

La importancia del HPC y el Big Data en la Investigación e Innovación

Ing. Emilio Chan Moya, MAP, PMP
ITCR / Grupo MCS / PTCR

Uruguay, 17 de Setiembre de 2015

CO-SUMMIT 2015

SMART INDUSTRY:
IMPACT OF
SOFTWARE
INNOVATION

ITEA 3



DAIMLER

Software complexity is growing exponentially



Director eDrive & Software
Technologies | Daimler AG



744,272,682

Tweets sent today

view how many in 1 second

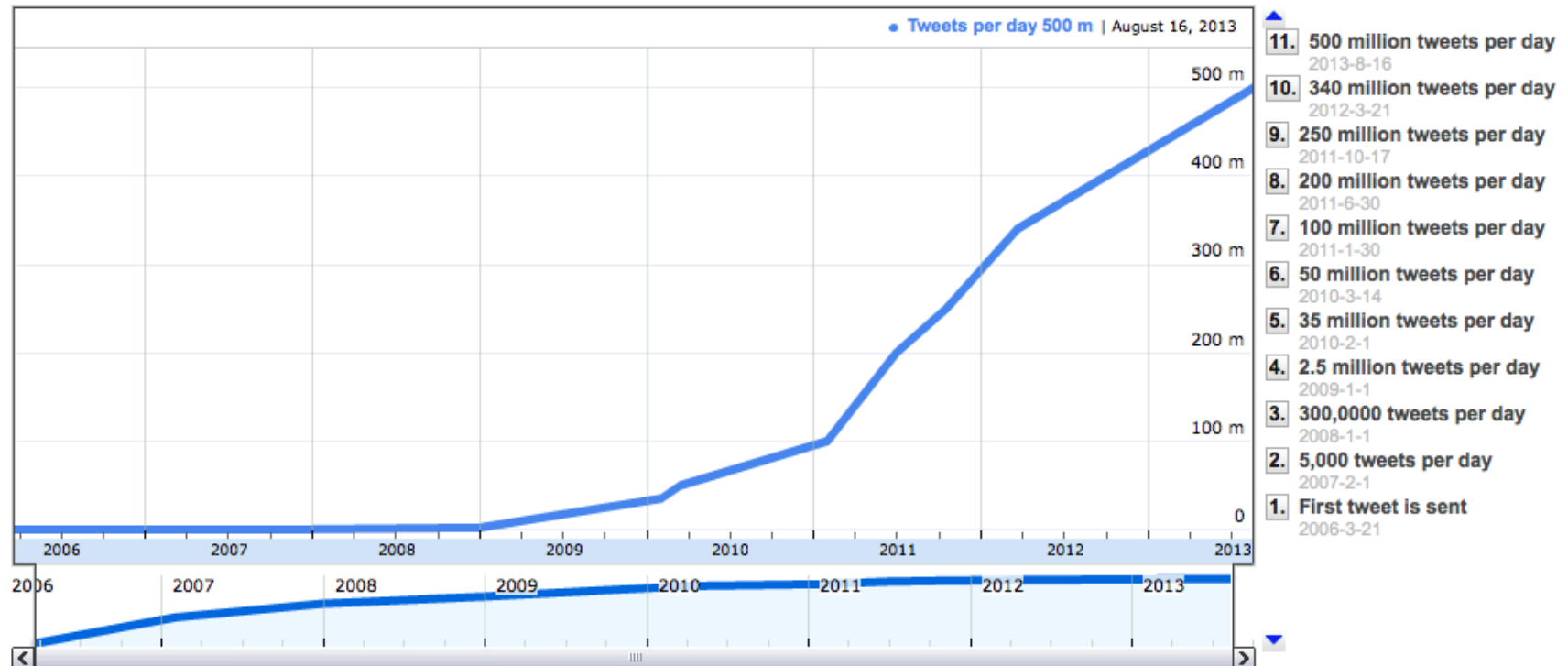
Number of Tweets (current and historical)

