

THE NEXT
BIG
THING

Introductory Data Science and Analytics

Rick Bahague

Co-Founder, PyData Philippines

Member, Computer Professionals' Union

HELLO!



Member
Computer Professionals' Union

Co-Founder
PyData Philippines



GITHUB.COM/RICKBAHAGUE



MEDIUM.COM/@RBAHAGUEJR

THE NEXT
BIG
THING



Radio Engineering Circle

A long time ago in a galaxy far,
far away....

THEORETICAL PHYSICS GROUP
NATIONAL INSTITUTE OF PHYSICS
UP DILIMAN

*TIME PROBLEM IN QUANTUM
MECHANICS, EIGENVALUE PROBLEMS,
SCIENTIFIC COMPUTING*

THE NEXT
BIG
THING

Source https://en.wikipedia.org/wiki/Star_Wars_opening_crawl#/media/File:Opening_crawl.jpg

PHYSICAL REVIEW LETTERS

Highlights Recent Accepted Collections Authors Referees Search Press About ⚡

Confined Quantum Time of Arrivals

Eric A. Galapon, Roland F. Caballar, and Ricardo T. Bahague Jr
Phys. Rev. Lett. **93**, 180406 – Published 29 October 2004; Erratum Phys. Rev. Lett. **101**, 169901 (2008)

[Article](#) [References](#) [Citing Articles \(35\)](#) [PDF](#) [HTML](#) [Export Citation](#)

PHYSICAL REVIEW A

covering atomic, molecular, and optical physics and quantum information

Highlights Recent Accepted Authors Referees Search Press About ⚡

Confined quantum time of arrival for the vanishing potential

Eric A. Galapon, Roland F. Caballar, and Ricardo Bahague
Phys. Rev. A **72**, 062107 – Published 6 December 2005; Erratum Phys. Rev. A **78**, 049902 (2008)

[Article](#) [References](#) [Citing Articles \(19\)](#) [PDF](#) [HTML](#) [Export Citation](#)

A long time ago in a galaxy far,
far away....

TECHNICAL SPECIALIST
MODNET/PHCOLO (ISP)

DATABASE MANAGER
BCD PINPOINT DM (ADVERTISING)

NON-PROFIT WORK / WEB &
APPLICATION DEVELOPMENT

SENIOR "DATA SCIENTIST" (TELCO)

THE NEXT
BIG
THING

STAR WARS



Radio Engineering Circle

A long time ago in a galaxy far,
far away....



BE FAMILIAR WITH DIVERSE DATA SETS

THE NEXT
BIG
THING

STAR
WARS



Radio Engineering Circle

OUTLINE

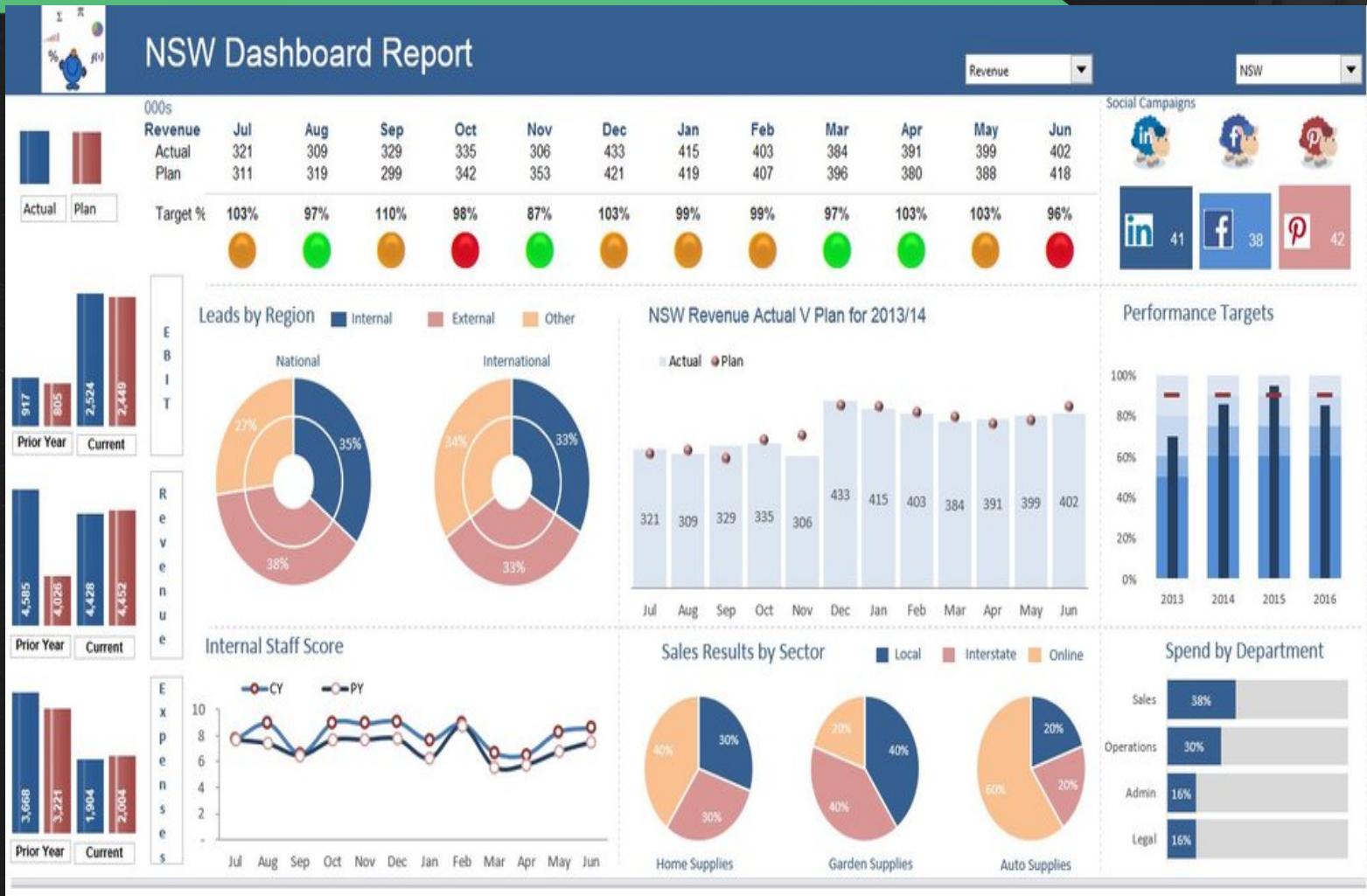
- Analytics, Big Data and Data Science
- Data Scientists
- Data Science and Big Data Applications
- Tools
- Development Track

