Streaming Data Analytics

Emanuele Della Valle

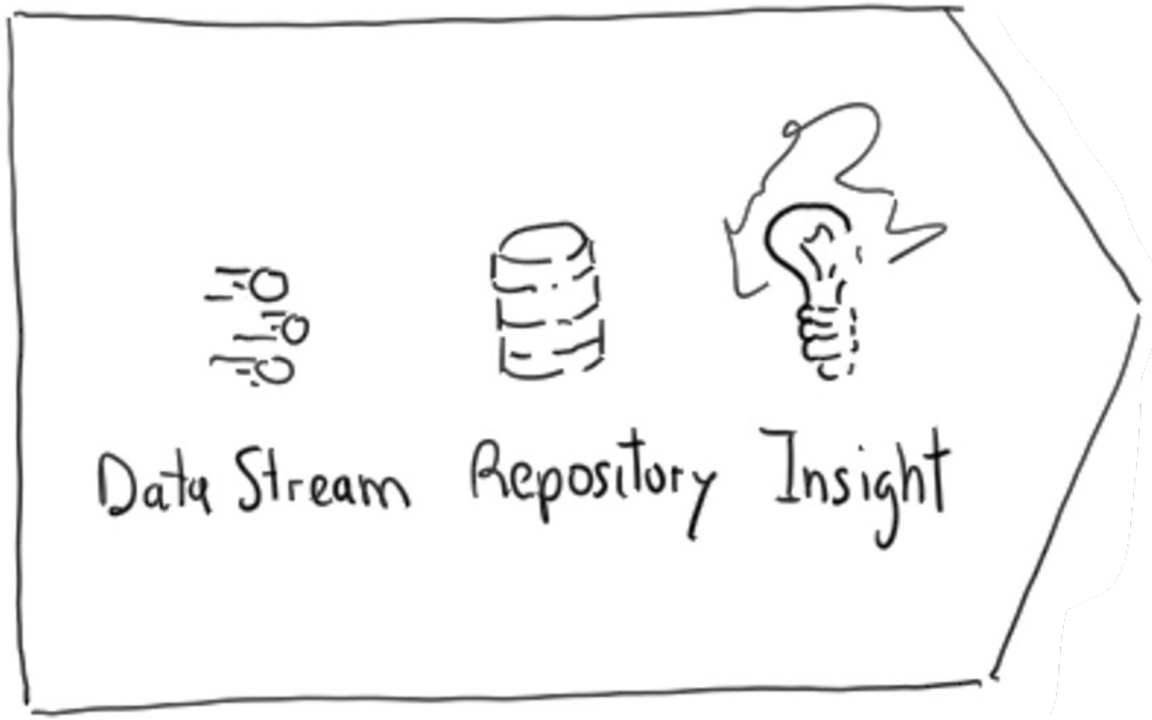
prof @ Politecnico di Milano
Co-founder @ Quantia Consulting
emanuele.dellavalle@polimi.it
http://emanueledellavalle.org

The vision

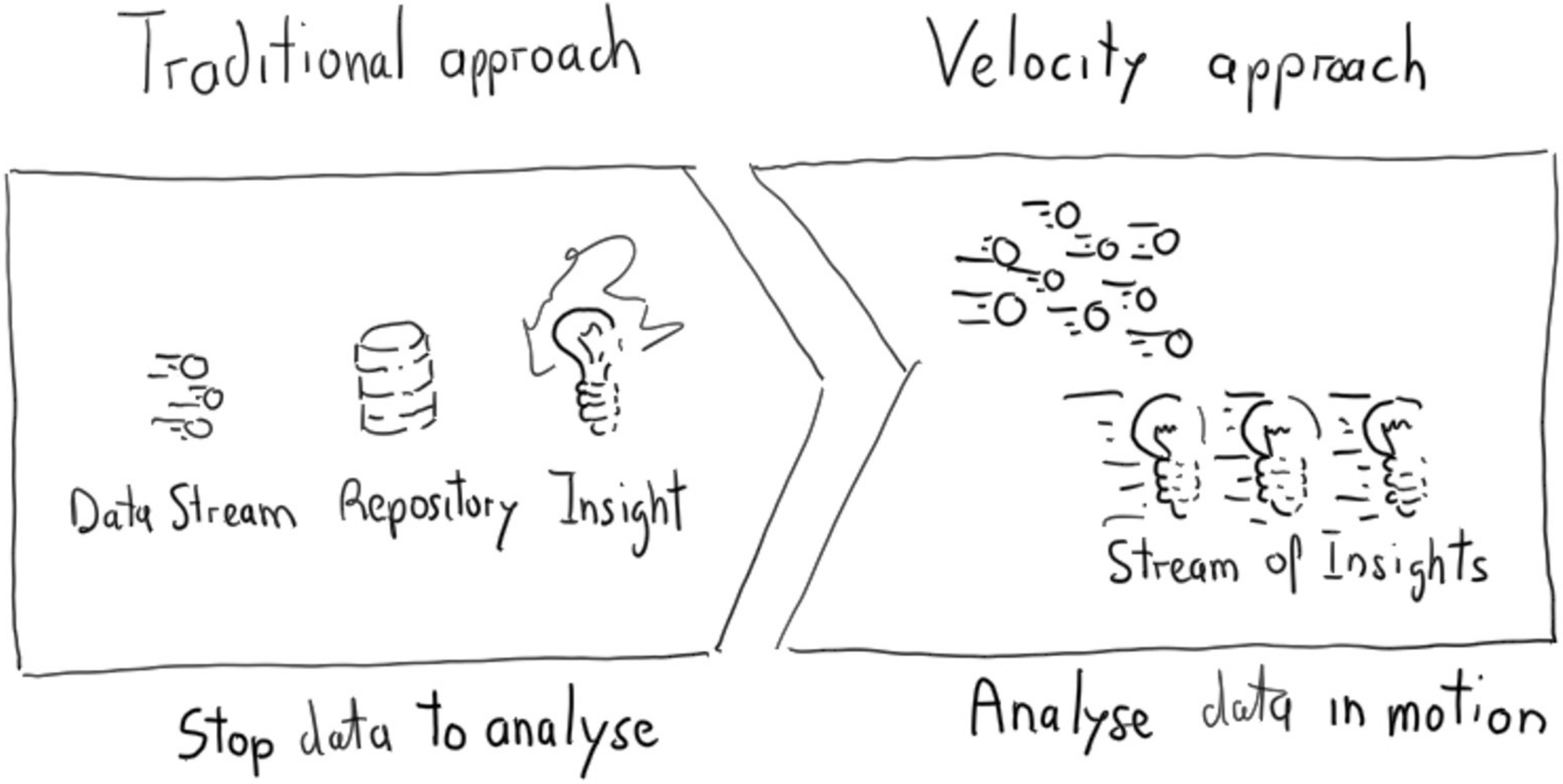
Continuity matters!


