software, data, clouds
and you

hello!

I AM JONATHAN FORBES

First AWS invoice on 2/8/2008.

14 cents.

You can find me & my friends at @WeArePhlo and I lurk at @elektrifi

“
…I want to help software engineers work more closely with data using modern cloud services and tools…

my core assumption

you want to **extract business value**

as quickly as possible and

for as little money as possible.

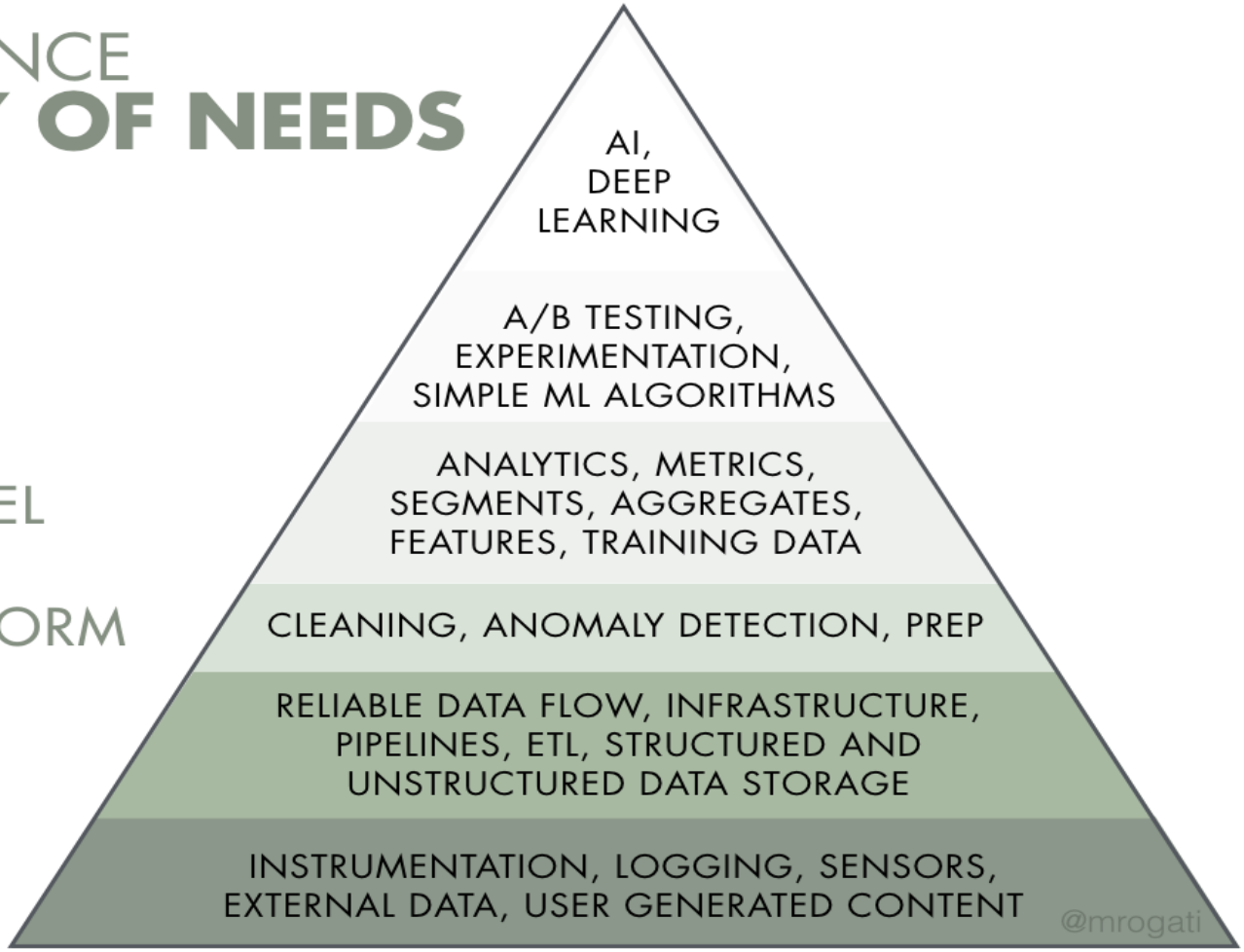
BIG CONCEPT #1

it's all about the data, right?

a starting point

why ~~is this~~ are you important?

