

Uplifting Big Data

Tony Shan

✉ mail@tonyshan.com

🐦 @TonyShan #UpliftBigData

October 20, 2016





促进软件开发领域知识与创新的传播



关注InfoQ官方信息
及时获取QCon软件开发者
大会演讲视频信息



[北京站] 2016年12月2日-3日

咨询热线: 010-89880682



[北京站] 2017年4月16日-18日

咨询热线: 010-64738142



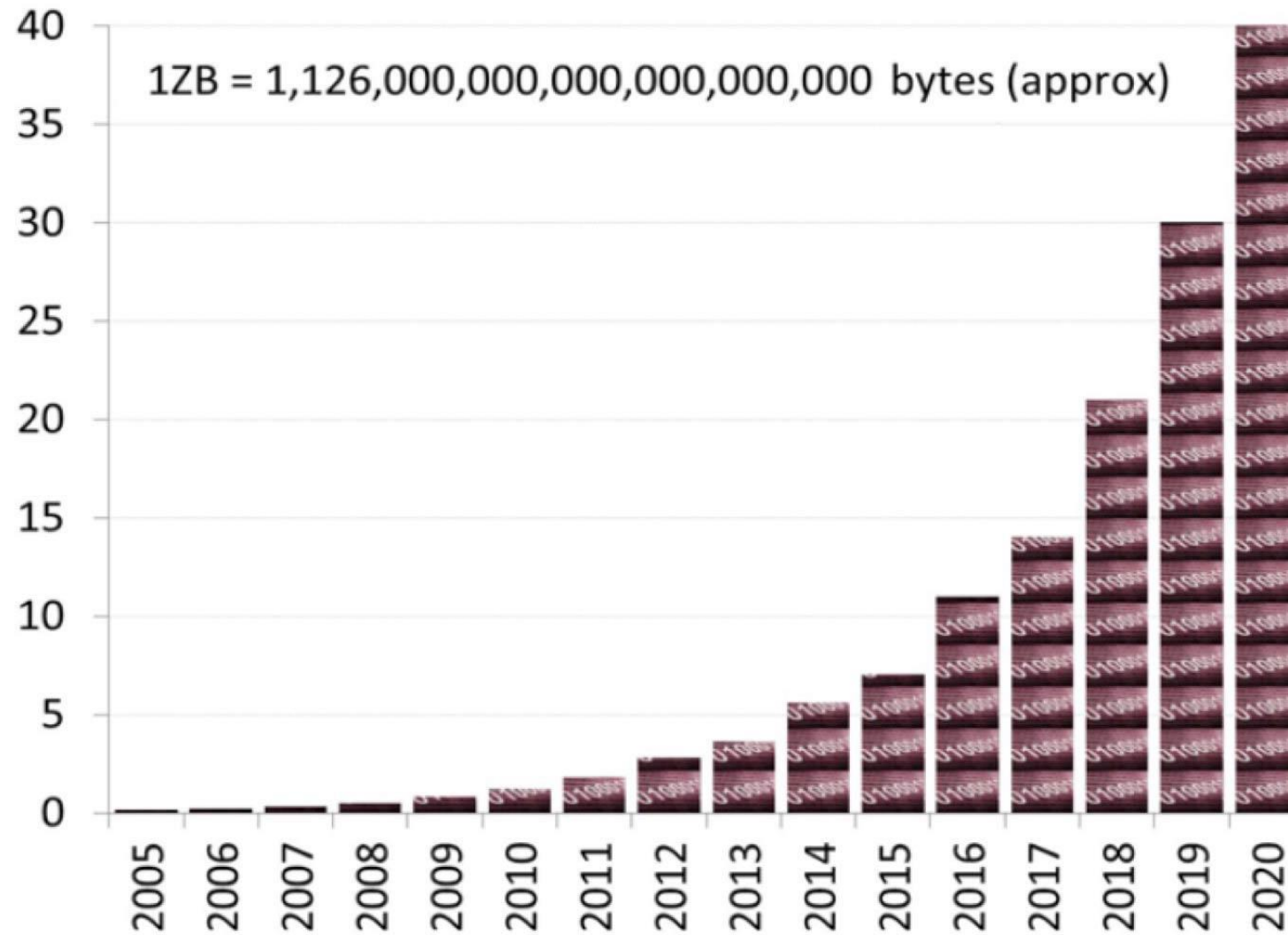
Table of Contents

- State of Big Data
- Big Data Term
- Issues
- Why Rethink
- Trending
- Anatomy
- Best Practices





All Global Data in Zettabytes



Source: <http://www1.unece.org/stat/platform/display/msis/Big+Data>



Ride the Big Data Wave

90%

of the world's existing
data has been created
in the last

2 YEARS

By 2020

35 Zettabytes worth of data will be created

*35 Zettabytes = 35,000,000,000,000,000,000 bytes

1/3 of that data will be stored in or will have
passed through the cloud



10,000

Libraries of Congress
worth of data are
created annually by
U.S. companies



80,000 km

is the distance past the moon the
current global storage capacity for
digital information stored on a CD
stack would stretch



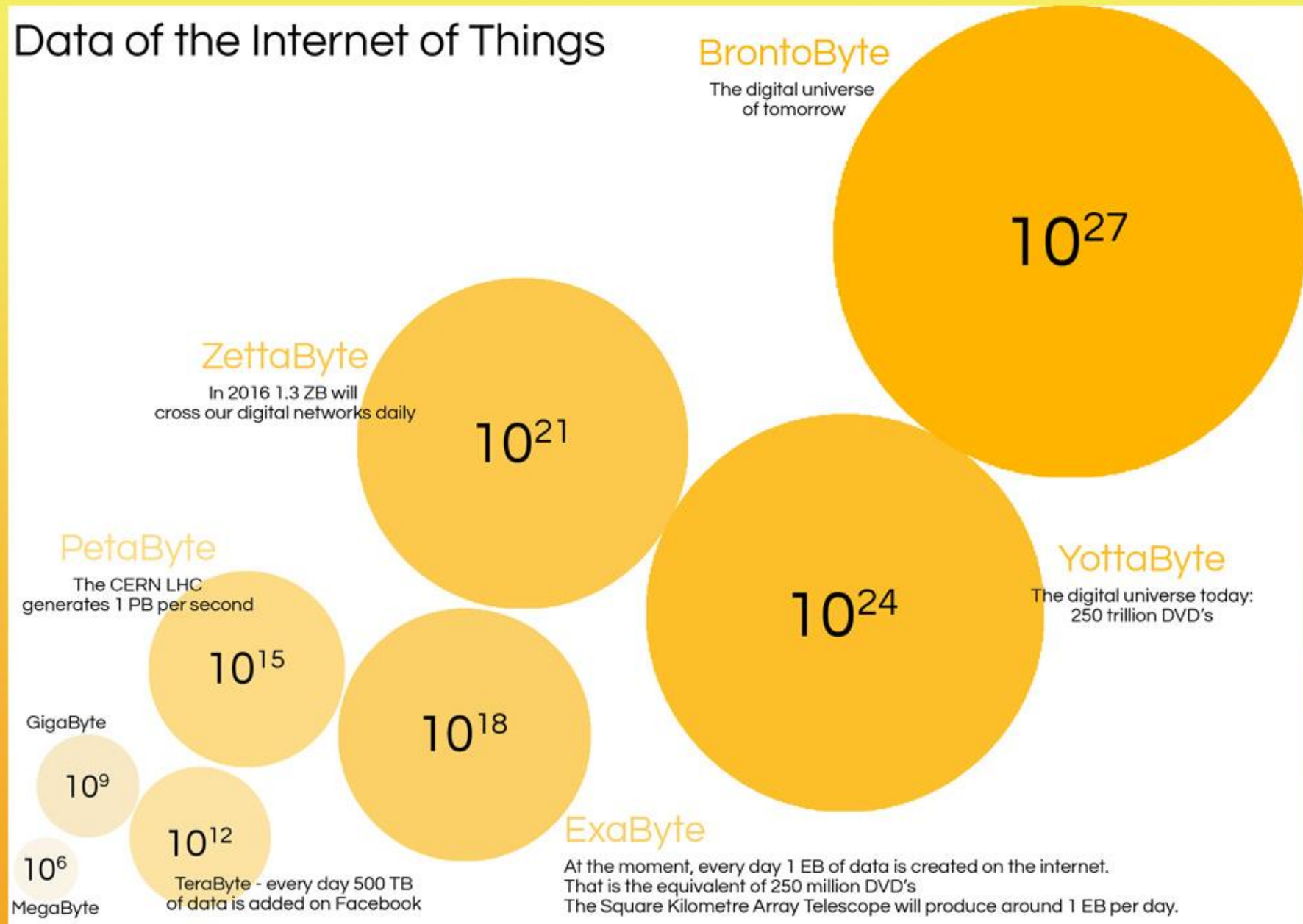
Every 2 days as much
information is created as
from the dawn of
civilization up until 2003