Emanuele Della Valle - http://emanueledellavalle.org

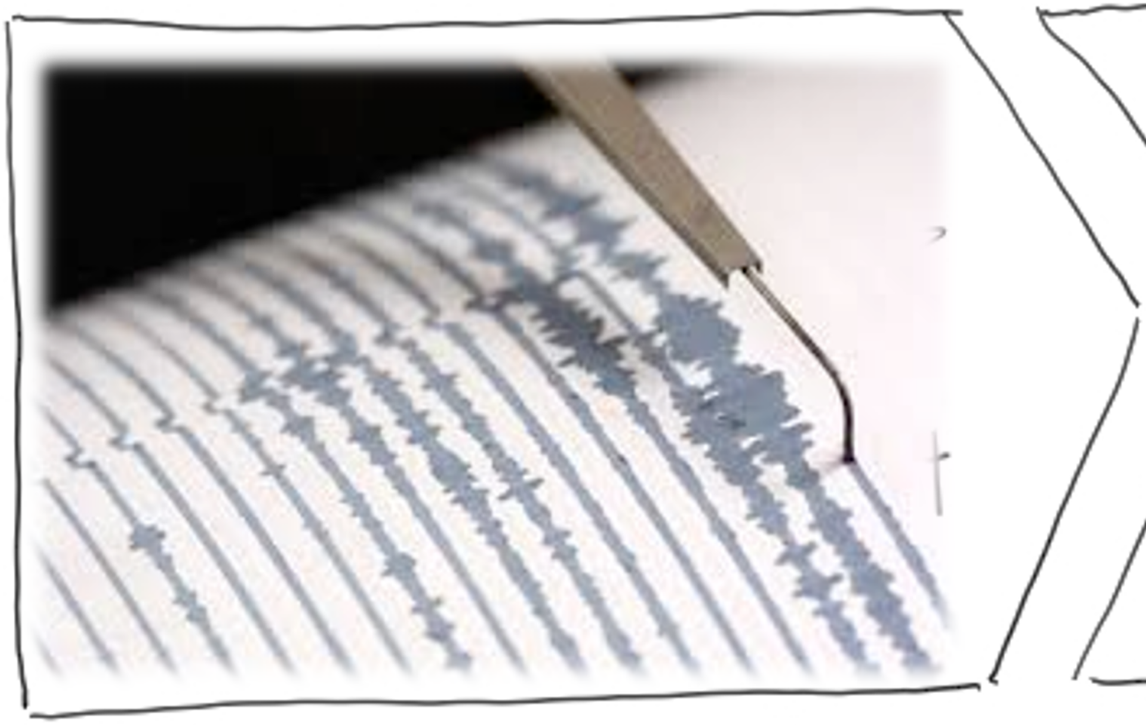
Traditional approach



Stop data to analyse


Traditional approach
Velocity approach
Data Stream
Repository
Insight
Stream of Insights
Stop data to analyse
Analyse data in motion

Traditional approach



Stop data to analyse

Velocity approach

Analyse data in motion

The key ingredients

The key ingredients

1. Sensor and actuators
2. Connectivity
3. Streaming Data Engineering
4. Streaming Data Science
5. People & Processes

The key ingredients

1. **Sensor and actuators** (*as sources and sinks, not covered in the course*)
2. Connectivity
3. Streaming Data Engineering
4. Streaming Data Science
5. People & Processes

