

# Đào tạo nhân lực và sản xuất thông minh thời chuyển đổi số

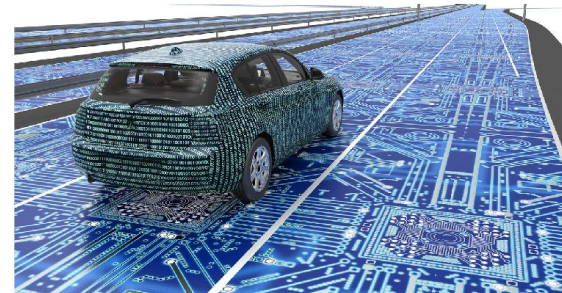
Nguyễn Thanh Tùng  
Trường Đại học Thủy Lợi  
[tungnt@tlu.edu.vn](mailto:tungnt@tlu.edu.vn)



# Thời chuyển đổi số

## Digital transformation time

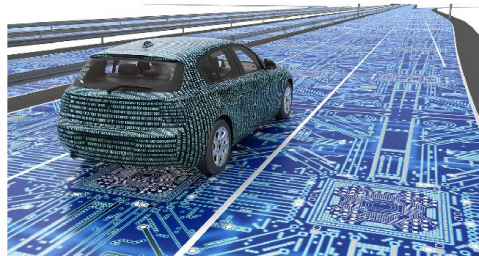
Thế giới các thực thể



# Thời chuyển đổi số



Thế giới các thực thể



Thế giới thực thể – không gian số  
Physical-cyber systems

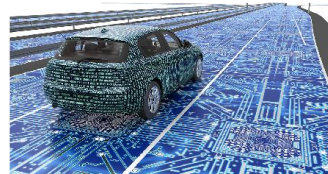
# Thời chuyển đổi số



Thế giới các thực thể

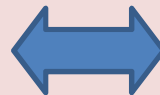


Thí dụ: 1 triệu  
xe kinh doanh  
trên toàn  
quốc đã gắn  
GPS,  
~500GB/day.



Thế giới thực thể – không gian số  
Physical-cyber systems

Mọi điều xảy ra trên  
thế giới các thực thể



Tính toán, điều khiển  
trên không gian số

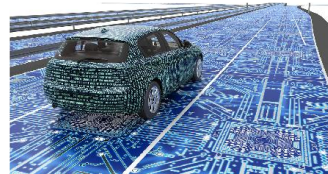
# Thời chuyển đổi số



Thế giới các thực thể

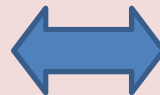


Thí dụ: 1 triệu  
xe kinh doanh  
trên toàn  
quốc đã gắn  
GPS,  
~500GB/day.



Thế giới thực thể – không gian số  
Physical-cyber systems

Mọi điều xảy ra trên  
thế giới các thực thể



Tính toán, điều khiển  
trên không gian số



Thay đổi phương thức sản xuất

# Đột phá của công nghệ số

- **Điện toán đám mây:**

Môi trường

- **Dữ liệu lớn:**

Năng lượng

- **Internet vạn vật:**

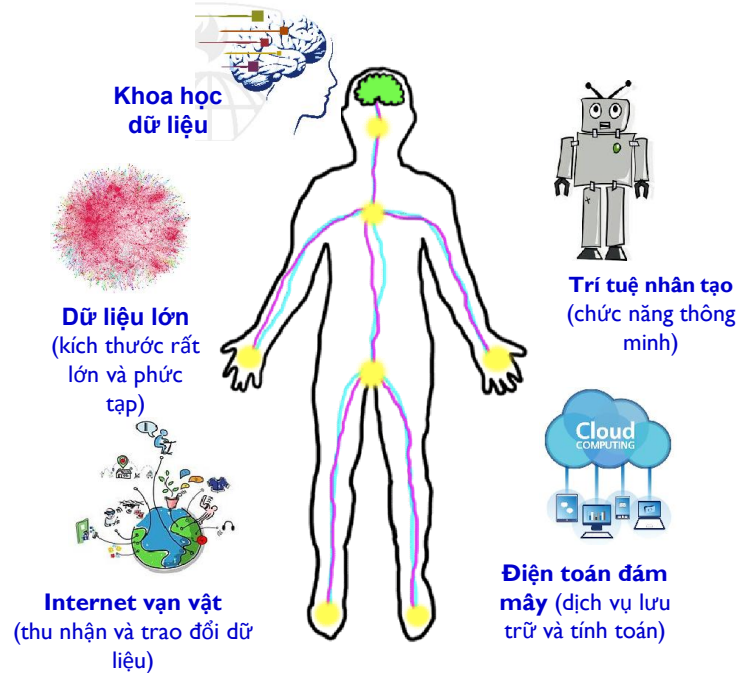
Thần kinh & huyết mạch

- **Trí tuệ nhân tạo:**

Chức năng thông minh

- **Khoa học dữ liệu:**

“Bộ não” phân tích dữ liệu để hỗ trợ quyết định và hành động.



