Practical data science use cases in the 4th industrial revolution

Stefan Steffen

AGENDA

1. **The critical context that defines our future landscape**
   - Exponential change in resources
   - Acceleration in algorithms
   - Speed of light limitations

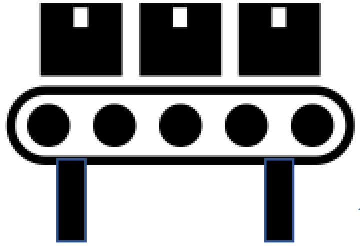
2. **Practical use cases to capitalize on this disruptive change**
   - Machine learning in telecoms
   - Computer vision in retail customer experience
   - Data science and IOT in smart farming
   - Cassandra and Kafka in supply chain

01 The critical context that defines our future landscape

01 From steam power to connected devices

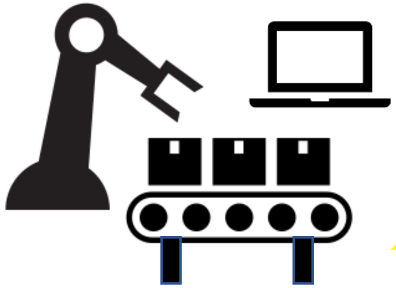
1st Industrial Revolution

Mechanisation of production supported by water and steam power



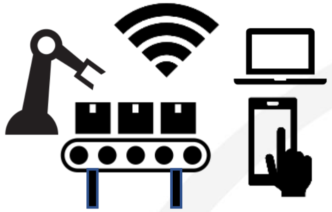
2nd Industrial Revolution

Mass production drives the 2nd industrial revolution with the help of electrical power



3rd Industrial Revolution

Further automation of the production process by combining electronics and computers



4th Industrial Revolution

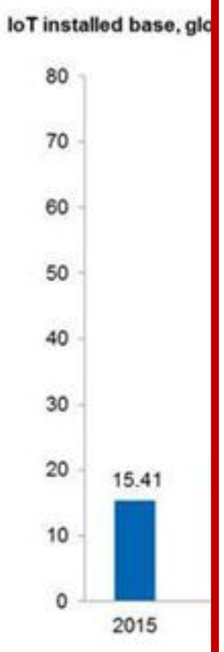
The advent of big data, connected devices, robotics and the move towards artificial intelligence fuels the 4th industrial revolution

i. Prof Klaus Schwab, *The Fourth Industrial Revolution* (WEF, 2016)

