HANDS-ON INTRODUCTION TO ELASTICSEARCH

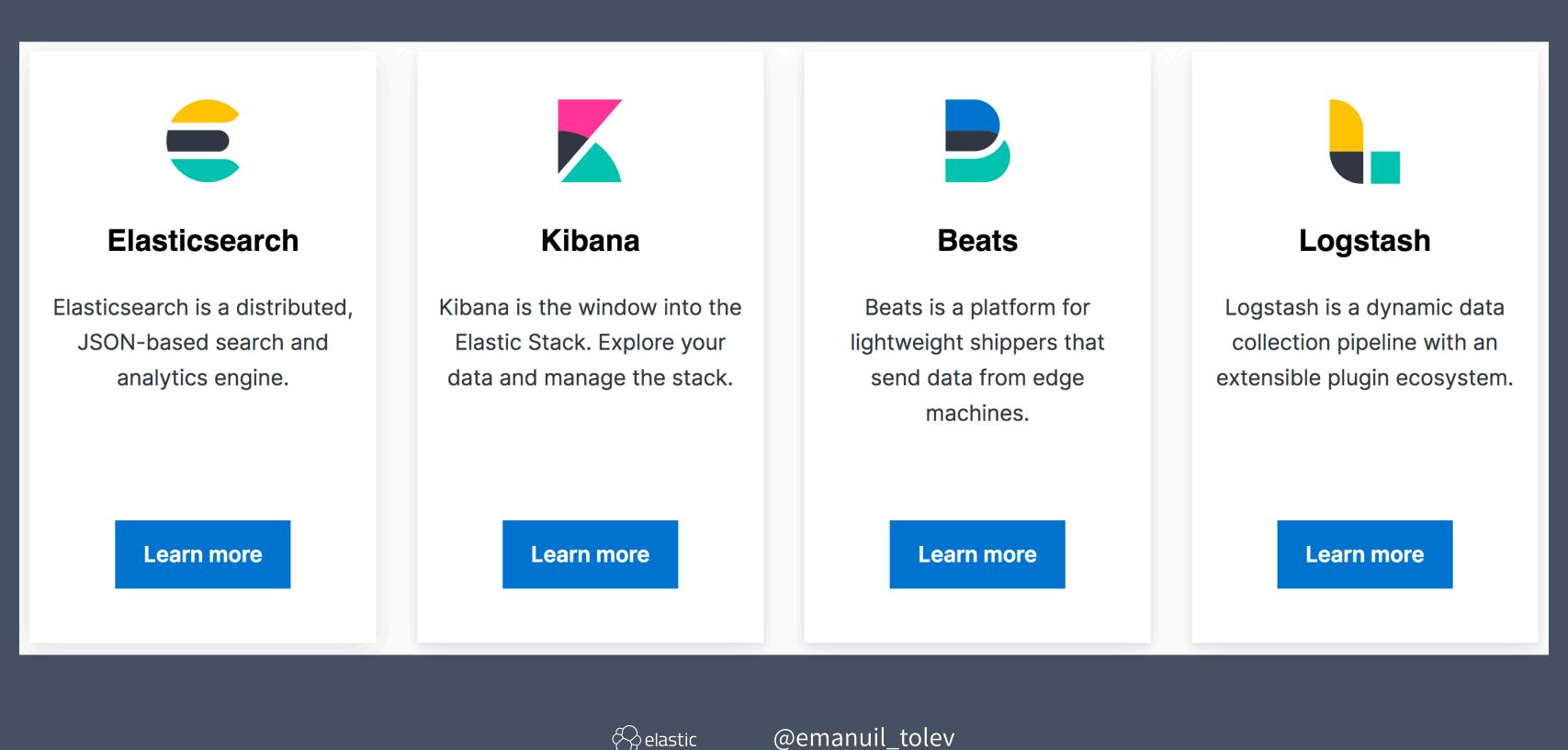## EMANUIL TOLEV
@EMANUIL_TOLEV

elastic   @emanuil_tolev

COMMUNITY ENGINEER

elastic   @emanuil_tolev

Elasticsearch

Elasticsearch is a distributed, JSON-based search and analytics engine.

[Learn more]

Kibana

Kibana is the window into the Elastic Stack. Explore your data and manage the stack.

[Learn more]

Beats

Beats is a platform for lightweight shippers that send data from edge machines.

[Learn more]

Logstash

Logstash is a dynamic data collection pipeline with an extensible plugin ecosystem.

[Learn more]

elastic   @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?

HOW BIG IS YOUR JUPYTER NOTEBOOK?

elastic   @emanuil_tolev

HOW LONG IS A PIECE OF STRING...

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. So is computation from a Jupyter notebook.

^ 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. Elastic sits at a balance between easy to distribute and (less) features than Jupyter which has all of Python, R and Julia.

ELASTICSEARCH SCALES HORIZONTALLY – YOU CAN ADD MORE PIECES

& 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

> SEARCH (ONLINE SHOP FRONT, ADMIN TOOLING, RECOMMENDATION SYSTEMS, ...)
> ANALYTICS (BUSINESS METRICS)
> WEB APP MONITORING: "LOGGING, METRICS, TRACES"
> INTERACTIVE DASHBOARDING
> UNSUPERVISED LEARNING FOR ANOMALY DETECTION (PAID! BUT GOOD)
> NOW SECURITY TOO

elastic   @emanuil_tolev

If you're a data scientist, you may be asked to analyse any of this data, so it's useful to know what it's typically used for. You can google for most of these and get help from your peers.

THE SEARCH EXPERIENCE

> IT'S MORE THAN A BACK-END TO INSTALL AND CONNECT TO
> > FULL-STACK VIEW: GREAT UX IS THE GOAL
> > > RELEVANCE, RANKING, SPEED
> > OF COURSE YOU DO ALSO NEED A BACK-END :)

Just in case you do get asked to build a search experience of some kind. Surprisingly common task.

TODAY

> SETUP AND CONNECT
> INDEX (PUT DATA IN)
> QUERY (GET DATA OUT - SEARCH AND AGGREGATE)
> USE KIBANA TO CREATE A FEW VISUALISATIONS
> ET VOILA!

elastic   @emanuil_tolev

HTTPS://GITHUB.COM/EMANUIL-TOLEV/ES-HANDS-ON-INTRO

> FOR THOSE WHO HAVEN'T DONE IT YET, DOCKER-COMPOSE UP AND THEN RUN ./BUILD.SH TO INJECT SAMPLE DATA

> FOLLOW DEMO-CONSOLE.TXT IN HTTP://LOCALHOST:5601/APP/KIBANA#/DEV_TOOLS/CONSOLE

> LATER WE'LL PLAY WITH KIBANA, THE VIS GUI!

THE CRUX OF THIS IS THAT YOU LEARN HOW

elastic   @emanuil_tolev

Copy paste lines 1 by 1, we'll go through them later.

QUESTIONS?

EMANUIL TOLEV
@EMANUIL_TOLEV

ETOLEV@ELASTIC.CO

elastic   @emanuil_tolev