



# **DỮ LIỆU LỚN (BIG DATA)**

# TỔNG QUAN VỀ DỮ LIỆU LỚN

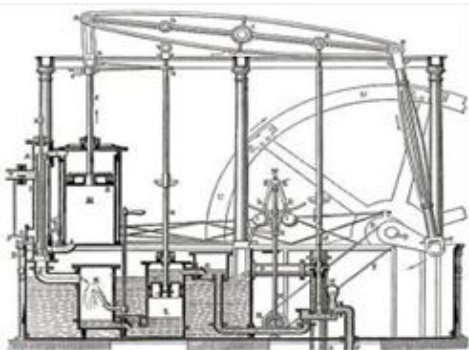


# NỘI DUNG

---

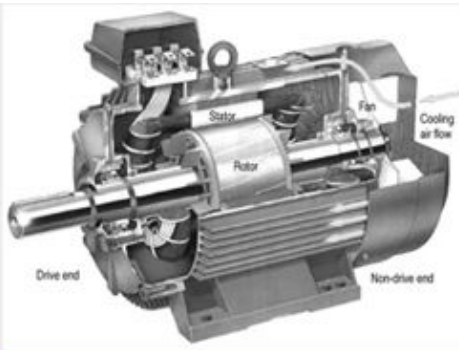
- Dữ liệu lớn/BigData (DLL) là gì?
  - Lớn như thế nào
  - Nguồn phát sinh DLL?
  - Cơ hội, thách thức của DDL?

# CÁCH MẠNG CÔNG NGHIỆP LẦN 4



Cách mạng công nghiệp lần thứ nhất về **sản xuất cơ khí** với máy dựa vào **động cơ hơi nước**.

Cuối thế kỷ 18 đầu thế kỷ 19



Cách mạng công nghiệp lần thứ hai về **sản xuất hàng loạt** với máy dựa vào **năng lượng điện**.

Cuối thế kỷ 19 đầu thế kỷ 20



Cách mạng công nghiệp lần thứ ba về **sản xuất tự động** với **máy tính, điện tử và cách mạng số hoá**.

Từ thập kỷ 70 của thế kỷ 20



Cách mạng công nghiệp lần thứ tư về **sản xuất thông minh** nhờ các **đột phá của công nghệ số**.

Bắt đầu từ bây giờ

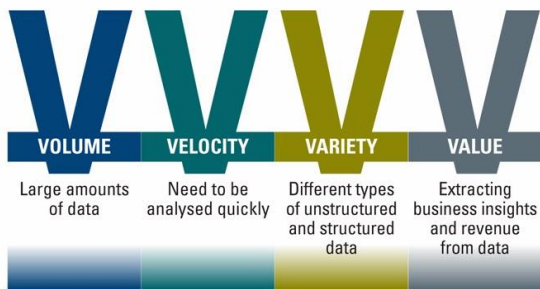
# DLL LÀ GÌ?



