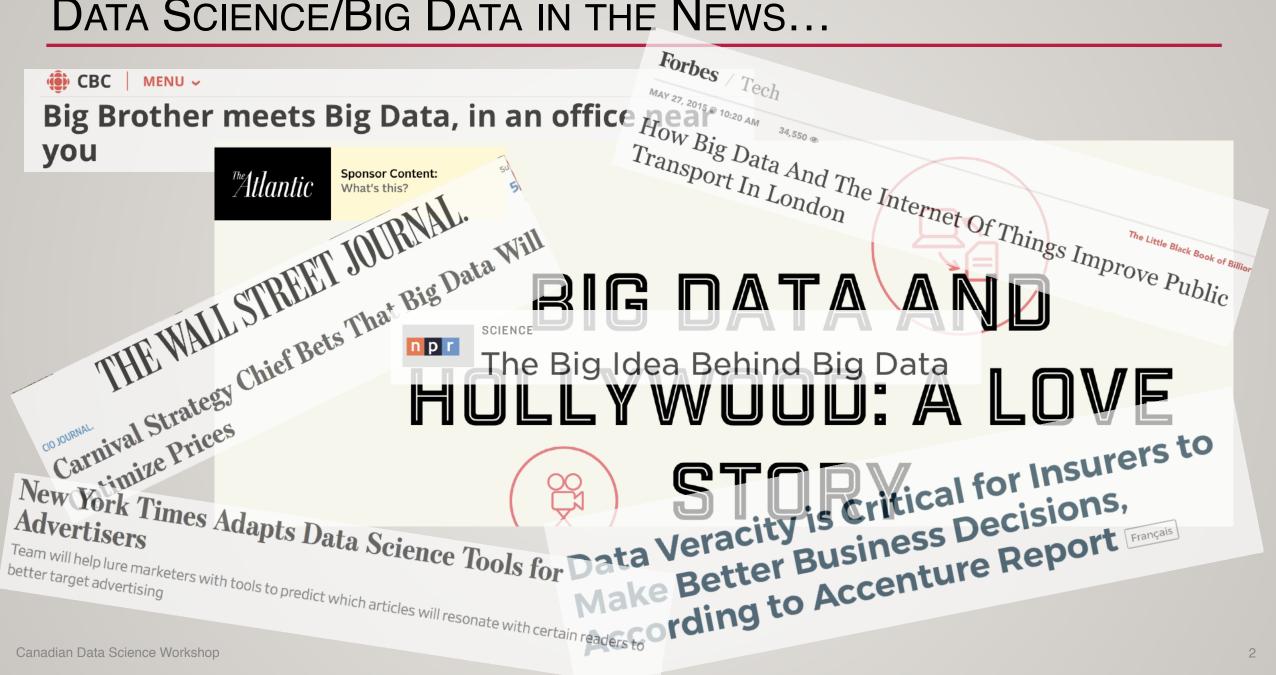
DATA SCIENCE ECOSYSTEM

M. TAMER ÖZSUNANCY REIDRAYMOND NGU. WATERLOOU. TORONTOUBC

DATA SCIENCE/BIG DATA IN THE NEWS...

