



Institute
and Faculty
of Actuaries

Opening the Box on Big Data and Probing the Fog of The Cloud

Tony Lovick – RPC Consulting

Tony.Lovick@rpc.co.uk

22 September 2016

Articulate
Sponsorship
Thought leadership
Progress
Community
Sessional Meetings
Education
Working parties
Volunteering
Research
Shaping the future
Networking
Professional support
Enterprise and risk
Learned society
Opportunity
International profile
Journals
Support

Agenda

- Big Data
 - What's new?
 - Why bother?
 - How do I do it
- Cloud Computing
 - Cloud as a Service
 - Can I really do that?
 - Everything must be in the Cloud
 - ... because it's cheaper



Institute
and Faculty
of Actuaries

Big Data – What's New?

Big Data is just Data:

- Velocity, Volume, Variety

011010011010011010011010011010011011010011010011010011010011010011010011010011010011010011010

Not Big Data

011010

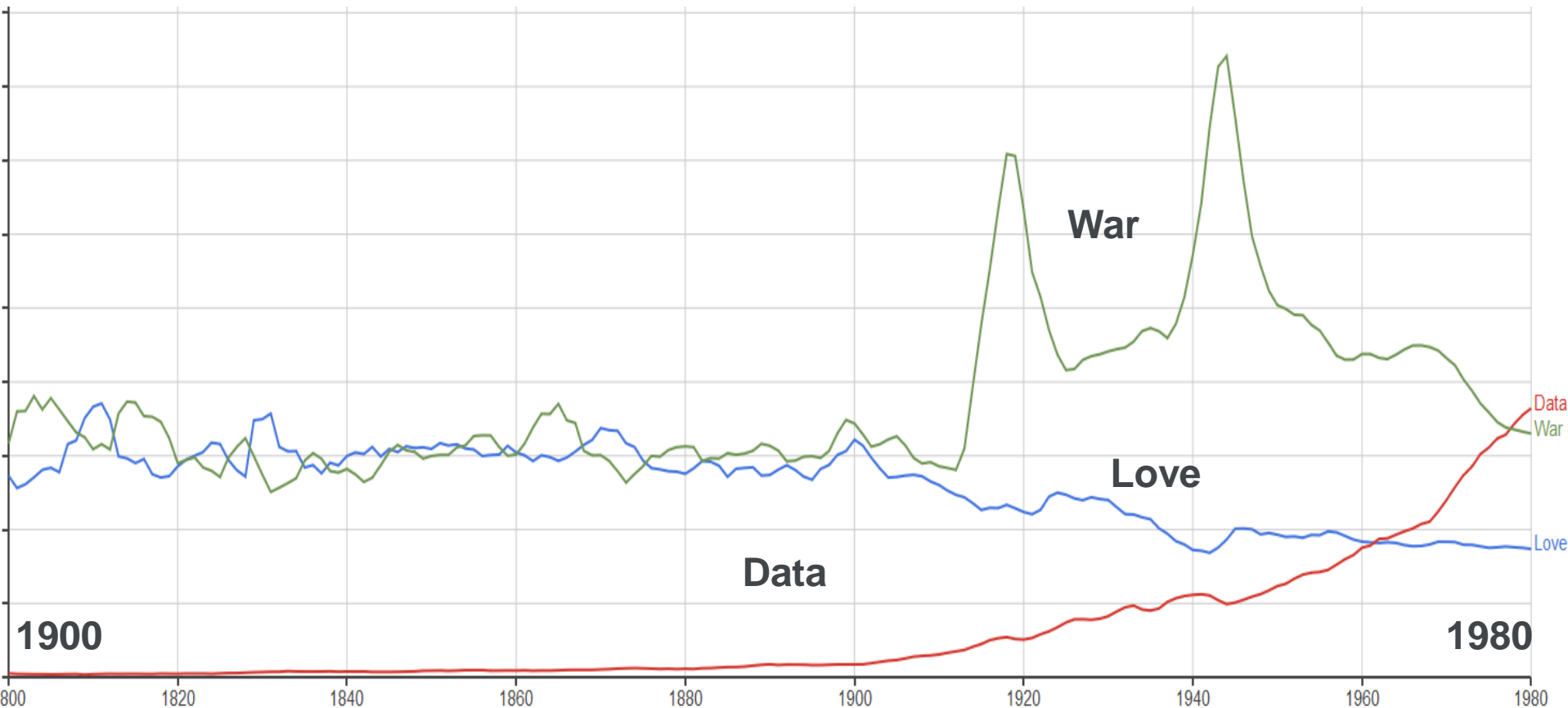
Big Data



Institute
and Faculty
of Actuaries

Big Data – Why Bother?