Eyes using cameras

Color Cameras

Fixed focus

auto focus

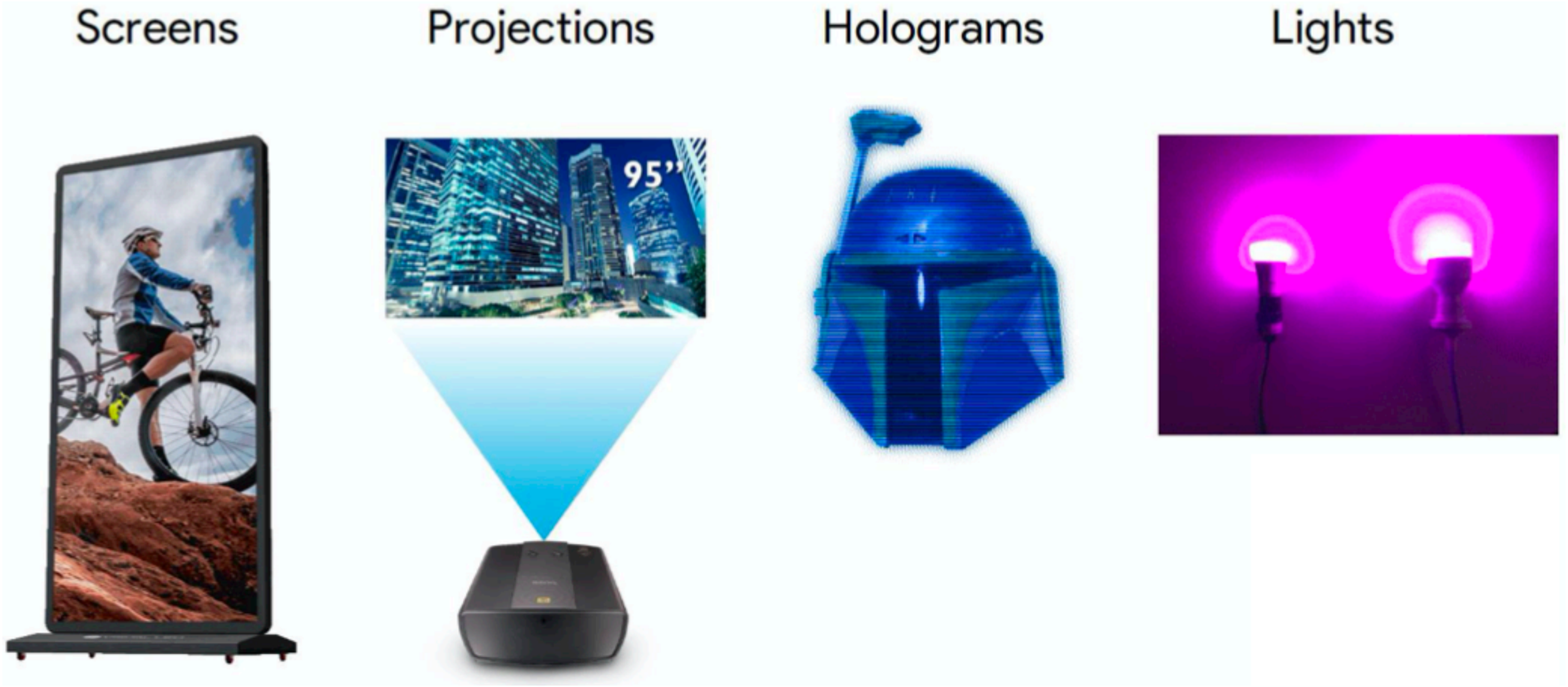
optical image stabilization

Color+Depth Cameras

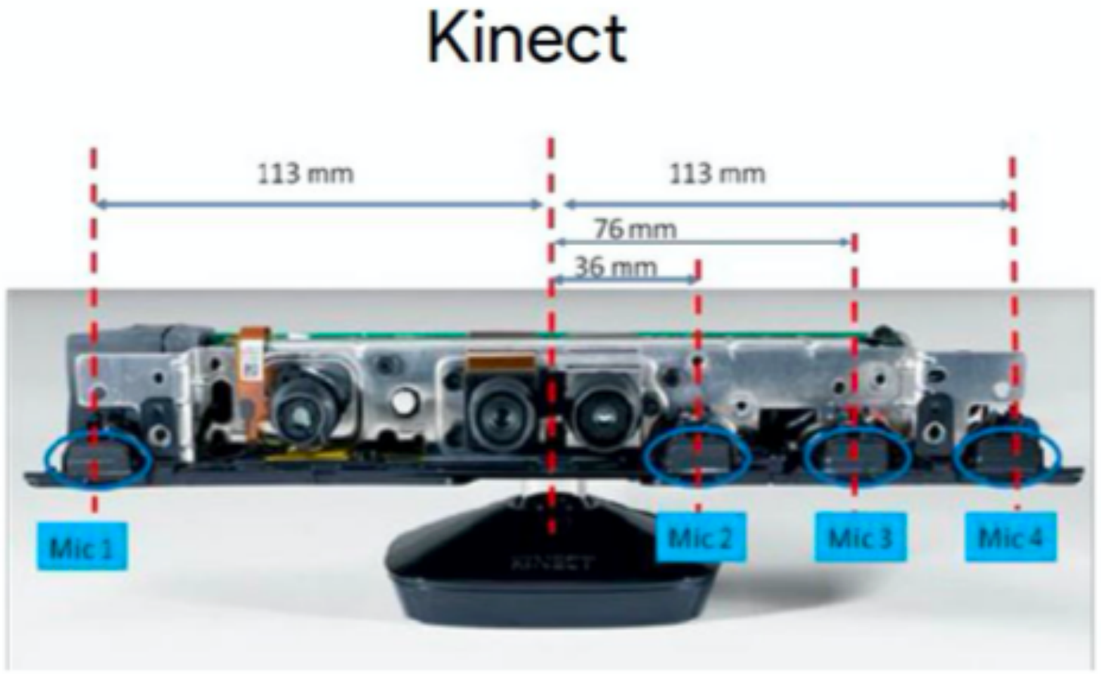
Short range

Long range

Lights to display



Ears using microphones

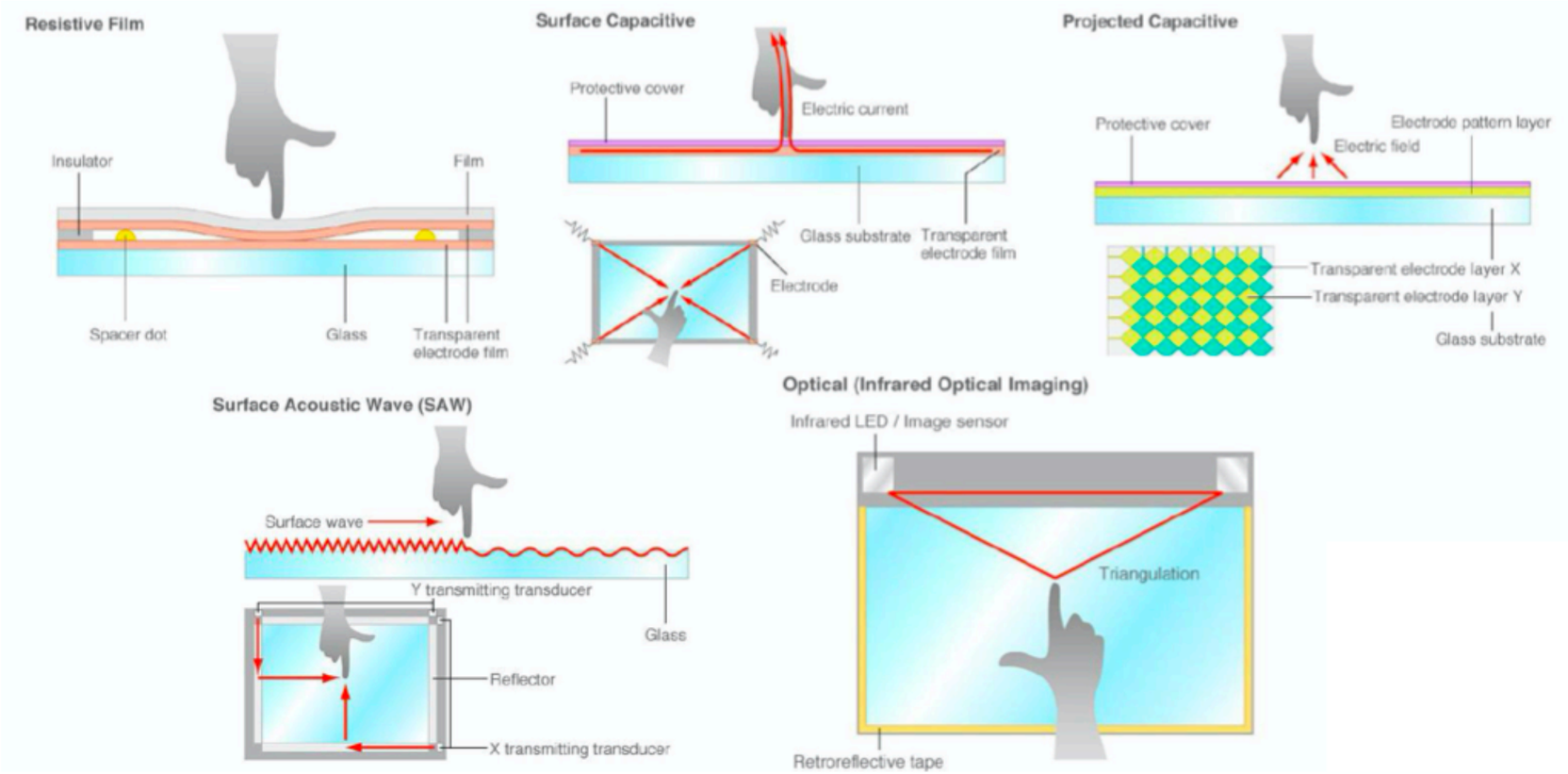


Microphone array

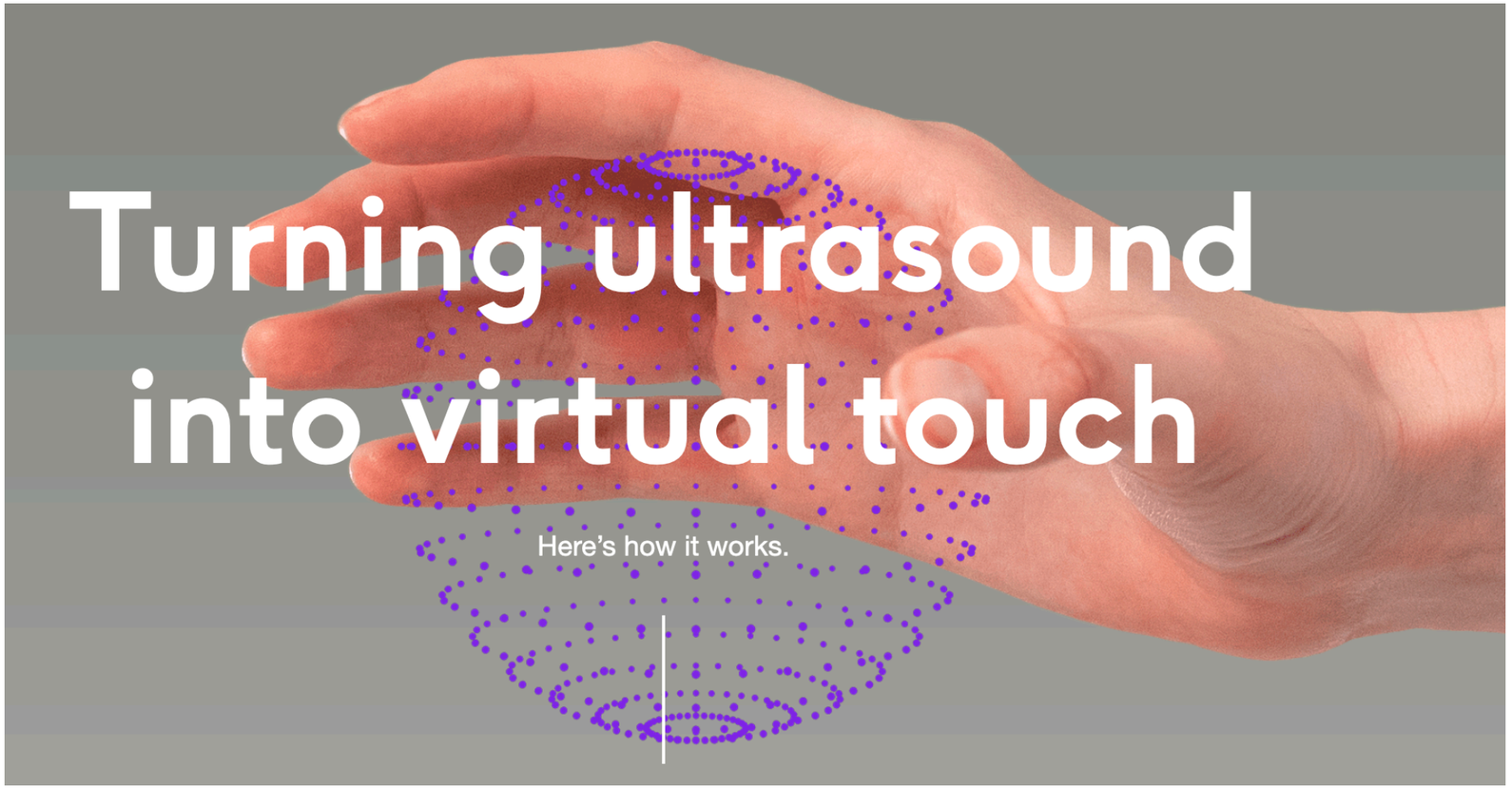
Microphone

Speakers to speak

