Science & Technology Facilities Council

UK Research and Innovation

Elastic Stack at RAL

Greg Corbett, James Adams

Scientific Computing Department

STFC Rutherford Appleton Laboratory

Elastic Stack at RAL

- Why we have an Elastic Stack
- Elastic Stack setup
- Combining data from multiple sources

Why we have an Elastic Stack

- Original use case
  - Monitoring/event logging for CASTOR
  - One day generated ~33 million events
- Now 12+ services use it
  - For monitoring, audit log storage and metric gathering
  - One day generates ~200 million events

Elastic Stack setup

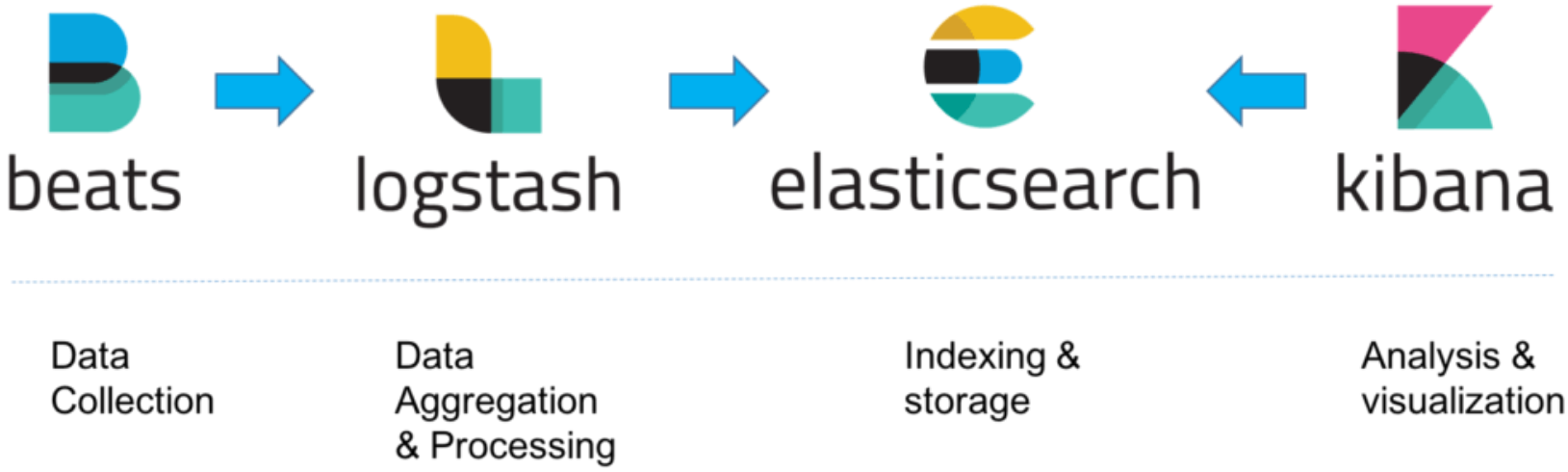
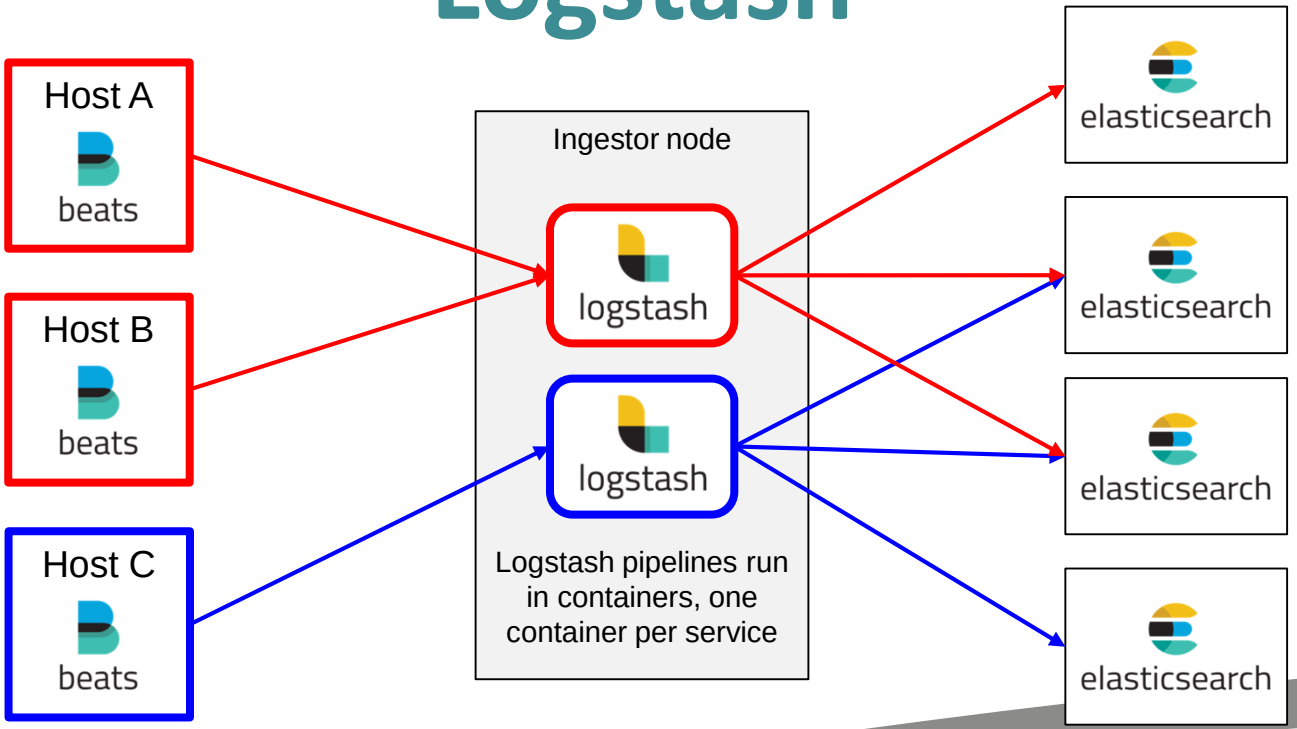


