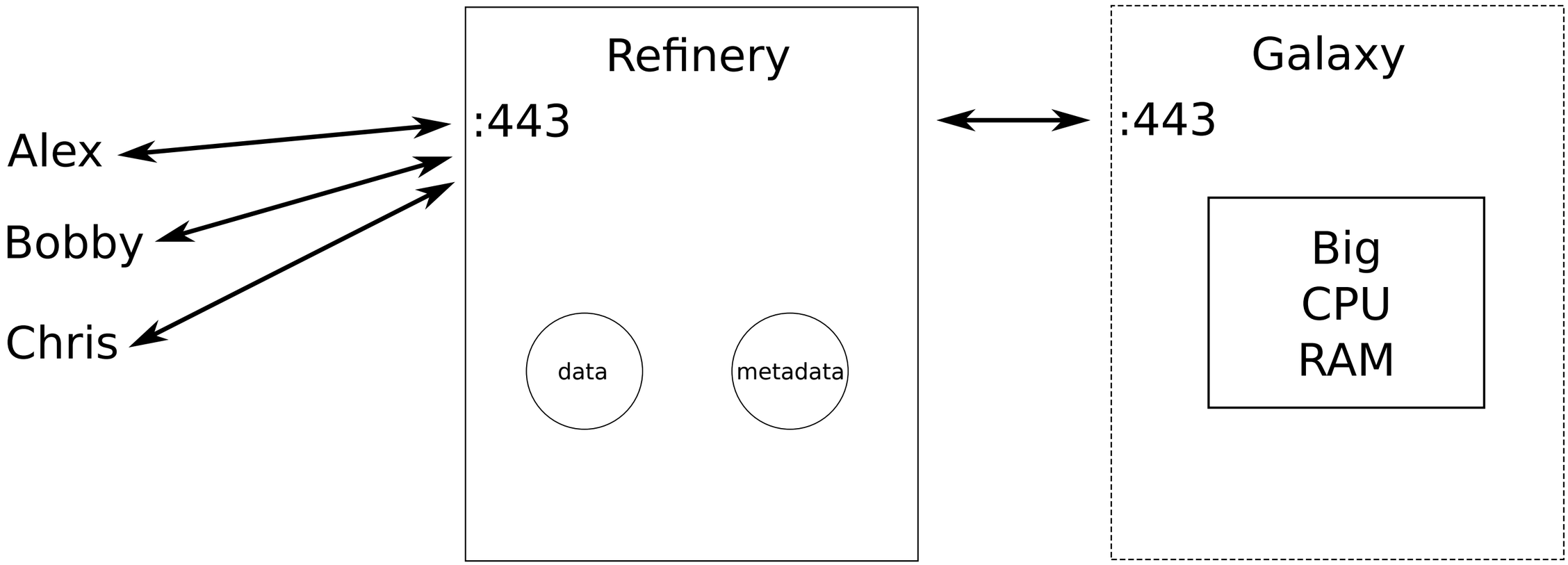
Refinery

A web based platform for
computational biology

Manage...
Analyse...
Visualise...