**Tuy mọi công nghệ đều tiến bộ, cách mạng công nghiệp lần thứ tư xảy ra chủ yếu do hội tụ của các công nghệ số có nhiều đột phá.**



# Big data

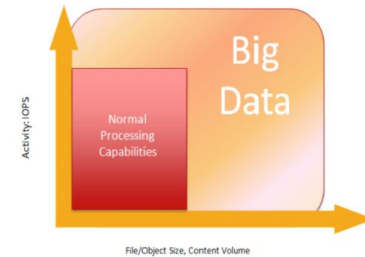
Big data refers to data sets that are **too large** and **too complex** to manage and analyze with traditional IT techniques.

**V**ariety: Complexity of data in many different structures, ranging from relational, to logs, to raw text

**V**elocity: Streaming data and large volume data movement

**V**olume: Scale from Terabytes to Petabytes ( $10^{15}$  bytes) to Zetabytes ( $10^{18}$  bytes)

**V**eracity: Accuracy and precision, truthfulness of the data.



**Data Scientist: The Sexiest Job of the 21st Century**

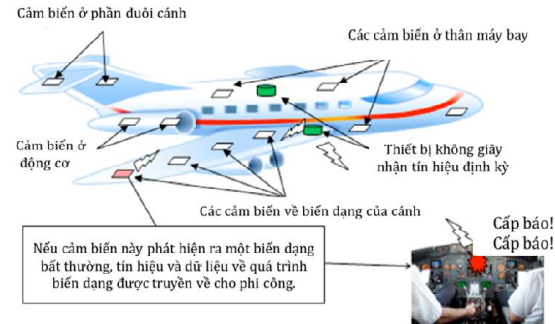
(Harvard Business Review, October 2012)