Growth Rate
















Curious Facts

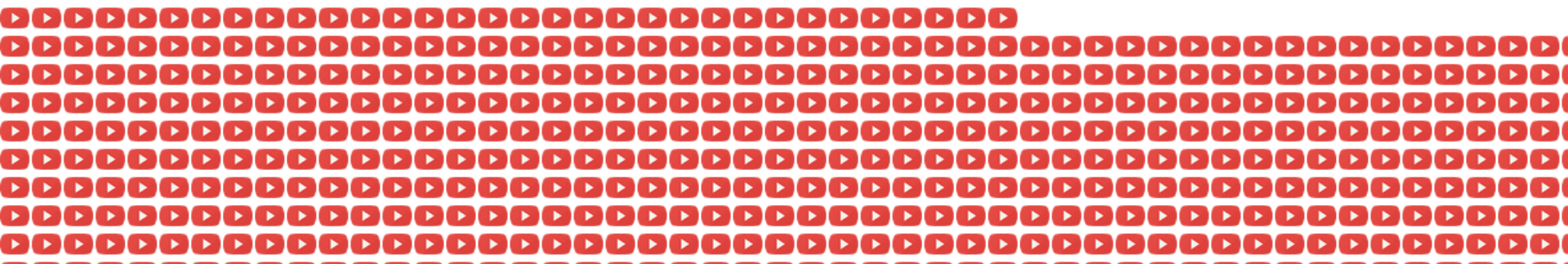
Sources and References

Every second, on average, around 6,000 tweets are tweeted on Twitter ([visualize them here](#)), which corresponds to over 350,000 tweets sent per minute, **500 million tweets per day** and around 200 billion tweets per year. The chart below shows the number of tweets per day throughout Twitter's history:

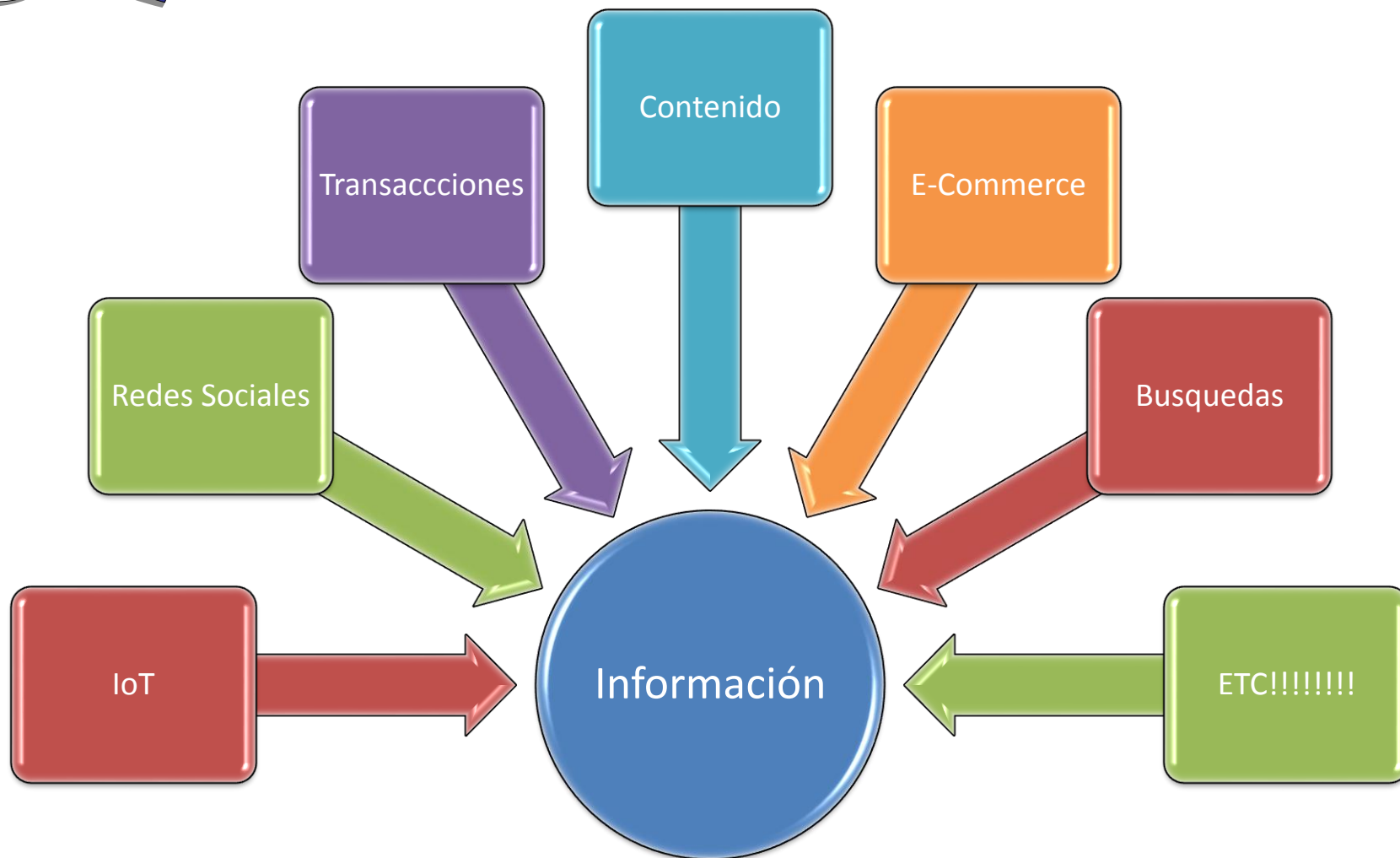


¿Qué puedes hacer en 15 segundos?

	10,089 Tweets sent in 1 second		150,780 Tweets since opening this page 0:00:15 seconds ago
	2,663 Instagram photos uploaded in 1 second		39,616 New photos since opening this page 0:00:15 seconds ago
	29,814 GB of Internet traffic in 1 second		453,680 Gigabytes since opening this page 0:00:15 seconds ago
	50,620 Google searches in 1 second		753,029 Searches since opening this page 0:00:15 seconds ago
	 = spam 2,427,068 Emails sent in 1 second		38,739,885 Emails since opening this page 0:00:15 seconds ago
	1,870 Skype calls in 1 second		28,216 Calls since opening this page 0:00:15 seconds ago
	107,832 YouTube videos viewed in 1 second		1,595,911 Videos since opening this page 0:00:15 seconds ago



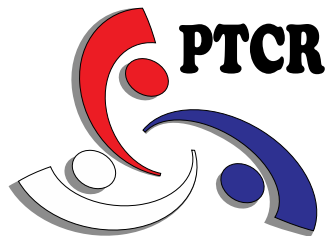
Fuentes de Información



A stylized logo composed of several curved, ribbon-like shapes and circles. The top-left shape is red, the middle-right shape is blue, and the bottom-left shape is white. Each shape has a slight 3D effect with a grey shadow. The text 'PTCR' is positioned to the right of the red shape, and 'BIG DATA' is positioned to the left of the white shape.

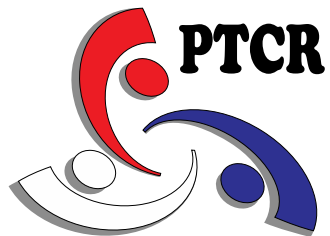
PTCR

BIG DATA



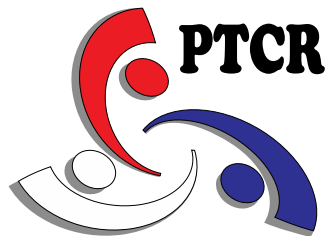
¿Para qué sirve?

- Permite generar información y conocimiento con base en información completa en tiempo real.



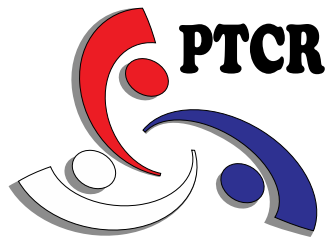
¿Cómo se origina?

Por la explosión en la cantidad (velocidad y frecuencia) y diversidad de datos digitales generados en tiempo real como resultado del rol cada vez mayor de la tecnología en las actividades diarias.



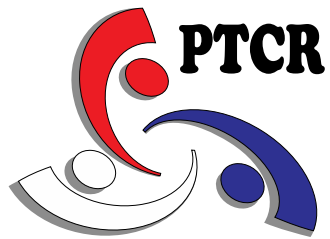
¿Qué es?