THE DATA SCIENCE **HIERARCHY OF NEEDS**



LEARN/OPTIMIZE

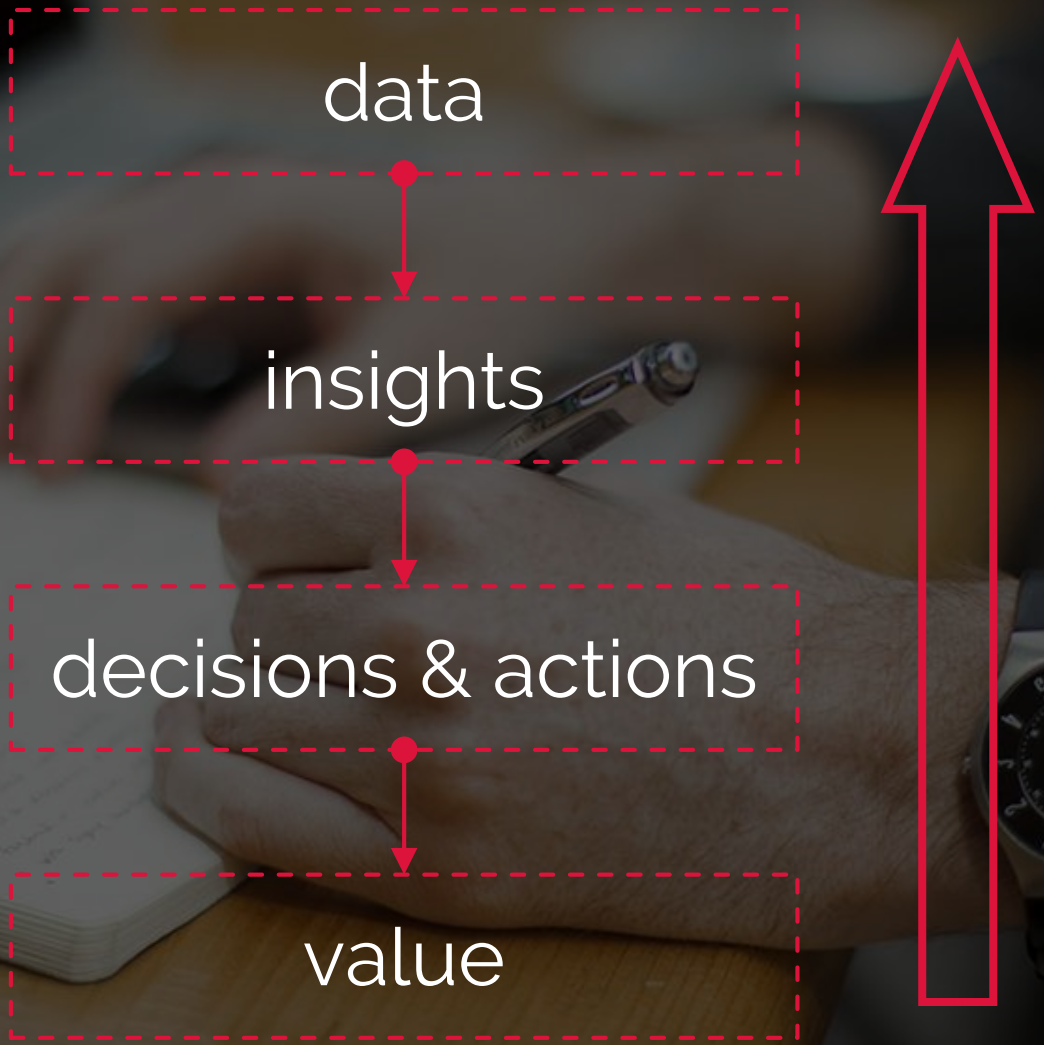
AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

