Fundación Sadosky

Big Data: ¡Qué Grande esta Charla!

Esteban Feuerstein

15 de Mayo de 2014

Ministerio de Ciencia, Tecnología e Innovación Productiva
Presidencia de la Nación

fundación SADOSKY
Investigación y Desarrollo en TIC

cessi ArgenTIna

CICOMRA

Titulos alternativos

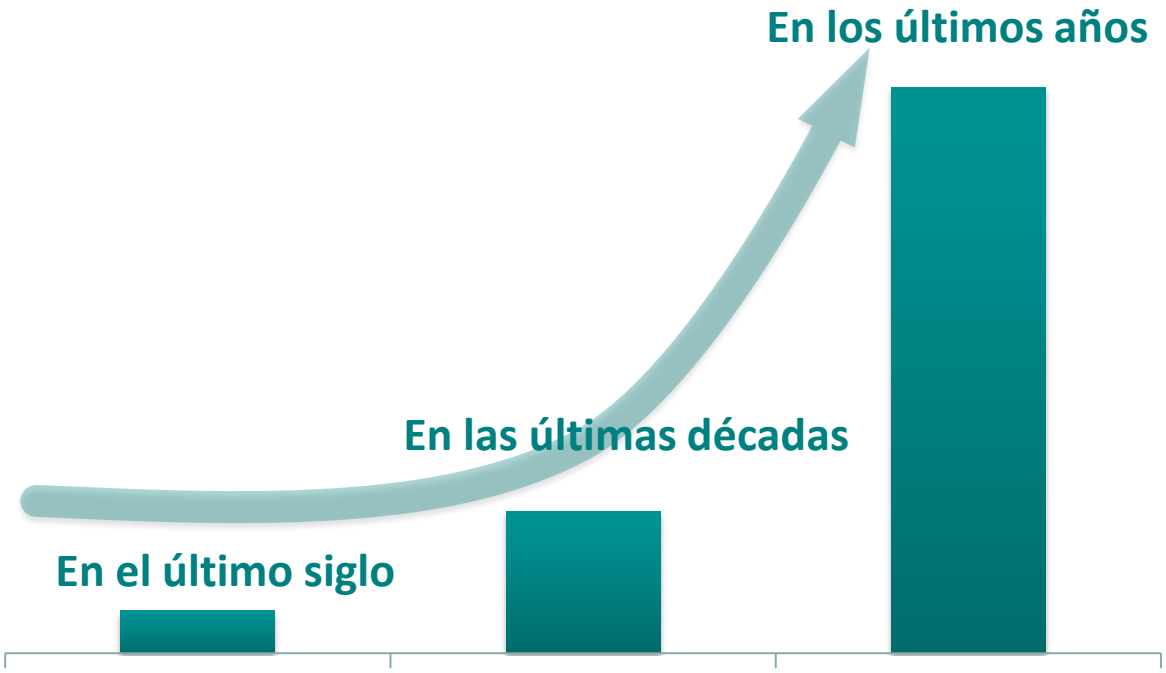
- Big Data ¿sexy o no?
- ¡Grande Pa!

¿De qué hablamos cuándo hablamos de Big Data?

- La disponibilidad de datos, provenientes de las más diversas fuentes, y la capacidad de almacenarlos y procesarlos, han crecido en forma extraordinaria (pero no exactamente la misma) en el último siglo.
- La disponibilidad de datos, provenientes de las más diversas fuentes, y la capacidad de almacenarlos y procesarlos, han crecido en forma extraordinaria (pero no exactamente la misma) en las últimas décadas.
- La disponibilidad de datos, provenientes de las más diversas fuentes, y la capacidad de almacenarlos y procesarlos, han crecido en forma extraordinaria (pero no exactamente la misma) en los últimos años.
- Volúmenes antes impensables, hoy aparecen por todos lados, junto con la capacidad de procesarlos.
- Y vamos por más…

¿De qué hablamos cuándo hablamos de Big Data?