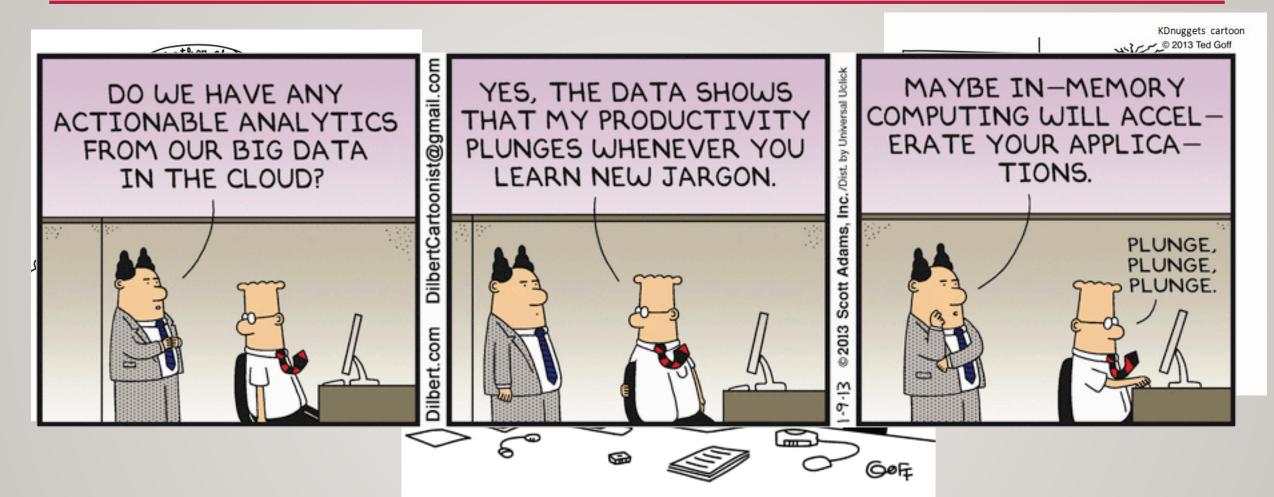


DATA SCIENCE EVERYWHERE!...



"You can't keep adjusting the data to prove that you would be the best Valentine's date for Scarlett Johansson."

DATA SCIENCE EVERYWHERE!...



"You can't keep adjusting the data to prove that you would be the best Valentine's date for Scarlett Johansson."

DATA SCIENCE EVERYWHERE!...



Valentine's date for Scarlett Johansson."

DATA SCIENCE VOCABULARY



 "Data science, also known as data-driven science, is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms, either structured or unstructured, similar to data mining."



WIKIPEDIA The Free Encyclopedia

- "Data science, also known as data-driven science, is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms, either structured or unstructured, similar to data mining."
- "Data science intends to analyze and understand actual phenomena with 'data'. In other words, the aim of data science is to reveal the features or the hidden structure of complicated natural, human, and social phenomena with data from a different point of view from the established or traditional theory and method."



WIKIPEDIA The Free Encyclopedia

firs in Canolication, Data Analysis



Fundamental Concepts and a Heuristic Exampl
Châie Hymbi

The Institute of Statistical Mathematics Sakuragaoka, Birijian 304 15-8 Sakuragaoka, Shibuya-ku Tokyo 150, Japan

What is Data Science ?*

Summary: Data Science is not only a synthetic concept to unify statistic, data analysis and their related methods but also comprises in results. It includes three planes, design for static, collection of data, and analysis on data. Fundamental concepts and various methods used on stare discussed with a heurostic example.

Statistics and data analysis have developed in their mains segmently and contributed to the development of scattering, theving their uncoper poperties. This does and various methods of statistics score-test (and developed energy for extension) that is and on the statistics in a score-test (and developed energy for extension). Multi-brancel Biometers, the development of mathematical statistics, has develop interf ways to the Biometers, the development of mathematical interaction, and develop interf ways to the statistics and development of mathematical interactions.

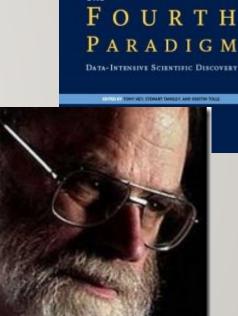
Consider the rest and there multiline of the rest of the multilinear states of the rest of the rest

* The roundtable discussion *Perspectives in classification and the Future of IPCS* was beid at the last Conference under the chairmanship of Professor H. -H. Bock. In this part discussion, I used the phase 'Data Science'. There was a question, "What is 'Data Science?" > Underly answered it. This is the training post of the present paper.

Canadian Data Science Workshop

- Fourth paradigm
 - "... change of all sciences moving from observational, to theoretical, to computational and now to the 4th Paradigm – Data-Intensive Scientific Discovery"