THE NEXT
BIG
THING



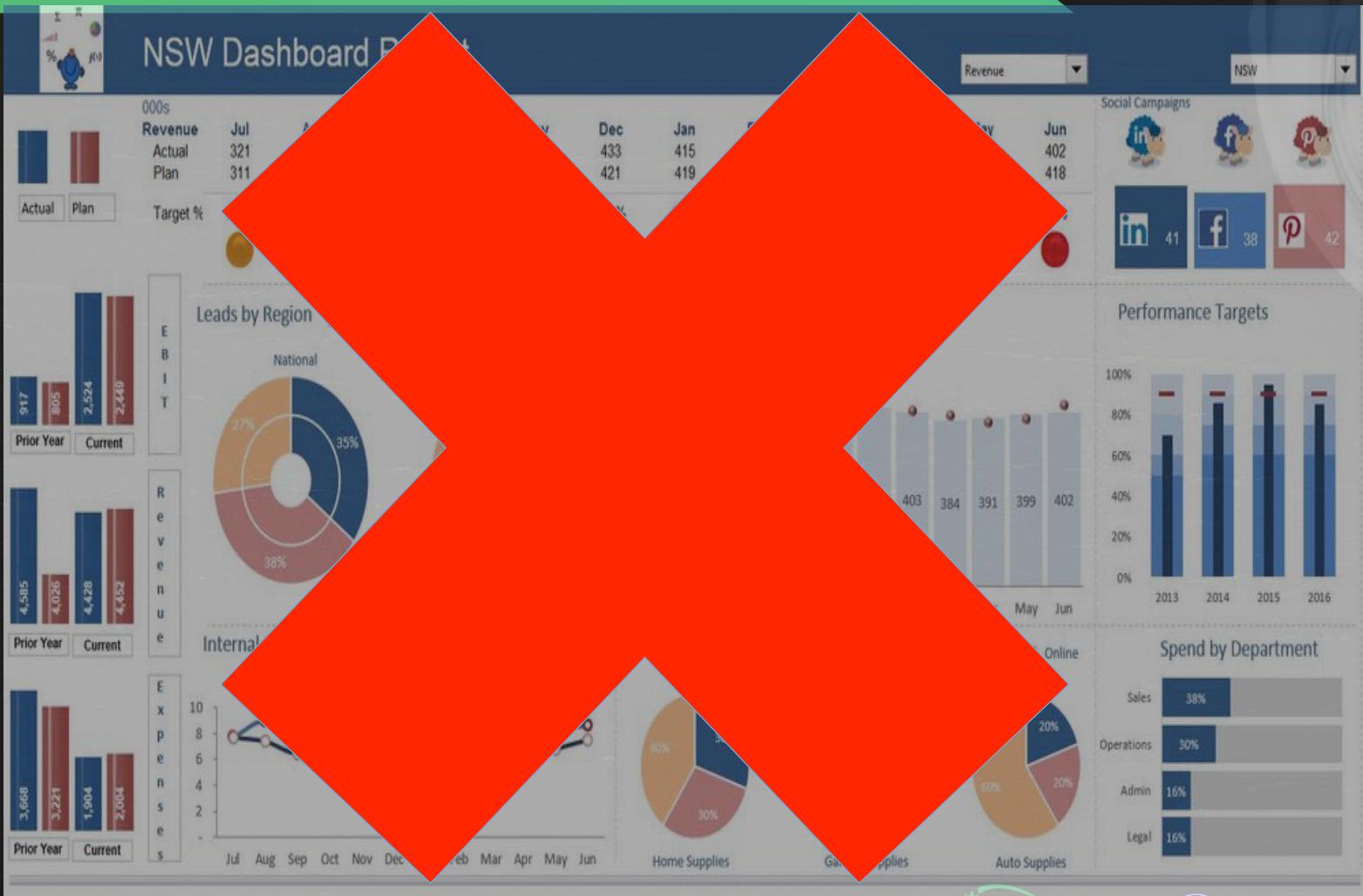
Radio Engineering Circle

WHAT IS ANALYTICS ?



Source: <https://static1.squarespace.com/static/57fc8a3d482e945c66724f7/l/585dfefcfc7fc506c9935130b/1482555000950/Excel+Dashboard+Report>

DASHBOARD != ANALYTICS



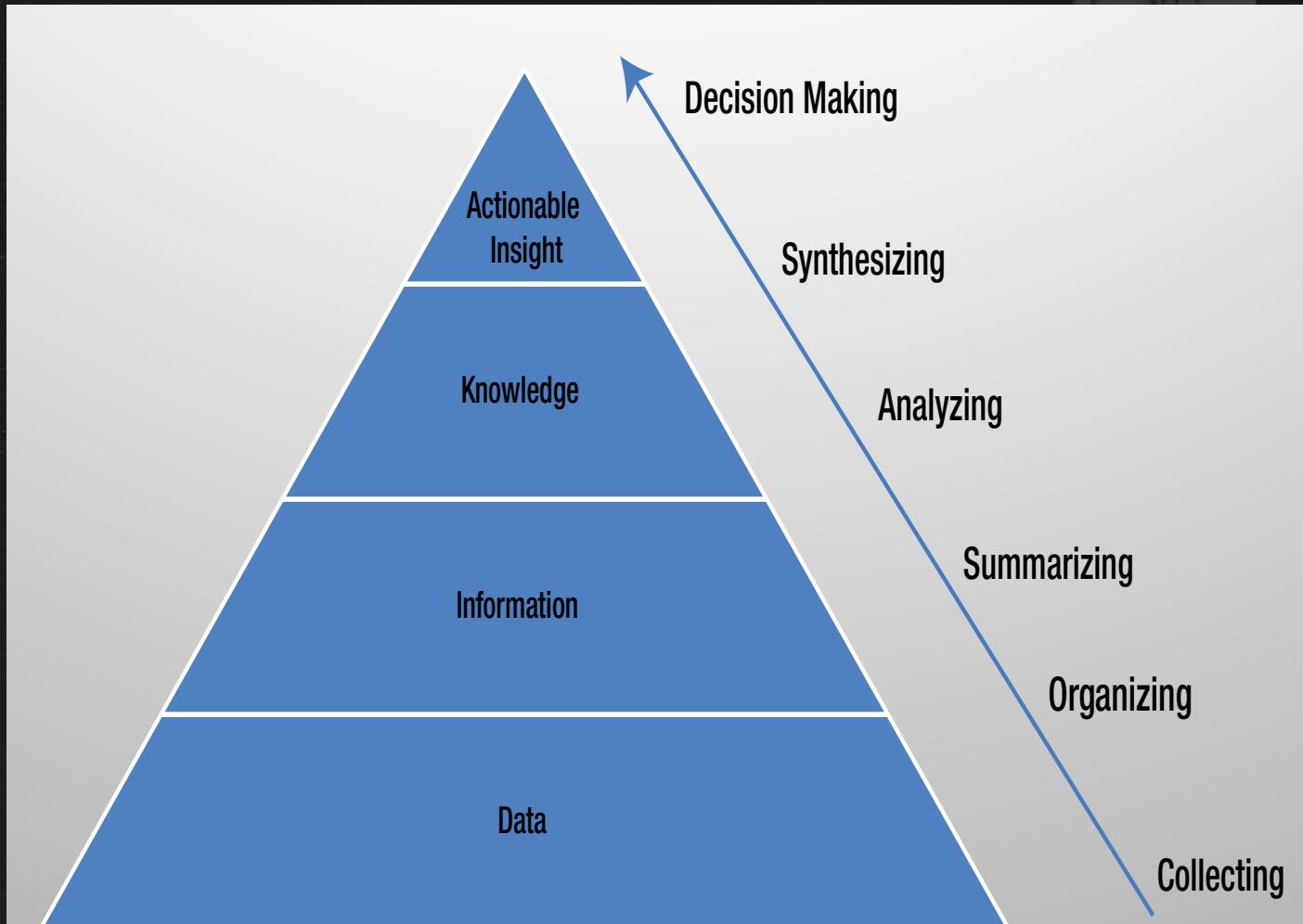
THE NEXT
BIG
THING

Source: <https://static1.squarespace.com/static/57fcf8a3d482e945c66724f7/l/585dfefcfc7c506c9935130b/1482555000950/Excel+Dashboard+Report>