a starting point



BIG CONCEPT #2

without data it's just an opinion
with data, change something for the better

today in software

apps && functions && containers && data

today in data

data + model(s) –compute→ prediction

we want

apps && functions && containers

&& data + model(s) –compute→ prediction

our goal is optimization

BLUE RIVER TECHNOLOGY

Optimize Every Plant

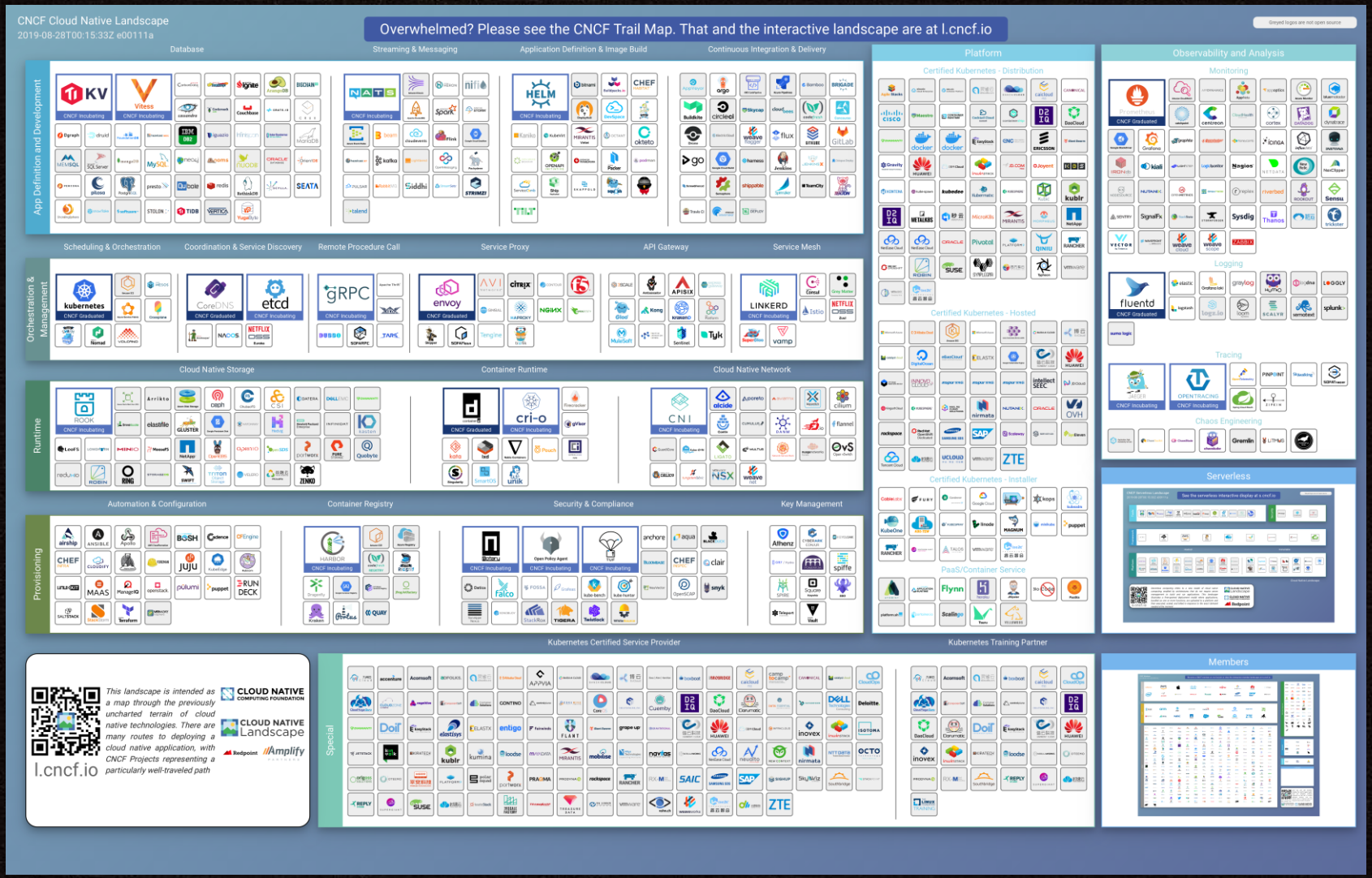
Blue River Technology is building the next generation of smart agriculture equipment. We're introducing See & Spray technology, enabling a world in which every plant counts.

