk-processor index`"
  - Index generated to /data/biocache-reindex
- SOLR admin tool
  - Use to swap between indexes
- Jenkins

Ongoing operation - skills

- Relational DB
- Linux admin
- Familarity with managing:
  - java apps
  - tomcat
  - apache