Radio Engineering Circle

ANALYTICS

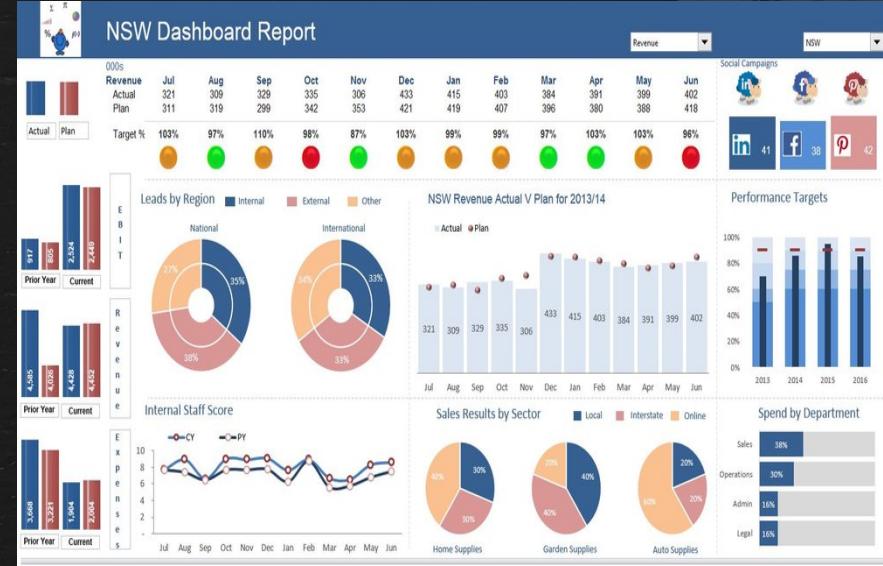


Source: Big Data Imperatives

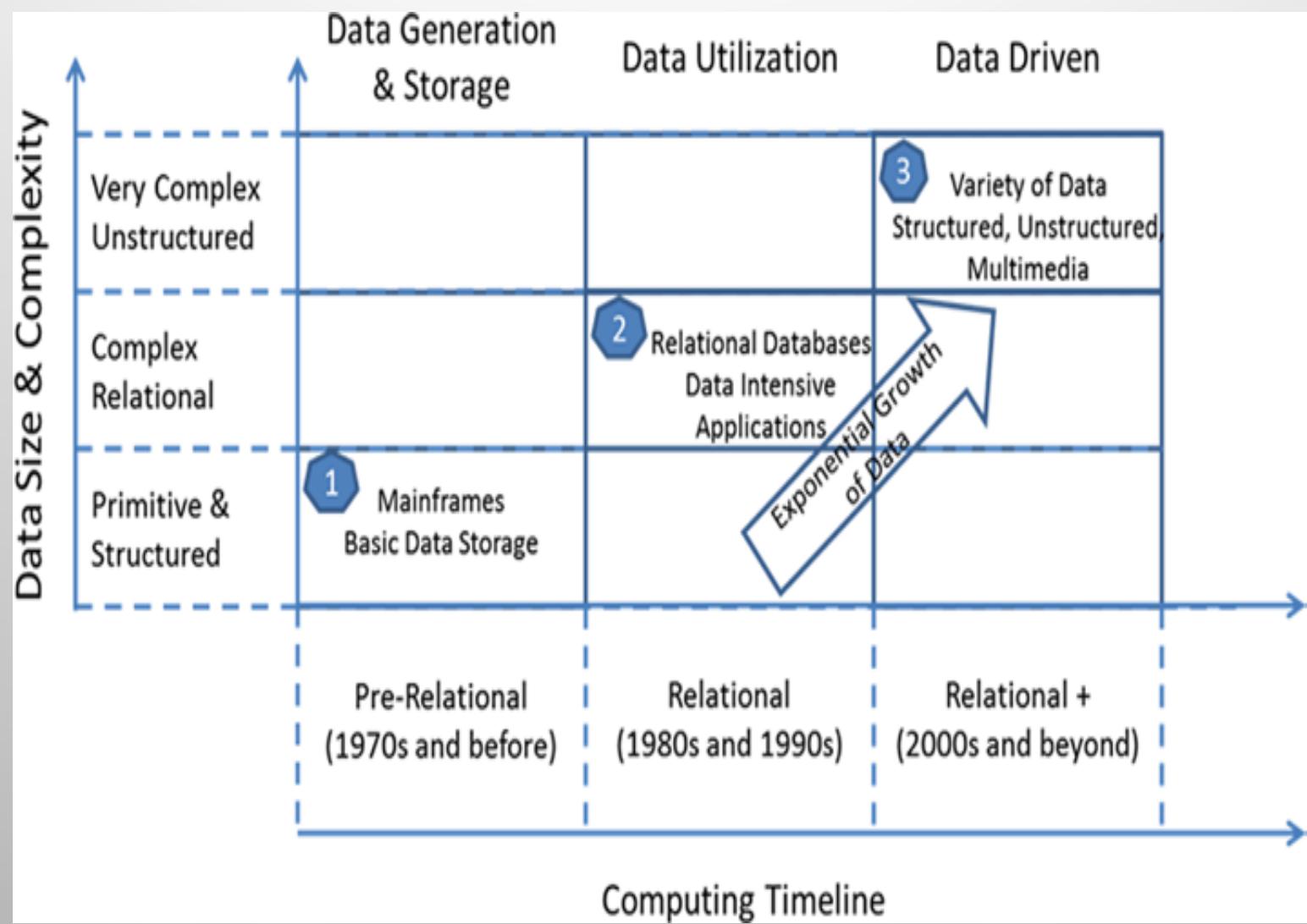
DASHBOARD IS A TOOL

OPERATIONAL ANALYTICS OR
EMBEDDED ANALYTIC

DATA DISCOVERY OR
EXPLORATORY ANALYTICS



Big Data



Big Data Value Across Industries

	Volume of Data	Velocity of Data	Variety of Data	Under -Utilized Data ('Dark Data')	Big Data Value Potential
Banking and Securities	High	High	Low	Medium	High
Communications & Media Services	High	High	High	Medium	High
Education	Very Low	Very Low	Very Low	High	Medium
Government	High	Medium	High	High	High

ANALYSTS & BIG DATA

INCREASING NEED FOR DATA ANALYSTS TO BE FAMILIAR AND SKILLED ON BIG DATA PROCESSING AND ANALYTICS



THE NEXT
BIG
THING

Source: <https://static1.squarespace.com/static/57fcf8a3d482e945c66724f7/l/585dfecfcf7c506c9935130b/1482555000950/Excel+Dashboard+Report>



Radio Engineering Circle

ANALYTICAL SKILLS

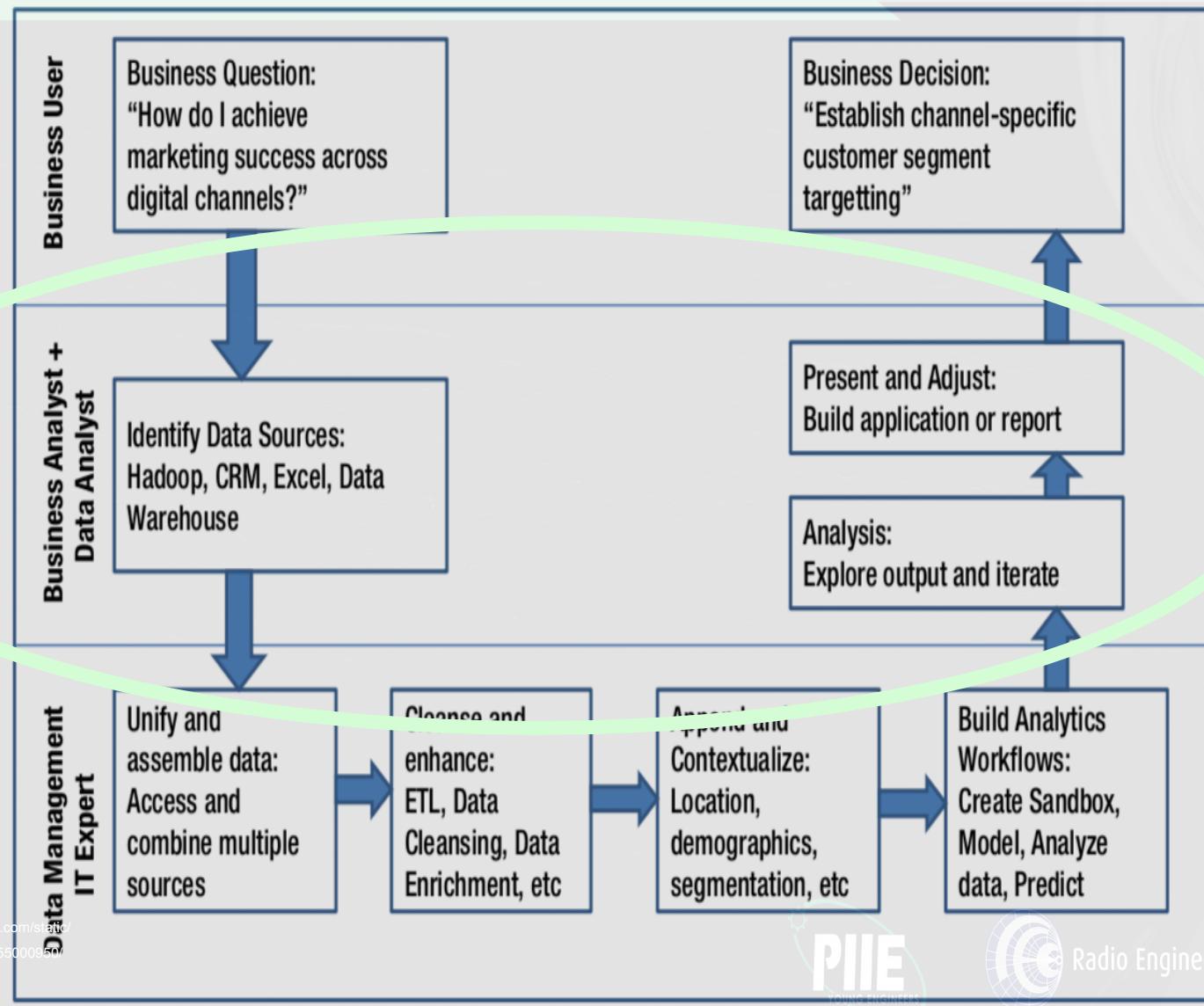
- CURIOSITY
- STATISTICS
- DATA CLEAN-UP OR CARPENTRY

