



Ingeniería Industrial

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

INTRODUCTION TO BIG DATA

Juan D. Velásquez
Felipe E. Vildoso Castillo



Ingeniería Industrial

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

CLASS 0

Juan D. Velásquez
Felipe E. Vildoso Castillo

COURSE OBJECTIVE

To introduce and give knowledge related to Big Data technologies to Industrial Engineering students in order that they can think new solutions to problems that may arise in all kinds of organizations.

ACTIVITIES

Three weekly hours of lectures.

One and a half weekly hours of assistantships.

Personal Assignments: three Tests (60%) and one Exam (40%).

3 Teamwork Assignments (2 presentation).

One U-test per each lecture.

Final Assessment Grade and Final Teamwork Grade must separately be over 4.0.

Final Course Grade: 50% Personal Assignments + 50% Teamwork Assignments.

SCHEDULE

Test 1 April 11th 2017

Test 2 May 9th 2017

Test 3 June 20th 2017

Teamwork 1 April 16th 2017 (20%)

Teamwork 2 May 14th 2017 (40%)

- Presentation May 23th 2017

Teamwork 3 June 27th 2017 (40%)

- Presentation June 27th 2017

SUGGESTIONS OR QUESTIONS

If you have any suggestion or question don't hesitate to contact any member of the teaching stuff.

BIBLIOGRAPHY

Nathan Marz y James Warren. “Big Data: Principles and best practices of scalable realtime data systems”. Manning 2015.

Foster Provost y Tom Fawcell. “Data Science for Business: What you need to know about data mining and data-analytic thinking”. 2013

Bernard Marr. “Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance”. 2015

Tom White. “Hadoop: The Definitive Guide”. O’Reilly 2015

Mat Brown. “Learning Apache Cassandra”. 2015

Corey L. Lanum “Visualizing Graph Data” 2016

Elijah Meeks “D3.js” 2015

Henry Brink, Joseph W. Richards, Mark Fetherolf “Real World Machine Learning” 2016



Ingeniería Industrial

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

*WHAT DO YOU THINK BIG
DATA IS?*

ACCORDING TO IBM

What is Big Data?



Big data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume and variety. To extract meaningful value from big data, you need optimal processing power, analytics capabilities and skills.

ACCORDING TO IBM



Competitive advantage

Data is emerging as the world's newest resource for competitive advantage.



Decision making

Decision making is moving from the elite few to the empowered many.



Value of data

As the value of data continues to grow, current systems won't keep pace.

ACCORDING TO IBM

4.4MILLION
data scientists
needed by 2015



UNSTRUCTURED DATA

Almost 80% of the Big Data is Unstructured Data.

Structured Data



0.103	0.176	0.387	0.300	0.379
0.333	0.384	0.564	0.587	0.857
0.421	0.309	0.654	0.729	0.228
0.266	0.750	1.056	0.936	0.911
0.225	0.326	0.643	0.337	0.721
0.187	0.586	0.529	0.340	0.829
0.153	0.485	0.560	0.428	0.628

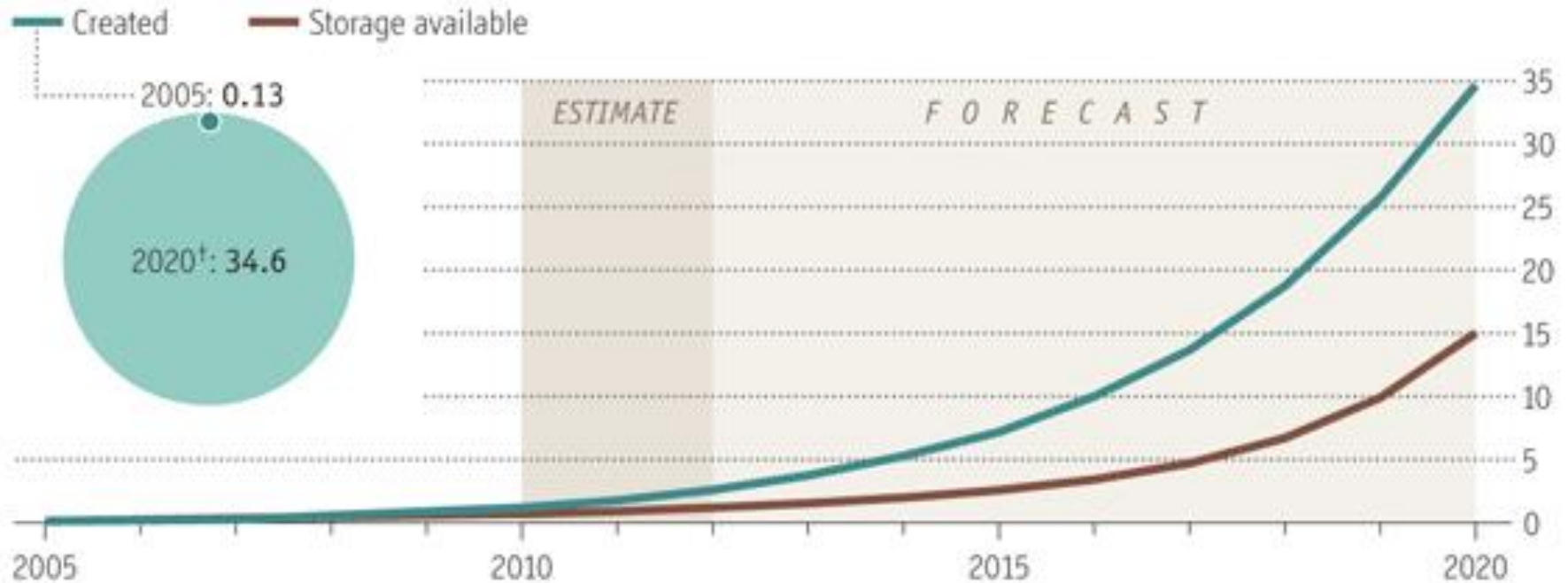
Unstructured Data



THERE ARE SO MUCH DATA!

Open wide

Global digital information
Zettabytes*



Source: IDC

*1 zettabyte=1 trillion gigabytes ¹Forecast

MULTIPLES OF BYTES

Decimal		
Value	Metric	
1000	kB	kilobyte
1000 ²	MB	megabyte
1000 ³	GB	gigabyte
1000 ⁴	TB	terabyte
1000 ⁵	PB	petabyte
1000 ⁶	EB	exabyte
1000 ⁷	ZB	zettabyte
1000 ⁸	YB	yottabyte

ONE ZETTABYTE

1.000.000.000 x



IN OTHER WORDS...

Big Data means data sets that are too large for traditional data processing systems.

Using Big Data technologies is associated with additional productivity growth.

<https://www.youtube.com/watch?v=ahZGEusG13A>

QUESTIONS?
SEE YOU ON THE NEXT CLASS!