

# B D y CAR/D

Fernando G. Tinetti

III-LIDI, Fac. de Informática, UNLP  
Comisión de Inv. Científicas, Bs. As.

[fernando@info.unlp.edu.ar](mailto:fernando@info.unlp.edu.ar)



# Teniendo en Cuenta la Invitación

"... desarrollar una hoja de ruta para ... estrategia de mediano y largo plazo ... nacionales"

+

"... presentar (su visión, ideas, propuestas o requerimientos en relación con Big Data"

---

Esta presentación

# Definición/es de Big Data

- Siempre subyacente: datos --> información
- Big Data... ¿más información?
- Definiciones:
  - Las 3/4 V: volumen, velocidad, variedad /veracidad<sup>(1)</sup>
  - Procesamiento de al menos  $10^{18}$  bytes (ExaB)<sup>(2)</sup>
  - Algunas centenas de MB, de entre 5 y 20 fuentes<sup>(3)</sup>
  - Todas parecen enfocadas en los datos, no en la utilidad de los mismos ... --> información, def.<sup>(\*)</sup>

# Big Data: Aspectos Relacionados

- Escala: datos sin precedentes<sup>(4)</sup>
  - Empleados (*data entry*) --> Usuarios --> máq...



<http://www.capocosmico.blogspot.com.ar/2012/11/monitoreando.html>

# Big Data: Aspectos Relacionados

- Escala: datos sin precedentes<sup>(4)</sup>
  - Empleados (*data entry*) --> Usuarios --> máq...
- Con/Semi/Sin estructura
- Captura, almacenamiento, transporte...
- Disponibilidad... ¿abierta? No es solo técnico
  - ¿Quién **tiene** los datos? ¿dónde?
- Procesamiento... transformación-->Utilización

# Procesamiento

- Múltiples tipos de tareas: clasificación
- ¿Por niveles de abstracción?<sup>(5)</sup>
  - Menor nivel --> más “reusable”, menos datos, más cercana a la fuente, quizás menos requerimientos en términos de  $O( )$
  - Mayor nivel-->mayor alcance, aplicación y usualmente también mayores requerimientos
- ¿Por área de aplicación? (EEUU – Departamentos<sup>(6-7)</sup>)
- **Desde CAR/D:** técnicas de optimización, distribución y paralelización, ¿I/O Computing?<sup>(8)</sup> – recursos
- Big Data no *existe* para procesamiento ni almacenamiento sino para resolver problemas (descubrir, revelar, generar inf.)

# Estrategias...

- Educación: planes y estudiantes (McKinsey<sup>(9)</sup>: en EEUU 140-190K por debajo para 2018)
- Ni “huir hacia adelante” ni “requisito: todo bien”
- Expertos: del área de aplicación + del área de conocimiento + **del área de CAR/D**
- Por problema, por disponibilidad de datos
- Datos a disposición --> creatividad
- Problemas: cuáles *podemos* y cuáles *tenemos*
- Definición en cuanto a datos por área

# Estrategias para CAR/D...

- Varias similares a las anteriores, *adaptadas*
- Promoción
  - Grados - Postgrados
  - Investigadores
  - Líneas de financiación
- Promover no siempre es “poner” ... ¿organizar?
- ¿Instalaciones de hardware? ¿SNCAD?
- Parafraseando: “si uds. lo llaman fantasear... me costaría discutirlo...” (pero en el *buen* sentido)