ACTUARIAL SOCIETY OF SOUTH AFRICA

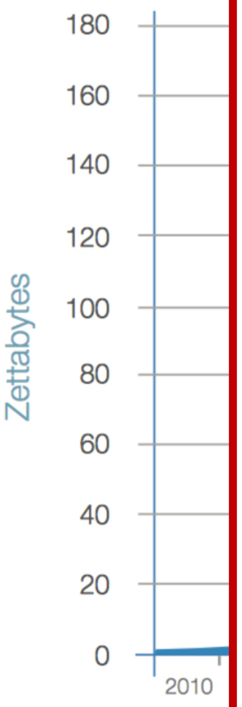
Connected devices

From



Proliferation of data

163Z

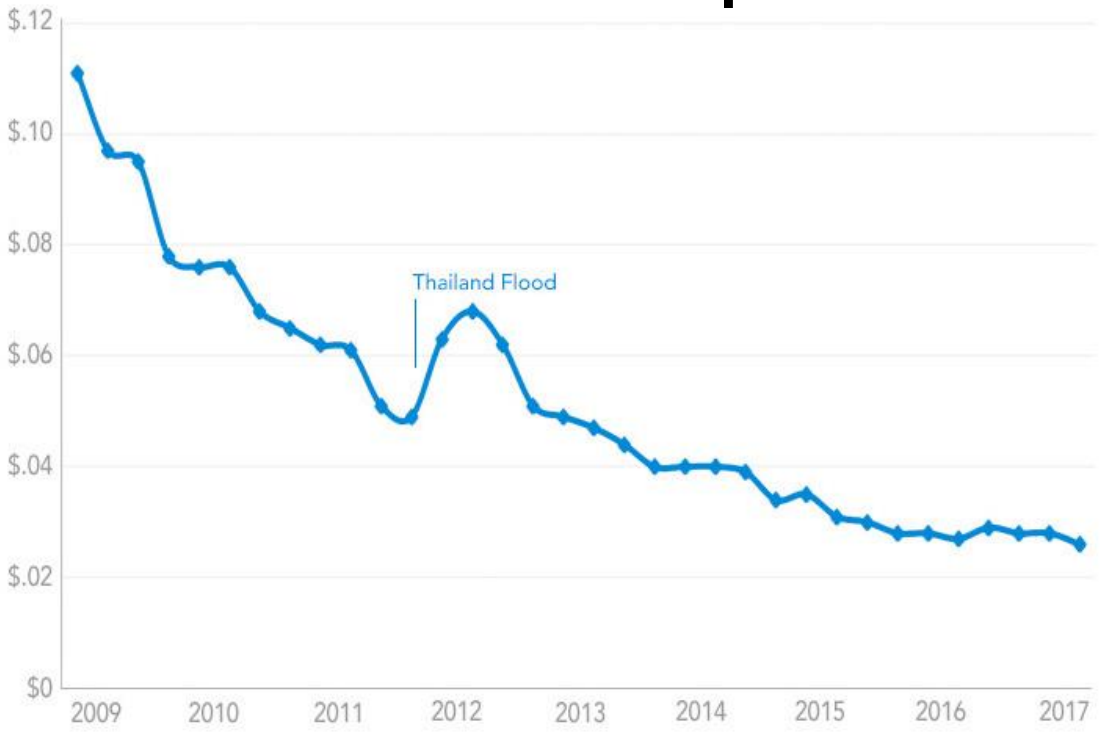


Processing power

Trill

Cost of Storage

Now less than R0.50 per Gb…



https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte/

01 Acceleration in algorithms

Machines

Google Cloud V

can see

Machines

amazo

(speech rec

can listen

Machines

Natural lan processin

can read

Machines

eviebot

Existor algorithm

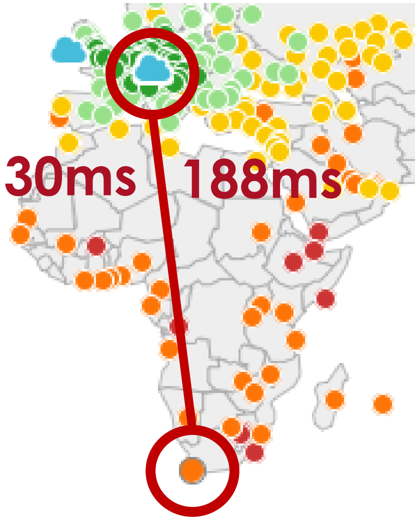
can speak

ACTUARIAL SOCIETY OF SOUTH AFRICA

Virtual reality

Our eyes produce over 5Gb of data per eye per second needing sub 7ms round trip latency to avoid illness

Latency



Perceptible lag once latency exceeds 100-200ms

Autonomous cars

5Tb of data generated per day with requirement for sub 5ms decisions

https://help.dyn.com/understanding-market-cities-and-cloud-performance/
Derek Collison – The future of Edge Computing

02 Practical use cases to capitalize on this disruptive change

02 Overview of Machine Learning in telecoms

ACTUARIAL SOCIETY OF SOUTH AFRICA

Network operations

- Network monitoring
- Fault detection
- Network optimisation
- Predictive maintenance

Product management

- Personalised marketing
- Product recommenders
- Churn prediction
- Dynamic discounting

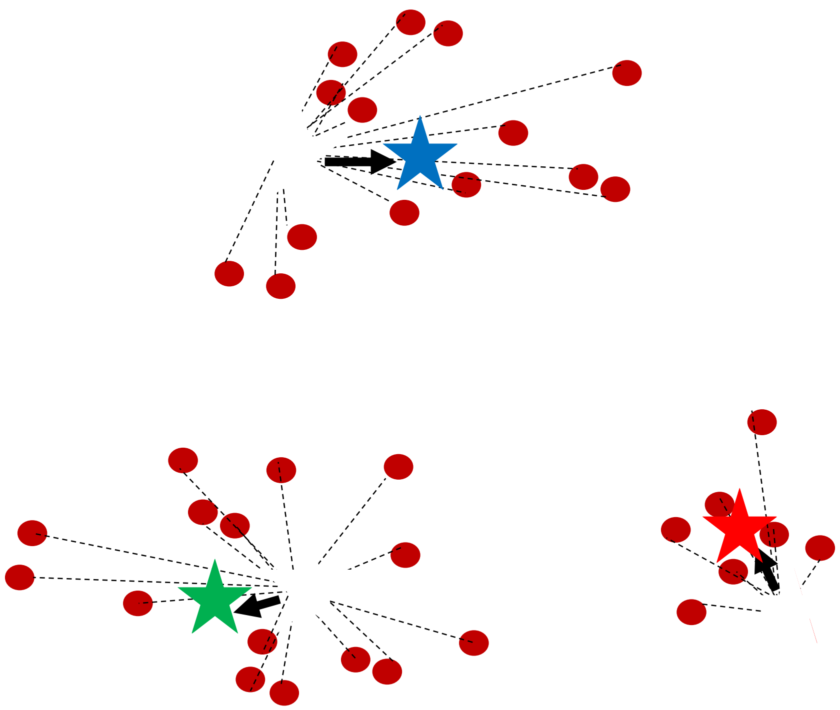
Service delivery

- Service chatbots
- Voice interfaces
- Predictive identification of service failures

Support functions

- Revenue assurance
- Fraud detection
- Cyber security anomaly detection

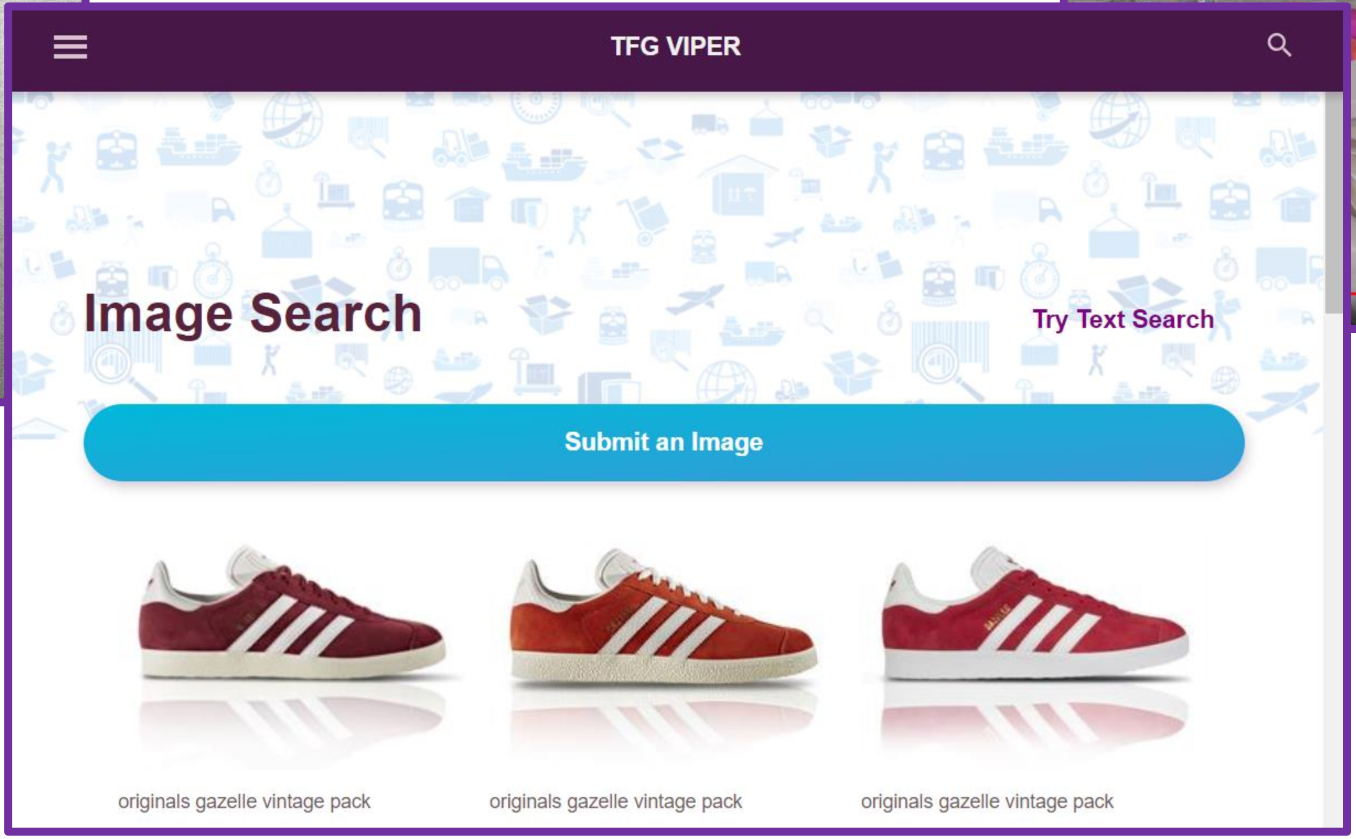
Basic K-means clustering for customer segmentation