Nguồn: <https://intellipaat.com/blog/author/naveen/page/3/>

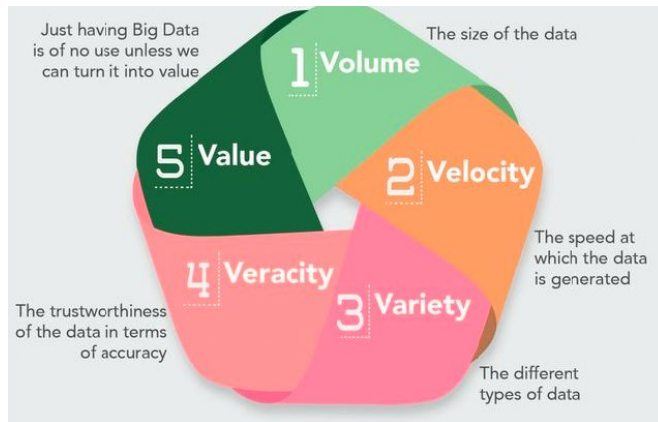
## Big Data: The four Vs

Volume, Velocity, Variety and Value



© World Newsmedia Network 2013

Nguồn: <https://newsbizblog.org/2013/11/14/big-data-the-four-vs/>

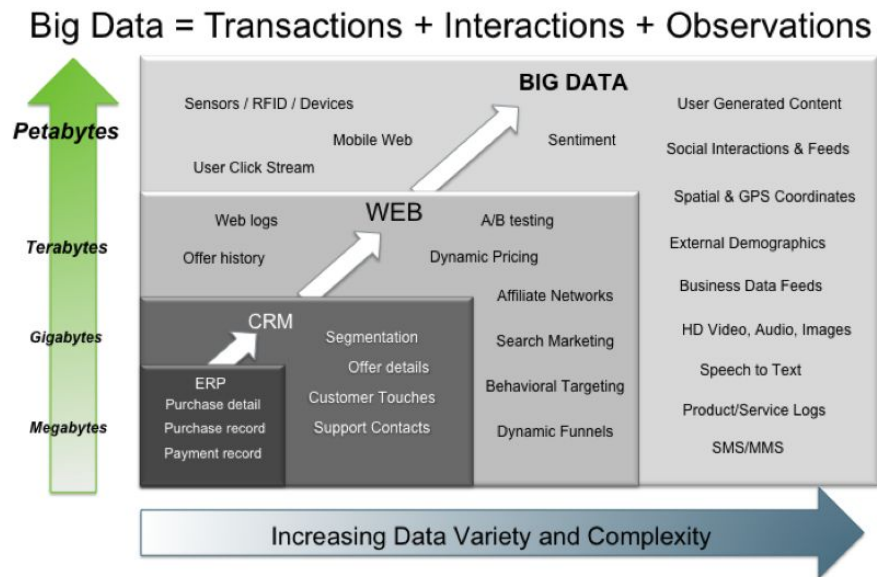


Nguồn: <https://elearningindustry.com/exploring-dark-side-of-big-data>

- Volume: Terabytes, Petabytes, Exabytes, Zettabytes, Yottabytes.
- Velocity: không ngừng chuyển động.
- Variety: nhiều định dạng khác nhau, không cấu trúc.
- Veracity: nhiều, không chính xác.
- Value: chuyển dữ liệu thành \$.

# DLL LÀ GÌ?

Là các tập dữ liệu rất lớn và/hoặc rất phức tạp, vượt quá khả năng xử lý của các phương pháp kỹ thuật tin học truyền thống.



Source: Contents of above graphic created in partnership with Teradata, Inc.

Nguồn: <http://aoife.dbsdataprospects.com/2016/04/07/big-data-big-decisions/>

# LỚN NHƯ THẾ NÀO?

Name	Equal to:	Size in Bytes
Bit	1 bit	1/8
Nibble	4 bits	1/2 (rare)
Byte	8 bits	1
Kilobyte	1,024 bytes	1,024
Megabyte	1,024 kilobytes	1,048,576
Gigabyte	1,024 megabytes	1,073,741,824
Terrabyte	1,024 gigabytes	1,099,511,627,776
Petabyte	1,024 terrabytes	1,125,899,906,842,624
Exabyte	1,024 petabytes	1,152,921,504,606,846,976
Zettabyte	1,024 exabytes	1,180,591,620,717,411,303,424
Yottabyte	1,024 zettabytes	1,208,925,819,614,629,174,706,176



# LỚN NHƯ THẾ NÀO?

---

TERABYTE: Ước lượng  
lưu trữ 200,000  
photos

Nguồn:

<https://www.theatlantic.com/technology/archive/2011/05/infographic-how-big-is-a-yottabyte/239034/>

