TBA2102 Introduction to Business Analytics

NUS | School of Computing
National University of Singapore
Leading The World With Asia's Best

Overview to Business Analytics

Growth in Big Data and Business Analytics (BDA)

Worldwide Business Analytics Services Spending[1]
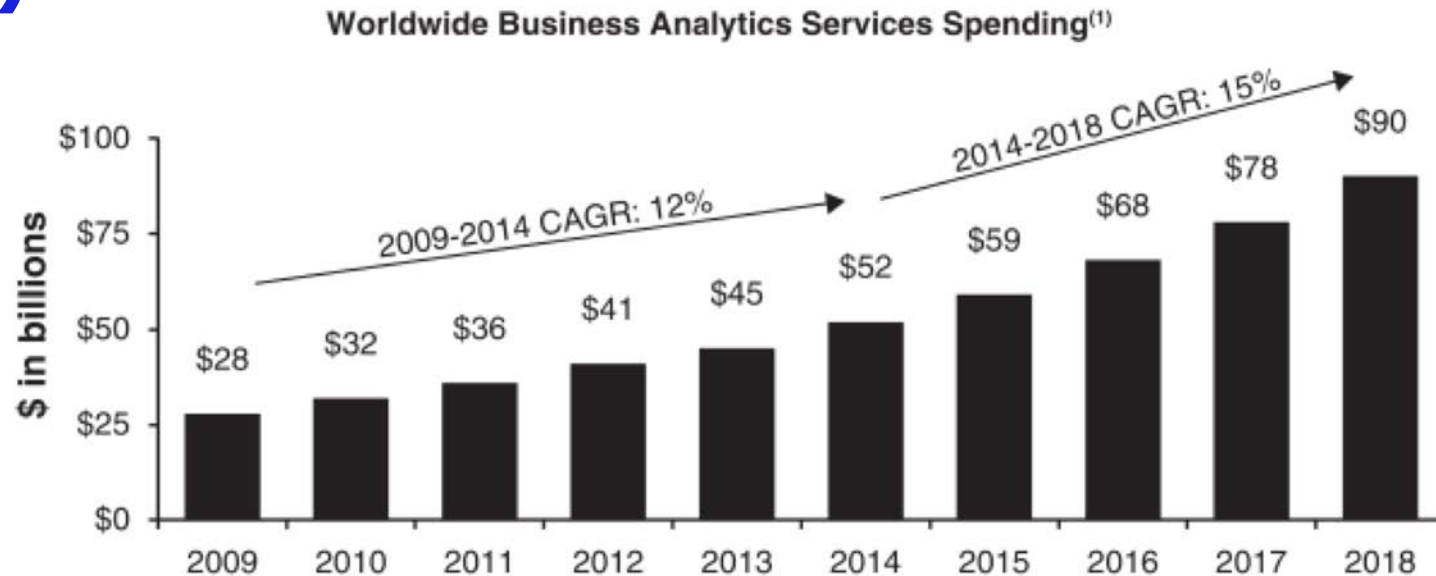
Revenue

Worldwide revenues for BDA
- **$150.8b** in **2017**
- **12.4% increase** over **2016.**
- will be more than **$210b by 2020**

Commercial purchases of BDA-related hw/sw/srvs

- **expected CAGR of 11.9% through 2020**

Industries with fastest growth in BDA spending

- **banking (13.3% CAGR)**
- **healthcare, insurance, securities & investment services, and telecommunications (12.8% CAGR each)**

2

Data Tsunami

We are generating more data than ever

Some big data facts

Over **50%** of US population owns a smartphone

10 billion mobile devices will be in use by 2020

294 billion emails sent every day

Over 1 billion Google searches every day

TRILLIONS of sensors monitor, track and communicate with each other, populating the Internet of Things with real-time data.

30+ petabytes of user-generated data stored, accessed and analyzed

230+ million Tweets each day

How much is much?

How Much Data is Produced Every Day?