Conjuntos de datos cuyo tamaño está más allá de la capacidad de las herramientas de software de bases de datos típicas para capturar, almacenar, gestionar y analizar información.



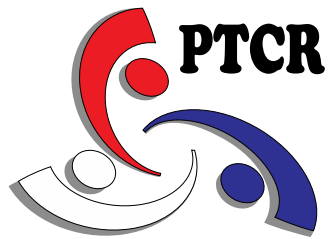
Concepto

- El concepto Big Data hace referencia a los sistemas que manipulan grandes conjuntos de datos, también conocidos como data sets .
- Principales cualidades: heterogeneidad, volatilidad.
- Mayores dificultades a la hora de trabajar a nivel de BIG DATA: volumen y velocidad.



Los retos...

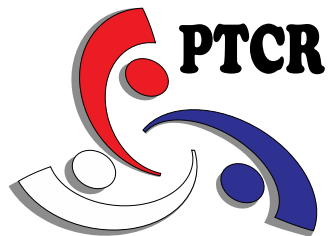
- Captura de datos .
- Almacenamiento de tales volúmenes de información.
- Capacidad de realizar búsquedas eficientes.
- Compartición .
- Posibilidad de llevar a cabo análisis efectivos.
- Visualización de los datos.



Implicaciones

Estamos en una Era caracterizada por la abundancia de datos.

- Ha alcanzado todos los sectores en la economía
- Los datos son un nuevo factor de producción y de ventaja competitiva



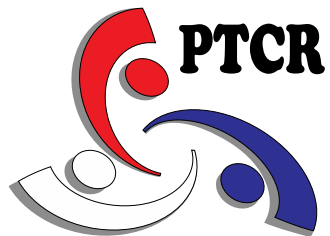
Implicaciones

Oportunidad:

- Aprender sobre el comportamiento humano para diversos fines.
- Creación de valor vía innovación, eficiencia y competitividad

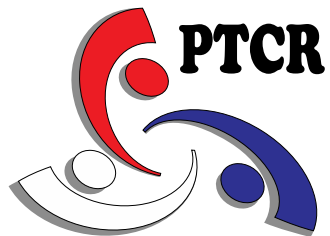
Nuevas formas de competencia y nuevos negocios

- Almacenamiento y gestión de datos.
- Análisis de datos empresariales. En 2010 se estimaba el valor de esta industria en más de \$ 100 mil millones, creciendo a casi un 10% al año



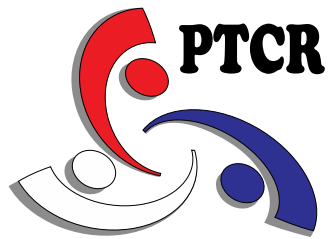
Creación de Valor...

- Segmentación de mercado y población para personalizar acciones
- Innovación en nuevos modelos de negocios, productos y servicios
- Mejora de productos existentes
- Desarrollo de nuevos productos (masa y personalización)



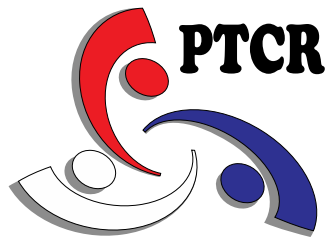
Creación de Valor...

- Nuevos modelos de servicio a nivel empresarial y gubernamental
- Apoyo a la toma de decisiones con software inteligente
- Transparencia y eficiencia por compartir datos
- Mejor y más oportuno análisis de desempeño de las organizaciones y ajustes en acción.



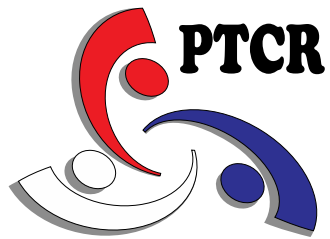
Analytics: capacidades

- La analítica de grandes datos se refiere a las herramientas y metodologías para transformar cantidades masivas de datos brutos en “datos sobre datos” con propósitos analíticos
- Se originó en las áreas de biología intensiva en cómputo, ingeniería biomédica, medicina y electrónica
- Algoritmos para detectar patrones, tendencias y correlaciones, en varios horizontes temporales, en los datos
- Uso de técnicas avanzadas de visualización: datos que hacen sentido