Skin to feel touches



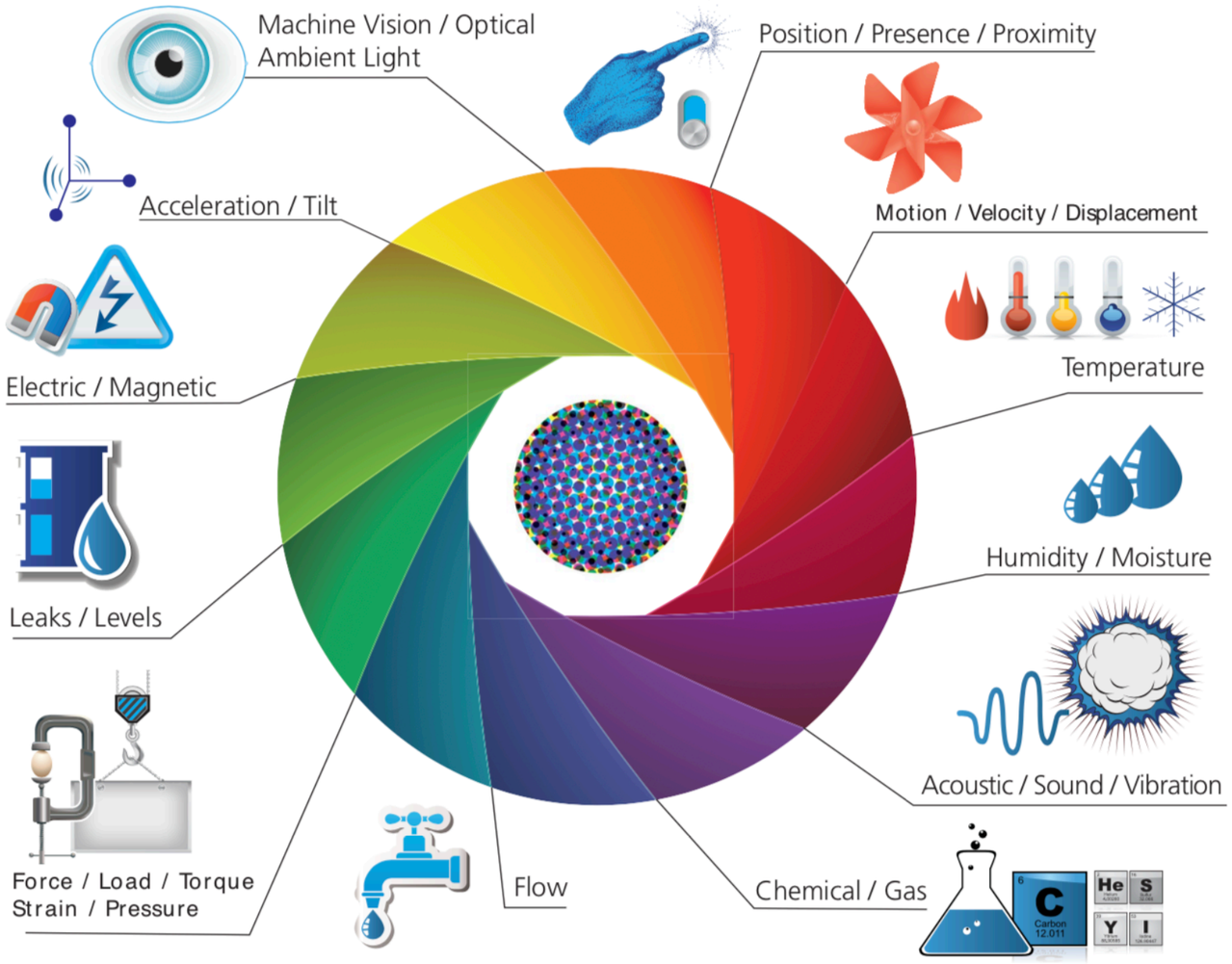
Ultrasounds to generate tactile sensations



Sensors and actuators

… and more!!!

- A **digital nervous system**
  - Orienteering capabilities using **GPS**, **compass**, …
  - Sensory organs that can measure **temperature**, **pressure**, …
- … and an increasing number of **actuators**
  - **harms** and **hands**
  - a **locomotor system**
  - …