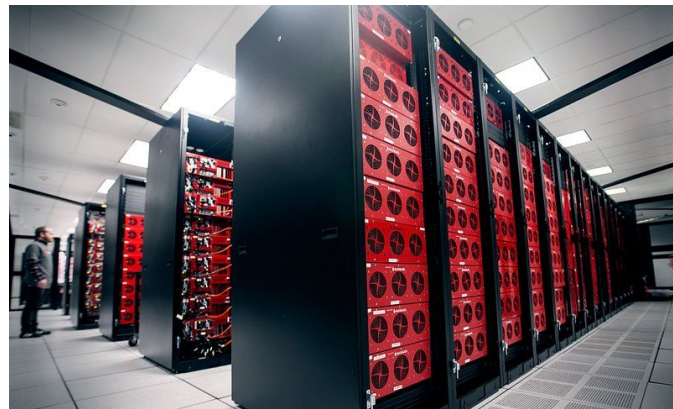




# LỚN NHƯ THẾ NÀO?

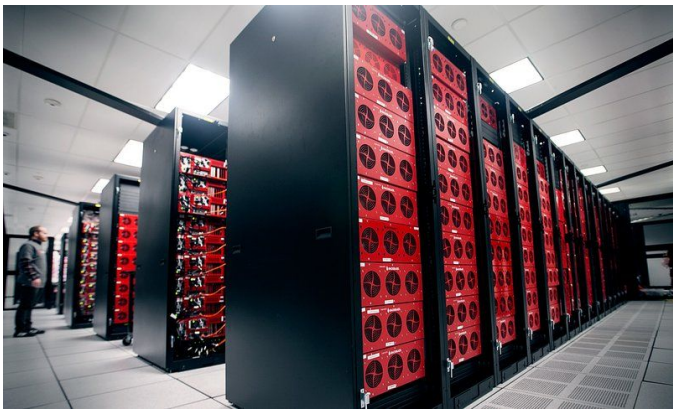
PETABYTE: ước lượng cần 16 Backblaze storage pods đặt trong 2 tủ.

1 Backblaze storage pod



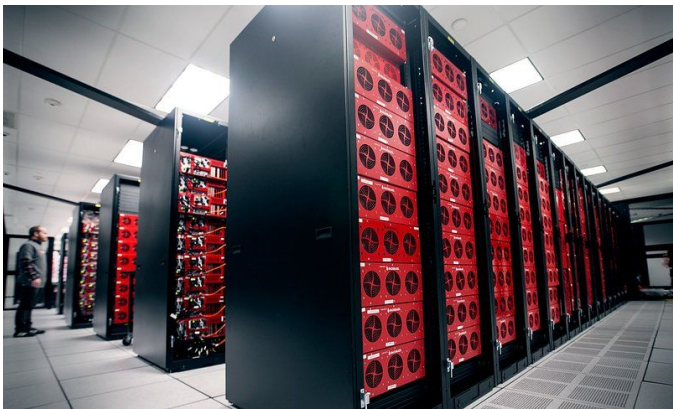
# LỚN NHƯ THẾ NÀO?

EXABYTE: cần khoảng 2000 tủ  
và đặt trong datacenter 4 tầng



# LỚN NHƯ THẾ NÀO?

ZETTABYTE: cần khoảng 1000 datacenter,  
20% quận Manhattan, New York



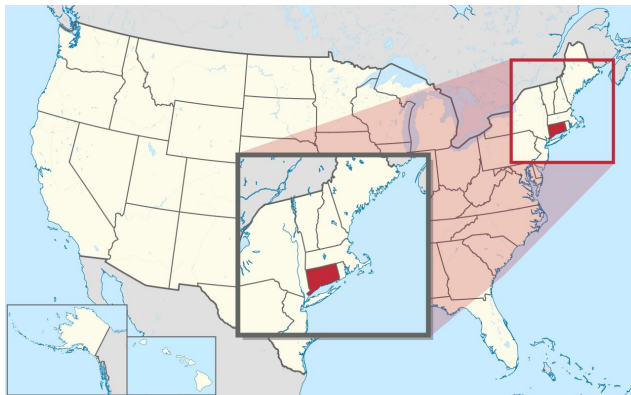
# LỚN NHƯ THẾ NÀO?

