Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"

cudi
Corporación Universitaria para el Desarrollo de Internet A.C.
Internet 2 - México

CIC
Centro de Investigación en Computación

Big Data: calling for a new scope in the curricula of Computer Science

Dr. Luis Alfonso Villa Vargas

23 de Abril, 2015, Puerto Vallarta, Jalisco, México

Big Data: beyond my project

▸ This talk is not about

  ▸ The relevance of Big Data

  ▸ My research Project

  ▸ Survey of Big Data

  ▸ The future of Big Data

▸ What this talk is really about

  ▸ The scope in the curricula of Computer Science

SCENARIO: Technology Dynamic

just: 20 years WEB

10 years Social Network (was launched facebook)

8 years was launched Tweeter

7 years Smartphones (users worldwide 1.75 billion in 2014)

4 years Tablets (iPad)

10 years there were no jobs:

Marketing digital, @-Commerce

Mobile apps design

Web development

1 Million new BLOGS / month

1 Million new users Twitter / day

The Information Age has Turned Into the Era of Big Data

- Humans generate 2.5 exabytes of data on a daily basis from different sources
  - Financial
  - Services
  - Social media
  - Mobile devices
  - Commercial transactions
  - Scientific
- More data cross the internet every second than were stored in the entire internet just 20 years ago.
- Every day, Google alone processes about 24 petabytes of data.
- It is estimated that Walmart collects more than 2.5 petabytes of data every hour from its customer transaction.

Data-driven decision tend to be better decision.

………We do not have better algorithms. We just have more data….. [1]

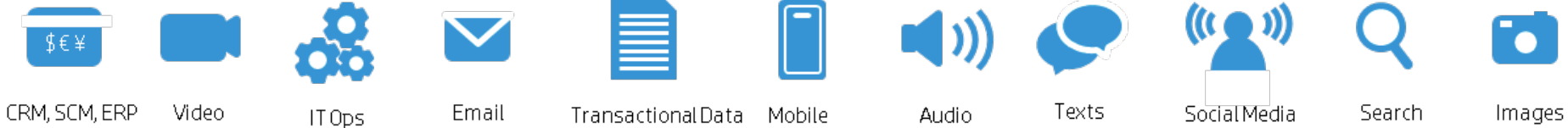……….each of us is now Walking Data Generator…..

[1]Google's director of research, Peter Norving, Harvard Business Review, october 2014

What´s New Here?

- The Big Data movement, like analytics before it, seeks to extract intelligence from data and translate that into business advantage. However, there are four key differences:
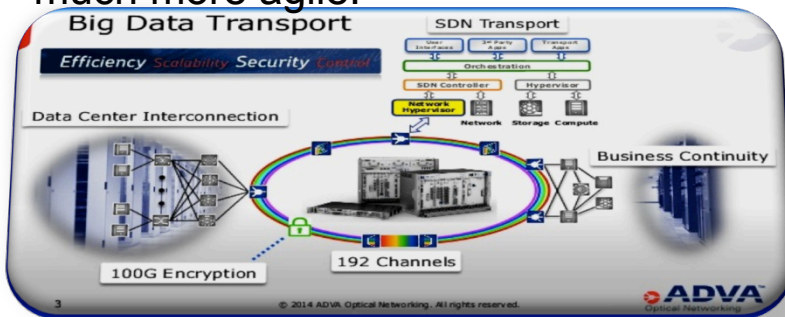
What´s New Here?
Data Path

Gathering: the speed of data creation is even more important than the volume
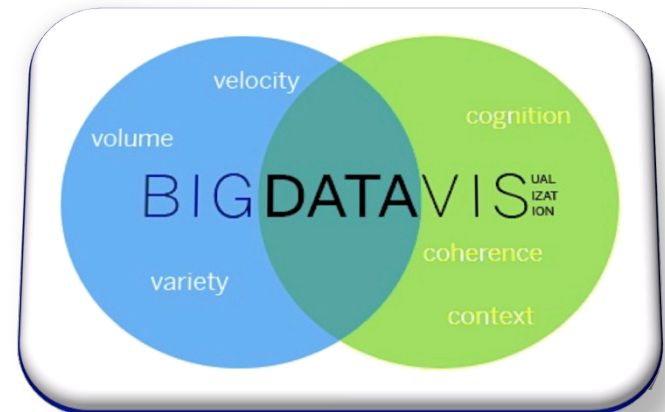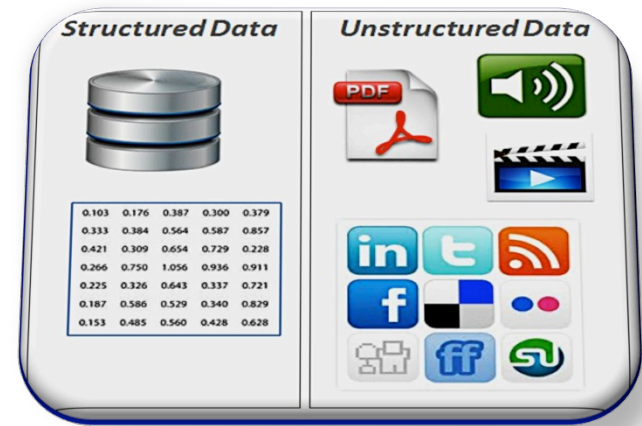

