



UNIVERSIDAD CATÓLICA
de Colombia

Uso de Herramientas Tecnológicas para Data Analytics y Big Data

Diego Alberto Rincón Yáñez MSc
darincon@ucatolica.edu.co

Afiliada a la Federación Internacional de Universidades Católicas (FIUC)
w w w . u c a t o l i c a . e d u . c o



Palabras de Koheleth Hijo de David, Rey en Jerusalem ~ 200 D.C.

*“No hay nada nuevo
bajo el Sol!”*

Ecclesiastes Chapter 1 verso 9

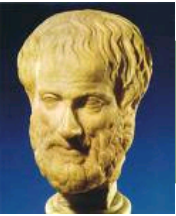




E-Ciencia



La Ciencia Evolucionaria



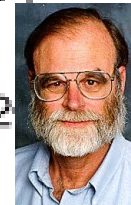
Empirical



Theory



Simulation



Data

$$y = e^x$$

<0

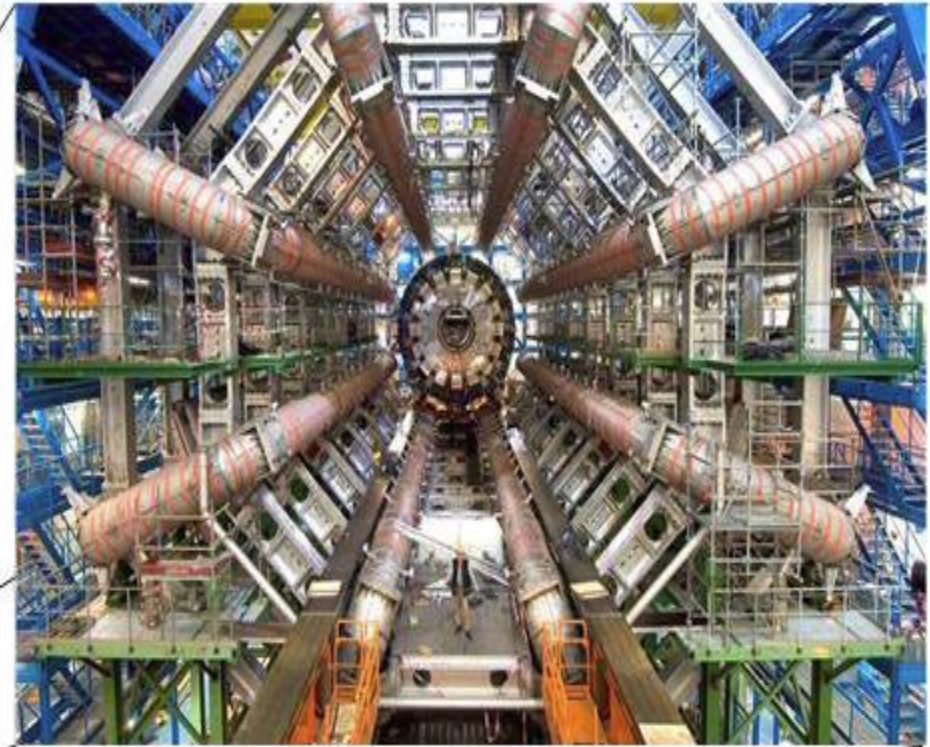
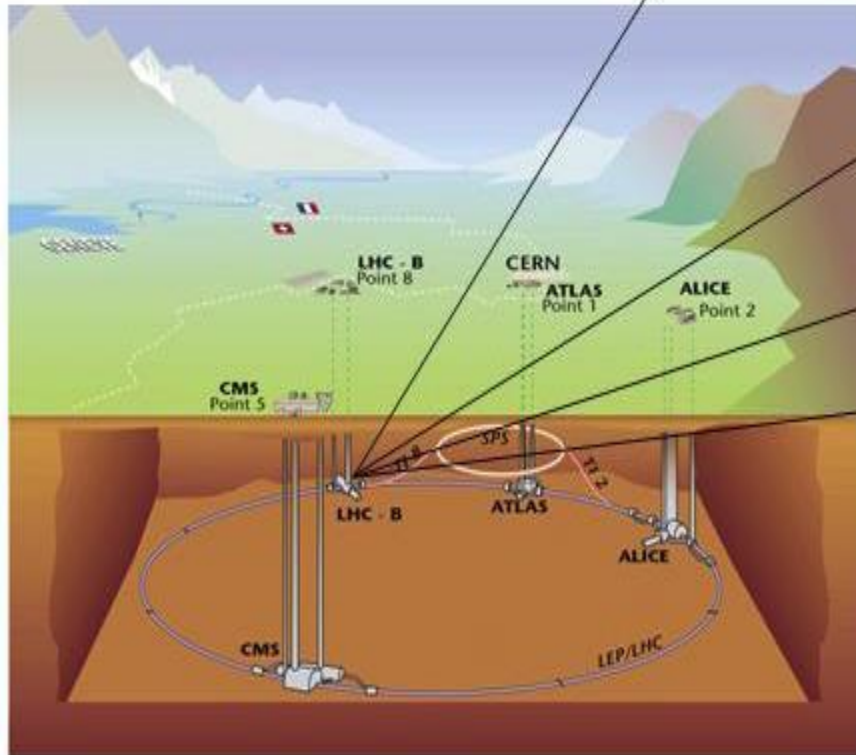
1700