THE NEXT
BIG
THING

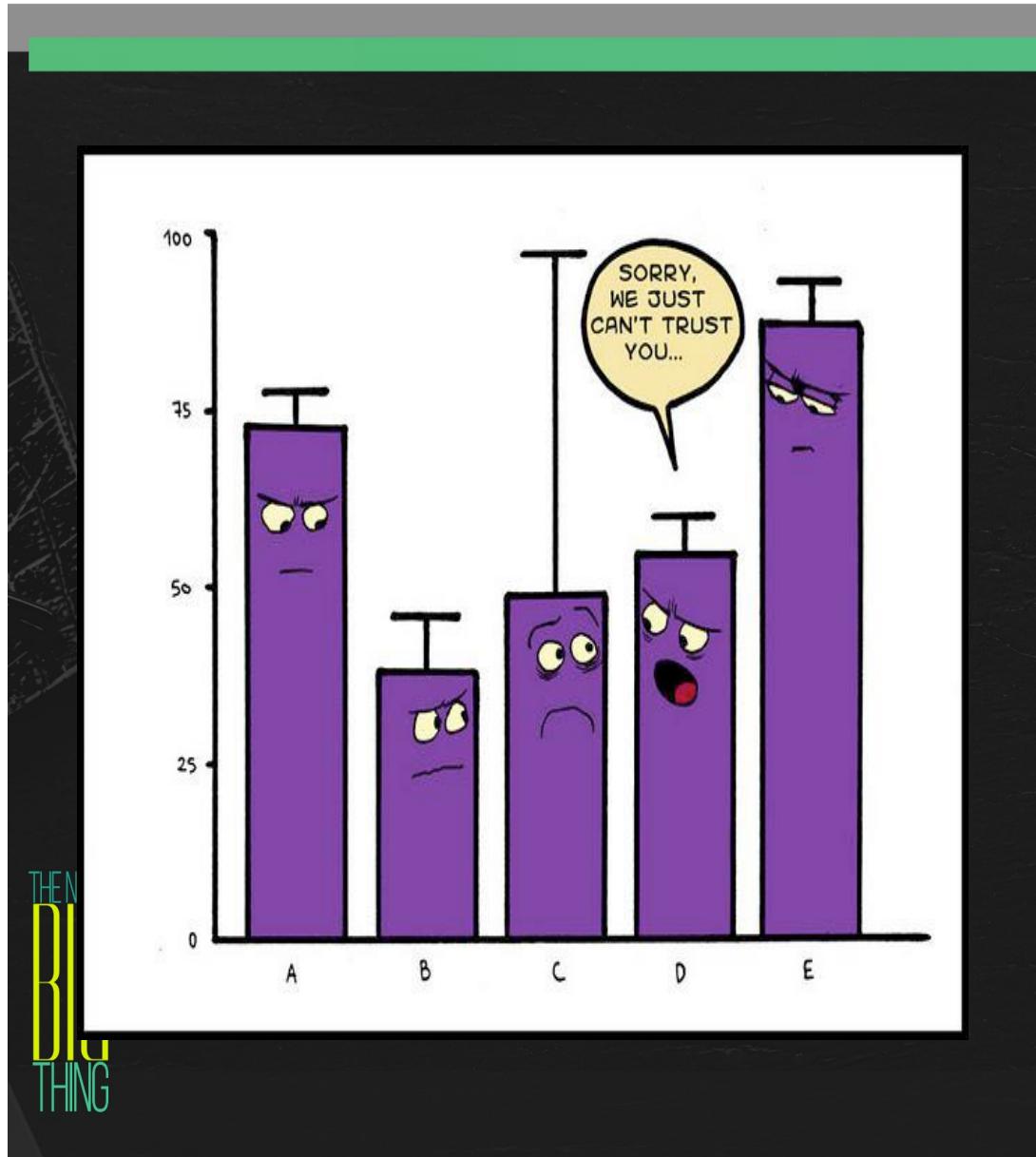
Source: <https://static1.squarespace.com/static/57fcf8a3d482e945c66724f7/l/585dfebcf7c506c9935130b/1482555000950/Excel+Dashboard+Report>

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	HIGHLY SIGNIFICANT
0.02	HIGHLY SIGNIFICANT
0.03	HIGHLY SIGNIFICANT
0.04	SIGNIFICANT
0.049	SIGNIFICANT
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	ON THE EDGE OF SIGNIFICANCE
0.07	HIGHLY SUGGESTIVE,
0.08	SIGNIFICANT AT THE P<0.10 LEVEL
0.09	SIGNIFICANT AT THE P<0.10 LEVEL
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥0.1	THIS INTERESTING SUBGROUP ANALYSIS

TRADITIONALLY,



DATA SCIENCE

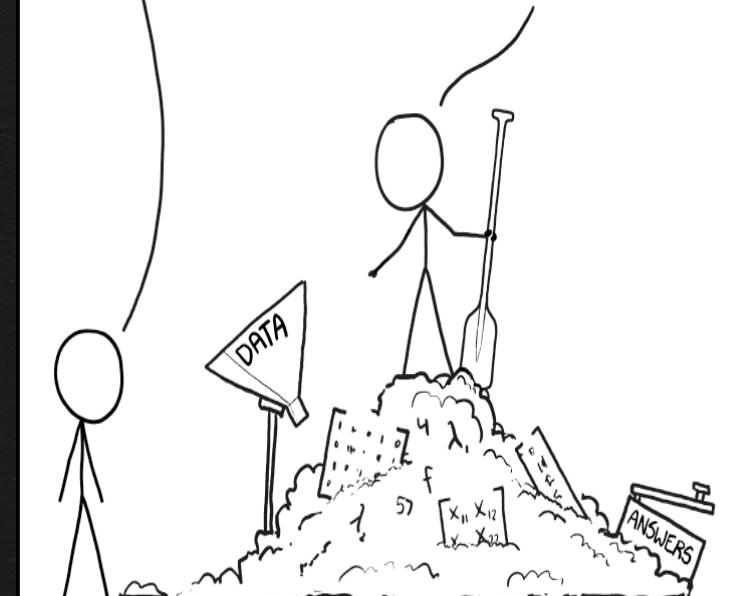


THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.



https://imgs.xkcd.com/comics/machine_learning.png

DATA SCIENCE

Science is a way of thinking much more than it is a body of knowledge

Carl Sagan.

Our perspective is that data science is the child of **statistics** and **computer science**. While it has inherited some of their methods and thinking, it also seeks to blend them, refocus them, and develop them to address the context and needs of modern scientific data analysis. This perspective is not new. Over 50 years ago, Tukey (7) defined “data analysis” as a broad endeavor, much broader than traditional mathematical statistics. In a sense, today’s data science, although set against a modern backdrop, is cast from Tukey’s original.

Biel & Smith, PNAS , <https://doi.org/10.1073/pnas.1702076114>



Source: <https://static1.squarespace.com/static/57fcf8a3d482e945c66724f7/l/585dfecfcf7c506c9935130b/1482555000950/Excel+Dashboard+Report>



Radio Engineering Circle

DATA SCIENCE

exploits the modern deluge of data for prediction, exploration, understanding, and intervention.

emphasizes the value and necessity of approximation and simplification.

values effective communication of the results of a data analysis and of the understanding about the world

THE NEXT
BIG
THING



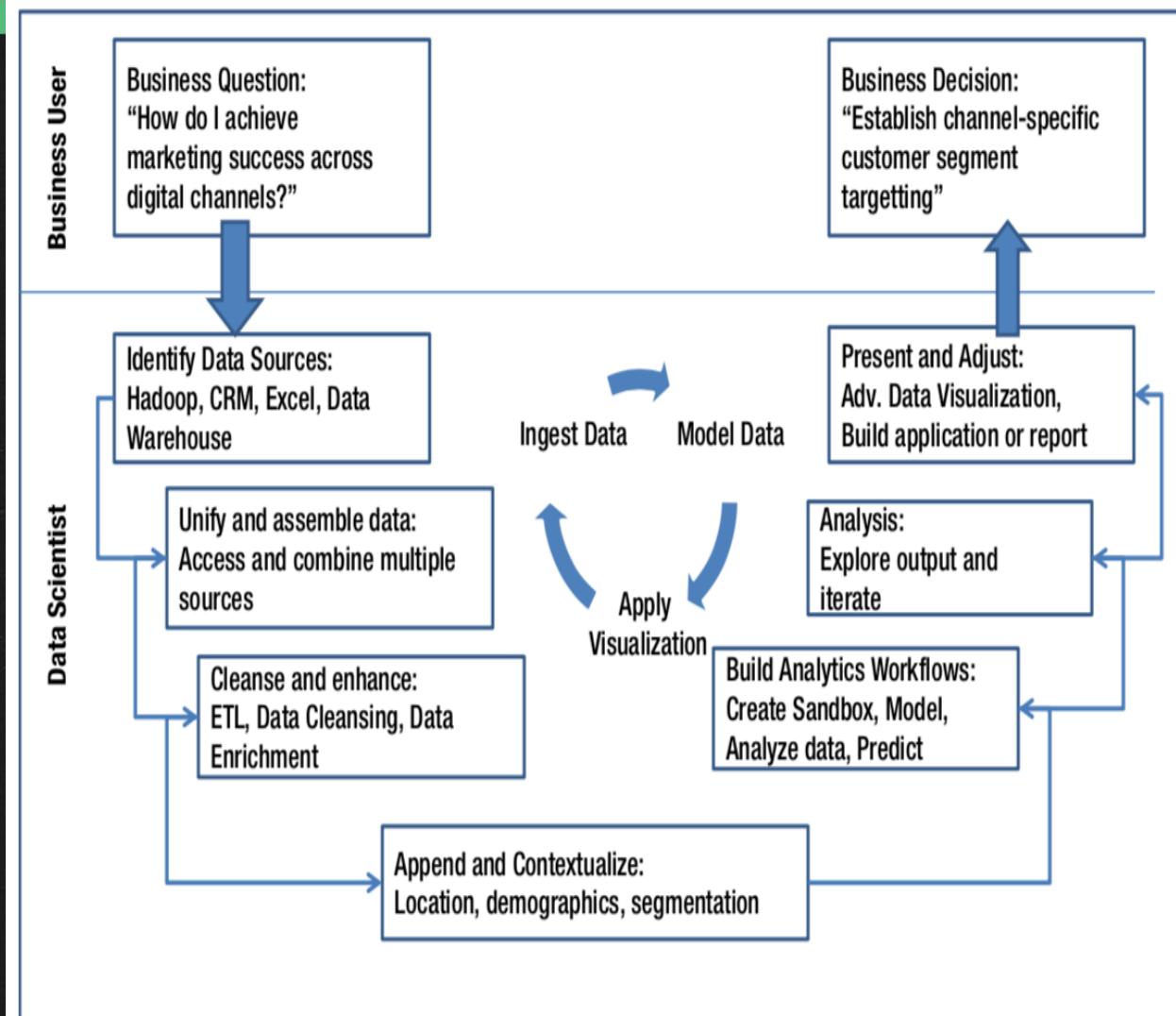
Radio Engineering Circle

DATA SCIENCE

prioritizes an understanding of the optimization algorithms and transparently managing the inevitable tradeoff between accuracy and speed