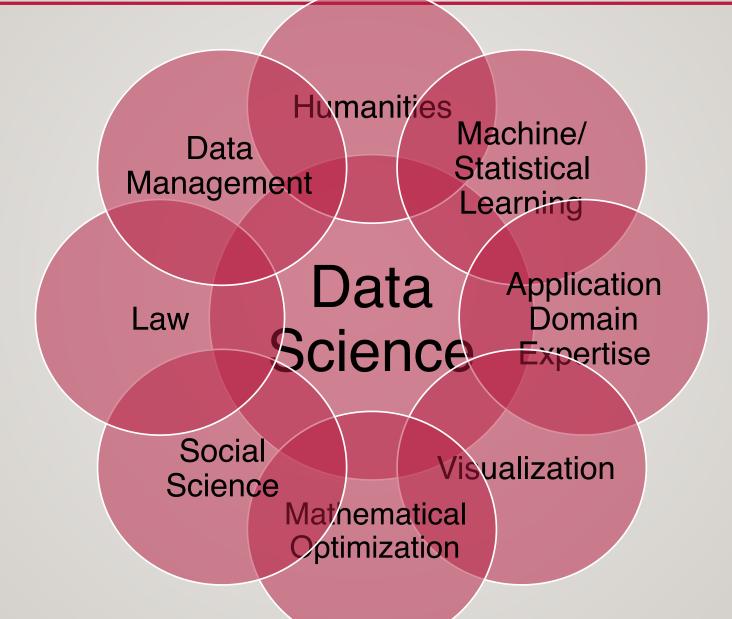




The

Need to solve a real problem using data... No applications, no data science.

DATA SCIENCE AS A UNIFIER

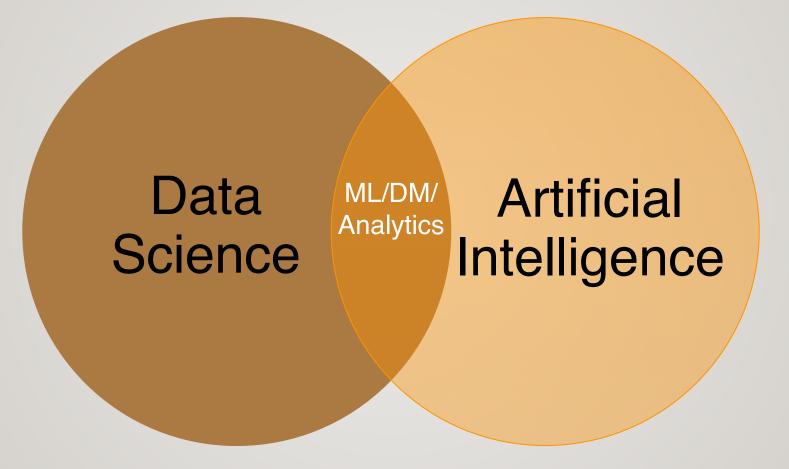


DATA SCIENCE AND BIG DATA

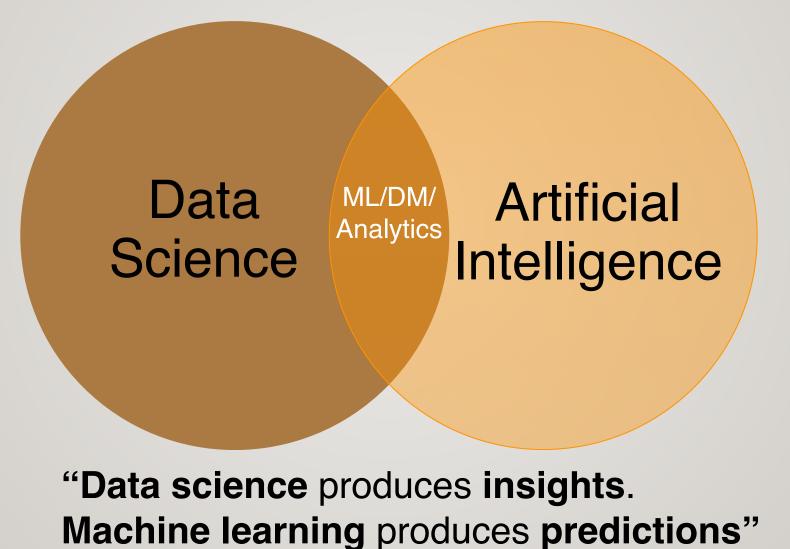
- They are not the "same thing"
- Big data = crude oil
 - Big data is about extracting "crude oil", transporting it in "mega tankers", siphoning it through "pipelines", and storing it in "massive silos"
- Data science is about refining the "crude oil"

Carlos Samohano Founder, Data Science London

DATA SCIENCE AND ARTIFICIAL INTELLIGENCE



DATA SCIENCE AND ARTIFICIAL INTELLIGENCE



- Fraud detection
 - Investigate fraud patterns in past data
 - Early detection is important
 - Before damage propagates
 - Harder than late detection
 - Precision is important
 - False positive and false negative are both bad
 - Real-time analytics