TOTAL
TAUROB
MERKLE

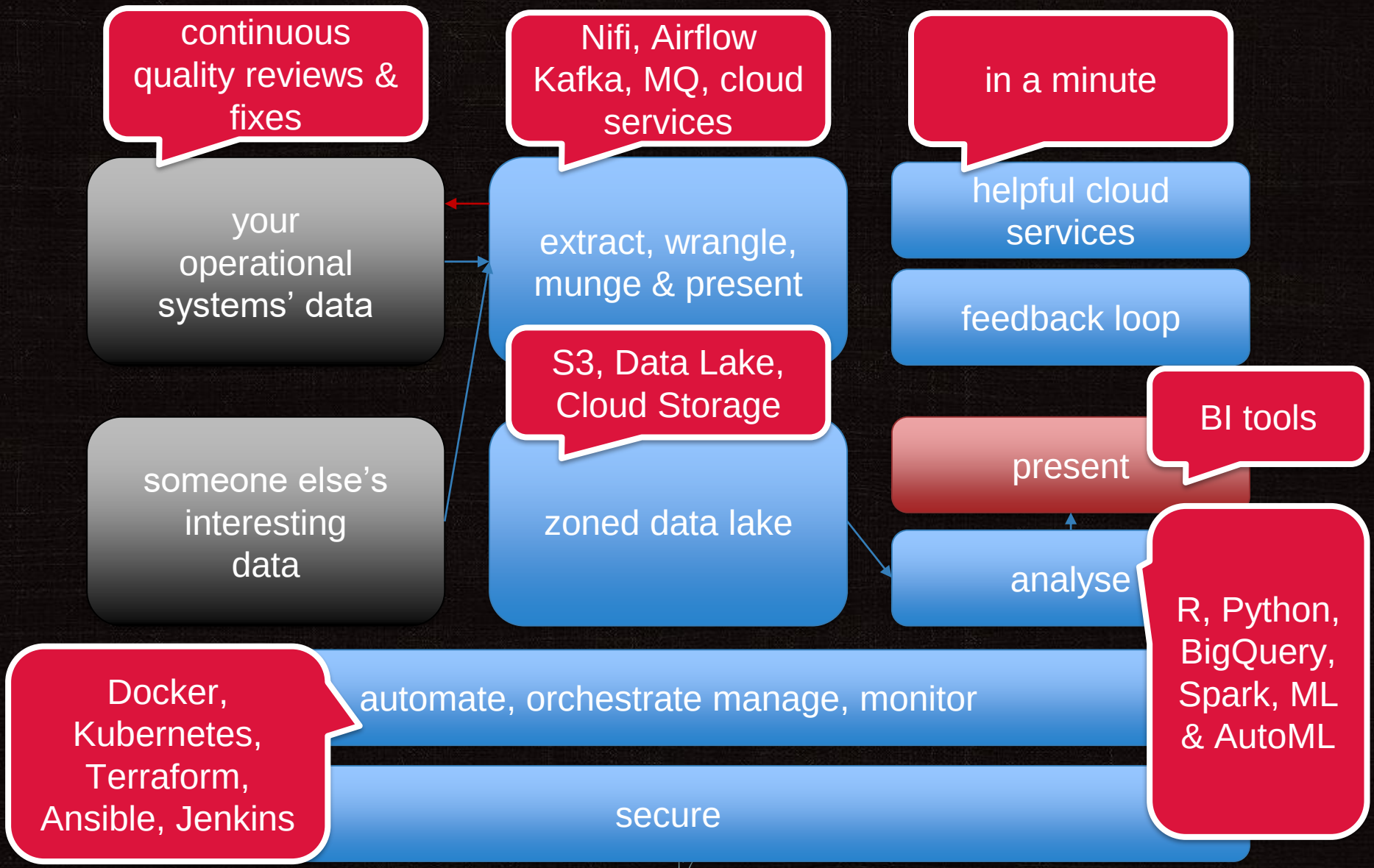
BIG CONCEPT #3

it's probably
already solved
in a cloud

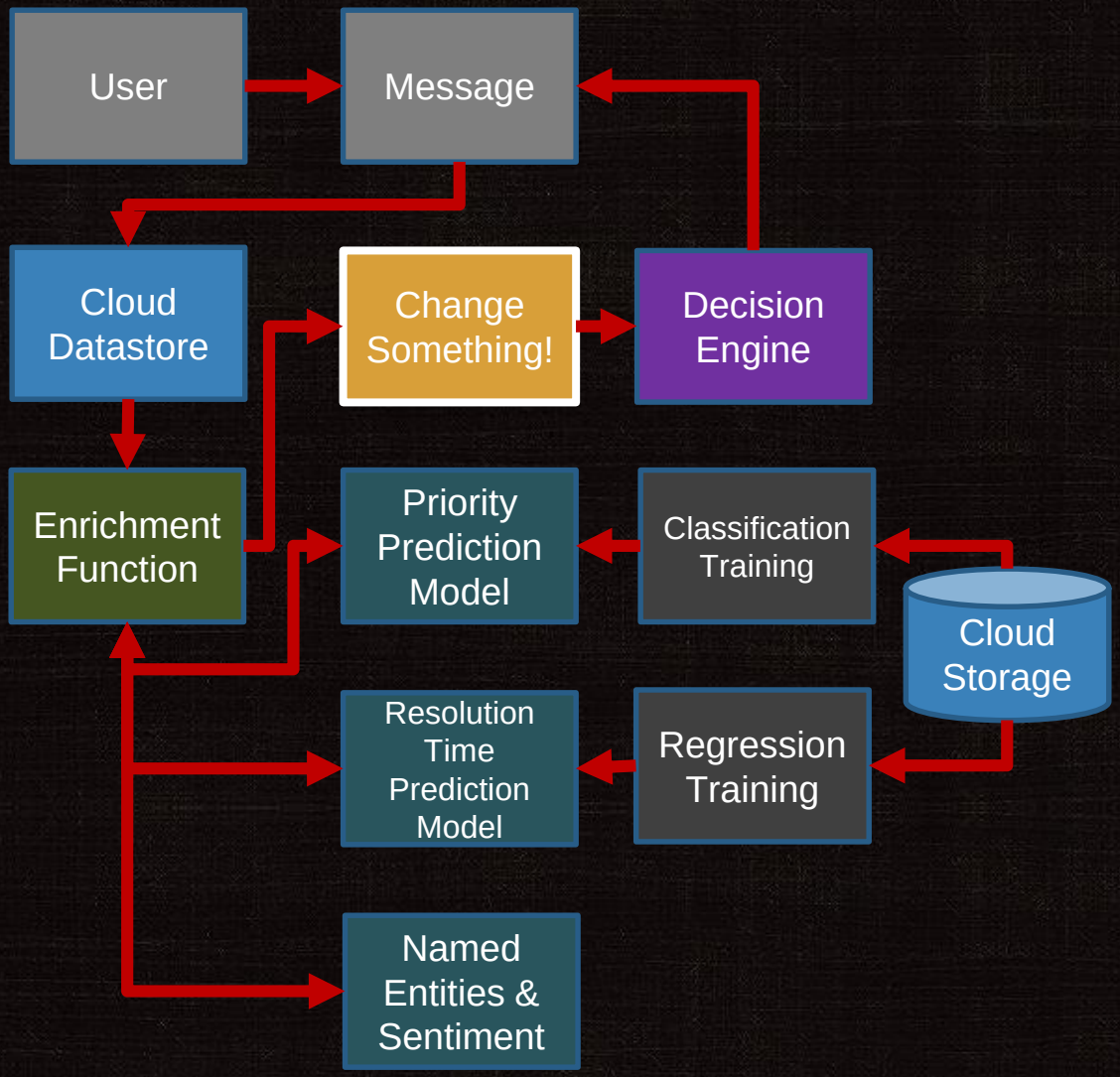
cncf landscape



for analytics, this works...



you could start with this



- AWS Lambda
- Google Cloud Functions
- Azure Functions
- Apache OpenWhisk
- Serverless framework

- TensorFlow
- Apache Unomi

- Data collection
- Model training
- Testing
- Versioning
- Deploying
- Publishing new versions

what can I do for...

£5 per day

- collect, join & analyse driver & passenger data
- two segmentation models run as serverless functions

- **value = at least £30K uplift**

£50 per day

- combine marketing, buyer & product data into single view of leads + customers
- auto-feed touchpoints & media strategy