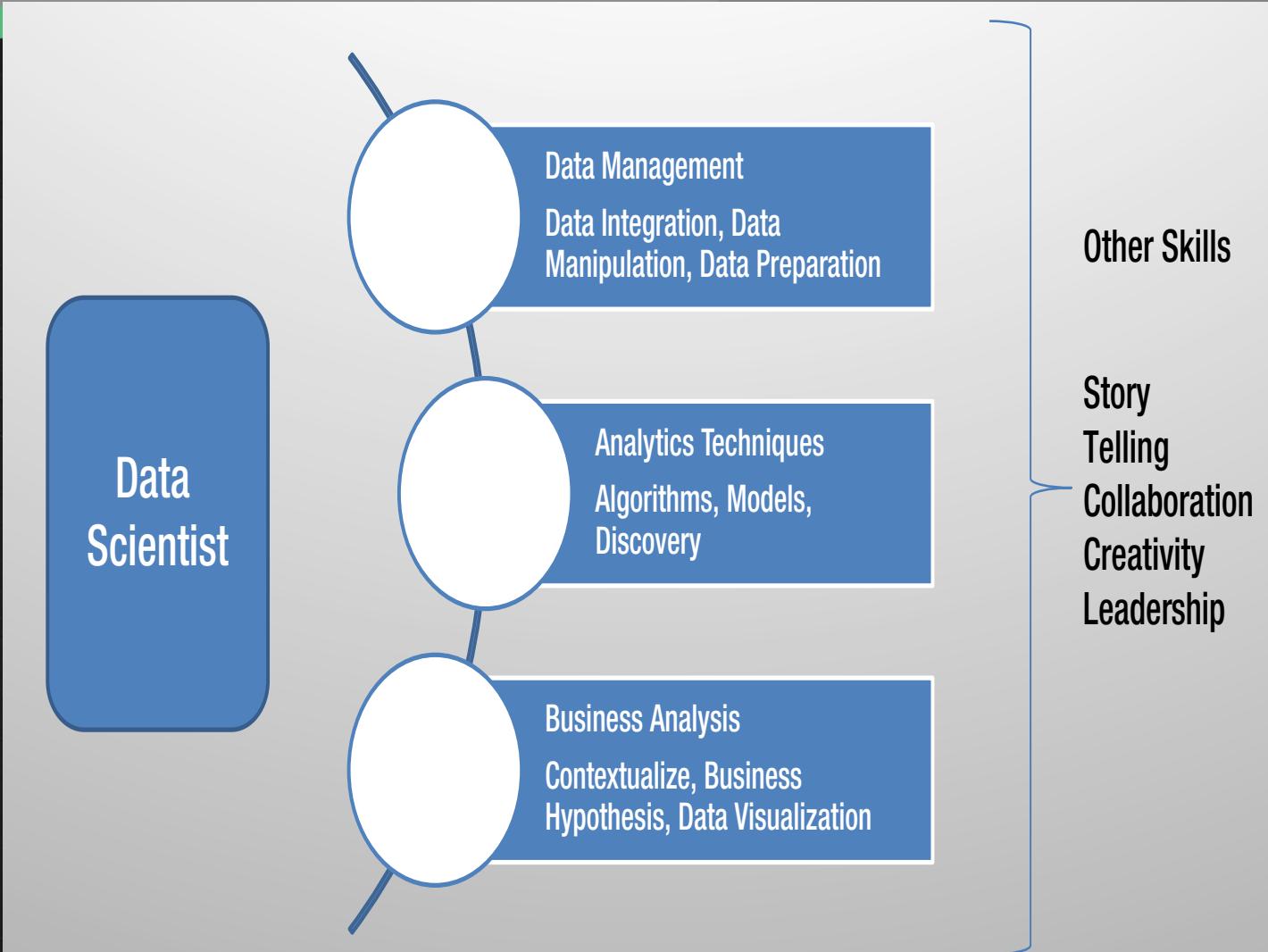
promotes domain-specific analyses, where data scientists and domain experts work together to balance appropriate assumptions with computationally efficient methods

BIG DATA & DATA SCIENCE



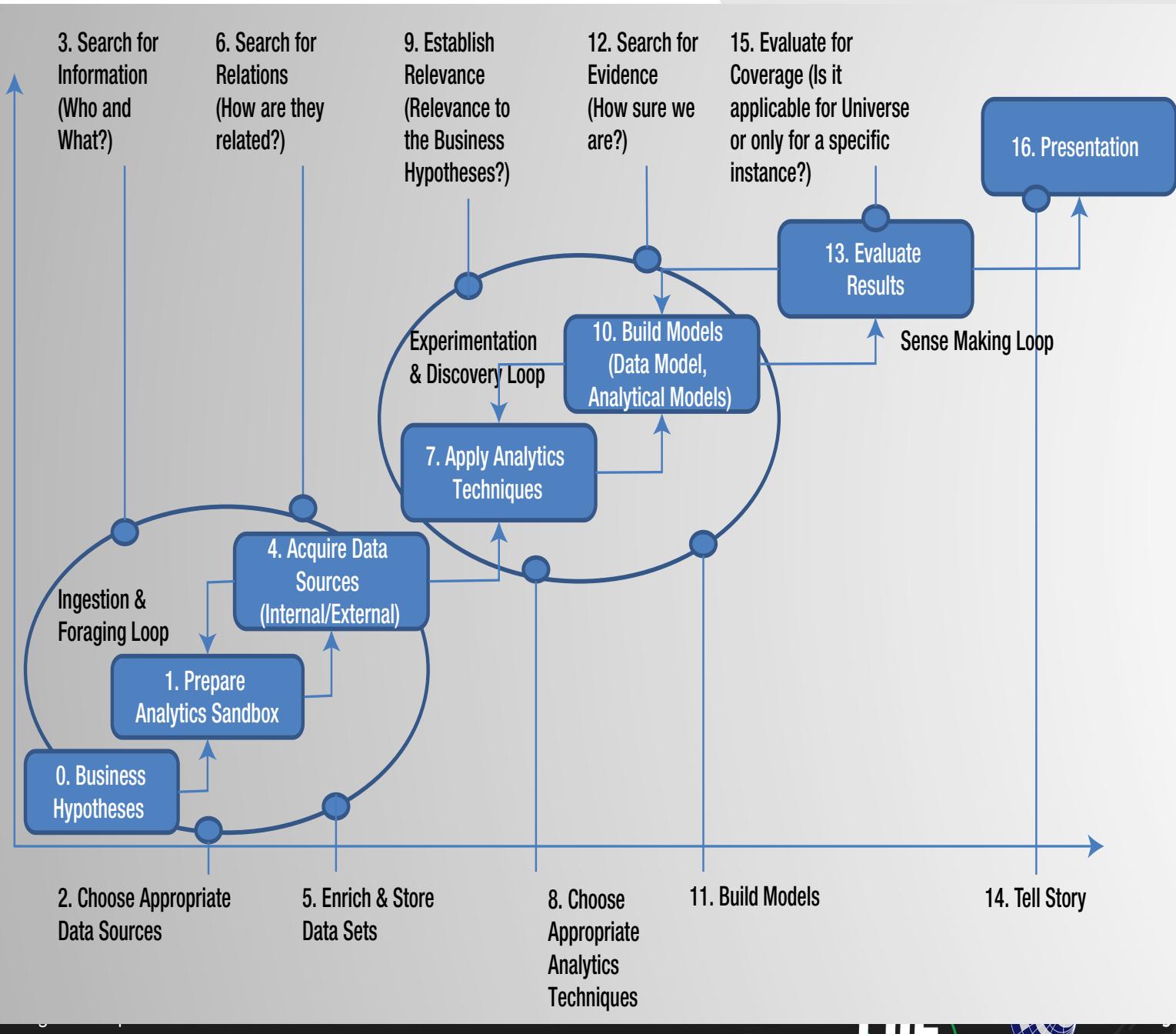
Source: Big Data Imperative

DATA SCIENTISTS



THE NEXT
BIG
THING

Source: Big Data Imperatives



ARE WE THIS GUY?

