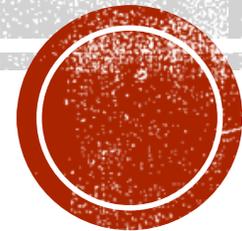




UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

EVOLUTION OF TRAINING IN BIG DATA IN FRANCE

France-Russia Workshop
October 26, 2018



Marc-Arthur DIAYE



UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

- **Analyst Consumer Insight**
- **Big data analyst**
- **Business Intelligence analyst**
- **Research analyst in statistics and computer-based decision support**
- **Consultant in statistiscis**
- **Data analyst**
- **Data Mining analyst**
- **Data scientist**
- **Data architect/Data Engineer**
- **(...)**





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

- In 2013, when big data started to become a strategic issue, the French government (of President Holland) asks for a strategic roadmap.
- This roadmap was not built by bureaucrats.
- Instead it was written by :
- Francois BOURDONCLE : Founder and Chief Strategist of Exalead (a search engine company) and currently CEO of FB & Cie
- Paul HERMELIN : CEO of Capgemini
- See « *Nouvelle France Industrielle : Big data, feuille de route* », July 2014.





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

- The Bourdoncle-Hermelin report includes seven recommendations.
- The first of these recommendation was to develop and improve the training data scientist in France.
- According to the report, France education system is already data scientist prone because French universities and Grandes Ecoles train well in mathematics, statistics and computer sciences.
- This means that France does not need to start from scratch and can adapt the existing Msc and bachelors, in order to make them more market-oriented.





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

- On the ground however, things were very different.
- The first true Msc in data science (for instance the one of Telecom ParisTech) have been created from scratch.
- The reason was that the existing Masters of science perform well and so Professors who led these Masters do not understand why they need to make their Msc more data science oriented.
- Moreover at this period, not every body in the academic community are convinced about Big Data.
- In the academic community, the computer scientists have immediatly understand the importance of Big Data/Data Science.
- This is may be the reason why one of the first true Msc in data science was set up in September 2013 at Telecom ParisTech, a French Engineering « Grande Ecole » specialized in Telecommunication. This master was created in Partnership with some French big names like Thales, Safran, Airbus, or Orange.





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

- However it is important to stress that the Msc created by Telecom ParisTech was what we call in France, Specialized Msc.
- Such masters are not part of the regular training process of engineers. Students who attend the Specialized Msc are not the regular engineers of Telecom ParisTech.
- Moreover the fees are quite high (in the French education system). For instance, the fees for the Msc in Big Data of Telecom ParisTech is about 19000 euro.
- This means that from the beginning, the Specialized Msc in big data of Telecom ParisTech self financed itself. Despite the fact that Telecom ParisTech is a public « Grande Ecole ».
- Despite the evident skepticism, this Specialized Msc was a success. The demand from firms of its students was very high.





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

- Given this success, some new actors enter in the market, either by adapting some existing masters and by creating a new one from scratch.
- For instance, in 2015, data science becomes a part of the regular training of the engineers from the most famous « Grande Ecole » in Statistics called Ecole Nationale de la Statistique et de l'administration Economique (ENSAE).





REMARK

- **Some old masters in data science do already exist.**
- Master Vision Apprentissage (MVA), from Ecole Normale Supérieure Paris-Saclay
- Specialised in ML and Image recognition.
- Exists since almost 20 years.

- Master TIDE from University Paris 1 Pantheon-Sorbonne
- Specialised in data management
- Exists since 40 years.

- (and many others)





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

■ **Conclusion**

- Four years after the Bourdoncle-Hermelin report, France has almost completely achieved his plan to have an efficient training system in data science.
- Data science is today part of the regular training of the engineers from of high ranked « Grandes Ecoles »; or these « Grandes Ecoles » have their own Msc (specialized or not) in data science.
- In prominent French Universities, the existing Masters in Statistics, Computer science, Quantitative finance, Fintech, Quantitative economics, ... have been transformed in order to make them more data science oriented.





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

- University and Engineering « Grandes Ecoles » masters in data science :
 - Professional Msc
 - Research Msc
- Business schools (like HEC, ESCP, ...) has also developed some Masters, MBA in data science.
- Rise of Specialized Certificates : short trainings whose role is to train on a specific aspect of data science (for instance, in ML, in Deep Learning, in DataViz, in Law, ..)





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

- On the job training : firms sent their employees to train in data science
- Since two years, Pole Emploi (a French agency in charge of unemployed people) has set up a huge plan to train in data science, some unemployed workers with a degree in science. This plan consists in a two-months intensive training of some pre-selected unemployed workers.
- On-line-training : [Datascientest.com](https://www.datascientest.com)





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

REALLY DATA SCIENTIST ?

- Actually a so-called « data scientist » is sometimes a data analyst or a data architect.



Data
analyst

Data
scientist

Data
architect





DATA ANALYST

- A data analyst is able to apply some statistical models using a data set (structured data or unstructured data) with some software like R, Tableau, Qlick, ..
- The data analyst is able to interpret (using his skills in a specific field: finance, marketing, economics, ...) in order to give a market value to his statistical analyses.
- However the data sets used by the data analyst have to be provided to him.





DATA SCIENTIST

- In the best cases, a data scientist can do everything that the data analyst can do.
- In addition, he/she can create his own statistical tools because he/she is, in addition to have perfect skills in theoretical statistics, a good developer (in Python, Scala, ...).
- He/she is able to estimate a statistical model on the server version of the statistical software.
- Moreover he/she understands data architecture.
- A top data scientist is, in addition of the above skills, also good in some fields of applied mathematics like graph theory or combinatorial optimization algorithms and complexity.





UNIVERSITÉ PARIS 1

PANTHÉON SORBONNE

DATA ARCHITECT

- The role of the data architect/data engineer is design (optimally) the system of data architecture.
- Moreover he/she is a good project manager in computer science.
- A data architect understands the needs of the data scientists and data analysts.
- Ideally he/she is also a data scientist.





- Actually about 65% of the so-called « data science » trainings (in France) train for a data analyst job.
- About 25% : data scientists
- About 10% : data architects





- **This means that there is a shortage of data architects.**
- **There is also a shortage of Big data Project Managers.**



O O F F R F D F M A N D F F N
 F C C N B C M F J R Q F I S C J A J Q S C J Q Q
 V F A P P P Q Q V F F

4 ground skills

| | |
|--|---|
| 1 ST GROUND: DOMAIN OF APPLICATION | Security, Image recognition, Finance, Marketing, Economie, ... |
| 2 ND GROUND: MODELS ET TRADITIONAL STATISTICAL SOFTWARE | Times series, Markov network, Scoring, Machine learning, ... |
| 3 ND GROUND: PROGRAMMING LANGUAGES | Java, R, Python, Scala, ... |
| 4 ND GROUND: TOOLS | Tableau, Alteryx, Hadoop, Hive, Spark, Splunk, ELK, ... |

