

Extensión de un batch system local con Amazon Web Services + Amazon Elasticsearch

Jordi Casals, Carles Acosta, Fernando López, Vanessa Acín

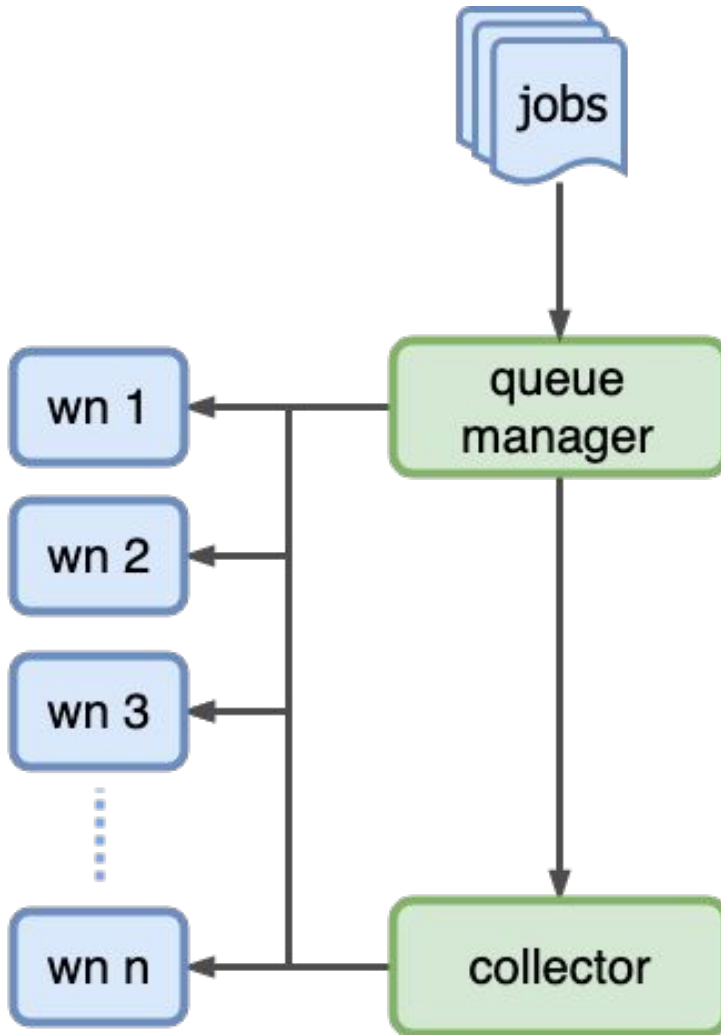
*Grupos de Trabajo de RedIRIS - 28 Mayo
Sevilla*