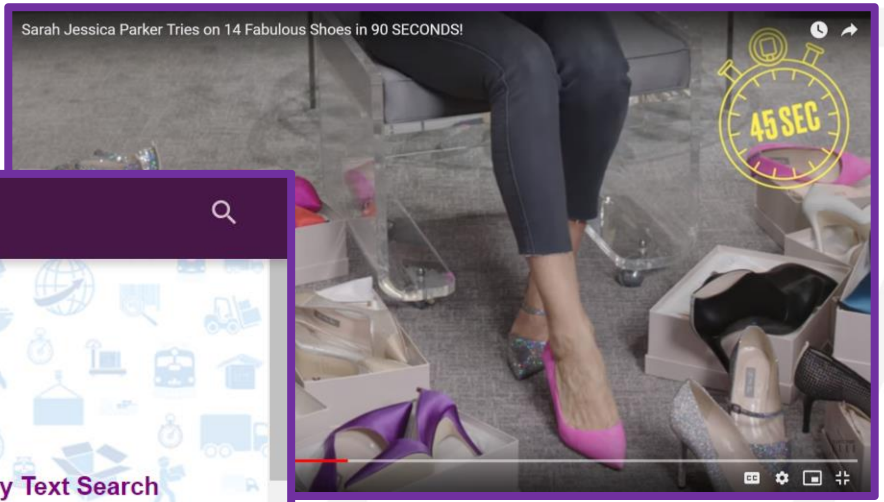


1. Randomly initialize the cluster centroids
2. Assign the data points to the closest cluster centroid
3. Move the cluster centroid to the mean of the assigned data points
4. Repeat until there is no change in the clusters
5. Run K-means many times starting with different random centroids
6. Examine the clusters distortion by reviewing the sum of squared differences

Manu Jeevan: Possibly the simplest way to explain K-Means algorithm

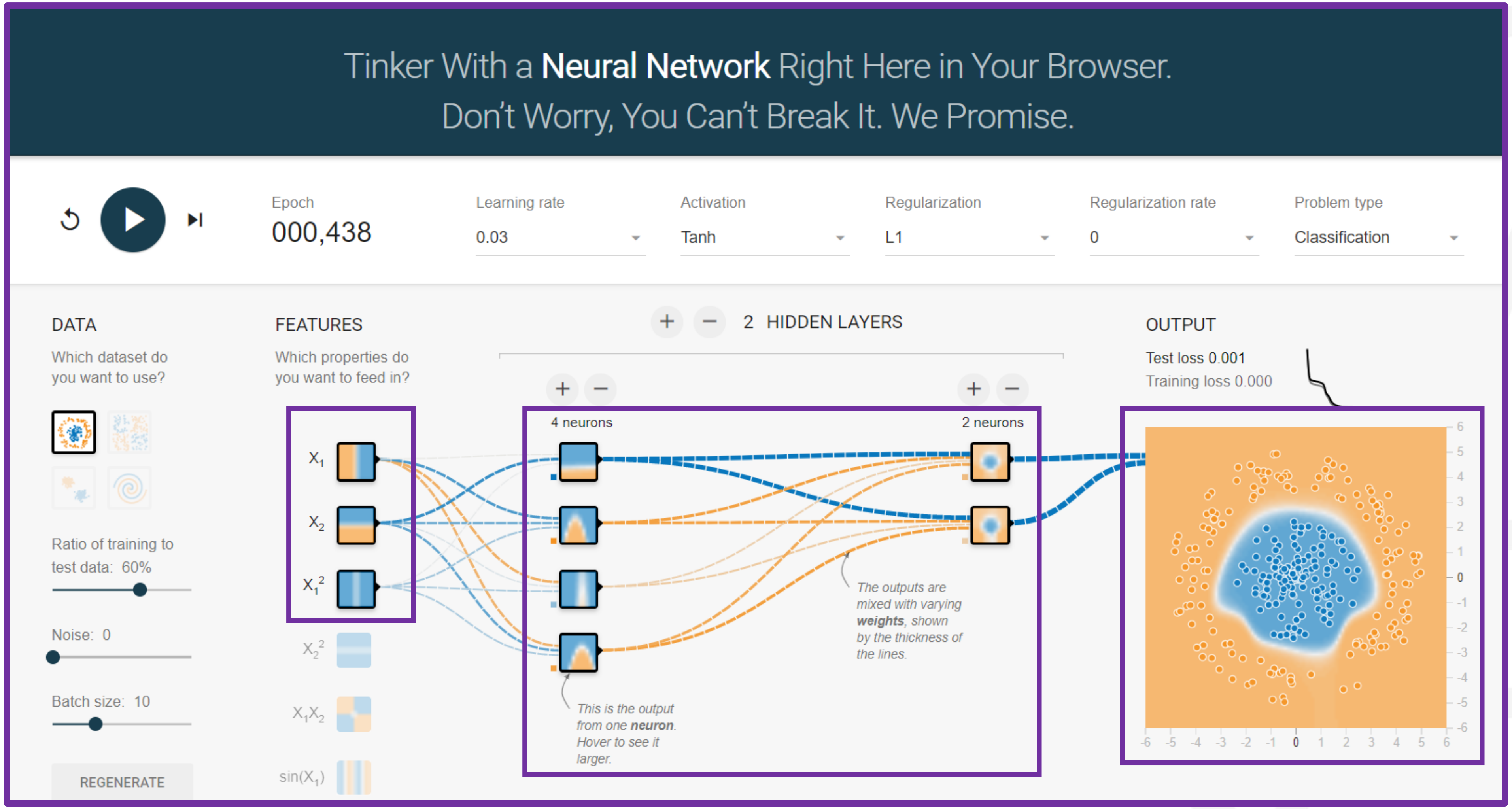
Computer Vision in retail customer experience