Real-time or nearly real-time information makes it possible for a company to be much more agile.

Big Data analytics are dependent on extensive storage capacity and processing power.

Real time information to understand the environment, to create new products and services.

Learning By Doing Approach

- Big data education requires joint efforts and collaboration between academia and industry.

- Industry perspective are vital as they will the ones employing the graduates of our programs.

- Identifying best practices for pedagogy to produce well prepared students for a career in the industry.

- Although selected skills as being foundational, it was recognized that a curriculum would not provide all the skills a data scientist would need for many big data projects.

Technology and Human Capital

- When talking about human capital often refer to a stock of knowledge and skills useful for production.

- Human knowledge is cumulative, the more knowledge is owned, greater the knowledge a person can achieve.

- The dynamics of economic growth is closely related to the educational levels of individuals.

- Better educated individuals are better able to discriminate between good and bad ideas, better able to solve problems, encouraging more innovation in enterprises and are more willing to assimilate innovations coming from outside.

- Knowledge is also extensive, knowledge is more productive when it operates in an ecosystem of high human level.

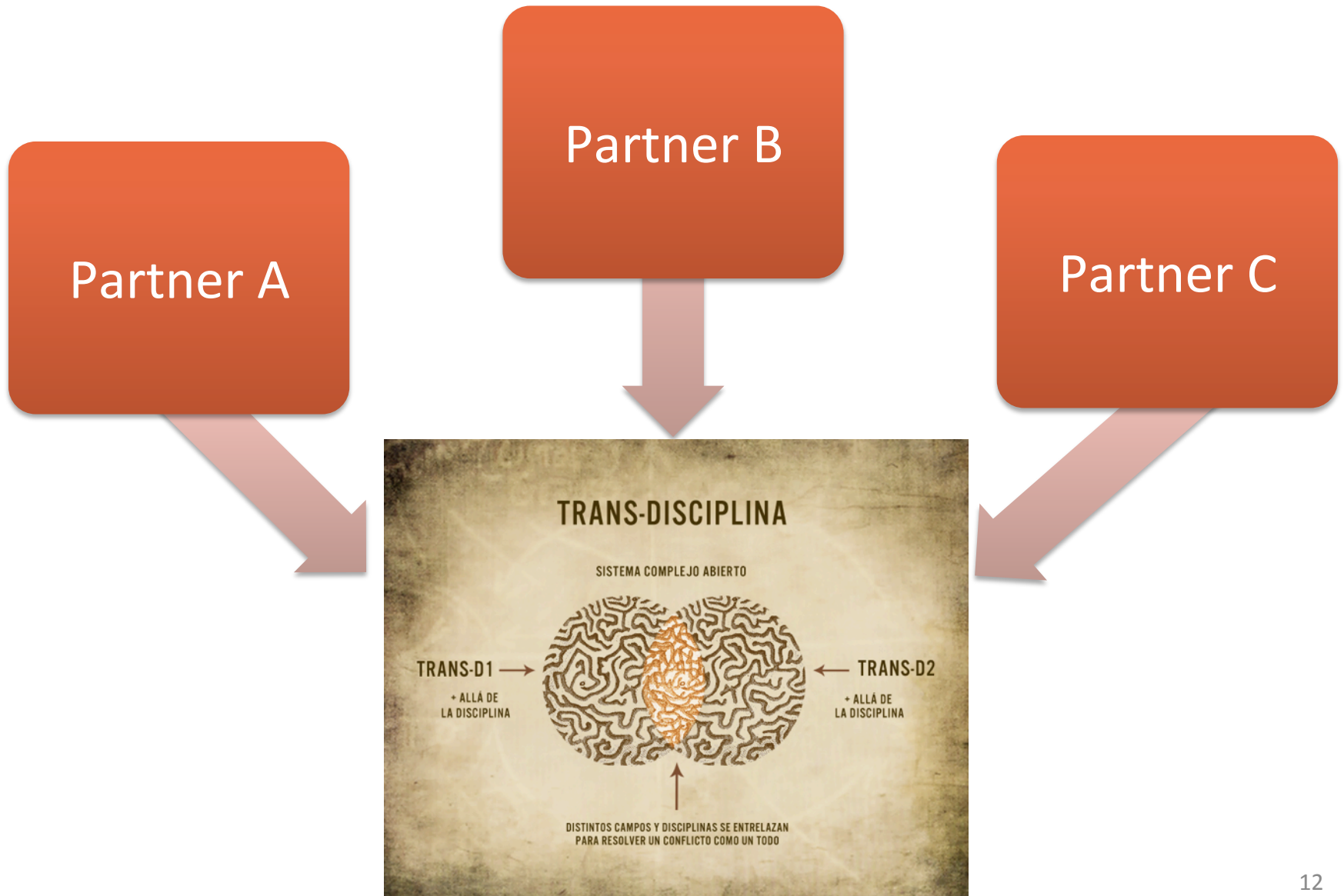**Technology and Computer Science Programs**