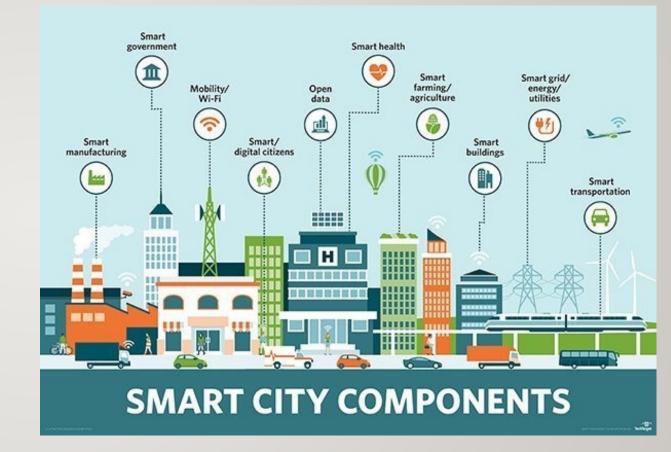
- Recommender systems
 - The ability to offer unique personalized service
 - Increase sales, click-through rates, conversions, …
 - Netflix recommender system valued at \$1B per year
 - Amazon recommender system drives a 20-35% lift in sales annually
 - Collaborative filtering at scale



- Predicting why patients are being readmitted
 - Reduce costs
 - Improve population health
 - Find the "why" behind specific populations being readmitted
 - Data lakes of multiple data sources
 - Investigate ties between readmission an socioeconomic data points, patient history, genetics, ...



- "Smart cities"
 - Not well-defined



- "Smart cities"
 - Not well-defined



lillisphotography / Getty / Emily Jan / The Atlantic

Stop Saying 'Smart Cities'

Digital stardust won't magically make future cities more affordable or resilient.

TECHNOLOGY

BRUCE STERLING | FEB 12, 2018

- "Smart cities"
 - Not well-defined
 - Generally refers to using data ICT to
 - Better plan communities
 - Better manage assets
 - Reduce costs
 - Deploy open data to better en with community



lillisphotography / Getty / Emily Jan / The Atlantic

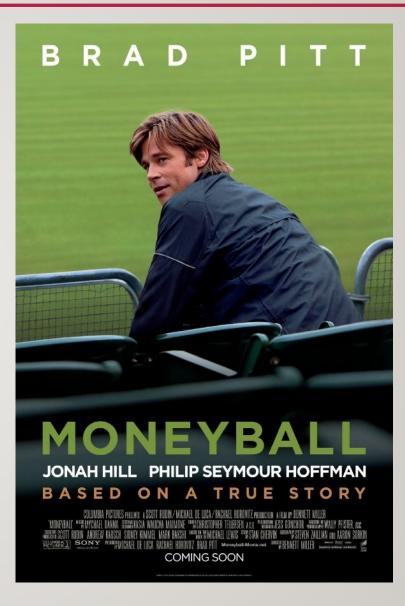
Stop Saying 'Smart Cities'

Digital stardust won't magically make future cities more affordable or resilient.