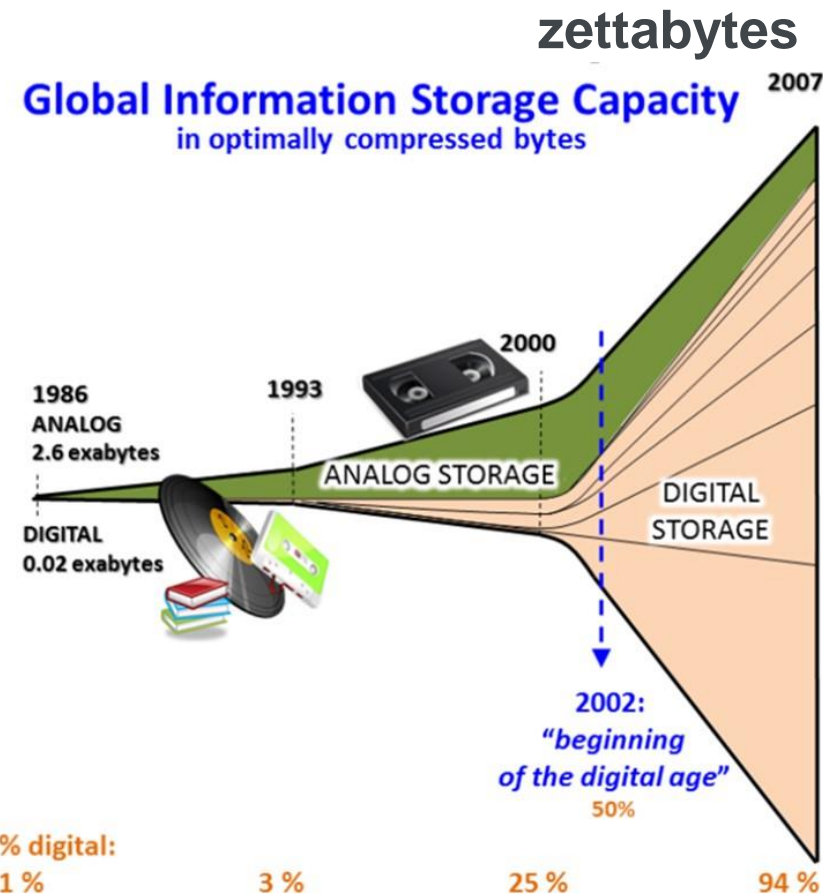
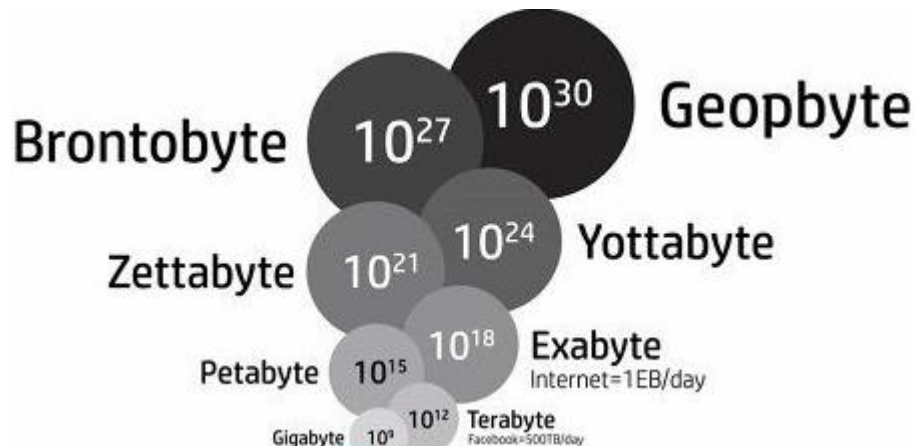
Interest in Data has been increasing steadily:



Big Data – What's New?

20 Years, 100x Data:

- The connected world now generates data at every step



Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65. <http://www.martinhilbert.net/WorldInfoCapacity.html>

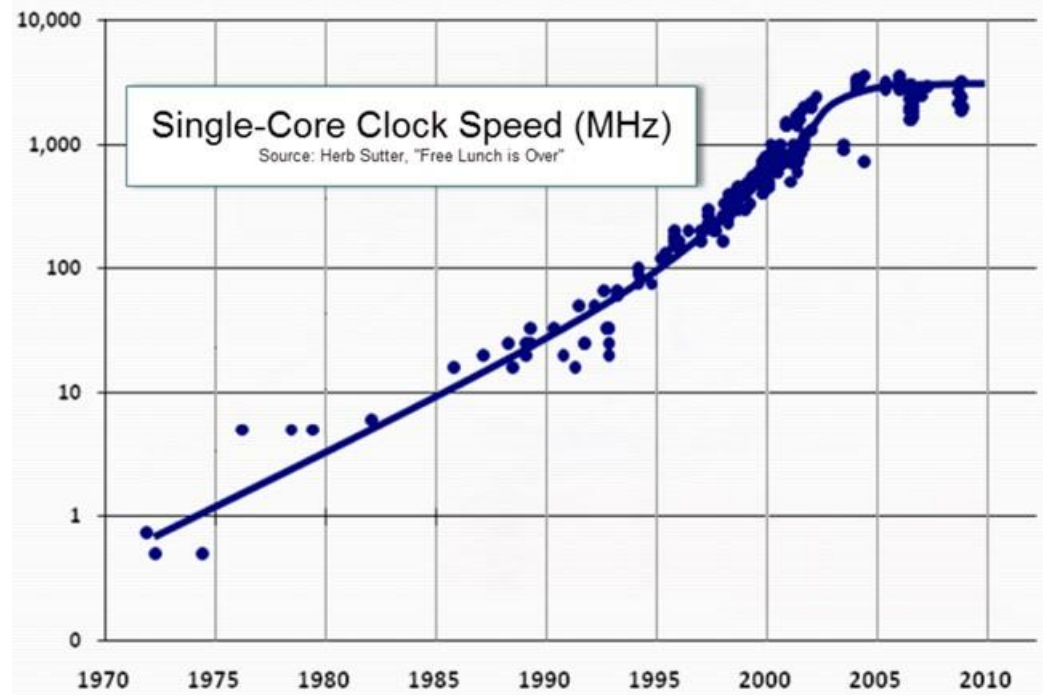


Institute
and Faculty
of Actuaries

Big Data – What's New?

Silicon Melting:

- Moore's Law for single cores effectively ended in 2006, due to heat dissipation and quantum limits. Since then, clock speeds have barely increased.
- The response is multi-core and more exotic forms of supercomputing hardware like GPUs, vector processors, coprocessors, FPGAs.



Institute
and Faculty
of Actuaries

Big Data – What's New?

Big Data is about the customer:

- Company
 - Transactions
- Customer
 - Messages
 - Browsing
 - Social Media
- Sensor
 - Wearables
 - Connected Car
 - Environment



Institute
and Faculty
of Actuaries

Big Data – Why Bother?

Focus on a business outcome:

- Better Faster Decisions
- Customer Centric
- Operational Efficiency
- Cost Reduction



BIG DATA
IS THE NEXT WAVE OF
PERFORMANCE IMPROVEMENT
**AND BUSINESS
TRANSFORMATION**



Institute
and Faculty
of Actuaries

Big Data – How do I do it?