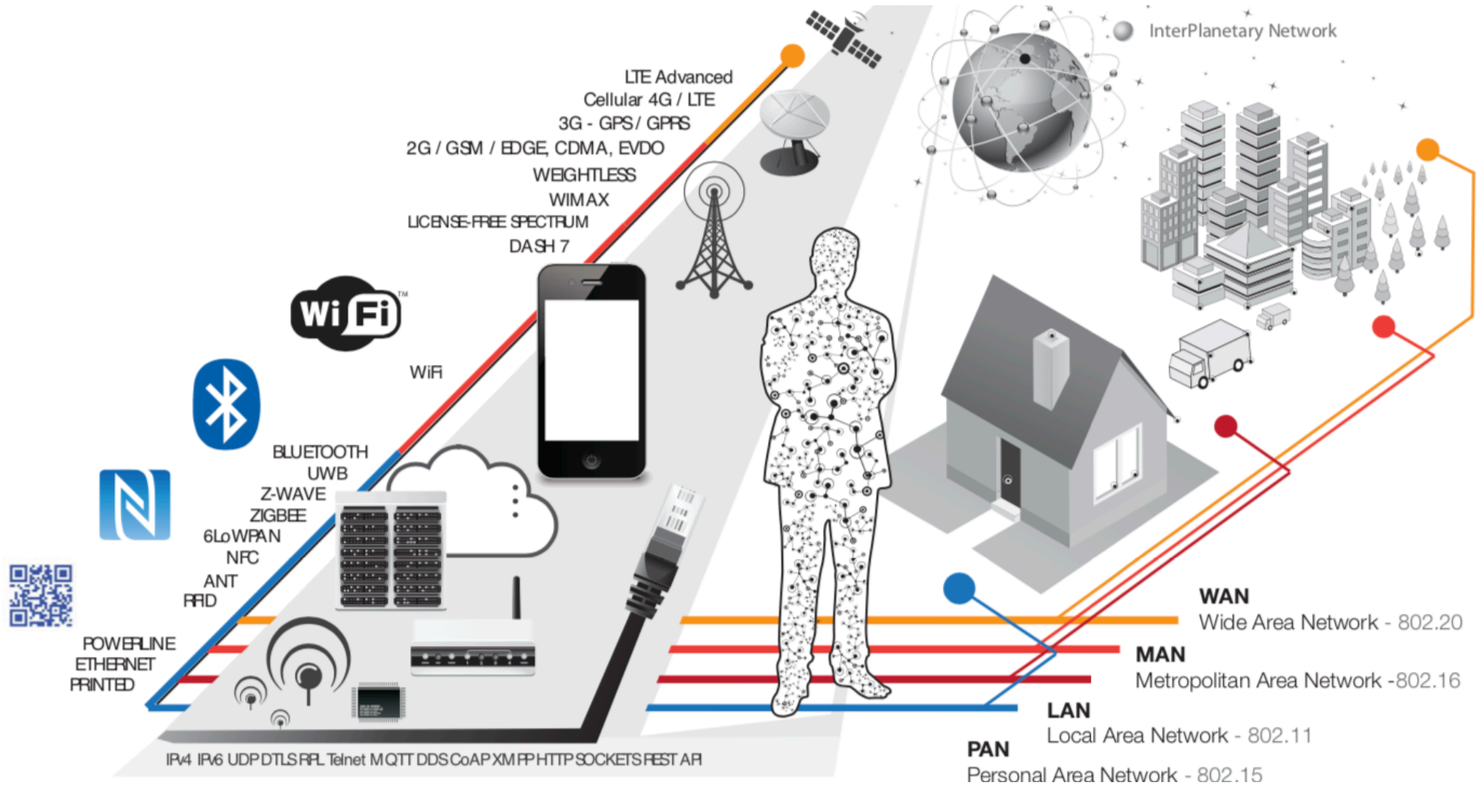


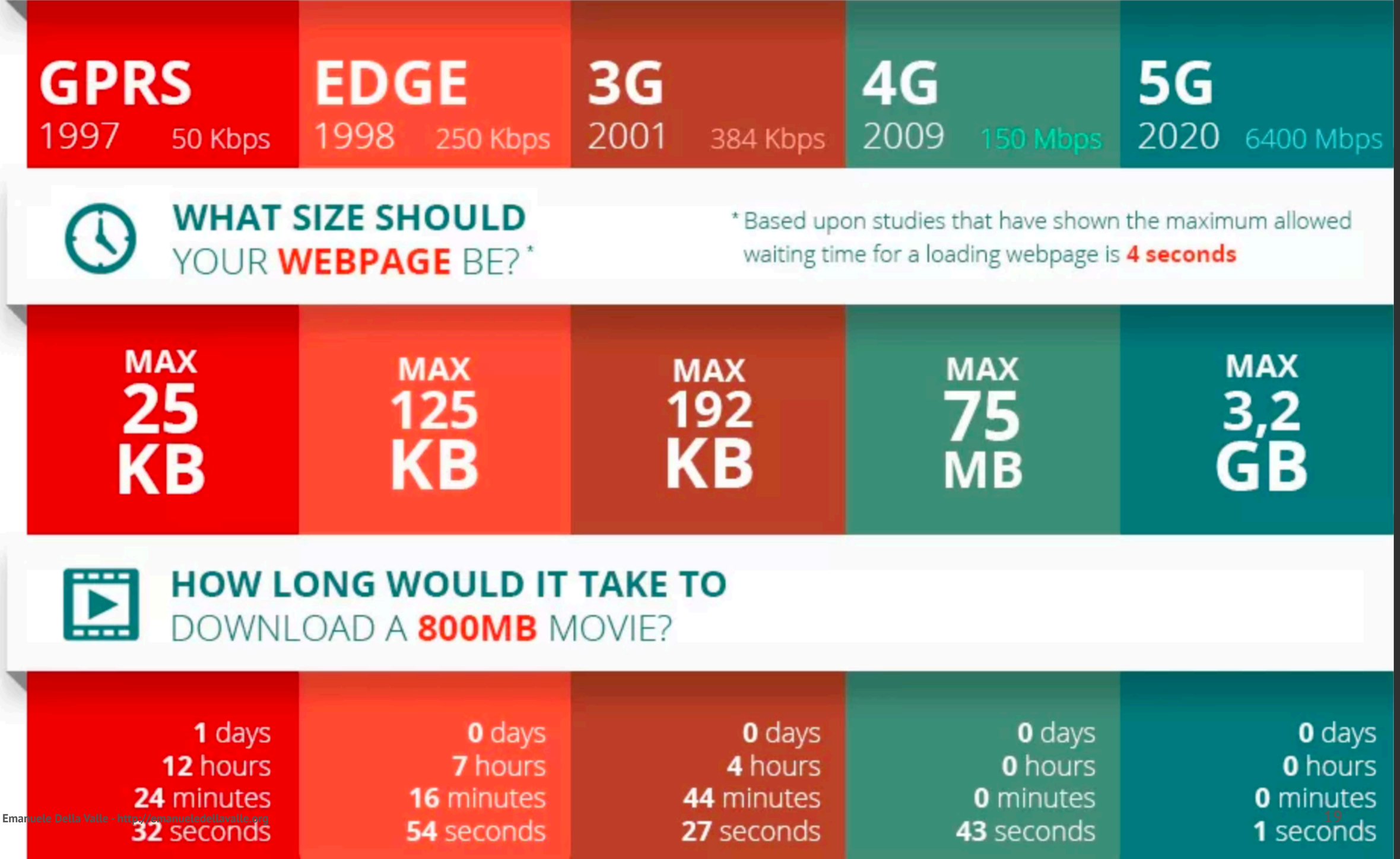
The key ingredients

1. Sensor and actuators
2. **Connectivity** (*as enabling and constraining factor, not covered in the course*)
3. Streaming Data Engineering
4. Streaming Data Science
5. People & Processes