2.5 Exabytes are are produced every day

Which is equivalent to:

- ♪ 530,000,000 millions songs
- 📱 150,000,000 iPhones
- 💻 5 million laptops
- 📚 250,000 Libraries of Congress
- ▶ 90 years of HD Video

Total amount of data in the world

4.4 zettabytes in 2013
~44 zettabytes in 2020

Kilobyte (1000bytes) ➤ Megabyte ➤ Gigabyte ➤ Terabyte ➤ Petabyte ➤ Exabyte ➤ Zettabyte ➤ Yottabyte ➤ Xenottabyte ➤ Shilentnobyte ➤ Domegemegrottebyte
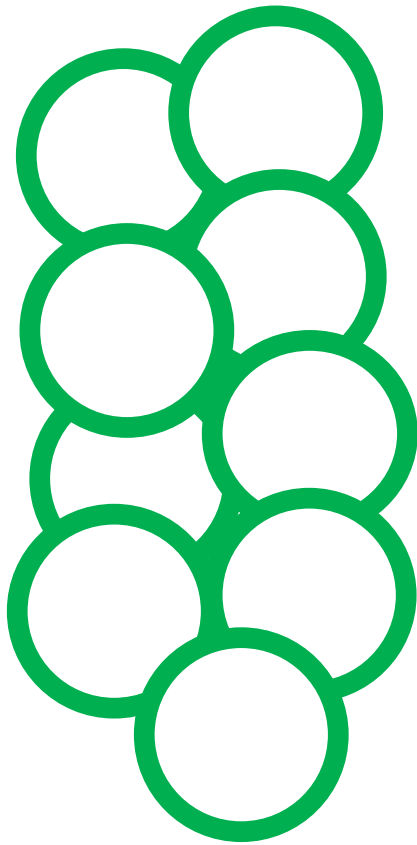
Yet, recent research has found that less than 0.5 percent of the data is actually being analyzed for operational decision making.

(Appdeveloper Magazine, 2016)
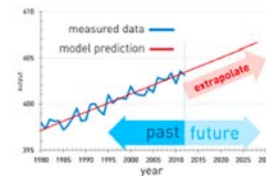
Why?