Data

Web App


Refinery
:443
Galaxy
:443
Alex
Bobby
Chris
Big
CPU
RAM
data
metadata

Inside Refinery

apache2 web requests

Django web framework

Python programming language

Celery scheduling

PostgreSQL persistent data

solr indexing

neo4j graph database

rabbitmq messaging

nodejs programming language

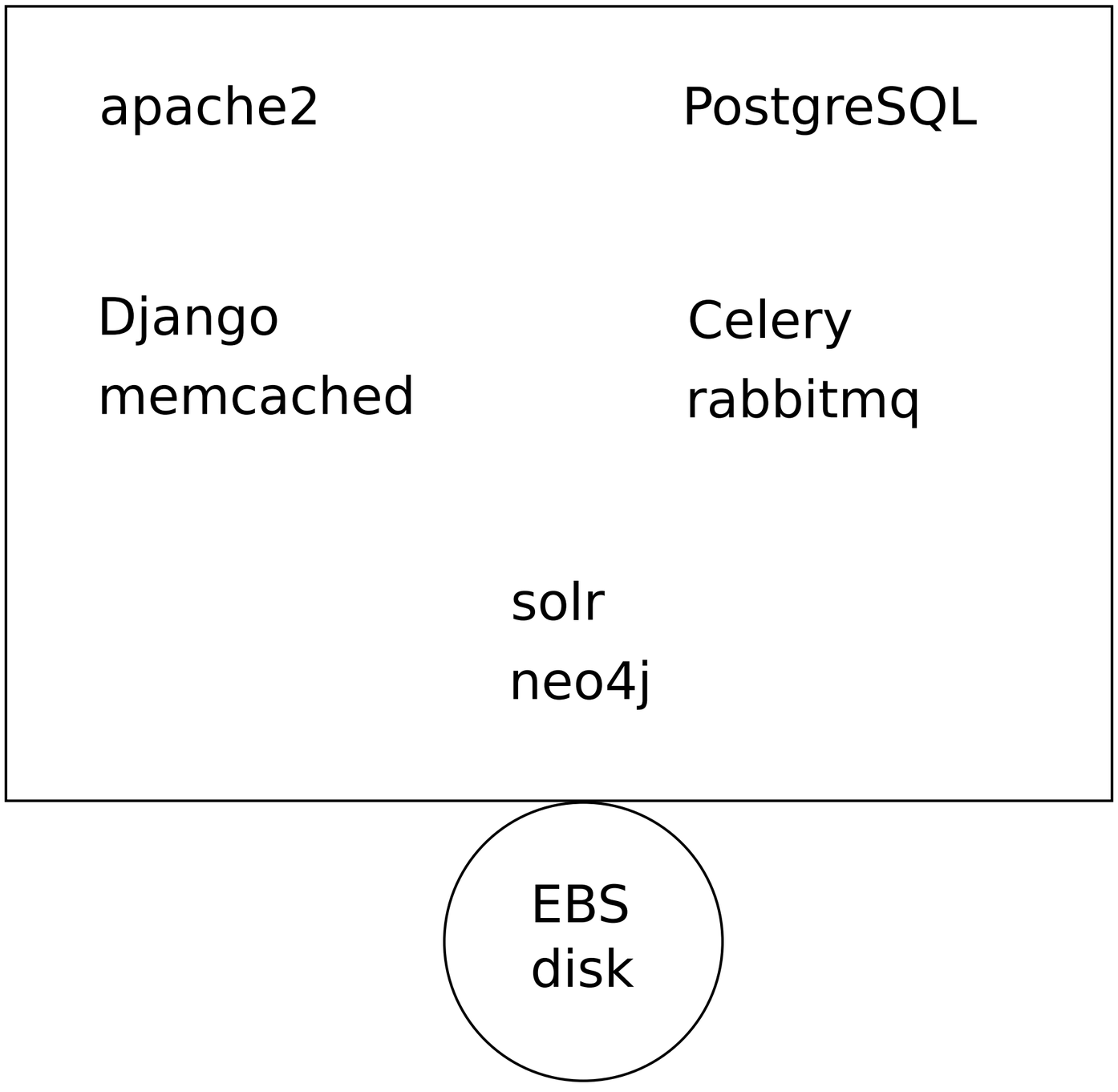