- Pertinence means movement.
- Strong connection with the technology: at least at the same velocity.
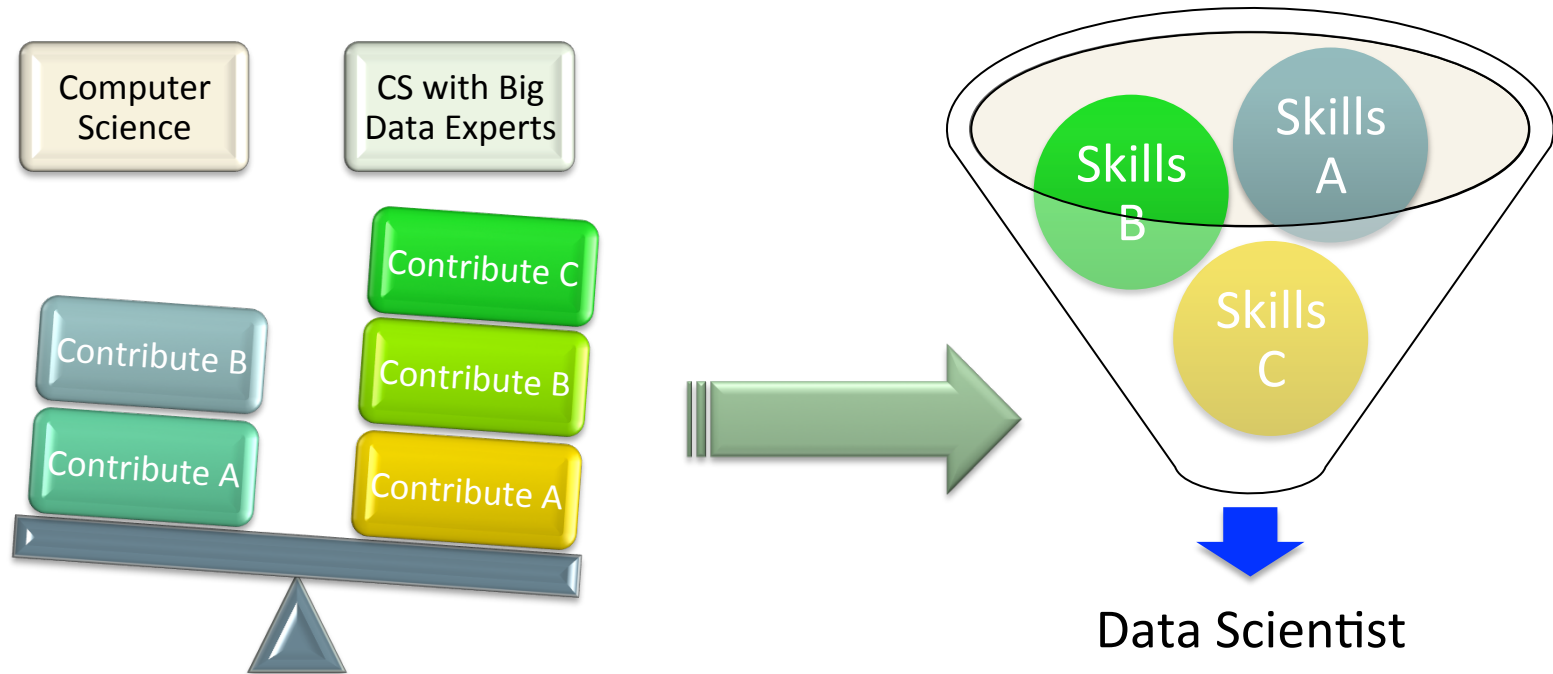- Dynamic curricula in computer science.