# Dữ liệu lớn có thể rất nhỏ. Không phải mọi tập dữ liệu to đều lớn

*Big data can be very small. Not all large datasets are big*

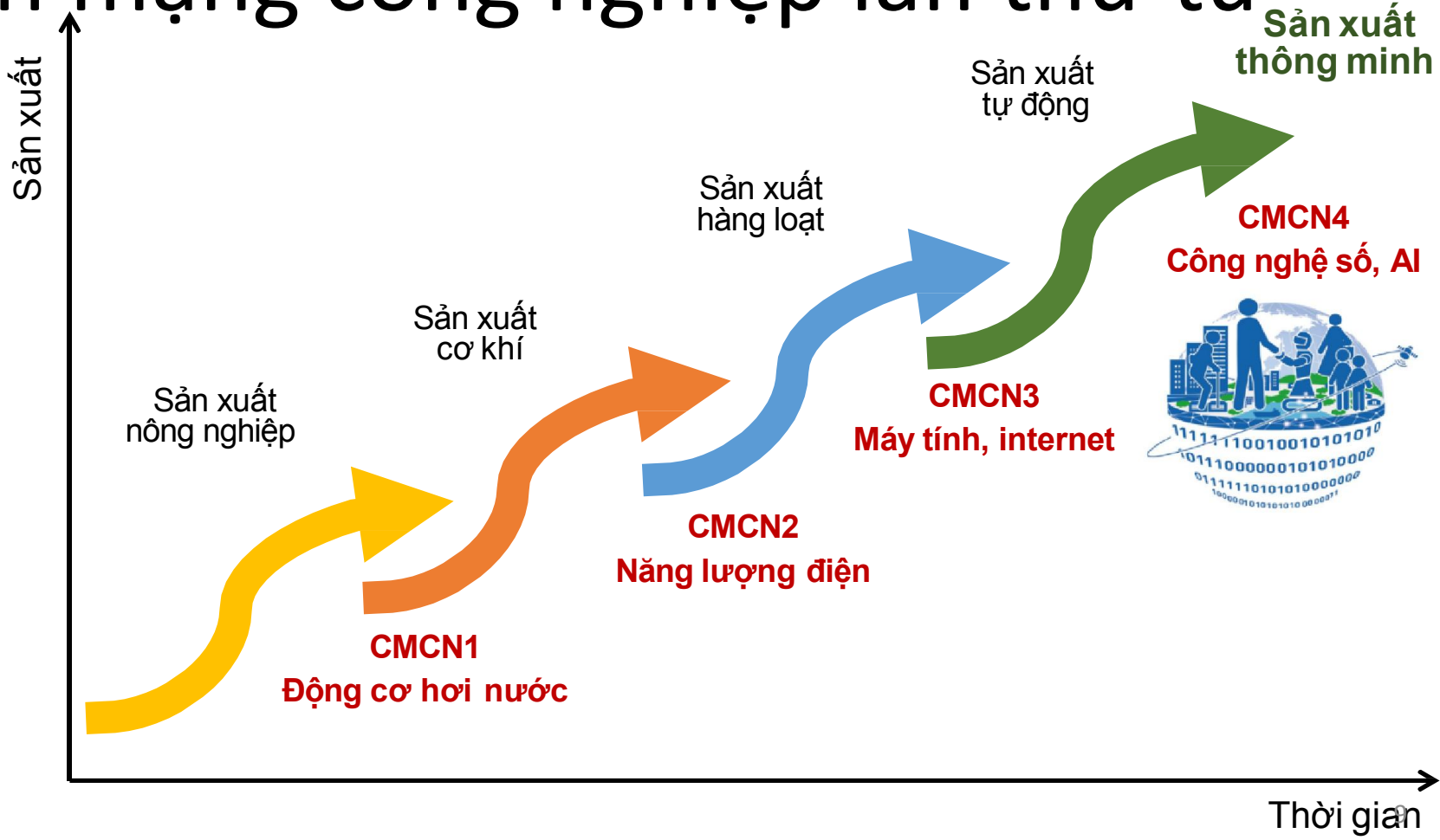
- **Big** liên quan tới sự **phức tạp** nhiều hơn tới **kích thước lớn**.
- **Dữ liệu lớn** nhưng lại nhỏ
  - Lò hạt nhân, máy bay... có hàng trăm nghìn sensors → sự phức tạp của việc tổ hợp dữ liệu các sensors này tạo ra?
  - Dòng dữ liệu của tất cả các sensors là lớn mặc dù kích thước của tập dữ liệu là không lớn (một giờ bay: 100,000 sensors x 60 minutes x 60 seconds x 8 bytes < 3GB).
- Tập dữ liệu **to nhưng không lớn**
  - Số hệ thống dù tăng lên và tạo ra những lượng khổng lồ dữ liệu nhưng đơn giản.



0001010101001001100010101010|  
0011000101010100100110001010|  
1001001100010101010010011000|  
1010100100110001010101001001|  
0010101010010011000101010100|  
0110001010101001001100010101|  
0010011000101010100100110001|  
0101001001100010101010010011|



