

“Sistemas de recomendação,
reflexões e **interação** em
tempos de Big Data”

Aula Inaugural

MARTA MATTOSO

Programa de Sistemas e Computação
COPPE, UFRJ

Organização

- Boas Vindas
- Apresentação
- Recomendações, reflexões e interação em tempos de Big Data
- Contribuição em Pesquisa
- Conclusão

**Boas vindas ao PPCIC
CEFET-RJ !!!**



Trajetória em Equipe

Seminário do Grupo de Pesquisas em Distribuição e Paralelismo em Dados

Agosto de 2014, COPPE/UFRJ

Apresentação e financiamento

- Bolsista PQ CNPq, desde 1996, atualmente no nível **1B**
- Cientista do Nosso Estado pela FAPERJ desde 2012
- Coordenação de 7 projetos de pesquisa na faixa > R\$100.000
- 30 artigos em periódicos internacionais indexados todos no estrato superior do *Qualis* CC (A1, A2 e B1)
- 58 dissertações de mestrado e 18 de doutorado
- Paraninfa e Homenageada na Engenharia de Computação e Informação- UFRJ ('09, '11, '12, 15)
- Homenageada pela CE-BD da SBC em 2005
- Homenageada pelas mulheres da CE-BD (WomBD) em 2014



Universidade Federal
do Rio de Janeiro
Escola Politécnica



Sociedade Brasileira
de Computação



**RECOMENDAÇÕES, REFLEXÕES E
INTERAÇÃO EM TEMPOS DE
“BIG DATA”**

Data Scientist: The Sexiest Job of the 21st Century

[Thomas H. Davenport](#) [D.J. Patil](#)

From the October 2012 Issue





- “The **first** one is one size no longer fits all, and that is going to give heartburn to the current [relational database] elephants.
- The **second** one is "big data" means three different things and you've got to remember which one you're talking about.
- The **third** one is that ACID [atomicity, consistency, isolation, durability] is a really good idea, so don't throw the baby out with the bathwater.
- The **fourth** thing is, Think memory. It's the new disk.
- The **last** thing is that -- in stages -- the cloud is really the answer to save money.”

What do you think are some of the key trends shaping the world of data management in **2012**?

Mineração de Dados

Marta Mattoso FOAF Follow

ALIAS: Marta L. Queiros Mattoso

Associate Professor
COPPE / Computer Science Federal University of Rio de Janeiro
Programa de Sistemas, COPPE/UFRJ Caixa Postal 68511 21941-972 Rio de Janeiro - RJ Brazil

P +55-21-2562-8694
F +55-21-2290-8626
marta@cos.ufrj.br

[EDIT AVATAR](#)

[EDIT PROFILE](#)

Share to: 0

Marta-Mattoso's Ego Community: [Rebuild](#)

Expertise:

- Reutilisation de I / Une application (15)
- XML Data (10)
- Web Data / Database design (7)
- Grid Computing / Parallel Data (6)
- Parallel algorithm (5)
- Coordination model / Applications Track Editorial (5)

Re-calculate Statistics ?

Bio Click [Here](#) to add Bio information.

Education Click [Here](#) to add Education information.