Connectivity

To transport sensory observations and actuation signals



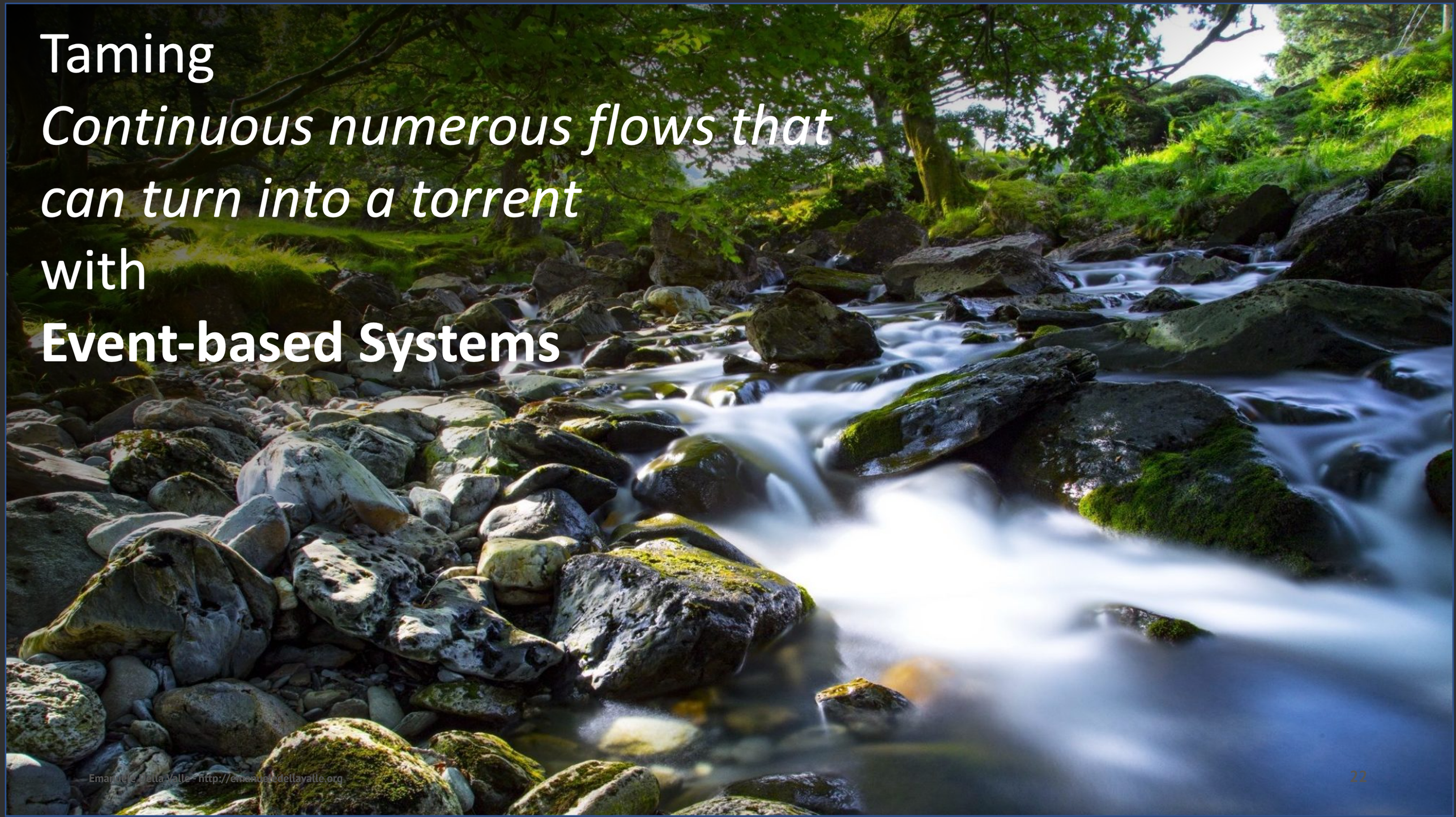

GPRS
1997
50 Kbps
EDGE
1998
250 Kbps
3G
2001
384 Kbps
4G
2009
150 Mbps
5G
2020
6400 Mbps
WHAT SIZE SHOULD
YOUR WEBPAGE BE? *
* Based upon studies that have shown the maximum allowed waiting time for a loading webpage is 4 seconds
MAX
25
KB
MAX
125
KB
MAX
192
KB
MAX
75
MB
MAX
3,2
GB
HOW LONG WOULD IT TAKE TO
DOWNLOAD A 800MB MOVIE?
1 days
12 hours
24 minutes
32 seconds
0 days
7 hours
16 minutes
54 seconds
0 days
4 hours
44 minutes
27 seconds
0 days
0 hours
0 minutes
43 seconds
0 days
0 hours
0 minutes
1 seconds

The key ingredients

1. Sensor and actuators
2. Connectivity
3. **Streaming Data Engineering** (*core topic of the course*)
4. Streaming Data Science
5. People & Processes

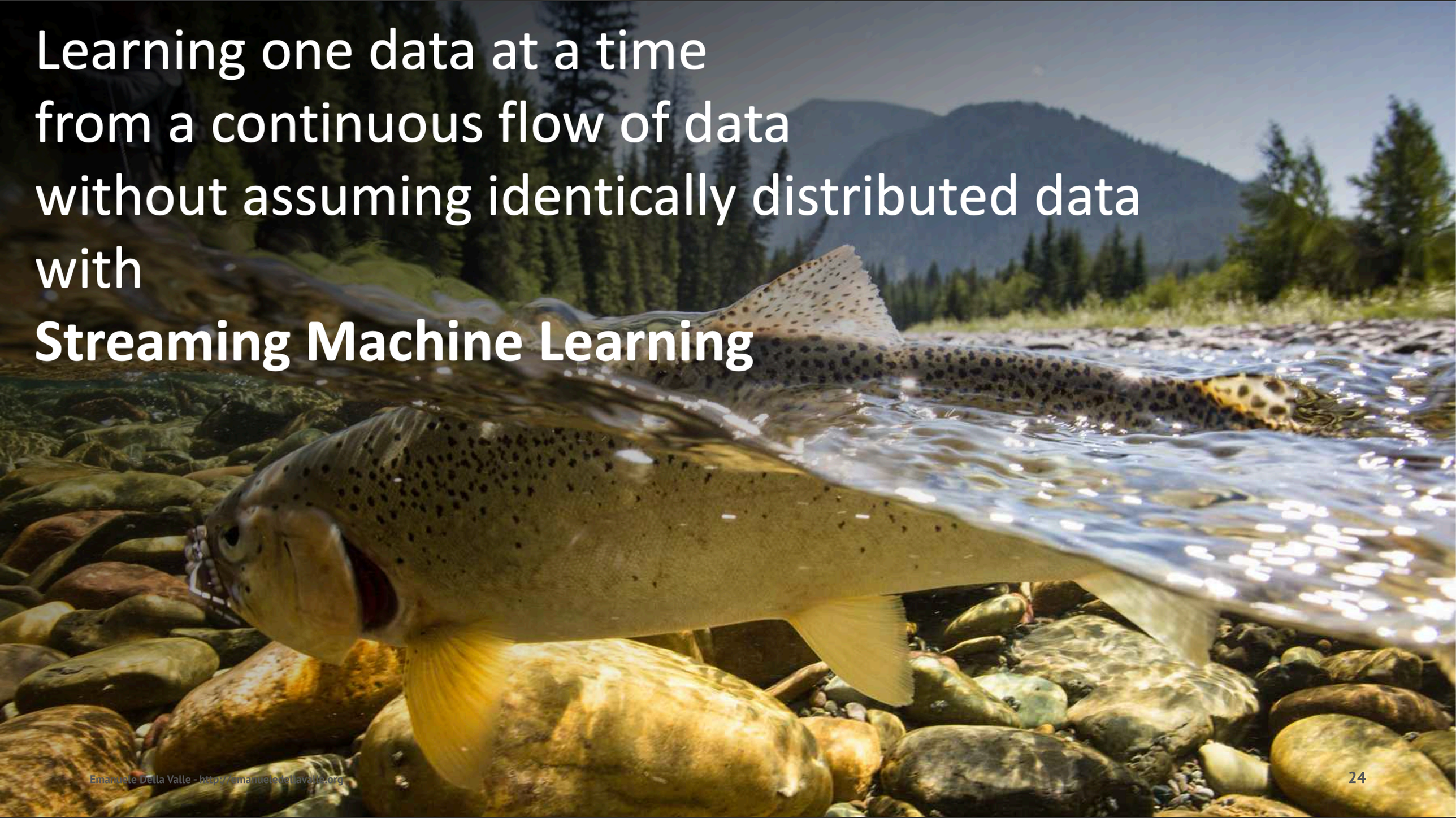

Taming
Continuous massive flows
than you cannot stop
with
Data Stream Mng. Systems

Taming
Continuous numerous flows that
can turn into a torrent
with
Event-based Systems



The key ingredients

1. Sensor and actuators
2. Connectivity
3. Streaming Data Engineering
4. **Streaming Data Science** (*core topic of the course*)
5. People & Processes