DATA SCIENCE TEAM

STATISTICS
COMPUTER SCIENCE
NATURAL SCIENCES
ENGINEERING

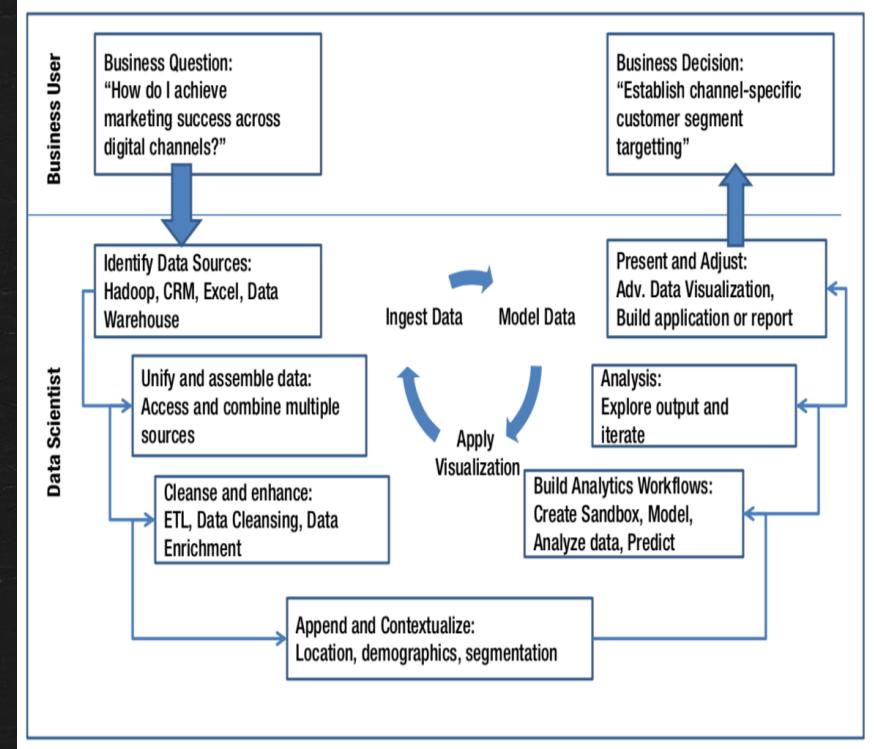
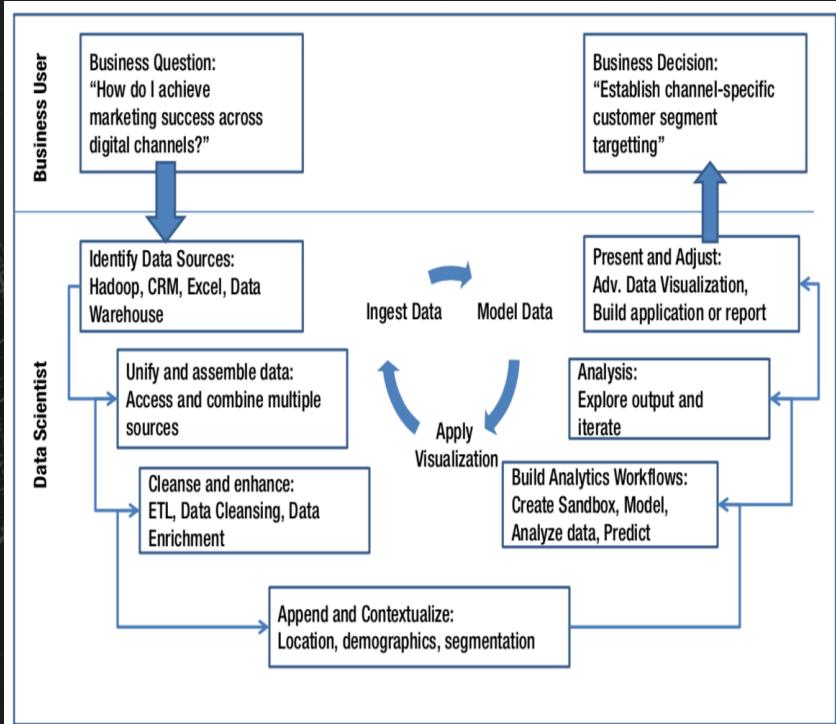
THE NEXT
BIG
THING

Source: Big Data Imperatives



Radio Engineering Circle

DATA SCIENTISTS & ANALYSTS



Data Science and Big Data Applications

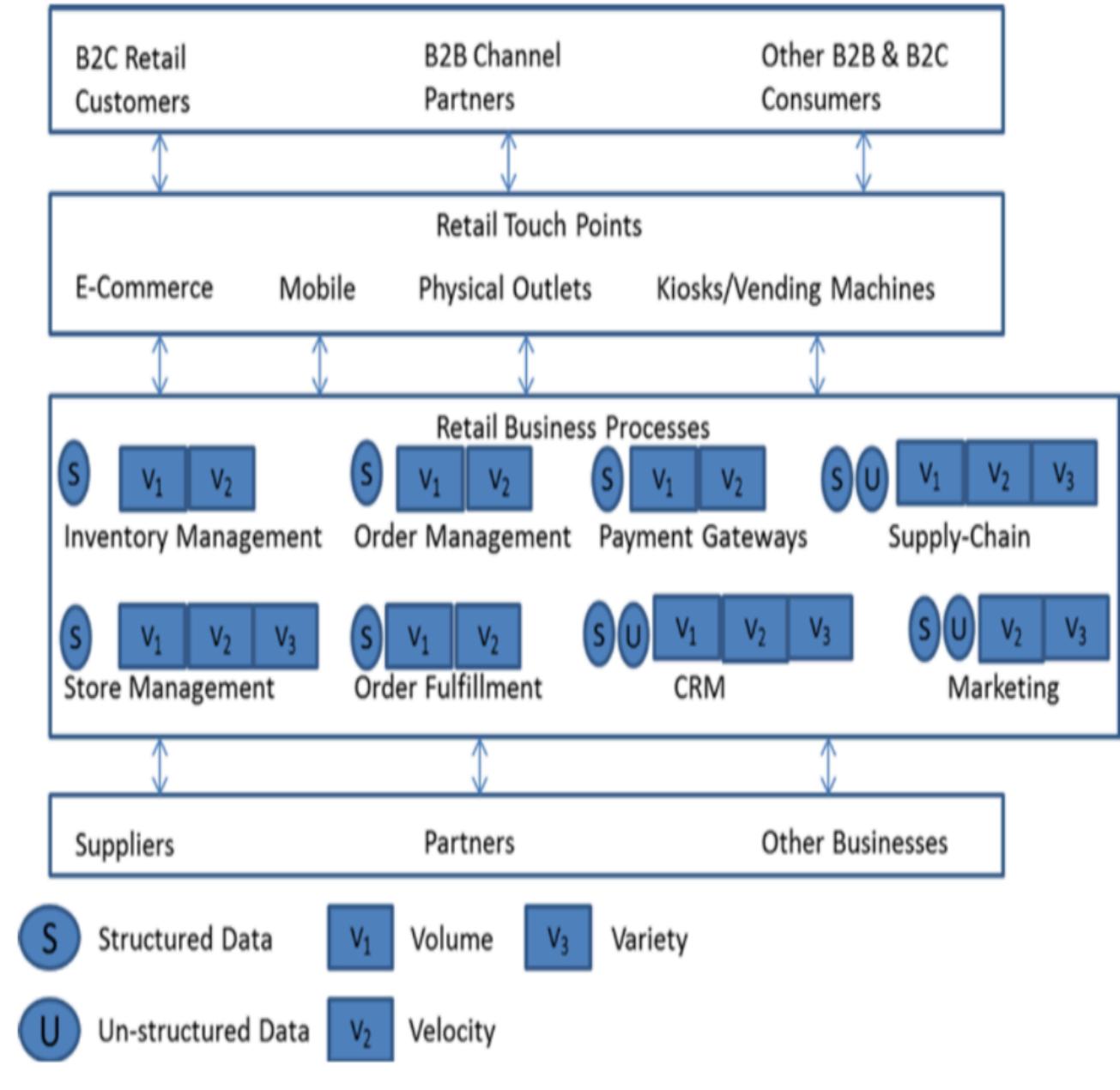
Retail	Manufacturing		
✓ Customer Relationship Management ✓ Store Location & Layout	✓ Fraud Detection & Prevention ✓ Supply-Chain optimization ✓ Dynamic Pricing	✓ Product Research ✓ Engineering Analysis ✓ Predictive Maintenance	✓ Process & Quality Metrics ✓ Distribution Optimization
Financial Services		Media & Telecommunications	
✓ Algorithmic Trading ✓ Risk Analysis	✓ Fraud Detection ✓ Portfolio Analysis	✓ Network Optimization ✓ Customer Scoring	✓ Churn Prevention ✓ Fraud Prevention
Advertising & Public Relations		Energy	
✓ Demand Signaling ✓ Targeted Advertising	✓ Sentiment Analysis ✓ Customer Acquisition	✓ Smart Grid ✓ Exploration	✓ Operational Modeling ✓ Power-Line Sensors
Government		Healthcare & Life Sciences	
✓ Market Governance ✓ Weapon Systems & Counter Terrorism	✓ Econometrics ✓ Health Informatics	✓ Pharmacogenomics ✓ Bioinformatics	✓ Pharmaceutical Research ✓ Clinical Outcomes Research