Required Skills of a Data Scientist

- Strong background in computer science is welcome.

- Required knowledge and skills focused on several areas including:
    - analytical, IT, business, domain, and communications,

- The analytical knowledge and skills included
    - data mining, statistics, text mining, optimization, sentiment analysis, network and graph analysis, econometrics, and predictive modeling.

- The IT knowledge and skills included
    - relational database and data warehousing, ETL, OLAP, dashboards and visualization, Hadoop and MapReduce environments, cloud computing, and all types of data management.

- Ability to take data and to be able to understand it, to process it, extract value from it, visualize it, and communicate it.

- Programming languages such as R, Java, Python and visualization tools such as D3 and GEPHI, mathematical modeling, graph and network theory, machine learning, algorithms, and data structures.

- Computer architecture, Interconnection networks, Parallel programing,

We Need Partners !!!

Partner A

Partner B

Partner C

TRANS-DISCIPLINA

SISTEMA COMPLEJO ABIERTO

TRANS-D1 →
+ ALLÁ DE
LA DISCIPLINA

← TRANS-D2
+ ALLÁ DE
LA DISCIPLINA

DISTINTOS CAMPOS Y DISCIPLINAS SE ENTRELAZAN
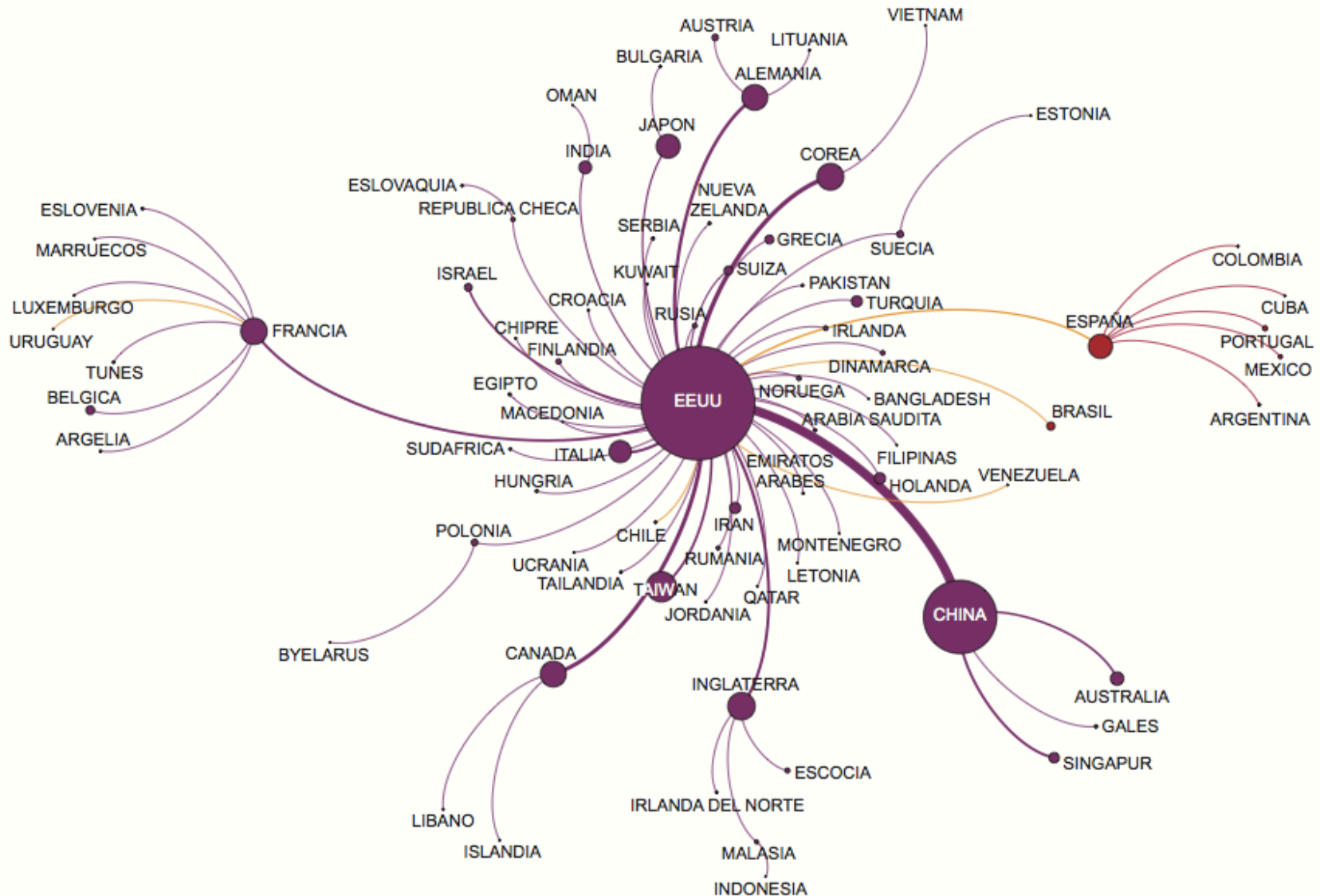PARA RESOLVER UN CONFLICTO COMO UN TODO