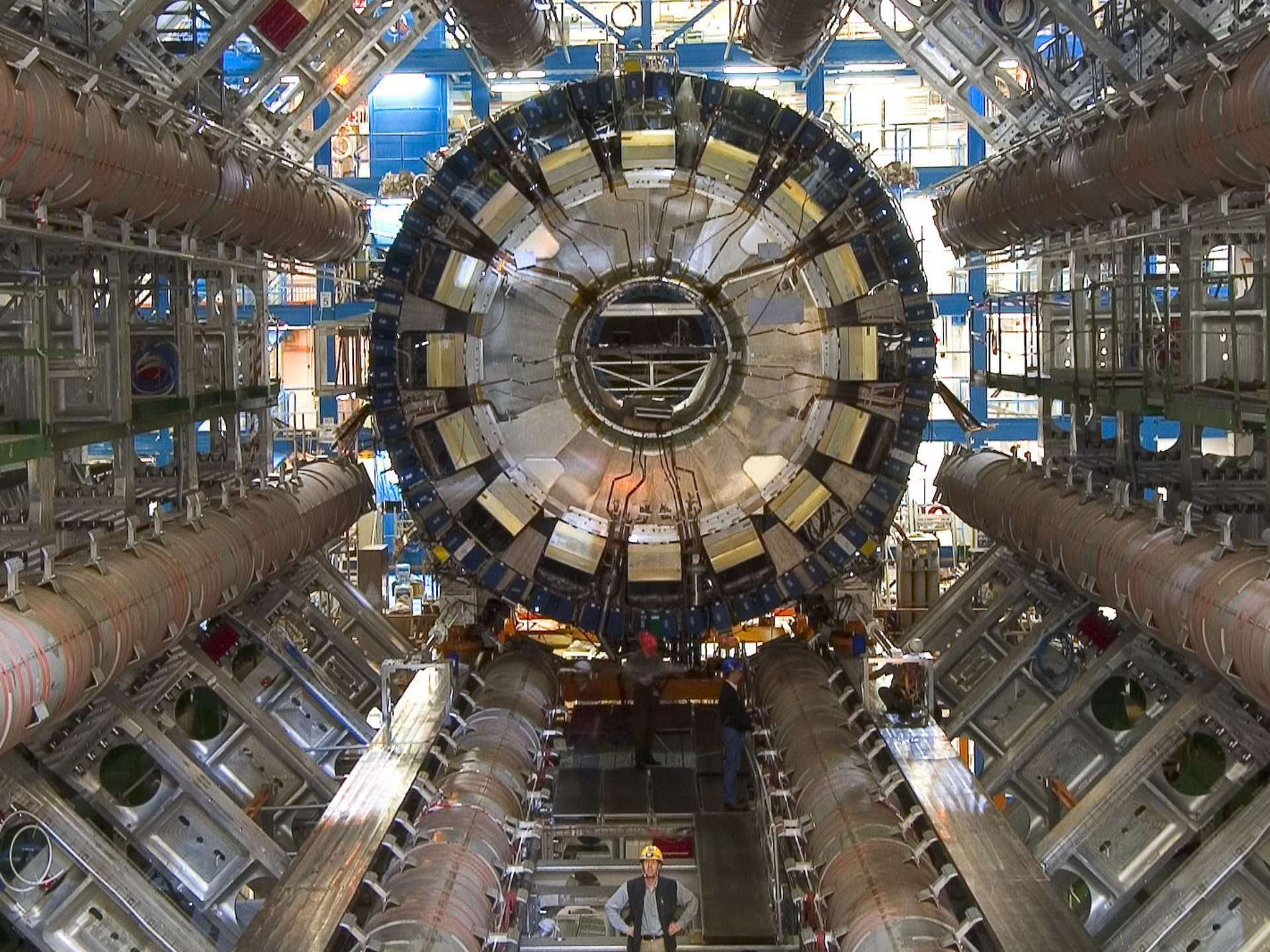
1950

1990

x

Y se hace GRANDE







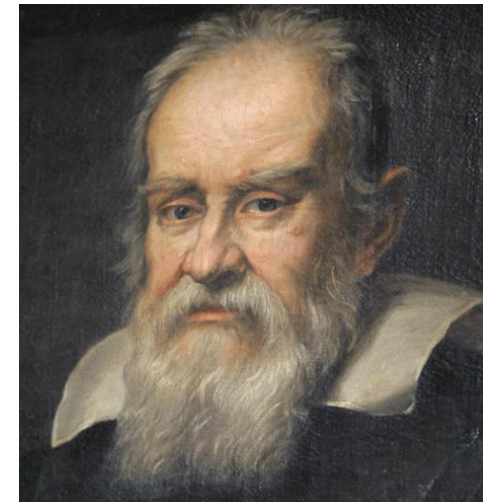
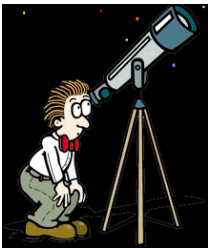
Pero nos quedamos CORTOS !!