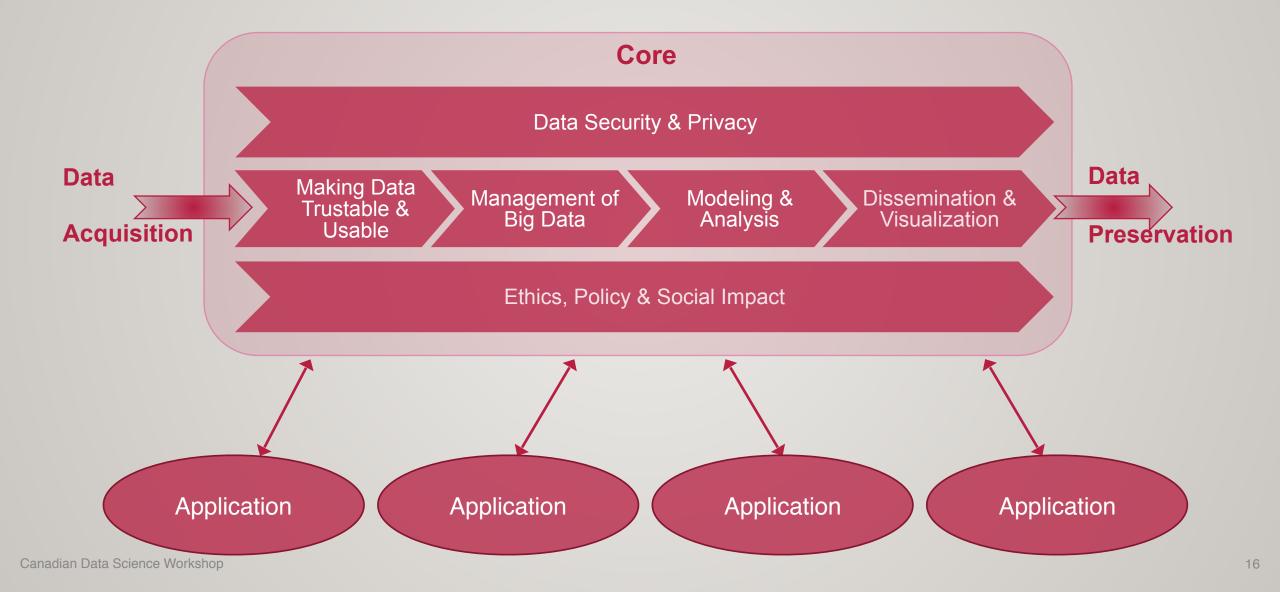
BRUCE STERLING | FEB 12, 2018

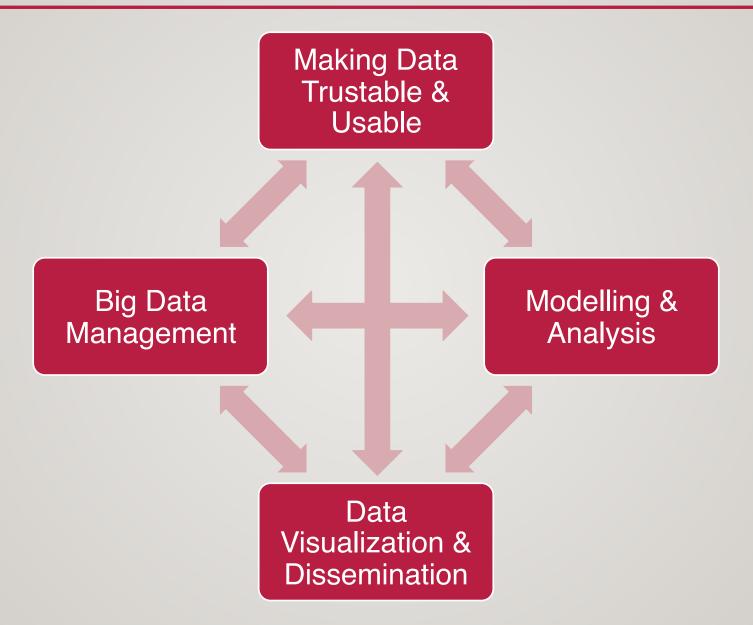
Canadian Data Science Workshop

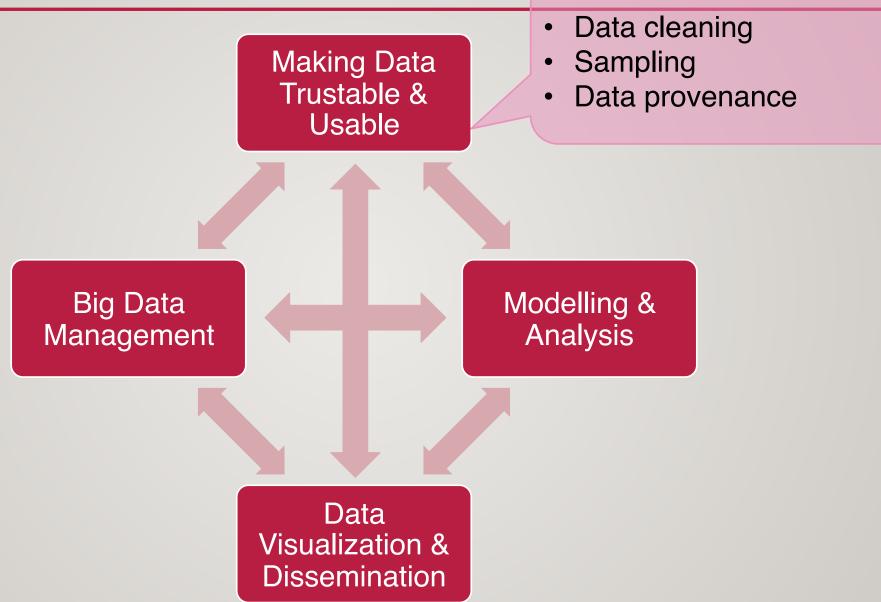
- Moneyball
 - How to build a baseball team on a very low budget by relying on data
 - Sabermetrics: the statistical analysis of baseball data to objectively evaluate performance
 - 2002 record of 103-59 was joint best in MLB
 - Team salary budget: \$40 million
 - Other team: Yankees
 - Team salary budget: \$120 million