Programs with a Partnership

- Homogenize quality faster
- Reduce the negative impact of the lack of human resources at the postgraduate
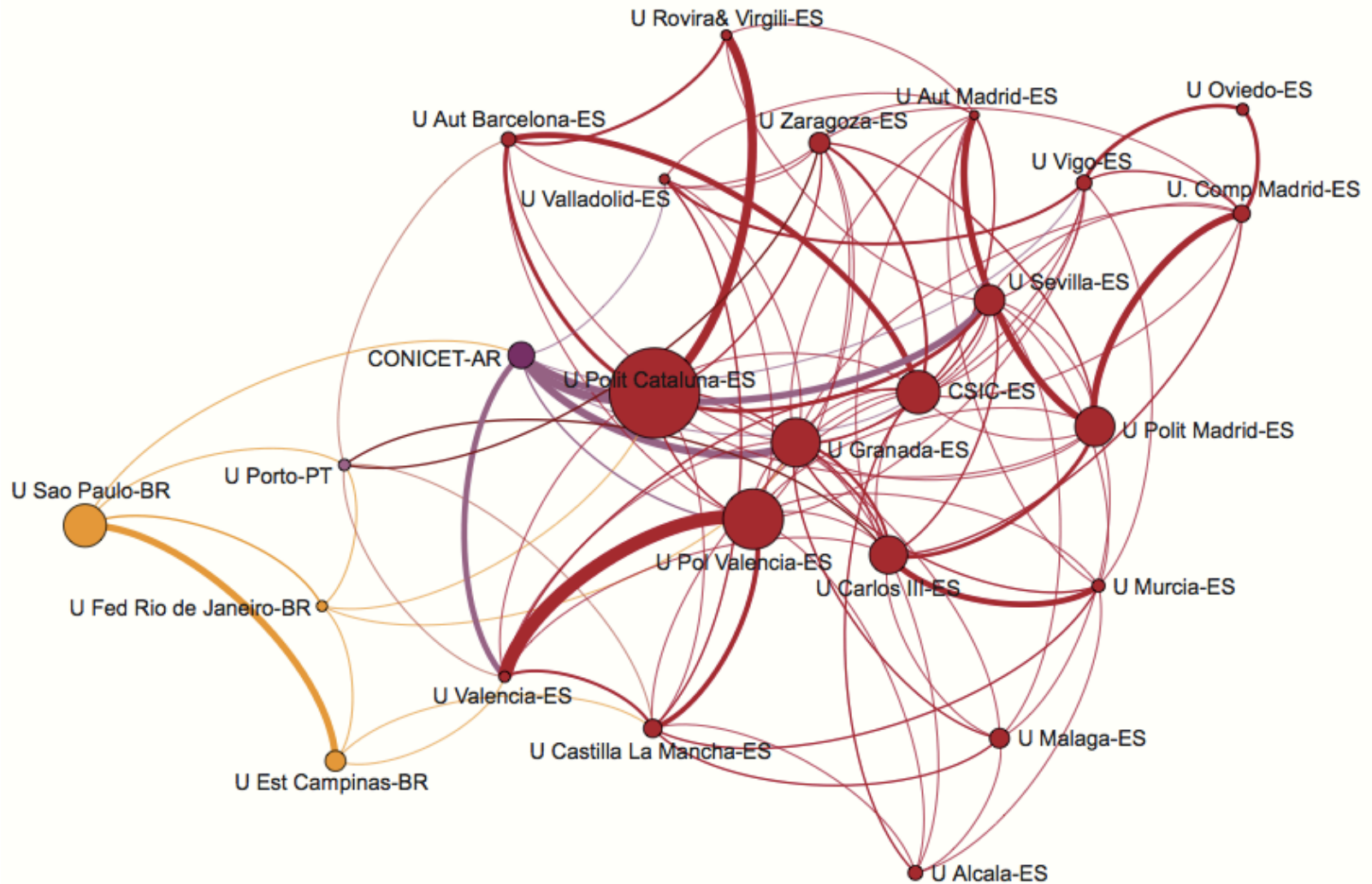- Promotes mobility students and faculty