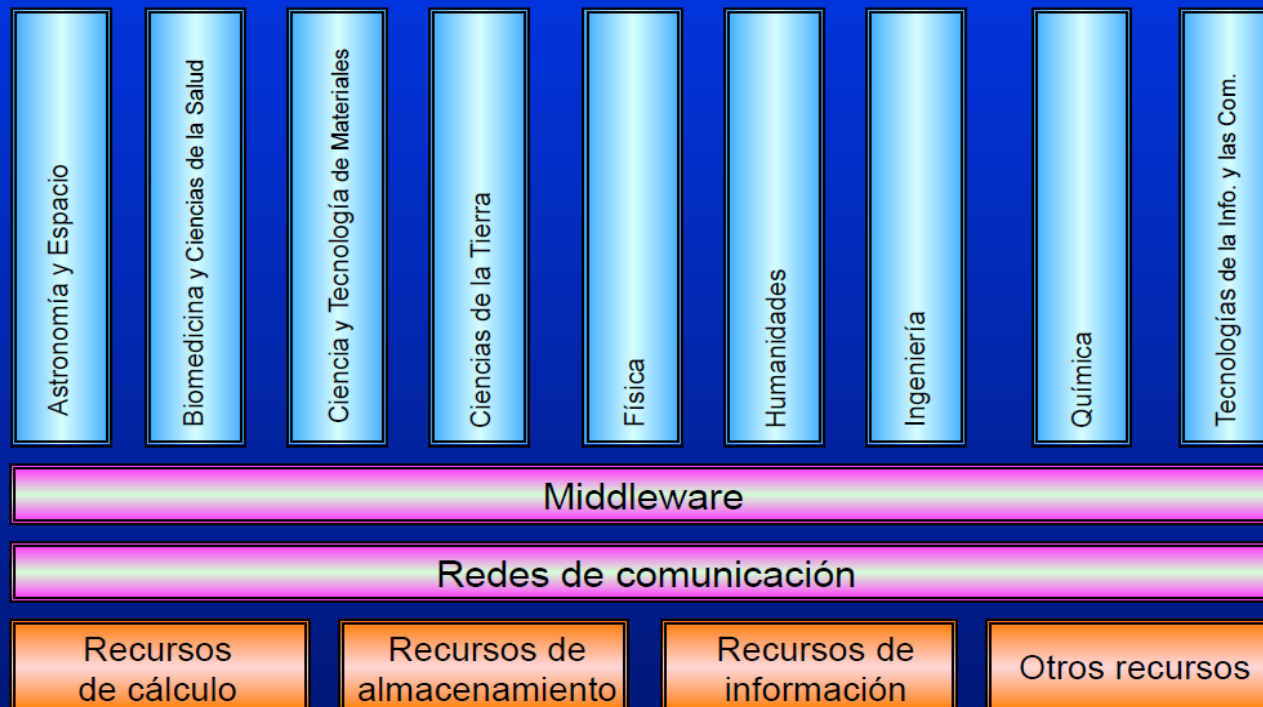
E-Ciencia

"e-Science is about global collaboration in key areas of science and the next generation of infrastructure that will enable it." - Dr John Taylor, Director General of Research Councils

"The term "e-Science" denotes the systematic development of research methods that exploit advanced computational thinking" - Professor Malcolm Atkinson, e-Science Envoy.



Taxonomía de la e-Ciencia



Víctor Castelo Gutiérrez

e-Infraestructuras en Europa



Big Data & Data Science



¿De donde la moda del Big Data?

Forbes
com

"big data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year..."

FT FINANCIAL TIMES
World business newspaper

"Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to 'understand where customers are, what makes them tick and what they want', says Deepak Advani, who heads IBM's predictive analytics group."

The New York Times

"At the World Economic Forum last month in Davos, Switzerland, Big Data was a marquee topic. A report by the forum, 'big data, Big Impact,' declared data a new class of economic asset, like currency or gold."



*"Data is the new oil."
Clive Humby*

THE WALL STREET JOURNAL

"Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers."

Forbes
com

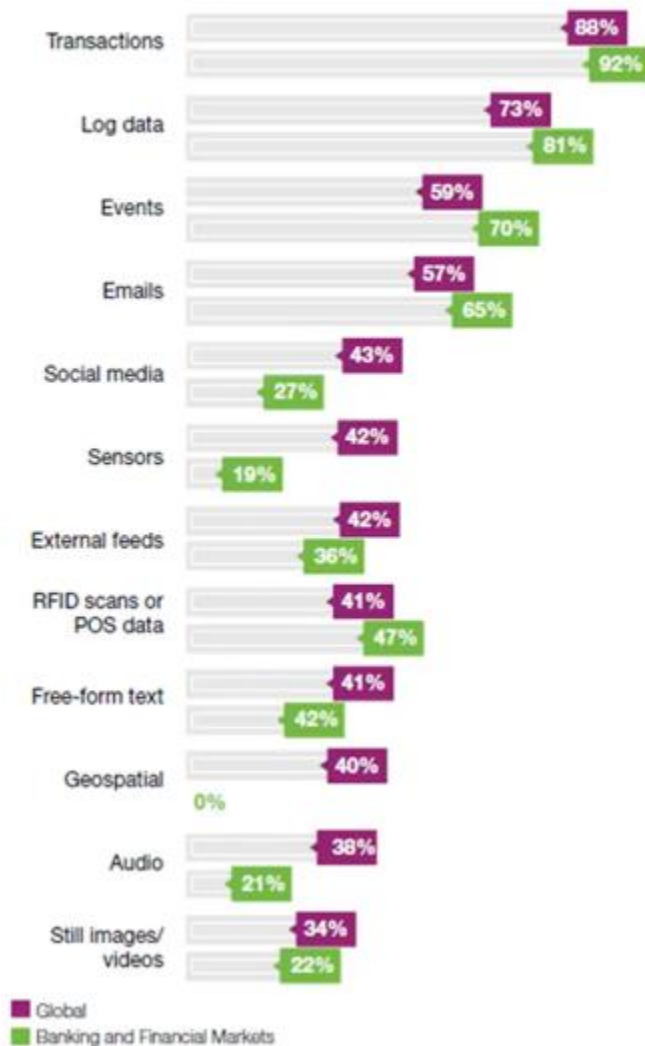
"...now Watson is being put to work digesting millions of pages of research, incorporating the best clinical practices and monitoring the outcomes to assist physicians in treating cancer patients."

¿Fuentes de Datos?