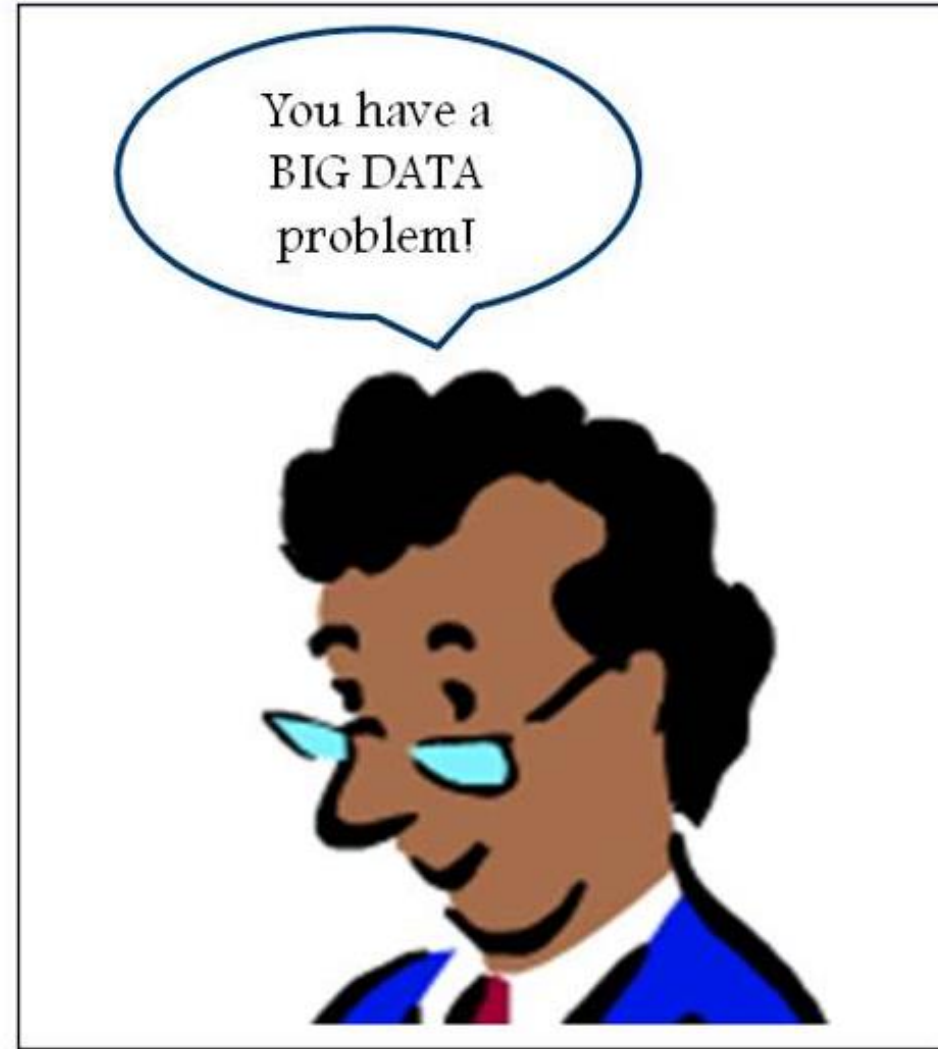


There are actually more
pieces of digital information
than stars in the universe



Data of the Internet of Things





Do You Have a Big Data Problem?



What is Big Data?



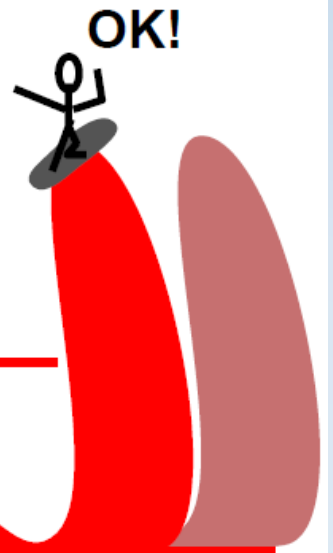
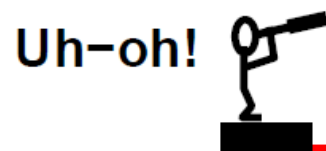
Who coined “Big Data” Term?



John Mashey was arguably the first to use the term Big Data in his public talks in late 1990s.

***Big Data ...
and the Next Wave of **InfraStress*****
John R. Mashey
Chief Scientist, SGI

***Technology Waves:
NOT technology for technology's sake
IT'S WHAT YOU DO WITH IT
But if you don't understand the trends
IT'S WHAT IT WILL DO TO YOU***



Father of “Big Data” Concept



Francis X. Diebold presented a paper, in which he states “Recently, much good science, whether physical, biological, or social, has been forced to confront—and has often benefited from—the ‘Big Data’ phenomenon. *Big Data* refers to the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage technology.”

Diebold, F.X. (2003),
“‘Big Data’ Dynamic Factor Models for Macroeconomic Measurement and Forecasting”
(Discussion of Reichlin and Watson papers), in M. Dewatripont, L.P. Hansen and S. Turnovsky (Eds.),
Advances in Economics and Econometrics, Eighth World Congress of the Econometric Society.
Cambridge: Cambridge University Press, 115-122.