A quick overview

1. Neurons in the brain
2. Artificial neural network
3. Input to Output layers
4. Hidden layers
5. Filters for edges & shapes
6. Pattern detection



https://playground.tensorflow.org/
Michael Nielsen: http://neuralnetworksanddeeplearning.com/index.html

A Neural Network Playground

https://playground.tensorflow.org/#activation=tanh&batchSize=10&dataset=circle&regDataset=reg-gauss&learningRate=0.03&regularizationRate=0&noise=0&networkShape=4,2&seed=0.2...

Tinker With a **Neural Network** Right Here in Your Browser.
Don't Worry, You Can't Break It. We Promise.

Epoch	Learning rate	Activation	Regularization	Regularization rate	Problem type
000,000	0.03	Tanh	None	0	Classification

DATA

Which dataset do you want to use?

Ratio of training to test data: 60%

Noise: 0

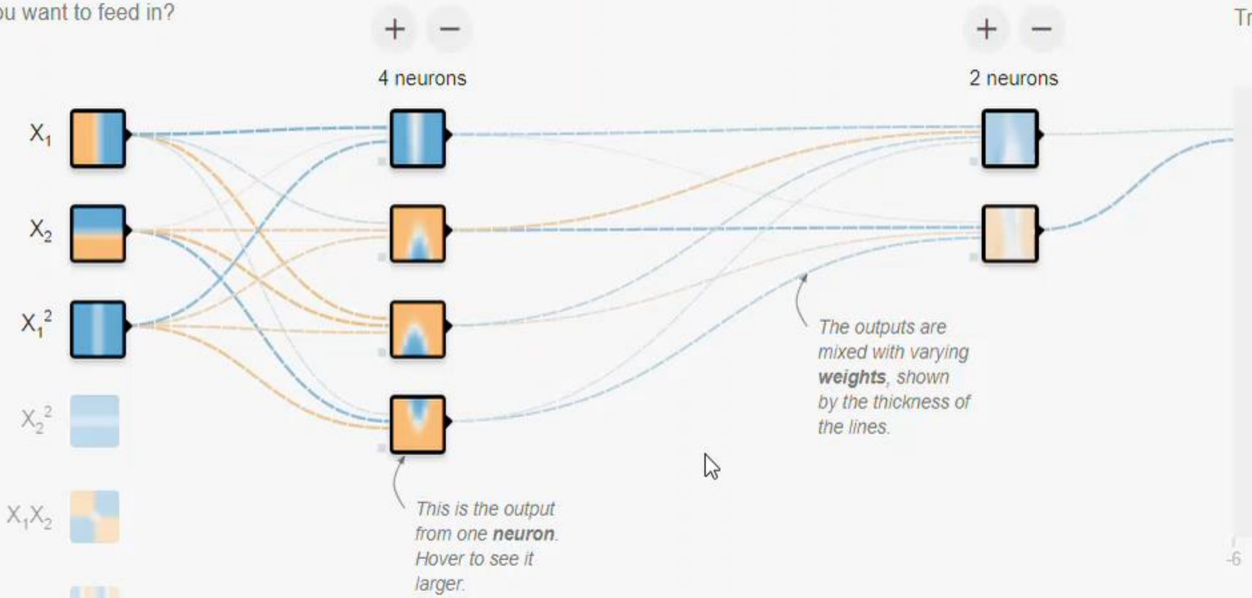
Batch size: 10

REGENERATE

FEATURES

Which properties do you want to feed in?