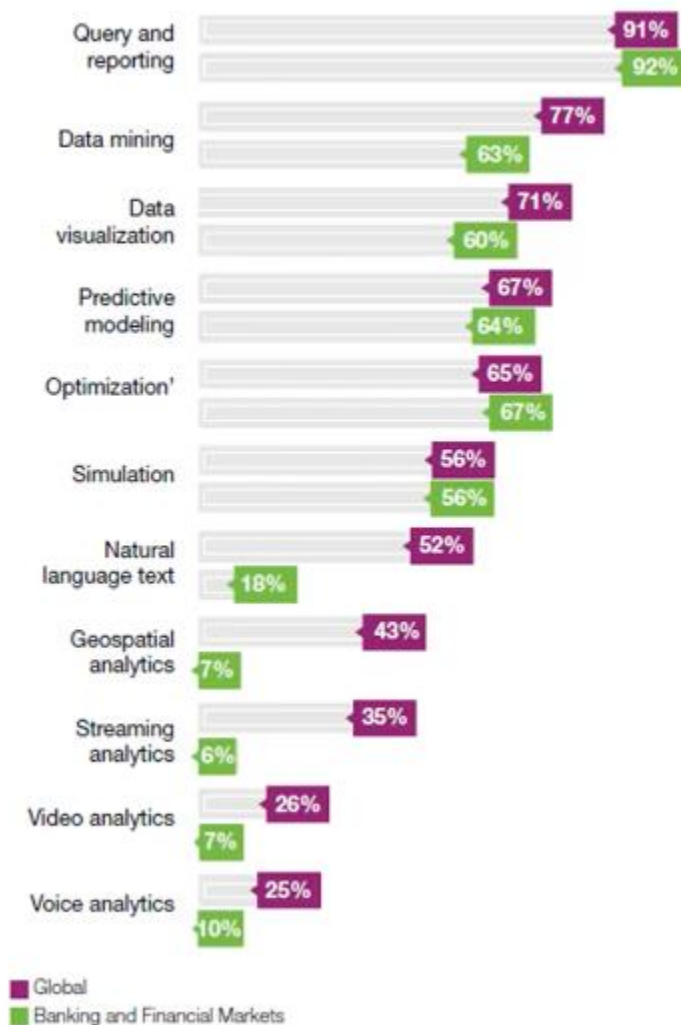


UNIVERSIDAD CATÓLICA
de Colombia

Big data sources



Analytics capabilities



Big Data is Not Only Hadoop!

- **Examples where Hadoop is not entirely applicable:**
 - **Cyber security, Stock market, Traffic control, Sensor information, monitoring trends in Social Media**
 - **What if your company has many silos of information, difficult to move to HDFS?**
 - **What about governance? Can we trust the source of this data?**