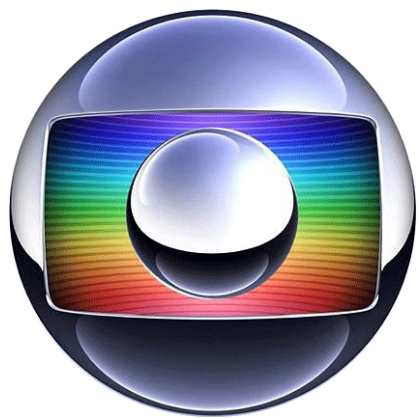
Research Interest

Web Services Web Search Scientific Workflows

Activity **Citations** **H-index** **G-index** **Diversity** **Sociability**

Análise de rede social acadêmica: <http://aminer.org/>

Novos modelos de negócio



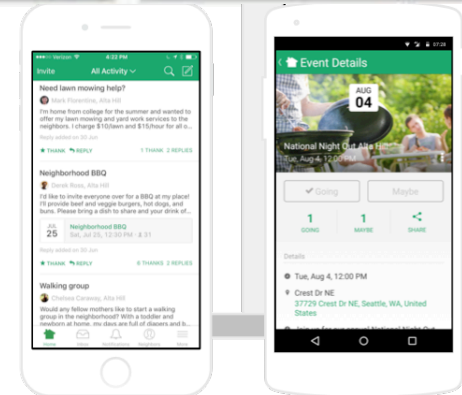
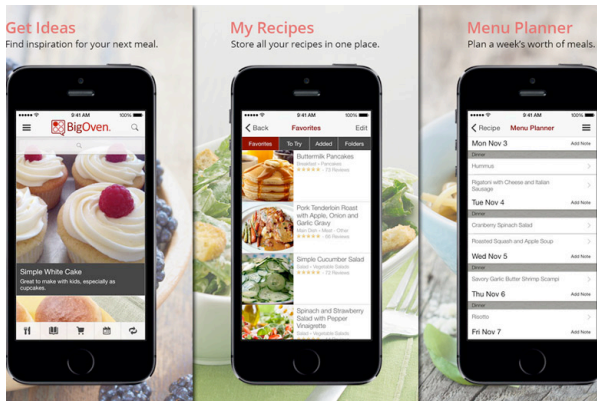
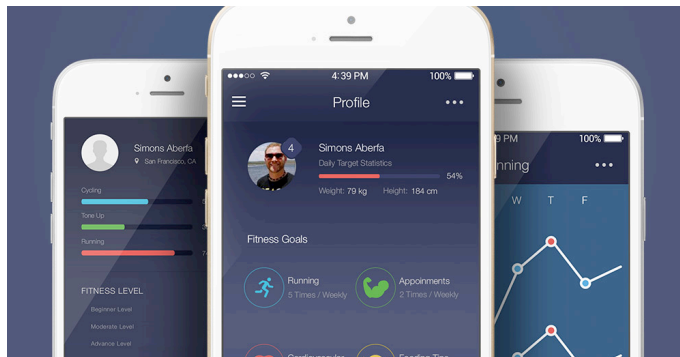
30 segundos: 400.000

Nichos 20.000

R\$ 1.000

R\$ 20.000.000





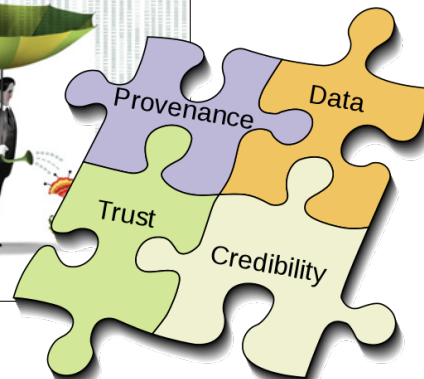
Para onde vai uma sociedade guiada por dados ?

Felix Naumann, "Reflexions of a data skeptic"

Mother's Home Cooking vs Street Food

JP Rangaswami, CDO @ Deutsche Bank

- Devices provide information:
 - from an ever more diverse range of sources, via ever more sensor types
 - that measures ever more of everything
 - that can be mashed-up in unforeseen ways



IPAW 2016 Keynote - Luc Moreau "Retrospect and Prospect of Provenance"

Proveniência

Proveniência de dados

- O termo proveniência de dados (*data provenance*) se refere aos dados que deram origem a dados resultantes de uma operação (Buneman et al. 2000)
- Também conhecido como linhagem de dados (*data lineage, lineage*) (Melton et al., 1995), descreve a história de processamento de um conjunto dados
- Bose & Frew (2005) identificam outros termos que aparecem na literatura que poderiam ser interpretados como proveniência de dados: *afiliação, genealogia de dados, dependência entre conjuntos de dados, arqueologia de dados, rastro de auditoria e histórico de derivação.*

Proveniência de dados

- O termo proveniência de dados (*data provenance*) se refere a fontes de consultas ou a serviços baseados no