Identifying Strategic Partners

- **Scientific Production in ICT**

Identifying Strategic Partners

- **Spain UPC is the leading university in ICT Production**
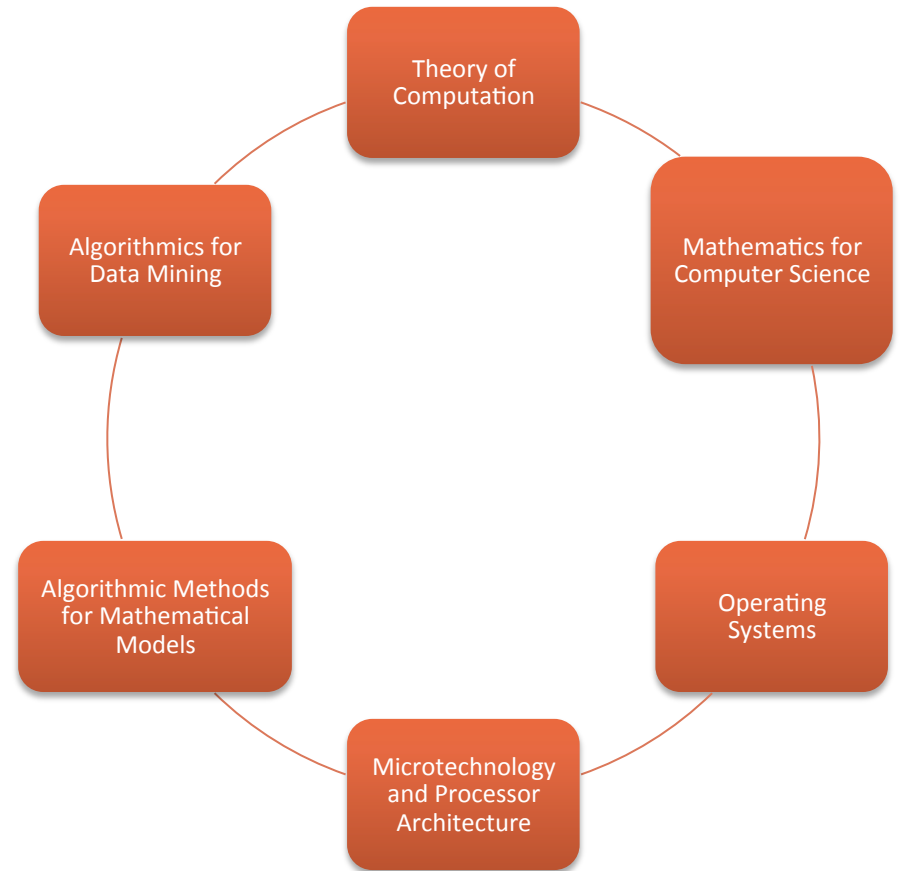
Top Ten Universities In Computer Science

- Computing Research Center at IPN began in 2010 a process of restructuring plans and programs of study, with three fundamental premises:
- International level;
- Close links with the industry sector;
- To promote the mobility of students.
- Restructuring plans and curricula
- Compare with Top Universities which offers equivalents programs.

| No.	Universidad	Programa(s) académicos
1	Harvard University	Master of Science and Master of Engineering in Computational Science and Engineering
2	Massachusetts Institute of Technology (MIT)	Master of Science (SM),
		Master of Engineering (MEng)
3	Stanford University	Master's Degree in Computer Science (MSCS)
4	Princeton University	Master of Science in Engineering (M.S.E.)
5	Oxford University	MSc in Computer Science
6	Yale University	Master of Science (MS) in Computer Science
7	University of Texas at Austin	MS in ECE w/ a concentration in Software Engineering
8	Bradford university	MS in Computing, Informatics and Media
9	California Institute of Technology (CALTECH)	MSc in Computer Science
10	Columbia University	MSc in Computer Science
11	University of California, Los Angeles	MSc in Computer Science
12	Texas A&M	MSc in Computer Science
		MSc in Computer Engineering

CORE COURSES

- Identify the core courses in these universities.