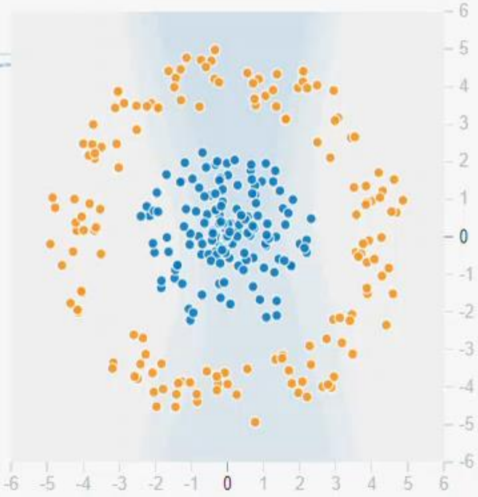
+ − 2 HIDDEN LAYERS



OUTPUT

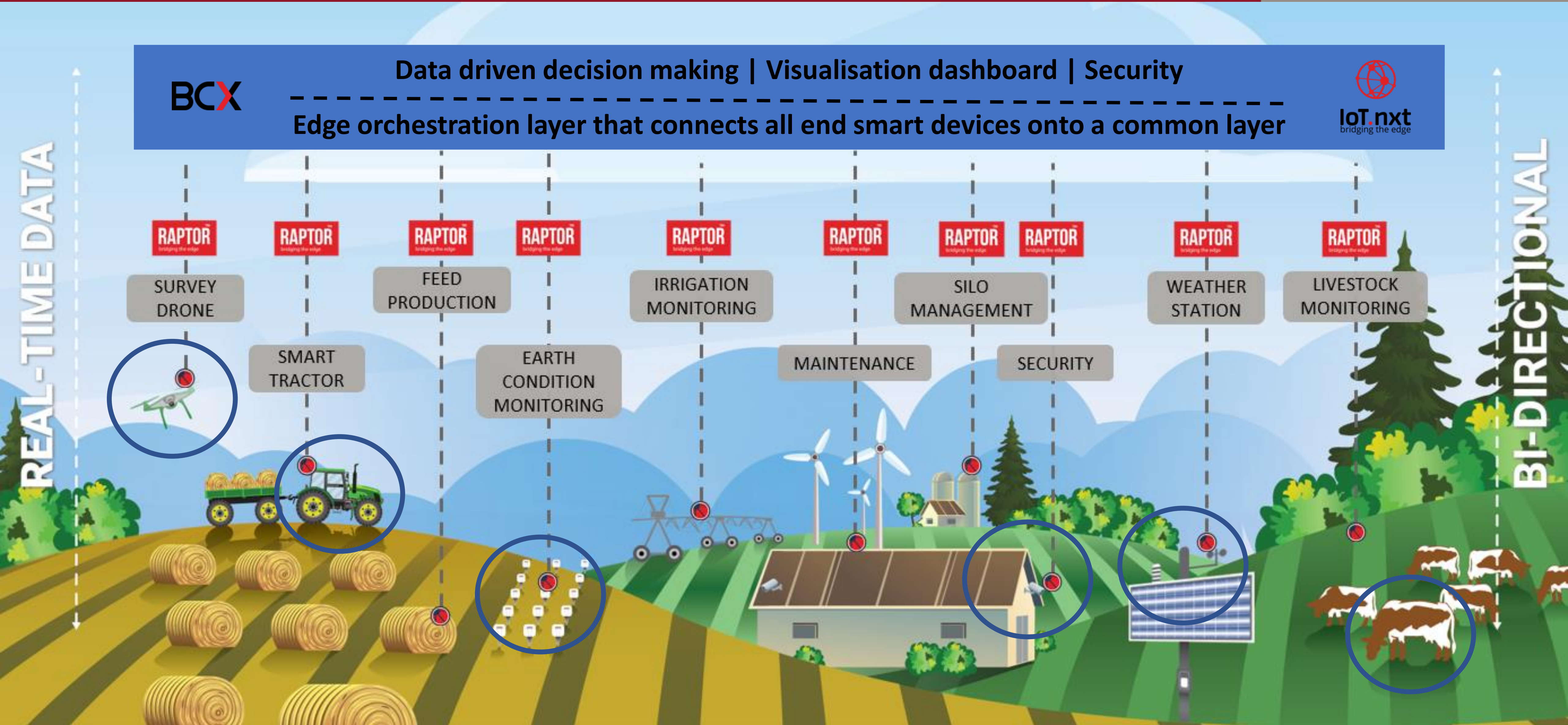
Test loss 0.479

Training loss 0.477

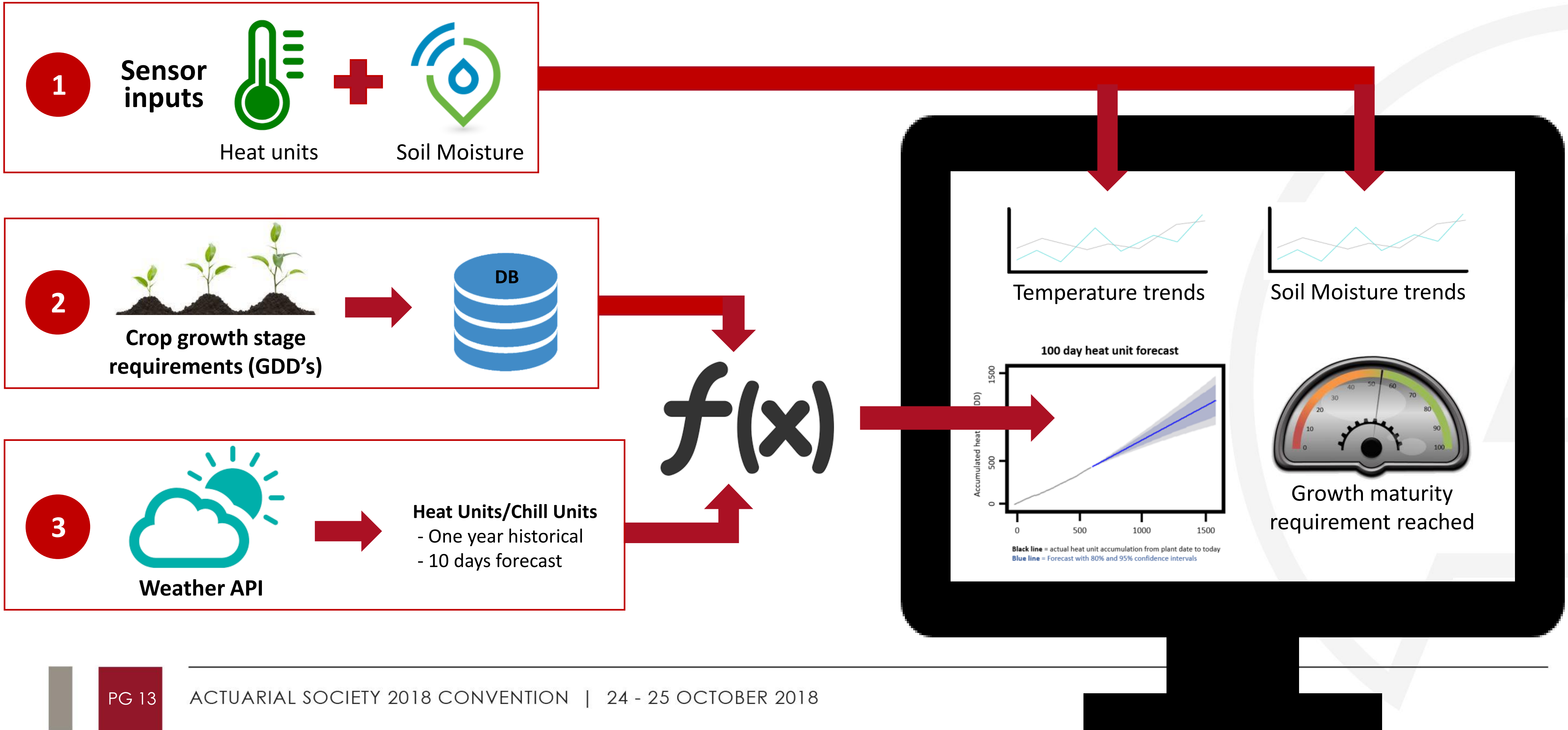


02 IOT applications in Smart farming

ACTUARIAL SOCIETY OF SOUTH AFRICA

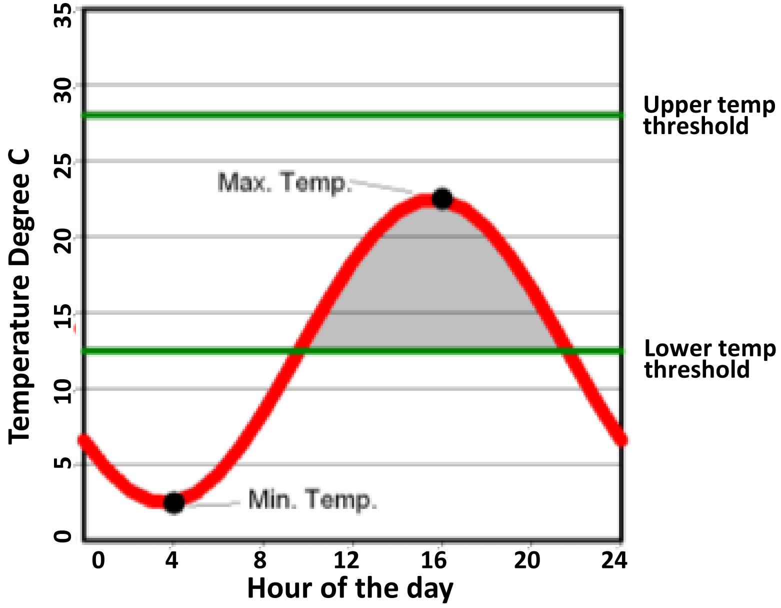


02 Crop growth forecasting using heat units & soil moisture

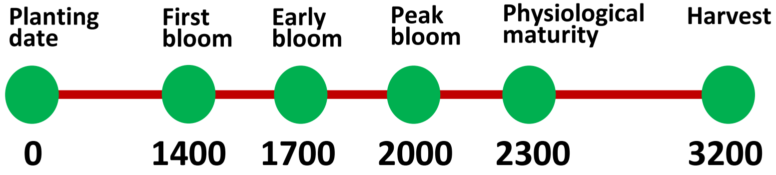


Crop growth forecasting using heat units & soil moisture

