Operation

Paris workshop - May 2015
Dave Martin
Agenda

- Server requirements
- Data management
  - ALA
  - GBIF
- Database backups, index regeneration
- Resourcing
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
● Familiarity with managing:
  ○ java apps
  ○ tomcat
  ○ apache