Source: Big Data Imperative

Application in Retail Industry

THE NEXT
BIG
THING

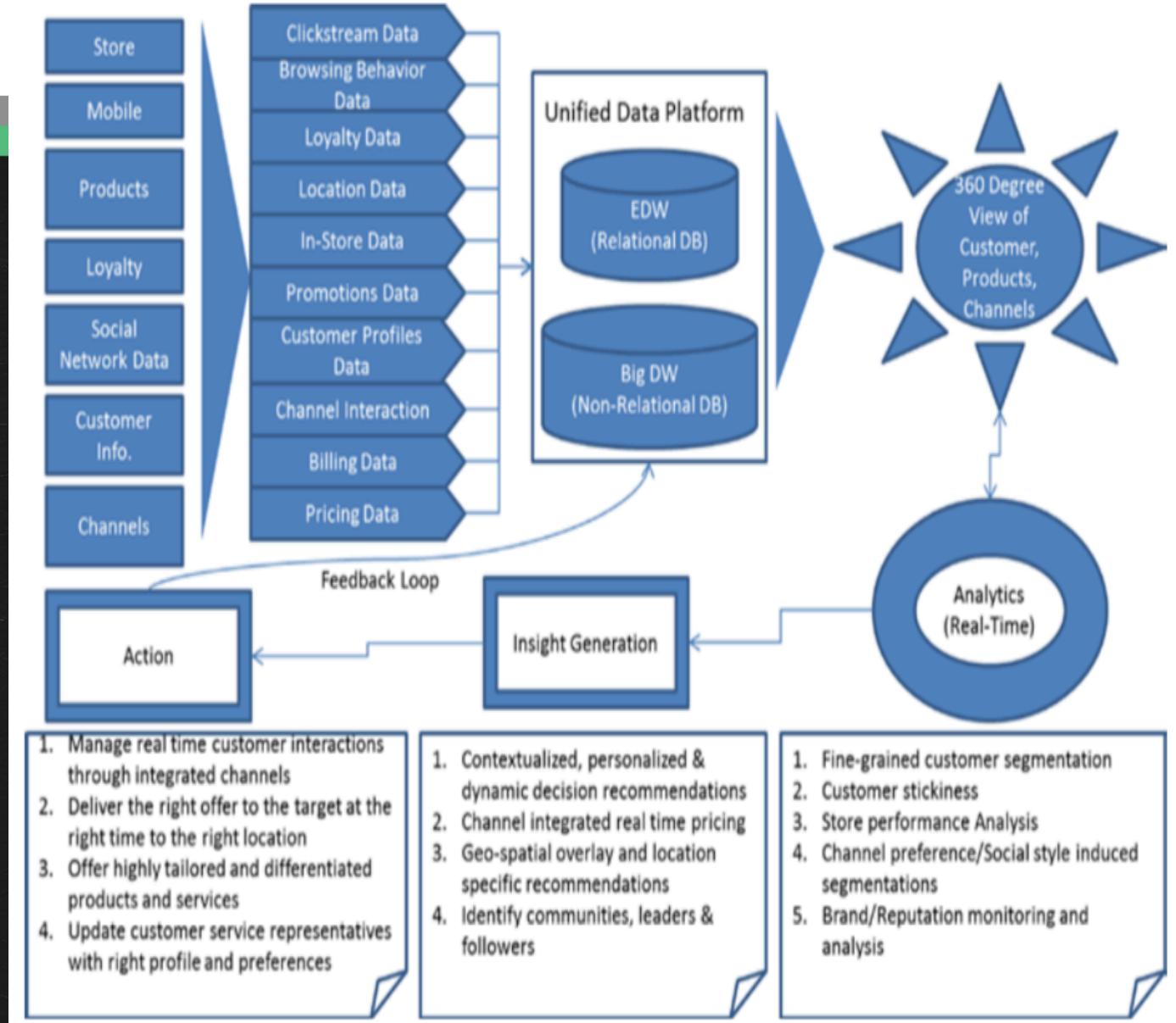
Source: Big Data Imperatives



Application in Retail Industry

THE NEXT
BIG
THING

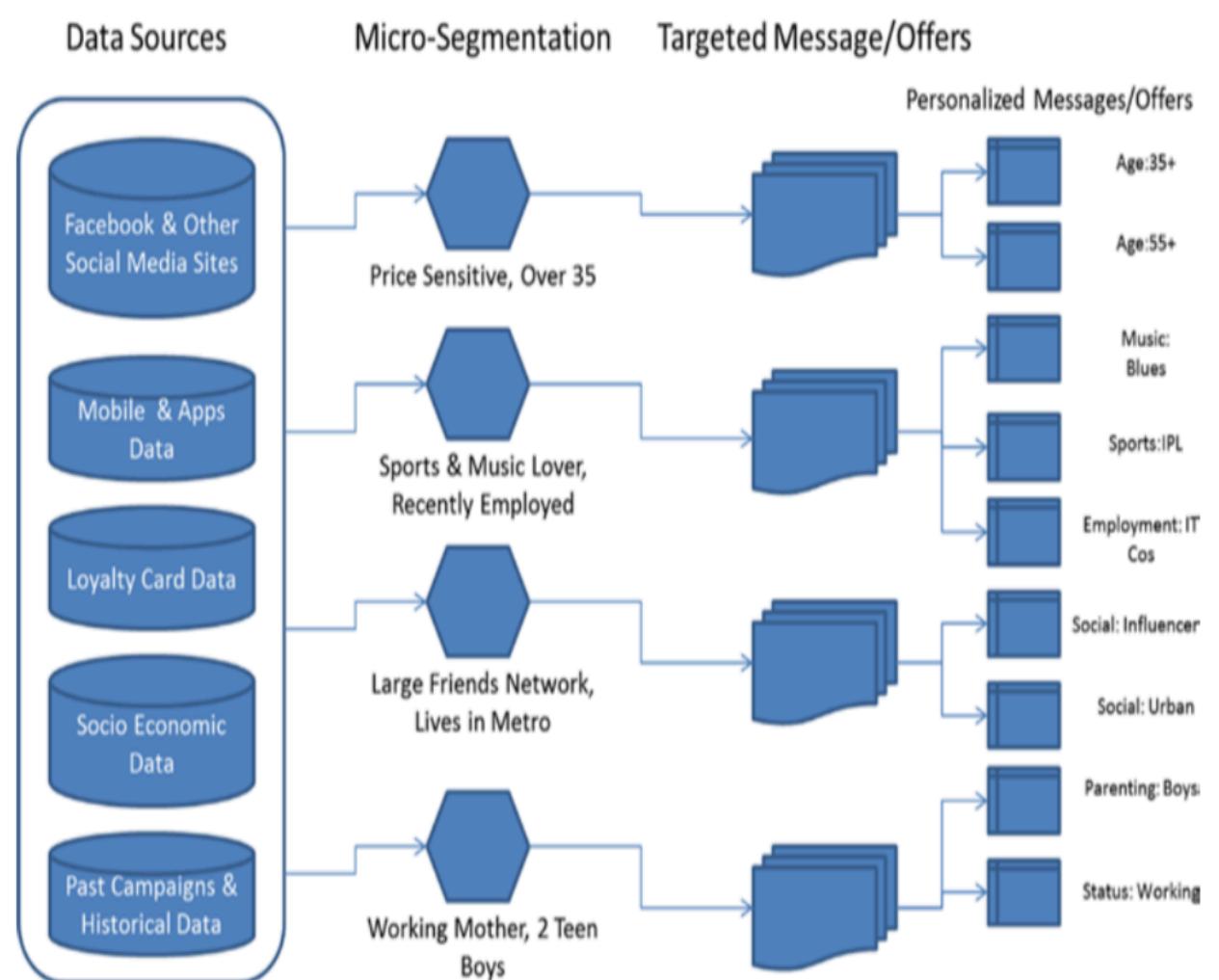
Source: Big Data Imperatives



Application in Retail Industry

THE NEXT
BIG
THING

Source: Big Data Imperatives

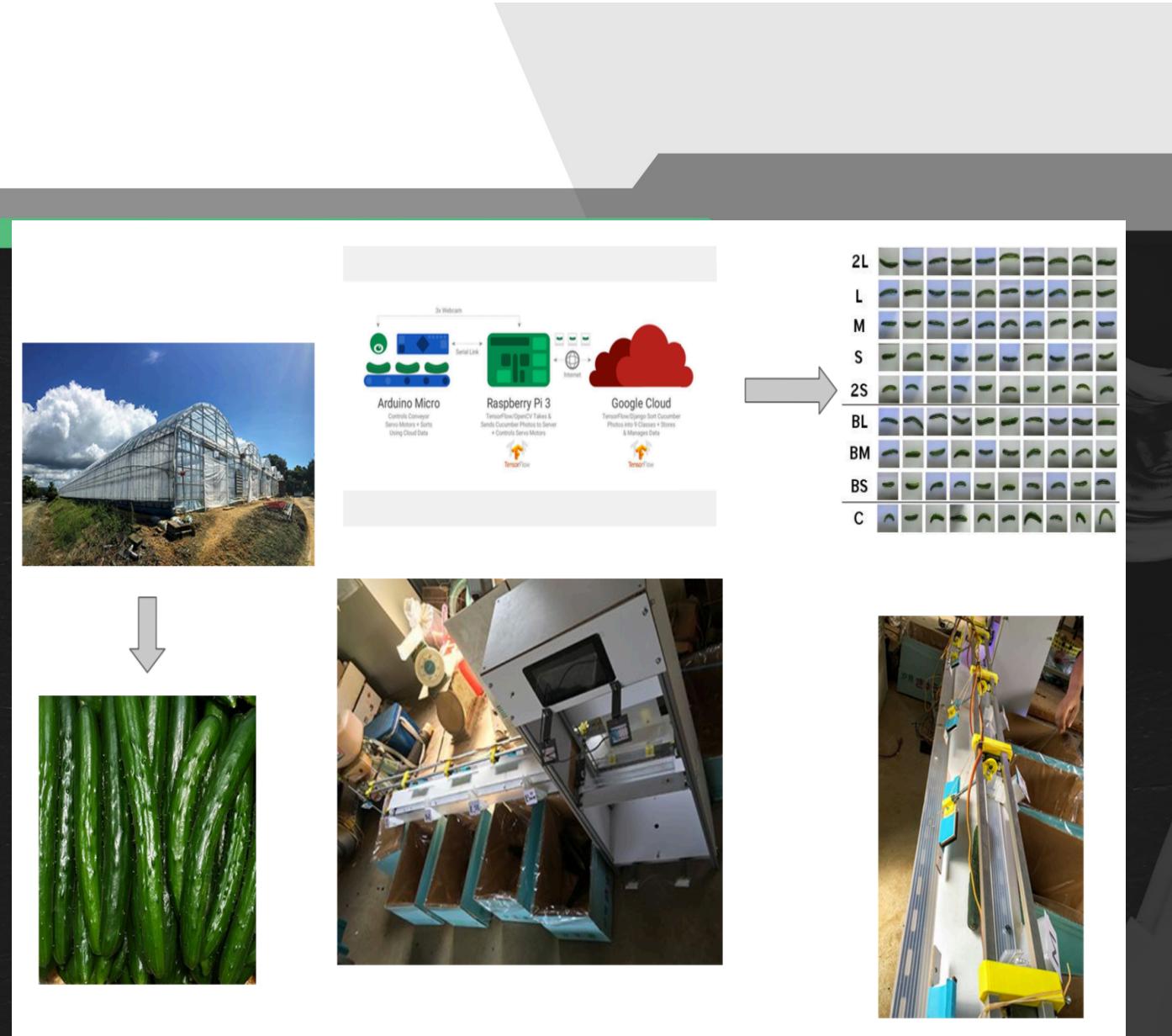


Radio Engineering Circle

Machine Learning Application

THE NEXT
BIG
THING

Source: Big Data Imperatives



Radio Engineering Circle

TOOLS



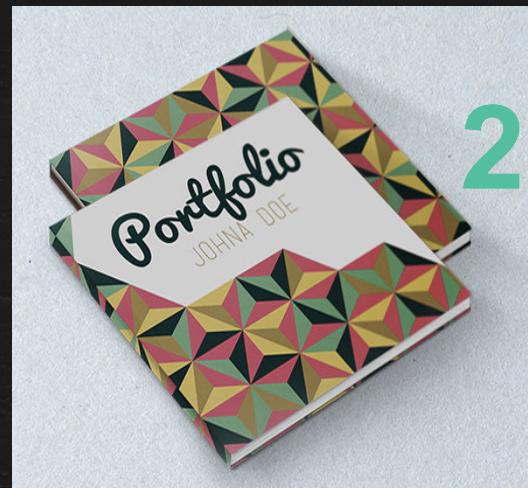
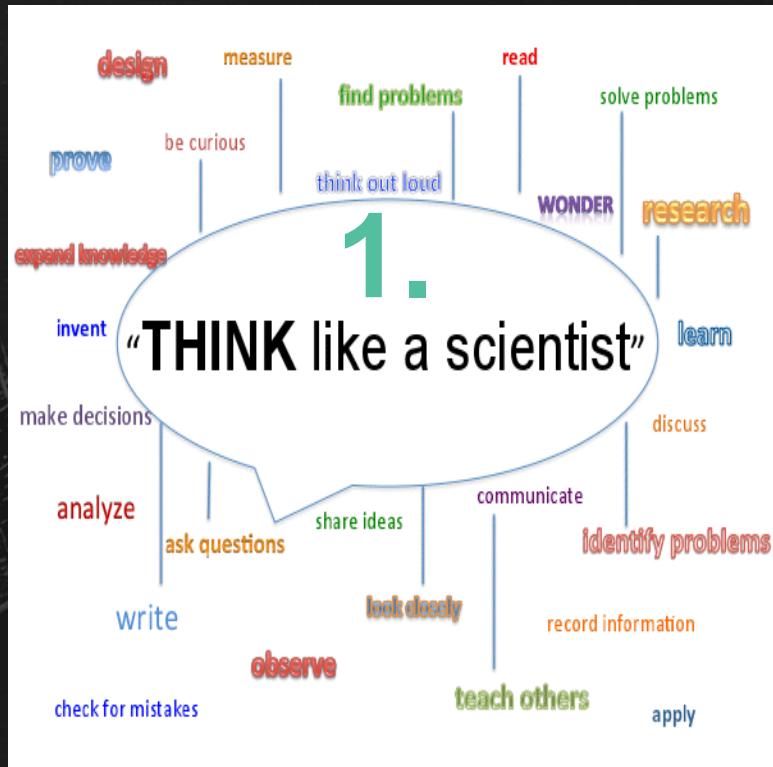
THE NEXT
BIG
THING

Source: <https://static1.squarespace.com/static/57fcf8a3d482e945c66724f7/l/585dfebcf7c506c9935130b/1482555000950/Excel+Dashboard+Report>



Radio Engineering Circle

DEVELOPMENT TRACK



2.

STATISTICS
DATA CLEAN-UP
A BIT OF LINEAR ALGEBRA
WORK WITH REAL WORLD PROBLEMS
BLOG

THE NEXT
BIG
THING

Source: <https://static1.squarespace.com/static/57fcf8a3d482e945c66724f7/l/585dfebcf7c506c9935130b/1482555000950/Excel+Dashboard+Report>



Radio Engineering Circle

DEVELOPMENT TRACK

Medium

rbahaguejr Nov 11, 2016 · 3 min read

How-to Declutter Your Data Science Workspace