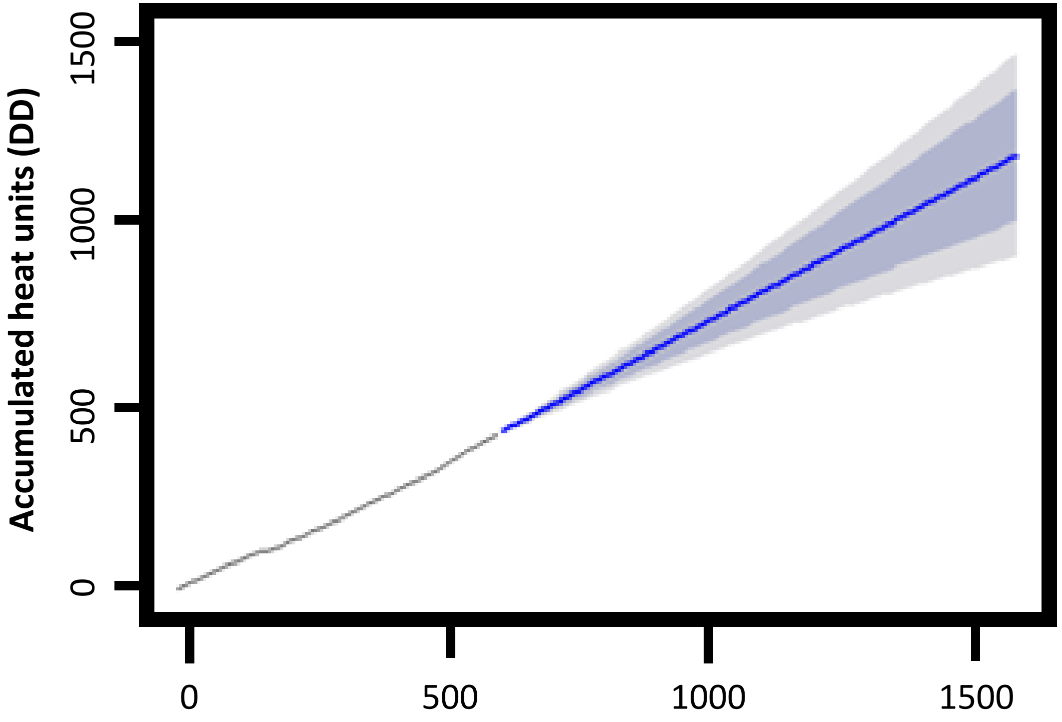
Single sine curve method



Heat units accumulated after planting



100 day heat unit forecast



Black line = actual heat unit accumulation from plant date to today

Blue line = Forecast with 80% and 95% confidence intervals

Paul W Brown: Heat Units – University of Arizona Cooperative Extension

02 Real-time supply chain stock visibility

Distribution Centre
- **Planning**
- **Logistics**