Problemas

- Disponibilidad de datos: asimetrías
 - *Las redes sociales generan datos abiertos*
 - *Los gobiernos los están abriendo, pero lentamente*
 - *Los datos de empresas siguen cerrados*
- Diferentes capacidades de buscar y analizar datos
- Falta de incentivos para compartir datos
- Privacidad y los límites al anonimato de conjuntos de datos
- Una buena parte de las nuevas fuentes de datos reflejan sólo percepciones, intenciones y deseos
- *Apophenia*: ver patrones donde no hay; cantidades masivas de datos abren conexiones en todos los sentidos (error de Tipo I)



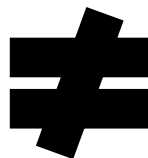
Big Data y Analytics

- ¿Qué sucede cuando las técnicas de análisis tradicionales se encuentran con sus límites?
- ¿Cuándo llega el momento en que la minería de datos no aporta las soluciones esperadas?
- ¿Cómo se enfrentan al desafío de los grandes datos y su expresión más desestructurada?

Romper el paradigma

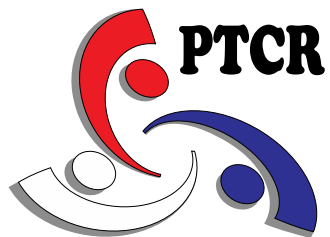
ANALISIS TRADICIONAL

- PERSONAS (SKILLS)
- EQUIPOS
- TECNICAS
- ENTEDIMIENTO



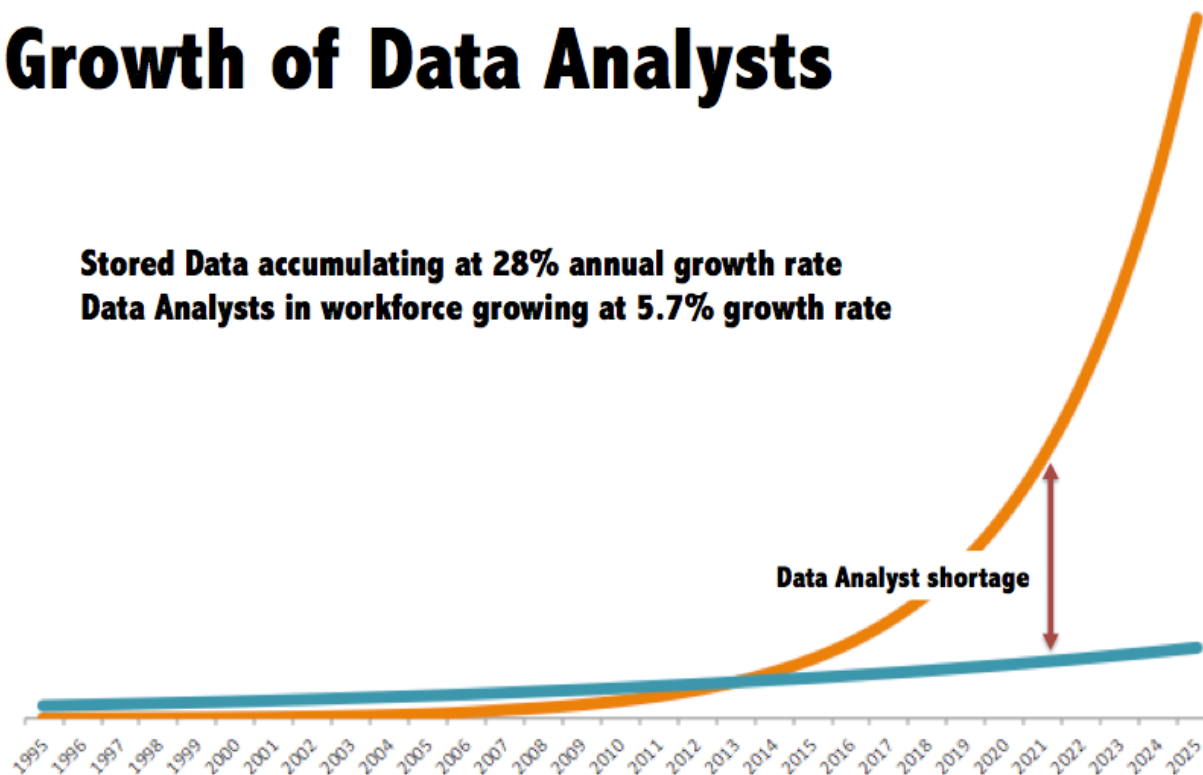
BI / BIG DATA ANALYSIS

- PERSONAS (SKILLS)
- EQUIPOS
- TECNICAS
- ENTEDIMIENTO

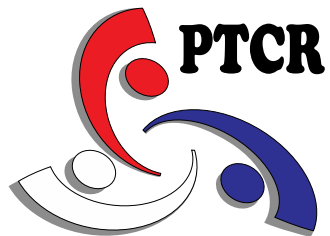


Growth of Data vs. Growth of Data Analysts

Stored Data accumulating at 28% annual growth rate
Data Analysts in workforce growing at 5.7% growth rate

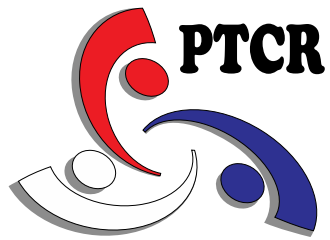


Datos almacenados se acumulan con una tasa de crecimiento anual del 28%, mientras que la fuerza de trabajo correspondiente a analistas de datos tan solo aumenta un 5,7% por año.