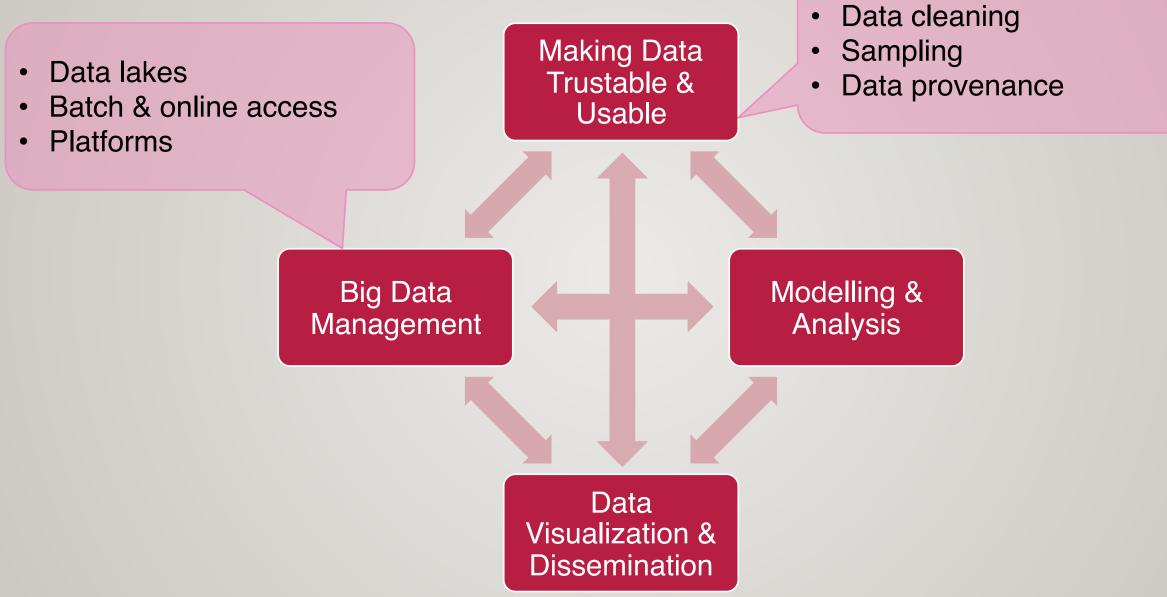


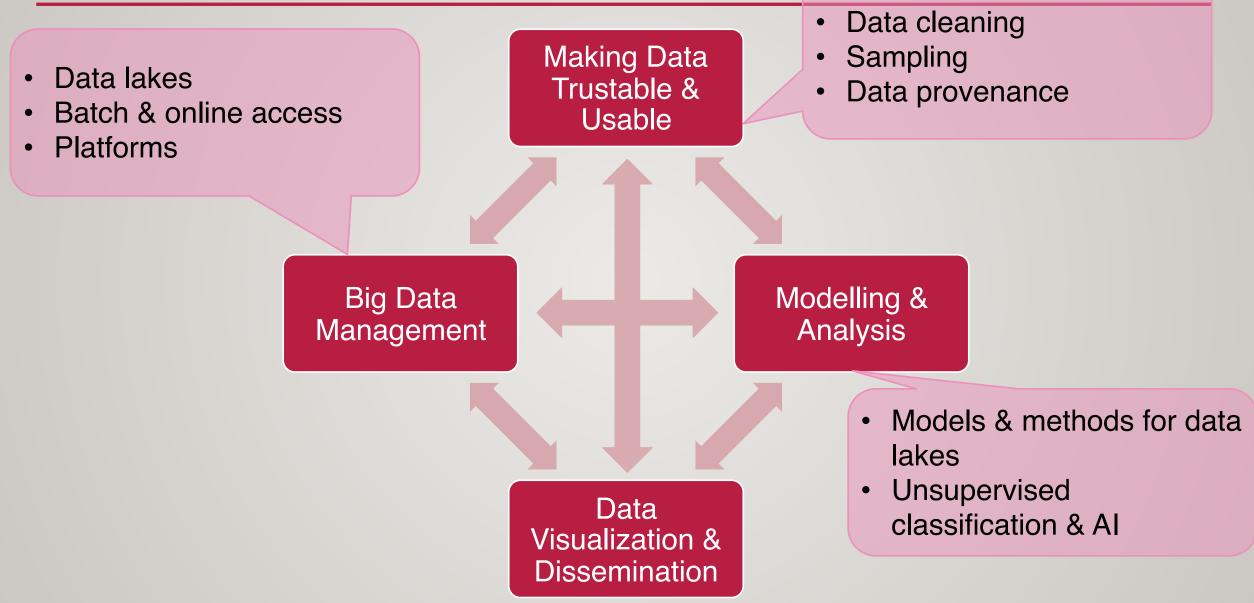
HOLISTIC APPROACH TO DATA SCIENCE

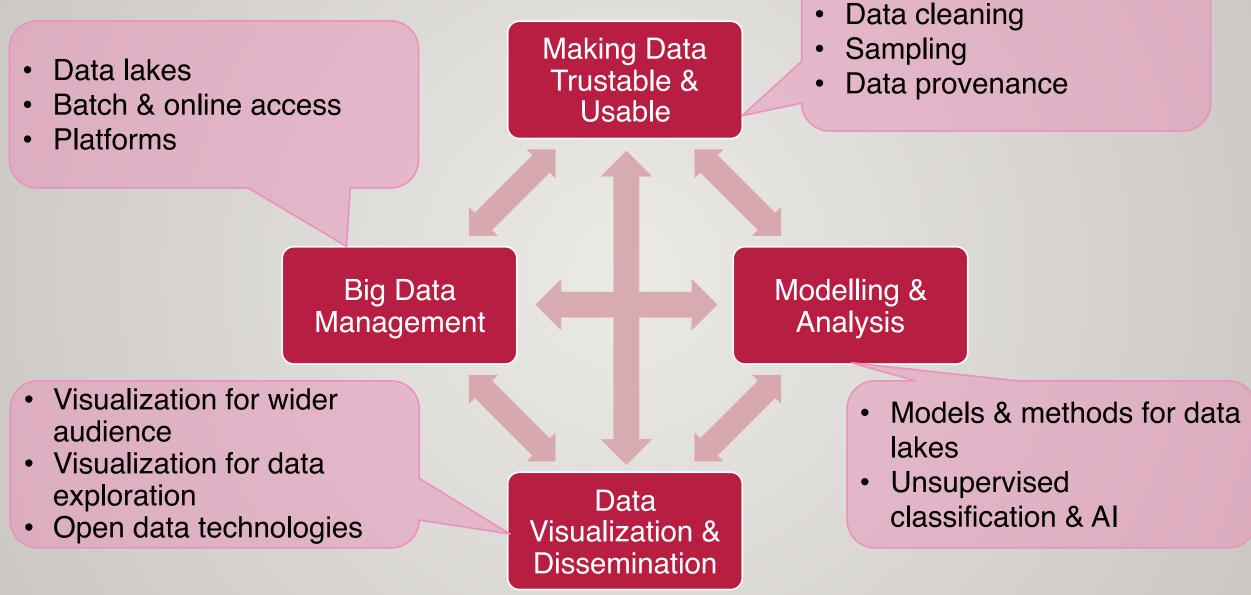












Data lakesBatch & online access	Making Data Trustable & Usable	Data cleaningSamplingData provenance
Platforms	 DM support for provenance Data preparation for big 	
	data management ta Cleaning for data	odelling & Analysis
 Visualization for wider audience Visualization for data exploration Open data technologies 	 ML for DM Visual analytics Data Visualization & Dissemination 	 Models & methods for data lakes Unsupervised classification & Al