

Managing Workflows in a Big Data World

Universidad Complutense, Madrid Facultad de Informática April 21, 2017

> Dr Stelios Kapetanakis University of Brighton



Speakers Intro

Dr Stelios Kapetanakis

- Associate Professor in Business Intelligence
- Head of the Knowledge Engineering Group, CEM, Brighton
- Consultant for Banking, UK Railways, US Healthcare
- Consultant for Clarksons, Airbus

• What if data could talk to us..



Contents

- Big Data
- Business Workflows
- Real-life Workflows
- Intelligent management of Workflows





Revision

byte	b	8 bits	1 byte
kilobyte	Kb	1024 bytes	1024 bytes
megabyte	MB	1024 KB	1 048 576 bytes
gigabyte	GB	1024 MB	1 073 741 824 bytes
terabyte	ТВ	1024 GB	1 099 511 627 778 bytes
petabyte	PB	1024 TB	1 125 899 906 842 624 bytes
exabyte	EB	1024 PB	1 152 921 504 606 846 976 bytes
zetabyte	ZB	1024 EB	1 180 591 620 717 411 303 424 bytes
yottabyte	YB	1024 ZB	1 280 925 819 614 629 174 706 176 bytes
brontobyte	BB	1024 YB	1 237 940 039 285 380 274 899 124 224 bytes
geopbyte	GB	1024 BB	1 267 650 600 228 229 401 496 703 205 376 bytes

www.wipro.com



Big Data is data that is too large, complex and dynamic for any conventional data tools to capture, store, manage and analyze.

The right use of Big Data allows analysts to spot trends and gives niche insights that help create value and innovation much faster than conventional methods.

0







21/04/2017

The "three V's", i.e the Volume, Variety and Velocity **O** CASE STUDY - Healthcare of the data coming in is what creates the challenge. 0 VOLUME \$300 billion is >3,500 Amount of >2,000 **Big** Data the potential NORTH annual value stored EUROPE > 250 AMERICA across the to Healthcare world (in CHINA petabytes) \$165B >400 >200 CLINICAL **JAPAN** >50 MIDDLE INDIA >50 EAST LATIN \$9B \$108B AMERICA PUBLIC 0 VARIETY VELOCITY PUBLIC HEALTH \$47B SURVEILLANCE AND RESPONSE Tube ACCOUNTS PEOPLE TO SYSTEMS MACHINE 20 PEOPLE MACHINE 2.9 50 ADVANCED FRAUD TO PEOPLE **TO MACHINE** ARCHIVES, MEDICAL DETECTION: MILLION HOURS DEVICES, DIGITAL TV. NETIZENS, VIRTUAL SENSORS, GPS DEVICES MILLION PERFORMANCE E-COMMERCE, SMART COMMUNITIES, EMAILS OF VIDEO BAR CODE SCANNERS, BASED DRUG CARDS, BANK CARDS, TWEETS SOCIAL NETWORKS. SURVEILLANCE CAMERAS, SENT EVERY UPLOADED PRICING WEB LOGS ... COMPUTERS, MOBILES.. SCIENTIFIC RESEARCH. SECOND EVERY MIN PER DAY 0 **O** VALUE 40% **PRODUCTIVITY INCREASE** SALES INCREASE PROJECTED RETAIL 49% \$9.6B GROWTH IN GLOBAL CONSULTING 39% \$5.0E DATA CREATED PER YEAR AIR TRANSPORTATION 21% \$4.3E Π CONSTRUCTION 20% \$4.2B The estimated size of the digital universe in 2011 was FOOD PRODUCTS 1.8 zettabytes. It is predicted that between 2009 and 20% \$3.4B 2020, this will grow 44 fold to 35 zettabytes per year. STEEL O A well defined data management strategy is essential 20% \$3.48 to successfully utilize Big Data. AUTOMOBILE 19% INDUSTRIAL INSTRUMENTS 18%) S1. Sources - O Reaping the Rewards of Big Data - Wipro Report @ Big Data: The Next Frontier for Innovation, Competition and Productivity - McKinsey Global Inditate Report @comScore, Radicati Group @ Measuring the Business Impacts of Effective Data - study by University of Texas, Austin @US Department of Labour. PUBLISHING 18% \$0.8 DO BUSINESS BETTER TELECOMMUNICATIONS 17% \$0.4B NYSE WIT I OVER 130,000 EMPLOYEES I 54 COUNTRIES I CONSULTING I SYSTEM INTEGRATION I OUTSOURCING

Complutense, Madrid



0

TRANSPARENCY IN

CLINICAL DATA AND

CLINICAL DECISION

AGGREGATION OF

PATIENT RECORDS.