“Big Data” Dynamic Factor Models for Macroeconomic Measurement and Forecasting

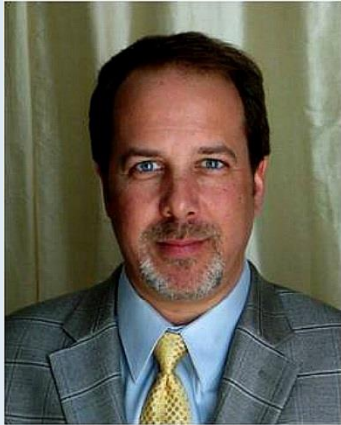
Francis X. Diebold

**University of Pennsylvania
and NBER**

First Version, July 2000
November 28, 2000

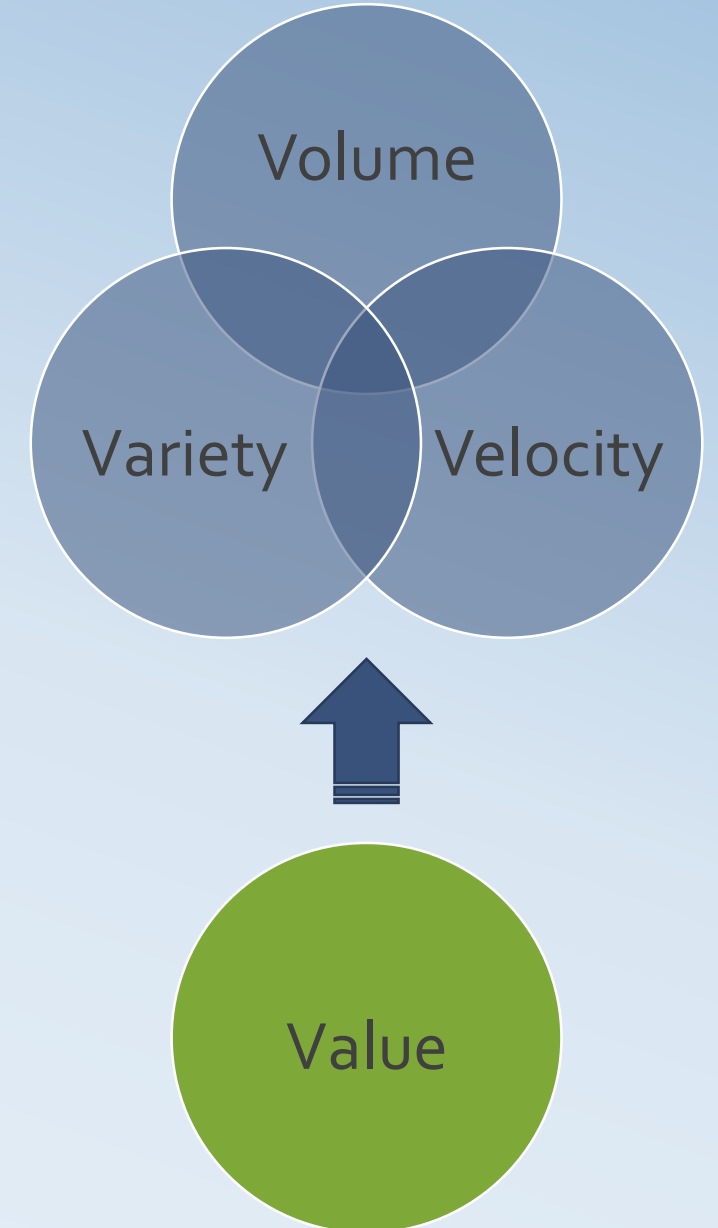


Big Data Definition

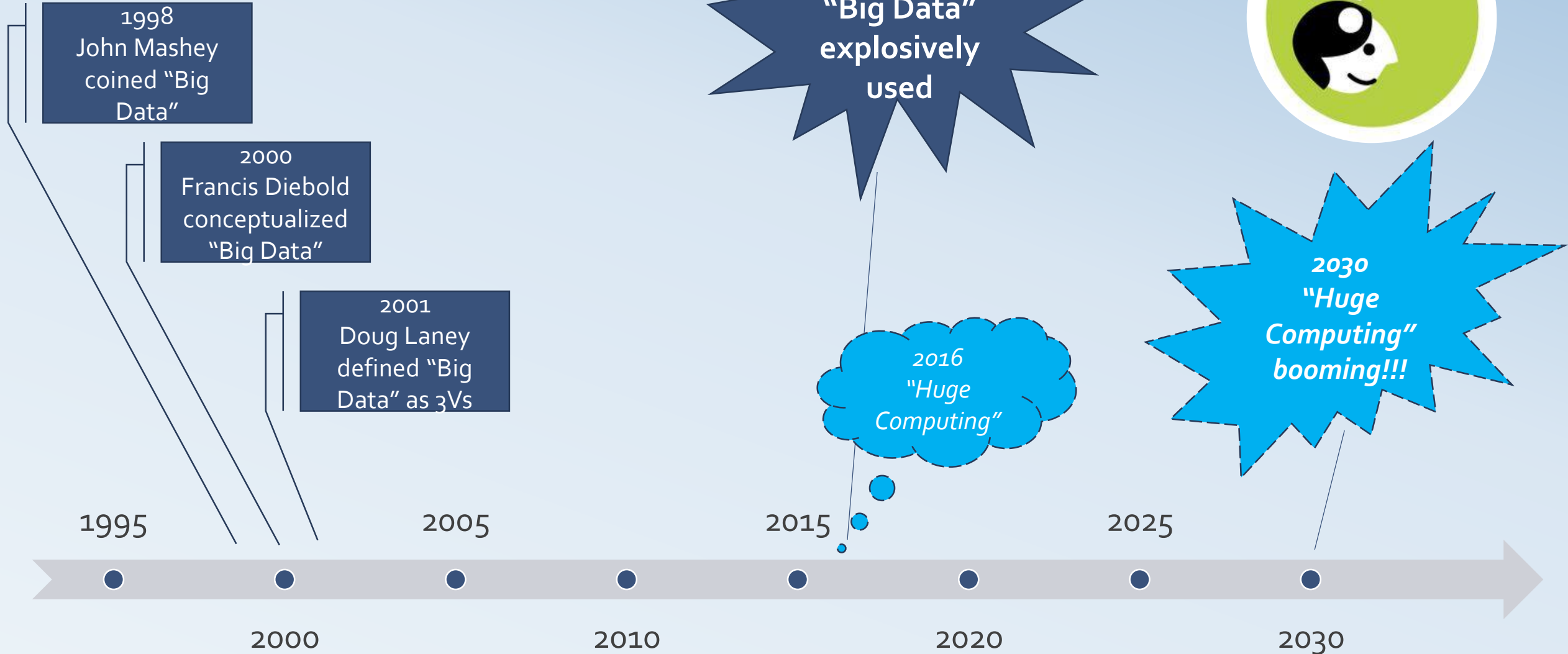


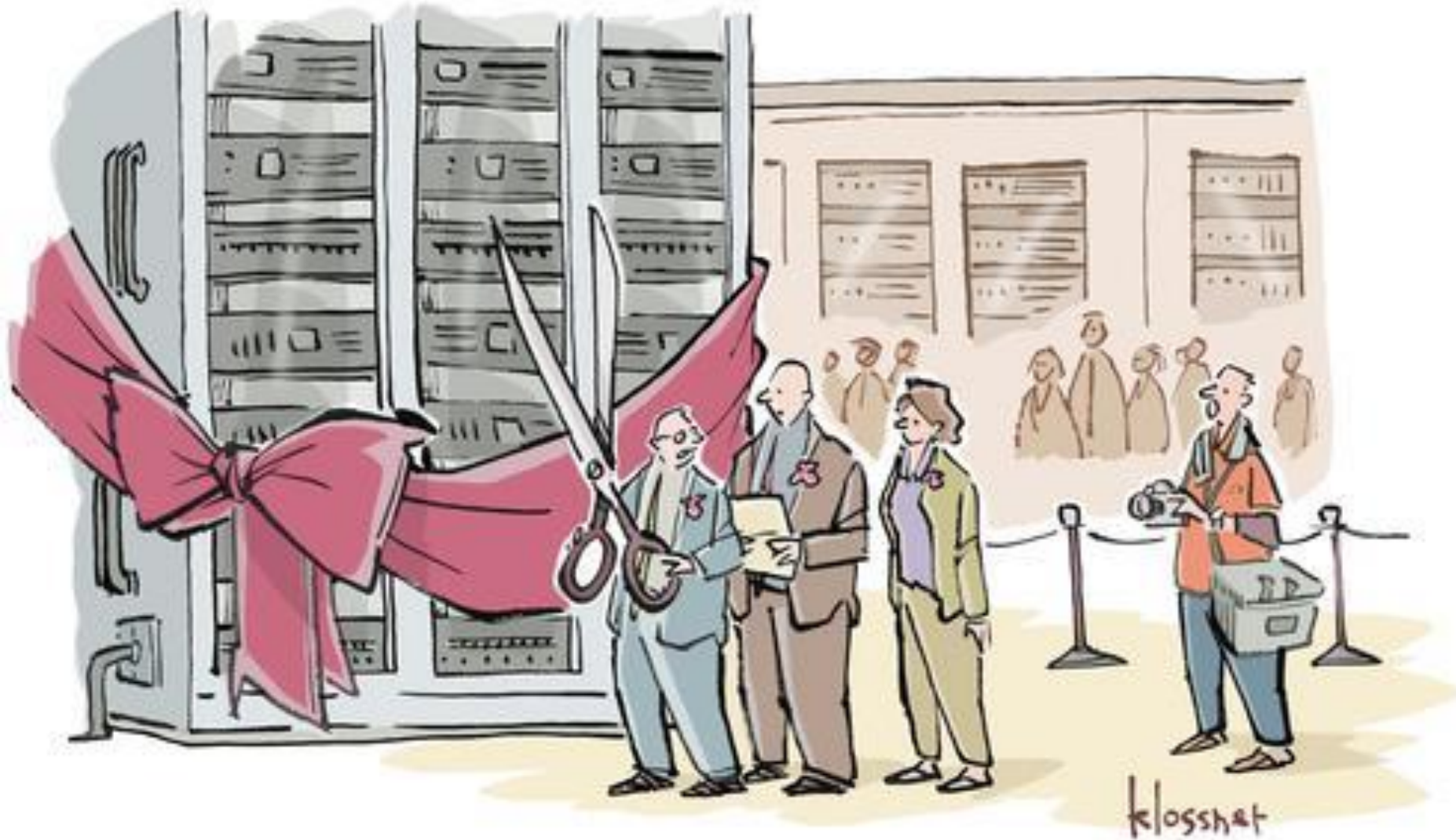
The definition of Big Data is typically traced back to the 3Vs model defined by Doug Laney in **2001**: Volume, Velocity, and Variety.

The fourth V was later added in different fashions, such as "Value" or "Veracity".



Valid?





"IS THIS A GOOD TIME TO TELL YOU I
DON'T KNOW WHAT 'BIG DATA' MEANS?"



Issues



The NIST report of “Big Data Interoperability Framework” lists the problem of multiple descriptions about Big Data in fragmented and incoherent efforts.



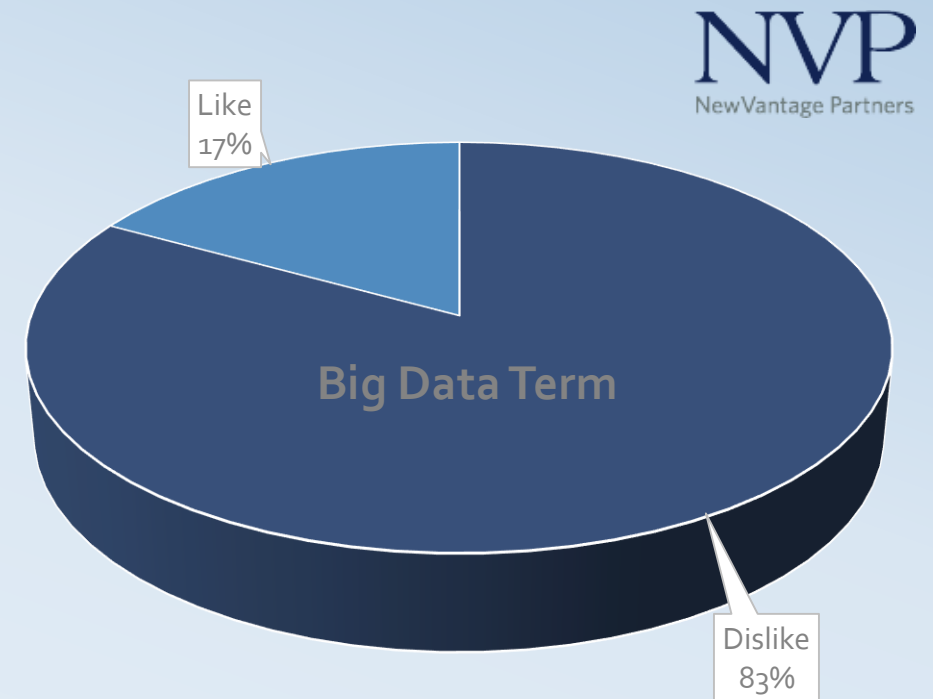
Even the original author of 3Vs admitted that he was simply writing about the burgeoning data in the data warehousing and business intelligence world.



Report by NewVantage Partner

83% of the executives surveyed thought that the term of Big Data was overstated, confusing, or misleading.

They liked the concept, but hated the phrase.



Expert's View

- There is Just One V in Big Data.
- I do think that big data can be better explained by adding a few more V's.
- Nobody likes the Big Data term and almost everybody wishes for a better, more descriptive name for it.



