

The modern data flow

Data, Applications & the Business

Manh NGUYEN, VnInfo

20 September 2018

Introduction and Motivation

The digital portfolio structure

Conclusion

Introduction and Motivation

Introduction

- ▶ Your profiles ? A little survey to start
- ▶ My profile:
 - ▶ Maths & Engineer background
 - ▶ 4 years in Economic & Finance Studies
 - ▶ 3 years as Data Scientist
 - ▶ Still not a specialist Data Scientist, good enough to get more specialised

Context and Motivation

- ▶ **The context:** The digital revolution has been changing the landscape of our society
 - ▶ Information versus Data
 - ▶ Digitalisation Process: Smartphones, Social Networks & IoTs
 - ▶ **Humans** and **Machines** together
 - ▶ New jobs and old jobs evolving: Business Analyst, BI, Data Engineers, Data Architect, Data Scientist, Data Analyst etc
 - ▶ Transition from model-centric to data-centric
- ▶ **The motivation:**
 - ▶ A data-centric discussion about data assets, applications and the business
 - ▶ A look in the past, the present to look further into the future

The digital portfolio structure

The data assets

- ▶ Data as in IT which is almost everything:
 - ▶ Big data with the 4Vs: volume, variety, veracity and velocity
 - ▶ Structured vs Non-structured or SQL vs No-SQL database or CSV vs JSON
 - ▶ Good quality and bad quality
 - ▶ Machine friendly vs human friendly
 - ▶ Raw data vs distilled data (most of the time aggregated data)
 - ▶ Internal vs external or private vs open data
 - ▶ Storage, Extraction, Search, Data Mining, Data Science
- ▶ Of course, there are still a lot of information which has not appeared as IT data form, so out of IT reach

Digital assets (intelligence)

- ▶ Intelligence is the information whose value lies between data and knowledge. Intelligence is a bit of information by processing data locally. It is actionable but might not be able to apply more broadly as knowledge
- ▶ Intelligence as Machine Learning Model

Digital assets (knowledge)

- ▶ Knowledge is a bit of information after processing data across larger domains and through larger time frames. Think of knowledge as scientific knowledge

The evolution of digital assets

- ▶ More tangible
- ▶ More intergrated
- ▶ More interactive
- ▶ More messy

Some historical perspectives

- ▶ The whole flow of information processing is not new:
 - ▶ Story about Copernicus, Galileo, Kepler and Newton: with the telescope and gravitational theory
 - ▶ Or Gutenberg with the invention of the printer
 - ▶ etc
- ▶ All information now just flows through digital means. Machines are computers, CPU, GPU.

The flow of digital assets

The different corresponding jobs

- ▶ Data engineer
- ▶ Data admin
- ▶ Data architect
- ▶ Data scientist
- ▶ Data analyst
- ▶ Business Analyst
- ▶ Project manager

Conclusion

What we are heading to then ?

- ▶ Humans and machines collaboration to serve human society
- ▶ Human as a social animal