Datalytyx

What really is "big data"?

Big Data, the Internet of Things, Cloud and all that

Peter Chase

18th April 2018

Datalytyx two minute overview

- **50 staff & growing | 30+ specialist associates**
  - 8 years in business – privately owned
  - Mix of Data Architects, Data Engineers & Data Scientists
- **Focus on delivering:**
  - Data Centric Solutions, Data Quality and MDM
  - Data Engineering & Data Analytics
  - Real-time data pipelines & processing
- **Head Office in Central London**
  - Second office in Surrey
  - Offshore team
- **High profile customers & industry leading technology partners**
  - Snowflake Rockies Partner, 2018
  - Talend Partner of the Year 2016, 2017 and 2018

Everyone’s talking about it…

STOP!

Let’s just slow down a minute

- **We’ve had data for ages - what’s different now?**
  - The data explosion
  - The ingredients: data sources, the Internet, processing/storage
  - The result
- **How data solutions are evolving**
- **What should a business consider?**
- **Defining some terms**
- **Points to take away**

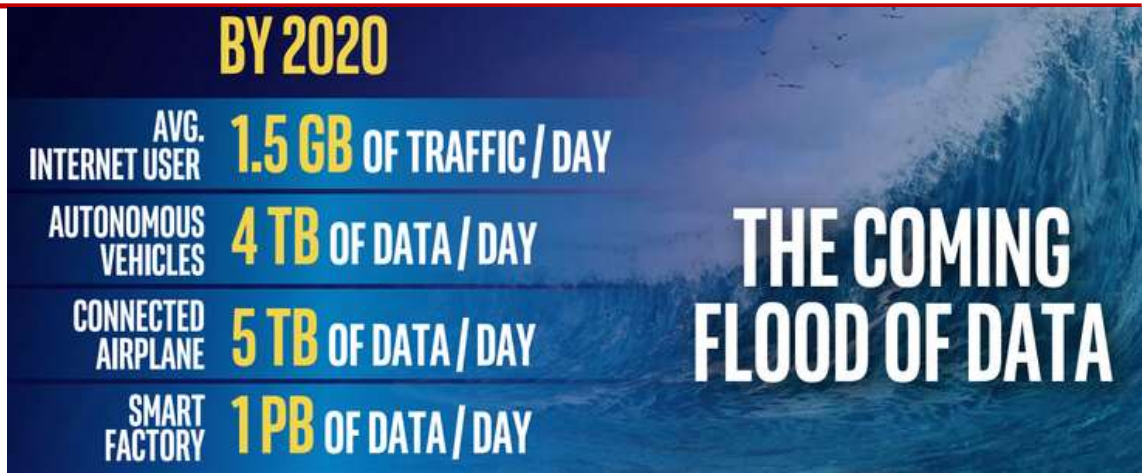
Datalytyx

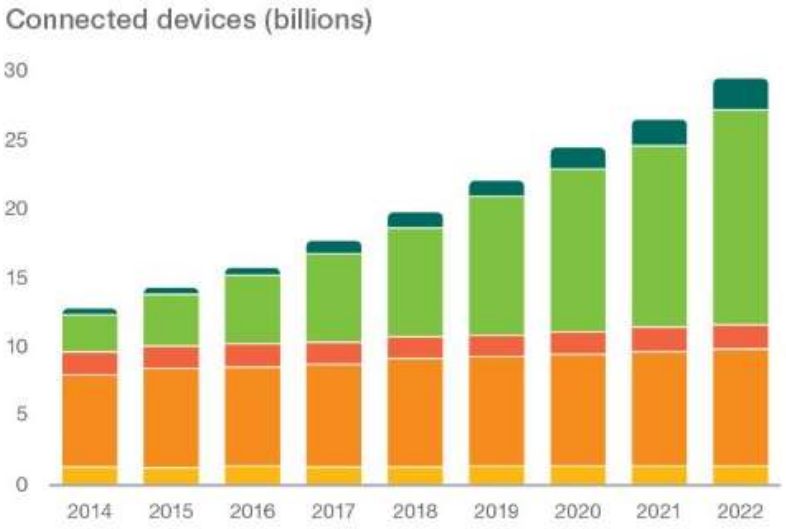
We've had data for ages...

...what's different now?

The data explosion

More data was created in 2017 than the previous 5,000 years of humanity





	2016	2022	CAGR
Wide-area IoT	0.4	2.1	30%
Short-range IoT	5.2	16	20%
PC/laptop/tablet	1.6	1.7	0%
Mobile phones	7.3	8.6	3%
Fixed phones	1.4	1.3	0%
	16 billion	29 billion	10%