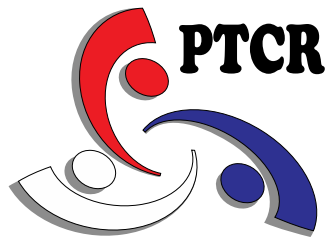
Perfil del Científico de Datos

- ***Transversal e integrador de disciplinas***
 - Matemáticas
 - Estadística
 - Programación
 - Aprendizaje autom.
 - Minería de textos
 - Procesamiento de lenguaje natural
 - Visualización de datos
 - Big Data
 - Area de dominio
- ***Demanda de profesionales a nivel de posgrado***
 - Farmacéutica
 - Software
 - Internet
 - TIC y servicios
 - Biotecnología
 - Finanzas y negocios
 - Salud
- ***Para 2018 se requerirá***
 - 140,000 a 190,000 científicos de datos
 - 1.5 millones de gerentes y analistas que entiendan del uso de Big Data y Analítica.



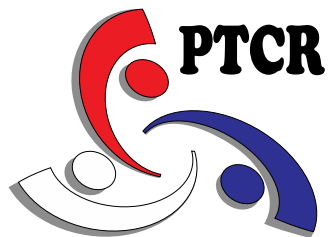
Computación de Alto Rendimiento

- **H P C (High Performance Computing)**
- La computación de alto rendimiento (HPC) es el uso de procesamiento paralelo para ejecutar aplicaciones avanzadas de manera eficiente, confiable y rápida.
- Usuarios frecuentes HPC: investigadores científicos, ingenieros e instituciones académicas.
- El término se usa, a veces, como sinónimo de supercomputadoras.



Computación de Alto Rendimiento

- **H P C (High Performance Computing)**
- El término se aplica en especial sistemas que operan arriba de un teraflops (10^{12}).
- Algunas supercomputadoras trabajan a más de un petaflops (10^{15}).



Para ponernos en perspectiva

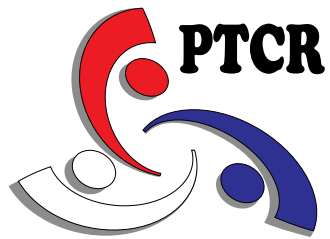
Procesador	Spec	GFLOPS
Core i7 5960X	8 core @ 3.0GHz AVX2	354
Dual Xeon E5 2687W	16 cores @ 3.2GHz AVX	345
Core i7 5930K	6 cores @ 3.5GHz AVX2	289

Computadora	Pais	Rmax (PFLOPS)	RPeak
Tianhe-2	China	33.863	54.902
Titan	EUA	17.590	27.113

GIGA (10^9)