• Amar Nadig



• Mark Van Rijmenam



• Tom Davenport





Time to Rethink Big Data



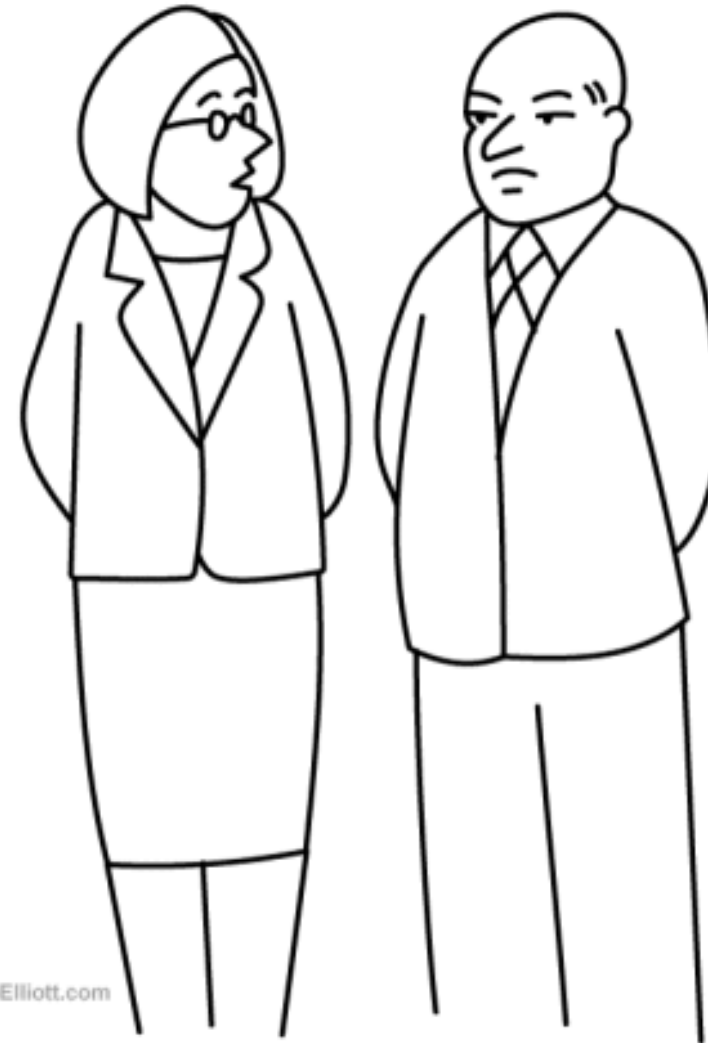
The big problem of Big Data is that the term ineffectively describes today's phenomenon and is becoming outdated for the new paradigm.

It is necessary to deal with today's data challenges that are beyond what the OLD Big Data term refers to.



Why Rethink?

“So far, our Big Data investments have just made your stupidity more scalable...”



WE'VE DECIDED
TO TAKE BIG
DATA TO THE
NEXT LEVEL...