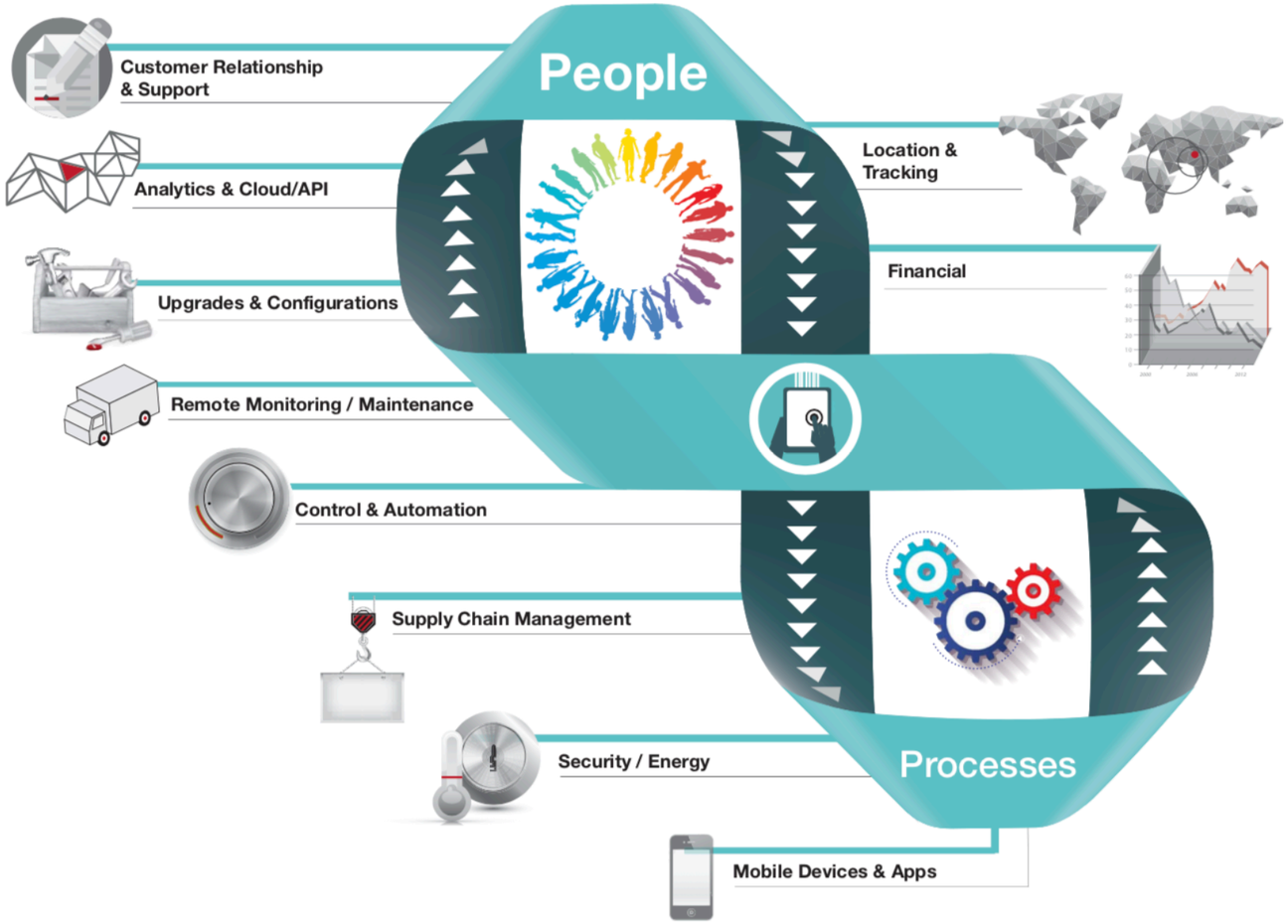

Learning one data at a time
from a continuous flow of data
without assuming identically distributed data
with
Streaming Machine Learning


Explaining the past and forecast the future
of a continuous flow of data
without assuming data independence
with
Time Series Analytics

The key ingredients

1. Sensor and actuators
2. Connectivity
3. Streaming Data Engineering
4. Streaming Data Science
5. **People & Processes** (*to stress real-world applicability*)

People & processes



Case studies

Learning Thermostats

Save resources and money on your heating bills by adapting to your usage patterns and turning the temperature down when you're away from home.

E.g., Nest

Self-driving cars

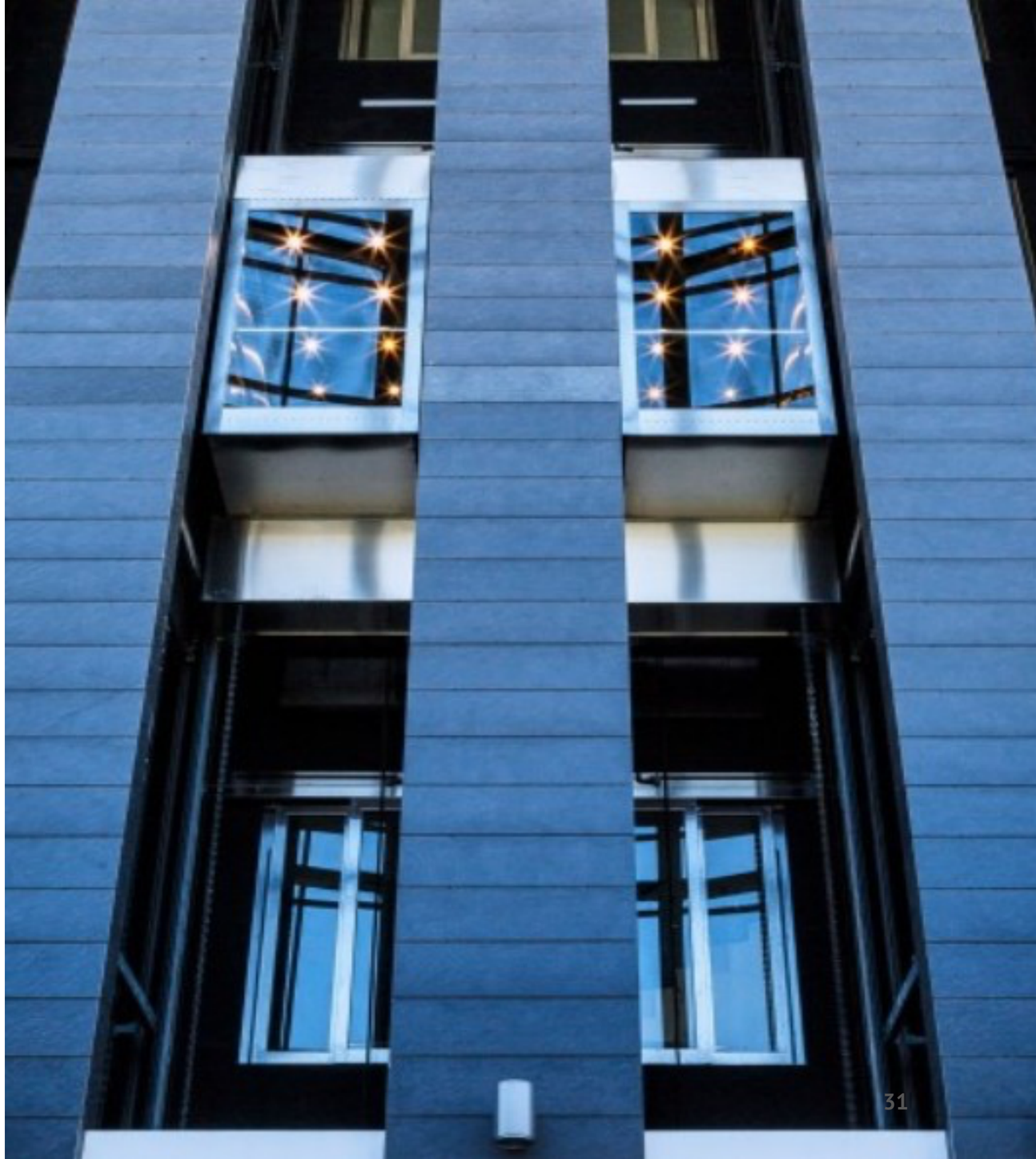
From Google Car research to Tesla autopilot and Vitibot

E.g., A movie I assembled

Smart connected industrial products

The data collected from these products are analyzed to inform decision-making, enable operational efficiencies and continuously improve the product's performance.

E.g., thyssenkrupp's Max elevator



Equipment as a service

Pioneered in 1997 by Rolls-Royce, which lets airlines pay for their engines based on the number of flight hours. It is now a standard business model in many industrial settings.