Image source: https://logz.io/learn/complete-guide-elk-stack/

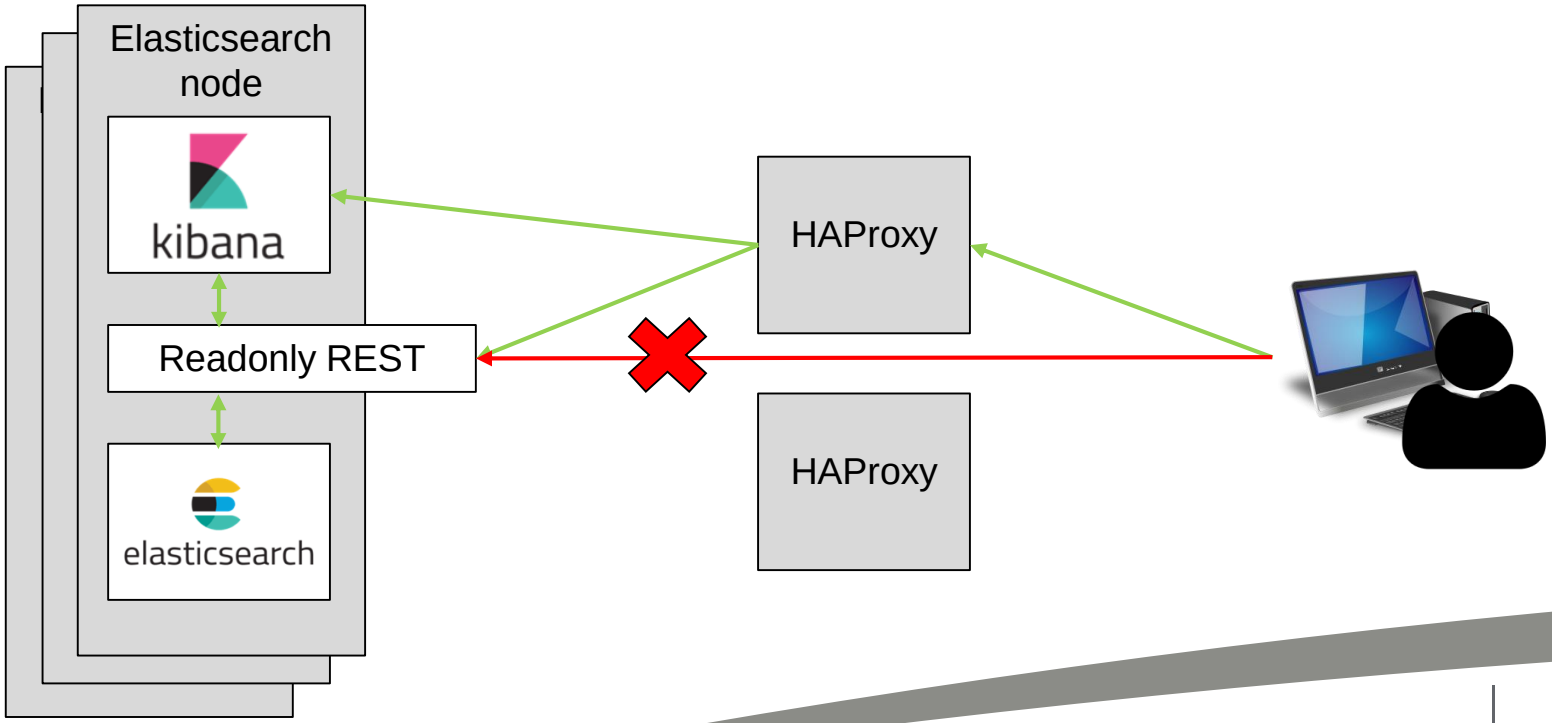
Logstash


Host A
beats
Host B
beats
Host C
beats
Ingestor node
logstash
logstash
Logstash pipelines run in containers, one container per service
elasticsearch
elasticsearch
elasticsearch
elasticsearch

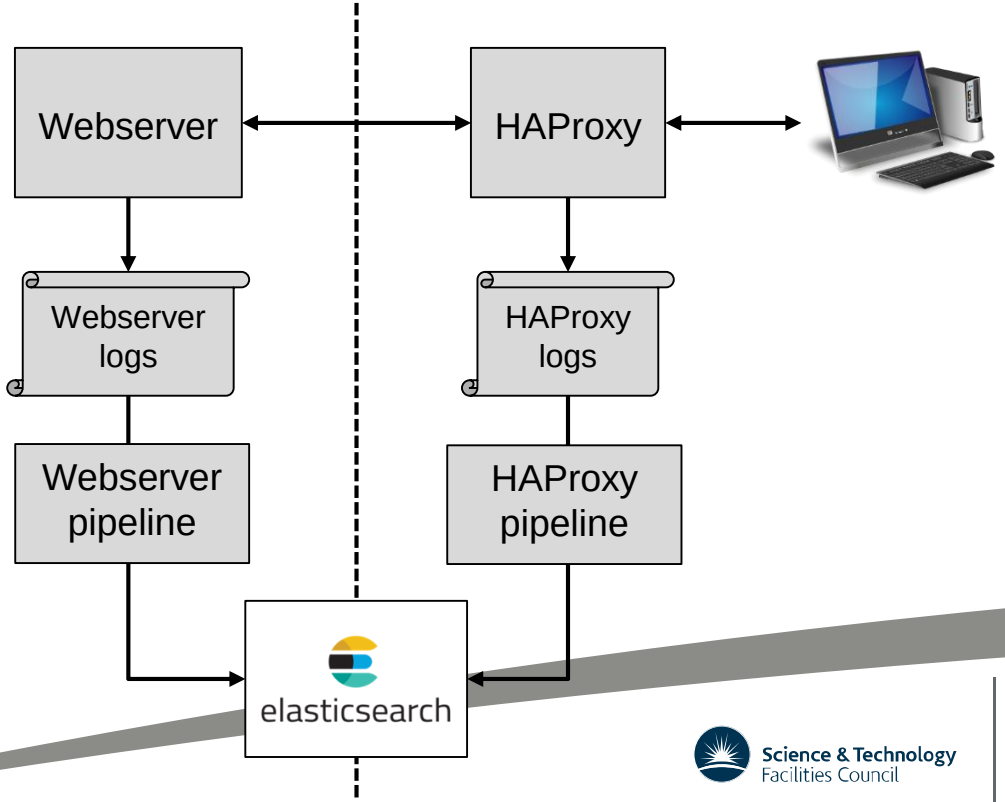
Elasticsearch

- Optimized for ingestion, trading off refresh rate for throughput
  - Ingesting thousands of events per second as opposed to a handful of searches every few minutes
- 3 shards per index, keeps number of shards per node to less than 20 per GB of heap (~600)
- We use Readonly REST to manage ACLs and enforce HTTPS