- La disponibilidad de datos, provenientes de las más diversas fuentes, y la capacidad de almacenarlos y procesarlos, han crecido en forma extraordinaria



- Volúmenes antes impensables, hoy aparecen por todos lados, junto con la capacidad de procesarlos.
- Y vamos por más…

Algunos grandes números

01234

56789

Algunos grandes números

- 1 Bit = Binary Digit
- 8 Bits = 1 Byte
- 1000 Bytes = 1 Kilobyte
- 1000 Kilobytes = 1 Megabyte
- 1000 Megabytes = 1 Gigabyte
- 1000 Gigabytes = 1 Terabyte
- 1000 Terabytes = 1 Petabyte
- 1000 Petabytes = 1 Exabyte
- 1000 Exabytes = 1 Zettabyte
- 1000 Zettabytes = 1 Yottabyte
- 1000 Yottabytes = 1 Brontobyte*
- 1000 Brontobytes = 1 Geopbyte*

* :Nombres no oficiales

- 1 Terabyte = 1000 copias de la Enciclopedia Británica
- 10 Terabytes = toda la colección impresa de la Biblioteca del Congreso de USA
- 1 Petabyte = 1000 Terabytes = 20 millones de archivos de 4 cajones llenos de texto
- 5 Exabytes = 5000 Petabytes = todas las palabras habladas por la humanidad desde su principio
- 1 Yottabyte = 1000000 Exabytes = la totalidad de Internet
- Se necesitarian aproximadamente 11,000,000,000,000 años para descargar un archivo de 1 Yottabyte

Gentileza Gabriel Taubin

Algunos grandes números

Bit
Byte
Kilobyte
Megabyte
Gigabyte
Terabyte
Petabyte
Exabyte
Zettabyte
Yottabyte

- 1 Terabyte = 1000 copias de la Enciclopedia Británica
- 10 Terabytes = toda la colección impresa de la Biblioteca del Congreso de USA
- 1 Petabyte = 1000 Terabytes = 20 millones de archivos de 4 cajones llenos de texto
- 5 Exabytes = 5000 Petabytes = todas las palabras habladas por la humanidad desde su principio
- 1 Yottabyte = 1000000 Exabytes = la totalidad de Internet
- Se necesitarian aproximadamente 11,000,000,000,000 años para descargar un archivo de 1 Yottabyte

Gentileza Gabriel Taubin

¿Dónde hay grandes datos?

- El colisionador de partículas del CERN generó **40.000.000.000.000 B/Seg (40 TB/Seg)** en 2012.

¿Dónde hay grandes datos?

- Un Airbus A380 genera **640 TB** por vuelo.

¿Dónde hay grandes datos?

- Twitter genera **12 TB** de datos por día.

¿Dónde hay grandes datos?

- La bolsa de Nueva York genera **1 TB** por día.

¿Dónde hay grandes datos?

- Walmart utilizaba **30.000 millones** de sensores RFID en 2012.

Walmart

¿Dónde hay grandes datos?

- Una cosechadora genera **5.000 datos por hectárea**, en cada pasada. Un dron genera **50.000 datos por hectárea**, en cada pasada. En Argentina hay alrededor de **30.000.000 de hectáreas** cultivadas.

¿Dónde hay grandes datos?

- El sistema SUBE está funcionando en **18.000 colectivos**, cada uno dispone de GPS y podría registrar su ubicación cada pocos segundos. Además, hay ~**11.000.000 de transacciones** de uso por día.

¿Qué hay de nuevo, viejo?

- ¿Pero qué hay de nuevo? ¿Solamente prefijos?
- No sólo **volumen**, sino **variedad** y **velocidad**. Algunos agregan **valor**, y otros **veracidad**
- ¿Entonces qué hay de nuevo?
- Ante todo: un cambio de paradigma
  - Por un lado, pasamos de un modelo en el que algunos generaban data, y los demás la consumían….a un modelo en el que todo genera data, y todos la consumimos
  - Pero por otro lado, de un modelo de **experimentos para probar hipótesis** (realizar hipótesis, diseñar experimento, recolectar datos, realizar experimento, validar hipótesis)…a otro en el que podemos **descubrir conocimiento en los datos** (recolectar datos….analizarlos, visualizarlos, sacar conclusiones)
- ¿Qué implica eso?
  - Nuevos nombres para cosas viejas
  - Nuevas dificultades para problemas viejos
  - Nuevas soluciones para problemas viejos
  - Nuevos problemas