E.g., Heller's insurance for its machines

Smart logistics

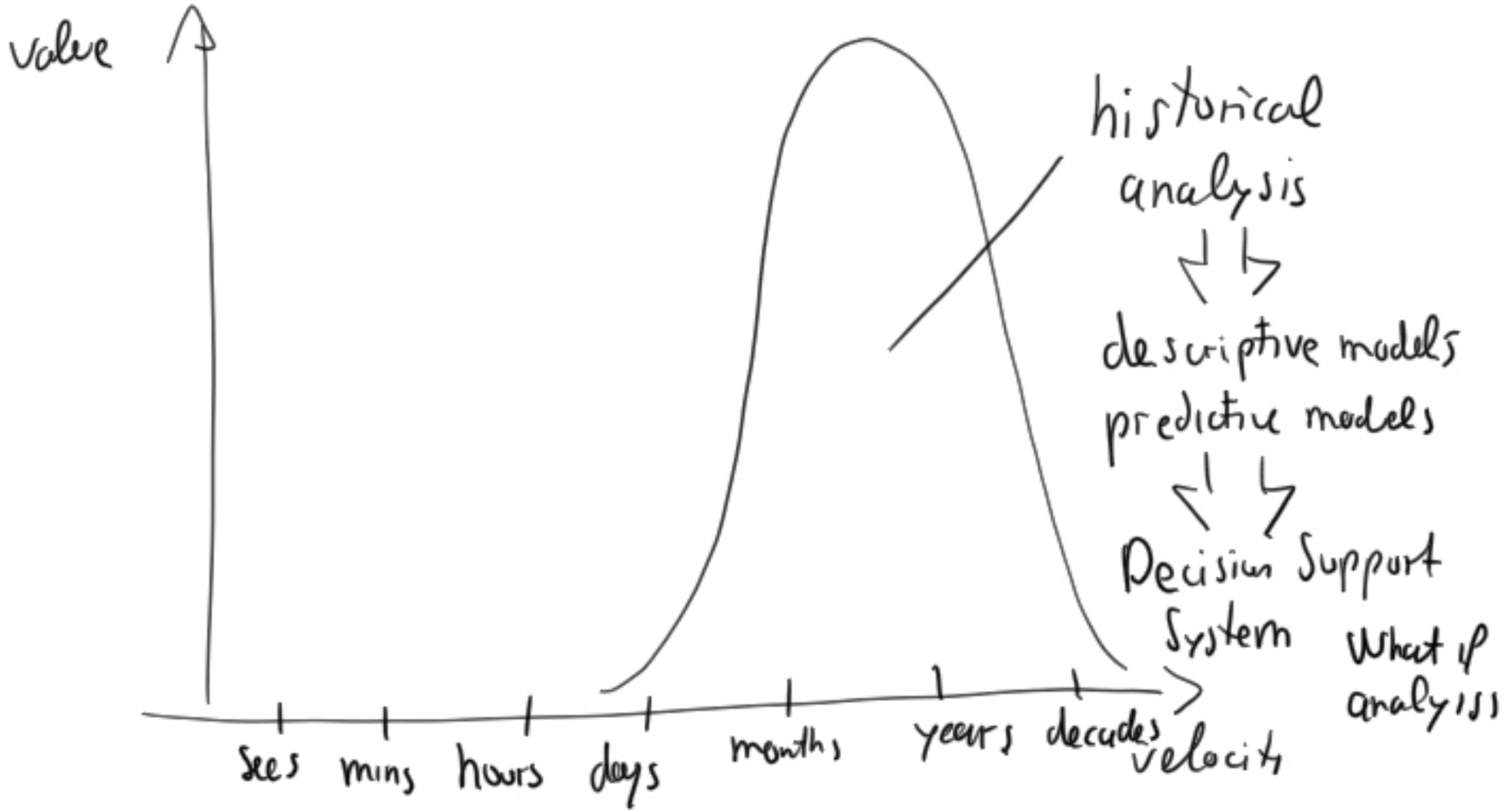
They are employing innovative technology to increase efficiency, effectiveness, and security by making logistics more environmentally sustainable, economically efficient, and capable of handling increasing traffic.

E.g., Hamburg Port Authority's Smart Port

The ultimate goal

Better decision making

Better decision making



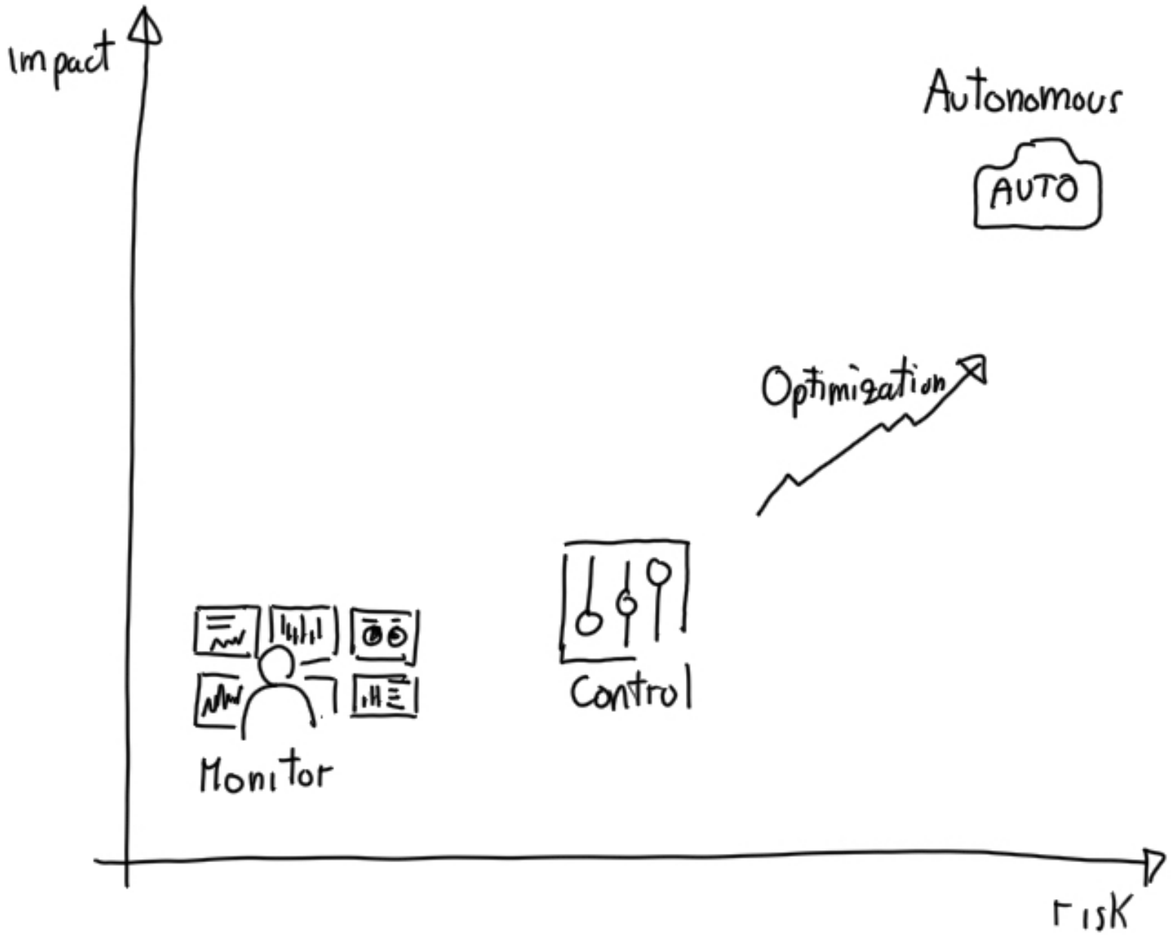
Better decision making

Better decision making

E.g.:

— reduce costs
— improve efficiency
— create innovative products
— new revenue streams

Streaming Data Analysis Maturity Model



Conclusions

It's about time

- **Continuous ingestion**
  - Events instead of tables of data
  - Publish / Subscribe middlewares instead of blob storage solutions
- **Continuous data preparation**
  - Stream- and Event-first data management
  - Continuous processing, no blocking operations
  - Incremental algorithms are a must
- **Reactive answers**
  - minimize latency while keeping high throughput
  - stream of answers

It's about change and temporal dependencies

- **The identically distributed assumption does not hold**
  - Traditional ML hardly applies
  - Change detection plays a key role
  - Streaming Machine Learning is a must
- **The data independence assumption does not hold**
  - Traditional ML hardly applies
  - Time Series Analytics is a must

It's at scale

- Trillions of concurrent messages
- Billions of sensors/actuators
- Often disconnected

- Whenever possible
  - **stay decentralized**
  - **do it incrementally**
  - **minimize resource footprint**
  - **do it at the edge**
- Bring data together only if merging provides a real benefit

Streaming Data Analytics

Emanuele Della Valle

prof @ Politecnico di Milano
Co-founder @ Quantia Consulting
emanuele.dellavalle@polimi.it
http://emanueledellavalle.org