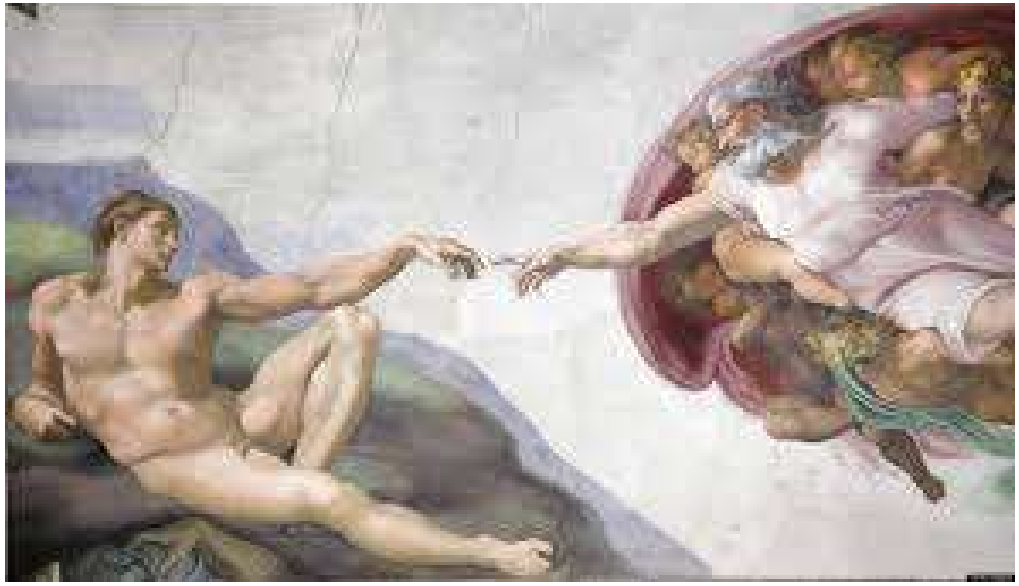
Datalytyx

The ingredients

- **Just imagine if:**
  - there was limitless storage and processing power everywhere, and
  - you could receive or send data from anywhere to anywhere else instantly
- **There's no need to imagine – for most of us this is already true**
- **And it's also amazingly cheap**
- **Let's look at how this has happened:**
  - Ways to make data
  - Communication
  - Processing & storage

The ingredients

- **Just imagine if:**
  - there was limitless storage and processing power everywhere, and
  - you could receive or send data from anywhere to anywhere else instantly
- **There's no need to imagine – for most of us this is already true**
- **And it's also amazingly cheap**



Ways to make data

1990's

2000's

2010's

Datalytyx

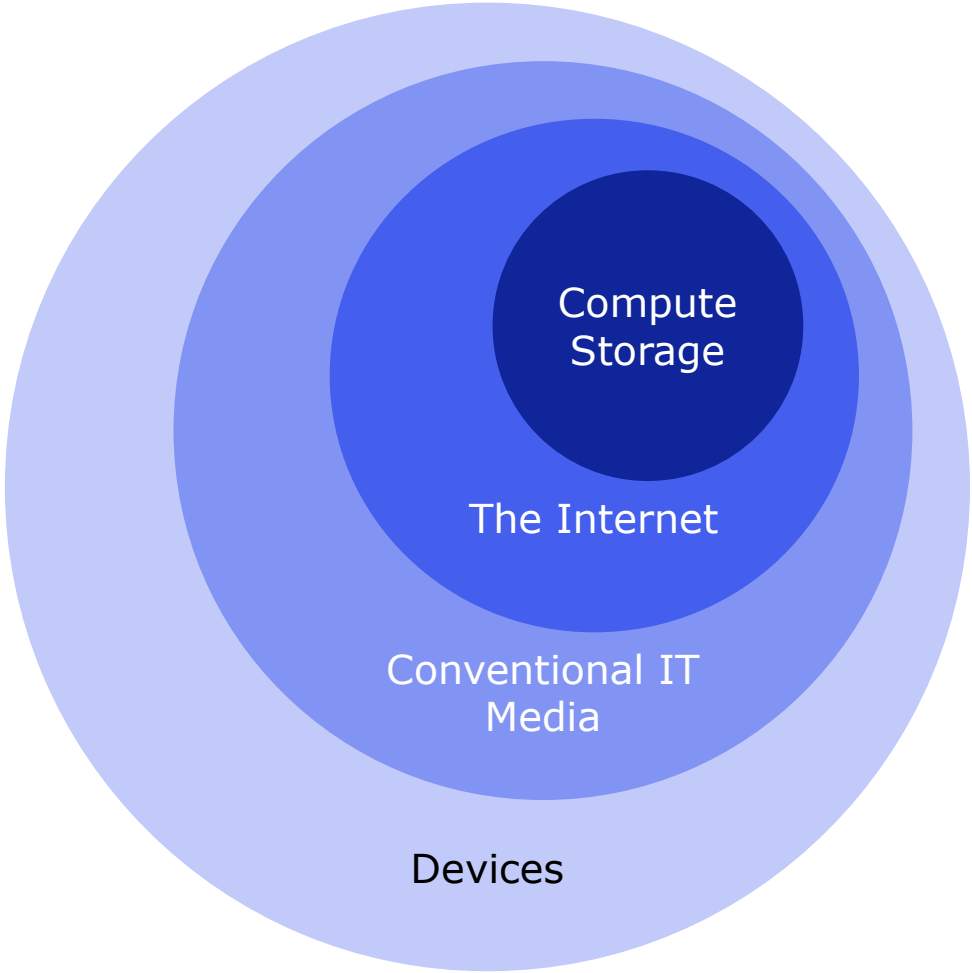
The Internet

- Born "Arpanet", the Internet emerged as de facto interconnect for everything
- Everything's gone digital:
  - The written word:
    - Email boom: 1990 to 2000
    - Web boom: 1995 to 2005
  - Audio:
    - Skype (2000)
    - MP3 / iTunes (2007)
    - DAB radio / Radio players
  - Video:
    - Freeview (2001)
    - BBC iPlayer (2003)
    - Netflix
- Throughout, the telco's have kept up, supplying bandwidth to meet demand