Ejemplos, por dominios

- Ciencias duras
  - Bases de datos de astronomía, genómica, ambiente…
- Humanidades y ciencias sociales
  - Libros escaneados, documentos históricos, datos de interacciones sociales, censos, GPS, teléfonos móviles…
- Negocios/Comercio/Finanzas
  - Ventas corporativas, mercado de valores, CRM…
- Entretenimientos/sociales
  - Video, música, redes sociales…
- Medicina/Salud
  - RMN, TC, análisis clínicos, historias clínicas…
- Transporte
  - GPS, tránsito, semáforos…
- Gobierno/Política
- Producción
  - Redes de sensores, imágenes satelitales, internet of things…
- Muchos "…" y todos los cruces!

Ejemplo: Transporte

- Podemos recolectar datos de transporte
  - GPS de los vehículos (ubicación, destino, tiempo)
  - Cámaras/sensores en el terreno
  - Existen medios inalámbricos para “subirlos a la nube”
- Podemos analizarlos
  - On-line u off-line
  - En forma genérica/agregada o personalizada
- Y utilizarlos para optimizar/planificar
  - Rutas congestionadas/ruteo online
  - Dónde está la gente ahora (por ej. para mandar taxis)
  - Simulaciones: qué va a pasar el día del partido, qué va a pasar si se inunda cierta esquina, etc.

Transporte: Caso de estudio de Singapur

- 26000 taxis total, 16000 para el estudio. Envían ID, ubicación GPS, velocidad, status, hora a intervalos de ≈ 0.5-1 minuto, durante un mes. Total: 512M datos, 33 GB
- Limpieza de datos, geolocalización en los caminos, …
- Aplicaciones: taxis ocupados vs. taxis circulando, Localización de paradas, Visualización del tráfico / Detección de patrones a partir de la velocidad de movimiento, Detección de congestiones, ruteo on-line
- Issues técnicos
  - En Singapur existen 10000 sensores de tráfico en tiempo real, que permitieron ajustar la curva de la data originada en los taxis y generalizarla (taxis llenos no van a la misma velocidad que taxis vacíos, etc.): regresión, modelos de Markov, …
  - En Buenos Aires, podríamos usar la data del SUBE. Los colectivos siempre van a la más alta velocidad posible (¿☹ o ☺?). ¿Para qué más podríamos usar esa data?

Ejemplo: Visualización de Twitter

- 500 millones de tweets por día
- Mucho más que solamente 140 caracteres (coordenadas, timestamp, información de usuario y seguidores, información de Respuestas, Hashtags, dispositivo/plataforma utilizada)
- Volúmenes y frecuencias masivas, hacen falta nuevas herramientas para visualizar
- Se quiere correlación con otros datos, internos o externos (ej. Preferencias de marcas vs. ingreso medido según censo)
- Y también análisis "profundo" de contenidos
  - ¿De qué producto, persona o espectáculo están hablando?
  - ¿Qué opinión se está expresando ("sentiment analysis")?
- Ej. Aplicación desarrollada por Guido de Caso en el DC: predicción de granizo
- Ej. 2: http://mapd.csail.mit.edu/tweetmap-desktop/

Ejemplo: Medicina

- Disponibilidad de datos médicos
  - No sólo de eventos o estudios hechos durante internaciones o crisis sino también información sobre la "vida normal" de los pacientes
- Datos que antes se TIRABAN (en la época del papel, pero también electrónica), hoy se conservan.
- Datos agregados, de distintas instituciones, por iniciativa del sector público.
- Las técnicas tradicionales de investigación clínica
  - No escalan a estas magnitudes (millones de casos, miles de estudios, estudios muy "pesados" (imágenes, videos, etc.))
  - Fueron diseñadas para testear hipótesis, no descubrir nuevo conocimiento
  - Requieren estudios clínicos adicionales, que pueden ser costosos