- **value = sell 500 more products per year**

£500 per day

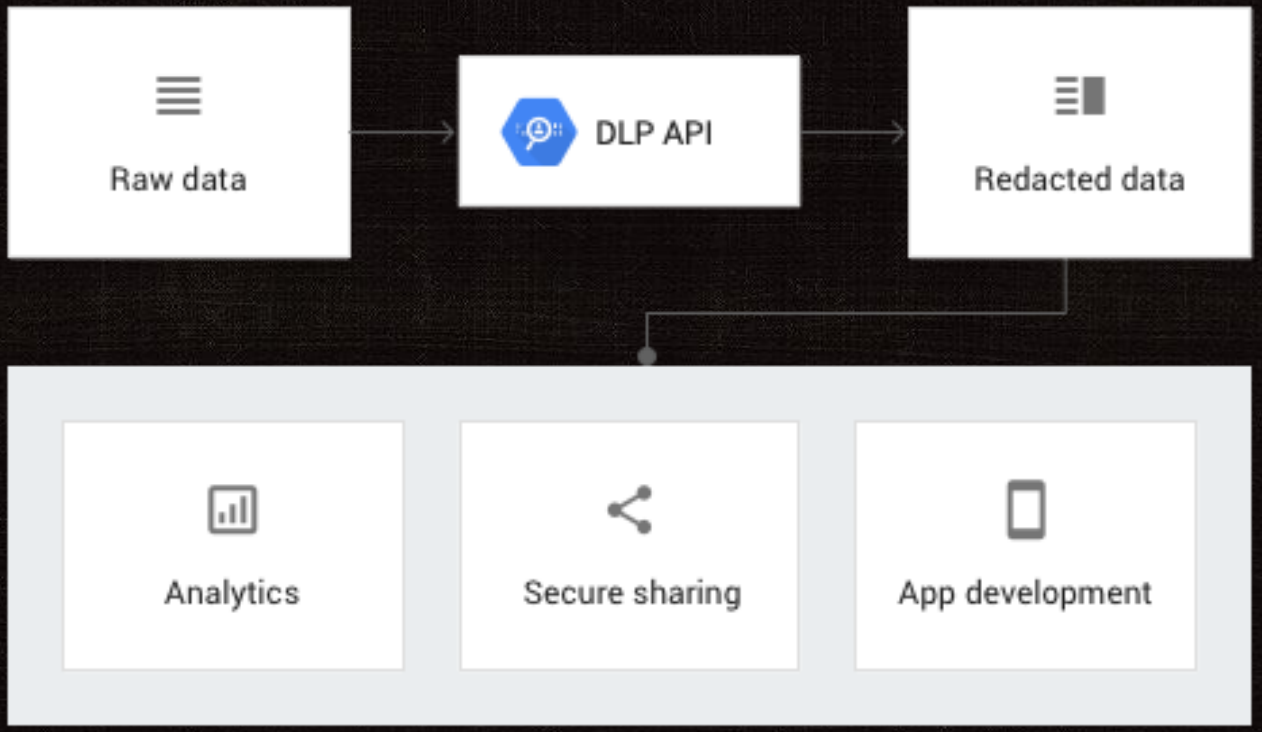
- run end-to-end analytics with 1800+ attribute data mart
- support 6 analysts on Spark clusters

- **value = £30M uplift online sales**

some useful native cloud services

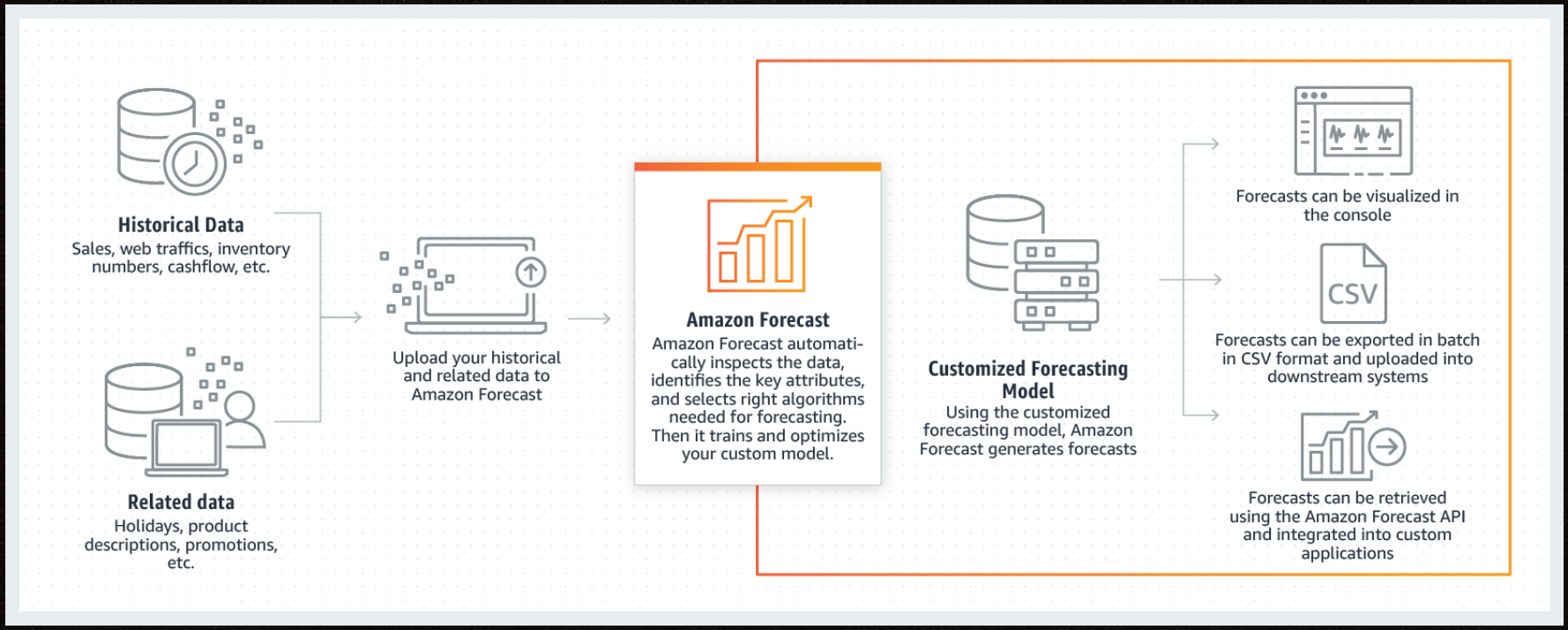
- **google cloud data loss prevention api**
- **aws forecast**
- **azure text analytics**
- google cloud healthcare api
- aws ground station
- azure iot hub