ZETTABYTE: cần khoảng 1000 datacenter,  
20% quận Manhattan, New York

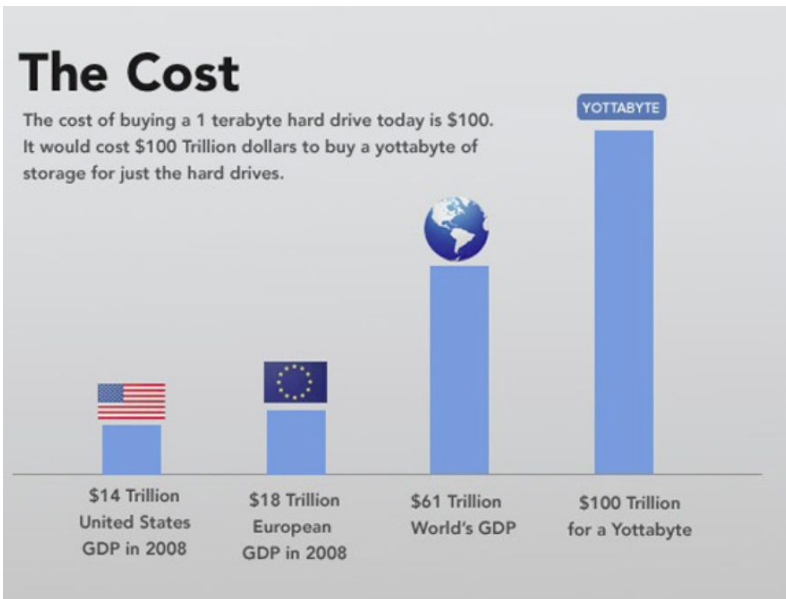


# LỚN NHƯ THỂ NÀO?

YOTTABYTE: cần khoảng 1.000.000 datacenter, lấp đầy 2 bang Delaware và Rhode Island nhỏ nhất, Mỹ



# LỚN NHƯ THẾ NÀO?



- Khoảng 100 USD cho Terabyte
- Cần 100 nghìn tỷ USD cho 1 Yottabyte.
- GDP của US trong 2008, 14 nghìn tỷ USD.
- GDP châu âu trong 2008, 18 nghìn tỷ USD.
- GDP Cả thế giới, 61 nghìn tỷ USD

# LỚN NHƯ THẾ NÀO?

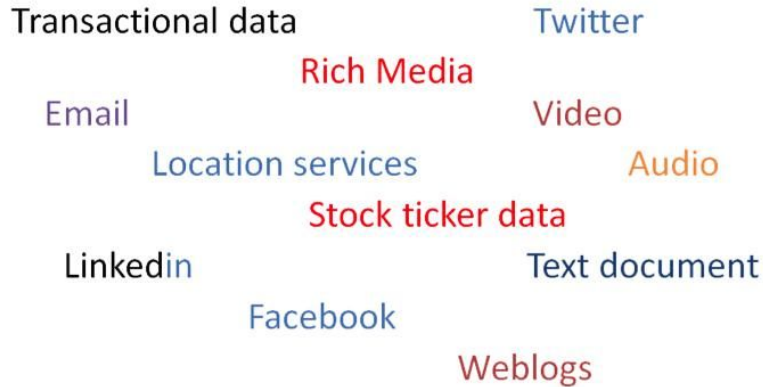
- Theo báo cáo của IDC được tài trợ bởi Seagate, sẽ không có gì ngạc nhiên vào năm 2025, chúng ta sẽ bị “tràn ngập” bởi dữ liệu, khoảng 163 ZB, gấp 10 lần so với năm 2016 (khoảng 16 ZB)
- 2025, kỷ nguyên của dữ liệu