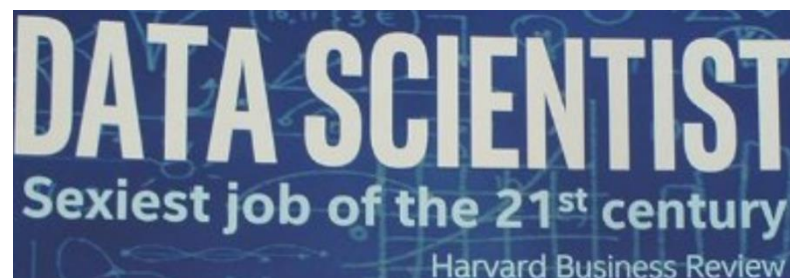
The challenge is to understand it:



Big Data – How do I do it?

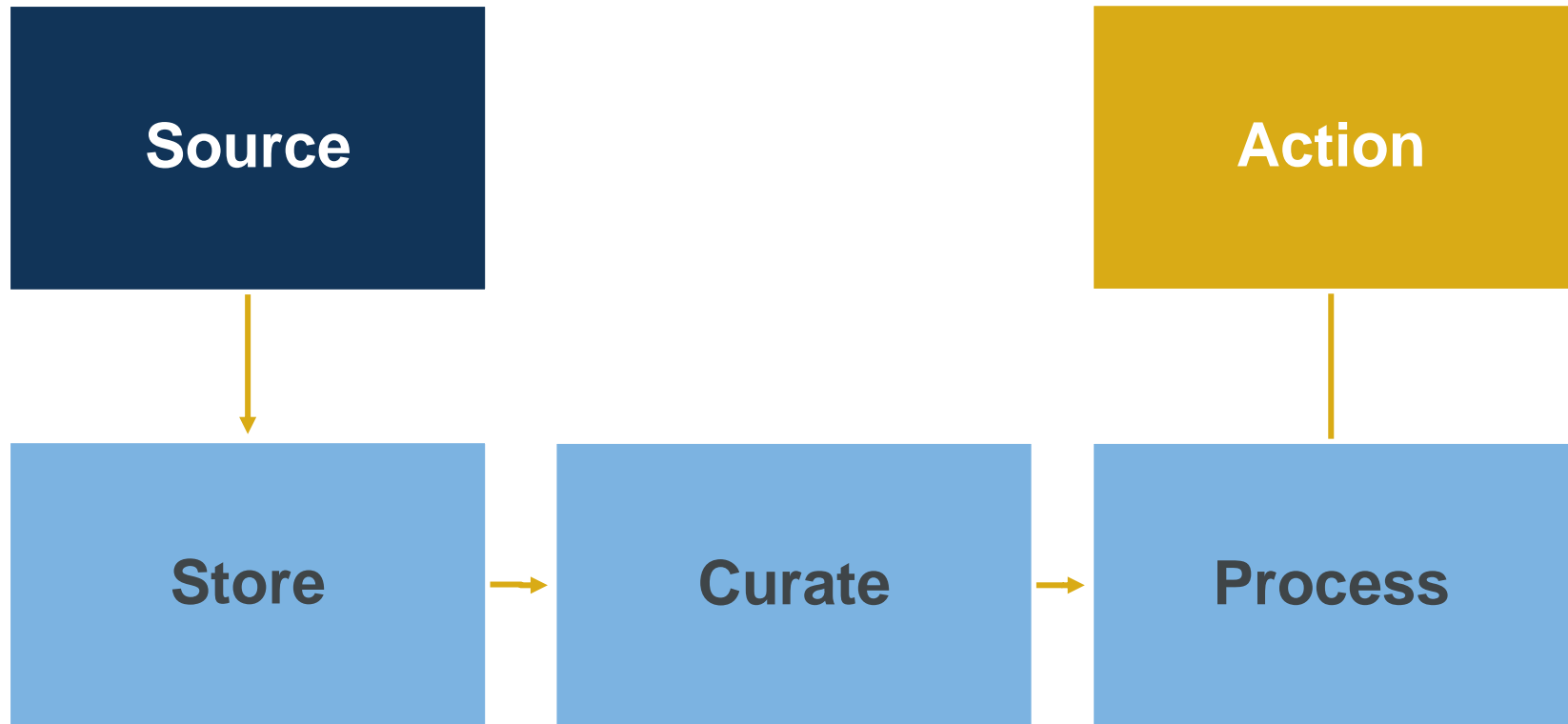
Data, Systems and People:

- Big Data's power does not erase the need for vision or human insight



Institute
and Faculty
of Actuaries

Big Data – How do I do it?



Keeping a focus on the outcome:

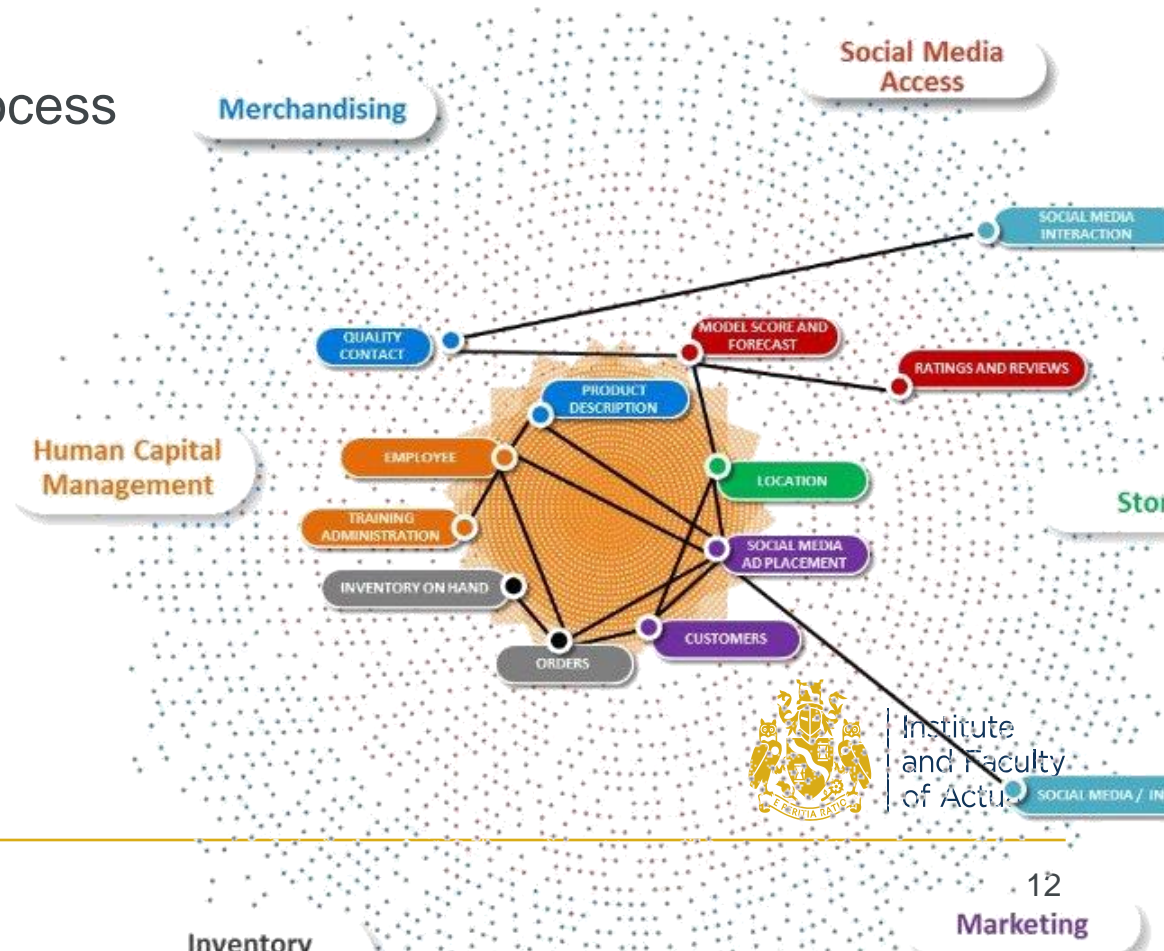


Institute
and Faculty
of Actuaries

Big Data – How do I do it?

Tightly couple a mix of :

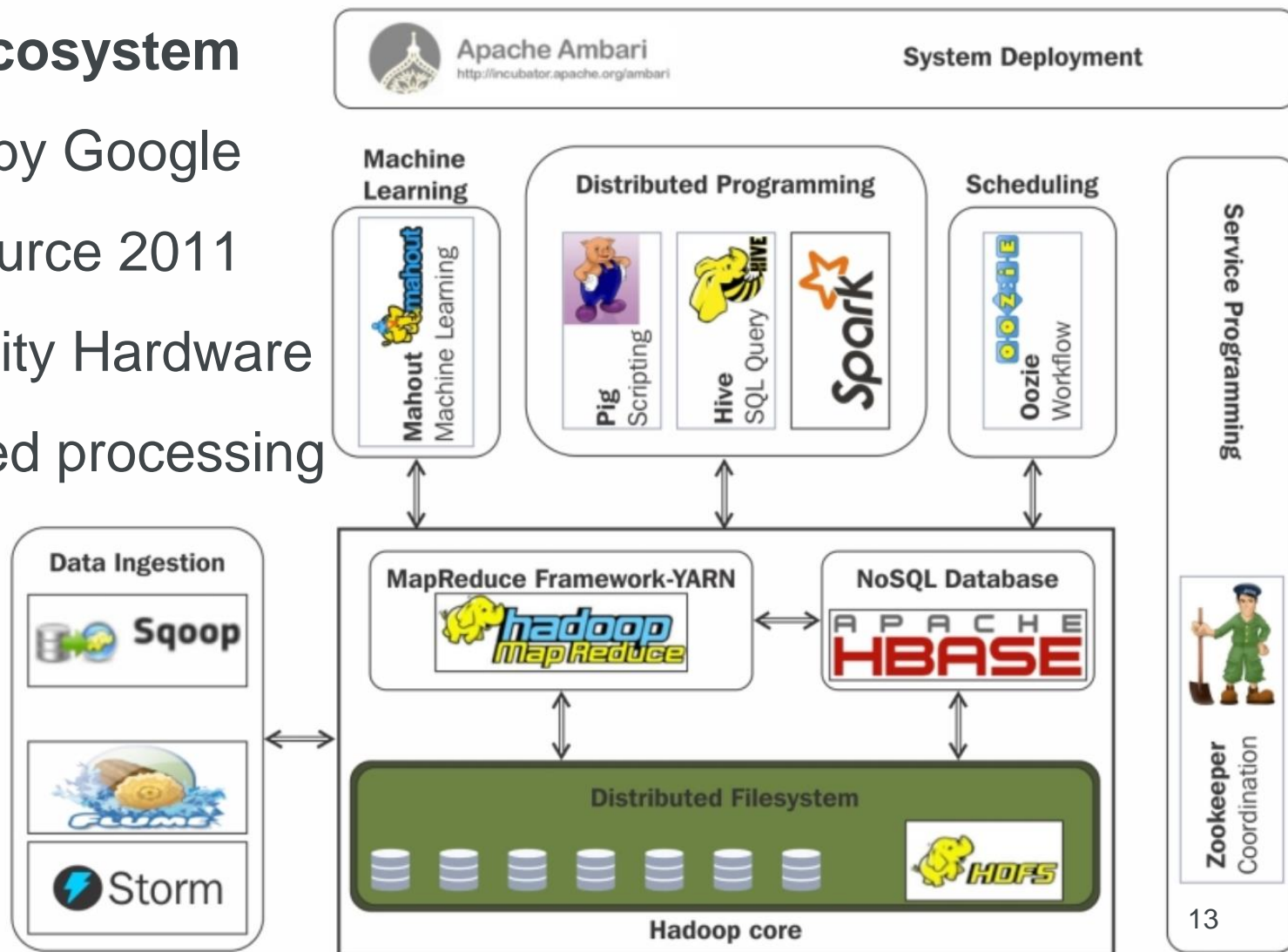
- Production data
 - Tightly coupled to process
- Research data
 - Curated for mining
- Peripheral data
 - Stored for later use



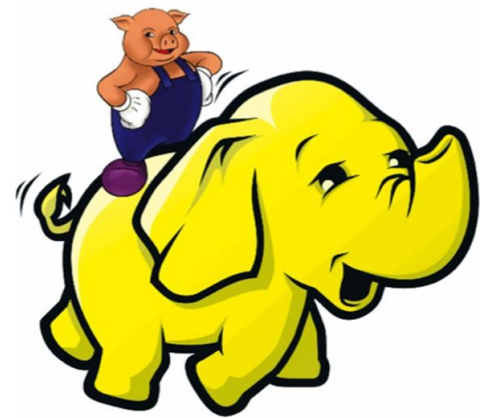
Big Data – How to do it?

