elastic

Introduction into Elasticsearch Ingest Node

Alexander Reelsen
@spinscale

What?

- Elasticsearch did not have any possibility to enrich JSON before indexing
- Logstash usually takes over the part of document enrichment
- Getting apache logs required a full ELK setup
- Getting data from a beat to Elasticsearch required logstash in between

- What if we had a little bit of enrichment power in Elasticsearch?

Will logstash be replaced?

elastic

Definitions

- Pipeline
  - Guide to document enrichment
  - Stored inside ClusterState
  - Index operations can have a pipeline configured
  - A pipeline consists of a series of processors
- Processor
  - A single step to change a document
  - Configurable as part of a pipeline

APIs

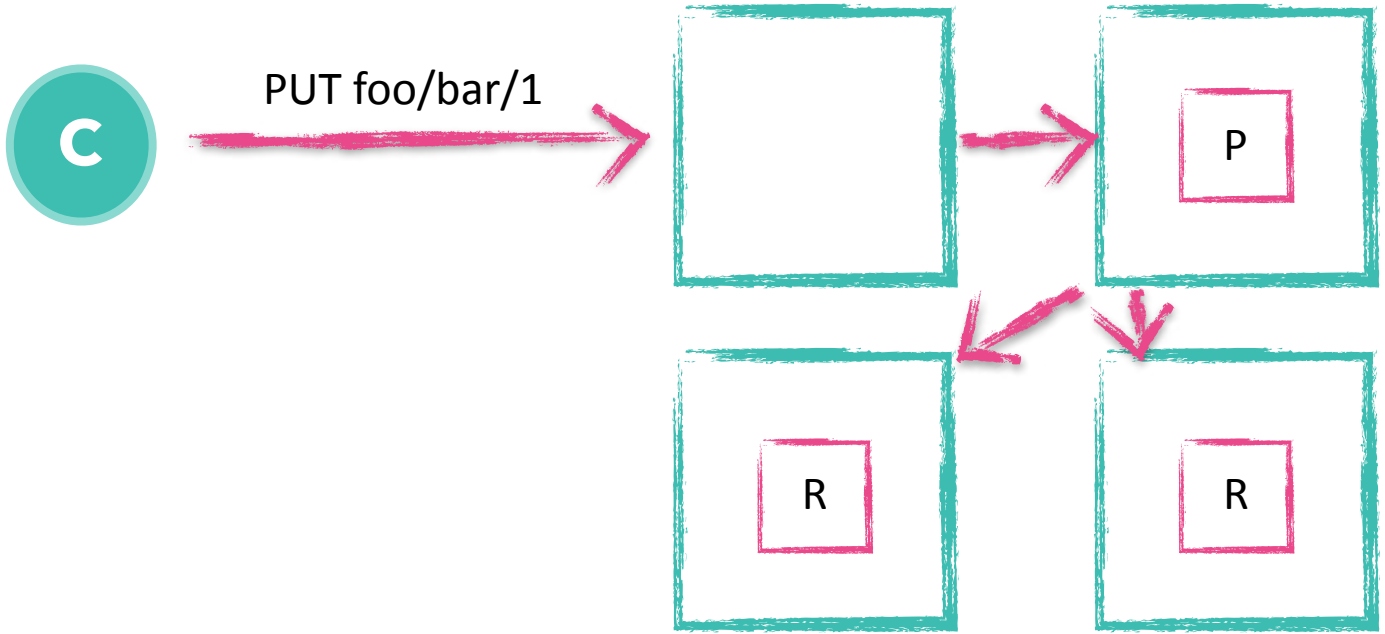
- `PUT _ingest/pipeline/my-pipeline-id`
- `GET _ingest/pipeline/my-pipeline-id`
- `DELETE _ingest/pipeline/my-pipeline-id`
- `POST _ingest/pipeline/_simulate`

Processors

- Append, Convert, Date, Date Index Name, Fail
- Foreach, Grok, Gsub, Join, JSON, KV, Lowercase
- Remove, Rename, Script, Set, Split, Sort, Trim, Uppercase, Dot Expander