Medicina: predicción de patologías

- Ejemplos: predicción de complicaciones post-quirúrgicas, predicción de enfermedades cardiovasculares.
- Mejores decisiones sobre qué tratamiento seguir con determinado paciente
- Mejor gestión de prestadores (matching paciente-prestador, optimización de recursos)
- Modificar tratamientos a partir del aprendizaje (por ejemplo, profilaxis)
- Issues técnicos
  - Datos realmente grandes: decenas de millones de pacientes, miles de MB por pacientes. El paralelismo puede ayudar, pero no si los algoritmos son por ejemplo cúbicos
  - Al mismo tiempo, la alta dimensionalidad, hace que los datos que soportan cada caso no sean suficientes.
  - Datos realmente variados: señales, estudios de laboratorio, imágenes, información genómica, lenguaje natural….
  - Datos incompletos, generalmente incorrectos, resultados ambiguos
  - La “verdad” es casi siempre imposible de verificar (¡es una predicción de riesgo!)
  - Datos variables de hospital a hospital

Ejemplo: Finanzas

- USA, octubre 2013: promedio de deuda con la tarjeta de crédito: US$ 15.000. TEA: 15%
- ¿Por qué tasas tan altas? ;-)
- 6,7% del total va a pérdida (1Q13, otros periodos es peor, hasta 10%)
- Los métodos de scoring son completamente insensibles al contexto, por ejemplo a las crisis económicas
- Data: todas las transacciones bancarias, saldos de las cuentas e información crediticia (niveles de ingreso, etc.) – En total 10 TB para un subconjunto del 1% del total.
- Objetivo: dado un individuo *j* con carácterísticas X_j, estimar la probabilidad de que *j* no pague su resumen de tarjeta de crédito.
- Las características incluyen tanto características individuales como factores "macro", y las interacciones entre ambos tipos.
- Datos de alta dimensionalidad (muchas "características"), y gran tamaño hacen imposible aplicar técnicas más tradicionales de Machine Learning.

Un caso con dos caras

- En 2009, investigadores de Google anunciaron un gran resultado en la revista Nature:
- Sin ningún tipo de dato médico, fueron capaces de seguir el brote de gripe a lo ancho de EEUU.
- Y más rápido que los centros de prevención nacionales: un día de demora vs. una semana.
- La clave: la correlación entre los términos que la gente buscaba, y los síntomas de la gripe.
- Cuatro años más tarde, la mala noticia: la capacidad de predicción del modelo se degradó fuertemente: mientras la data de Google apuntaba a una epidemia severa, los datos oficiales lo negaban.
- ¿Quién estaba equivocado? Google, por un factor 2
- ¿El problema? La clásica lucha entre correlación y causalidad puede dar una explicación: los patrones de búsqueda cambiaron, aunque no sepamos por qué.

Perspectiva de impacto – áreas tecnológicas

- Twelve potentially economically disruptive technologies (Fuente: McKinsey Global Institute analysis)
  - Mobile Internet
  - Automation of knowledge work
  - The Internet of Things
  - Cloud technology
  - Advanced robotics
  - Autonomous and near-autonomous vehicles
  - Next-generation genomics
  - Energy storage Devices
  - 3D printing
  - Advanced materials
  - Advanced oil and gas exploration and recovery
  - Renewable energy
- ¿Cuántas de ellas tienen que ver / generan Big Data?

¿Qué dicen los que tienen la sartén por el mango?

- McKinsey considera que los datos [“Big Data”] se han convertido en un factor de producción, junto al capital físico y al humano (The Economist, 2011.)
- “Datos son materia prima vital para la economía del conocimiento, así como lo fueron el carbón y los minerales férricos durante la Revolución Industrial. ” (Lohr, 2011.)
- “Las mismas precondiciones que permitieron que olas previas de innovación en IT promovieran productividad, es decir, innovaciones tecnológicas seguidas de la adopción de innovaciones de gestión complementarias, se están dando para bases masivas de datos, y es de esperar que “Big Data” y las capacidades analíticas avanzadas que genera tengan al menos tanto impacto permanente en la productividad como otros tipos de tecnología” (McGuire et al., 2012.)
- The World Economic Forum (2012) declaró que los datos son una nueva clase de activo económico, como oro o moneda.
- Un buen ejemplo del uso productivo de datos es el aumento de productividad en la industria del petróleo (Forbes, 2013.) IBM (2013) presenta estos aumentos en otras industrias con valores de entre 20% (salud, reducción de mortalidad de pacientes) y 90-99% (industria de telecomunicaciones y provisión de generadores de energía).