# Cánh mạng công nghiệp lần thứ tư



H.T. Bảo

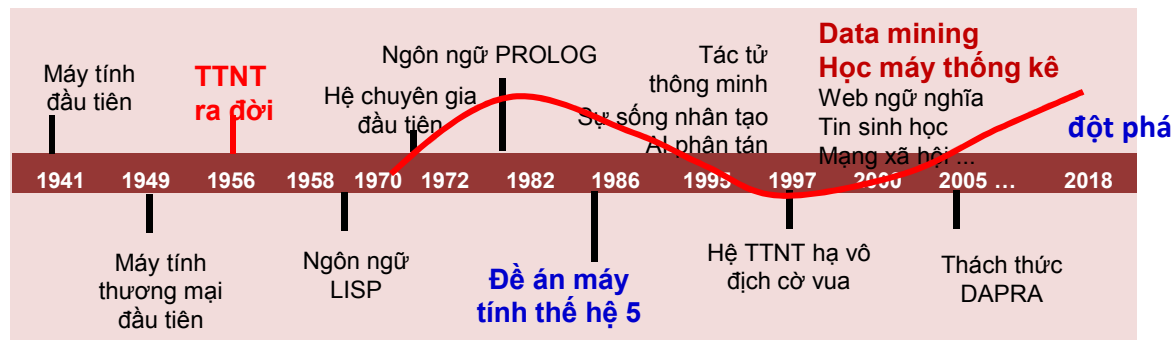


Thời gian

# Trí tuệ nhân tạo - Artificial Intelligence

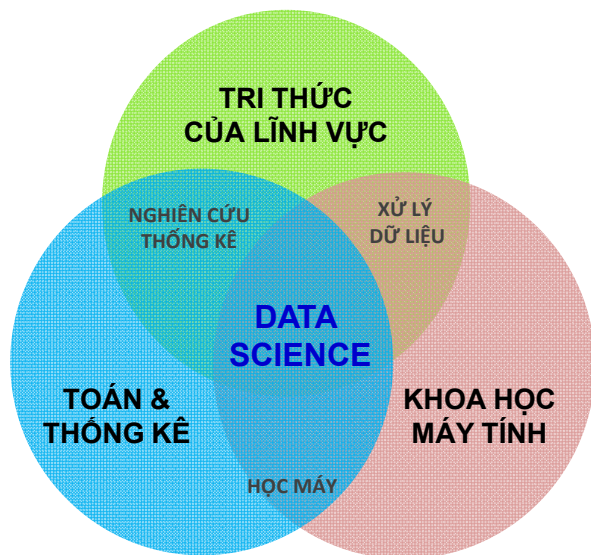


- Lĩnh vực làm cho máy (máy tính) hoạt động như có trí thông minh của con người (suy luận, giải quyết vấn đề, nhận thức, hiểu ngôn ngữ, học tập...).
- AlphaGo, hiểu ngôn ngữ, nhận dạng tiếng nói, chẩn đoán ung thư, ô-tô tự lái...



AI tăng trầm 60 năm, đột phá gần đây nhờ học máy (machine learning).

# Khoa học Dữ liệu



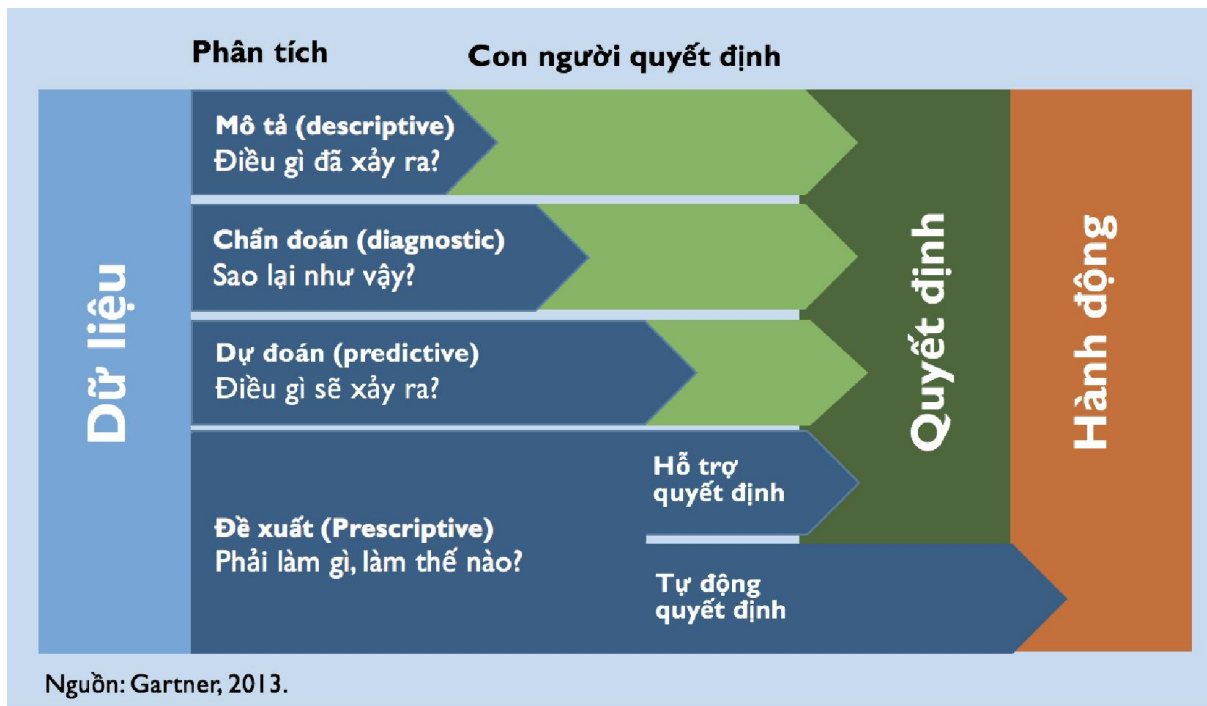
Khoa học về phân tích dữ liệu

“In God we trust.  
All others bring  
data”.  
“Ta tin Thượng đế.  
Ngoài ra, là dữ  
liệu”.  
W.E. Deming