Flow of Data

DATA SOURCES

ANALYTICS

Data Cleansing

Analyze

Import Data

Export Data

Merge Data Sets

Verify & Enrich

Rebuild Missing Data

De-Duplicate

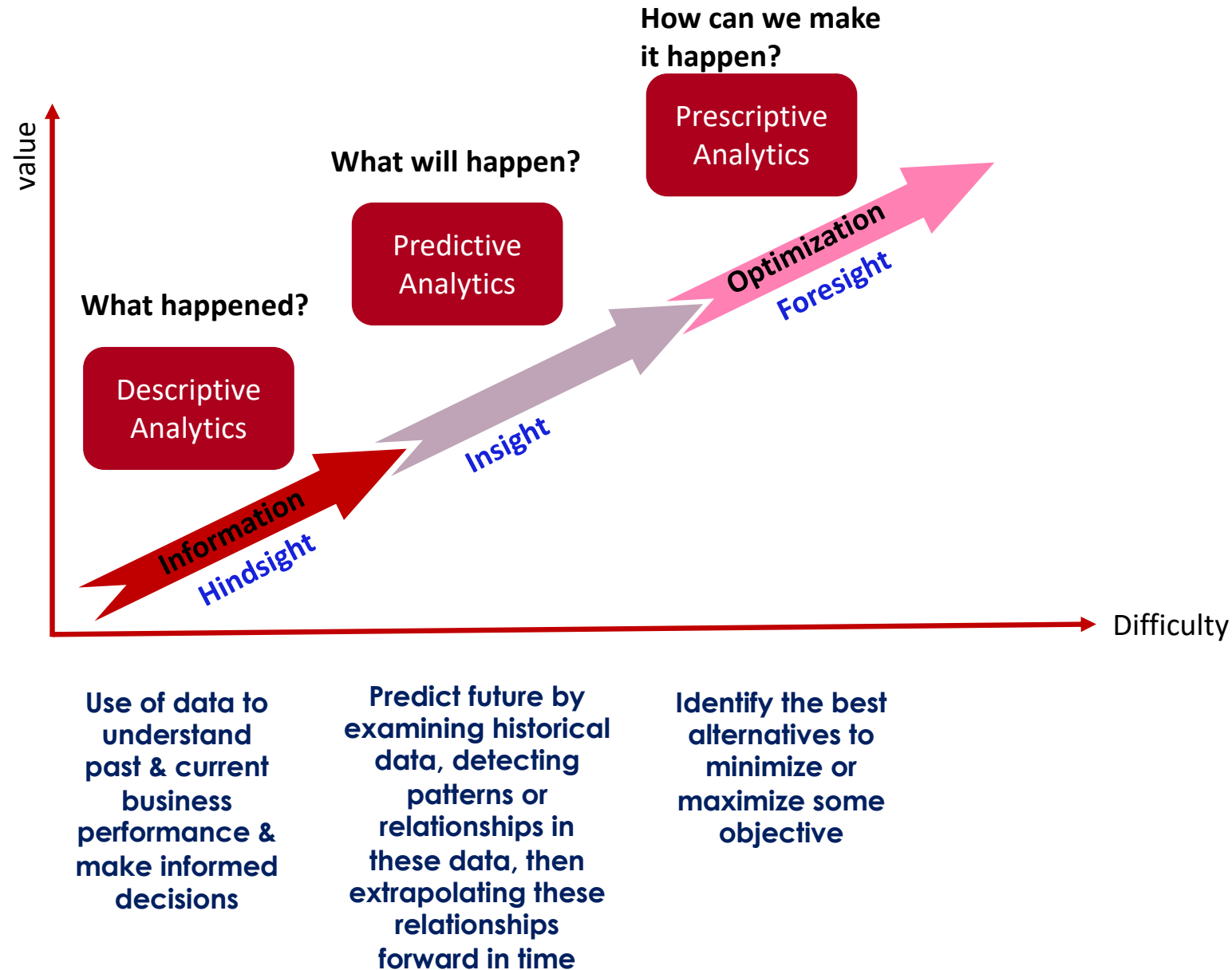Standardise Data

Normalise Data
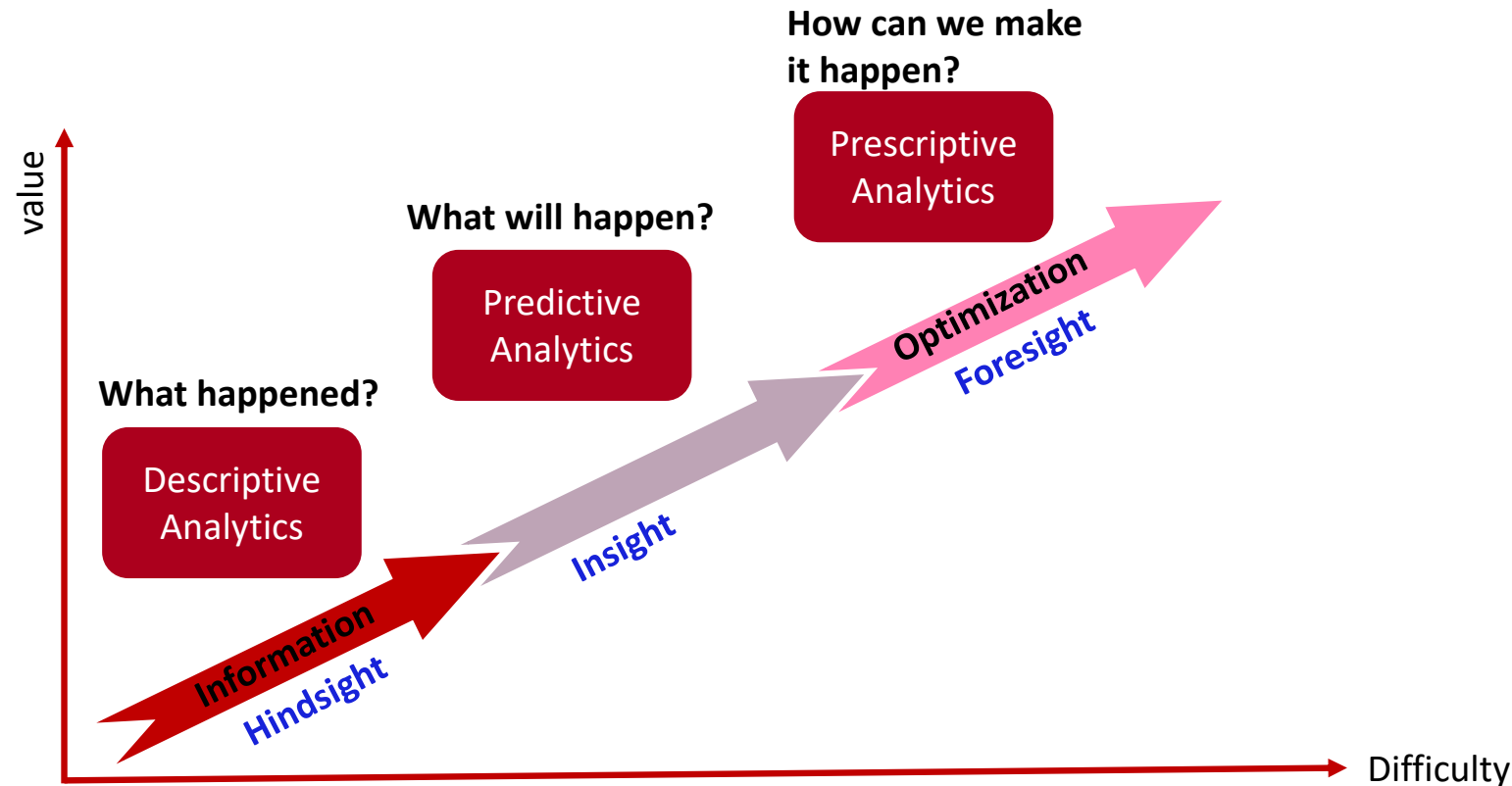
Business Analytics is the use of:

- data,
- information technology,
- statistical analysis,
- quantitative methods, and
- mathematical or computer-based models

to help managers gain improved insight about their business operations and make better, fact-based decisions.

Scope of Business Analytics

value

What happened?

Descriptive Analytics

What will happen?

Predictive Analytics

How can we make it happen?

Prescriptive Analytics

Information

Hindsight

Insight

Optimization

Foresight

Difficulty

Use of data to understand past & current business performance & make informed decisions

Predict future by examining historical data, detecting patterns or relationships in these data, then extrapolating these relationships forward in time

Identify the best alternatives to minimize or maximize some objective

EG: No show in hospital specialist clinics

How can we make it happen?

Prescriptive Analytics

What will happen?

Predictive Analytics

What happened?

Descriptive Analytics

value

Optimization

Foresight

Insight

Information

Hindsight

Difficulty

- **What is the demographic of patients who did not show up?**
- **What is the difference in no-show rate across the year? (any seasonal effect?)**

- **Which patients are unlikely to show up tomorrow?**

- **How many patient visit slots should be allocated on a given day?**

Processes involved in using Analytics for problem solving

| 1. Recognizing a problem | 2. Defining the problem | 3. Structuring the problem | 4. Analyzing the problem | 5. Interpreting results & making a decision | 6. Implementing the solution |
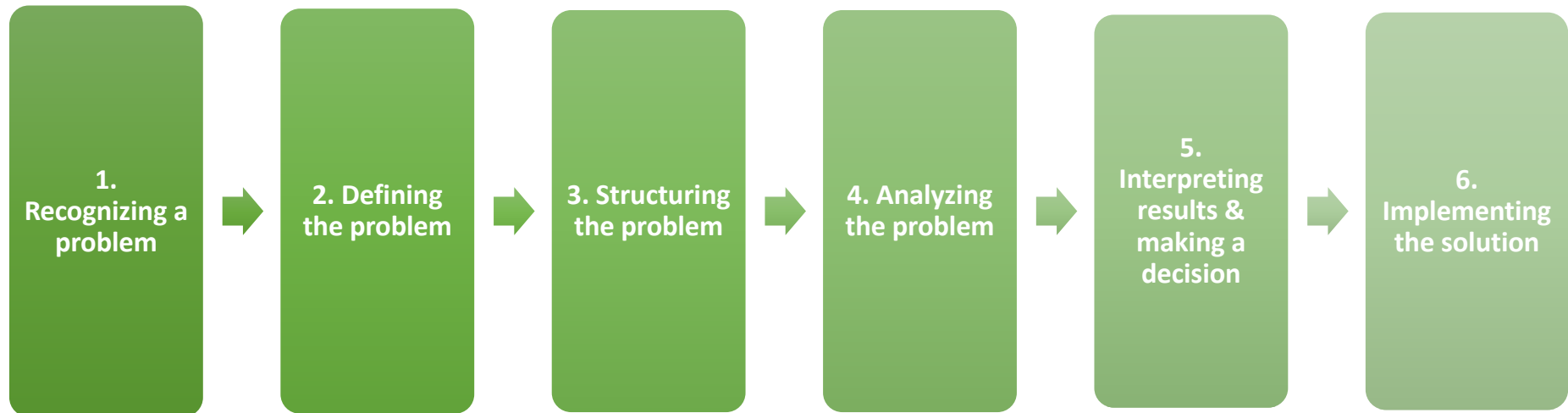
Data for Business Analytics

▶ **Data: numerical or textual facts and figures that are collected through some type of measurement process.**

▶ **Information: result of analyzing data; that is, extracting meaning from data to support evaluation and decision making.**

Data Sets and Databases

- **Data set - a collection of data.**
  - **Examples: Marketing survey responses, a table of historical stock prices, and a collection of measurements of dimensions of a manufactured item.**