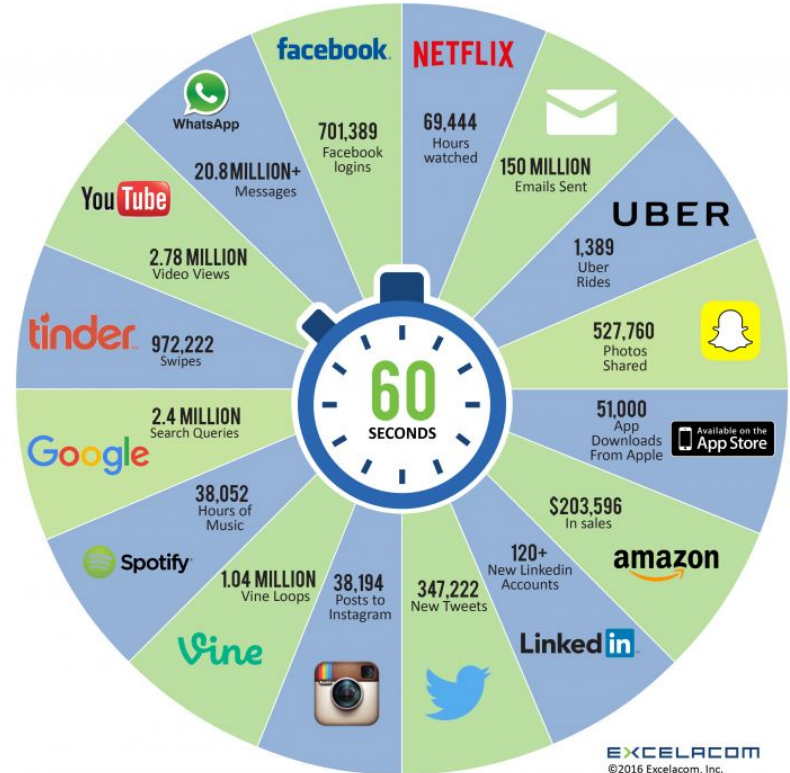
# NGUỒN SINH RA DLL?

## Variety of Big Data



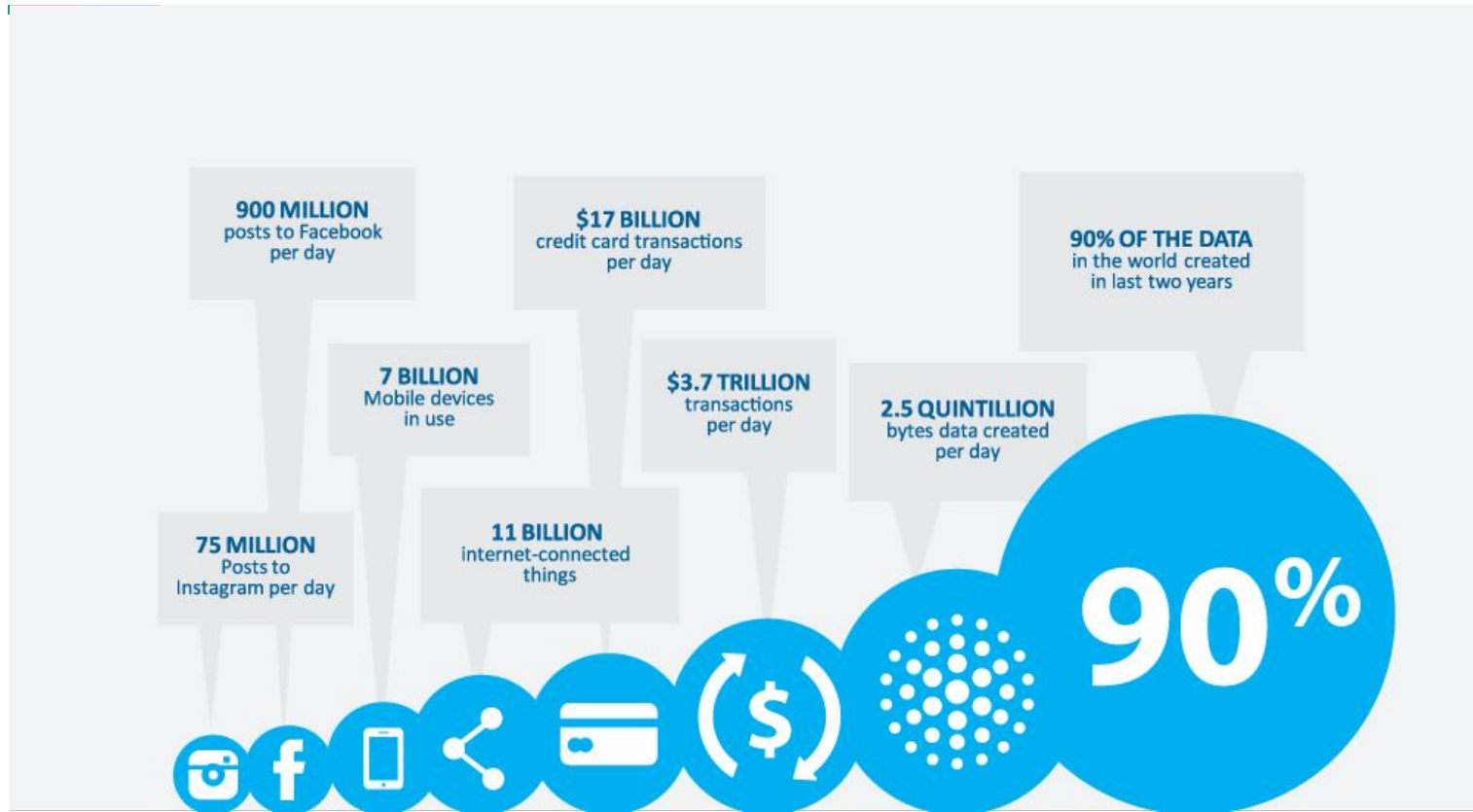
Nguồn: <https://www.linkedin.com/pulse/20140613120357-18324984-can-big-data-make-small-enterprises-big/>