- laboratories used this study to compare our course with respect to those offered in the top universities.

Theory of Computation

Mathematics for Computer Science

Operating Systems

Microtechnology and Processor Architecture

Algorithmic Methods for Mathematical Models

Algorithmics for Data Mining

Curriculum Dynamic And State-of-the-Art

- **Core Course**. It is the minimum knowledge to be integrated into a curriculum on computer science. Any student that seeks to graduate from a master computer must necessarily take these courses.

- **Major Research Lines**. The CIC has 11 laboratories where each cultivated some of the branches of Computer Science. As part of the curriculum each laboratory define what are the main courses each student must study according to their specialty and thesis project are.

CORE COURSES

SOFT SKILL

MAJOR RESEARCH LINES LABORATORIES

INDUSTRIAL SKILLS

THESIS PROJECT
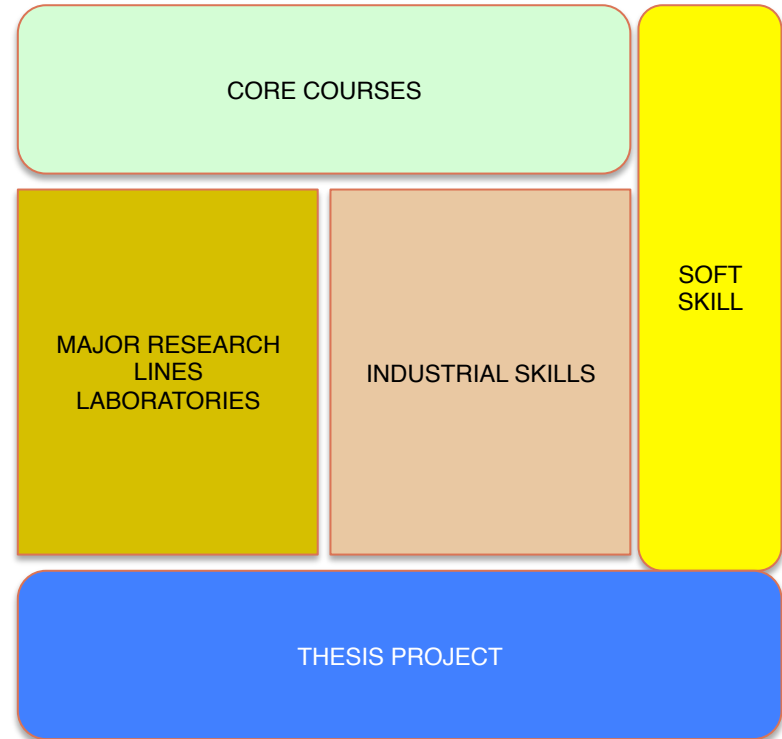
Curriculum Dynamic And State-of-the-Art

- **Industrial Skills**. As part of the curriculum, restructuring of CIC considered a number of courses that will be defined by the same ICT industry. The structure of these courses was developed with "Mexico First of CANIETI"who link us and support logistically and financially, certifying the skills required by ICT companies in Mexico.

- **Management Skills and Soft Skills**. Skills that would give them a chance to meet the demand for managerial profiles required by the industry.