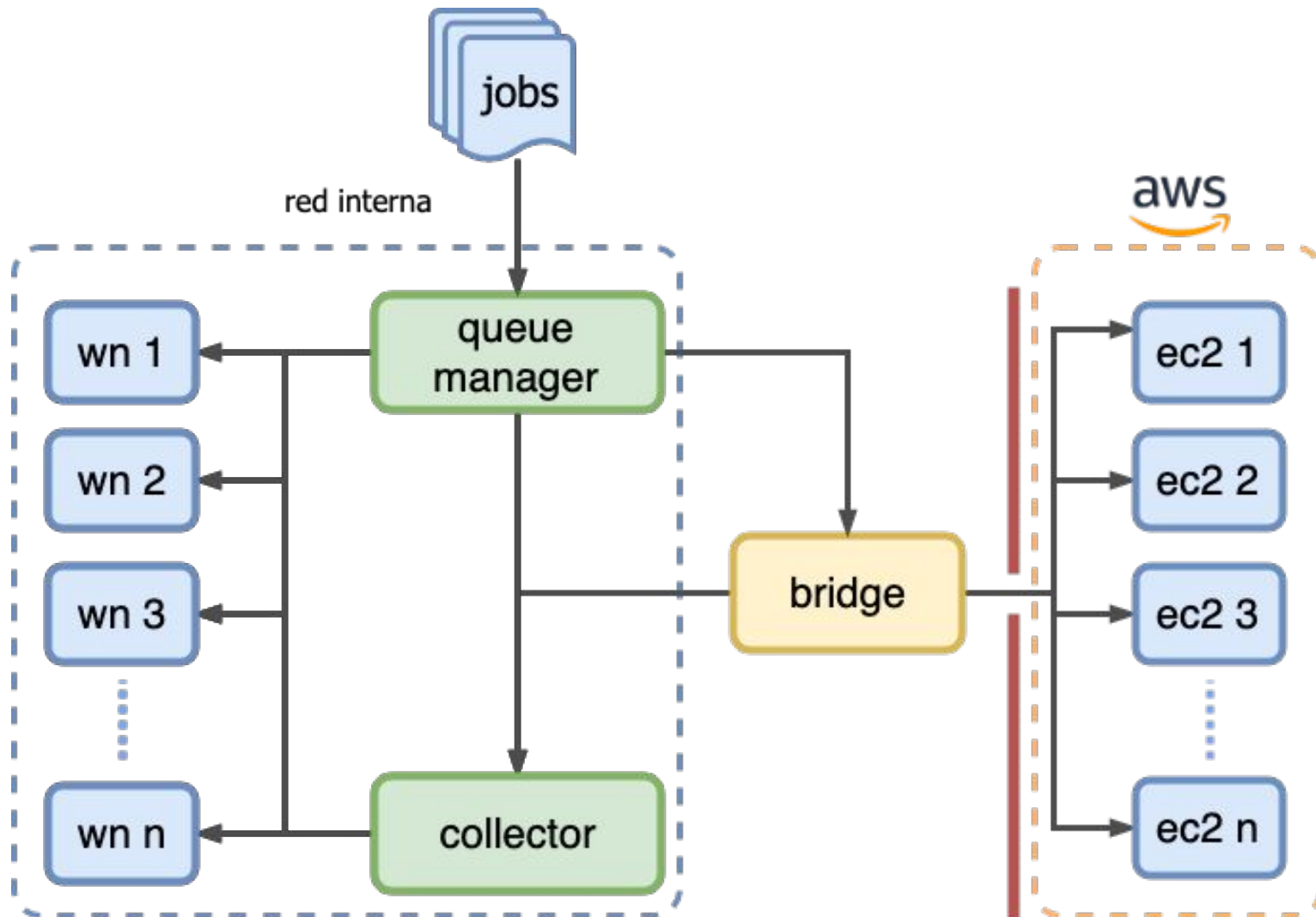
- 22 personas
- Tier 1 del **L**arge **H**adron **C**ollider (ATLAS, CMS, LHCb)
- Tier 2 y Tier 3 de ATLAS
- Proyectos cosmología y astronomía (MAGIC, CTA, PAU, Euclid, DES)
- Disco → 10PB
 - dCache, sistema de ficheros distribuido
- Cinta magnética → 26PB
 - Robot Storagetek SL8500 con software Enstore
- Computación → 300 servidores, ~8200 slots
 - PBS
 - **HTCondor**
- Big Data → Hadoop 16 nodos, 448 cores, 287 TB

- Evaluar el coste de uso de recursos cloud
 - Especialmente el uso de Spot instances de AWS
 - Probar la integración de un entorno cloud con el batch system del centro
 - HTCondor dispone de integración nativamente
- Evaluar mover el clúster de Elasticsearch a un entorno gestionado

- Experimentos LHC pueden llegar a “monopolizar” los recursos
 - ~ 7000 slots ocupados y ~ 8000 en cola constantemente
 - Experimentos más pequeños pueden verse afectados
- Posibilidad de crecer puntualmente
 - Sin necesidad de comprar y configurar máquinas
- HTCondor ofrece comando para conectar “fácilmente” con AWS
 - Especial interés en el uso de **spot-instances**





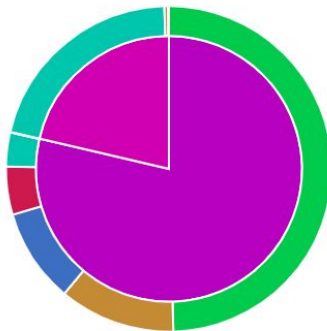


- Elasticsearch es una herramienta de búsqueda de texto completo y de manera distribuida
 - Muy útil para análisis de logs (*entre otras cosas*)
- Kibana como frontend para visualizar los datos
- Puede recibir datos de un servidor de Logstash
 - Recolecta, parsea e indexa logs
- Servicio totalmente gestionado por Amazon Web Services
 - Versión, número y tipo de instancias, tamaño de disco y tipo de acceso (Público privado o limitado por rango de IP)
 - Nuestro clúster local de Elasticsearch nos daba algunos problemas de rendimiento y de durabilidad de datos
 - Valorar si vale la pena usar el servicio o pagar uno gestionado
- Más de 1.000.000.000 de líneas de log en ~ 650GB de disco