Hadoop Ecosystem

- Inspired by Google
- Open Source 2011
- Commodity Hardware
- Distributed processing



Big Data – How do I do it?



Time Rich – Cash Poor:

- Functional, but focussed on the Linux command shell



```
hive> CREATE TABLE store_master (store_cd S
>                                city STRIM
>                                ROW FORMAT DELIMITED FIELDS TERMIN
>                                TBLPROPERTIES("skip.header.line.co
OK
Time taken: 1.558 seconds
hive> LOAD DATA INPATH '/user/vijaya/store_
Loading data to table default.store_master
Table default.store_master stats: [numFiles
OK
Time taken: 1.317 seconds
hive> █
```

```
[root@sandbox ~]# pig
WARNING: Use "yarn jar" to launch YARN applications.
16/01/27 21:19:14 INFO pig.ExecTypeProvider: Trying ExecType
16/01/27 21:19:14 INFO pig.ExecTypeProvider: Trying ExecType
16/01/27 21:19:14 INFO pig.ExecTypeProvider: Picked MAPREDUC
2016-01-27 21:19:14,291 [main] INFO org.apache.pig.Main - A
15.0.2.3.2.0-2950 (rexported) compiled Sep 30 2015, 19:39:20
2016-01-27 21:19:14,291 [main] INFO org.apache.pig.Main - L
s to: /root/pig_1453929554288.log
2016-01-27 21:19:14,323 [main] INFO org.apache.pig.impl.uti
otup file /root/.pigbootup not found
2016-01-27 21:19:14,970 [main] INFO org.apache.pig.backend.
ne.HEExecutionEngine - Connecting to hadoop file system at: h
works.com:8020
grunt> █
```

Big Data – How do I do it?

SQL coding will help:

- Apache Impala is the open source SQL query engine for Hadoop
- An Apache incubator project
- Supported by Cloudera, MapR, Oracle, Amazon
- Development planned for 2016 ...
 - 20x gains via multi-core joins



Institute
and Faculty
of Actuaries

Big Data – How do I do it?

Apache Spark:

- An in-memory accelerator for running programs 100x faster than Hadoop MapReduce
- Originally from Berkeley's AMPLab
- Released 2014, 2.0 July 2016

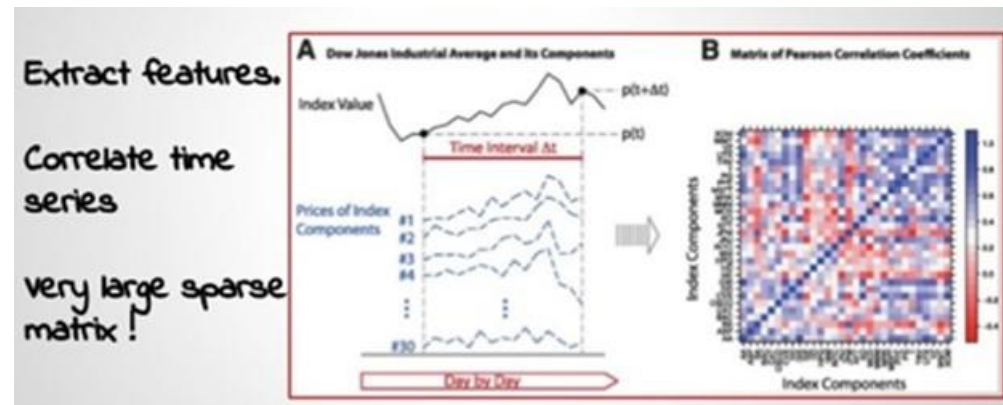
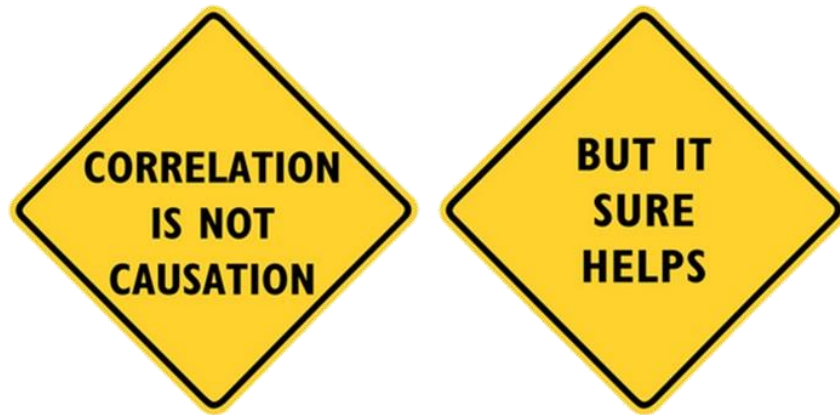


Institute
and Faculty
of Actuaries

Big Data – How do I do it?