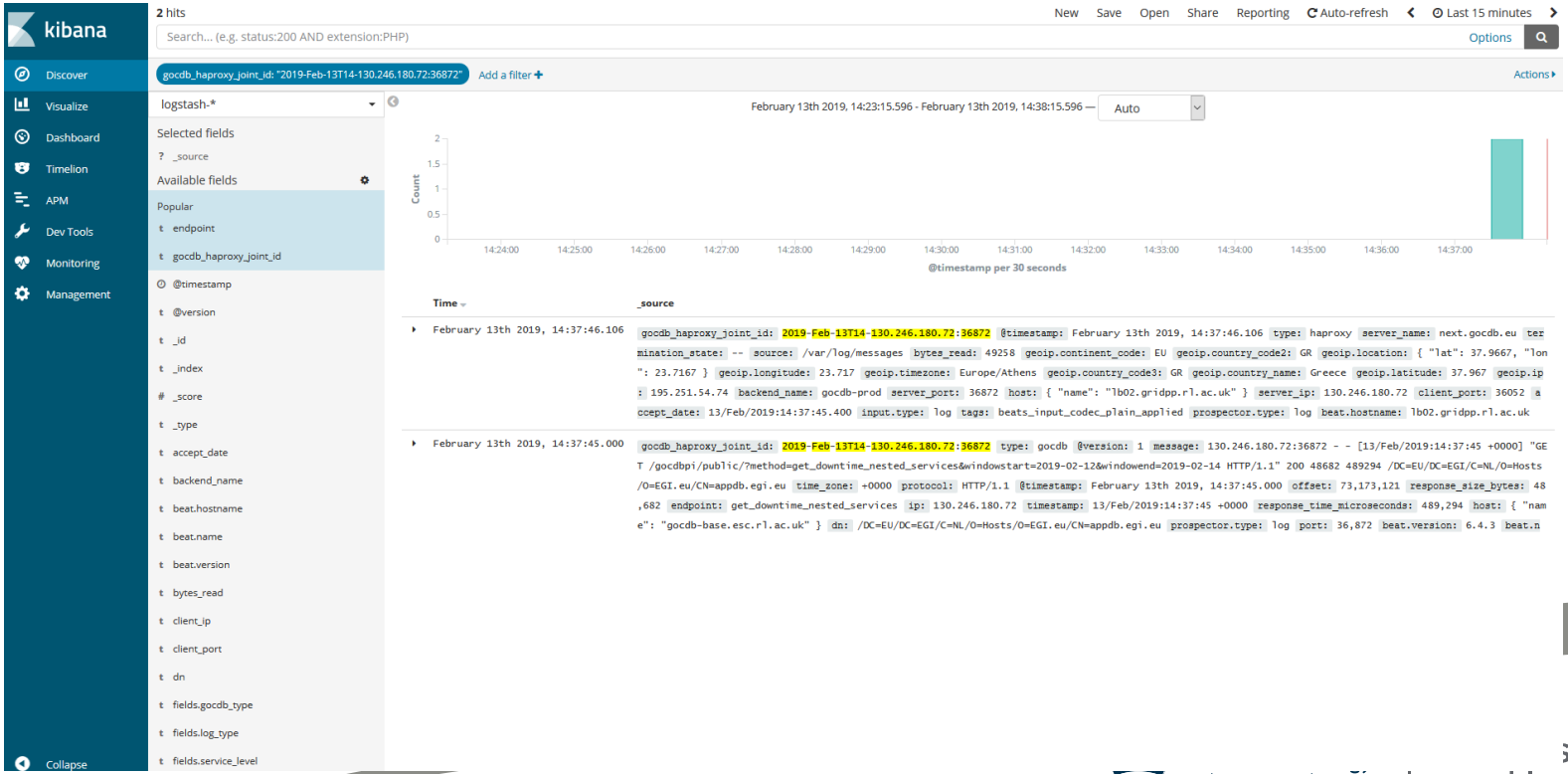
Elasticsearch and Kibana



Data from Multiple Sources



Data from Multiple Sources



Facilities Council

search and Innovation

Questions?

Further Reading

Scaling Elasticsearch

https://indico.cern.ch/event/391769/contributions/1827758/attachments/784181/1074960/2015-06-02-hepsysman-jrha-elasticsearch-scaling.pdf

How many shards should I have in my Elasticsearch cluster?

https://www.elastic.co/blog/how-many-shards-should-i-have-in-my-elasticsearch-cluster