“dado que descreve dado” é metadado

Proveniência descreve dado

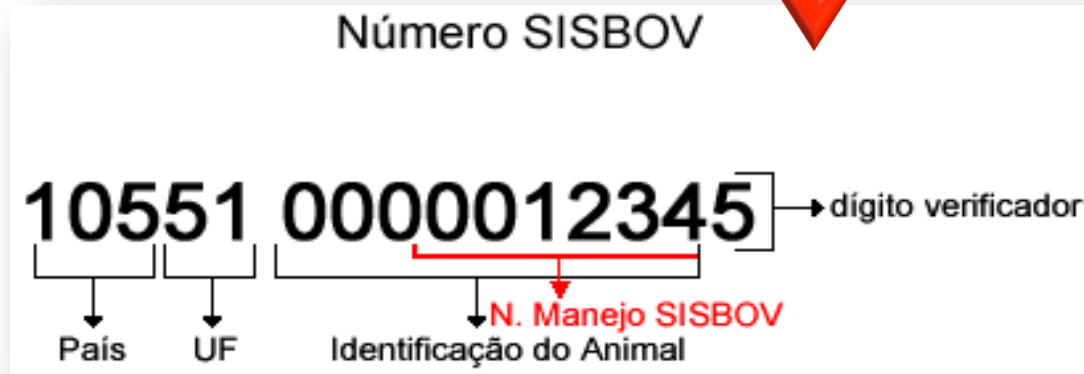
Logo...

Proveniência é metadado.

Como proveniência de dados: *afiliação, genealogia de dados, dependência entre conjuntos de dados, arqueologia de dados, rastro de auditoria e histórico de derivação.*