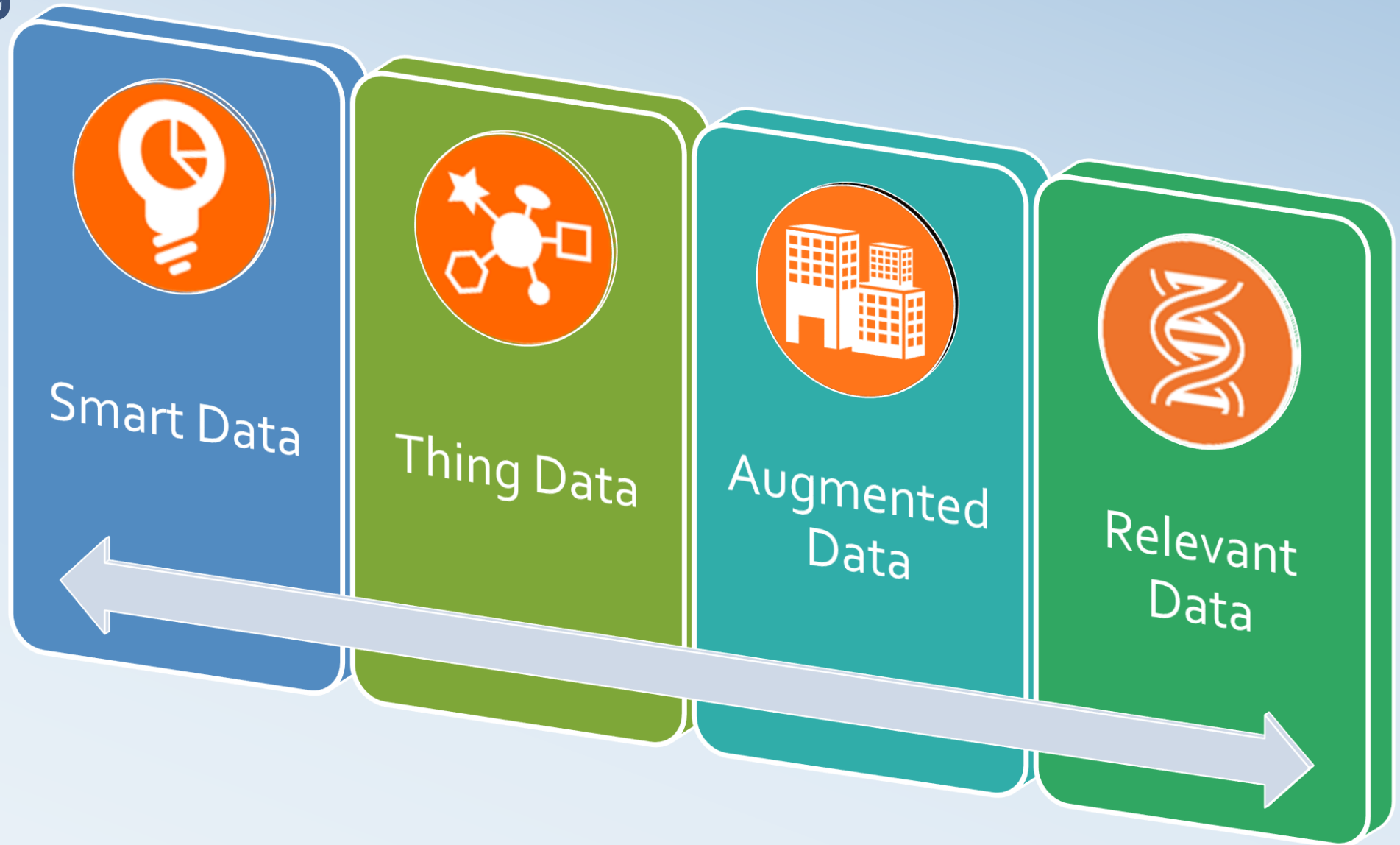
**HUMONGOUS
DATA**



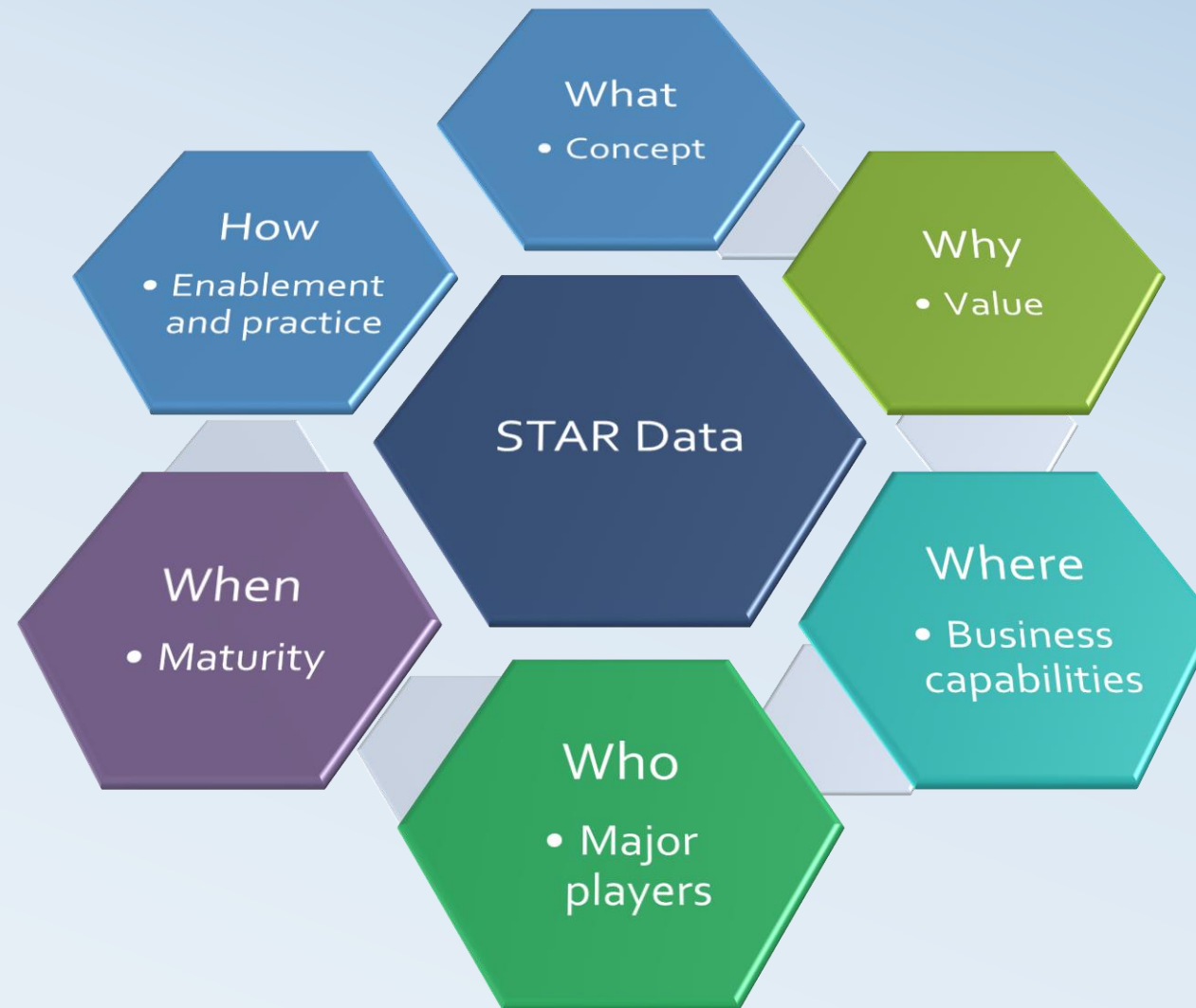
© D.Fletcher for CloudTweaks.com







Trending



Anatomy

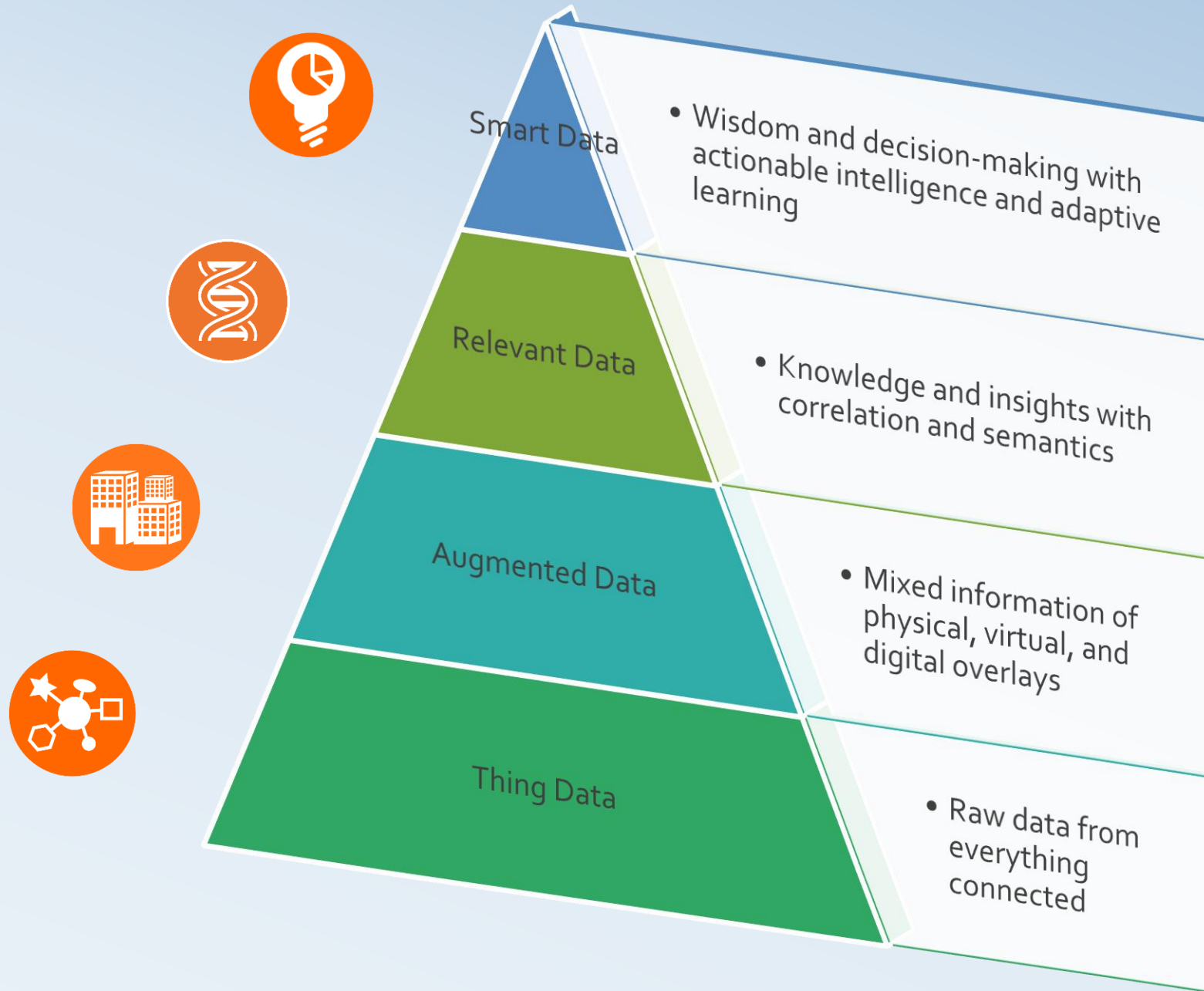


STAR Data

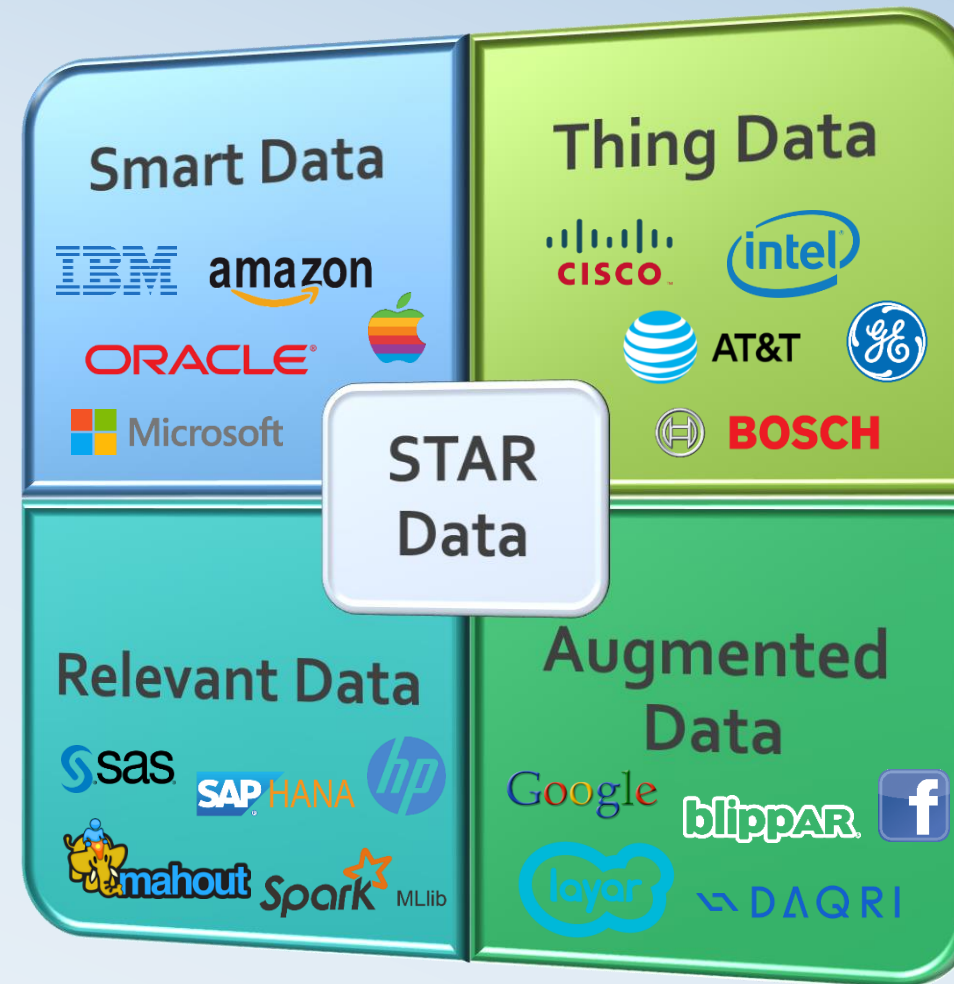
	Description	Application Areas
Smart Data 	Filter out the noise and hold the valuable data. The cognitive computing makes solutions intelligent and self-improving. Knowledgeable reasoning results in more sound decisions and better customer engagements.	Intelligent content personalization leverages all the data that are accumulated from B2C, B2B and C2C channels, to not only optimize the display of customized contents, but also heighten the user experience.
Thing Data 	Data from connected things like devices, mostly machine generated data. IoT Data is raw in nature, small in size, frequent in transmission, and large in volume.	A physical object like a sensor monitors a specific condition such as location, vibration, motion and temperature to collect data. IoT-enabled objects share information about their condition and the surrounding environment with others.
Augmented Data 	Audio, visual and interaction data for Augmented Reality (AR) and Virtual Reality (VR) in a physical 3D or cyber-physical converged environment. The vision, NLP, gesture and emotion recognition enable the surrounding real world become more interactive and digitally manipulatable.	AR enhances product previews by allowing a customer to view what's inside a product's packaging without opening it. It is used to let a doctor look inside a patient by combining one source of images such as an X-ray/ultrasound with another such as video of an endoscope to inspect a tumor.
Relevant Data 	Data relationship is critical to identify pertinence in the data set, which leads to profounder understanding of seemingly unrelated events and sequence. The focus is to contextualize each bit of data with its own specific semantics in predictive and preventive analytics.	Social analytics, as an example, empower businesses to distill the hidden values behind the reams of social data and activities, in order to glean actionable insights about the user's preference, behavior, habits and influence.