## 2016 What happens in an INTERNET MINUTE?



Nguồn: <http://www.excelacom.com/resources/blog/2016-update-what-happens-in-one-internet-minute>

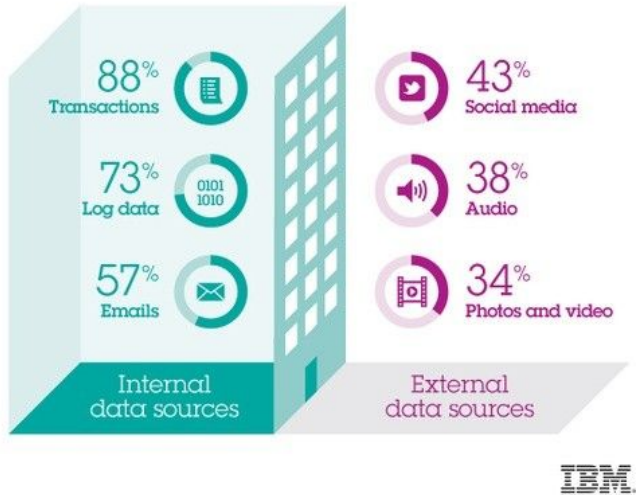
# NGUỒN SINH RA DLL?



# NGUỒN SINH RA DLL?

## Where does big data come from?

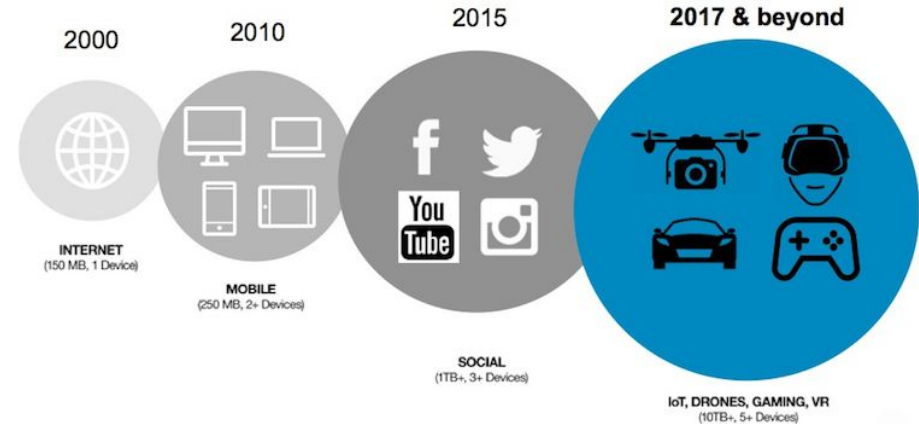
Most big data efforts are currently focused on analyzing internal data to extract insights. Fewer organizations are looking at data outside their firewalls, such as social media.