- **Database - a collection of related files containing records on people, places, or things.**
  - **A database file is usually organized in a two-dimensional table, where the columns correspond to each individual element of data (called *fields*, or *attributes*), and the rows represent records of related data elements.**

Eg of a Sales Transaction Database File

	A	B	C	D	E	F	G	H
1	Sales Transactions: July 14							
2								
3	Cust ID	Region	Payment	Transaction Code	Source	Amount	Product	Time Of Day
4	10001	East	Paypal	93816545	Web	$20.19	DVD	22:19
5	10002	West	Credit	74083490	Web	$17.85	DVD	13:27
6	10003	North	Credit	64942368	Web	$23.98	DVD	14:27
7	10004	West	Paypal	70560957	Email	$23.51	Book	15:38
8	10005	South	Credit	35208817	Web	$15.33	Book	15:21
9	10006	West	Paypal	20978903	Email	$17.30	DVD	13:11
10	10007	East	Credit	80103311	Web	$177.72	Book	21:59
11	10008	West	Credit	14132683	Web	$21.76	Book	4:04
12	10009	West	Paypal	40128225	Web	$15.92	DVD	19:35
13	10010	South	Paypal	49073721	Web	$23.39	DVD	13:26

Records

Entities

Fields or Attributes

13

Structured vs Unstructured Data

Customer reviews

★★★★☆ 2,729

3.9 out of 5 stars ▾

5 star	▬▬	46%
4 star	▬	22%
3 star	▬	14%
2 star	▪	10%
1 star	▪	8%

See all verified purchase reviews ›
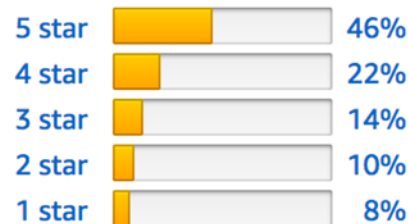
Share your thoughts with other customers

Write a customer review

Top customer reviews

★★★★★ **Wrath of Khan bumped down to no 2 best Star Trek films**

By Kindle Customer on July 23, 2016
Format: Amazon Video

What movie were the other reviewers watching? I'm a 47 year old super Trek fan and I was honestly worried about the franchise after Into Darkness. Pegg, Jung, Lin and the rest of the team knocked it out of the park. This was a fantastic, original, story. Idris Elba was phenomenal and the writers gave Karl Urban's Bones a solid role rather than only the platitudes that were a fraction of what made up Deforest Kelly's Bones. Roddenberry's vision of humanity and coexisting with those different than us was completely restored, and then some. In addition the special effects were not only awesome but original, particularly the Yorktown space station. This movie was well over 2 hrs but felt like 5 minutes. I hope these writers and director come back for another one.
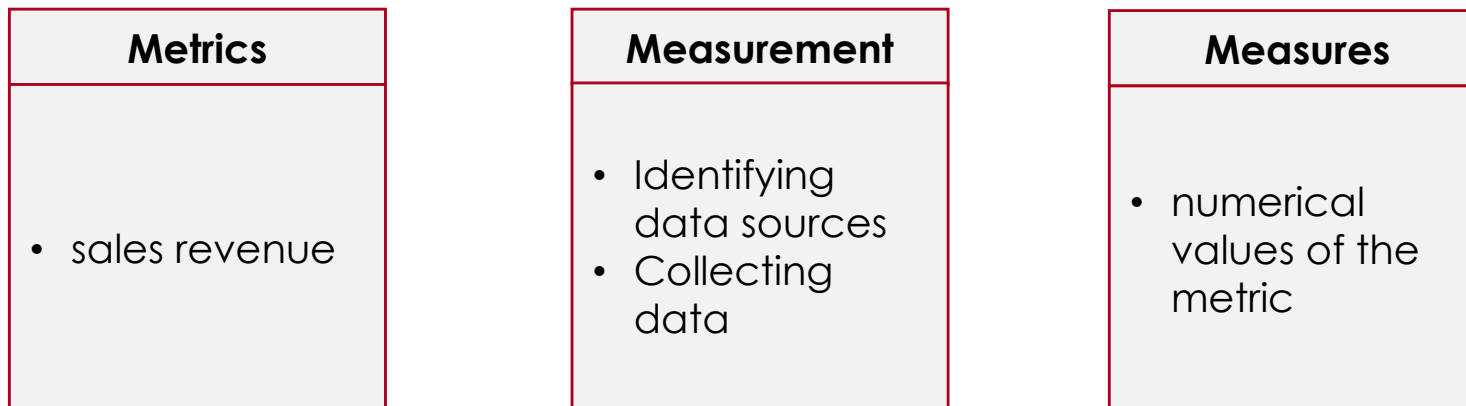
Big Data

- Massive amounts of business data from a wide variety of sources, much of which is available in real time, and much of which is uncertain or unpredictable.