¿Y los que cortan el bacalao?

- En mayo de 2012, en la Cloud Computing Conference, organizada por Goldman Sachs, Shaun Connolly de Hortonworks presentó el concepto de Big Data como “The New Competitive Advantage”. Connolly articuló siete razones para su aserción: dos de negocios, tres de tecnología y dos financieras:
  - Business reasons
    - 1. New innovative business models become possible
    - 2. New insights arise that give competitive advantages
  - Technological reasons
    - 3. The generation and storage of data continue to grow exponentially
    - 4. We find data in various forms everywhere
    - 5. Traditional solutions do not meet new demands regarding complexity
  - Financial reasons
    - 6. The costs of data systems continue to rise as a percentage of the it budget
    - 7. New standard hardware and open-source software offer cost benefits …”

¿Qué leen todos ellos?

- *Harvard Business Review* - Número dedicado a Big Data (oct. 2012)

- **Data Scientist: The Sexiest Job of the 21st Century** by Thomas H. Davenport and D.J. Patil

- **Big Data: The Management Revolution** by Andrew McAfee and Erik Brynjolfsson
  - Versión libre: “los sistemas tradicionales presentes en las organizaciones, fueron construidos para brindar resultados en modalidad batch, y no tiempo real como se requiere en el contexto Big Data. ¡Esto implica un desafío en términos del management también!

¿Entonces, qué hay de nuevo?

- Nuevos nombres para cosas viejas
- Nuevas dificultades para problemas viejos
- Nuevas soluciones para problemas viejos
- Nuevos problemas
- ¡Nuevas buzzwords!

- Y no vale la pena discutir, en cada caso, en qué categoría cae lo que tenemos a mano (lo más probable es que caiga en varias)

Algunas dimensiones de la problemática

- Adquisición y preprocesamiento de datos
- Complejidad de Datos
- Modelado, análisis, aprendizaje y extracción de conocimiento
- Simulación y visualización
- Seguridad y privacidad

Adquisición e integración de Datos

- Algunos principios clásicos de las bases de datos como
  - La posibilidad de crear un esquema unificado, con nombres y tipos de datos unificados para describir determinados conceptos,
  - y la resolución o unificación de entidades referenciadas en múltiples data sets para identificarlas unívocamente,

  constituyen un reto distinto en un contexto con cientos o miles de fuentes de datos, y miles de millones de registros.
- Ejemplo (con pocas fuentes)
- Ahora imaginemos miles de productores agrarios mandando información sobre el efecto de la aplicación de determinado fertilizante en su rinde.

Procesamiento

- El paralelismo, los sistemas distribuidos, la computación de alto rendimiento (HPC) ya existían, pero:
  - el paralelismo masivo,
  - la posiblilidad de aprovechar el poder de muchos procesadores, mucha memoria,
  - datos que llegan en modo espasmódico

  dieron lugar a:
  - nuevos conceptos informáticos (Hadoop, MapReduce, Spark y otras plataformas de cómputo, cada uno con sus limitaciones y sus desafíos inherentes para mejorar su performance)
  - nuevos conceptos económicos, como la computación en la nube, la elasticidad, y el pay-as-you-go
  - el surgimiento de nuevos actores de peso (por ejemplo Amazon)

Procesamiento – Ejemplos de conceptos nuevos

- GPUs de punta proveen
  - 250 GB/segundo de ancho de banda (5X microprocesadores convencionales)
  - 4 Teraflops de poder de cómputo (10X microprocesadores multi-core convencionales)
  - 7 a 70X speedup en operaciones de base de datos
- Desafíos
  - Memoria limitada en las GPUs (pero en ascenso)
  - Ancho de banda limitado entre la CPUy las GPUs (pero va a cambiar)
- “Shared Nothing” Processing: Múltiples GPUs, con data particionada entre ellas, junto a
- “Parallel plumbing”: Paralelizar las operaciones y luego pensar en SQL tradicional…..