Lack of stock visibility from the Distribution Centre to Stores

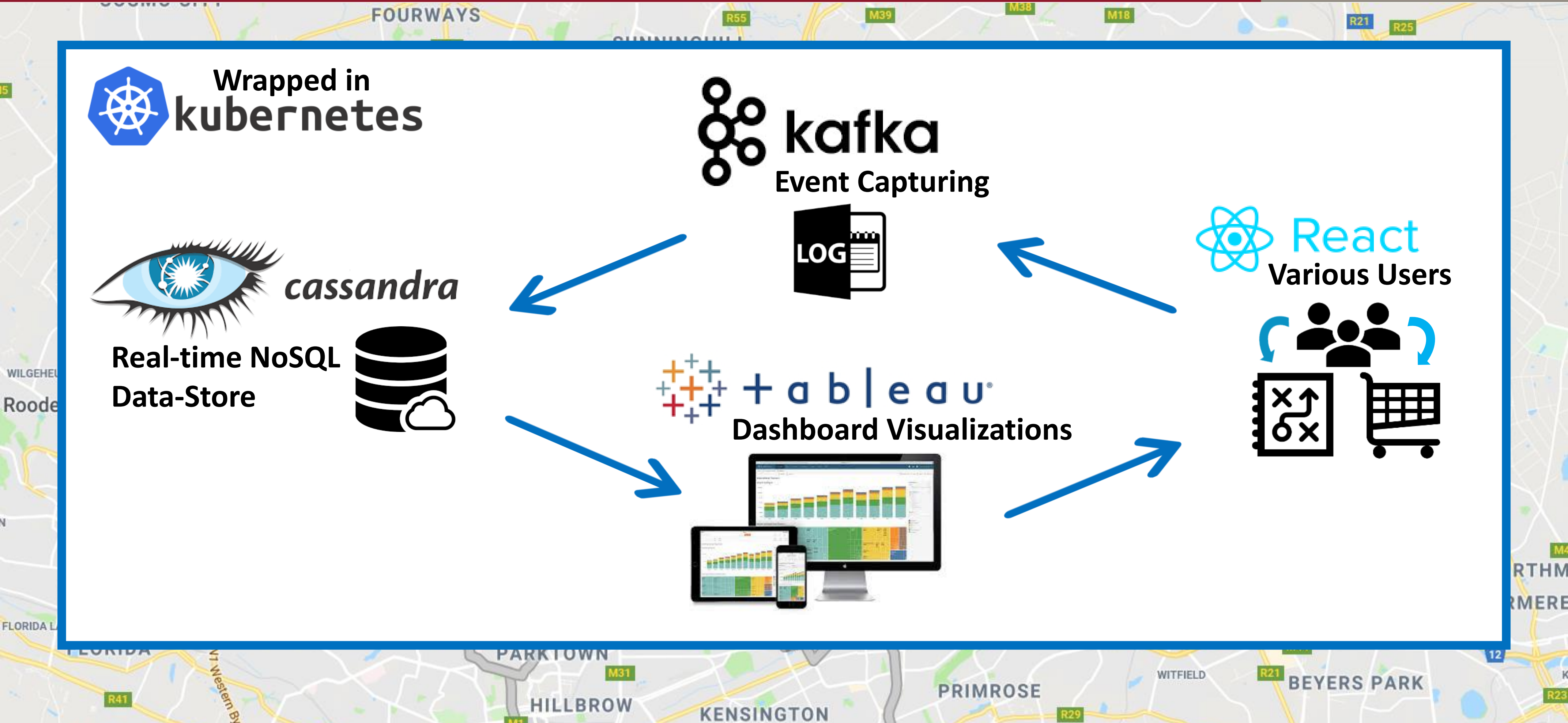
Inefficient movement of stock through supply chain

Inadequate allocation of stock causing lower trade

Poor customer experience

Multiple stores
- **Operations**
- **Floor staff**

02 An architecture view of a stock visibility deployment



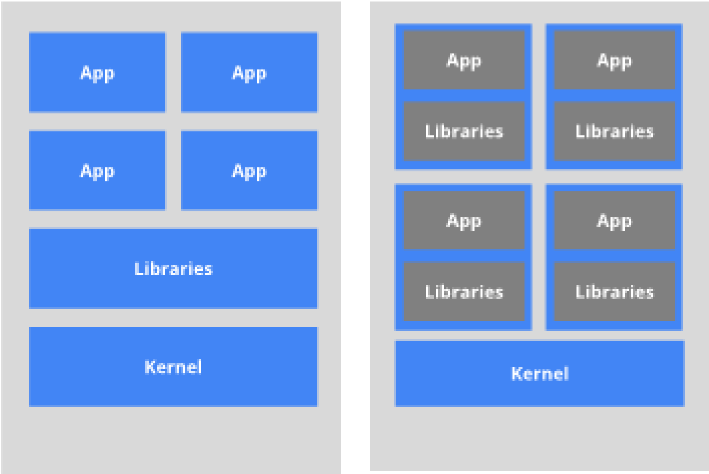
02 Understanding the components

Kubernetes overview

Kubernetes can be thought of as:

- A container platform
- A microservices platform
- A portable cloud platform

Old way: Applications on host **New way:** Deploy containers

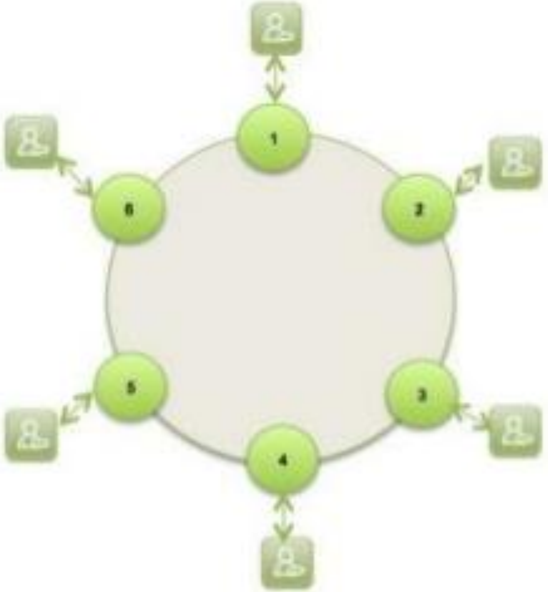


Heavy-weight, non-portable. Relies on OS package manager.

Small and fast, portable. Uses OS-level virtualisation.

Cassandra overview

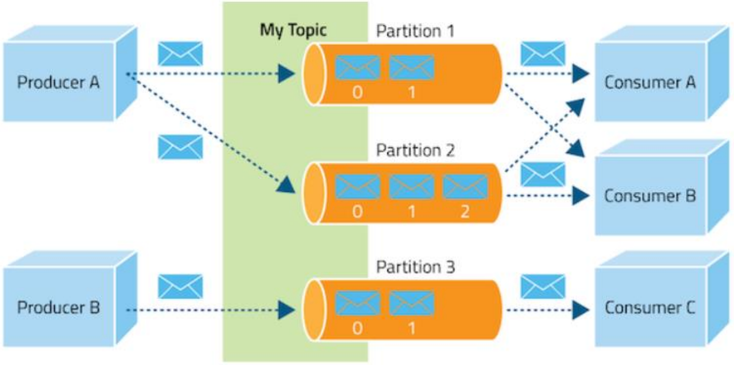
Apache Cassandra is a high performance, extremely scalable, fault tolerant (i.e. no single point of failure), distributed post-relational database solution.



Kafka overview

Apache Kafka is *a distributed streaming platform*.

- Publish and subscribe to streams of records.
- Store streams of records in a fault-tolerant way.
- Process streams of records as they occur.



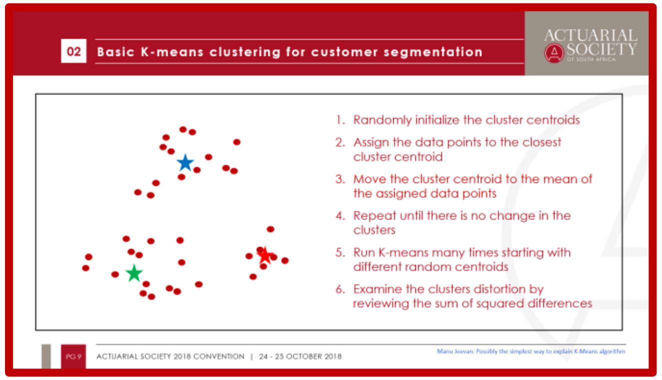
Kafka is generally used for two broad applications:

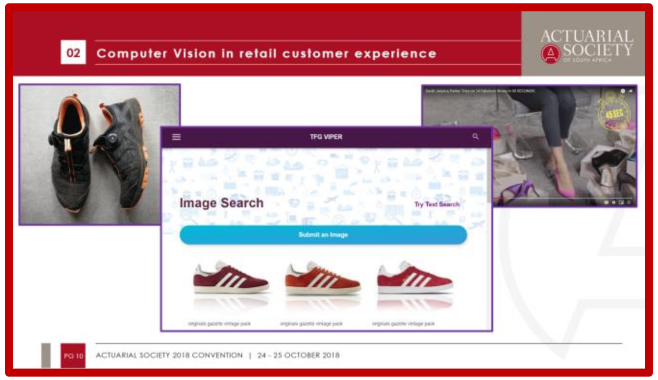
- Building real-time streaming data pipelines that reliably get data between systems or applications
- Building real-time streaming applications that transform or react to the streams of data

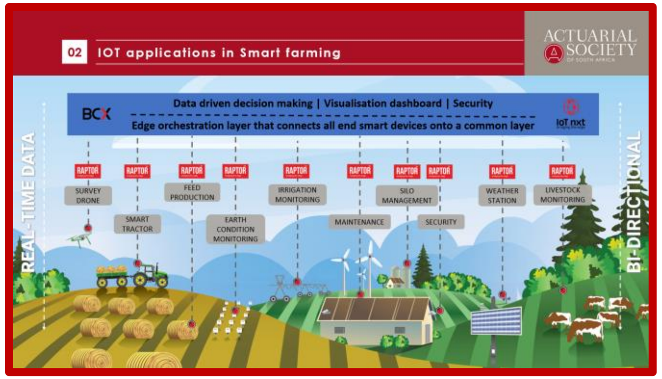
https://kubernetes.io/docs/concepts/overview/what-is-kubernetes/
Datastax Cassandra Tutorials: Apache Cassandra Overview
https://www.cloudera.com/documentation/kafka/1-2-x/topics/kafka.html

The critical context that defines our future landscape

An exponential increase in resources, an acceleration in algorithms, and speed of light limitations driving computing to the edge.







There is practical data science application in every business.

Make sure you capitalize on this disruptive change.