Kết hợp của Toán học và Tin học là cốt lõi của Khoa học dữ liệu

# Khoa học dữ liệu: Cơ sở của quyết định



# Hạ tầng của sự phát triển?

Hành động trong thế giới  
vật lý của các thực thể



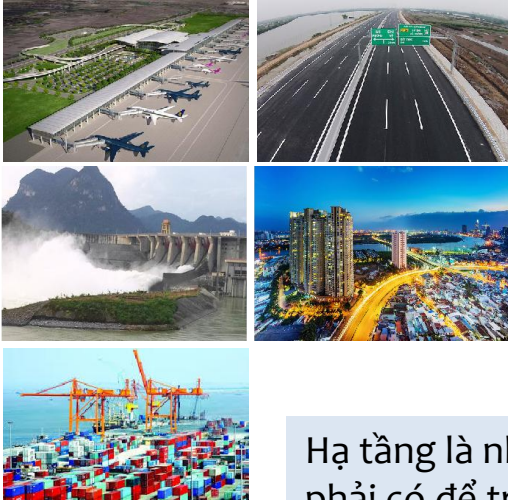
Hạ tầng là những thứ cần nhất  
phải có để trên đó ta xây dựng  
và phát triển những thứ khác.

# Hạ tầng của sự phát triển?

Hành động trong thế giới vật lý của các thực thể



Tính toán, điều khiển trên không gian số



**Hạ tầng  
trong thế  
giới số là gì  
và ta đã có  
gì?**

Hạ tầng là những thứ cần nhất phải có để trên đó ta xây dựng và phát triển những thứ khác.



# Thành phần trọng yếu của hạ tầng số?



- Xác định và xây dựng các thành phần chính của hạ tầng số.
- Lộ trình xây dựng phải hợp lý, không phải phá làm lại.

# NHÂN LỰC SỐ

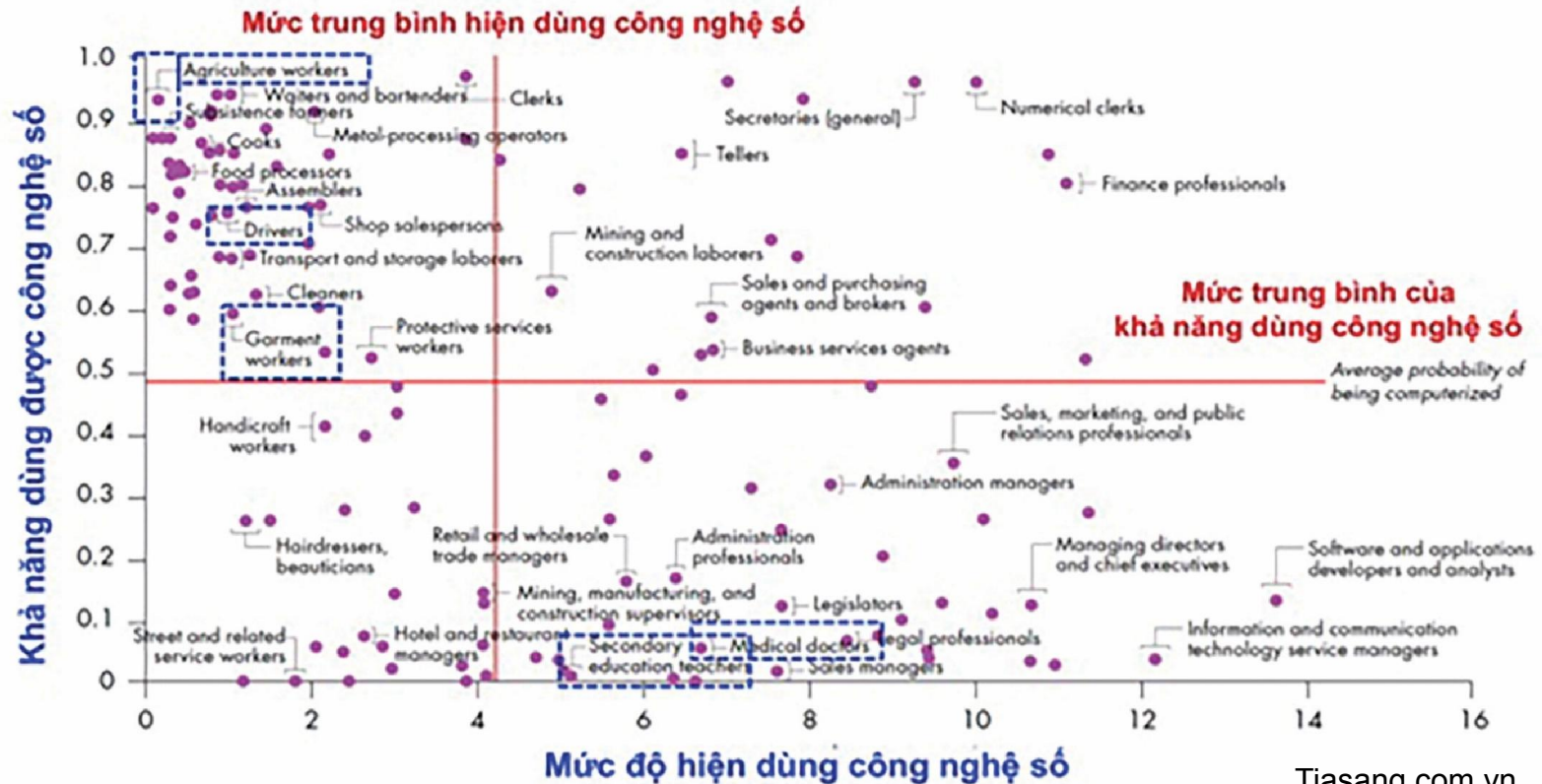
Xác suất	Nghề nghiệp có khả năng tự động hóa
0.99	Nhân viên tiếp thị từ xa (telemarketers)
0.99	Nhân viên kỹ thuật thư viện
0.98	Người định giá bảo hiểm
0.98	Trọng tài thể thao, các viên chức thể thao khác
0.98	Thư ký pháp luật
0.97	Chủ khách sạn, quán ăn, quán cà-phê
0.97	Người môi giới bất động sản
0.97	Nhà thầu lao động nông nghiệp
0.96	Thư ký và trợ lý hành chính ngành luật, ngành y
0.94	Nghề chuyển phát nhanh