- Plugins: useragent, geoip, attachment

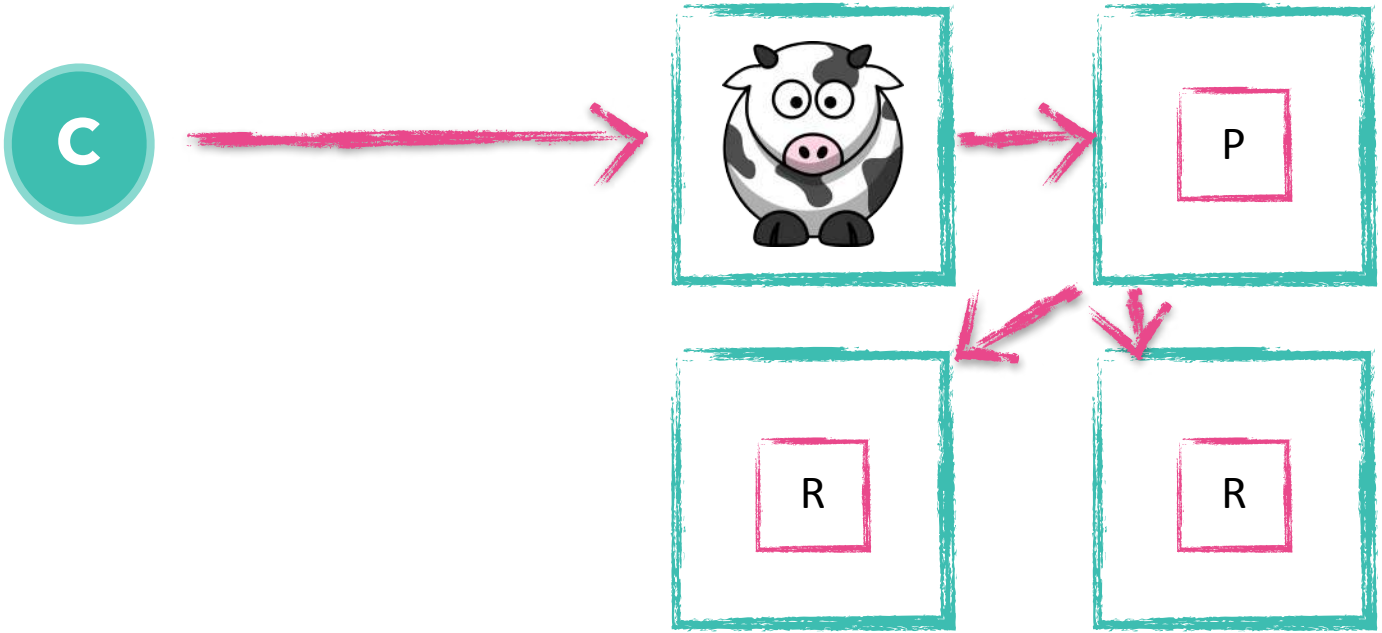
Ingestion inside of a cluster



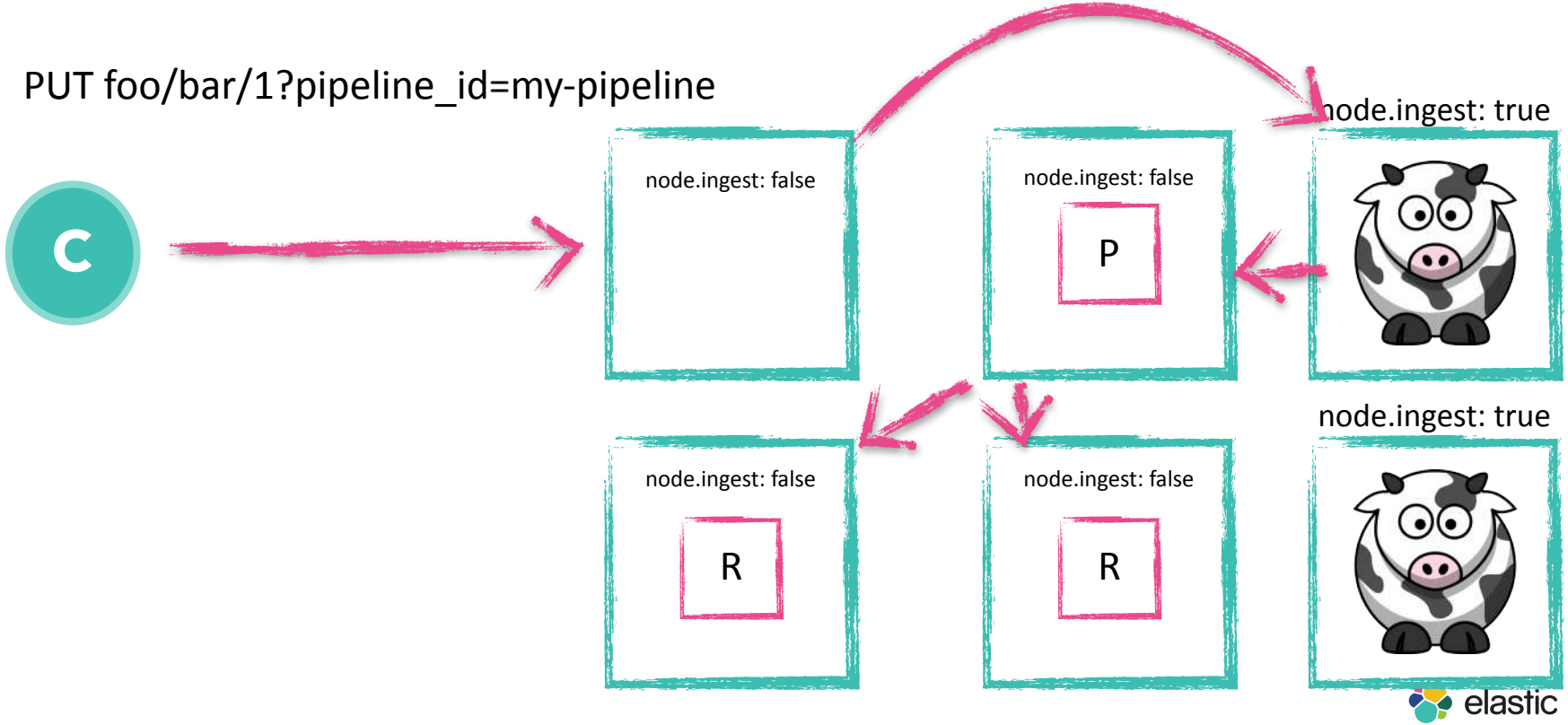
elastic

Ingestion inside of a cluster

PUT foo/bar/1?pipeline_id=my-pipeline



dedicated ingest nodes



elastic

elastic

Demo: Using pipelines

elastic

Writing your own processor

Writing your own processor

- Processors can be written as own plugins
- Use any JVM language
- Processors are fully unit testable!
- Beware of the security manager!

elastic

Demo: Writing your own processor

elastic

Further reading

Further reading

https://speakerdeck.com/elastic/ingest-node-enriching-documents-within-elasticsearch

https://www.elastic.co/guide/en/elasticsearch/reference/master/ingest.html

https://www.elastic.co/blog/ingest-node-a-clients-perspective

https://www.elastic.co/guide/en/elasticsearch/reference/master/ingest-apis.html

https://www.elastic.co/blog/new-way-to-ingest-part-1

https://www.elastic.co/blog/ingesting-and-exploring-scientific-papers-using-elastic-cloud

https://www.elastic.co/blog/writing-your-own-ingest-processor-for-elasticsearch

https://github.com/spinscale/elasticsearch-ingest-opennlp

https://github.com/spinscale/elasticsearch-ingest-langdetect

https://github.com/spinscale/cookiecutter-elasticsearch-ingest-processor

Questions?