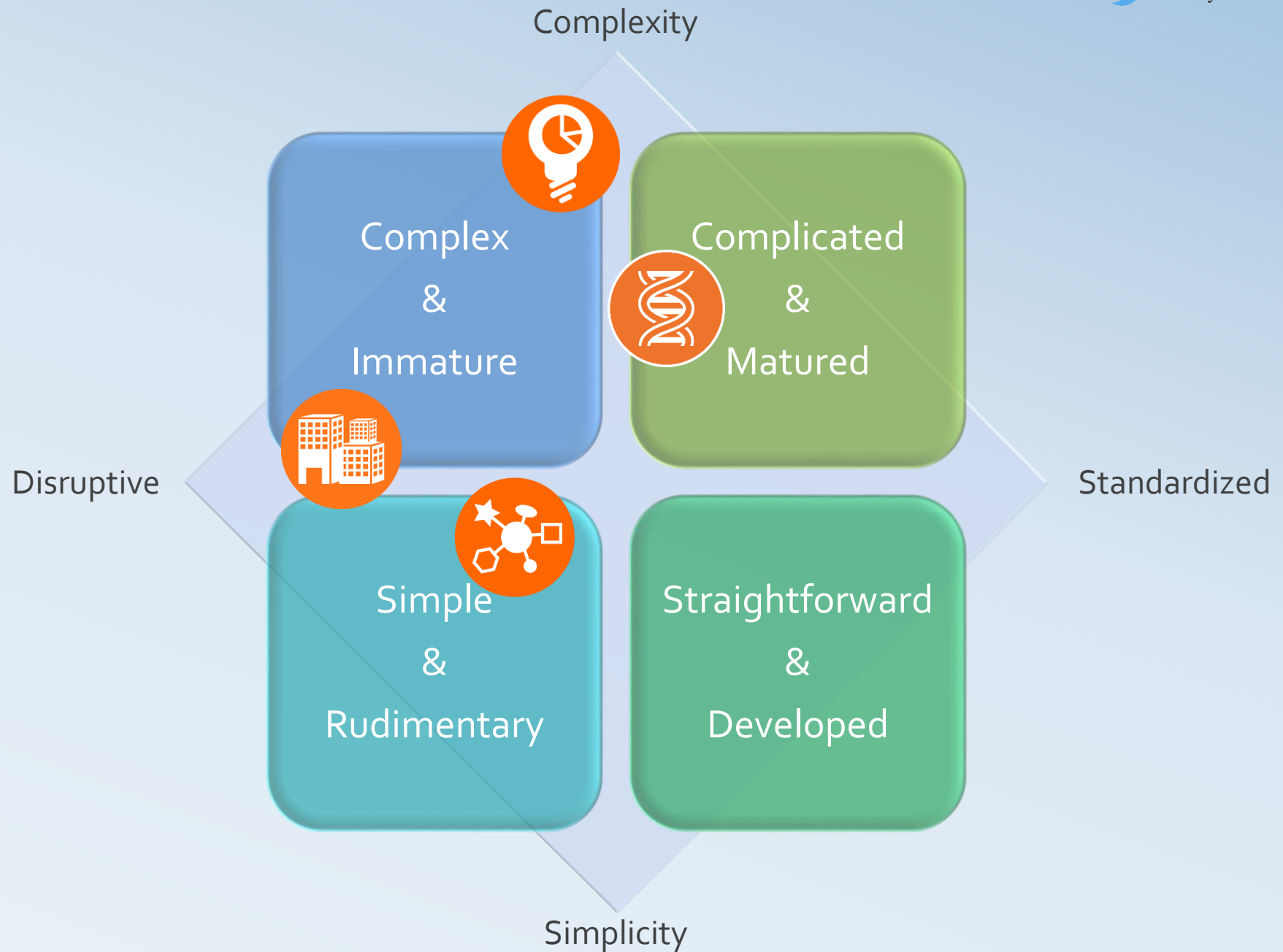
Why



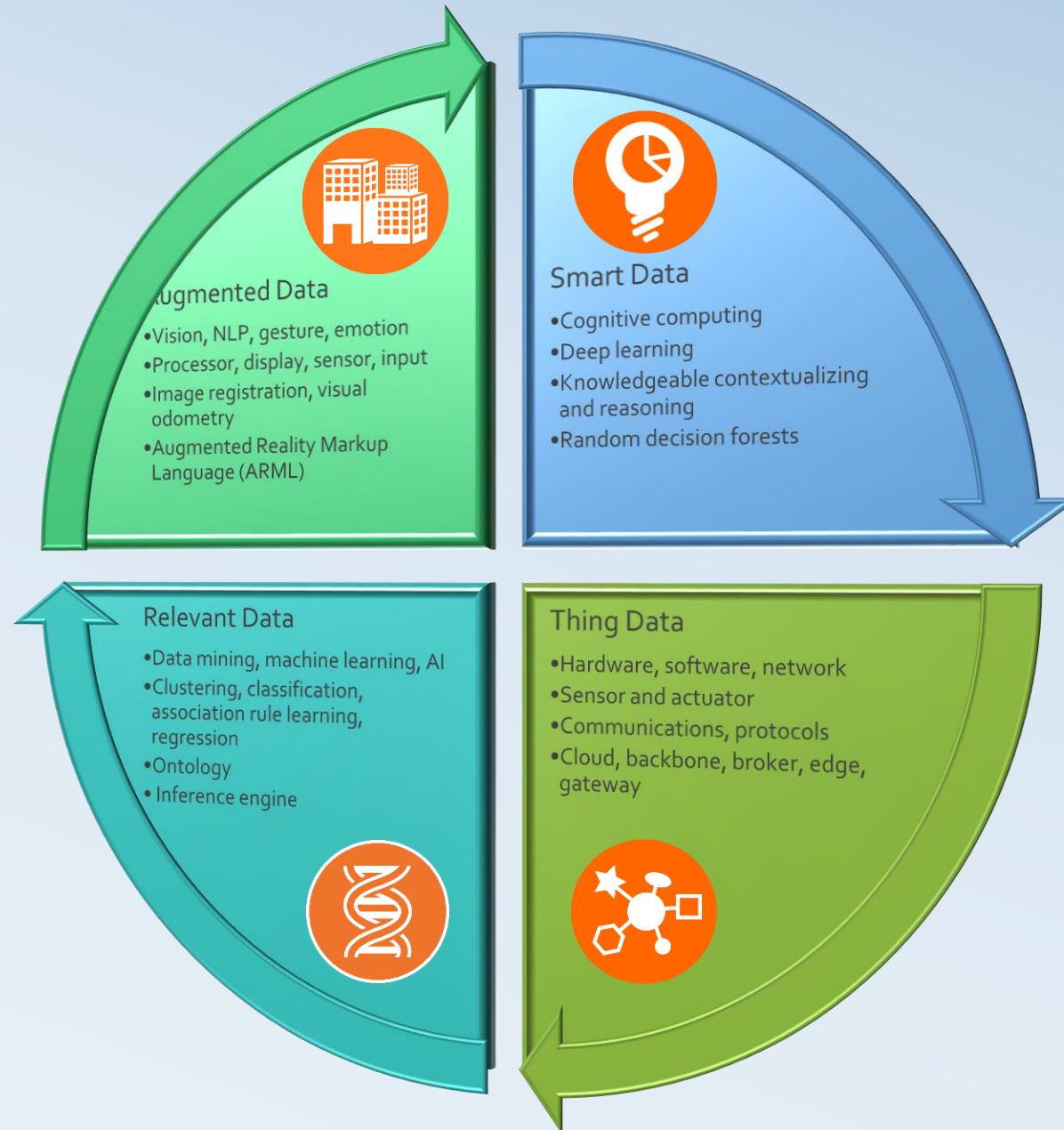
Who



When



How



IoT Protocols

Framework

- Alljoyn
- IoTivity
- Weave
- Homekit

Semantic

- JSON-LD
- Web Thing Model

Device Management

- TR-069
- OMA-DM

Data

- MQTT
- CoAP
- AMQP
- Websocket
- Node

Discovery

- Physical Web
- mDNS
- DNS-SD

Transport

- Wifi
- Bluetooth
- LPWAN

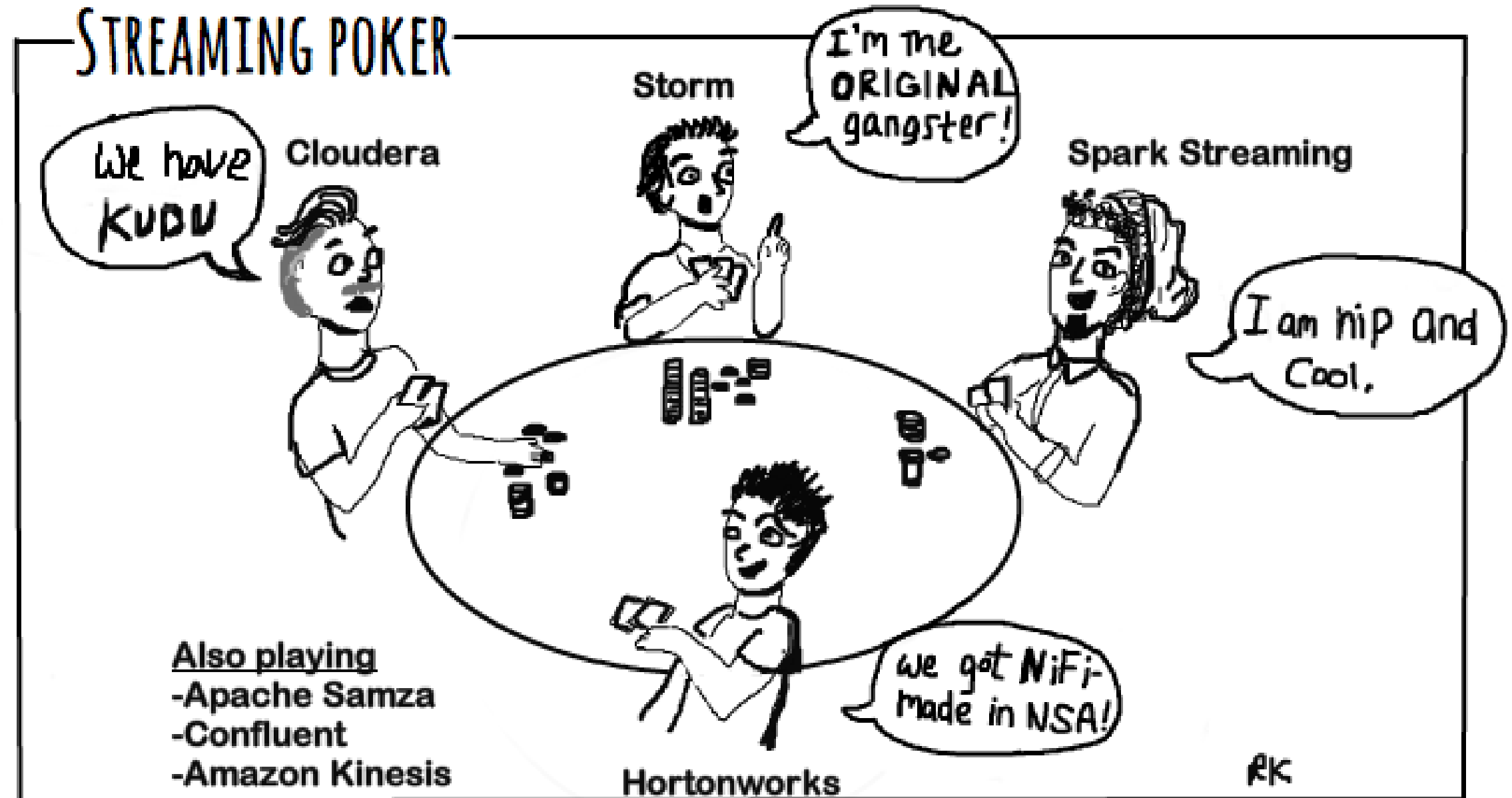
Identification

- EPC
- uCode
- IPv6
- URIs