CORE COURSES

SOFT SKILL

MAJOR RESEARCH LINES LABORATORIES

INDUSTRIAL SKILLS

THESIS PROJECT
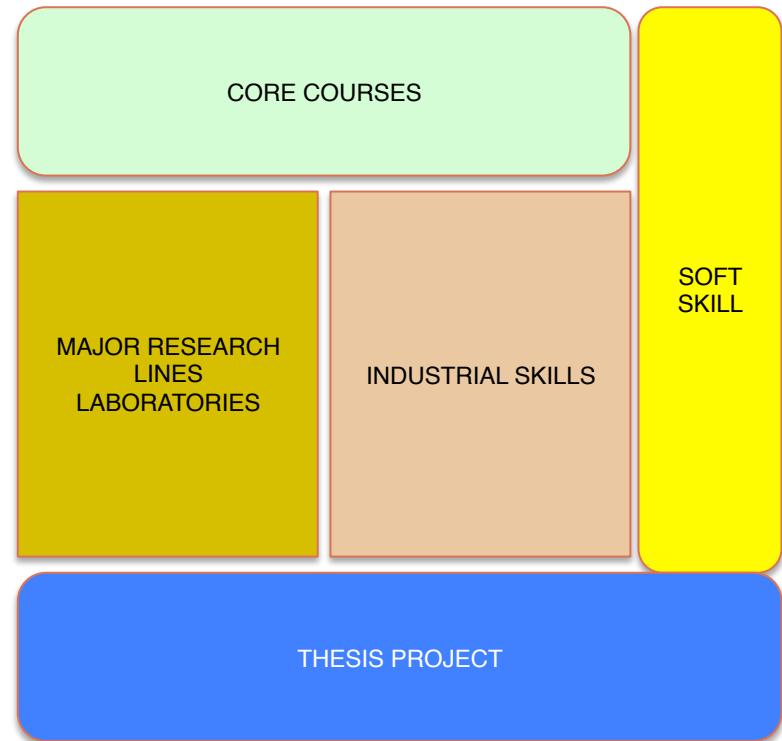
Preliminary Results of the Model

- Skills Training and Certification Required for Industry.

Educación a lo largo de la vida

2014
1216 Participantes
84 Cursos

Programa de Certificación coordinado con la Cámara Nacional de la Industria Electrónica de Telecomunicaciones y Tecnologías de la Información.

Certificaciones

2014
252 Certificaciones

- LINUX Administrator
- Social Business
- HTML 5 para móviles
- Programa Junior para Android

MEX1CO FIRST

CANIETI
www.canieti.org

20

Preliminary Results of the Model

- Interest of transnational corporations.

- Some of them are already doing recruiting students directly in our facilities.

- Highlights knowledge in Artificial Intelligence, Social Networks, Cyber Security, Embedded Systems, Data mining, Big Data.

- Certified programming tools.

Empresas
Microsoft
Intel
Oracle
Ericsson
Google
Apple
Banco de México
Bolsa de Valores México
Thales
LG

Preliminary Results of the Model

- Master double degree with the Polytechnic University of Catalonia and super computing center of Barcelona.

- The first generation of students is completing the year of stay.

- Supercomputing and Computer Architecture.
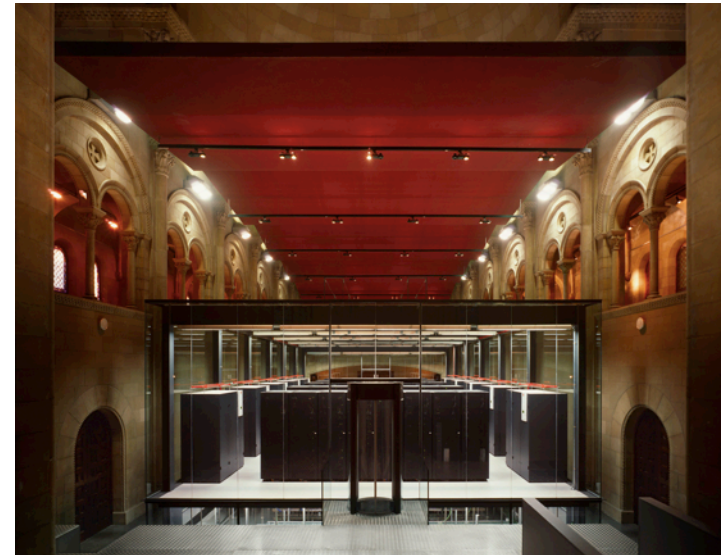
Concluding Remarks

- Successful adoption of Big Data technologies is a result of mutual enrichment between science and industry.

- Companies and professionals with the ability to work with and extract insight from the explosion of electronic information will be better aligned to make strategic decisions about their business models giving them a competitive advantage.

- Big Data technologies creates demand for continuous and dynamically adopted for preparing new specialist an training for new skills.

- There are a lot of things going on in computer science and in order to prepare student to get better jobs and higher salaries, we need curriculum dynamic and state-of-the-art.

- It was recognized that a curriculum would not provide all the skills a data scientist would need for many big data projects.

Muchas gracias