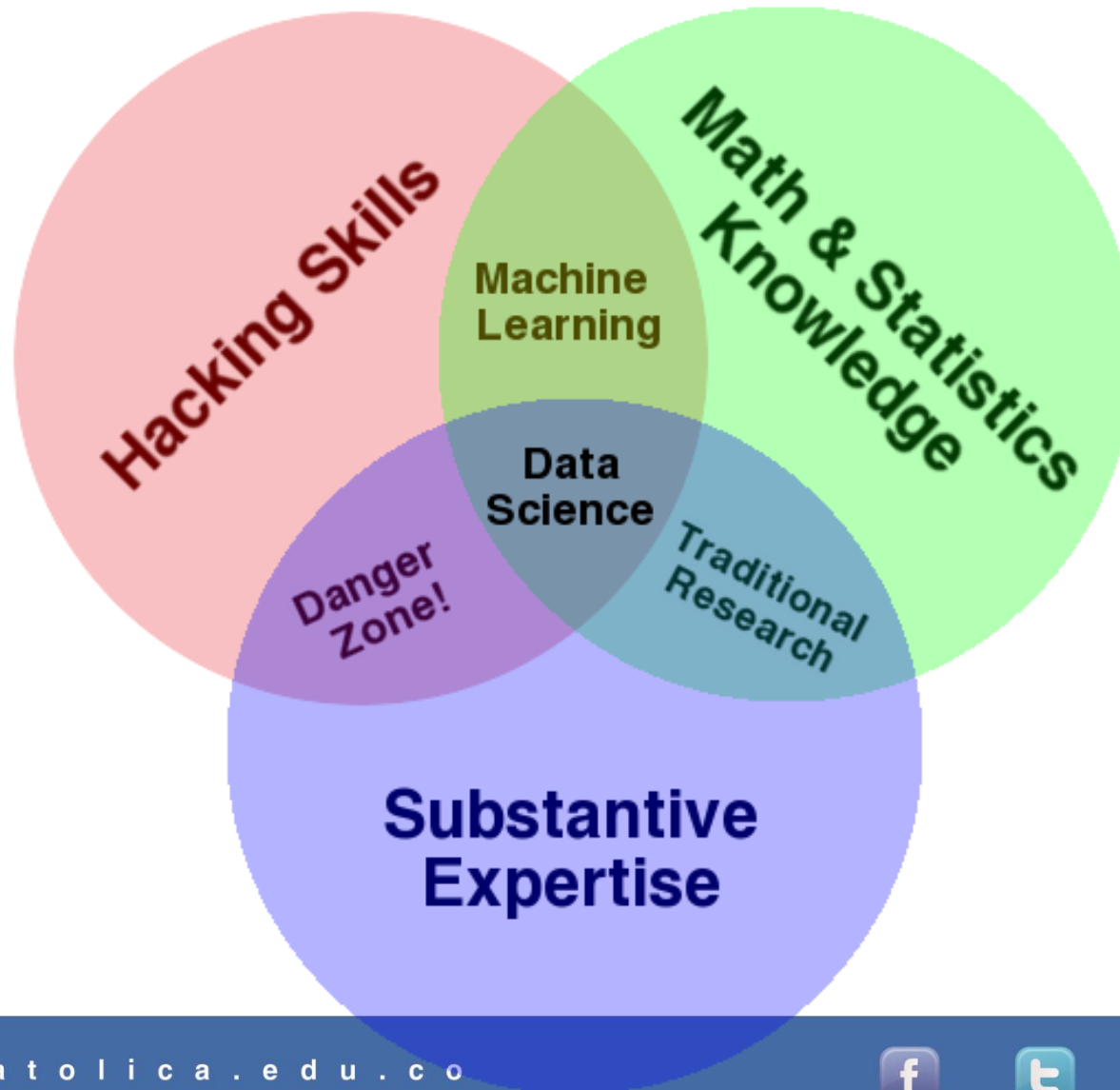
¿De donde la moda del Big Data?