Ways to Process Data

- **Processing/storage**
  - Moore's law has continued
  - Ultra cheap memory/disk
  - Amazing amounts of CPU power
- **Virtualization**
  - Allowing access to this power on an on-demand basis
  - Ability to create web-based accessible services, charged as you use it
    - Amazon – opened up their retail operation's compute resources
    - Microsoft – opened up Office 365 service / Azure
    - Google – opened up their search compute/storage resources

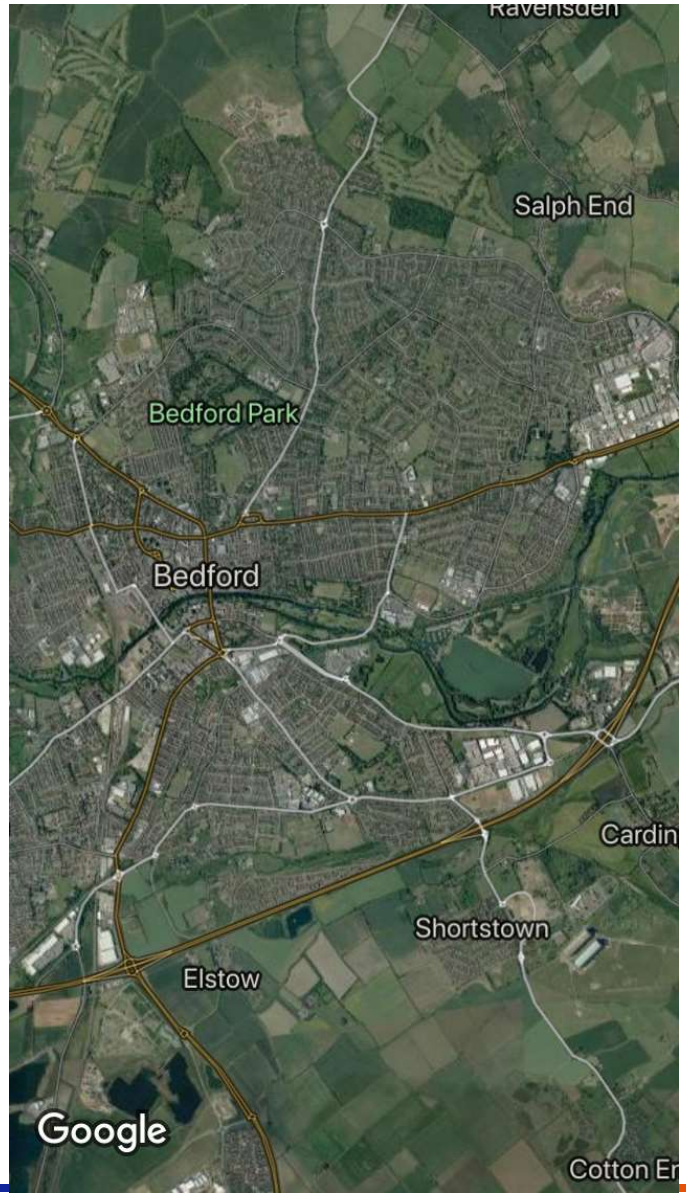
The result

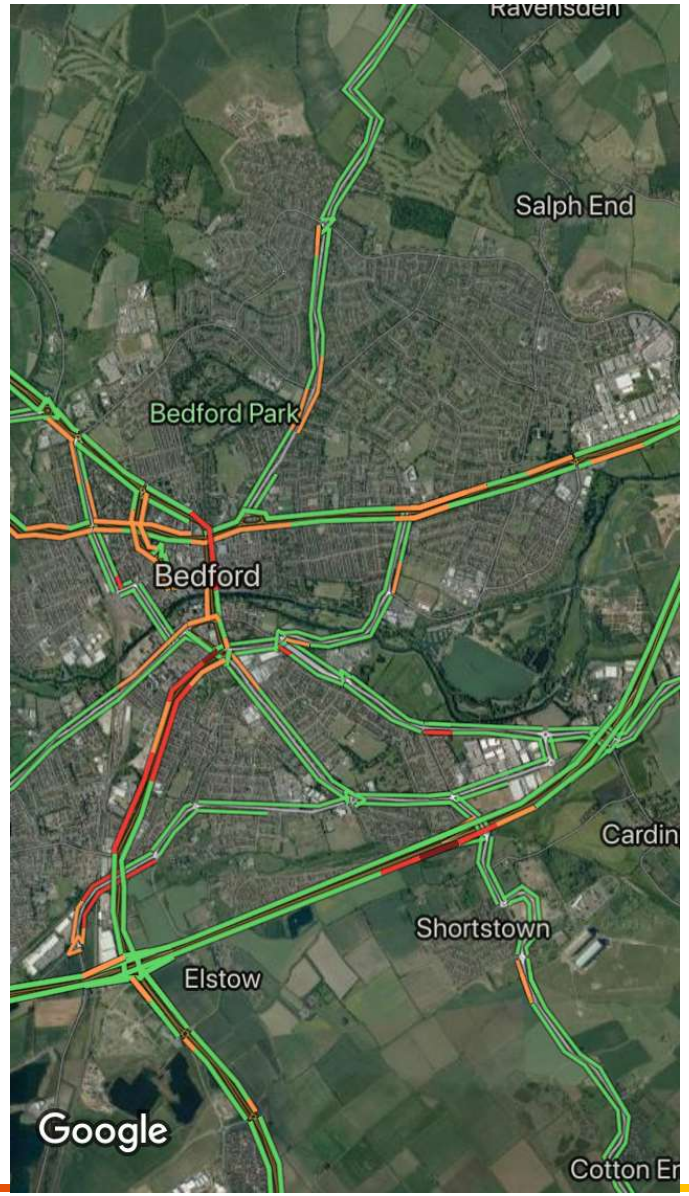


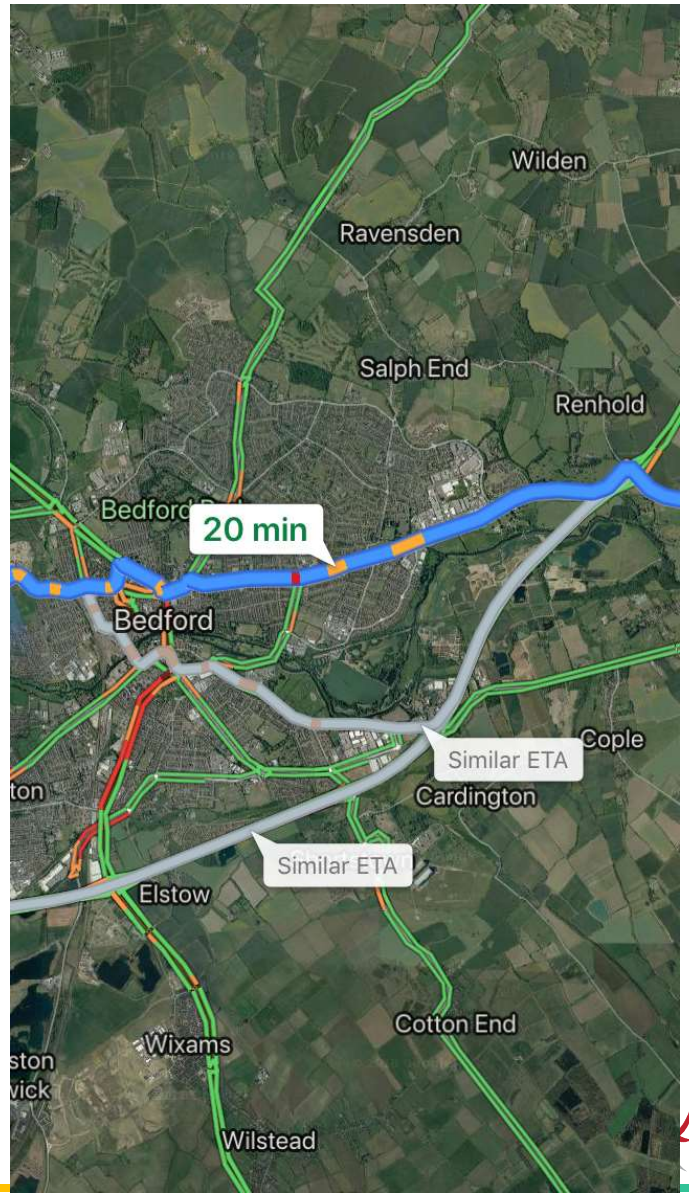
Datalytyx

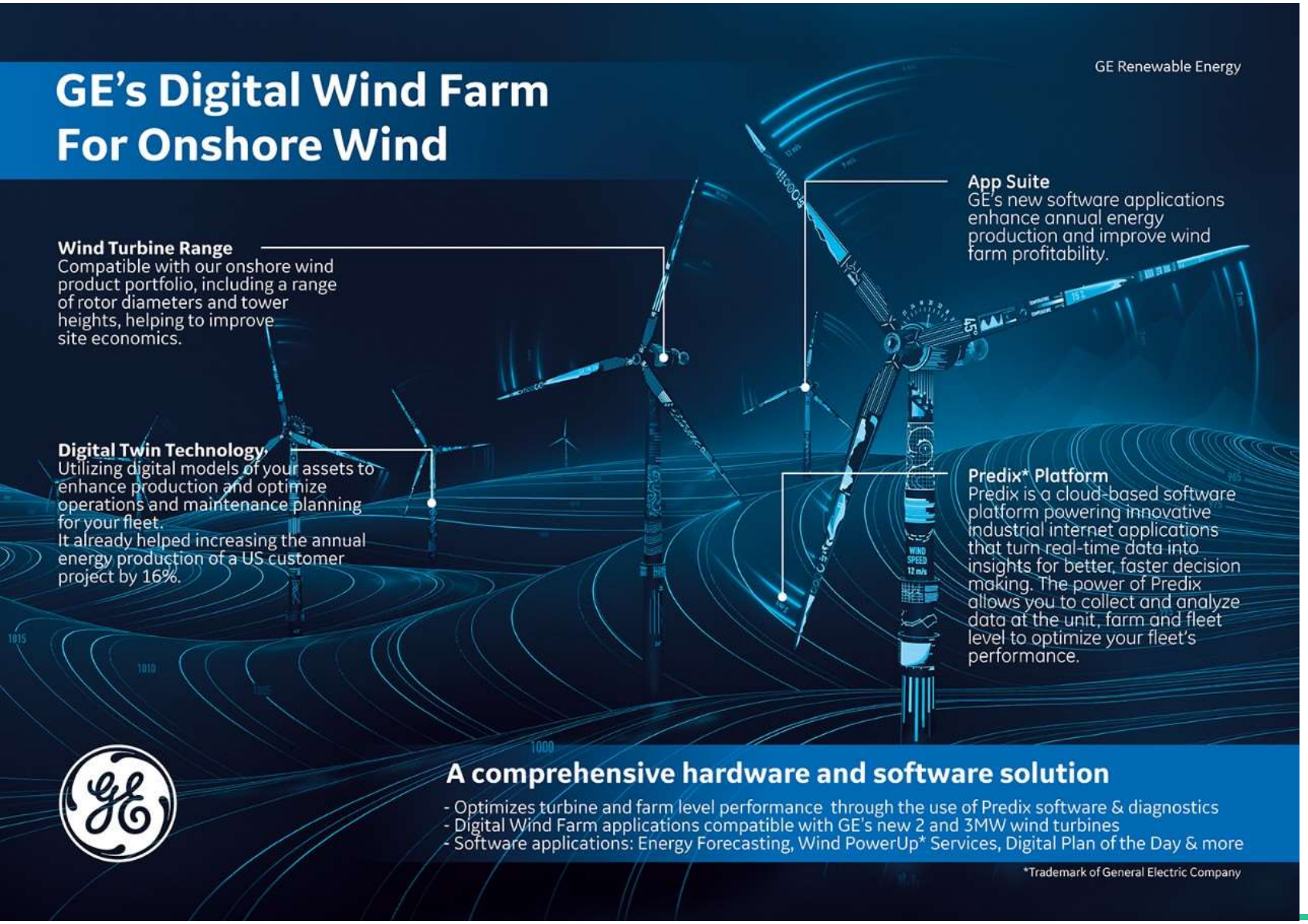
Datalytyx

How data solutions are evolving


Salph End
Bedford Park
Bedford
Shortstown
Elstow
Google


Salph End
Bedford Park
Bedford
Shortstown
Elstow
Google


Wilden
Ravensden
Salph End
Renhold
20 min
Bedford
Similar ETA
Cople
Cardington
Similar ETA
Elstow
Cotton End
Wixams
Wilstead



Wind turbine telemetry

Captured, monitored and used to direct maintenance crews

Only visit turbines when necessary

"Predictive maintenance"