ONLINE PLATFORMS

AND COMMUNITIES

SUPPORT

\$5B

BUSINESS

RESEARCH AND

DEVELOPMENT:

PERSONALIZED

CLINICAL TRIAL

5%

PROJECTED

IN GLOBAL IT

WIPRO

GROWTH

SPENDING

PER YEAR

MEDICINE;

DESIGN

MODEL

Source: technophenia.com



Big Data what is really..

- Big Data is a marketing term
- Since 2012 everybody has.. some..
- Big Data assumes that bigger is better
- "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization"

Gartner, 2012





Traditional Workflows





Automated Workflows





Managing workflows

Categories

- Flow control
- In-transit visibility
- Processes

Components

- Input(s)
- Transformations
- Outputs



• Expert systems



Workflow Management Systems

- Over the last years standards have been developed in enterprise (W3C standards, BPEL, BPMN, XML, XPDL)
- Recipe of success:
 - Broadly used.. Manually used..
 - Different actors, systems, aims..
 - Coordinate their actions towards a set target
 - Large volumes of data
- Need for intelligent systems
 - Assist with the management
 - Make desirable, feasible DSS
 - Identify volumes and repeated patterns
 - Give ground to AI technologies and in particular CBR



Challenges

Human – related workflows

- Uncertainty
- Inconsistency
- Incompleteness
- Large volumes of workflow instances





Big Data -> Big Problems



• The problems really start when you **reach back into the tree and change something**; for example, a new business strategy is defined which invalidates the interpretation and everything downstream of that





Intelligent Workflow Monitoring

- Data Mining techniques
- Machine Learning
- Artificial Intelligence
- Case-based Reasoning



Workflows as cases









CBR Mechanics



Source: Eremeev &. Vagin, 2011





Time Mechanics

• GTT: General Time Theory [Ma & Knight, 1994]: Improved General Theory of Time



• Maximum Common Sub-graph [D. Conte et al., 2007]

$$S(G,G') = \frac{(\sum_{\substack{matches \\ C,C' \\ in \\ MCSG}} \sigma(C,C'))^2}{count(G).count(G')}$$



Case study: UK Rail Industry

Q: How train operational data could be used in order to provide information/insights of trains' behaviour and performance?

• Visualise RCM data to understand the granularity of delays

 Understand how trains reacted over signals, system failures and other various unexpected situations.





British National Rail

- Case & Solution Architecture
- Process
- Implementation time
 4 years (4 phases)
- Future perspectives
 - Now on phase 3
 - 4 phases (Visualisation,
 - .. Monitoring, Data Mining, Simulation)





Real Challenges

- Data silos (dlis, las, lis)
- Application silos (apis)
- Library style data management
- Project, corporate, or master?
- Never fixing the data (gps)
- Big data vs "lots of data"
- Decide by PowerPoint
- We do what everybody else does



The End



Thank you!

Contact Details

Dr Stelios Kapetanakis

email: s.kapetanakis@brighton.ac.uk linkedin: stelios-kapetanakis-ba816440





References

- [Agorigianitis, 2017]: Agorgianitis, I., Kapetanakis, S., Petridis, M., Fish, A. (2017) Business Process Workflow Monitoring using Distributed CBR with GPU Computing, In the 30th International FLAIRS Conference, Florida, Marco Island, pp.48-57, May 2017
- [Chiu, 2008]: Chiu, D. (2008), Business Process and Workflow Management
- [D. Conte et al., 2007]: Conte, D., Foggia, P., Vento, M., (2007) Challenging Complexity of Maximum Common Subgraph Detection Algorithms: A Performance Analysis of Three Algorithms on a Wide Database of Graphs, Journal of Graph Algorithms and Applications, 11(1) pp. 99–143
- [Eremeev &Vagin, 2011]: Eremeev, A.P., Vagin, V. N. (2011) Common Sense Reasoning in Diagnostic Systems, Efficient Decision Support Systems - Practice and Challenges From Current to Future, Prof. Chiang Jao (Ed.), ISBN: 978-953-307-326-2
- [Ma & Knight, 1994]: Ma, J., Knight, B. (1994) A General Temporal Theory, The Computer Journal, Vol.37(2), 114-123, 1994.