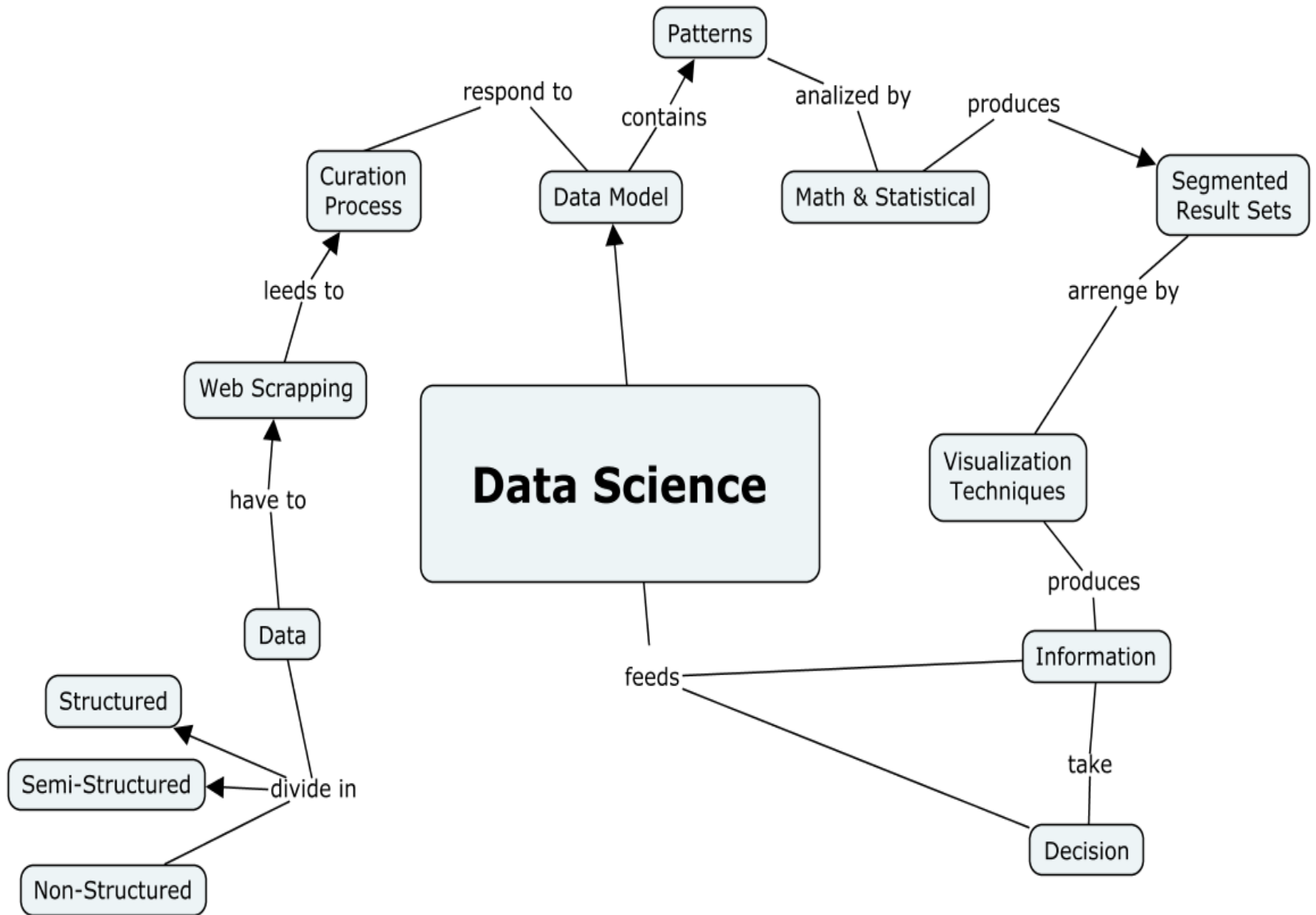
*“Big Data is like teenage sex;
everyone talks about it, nobody
really knows how to do it, everyone
thinks everyone else is doing it, so
everyone claims they are doing it.”*

Dan Ariely, Duke University Director of the Center for Advanced
Hindsight



¿Data Science?







Data Scientist


A data scientist is someone who is better at statistics than any software engineer and better at software engineering than any statistician.

Math



☐ Data cleaning and munging

DataWrangler^{alpha} Jupyter^{alpha}



 data.table dplyr




Stats



☐ Data visualization

 python

ggvis 

 vega



 mongoDB

ML



☐ Reporting

 + a b l e a u

 Spotfire[®]
TIBCO Software

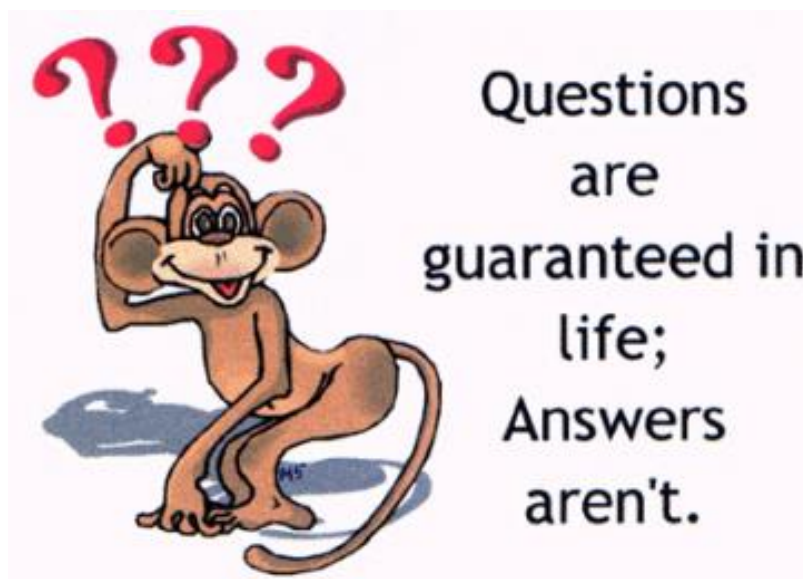
 R Markdown

MapReduce



Spark[★]

 hadoop



¿Preguntas?

Diego Alberto Rincón Yáñez MCSc.
Twitter: @d1egoprogram