Modelado, análisis, aprendizaje, extracción de conocimiento

- Data Analysis, Estadística, Machine Learning, Teoría de Grafos ya existían
- Nuevamente: las 3 Vs implican nuevos desafíos
- Paralelismo masivo: fuerza ¿bruta?
- Paralelismo multicore, y sus problemas de escalabilidad
- Algoritmos más eficientes
  - Particionamiento: operar en la parte de los datos que necesitamos
  - Sampling: operar en un subconjunto de los datos
  - Sumarización: operar en un resumen de los datos
  - Compresión: operar en los datos comprimidos

Modelado, análisis, aprendizaje, extracción de conocimiento

- Nuevas formas de análisis
- Nuevo enfoques para los experimentos
  - No sólo validar hipótesis sino descubrir
- Nuevas aplicaciones y modelos
  - Motores de recomendación
  - Procesamiento de lenguaje natural
  - Redes Sociales
  - Big Graphs

Almacenamiento y acceso

- Nuevas tecnologías fueron (están siendo) desarrolladas para procesar Big Data
  - Bases de datos “columnares” (column-oriented)
  - Sistemas orientados a procesamiento en GPUs
  - Sistemas de procesamiento en memoria principal
  - NoSQL, NewSQL con sus distintas interfaces, mecanismos de consistencia y performance
  - SQL masivo, queryng arrays y processing graphs.

Visualización

- Otra vez: los volúmenes y velocidades son masivos,
- Correlación con datos de localización u otras variables
- Hacen falta nuevas herramientas de visualización y manipulación interactiva
- Lectura recomendada: Presentación de Gabriel Taubin en las Jornadas de Definición Estratégica en Big Data, 2013.

Seguridad /Privacidad

- Vinculación entre paralelismo masivo y seguridad
- Algoritmos eficientes de encriptación (hay que encriptar muchos datos!)+ protocolos de acceso distribuido

mejorar

Conclusiones (¡no es la última transparencia!)

No todo es BD, ni tiene por qué serlo. Pero no hay dudas de que cada vez más aplicaciones tendrán la impronta BD, que el paradigma llegó para quedarse, que va a haber oportunidades de generar valor a partir de su aplicación, y que va a haber trabajo para quienes se dediquen a esto, desde la teoría y desde la práctica, desde el software y desde el hardware.

Más historietas

Desafíos para la Argentina

- Soberanía tecnológica: si no guardamos y procesamos nosotros nuestros datos, lo harán otros. Estamos ante una dimensión en la cual si no hacemos algo, seguirá aumentando la brecha tecnológica.
- Formación: la disponibilidad de profesionales formados va a ser uno de los cuellos de botella.
- Multidisciplinariedad: la formación requerida no corresponde a una única disciplina.
- Necesidad de construir una Infraestructura de Grandes Datos (no sólo hardware sino herramientas de base y middleware, además de la capacidad de análisis)
- Proyectos: impulsar el desarrollo de la actividad en el país. Generar masa crítica. No formar profesionales que se vayan al exterior.

Qué está haciendo la Fundación Sadosky - Historia

- Estudio de Prospectiva: ¡se viene BD!
- Grupo de trabajo con Mario Nemirovsky + Gabriel Taubin
- Jornadas con grupos de investigación, empresas y reparticiones estatales o para-estatales
- Documento estratégico

Visión: Transformar a la Argentina, en los próximos 5 años, en un actor relevante –como país "periférico"- líder en la región sudamericana, de la "revolución de los Grandes Datos

Qué está haciendo la Fundación Sadosky - Actualidad

- **Visión:** Que la Argentina sea un líder regional en la temática de Grandes Datos, considerada clave para la autonomía tecnológica, el desarrollo económico y social y la competitividad.
- **Misión/Estrategia:** La Fundación Sadosky será actor principal de ese camino a través de la creación de un Programa de Ciencia de Datos (PCD) dentro de su estructura, con un rol que se extiende tanto sobre el nivel estratégico como sobre el de gestión.
- **Estructura Organizacional:** tres personas, reportando directamente al Director Ejecutivo.