Nguồn: <http://www.ibmbigdatahub.com/infographic/where-does-big-data-come>

Data is being generated by people, machines, applications and a combination of all of these.

## THE NATURE OF DATA IS SHIFTING



Nguồn: <https://blog.seagate.com/business/the-evolution-of-smart-data-what-does-a-successful-iot-look-like/>

# Thách thức & cơ hội của DDL?



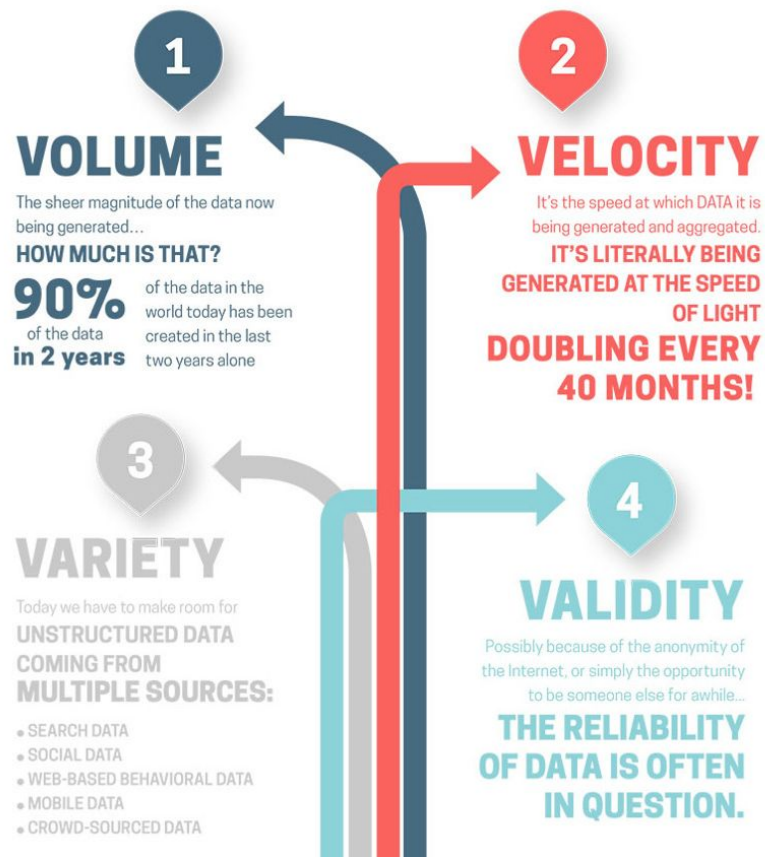
# Thách thức & Cơ hội của DDL?



Nguồn:

<http://www.exelanz.com/blogs/use-of-big-data-analytics-and-predictive-analytics-in-marketing-and-the-role-played-by-cloud-computing/>

Today we can access any data we want, but it doesn't provide insights to act upon. **WHY?** Let's look at the key challenges posed by big data:



Nguồn: <http://aoife.dbsdatapoints.com/2016/04/07/big-data-big-decisions/>

# Thách thức & Cơ hội của DDL?

---

- Chiến lược phát triển các quốc gia.
- Năm 2012, Chính quyền Mỹ công bố kế hoạch chi tiêu 200 triệu USD mỗi năm cho R & D liên quan DDL. Đây là một trong những chương trình R&D trọng điểm của chính quyền Mỹ.