Fonte: www.abczcertificadora.com.br/identificacao.php



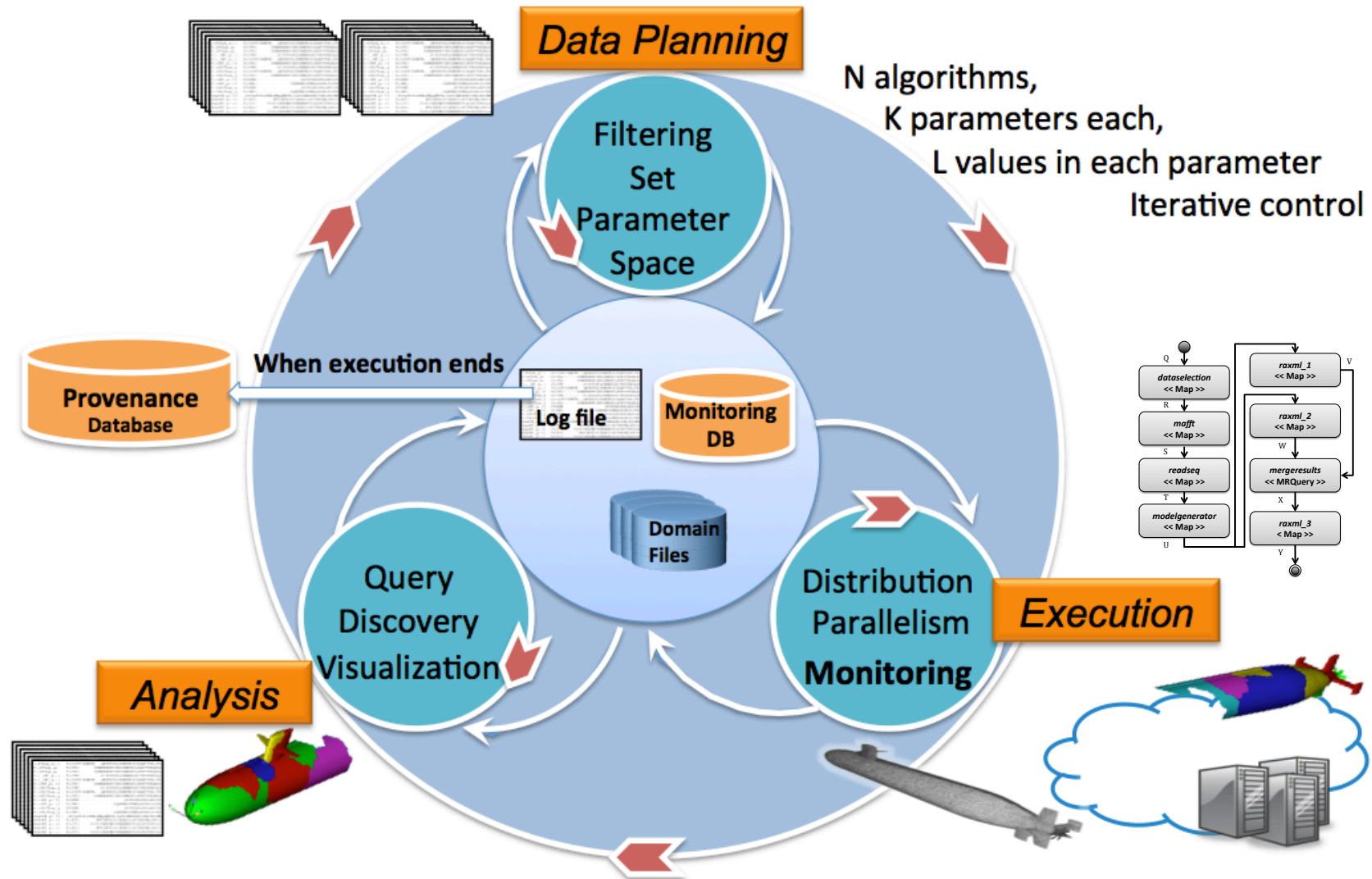
Proveniência no dia-a-dia



Importância da proveniência na Ciência

- **Identifica os dados** de entrada e saída de um experimento com sua origem
- Permite **verificar se um experimento foi realizado de acordo com procedimentos científicos** aceitáveis
- Contribui para assegurar a **qualidade dos dados**
- Útil **Proveniência é tão (ou mais) importante quanto os resultados** (Davidson, Freire, 2008)
- Útil **eventos** usados na produção de um resultado
- **Rastreio** de quem executou o experimento e quem foi o responsável pelos resultados (patente)

Exploratory experimentation



Algumas publicações

- “Dynamic steering of HPC scientific workflows: A survey”.
Future Generation Computer Systems (2015)
Mattoso, Marta ; Dias, J. ; Ocaña, K. ; Ogasawara, Eduardo ; Costa, F. ; Horta, F. ; Silva, Vitor ; Oliveira, D.
- “An Algebraic Approach for Data-Centric Scientific Workflows”,
PVLDB Journal (2011)
Ogasawara, E., Dias, J., Oliveira, D., Porto, F., Valduriez, P., Mattoso, Marta,
- “Data-centric iteration in dynamic workflows”,
Future Generation Computer Systems (2015)
Dias, J., Guerra, G., Rochinha, F.A., Coutinho, A.L.G.A., Valduriez, P., Mattoso, Marta

CONTRIBUIÇÃO EM PESQUISA – O QUE É ?

Contribuição em pesquisa

- Diferente da graduação
- Trabalho original
- Conhecer a literatura
- Conhecer os principais trabalhos e autores
- Resolver um problema real

Diferente da graduação

Não basta passar nas provas
Não há um roteiro a ser seguido
Aprovação da tese:



- Avaliação do texto da tese
- Avaliação da defesa oral
 - Ato formal
 - Apresentação da solução
 - Exposição oral

Contribuições dissertação - mestrado

- Contribuição
 - Resultados úteis para outras pessoas (passíveis de generalização)
 - Resultados com novidade, isto é, não dariam para ser afirmados sem o desenvolvimento do trabalho

Contribuições dissertação - mestrado

- O que não tem contribuição
 - Levantamento do estado da arte de uma área de pesquisa e comparação analítica das propostas existentes
 - Desenvolvimento de uma aplicação convencional sobre um software comercial
 - Desenvolvimento de um protótipo com técnicas já consagradas

Análise das soluções existentes

- **Fazendo buscas na Web**

(adaptado de Justin Zobel, [Writing for Computer Science](#) (Springer, third edition, 2015))

- “Use general-purpose search engines and special-purpose repositories to search for titles and abstracts of likely papers.
- Identify authors who publish in reputable journals and conferences, then look for their home pages.
 - See their technical reports
- **Ferramentas de busca**
 - “Google Academic”
 - Reputação de veículos (SciMago, Google, JCR)

Stonebraker's advice



- For **undergraduates** just getting into computer science, *"Learn how to code, and code well, because whatever you do is going to involve implementation."*
- For **Ph.D. students** trying to figure out where to focus their attention, he suggests talking to real-world computer users. *"They're happy to tell you why they don't like or do like any given technology, so they're a wonderful source of problems to work on."*

“Sistemas de recomendação,
reflexões e **interação** em
tempos de Big Data”

Obrigada!

MARTA MATTOSO

Programa de Sistemas e Computação
COPPE, UFRJ