Distributed Enabling Platforms (PAD)

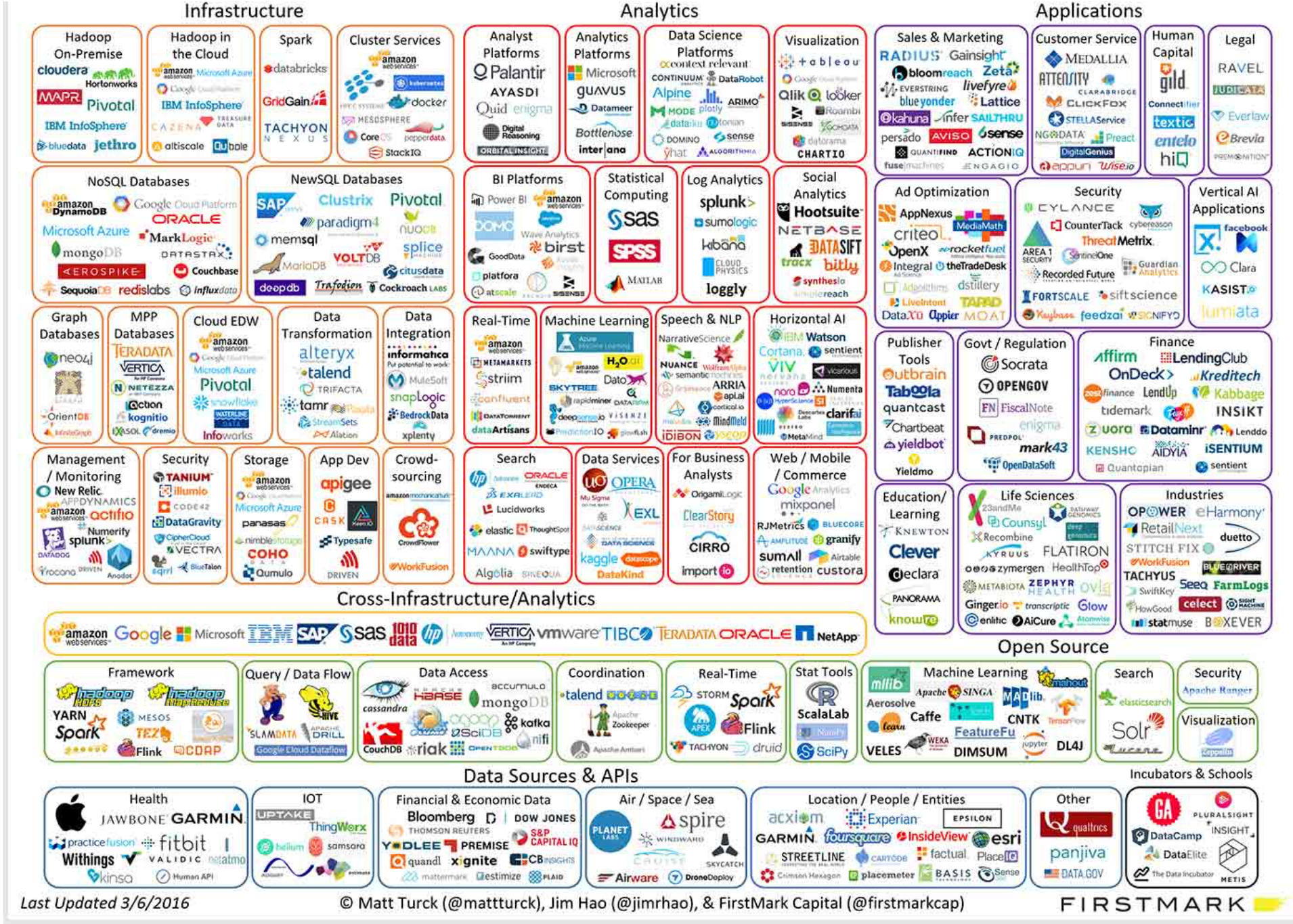
- Teacher(s) name: Nicola Tonellotto

  email: nicola.tonellotto@isti.cnr.it

  tel: 050 315 2967

  web: http://hpc.isti.cnr.it/~khast/

- Semester: 1st
- Exam mode: project + oral examination


Infrastructure
Hadoop On-Premise
Hadoop in the Cloud
Spark
Cluster Services
NoSQL Databases
NewSQL Databases
Graph Databases
MPP Databases
Cloud EDW
Data Transformation
Data Integration
Management / Monitoring
Security
Storage
App Dev
Crowd-sourcing
Analytics
Analyst Platforms
Analytics Platforms
Data Science Platforms
Visualization
BI Platforms
Statistical Computing
Log Analytics
Social Analytics
Real-Time
Machine Learning
Speech & NLP
Horizontal AI
Search
Data Services
For Business Analysts
Web / Mobile / Commerce
Cross-Infrastructure/Analytics
Applications
Sales & Marketing
Customer Service
Human Capital
Legal
Ad Optimization
Security
Vertical AI Applications
Publisher Tools
Govt / Regulation
Finance
Education/ Learning
Life Sciences
Industries
Open Source
Framework
Query / Data Flow
Data Access
Coordination
Real-Time
Stat Tools
Machine Learning
Search
Security
Visualization
Data Sources & APIs
Health
IOT
Financial & Economic Data
Air / Space / Sea
Location / People / Entities
Other
Incubators & Schools
Last Updated 3/6/2016
© Matt Turck (@mattturck), Jim Hao (@jimrhao), & FirstMark Capital (@firstmarkcap)
FIRSTMARK

Syllabus

- Design issues and solutions in very-large-scale distributed systems
- The objectives of this course are:
  - to develop an understanding of the typical issues of very large scale distributed systems;
  - to equip students with tools, best practices and common procedures to design, implement and program such systems, through understanding of algorithms and suitable theoretical models.
- List of topics:
  1. Introduction to large scale distributed systems.
  2. Cloud computing: introduction, service and deployment models, solutions.
  3. Infrastructure: virtualization, coordination, scalability, availability
  4. Programming: mapreduce model, APIs, patterns
  5. Data: data management, consistency, replication, fault tolerance.

Expected Outcomes

- Distributed Computer System Engineer
  - Analyze requirements
  - Understand design choices
  - Propose and implement solutions
- Large-scale Data Manager
  - Data warehousing
  - Analysis of solutions
- Big Data Analyst
  - Program applications to crunch terabytes of data
  - From numbers to texts to structured data

Thesis available

- Web search algorithms for efficient processing
- Distributed and replicated architectures for Web search
- Energy-efficient Web Search
- Green Information Retrieval
- Dynamic modeling of distributed systems
- Large-scale algorithms for data processing
- Efficient large-scale machine learning algorithms