- 4 Vs: volume, variety, velocity, and veracity

- Size is a "moving" target, depending on available technological capabilities

Metrics

▶ **Metric** - **a unit of measurement that provides a way to objectively quantify performance.**

▶ **Measurement** - **act of obtaining data associated with a metric.**

▶ **Measures** - **numerical values associated with a metric.**

Eg: Let's assess the performance of a retail store.

Metrics	Measurement	Measures
• sales revenue	• Identifying data sources • Collecting data	• numerical values of the metric

Types of Metrics

- **Discrete metric** - one that is derived from counting something.
  - Eg: a delivery is either on time or not; an order is complete or incomplete; or an invoice can have one, two, three, or any number of errors.
  - Some discrete metrics would be the proportion of on-time deliveries; the number of incomplete orders each day, and the number of errors per invoice.

- **Continuous metric** - based on a continuous scale of measurement.
  - Eg: involving dollars, length, time, volume, weight

Data Classification by Measurement Scales

▸ **Categorical (nominal) data** - **sorted into categories according to specified characteristics.**

▸ **Ordinal data** - **can be ordered or ranked according to some relationship to one another.**

▸ **Interval data** - **ordinal but have constant differences between observations and have arbitrary zero points.**

▸ **Ratio data** - **continuous and have a natural zero.**

EG: Classifying Data Elements

	A	B	C	D	E	F	G	H	I	J
1	Purchase Orders									
2										
3	Supplier	Order No.	Item No.	Item Description	Item Cost	Quantity	Cost per order	A/P Terms (Months)	Order Date	Arrival Date
4	Hulkey Fasteners	Aug11001	1122	Airframe fasteners	$ 4.25	19,500	$ 82,875.00	30	08/05/11	08/13/11
5	Alum Sheeting	Aug11002	1243	Airframe fasteners	$ 4.25	10,000	$ 42,500.00	30	08/08/11	08/14/11
6	Fast-Tie Aerospace	Aug11003	5462	Shielded Cable/ft.	$ 1.05	23,000	$ 24,150.00	30	08/10/11	08/15/11
7	Fast-Tie Aerospace	Aug11004	5462	Shielded Cable/ft.	$ 1.05	21,500	$ 22,575.00	30	08/15/11	08/22/11
8	Steelpin Inc.	Aug11005	5319	Shielded Cable/ft.	$ 1.10	17,500	$ 19,250.00	30	08/20/11	08/31/11
9	Fast-Tie Aerospace	Aug11006	5462	Shielded Cable/ft.	$ 1.05	22,500	$ 23,625.00	30	08/20/11	08/26/11
10	Steelpin Inc.	Aug11007	4312	Bolt-nut package	$ 3.75	4,250	$ 15,937.50	30	08/25/11	09/01/11
11	Durrable Products	Aug11008	7258	Pressure Gauge	$ 90.00	100	$ 9,000.00	45	08/25/11	08/28/11
12	Fast-Tie Aerospace	Aug11009	6321	O-Ring	$ 2.45	1,300	$ 3,185.00	30	08/25/11	09/04/11
13	Fast-Tie Aerospace	Aug11010	5462	Shielded Cable/ft.	$ 1.05	22,500	$ 23,625.00	30	08/25/11	09/02/11
14	Steelpin Inc.	Aug11011	5319	Shielded Cable/ft.	$ 1.10	18,100	$ 19,910.00	30	08/25/11	09/05/11
15	Hulkey Fasteners	Aug11012	3166	Electrical Connector	$ 1.25	5,600	$ 7,000.00	30	08/25/11	08/29/11

Categorical   Ordinal   Categorical   Categorical   Ratio   Ratio   Ratio   Ratio   Interval   Interval

19

Data Reliability and Validity

	Reliability (data is accurate and consistent)	Validity (data correctly measures what is it supposed to measure
A tire pressure gage that consistently reads several pounds of pressure below the true value	NO	YES
Number of calls to customer service desk (counted correctly) used to assess customer dissatisfaction	YES	NO
Customer rating on food quality is used to assess customer satisfaction	NO	NO