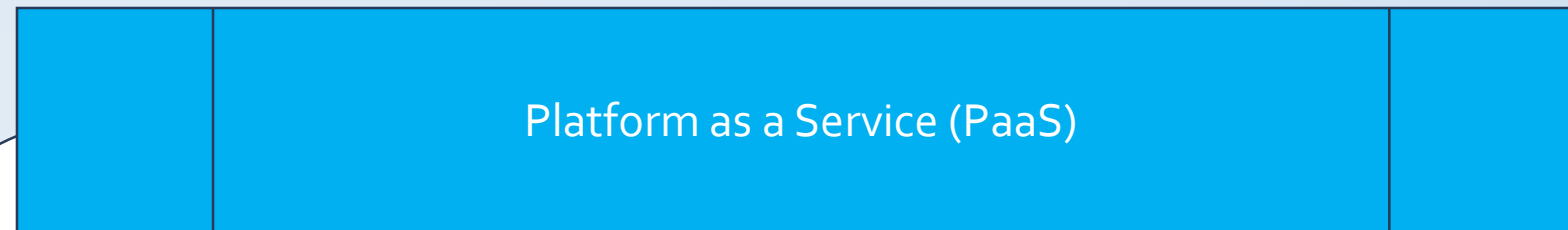
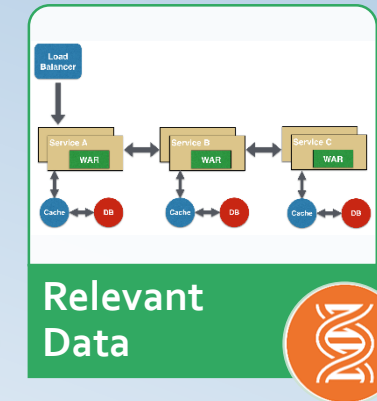
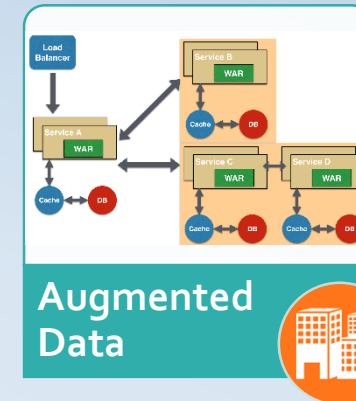
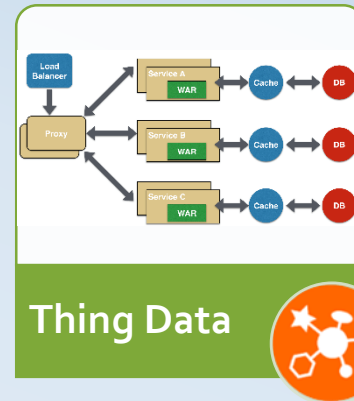
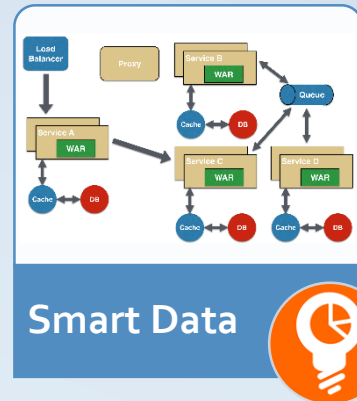
Infrastructure

- 6LoWPAN
- IPv4/IPv6
- RPL





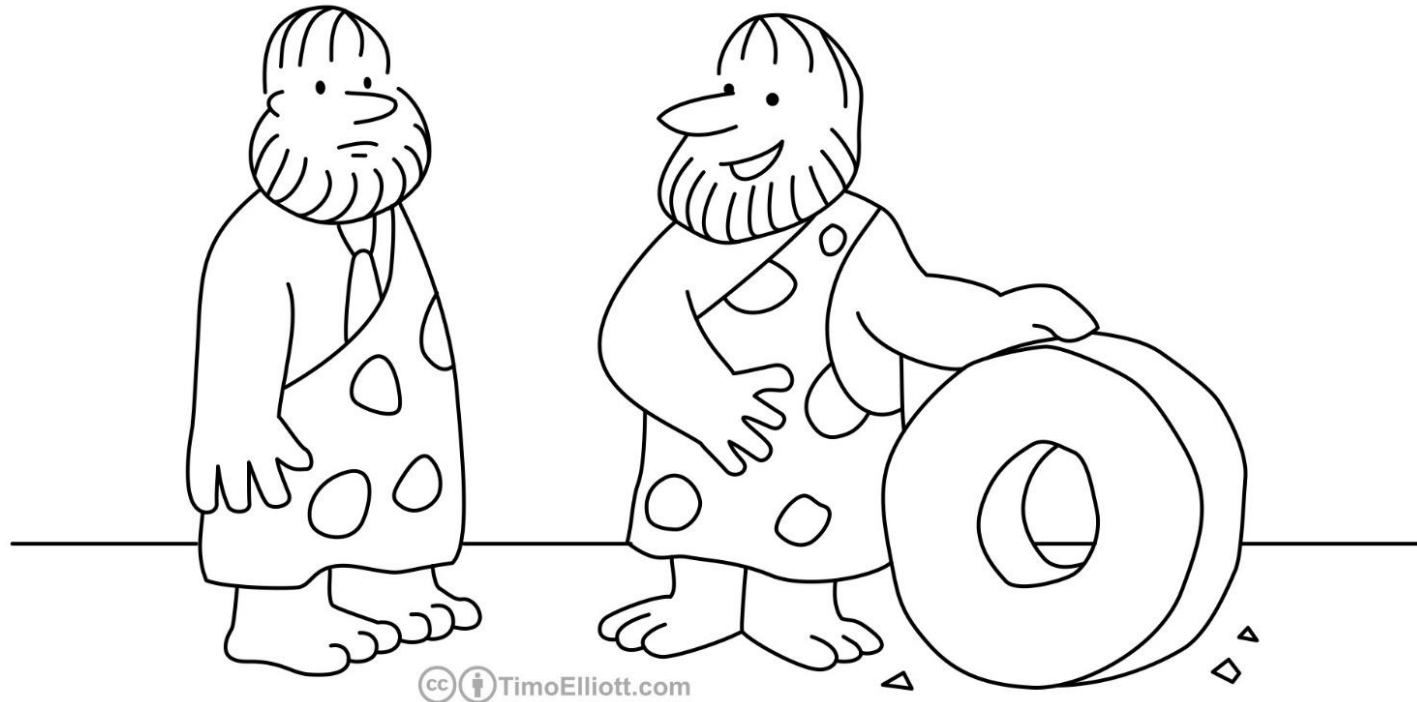
STAR Data Execution



Private, public, and hybrid cloud



Do STAR Data Wisely



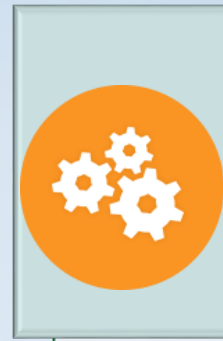
*“It does look similar—but this one
is powered by Hadoop”*