Angular frontend

JavaScript programming language

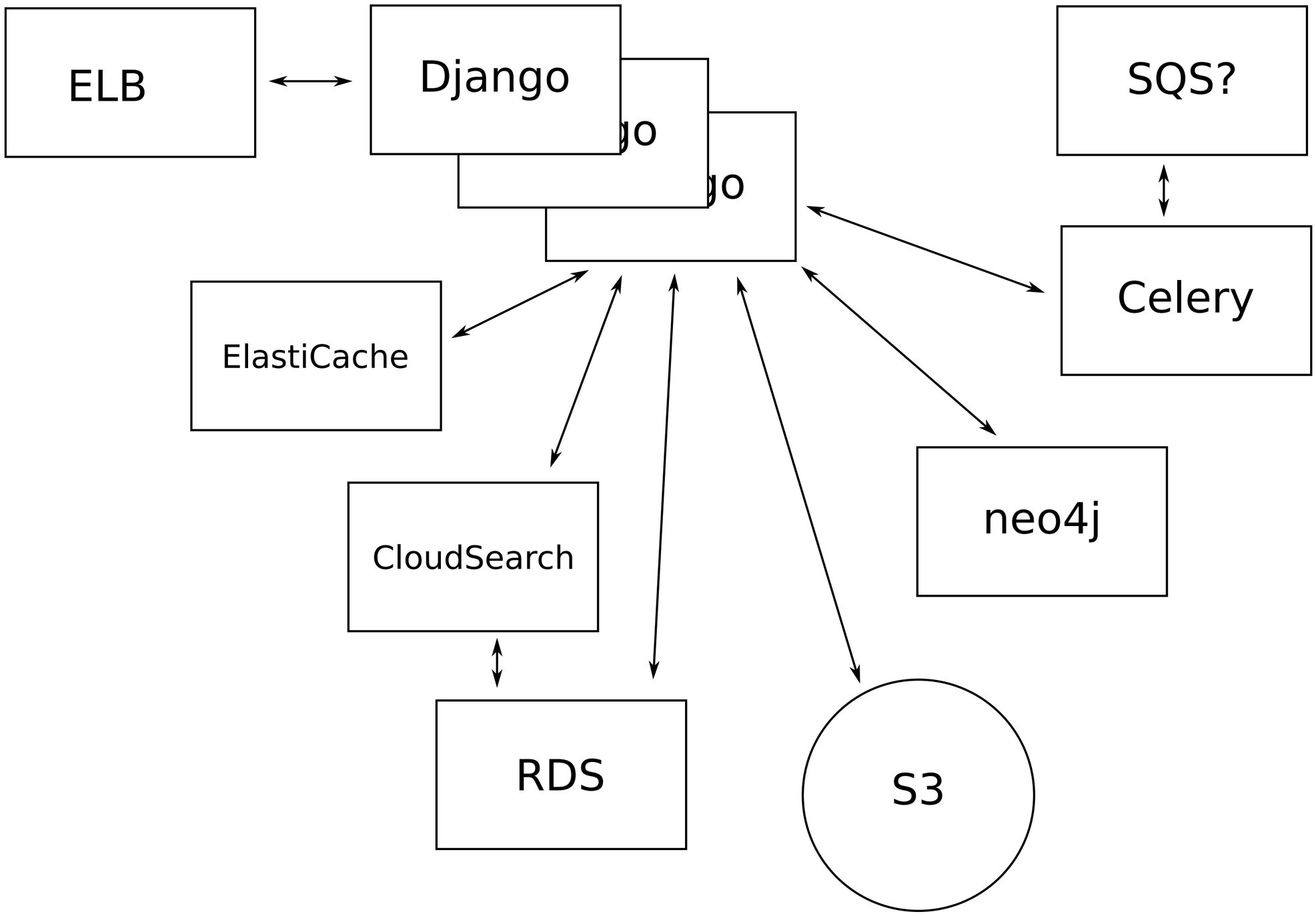
memcached caching

supervisord system services

On Amazon


apache2
PostgreSQL
Django
memcached
Celery
rabbitmq
solr
neo4j
EBS
disk

Amazon ∞


ELB
Django
go
go
SQS?
Celery
ElastiCache
CloudSearch
RDS
S3
neo4j

Amazon Minimal

apache2

Django
memcached

Celery