Muchas gracias

# Referencias

- <sup>(1)</sup> [http://en.wikipedia.org/wiki/Big\\_data#Definition](http://en.wikipedia.org/wiki/Big_data#Definition) y referencias (Ag. 2013):
  - Snijders, C., Matzat, U., & Reips, U.-D. (2012). 'Big Data': Big gaps of knowledge in the field of Internet. International Journal of Internet Science, 7, 1-5. [http://www.ijis.net/ijis7\\_1/ijis7\\_1\\_editorial.html](http://www.ijis.net/ijis7_1/ijis7_1_editorial.html)
  - Hogan, M. (3 Jan 2013). "Large Databases"
  - Douglas, Laney. "3D Data Management: Controlling Data Volume, Velocity and Variety". Gartner. Retrieved 6 February 2001.
  - Beyer, Mark. "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data". Gartner. Archived from the original on 10 July 2011. Retrieved 13 July 2011.
  - Douglas, Laney. "The Importance of 'Big Data': A Definition". Gartner. Retrieved 21 June 2012.
  - "What is Big Data?". Villanova University
  - Delort P., Big data Paris 2013 <http://www.andsi.fr/tag/dsi-big-data/>
  - Delort P., Big Data car Low-Density Data ? La faible densité en information comme facteur discriminant <http://lecercle.lesechos.fr/entrepreneur/tendances-innovation/221169222/big-data-low-density-data-faible-densite-information-com>
- <sup>(2)</sup> S. Kaisler, F. Armour, J. A. Espinosa, W. Money, Big Data: Issues and Challenges Moving Forward, 2013 46th Hawaii International Conference on System Sciences, (Congreso Clase "A"), <http://www.hicss.hawaii.edu/>
- <sup>(3)</sup> [http://www.customerthink.com/blog/my\\_datas\\_not\\_big\\_are\\_you\\_sure](http://www.customerthink.com/blog/my_datas_not_big_are_you_sure)
- <sup>(4)</sup> <http://www.youtube.com/watch?v=c4BwefH5Ve8>
- <sup>(5)</sup> Troy Blackman The 8 States of Big Data, September 2012, <http://www.altairci.com/the-8-states-of-big-data>
- <sup>(6)</sup> Executive Office of the President, Big Data Across the Federal Government, March 29, 2012, [http://www.whitehouse.gov/sites/default/files/microsites/ostp/big\\_data\\_fact\\_sheet\\_final\\_1.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final_1.pdf)
- <sup>(7)</sup> Unleashing the Power of Big Data, <http://www.whitehouse.gov/blog/2013/04/18/unleashing-power-big-data>
- <sup>(8)</sup> David Floyer, Real-time IO Centric Processing for Big Data, [http://wikibon.org/wiki/v/Real-time\\_IO\\_Centric\\_Processing\\_for\\_Big\\_Data](http://wikibon.org/wiki/v/Real-time_IO_Centric_Processing_for_Big_Data)
- <sup>(9)</sup> McKinsey Global Institute, Big data: The next frontier for innovation, competition, and productivity, [http://www.mckinsey.com/insights/business\\_technology/](http://www.mckinsey.com/insights/business_technology/)

# Otras Referencias

- <http://www.forbes.com/sites/davefeinleib/2012/07/24/big-data-trends/>
- <http://www.c3dmw.com/IRDC3DMW/DataMining1.jsp>
- <http://www.whitehouse.gov/blog/2013/04/18/unleashing-power-big-data>
- [WhiteHouse.gov/OSTP](http://www.whitehouse.gov/OSTP)
- <http://www.washingtonpost.com/wp-srv/special/politics/prism-collection-documents/>
- <http://www.youtube.com/watch?v=J1EKvWot-3c>
- <http://blog.sqrrl.com/post/46306669352/sqrrls-take-on-the-big-data-ecosystem>
- [http://www.sas.com/offices/latinamerica/argentina/noticias/notas\\_prensa/encuesta-estrategia-de-big-data-es-critica.html](http://www.sas.com/offices/latinamerica/argentina/noticias/notas_prensa/encuesta-estrategia-de-big-data-es-critica.html)
- <http://www.managementthinking.eiu.com/big-data.html>
- <http://www.scientificamerican.com/article.cfm?id=foldit-gamers-solve-riddle>
- <http://www.medgadget.com/2012/01/another-success-for-foldit-gamers-increase-enzyme-activity-by-a-factor-of-18.html>
- <http://www.gamesforchange.org/play/foldit/>
- <http://www.theguardian.com/world/2013/aug/23/nsa-prism-costs-tech-companies-paid>
- <http://www-01.ibm.com/software/data/bigdata/>
- [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)
- [http://www.webopedia.com/TERM/B/big\\_data.html](http://www.webopedia.com/TERM/B/big_data.html)
- <http://strata.oreilly.com/2012/01/what-is-big-data.html>
- <http://www.cra.org/ccc/resources/ccc-led-white-papers>, Challenges and Opportunities with Big Data
- Big Data, Big Science, Big Collaboration: Delivering Connected R&D for Better Value, March 2013, <http://www.idbs.com/pdfs/scientific-computing-big-science.pdf>
- [http://en.wikipedia.org/wiki/Open\\_data](http://en.wikipedia.org/wiki/Open_data) (Ag. 201) y referencias, tales como:
  - Auer, S. R.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. (2007). "DBpedia: A Nucleus for a Web of Open Data". The Semantic Web. Lecture Notes in Computer Science 4825. p. 722. doi:10.1007/978-3-540-76298-0\_52. ISBN 978-3-540-76297-3
  - Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 - Final Communique, <http://www.oecd.org/science/sci-tech/sciencetechnologyandinnovationforthe21stcenturymeetingoftheoecdcommitteeforscientificandtechnologicalpolicyatministeriallevel29-30january2004-finalcommunique.htm>