Xác suất	Nghề nghiệp ít khả năng tự động hóa
0.0028	Chuyên gia trị liệu
0.0040	Biên đạo múa
0.0042	Bác sĩ và bác sĩ phẫu thuật
0.0043	Nhà tâm lý
0.0055	Nhà quản lý nguồn nhân lực
0.0065	Nhà phân tích hệ máy tính
0.0077	Nhà nhân chủng học và nhà khảo cổ học
0.0100	Kỹ sư hàng hải và kiến trúc sư hải quân
0.0130	Người quản lý kinh doanh
0.0150	Giám đốc điều hành



# Nhân lực số

Báo cáo của Ngân hàng Thế giới vào năm 2016

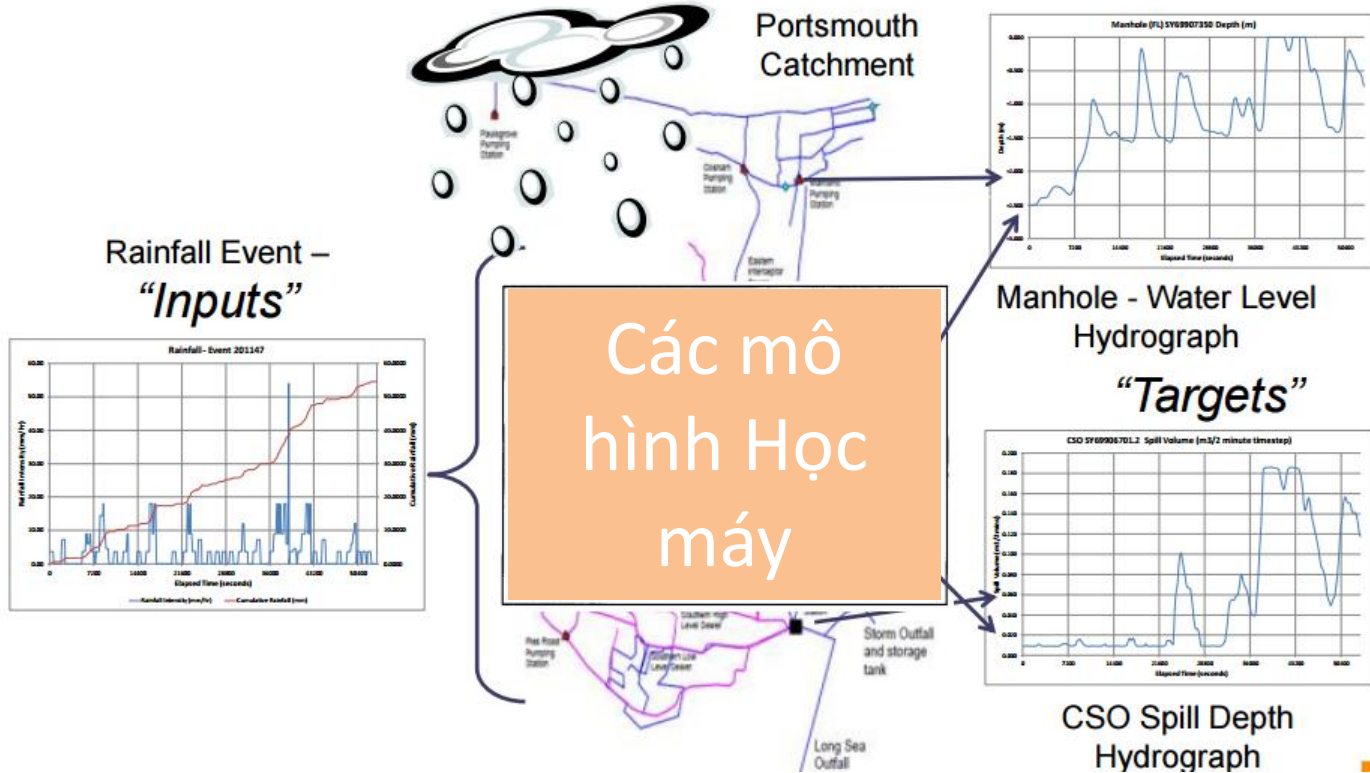


# Ngành thủy lợi và khoa học dữ liệu

- Bước đầu xây dựng được các CSDL quản lý: CSDL thiên tai, hồ chứa, đê và công trình trên đê,..
- Tương tác trực quan qua giao diện GIS
- Một số tồn tại:
  - Dữ liệu khó chia sẻ, manh mún và khó tích hợp để khai thác.