Working on a data science project is almost always equivalent to an amazing clutter in the working directory. Data scientists would most likely have the following materials dumped in their project working directory:

Read more...

4

rbahaguejr Nov 6, 2016 · 4 min read

Exploring AirBnB's Knowledge Repo: A Curated Knowledge Sharing Platform

Why is it needed?

Data Science researches and studies should have the following aims (From Scaling Knowledge at Airbnb):

THE
BIG
THING

Source: <https://static1.squarespace.com/static/57fcf8a3d482e945c66724f7/l/585dfecfcf7c506c9935130b/148255500950/Excel+Dashboard+Report>

RickBahague

Overview Repositories 38 Stars 12 Followers 8 Following 2

NEP2017
Exploration and Analysis of 2017 NEP
Jupyter Notebook GPL-3.0 Updated on Sep 7, 2016

IPredict
Forked from tedgueniche/IPredict
Sequence Prediction Framework
Java 10 Updated on Aug 13, 2016

dspop
Data Science Popularization
Jupyter Notebook ★1 2 GPL-3.0 Updated on Aug 13, 2016

PHAutomatedElection2016
On the issue of lack of basic safeguard of PH's Automated Election System for 2016
Updated on May 29, 2016

PIIE
YOUNG ENGINEERS

Radio Engineering Circle

Questions



[GITHUB.COM/RICKBAHAGUE](https://github.com/rickbahague)



[MEDIUM.COM/@RBAHAGUEJR](https://medium.com/@rbaahaguejr)

THE NEXT
BIG
THING



Radio Engineering Circle