Cloud & Big Data opportunities and risks

Patrick Valduriez











Cloud & Big Data: the hype!

TECHTOUR CLOUD & BIG DATA SUMMIT NOVEMBER 21 - 22, 2012



Behind the Hype?



Cloud & Big Data: a perfect marriage?

• But do they have the same directions ?



A Marriage of Convenience?



- Cloud and big data have different goals
 - Big data aims at added value and operational performance
 - Cloud targets flexibility and reduced cost
- But they can help each other by
 - 1. Encouraging organizations to outsource more and more strategic internal data in the cloud
 - 2. Get value out of it, e.g. by integrating their data with external data, through big data analytics at affordable cost

Cloud & Big Data: *big risks*

• Big data

- Complex data management landscape
- Big data can go bad
- Many others
- Cloud
 - Cloud can go bad
 - Data safety
 - Data privacy and security
 - Environmental impact
 - Energy consumption, cooling, etc.
- Combining cloud & big data magnifies the risks

Big Data Challenge



• The needs are in technology

- New architectures, algorithms, techniques
- AND technical skills
 - Experts in using the new technology and dealing with big data, e.g. big data scientists

Big Data Landscape



Copyright © 2012 Dave Feinleib

dave@vcdave.com

blogs.forbes.com/davefeinleib

When Big Data goes bad



November 5, 2013: 1:00 PM ET





How the models underlying today's supercomputing prowess are costing us its success.

By Joshua Klein



When Big Data goes bad - 1

• Excerpts:

Peter Lawrence's The Making of a Fly, a classic book in developmental biology, was listed on Amazon.com as having 17 copies for sale: 15 used from \$35.54, and two new from

Problem: over simplified models, but reality is complex!

profnath would raise their price to 0.9983 times bordeebook's listed price. Several hours later, bordeebook would increase their price to 1.270589 times profnath's latest amount.

When Big Data goes bad – 2

• Excerpts:

One t-shirt seller on Amazon.co.uk put up a shirt for sale emblazoned with the statement, "Keep Calm and Rape a Lot."

Problem: context-independent model, but context does matter!

"Keep Calm and Carry On," which enjoyed a brief mimetic popularity online) to make derivations that get dropped onto a template of a T-shirt and automatically get posted as an Amazon item for sale.

Their mistake was overlooking a single word in a list of 4,000 or so others.

When Big Data goes bad (for me)



When Big Data goes bad (for me)



 Object Technology de Mokrane Bouzeghoub, Georges Gardarin et Patrick Valduriez (5 juin 1997)

 EUR 1,46 d'occasion (5 offres)
 Livres anglais et étrangers: Voir l'ensemble des 22 articles

Problem: how do I get complete deletion of wrong information?





Ozzy Osbourne. Fucking Mad. Die Story zu seinen Songs. de Patrick Valduriez (30 juin 2003)

EUR 21,46 neuf (1 offre) EUR 6,26 d'occasion (2 offres) Livres anglais et étrangers: Voir l'ensemble des 22 articles

When Cloud goes bad

TECH 10/07/2013 @ 9:31PM 40 481 views

The NSA's Hugely Expensive Utah Data Center Has Major Electrical Problems And Basically Isn't Working

b 🌒 🗞 を 💿 🙁 🕕 21 comments, 10 called-out

+ Comment Now + Follow Comments

Well, this is good news for those with privacy concerns about the NSA and terrible news for those concerned about government spending.





When Cloud goes bad

• Excerpts:

Well, this is good news for those with privacy concerns about the NSA and terrible news for those concerned about government spending. The National Security Agency's new billion-dollar-plus data center in Bluffdale, Utah was supposed to go online in September, but the Wall Street Journal's Siobhan Gorman reports that it has major electrical problems and that the facility known as "the country's biggest spy center" is presently nearly unusable.

"The problem, and we all know it, is that they put the appliances too close together," a person familar with the database construction told FORBES, describing the arcs as creating "kill zones." "They used wiring that's not adequate to the task. We all talked about the fact that it wasn't going to work."

Cloud Safety





Cloud Computing Enterprise

Amazon's Cloud Crash Disaster **Permanently Destroyed Many Customers'** Data

HENRY BLODGET APR. 28, 2011, 7:10 AM 6 92,357 **P75**

Data Privacy: risks

Loss of control over data

- Data can be compromised by the cloud provider or third parties
 - In particular, using big analytics (e.g. data mining)
- Lack of transparency about where is the data
 - And for reliability reasons, data is replicated across sites
- Dependence on the cloud provider
 - SLAs typically for service operation, not privacy
 - And providers subcontract providers, in different countries
- Traditional risk management impossible
 - Two many participants: customers, internet providers, cloud providers, subcontractors
- Regulations
 - Will always lag behind technology and services
 - Anyone heard of french HADOPI law?
 - And how to enforce regulations at internet scale?

Where Can Regulation Help?

- Impose guarantees by cloud providers on customer data (in addition to QoS)
 - Safety, privacy, complete deletion
 - Logging and auditing
 - Transparency wrt where the data is stored
- Balance the power of cloud providers
 - To limit dominant positions
- Promote interoperability between cloud providers
 - Standard protocols
 - Open source software
- Promote users education and the "right to know"
 - Both legal and technical knowledge

What's next?