Datalytyx

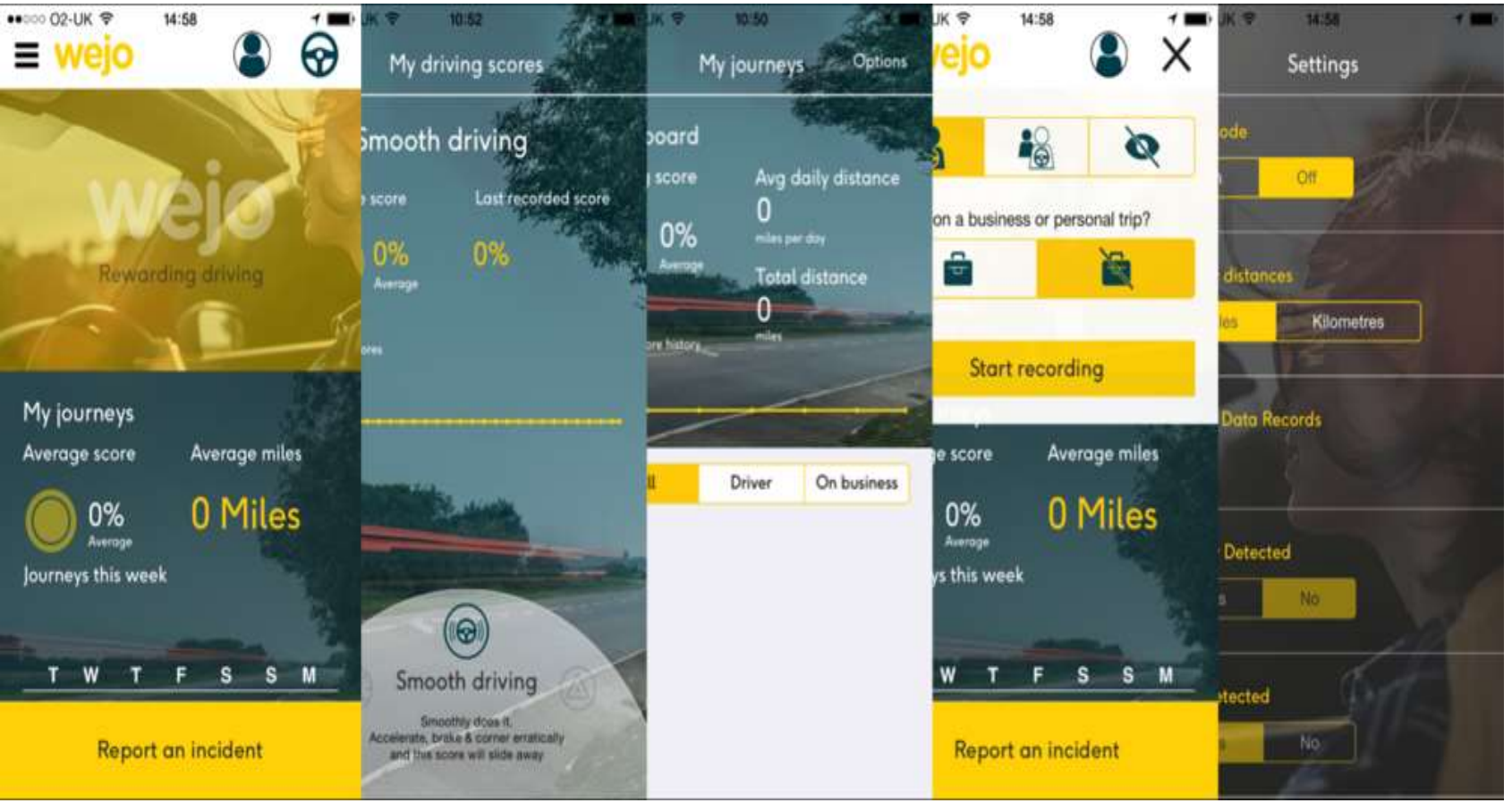
Jet engine telemetry

Captured, monitored and used to direct Rolls Royce maintenance engineers

RR sell engines as a service, not as assets

“Servitization”

Datalytyx



Wejo collect car location and driving behaviour

Anonymized and sent to insurers, retailers and manufacturers

Drivers receive rewards, e.g. free parking, coffee vouchers

"Rewarding driving"

Datalytyx

flybe.

Flybe crew receive daily itinery / plane assignments via iPad

Calculate load sheets for each flight

Log incidents

Integrated with back-end information systems by Datalytyx

Datalytyx

Increasing revenue and customer satisfaction across its worldwide customers by analysing player behavior patterns in real-time and delivering a personalised betting experience.

Relevant and personalised betting offers to 500,000+ customers

BETVICTOR

Datalytyx Managed Service for Integration, Analytics & Performance

Customer Data Sources

CORE BUSINESS DATA SOUCES

- Data warehouse
- Marketing
- SMS
- Text local
- Various Providers
- Etc.

BETVICTOR

OTHER POTENTIAL DATA SOURCES

- Financial Systems
- Cloud Systems
- Customer Satisfaction Systems
- 3rd Party System
- Web Logs & Click streams
- Machine generated
- Geolocation data
- Marketing automation systems

OTHER FLATFILE SOURCES

- (Spread Sheets)
- (CSV Export)
- (Text Delimited)