  - Chưa ứng dụng được AI & Khoa học dữ liệu theo xu hướng CMCN4.
  - Các ứng dụng trong ngành chưa kết hợp được các mô hình thủy văn, thủy lực và học máy để kết quả dự báo chính xác hơn.



# VD: Dự đoán mực nước



# Đào tạo và nghiên cứu KHDL

## Khoa CNTT, ĐHTL có 3 Labs:

- Modeling & Simulation Laboratory
- **Data Science Laboratory**
- Network Systems (IoT)





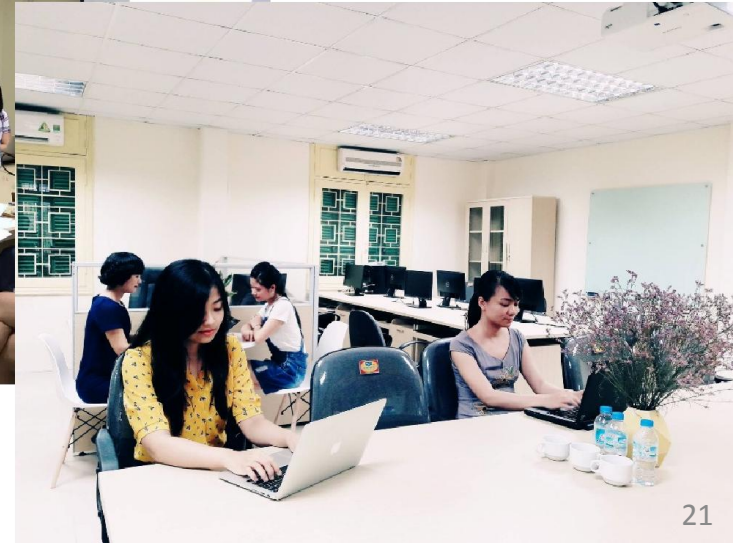
# Phòng thí nghiệm Khoa học dữ liệu (DS Lab)

Các Tiến sỹ nghiên cứu và đào tạo Khoa học dữ liệu

- Nguyễn Thanh Tùng
- Đặng Thị Thu Hiền
- Nguyễn Mạnh Hiển
- Lại Hiền Phương
- Trần Mạnh Tuấn
- Bùi Thu Cúc
- Trần Thị Ngân



cùng 15 thầy cô và sinh viên khác.