<https://gcn.com/articles/2012/05/07/ed-note-feds-bridge-big-data-gaps.aspx>

<http://data.nistep.go.jp/dspace/bitstream/11035/2901/1/NISTEP-STT046E-37.pdf>

<https://www.smartdatacollective.com/incredible-ways-big-data-used-us-government/>

# Thách thức & Cơ hội của DDL?

Phân tích DDL → hiểu hành vi người dùng, đo lường dựa trên phản hồi để hành vi người dùng có phải theo cách họ tiên đoán hay không.

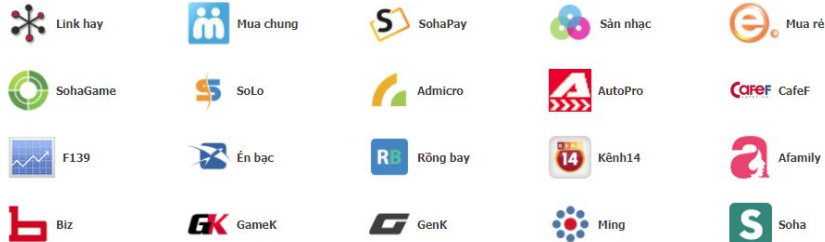




# Thách thức & Cơ hội của DDL?



Nguồn: <http://smartsocialcity.com/en/688-2/>



Nguồn: <https://vccorp.vn/>

# Thách thức & Cơ hội của DDL?



- Thủ tướng chỉ đạo khẩn trương tiếp cận cuộc cách mạng 4.0.
- Việt Nam đi trong cách mạng công nghiệp lần thứ 4 thế nào?
- TpHCM bắt đầu xây dựng đô thị thông minh.

## Nguồn:

[https://vnexpress.net/tin-tuc/khoa-hoc/trong-nuoc/thu-tuong-chi-dao-khan-truong-tiep-can-cuoc-cach-mang-4-0-3580577.html#ctr=box\\_topic\\_kinhdoanh\\_env\\_4\\_click](https://vnexpress.net/tin-tuc/khoa-hoc/trong-nuoc/thu-tuong-chi-dao-khan-truong-tiep-can-cuoc-cach-mang-4-0-3580577.html#ctr=box_topic_kinhdoanh_env_4_click)

[https://vnexpress.net/tin-tuc/khoa-hoc/viet-nam-di-trong-cach-mang-cong-nghiep-lan-thu-4-the-nao-3575368.html?vn\\_source=rcm\\_detail&vn\\_medium=khoahoc&vn\\_campaign=rcm&ctr=rcm\\_detail\\_env\\_4\\_click\\_khoahoc](https://vnexpress.net/tin-tuc/khoa-hoc/viet-nam-di-trong-cach-mang-cong-nghiep-lan-thu-4-the-nao-3575368.html?vn_source=rcm_detail&vn_medium=khoahoc&vn_campaign=rcm&ctr=rcm_detail_env_4_click_khoahoc)

<https://vnexpress.net/tin-tuc/thoi-su/tp-hcm-bat-dau-xay-dung-do-thi-thong-minh-3675717.html>

# Thách thức & Cơ hội của DDL?

---

- Lợi ích và chiến lược phát triển của quốc gia
- Cạnh tranh, phát triển, đổi mới doanh nghiệp
- Khởi nghiệp
- Khám phá mới cho khoa học
- Người dùng, người dân tiếp cận với xã hội thông minh.

# SÁCH THAM KHẢO CHO MÔN HỌC



1. Paul Zikopoulos, Chris Eaton. 2011. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data (1st ed.). McGraw-Hill Osborne Media.
2. Judith Hurwitz, Alan Nugent, Fern Halper, and Marcia Kaufman. 2013. Big Data for Dummies (1st ed.). For Dummies.

# Tóm tắt



1. Dữ liệu lớn/BigData (DLL) là gì?
  - a. Lớn như thế nào
  - b. Nguồn phát sinh DLL?
  - c. Cơ hội, thách thức của DDL?