DATA CLEANSING / MANAGEMENT

Various Extracts

Semi Clean Data

Various Extracts

Various Extracts

talend

Data Cleansing & Processing Rules

CSV

High Quality Data Lake

cloudera

Clean Data

Data Preparation, Calculations, Aggregations & MDM

talend

PERFORMANCE MANAGEMENT

SECURE FTP DROPZONE

Scorecards

Action, Issue, Risk,

KPIs, SLAs, OLAs, Contracts

DATA ANALYTICS

Regression Algorithms

Regularization Algorithms

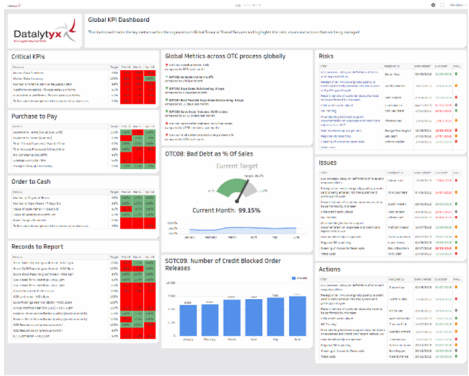
Decision Tree Algorithms

SAS, Python, R, SparkML

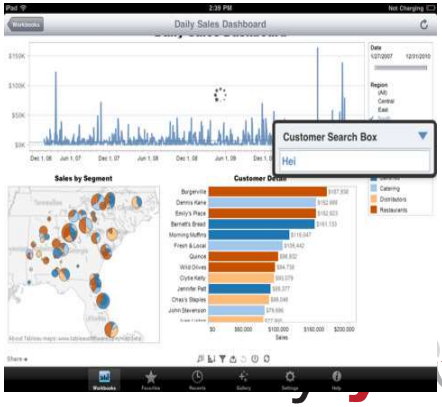
DATA VISUALISATION

Qlikview,Tableau, Zoomdata

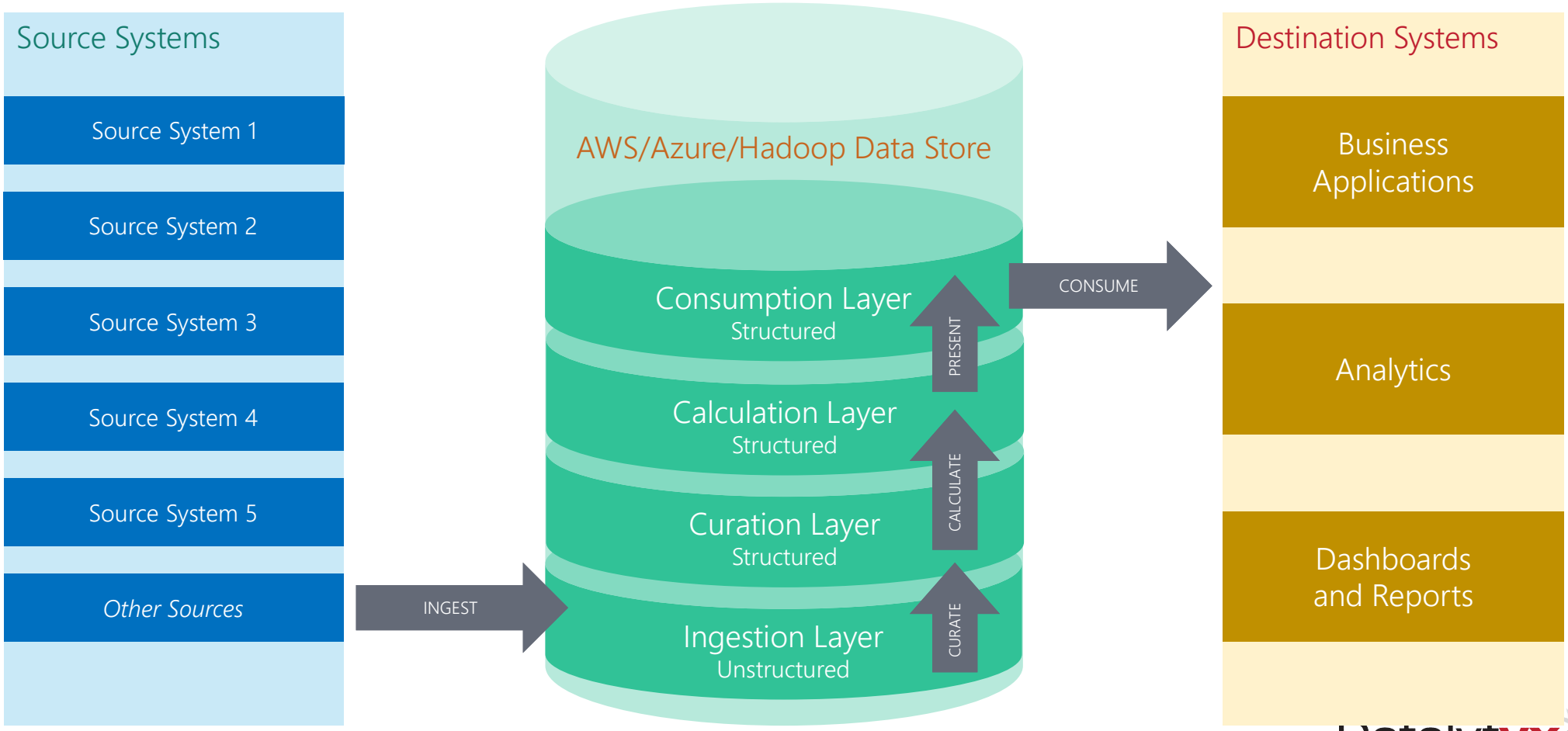
Performance Scorecard



Analytics Dashboards



Modern Data Warehouse Architecture



Datalytyx

Datalytyx

What should a business consider?

So what should a business consider?

Go Cloud: buy basic IT services from the cloud

Be "Data First"

Think about a data future

So what should a business consider?

- **Go Cloud: buy basic IT services from the cloud**
  - You don’t own your own generator to make electricity for your office
  - So why buy and manage your own servers?
  - Gone are the days of installing your own servers for email, intranet, etc.
  - Use the SAAS offering from all your business software providers
    - if they don’t have one, consider changing

So what should a business consider?