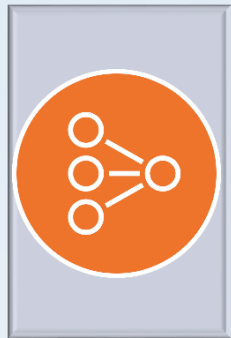
Best Practices



Convert



Condense



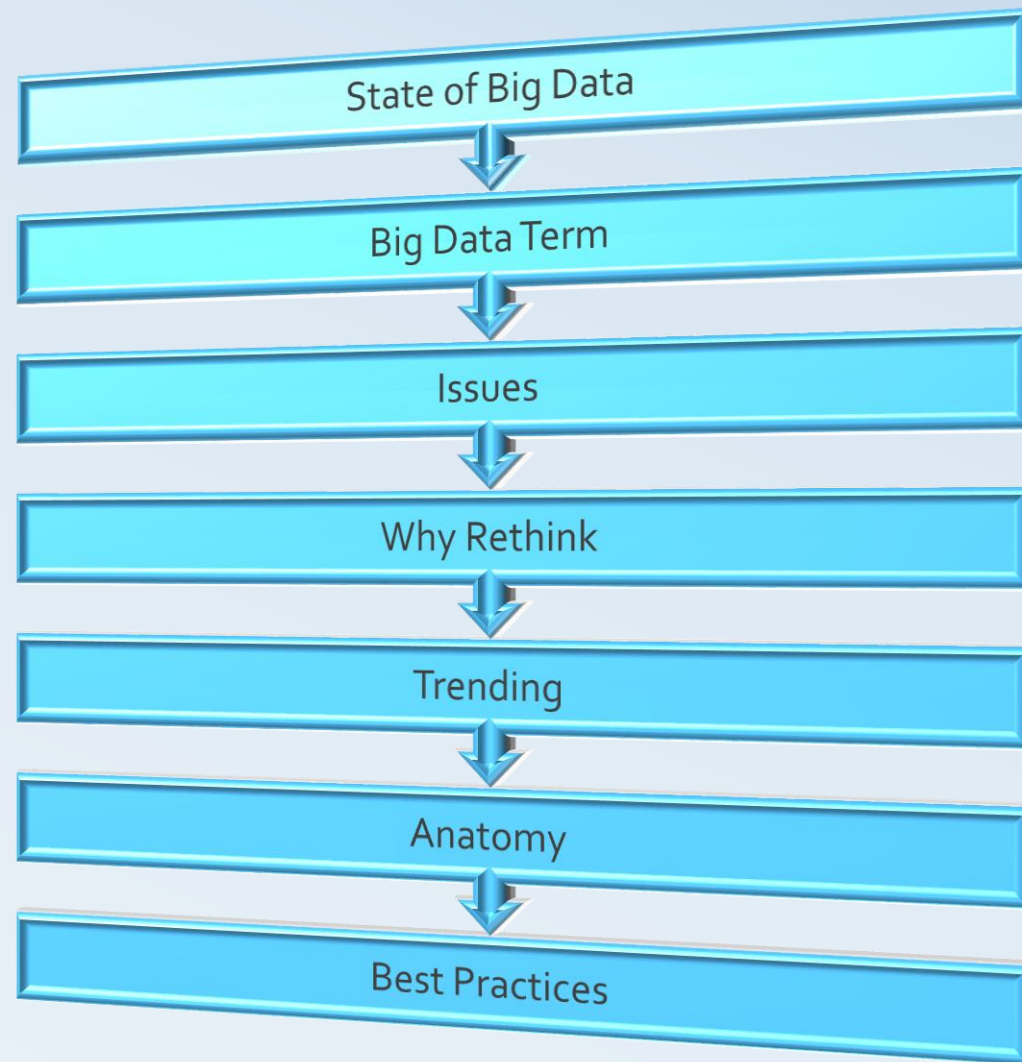
Converge



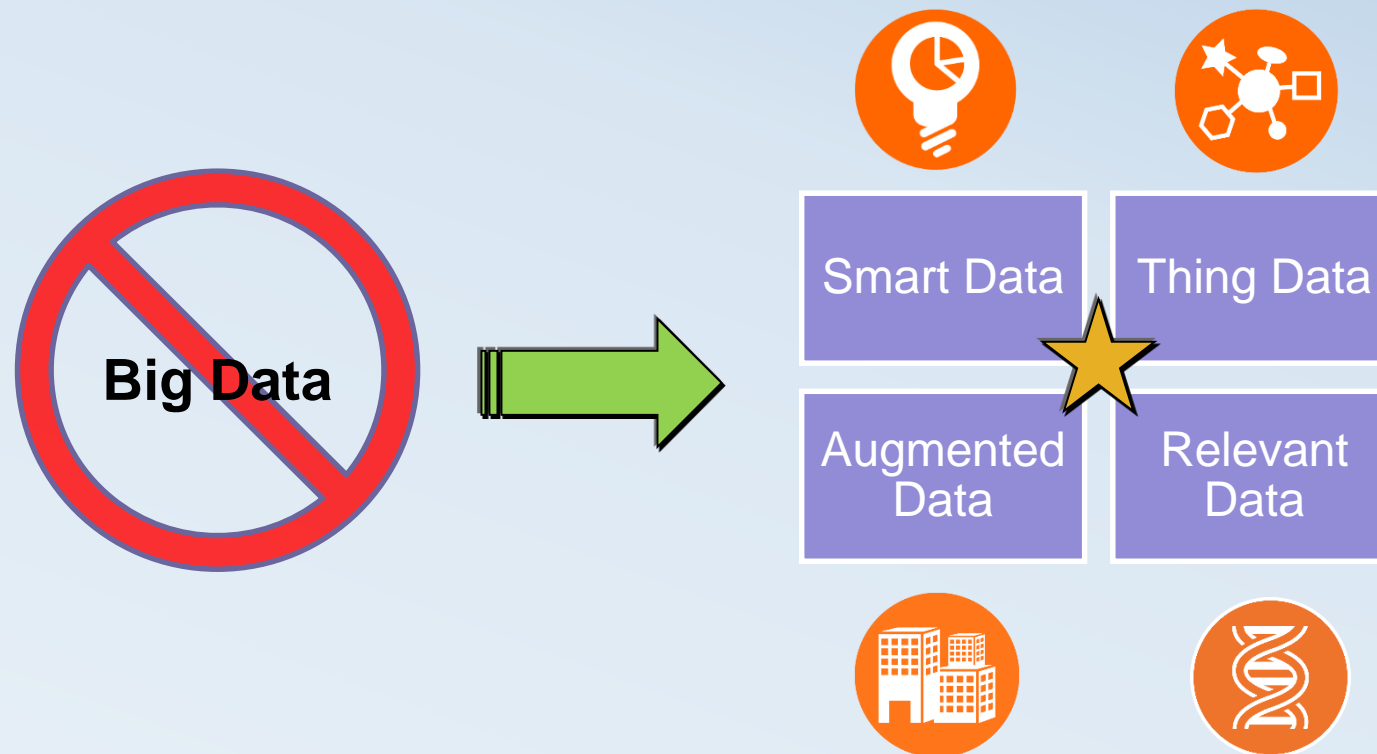
Convince

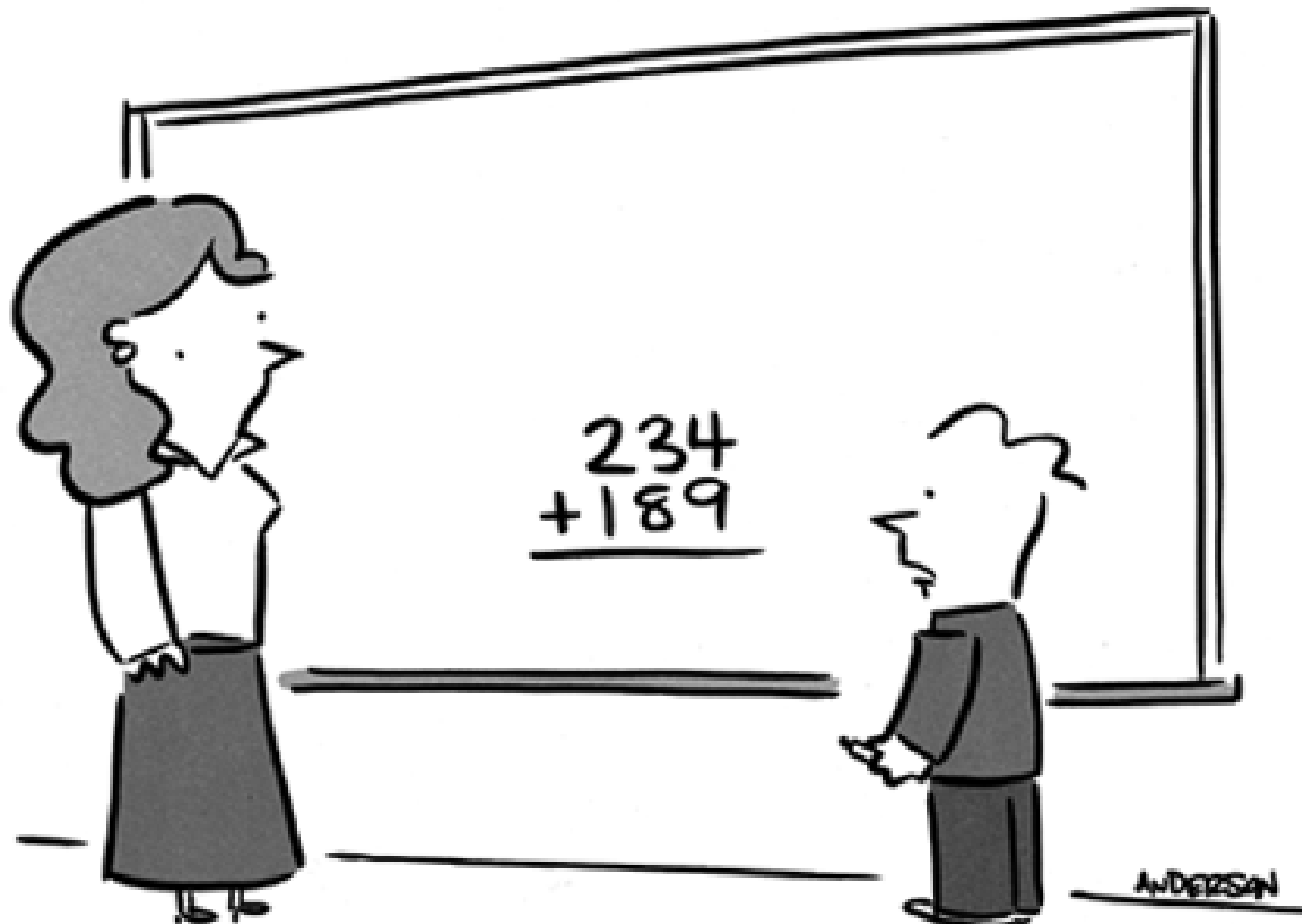


Summary



Key Takeaways





"Does this count as big data?"





© Tony Shan. Some rights reserved. All materials, contents and forms contained in this presentation are the intellectual property of the author and may not be copied, reproduced, distributed or displayed without author's express written permission.

Other streams of data and information from Internet are adapted and incorporated in the presentation for reference and illustration purposes. Some sources are not mentioned on the slides due to space and time constraints.

The author does not warrant, either expressly or implied, the accuracy, timeliness, or appropriateness of the information contained in this deck. The author disclaims any responsibility for content errors, omissions, or infringing material, and disclaims any responsibility associated with relying on the information provided in this document. The author also disclaims all liability for any material contained in other resources, used in or linked to this file.



Contact

mail@tonyshan.com