rabbitmq

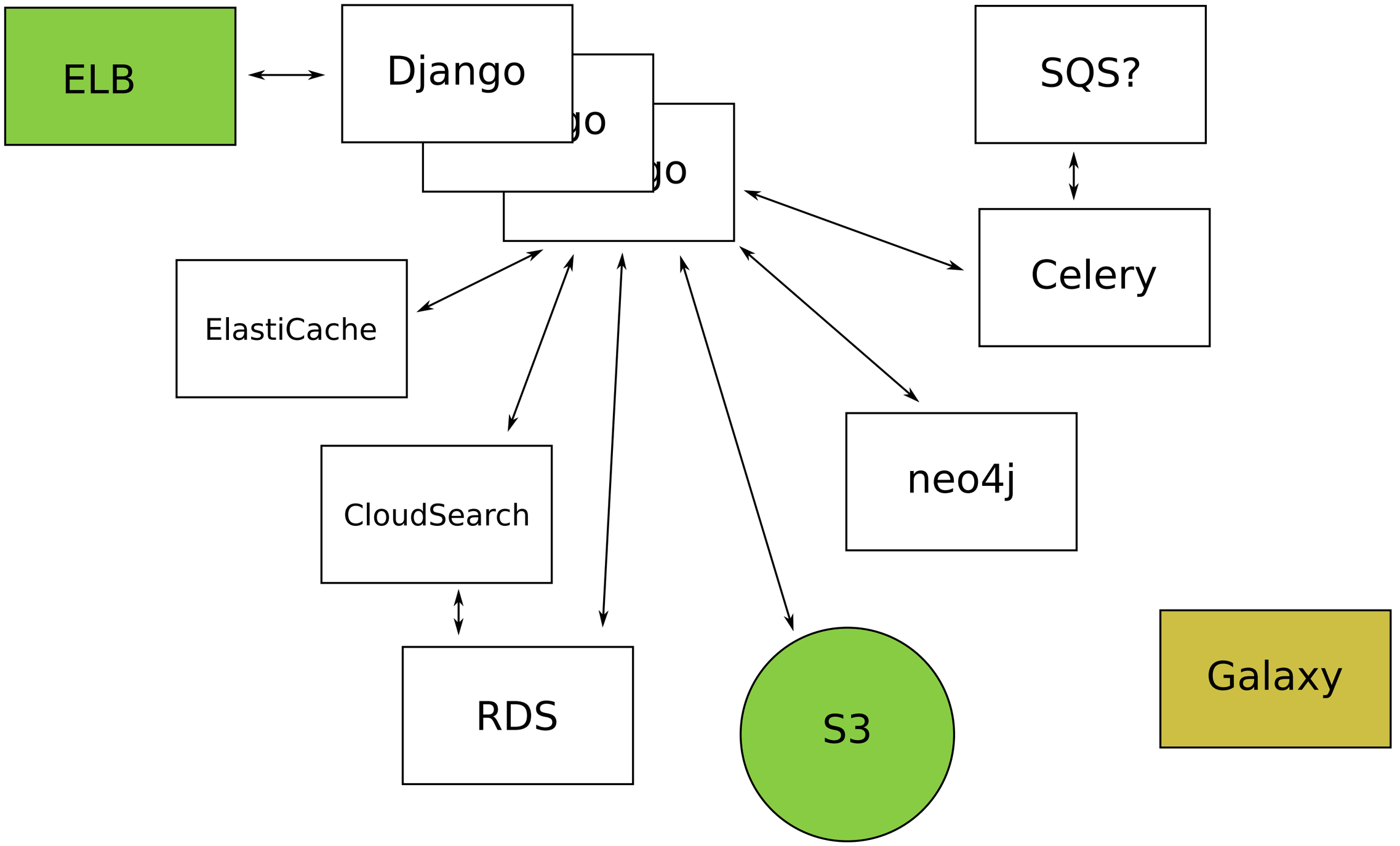
solr
neo4j

RDS

EBS
for code

EBS
for data

Amazon Priority



Risks

S3 - shuttling data in and out (and encrypting?) without losing it

Galaxy - reliable cloud burst

Backups - need to work even when managed by a bioinformatician

Monitoring - reliably tell when a Galazy job has finished/failed