- **Be "Data First"**
  - Consider how data can give you insight to your customers' needs
  - Pool data about customers into one place, to make a "single view of the customer"
  - Ask yourself if data can help you run your operations more efficiently
  - Be prepared for data protection questions and GDPR

So what should a business consider?

- **Think about a data future**
  - Can you add value by providing information (data) to suppliers or customers?
    - Power utility smart meters let customers see their usage statistics
    - Suppliers of Travis Perkins pay for an app which shows how well their product lines are selling
  - How could competitors use data to leapfrog your offerings?
    - If they could, then transform your own offerings before it's done to you
  - Consider turning selling assets into services: "Servitization"
    - Can you sell people a service, i.e. the value something provides, rather than the thing itself

Datalytyx

Speed masterclass in data terminology

Big Data

Data has a number of characteristics – all beginning with "V":

- Volume – how much of it is there, how much is being made?
- Velocity – is it moving, and if so how fast?
- Variety – does the data itself, and more importantly its format, change?
- Veracity – can we trust it to be correct?
- Value – what benefit can it bring?

Big Data is defined as:

- Large Volume
- Large Velocity
- Large Variety

The Internet of Things

A pretty recent phenomenon…

- **Brought on by the right conditions (see above)**
- **Multiplier effect:**
  - small packets across thousands / millions devices = **BIG DATA**

Also high Veracity – devices generally produce reliable, accurate data

Also high Value, for example:

- Geo-location (as above) for traffic management
- Asset telemetry: jet engines, wind turbines, anything hard to reach or expensive can be monitored from a distance, "predictive" maintenance
- Healthcare: patient sensors, remote diagnostics, targeted interventions

Cloud Computing

What Cloud computing isn't:

- A website
- Putting a PC on the Internet and "remote desktopping" to it to access an app
- Logging into a web server on the Internet which runs an application/suite

Cloud Computing is:

- Creating a virtual server by filling in a form and clicking OK
- Being able to change the RAM, disk or compute power up/down instantly
- Paying by the second for the CPU cycles used
- Calling a serverless function (written in JavaScript, Python or C#), e.g. to reformat a data value (AWS call these Lambda functions)
- Accessing a database managed as a service (e.g. Snowflake) specifying the amount of compute power to apply as part of the SQL query

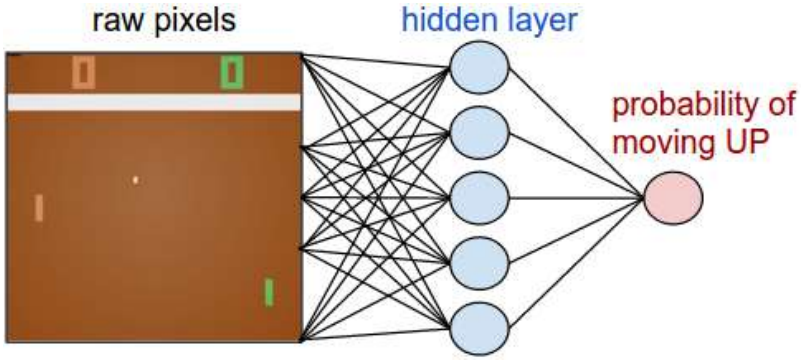
Machine Learning

Machine learning is:

- the repeated recording of varying conditions and resultant outcomes
- from which a computer can deduce the likely outcome of a future set of conditions

Example: Pong

1. Take in images from the game and preprocess them (remove color, background, downsample etc.).
2. Use the Neural Network to compute a probability of moving up.
3. Sample from that probability distribution and tell the agent to move up or down.
4. If the round is over (you missed the ball or the opponent missed the ball), find whether you won or lost.
5. When the episode has finished(someone got to 21 points), pass the result through the backpropagation algorithm to compute the gradient for our weights.
6. After 10 episodes have finished, sum up the gradient and move the weights in the direction of the gradient.
7. Repeat this process until our weights are tuned to the point where we can beat the computer.



Datalytyx

Artificial Intelligence

The apparent ability of a computer to perform a supposedly complex or "clever" task a human might do

- Play pong
- Play chess
- Weld a car body together
- Predict stock market movements
- Navigate a route across the country
- Drive a car

Do you really believe the computer is "clever"?

These are specific tasks computers have been specifically programmed to perform

Artificial General Intelligence

Artificial General Intelligence seeks to:

- combine machine learning
- and pre-programmed abilities to perform certain skills (AI)
- in order to provide a general purpose, adaptable machine suited to a particular environment or set of complex tasks

Examples:

- Minefield clearance
- Deep sea exploration / underwater maintenance
- Home help: Siri / Alexa / Google Assistant / Cortana (not yet, but you wait…)

Say, in five years time, you gave Alexa a body. What would you have?...

Datalytyx

Points to take away

Points to take away

- **Supply of the world's computing, storage and comms resources exceed demand**
- **The result is a boom in data generation and processing**
- **Previously inconceivable uses are possible, from airborne telemetry to intelligent cities to genetic analytics**
- **Businesses can easily collect and gain insight from data about their customers and services they consume**
- **Not doing so is simply to give ground to competitors**

Datalytyx

Q&A

Thanks for listening