google cloud data loss prevention



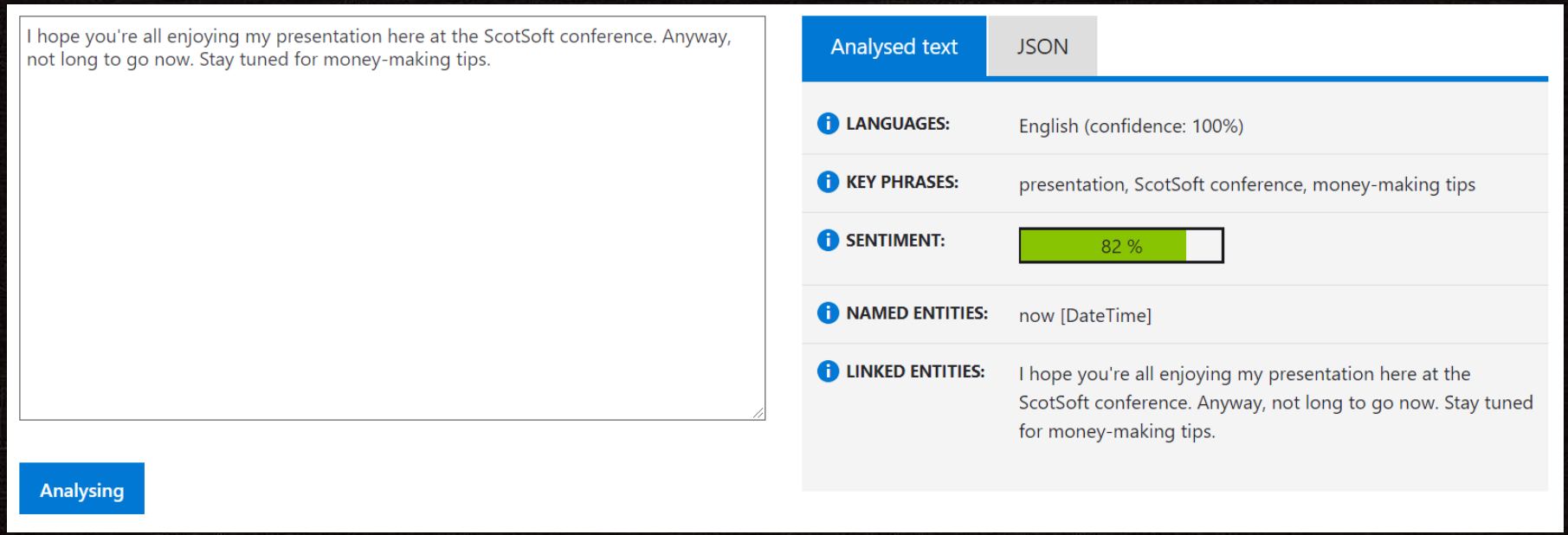
- 90 predefined data detectors, understands region & context
- redaction using masking, secure hashing, tokenization, bucketing, format-preserving encryption
- REST API from inside or outside of GCP & from mobile devices, IoT devices, browsers.