Total Written

212,837
Total Writes

Reads / Writes by Protocol



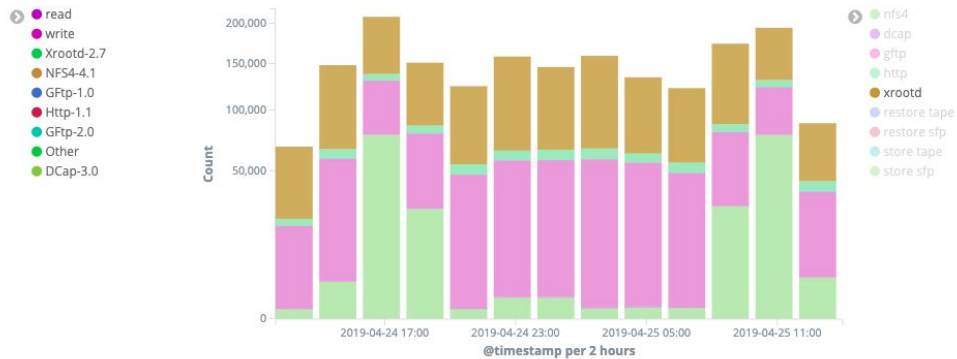
149.73TB
Total Written

Total Read

791,794
Total Reads

394.136TB
Total Read

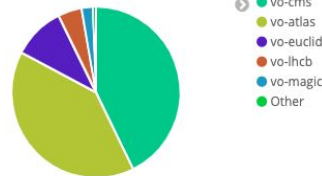
Events Over Time By Protocols



Top 10 Clients

No results displayed because all values equal 0.

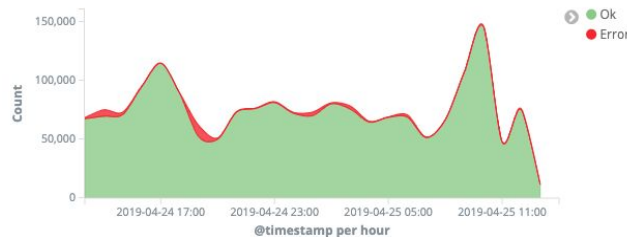
Top 5 Storage Groups



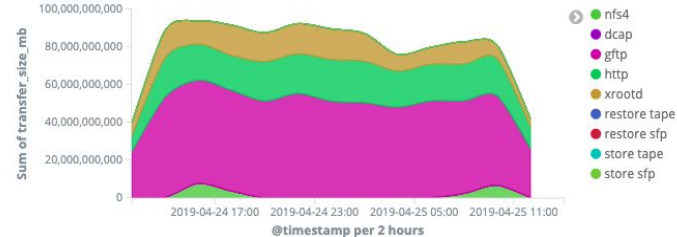
Top 10 Pools



Errors



Transfer Size by Protocols (MB)



- Modelo de contratación: bajo el acuerdo marco GÉANT Cloud IaaS
- Gestión desde el departamento IT:
 - Seleccionar la empresa → Sparkle
 - La empresa nos envía la documentación necesaria:
 - Call-Off Agreement (acuerdo marco)
 - Plantilla del contrato
- Gestión con nuestro departamento de contratación:
 - Documentación interna requerida:
 - Documentación que certifique que RedIRIS puede acogerse al acuerdo marco de GÉANT
 - Documentació acreditativa de que Sparkle fue una de las empresas seleccionadas por el concurso público de GÉANT
 - Datos de afiliación de nuestra entidad con RedIRIS.

¡Gracias!



RedIRIS



SPARKLE

Jordi Casals, Carles Acosta, Fernando López, Vanessa Acín

*Grupos de Trabajo de RedIRIS - 28 Mayo
Sevilla*