MapReduce:

- Map – The data stream to a feature, and process in parallel across the dataset
- Reduce – Aggregate the results between nodes



Institute
and Faculty of Actuaries

Cloud Computing

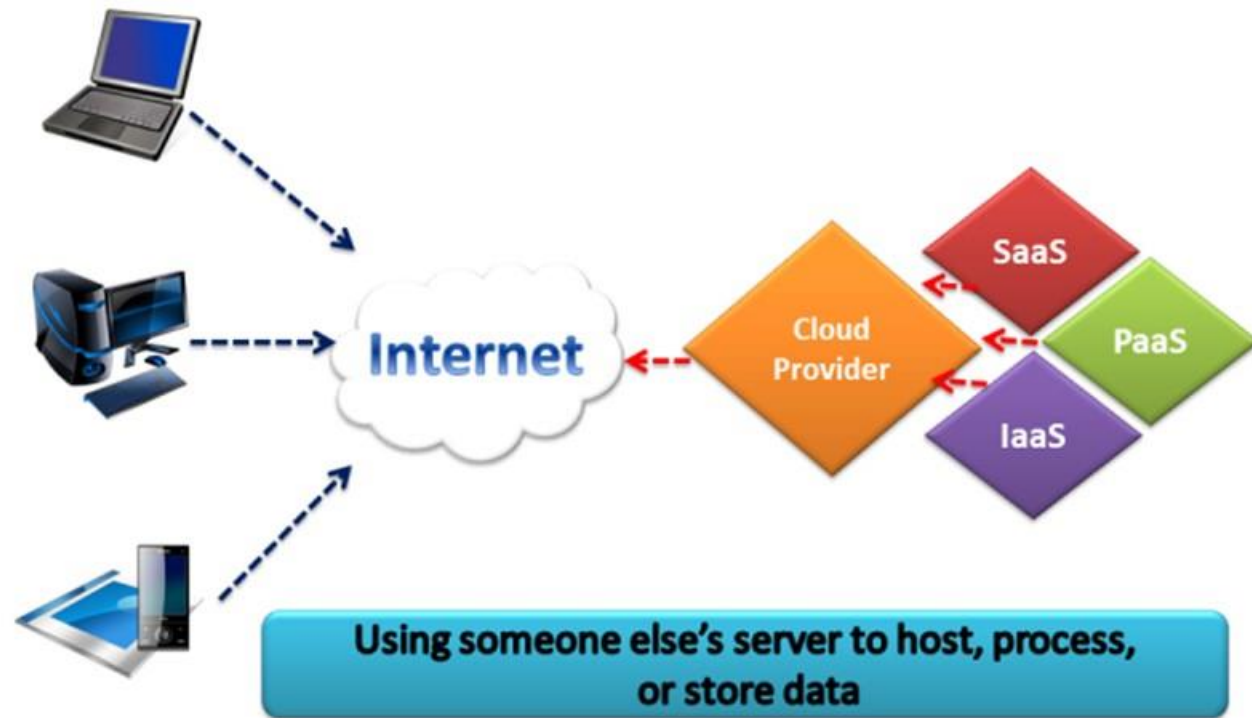
- Cloud as a Service
- Can I really do that?
- Everything must be in the Cloud
- ... because it's cheaper



Cloud Computing – as a service

- Infrastructure as a service
- Platform as a service
- Software as a service

What is the Cloud?



Cloud Computing – Can I really do that?

- Sensitive Data
- Privacy concerns
- ~~• Safe Harbor~~
- DPA



Cloud Computing – Everything must be in the Cloud

... said the IT team:

- No capital spend
- Flex usage down as well as up
- Service included



Institute
and Faculty
of Actuaries

Cloud Computing ... bec

Deals Laptops Desktops Printers Tablets Tablet Accessories Monitors Computer Accessories

ard drive"



Toshiba HDWC240EK3J1 4TB Stor.E Can

by Toshiba



500 customer reviews | 101 answered

RRP: £126.42

Price: **£89.98** & **FREE Delivery** in the UK. [Details](#)

You Save: **£36.44 (29%)**

Usually dispatched within 1 to 3 weeks.

This item can be delivered to UK - Mainland. [Details](#)

Dispatched from and sold by Amazon. Gift-wrap available.

Note: This item is eligible for **click and collect**. [Details](#)

29 new from **£89.98**

Colour Name: **Black**



me: 4 TB

3 TB

4 TB

5 TB

6 TB

able peace of mind
hard drive protection
capacity storage
ish design
er transfer speeds
[more product details](#)

Item	Cloud Quote	On-Site Claim	On-Site Actual
Processing	£2400	£70000	£1540
Disk	£9600	£76000	£90
Network	£0	£5000	£50
Total	£12000	£151000	£1650

Cloud Computing – Everything must be in the Cloud

The Skinny Drinking Straw problem:

- Notice that connectivity was free!
How can that be the case?
- If I could afford to keep data
in the cloud,
how could I get it there fast enough?

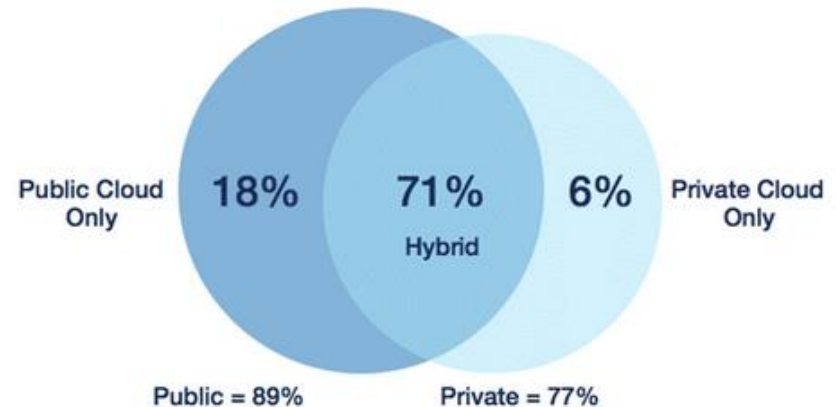


Cloud Computing – Everything must be in the Cloud

Only 1/5 practitioners agree:

- The rest have sensitive data private and secure

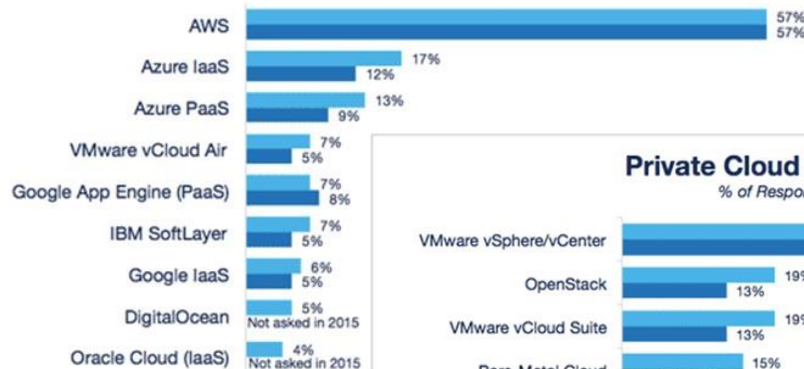
95% of Respondents Are Using Cloud



Source: RightScale 2016 State of the Cloud Report

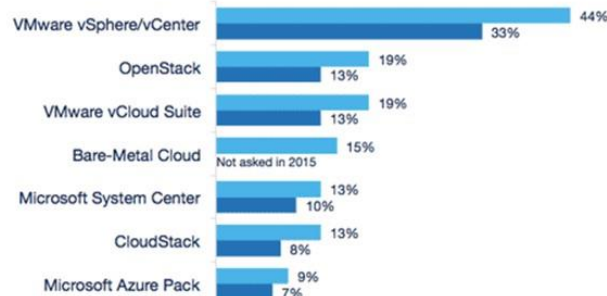
Public Cloud Adoption 2016 vs. 2015

% of Respondents Running Applications



Private Cloud Adoption 2016 vs. 2015

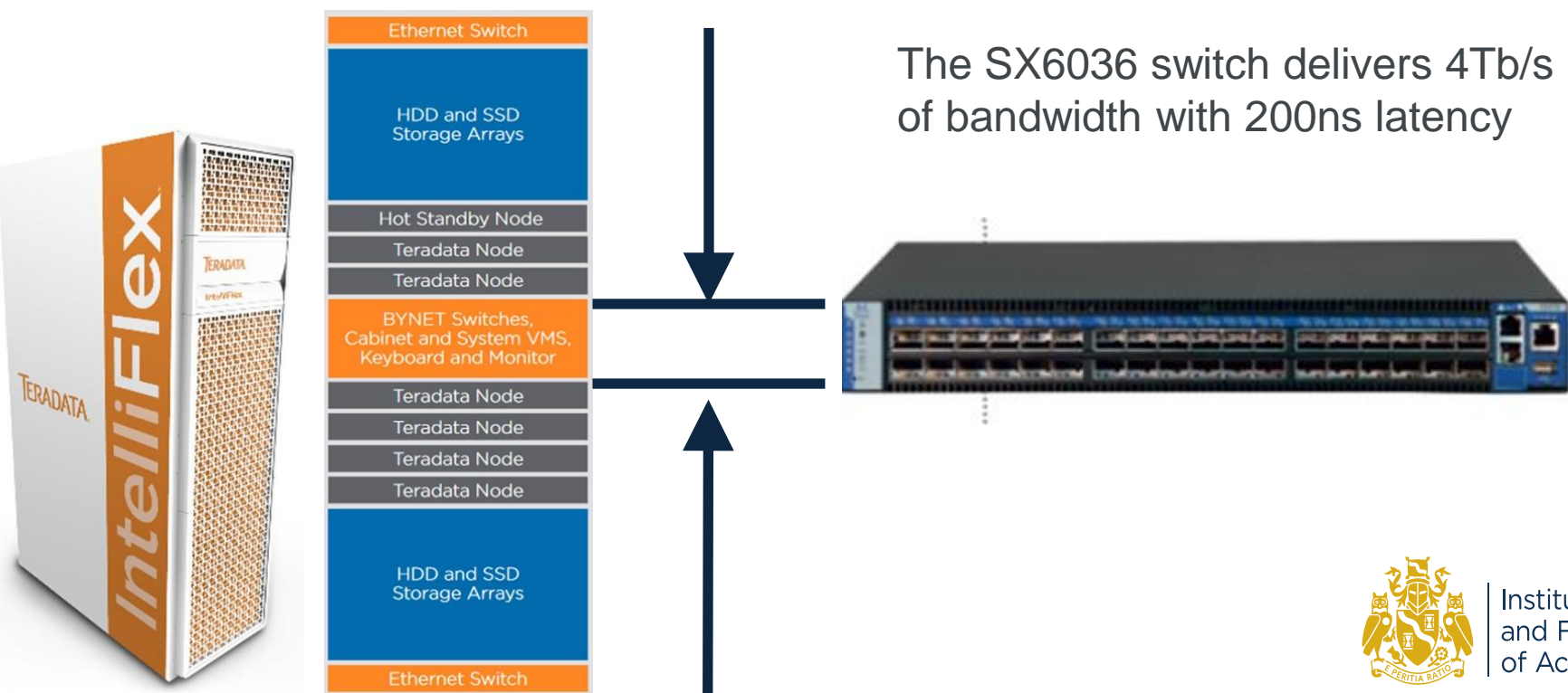
% of Respondents Running Applications



Institute
and Faculty
of Actuaries

Cloud Computing – Everything must be in the Cloud

Network Bandwidth is the key:

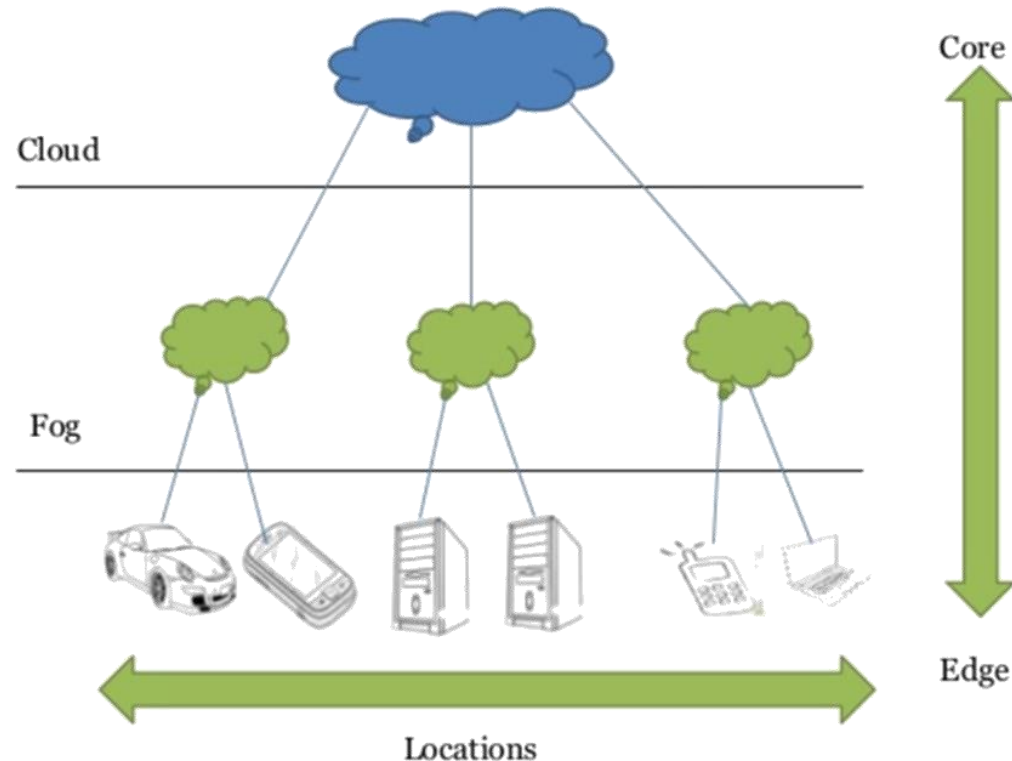


Institute
and Faculty
of Actuaries

Cloud Computing – Everything must be in the Cloud

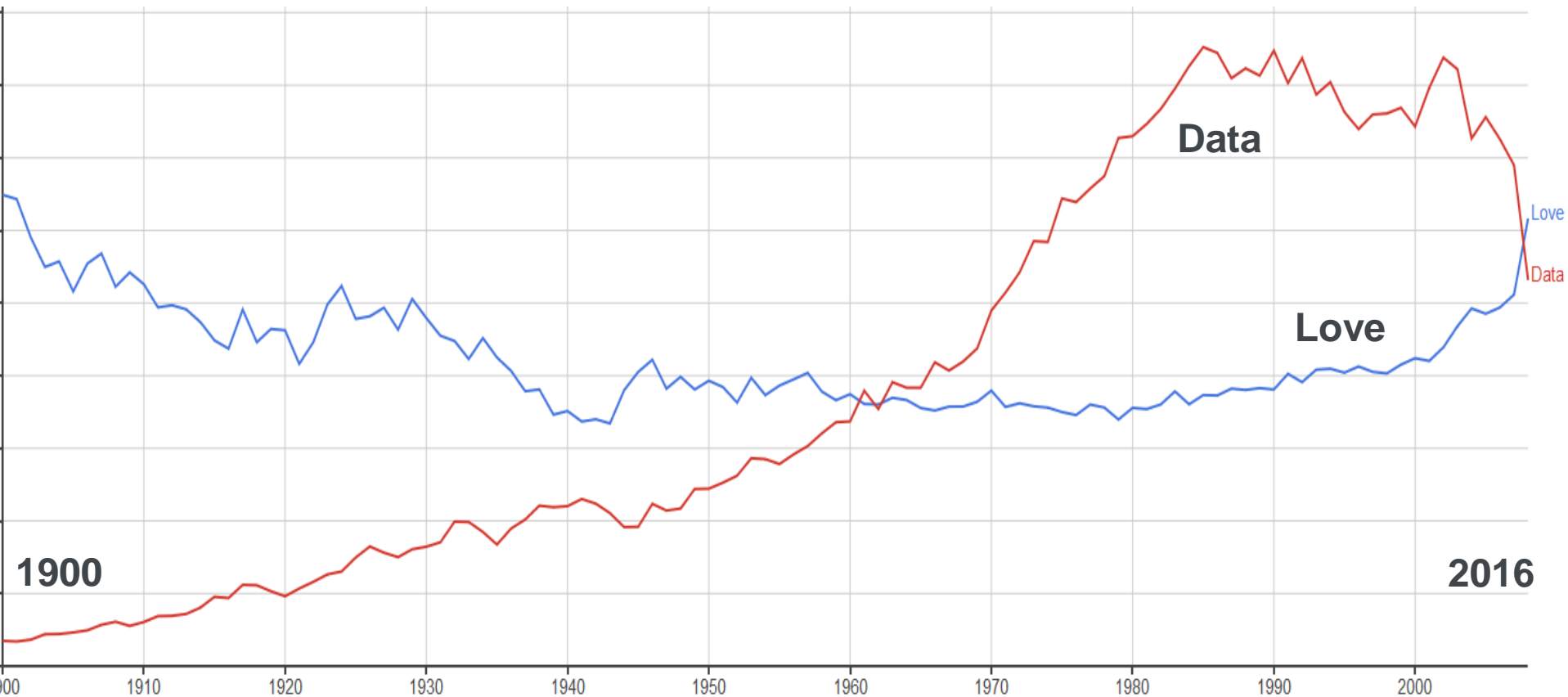
Fog Computing:

- Forget taking big data to the cloud and process in the Fog



Big Data – Why Bother?

And the good news is we have been cured:



Questions

Comments

The views expressed in this [publication/presentation] are those of invited contributors and not necessarily those of the IFoA. The IFoA do not endorse any of the views stated, nor any claims or representations made in this [publication/presentation] and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this [publication/presentation].

The information and expressions of opinion contained in this publication are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this [publication/presentation] be reproduced without the written permission of the IFoA [*or authors, in the case of non-IFoA research*].



Institute
and Faculty
of Actuaries



Institute
and Faculty
of Actuaries

Opening the Box on Big Data and Probing the Fog of The Cloud

Tony Lovick – RPC Consulting

Tony.Lovick@rpc.co.uk

22 September 2016

Articulate
Sponsorship
Thought leadership
Progress
Community
Sessional Meetings
Education
Working parties
Volunteering
Research
Shaping the future
Networking
Professional support
Enterprise and risk
Learned society
Opportunity
International profile
Journals
Support