aws forecast

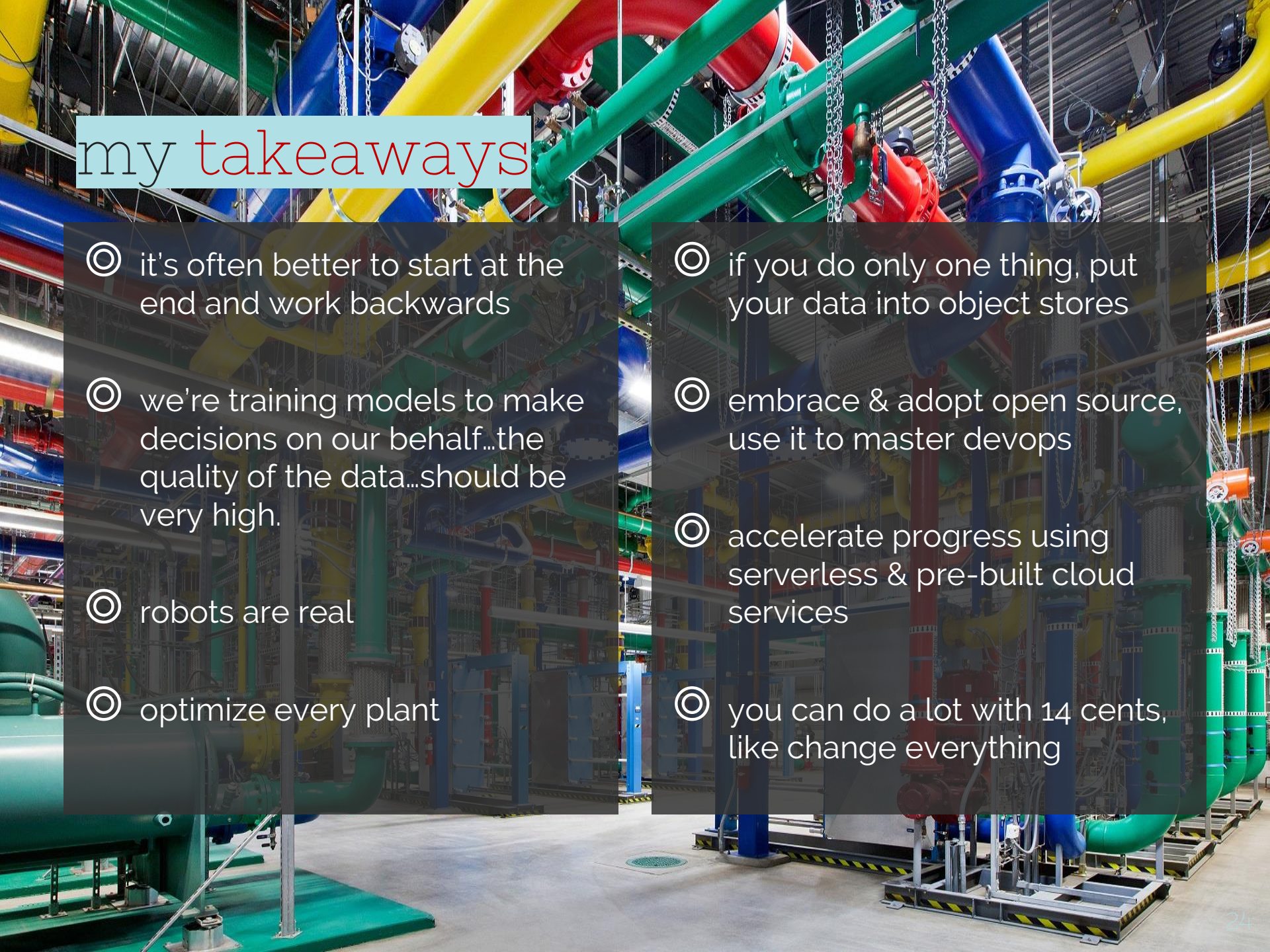


- end-to-end time-series forecasting
- greatly simplifies ML modelling
- predefined algorithms and AutoML

azure text analytics



- language detection, sentiment analysis
- key phrase extraction, named entity recognition
- runs in azure or on prem via docker container



my takeaways

- it's often better to start at the end and work backwards
- we're training models to make decisions on our behalf…the quality of the data…should be very high.
- robots are real
- optimize every plant
- if you do only one thing, put your data into object stores
- embrace & adopt open source, use it to master devops
- accelerate progress using serverless & pre-built cloud services
- you can do a lot with 14 cents, like change everything

where can I learn more?

- Andy Cobley @ University of Dundee
  - Data Lab-supported Data Engineering M.Sc.
- Databricks on Youtube
  - Spark, Delta and MLFlow
- Uber Engineering's Michelangelo
  - https://eng.uber.com/michelangelo/
- Google some keywords:
  - cncf, tfx, ONNX, netron, seldon.io, polyaxon
  - dscully, neil d. lawrence

thanks for coming!

Who are you again?

- I'm Jonathan Forbes

@elektrifi & @WeArePhlo

www.wearephlo.com

Credits

- hamlet @ slidescarnival
- carbon.now.sh
- mindmeister
- unsplash
- thenounproject
- fontsquirrel
- death to the stock photo (license)
- the pattern library
- nina zakharenko