BETTER DATA EXPLORATION AND VISUALISATION WITH ELASTICSEARCH AND KIBANA

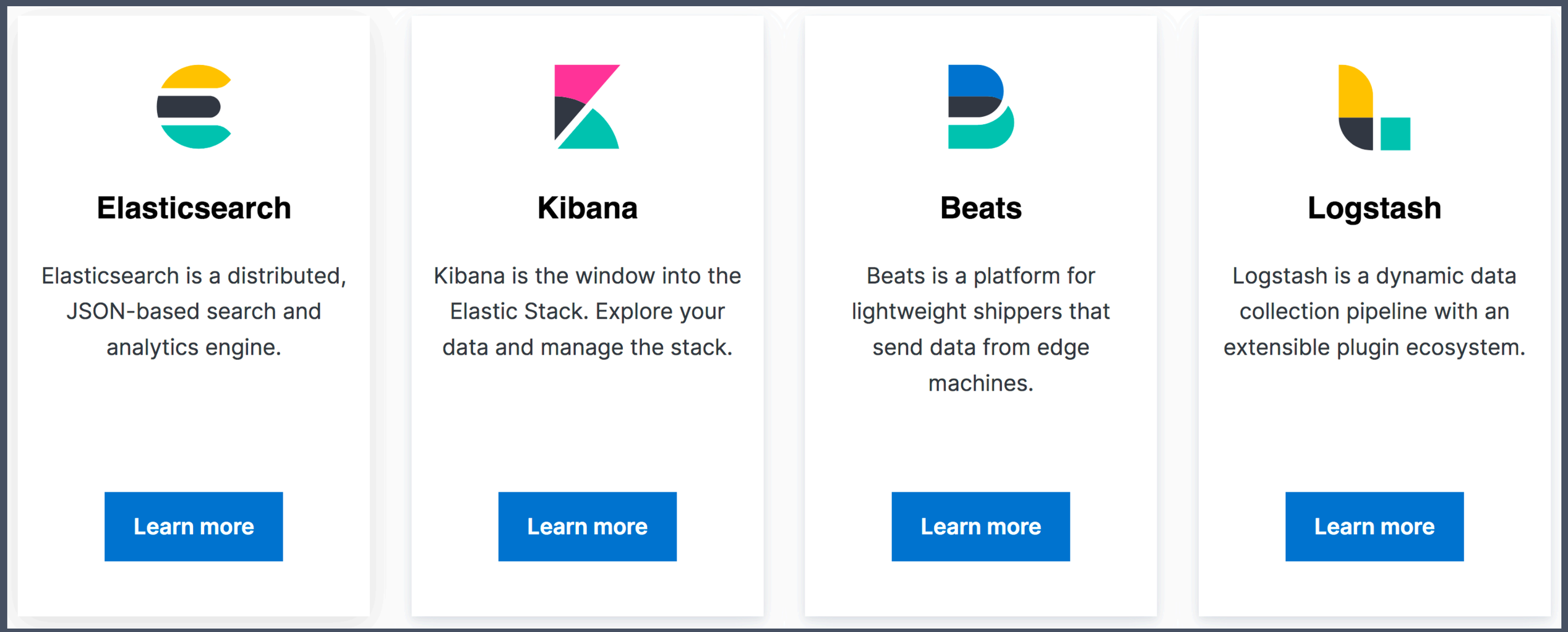
@EMANUIL_TOLEV

elastic @emanuil_tolev

Community engineer at Elastic.

WHAT IS THE ELASTIC STACK?

@emanuil_tolev



Large open source projects: Elasticsearch, Kibana, Beats and Logstash

^ At the heart is Elasticsearch, an open source search and data analytics engine.

^ Kibana is an open source visualisations and dashboarding tool

^ The rest support logging, metrics and tracing, a use case with 1000s of big company and millions of smaller users.

^ "Meet the open source tools that power experiences from the search for life on Mars to finding the best sushi in your neighborhood."

^ All very well, but what does it have to do with data science?

COMMON TOOLS: JUPYTER NOTEBOOKS

elastic @emanuil_tolev

Python, R, Julia

^ For Python pandas, numpy, scipy, scikit-learn, etc.

^ Let me ask you this:

HOW BIG IS YOUR NOTEBOOK?

@emanuil_tolev

HOW LONG IS A PIECE OF STRING.

elastic @emanuil_tolev

It depends. But generally, with recent software versions, it depends on your computer. On your own desktop or laptop.

^ A notebook is very difficult to distribute.

^ Some computations can be distributed with existing toolkits. Without getting into comparisons, some people have found such tooling difficult to configure, plus the leap from own laptop to production env is large. Not so with Elastic.

ELASTICSEARCH SCALES HORIZONTALLY - YOU CAN ADD MORE PIECES

elastic @emanuil_tolev

& thus memory/CPU/etc. Its strength is the easy addition and coordination with new nodes. Your computations can scale as much RAM as a whole cluster, but the interface and commands you use remain as if it were just one node on your laptop.

USE IN PRODUCTION

FOR DATA SCIENCE:

- INTERACTIVE DASHBOARDING
- UNSUPERVISED LEARNING
- ANALYTICS (BUSINESS METRICS)

USE IN PRODUCTION

OTHER INTERESTING APPLICATIONS:

- SEARCH (ONLINE SHOP FRONT, ADMIN TOOLING, RECOMMENDATION SYSTEMS, ...)
- WEB APP MONITORING: "LOGGING, METRICS, TRACES"
- NOW SECURITY TOO

demo of both

^ Interactive dashboarding: flights data.

^ Interactive dashboarding helps a lot with putting insights into production and keeping them there.

^ Interactive dashboarding: earthquakes

^ Unsupervised learning: explain ML and clearly state it is commercial.

- A FREE, EASY TO CONFIGURE TOOL THAT SPEAKS PYTHON AND R
- WE ARE EXPLORING HOW THE WHOLE STACK CAN HELP DATA SCIENTISTS

COME TALK TO ME!

elastic @emanuil_tolev

Talk about problems you're trying to solve, whether RAM has been a blocker for you, how you've found other distributed computation tooling.

THANK YOU!

@EMANUIL_TOLEV
ETOLEV@ELASTIC.CO