Qué está haciendo la Fundación Sadosky - Presente/Futuro

- Comunidad Argentina de Grandes Datos, con más de 60 participantes al día de hoy.
- Proyecto Faro "**Palenque**": una plataforma de grandes datos geolocalizados del agro (en conjunto con AACREA)
- Conglomerado de empresas y grupos de investigación para el desarrollo de aplicaciones para el Palenque
- Formación de Recursos Humanos
  - Track de Big Data en la ECI
  - Auspicio a las Jornadas de la Maestría en Data Mining (UBA)
  - Becas de formación en BD en el exterior
- Investigación
  - BD como área estratégica para los PICT
- Proyectos específicos: Bioinformática
- Soporte al desarrollo de un Cloud Nacional
- Soporte a iniciativas de (Grandes) Datos Abiertos
- Jornadas Argentinas de Big Data (2015?)
- Centro Nacional de Ciencia de Datos

Algo sobre Palenque

- En pocos días se anunciará (???)
- Participantes:
  - Fundación Sadosky
  - AACREA (*Asociación Argentina de Consorcios Regionales de Experimentación Agrícola)*
  - INTA
  - Ministerio de Agricultura
- En qué consiste:
  - Una plataforma y un ecosistema de aplicaciones que brindarán soluciones tecnológicas basadas en grandes datos a los productores agropecuarios.
- Beneficios esperados
  - Aprovechamiento de grandes volúmenes de datos existentes o posibles de ser generados
  - Aumento de la productividad a través de la agricultura de precisión (sensores, irrigación inteligente, siembra inteligente, fertilización inteligente, etc.), Simulación, Utilización de pronósticos meteorológicos, etc.
  - Generación de conocimiento público

Ministerio de Ciencia, Tecnología e Innovación Produc
Presidencia de

Y un anuncio: E

Ministerio de Ciencia, Tecnología e Innovación Productiva
Presidencia de la Nación

Y un anuncio: ECI 2014

Ministerio de Ciencia, Tecnología e Innovación Productiva
Presidencia de la Na

Y un anuncio: ECI 2

Ministerio de Ciencia, Tec e Innovació
Presiden

Y un anunc

Ahora sí

¡Big gracias!

Ejemplo: Adquisición e integración de Datos

FlightView

American Airlines Flight Number 119 (AA119)

FLIGHT TRACKER

6:15 PM

Departure
Airport:
Scheduled Time: 6:15 PM, Dec 08
Takeoff Time: 6:53 PM, Dec 08
Terminal - Gate: Terminal A - 32

ArrivalStatus: In Air
Airport:
Scheduled Time: 9:40 PM, Dec 08
9:42 PM, Dec 08
Estimated Time:
Track This Flight ve!
Time Remaining: 25 min
Terminal - Gate: Terminal 4 - 42
Baggage Claim: 4

9:40 PM

FlightAware

AAL119 (Track inbound flight)
(web site) (all flights)
American Airlines "American"
Alert Me

Aircraft	Boeing 737-800 (twin-jet) (B738/Q - track or photos)
Origin	Terminal A / Gate 32 / Newark Liberty Intl (KEWR - track or info)
Destination	Terminal 4 / Gate 42B / Los Angeles Intl (KLAX - track or info)
	Other flights between these airports
Route	ZIMMZ Q42 BTRIX Q480 AIR J80 VHP J80 MCI J24 SLN J102 ALS J44 RSK J64 POB RITVR1 (Decode)
Date	2011年 12月 08日 (Thursday)
Duration	5 hours 43 minutes
	20 minutes left
Progress	5 hours 23 minutes
Status	En Route (2,284 sm do 8 sm to go)
Distance	Direct: 2,451 sm ed: 2,458
Fare	$51.99 to $3,561 average: $241.96 (airline insight)
Cabin	First: Dinner / Economy: Food for sale

	Scheduled	7-day Average	Actual/Estimated
Departure	06:15PM EST	07:08PM EST	06:53PM EST
Arrival	08:33PM PST	09:17PM PST	*09:36PM PST*

6:15 PM

8:33 PM

Orbitz

American Airlines # 119

Leg 1: In Transit

Departs: Newark (EWR) View real-time airport conditions at

Gate: 32

Scheduled Estimated Actual
6:22p Dec 8 — 6:32p Dec 8

6:22 PM

Arrives: Los Angeles (LAX) View real-time airport conditions

Gate: 42B

Scheduled Estimated Actual
9:54p Dec 8 — 9:47p Dec 8

9:54 PM