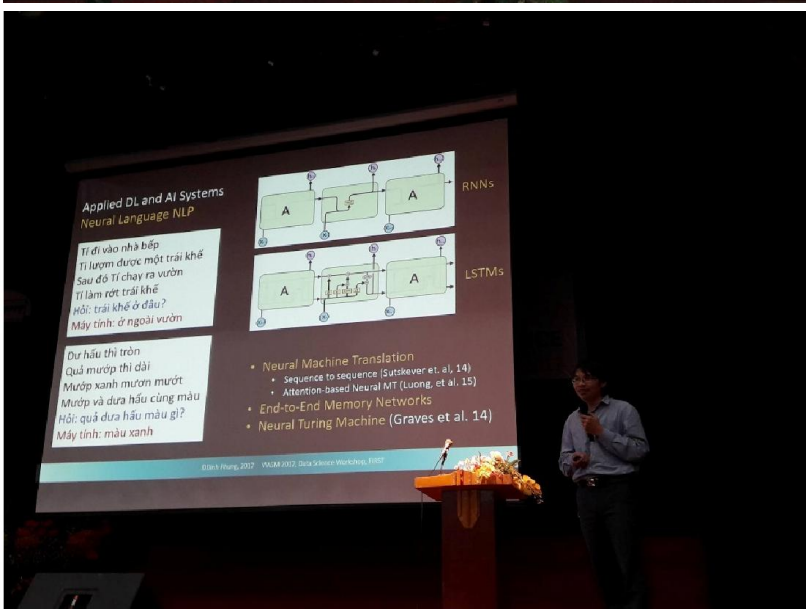


# Đào tạo Khoa học dữ liệu

1	Brief of statistics
2	Brief of linear algebra/optimization
3	Introduction to R and Python
4	Data preprocessing
5	Principles of Data Science
6	Evaluation of analysis results
7	Regression
8	Classification
9	Cluster analysis
10	Association analysis
11	Real-life problems 1
12	Outlier analysis/Anomaly detection
13	Reinforcement learning
14	Analysis of social networks
15	Analysis of time series data
16	Real-life problems 2

- Đào tạo sinh viên chính quy hai ngành CNTT và HTTT.
- Đào tạo các khoa học ngắn hạn (3-6 tháng)
- Khóa học chuyên đề cho Trường/Viện và doanh nghiệp.







# DSL Lab Opening ceremony 13.4.2018



# Trao đổi chuyên môn của DS Lab



# Ngành thủy lợi và khoa học dữ liệu

Ta có thể và nên đi thế nào trong CMCN4?

- Con đường và cách đi phụ thuộc nhiều yếu tố: thể chế chính trị, chính sách phát triển của Nhà nước, Bộ và Tổng cục, tình trạng kinh tế, xã hội, văn hoá, giáo dục và đào tạo, khoa học và công nghệ...
- Khi CMCN4 bùng nổ, mọi ngành nghề đều phải dựa nhiều hơn vào khoa học và công nghệ, vào hạ tầng dữ liệu.
- Thủy lợi hiệu quả và chính xác? Nông nghiệp và du lịch thông minh? Lựa chọn và làm chủ những công nghệ số và các công nghệ cao cần cho mình?



**Thực hiện được đến đâu sự thay đổi phương thức sản xuất trong việc ta muốn và cần làm?**





# Ngành thủy lợi và khoa học dữ liệu

- Xây dựng hạ tầng dữ liệu thủy lợi, tài nguyên nước và khoa học dữ liệu là 'hạ tầng cơ sở' để ngành thủy lợi tiếp cận CMCN4.
- Nhanh chóng đào tạo trong thời gian 3-5 năm một lực lượng về khoa học dữ liệu của ngành thủy lợi.
- Có lộ trình hợp lý và từng bước thực hiện. Trước hết cần đẩy mạnh số hoá và tạo các nguồn dữ liệu cơ bản.
- Sử dụng khoa học dữ liệu hiệu quả để đưa ra các quyết định đúng và sáng suốt trong các hoạt động của ngành.



# Một vài suy nghĩ

- Câu hỏi chung là mỗi lĩnh vực quản lý của Bộ, Ngành cần và có thể chuyển đổi thế nào trong thời chuyển đổi số?
- Những mô hình, tri thức, nguyên lý, phương pháp... chúng ta đang dùng tốt đến đâu và có thể tốt hơn không nếu gắn hơn với dữ liệu?
- Thủy lợi là lĩnh vực rất sâu và rất rộng. Liệu mỗi lĩnh vực của thủy lợi có cần và có thể gắn với ‘data- driven’ thế nào để sản xuất và quản lý điều hành “thông minh” hơn?
- Khoa học dữ liệu và hạ tầng dữ liệu thủy lợi (IoT hỗ trợ thu thập tự động) là chìa khóa để hiện đại hóa ngành trong xu thế CMCN4.



# Câu hỏi?