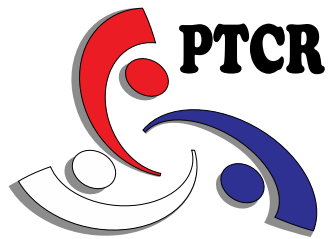
TERA (10^{12})

PETA (10^{15})



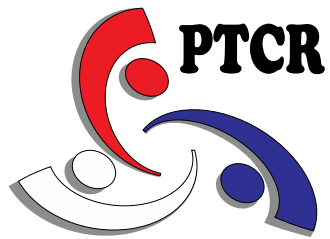
Aplicaciones de la HPC

- La **simulación de terremotos** para identificar áreas especialmente sensibles y predecir sus condiciones.
- **Modelado del clima.** Modelos computacionales pueden ser usados con datos viejos para evaluar su utilidad.
- **Modelaje de prototipos** físicos es caro y lleva mucho tiempo.
- **Manufactura digital.** El uso de la HPC (modelado, simulación y analítica) para definir productos y procesos manufactureros (The National Center for Manufacturing Sciences, NCMS).
- **Big data:** manejo de grandes cantidades de datos y de decisiones o rutinas complejas.



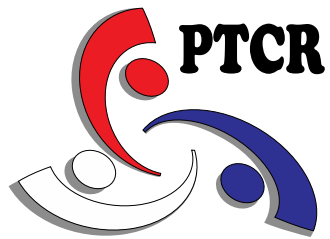
Beneficios en la innovación usando HPC

- El tiempo de ajuste en un laboratorio es de unos 9 meses, HPC puede reducir a menos de una semana.
- El análisis de un componente cuesta en promedio 50,000 USD en un laboratorio, mientras que mediante HPC se puede hacer por 3,000 USD.
- Prototipos virtuales y modelados en gran escala con base en HPC aceleran y racionalizan los procesos.
- Se mejoran la I&D, el diseño y la ingeniería, así como también los procesos de negocios (minería de datos, logística, CRM, etc.)



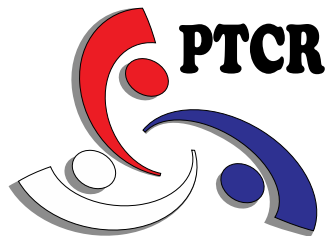
PREDICCIONES EN TICs 2015

- Principales predicciones de IDC en TIC para 2015
- Transición a la tercera plataforma tecnológica (informática móvil, servicios en la nube, Big Data, analítica y las redes sociales).
- IDC identificó esta plataforma en 2007 y predijo que se convertiría en el nuevo tercio del gasto mundial



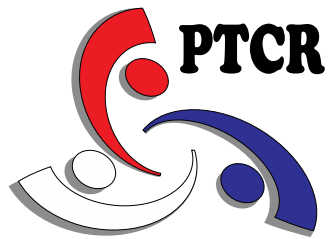
PREDICCIONES EN TICs 2015

1. El **gasto a nivel mundial en TIC crecerá un 3,8%** en 2015 a más de 3.800 MDD. Casi todo este crecimiento del gasto se centrará en las tecnologías de la tercera plataforma.
2. Los servicios de telecomunicaciones verán cómo los **datos inalámbricos serán los que más crecimiento registren** en este sector.
3. Los **dispositivos móviles y aplicaciones continuarán creciendo**. Las ventas de teléfonos inteligentes y tabletas alcanzarán 484.000 MDD (40% de todo el crecimiento del gasto de TI (con exclusión de los servicios de telecoms)).



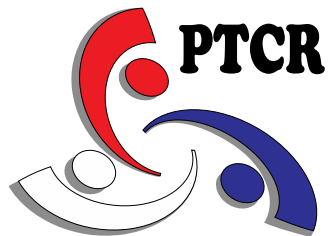
PREDICCIONES EN TICs 2015

4. **Servicios en la nube seguirán siendo un foco de actividad en 2015**
5. **Big Data y analítica verán importantes desarrollos en el 2015**
6. El **Internet de las cosas es uno de los aceleradores de innovación** más importantes para el crecimiento y la expansión de la tercera plataforma. La creación de más objetos conectados, impulsará el desarrollo de nuevas soluciones y un tercio del gasto en este mercado en 2015 se centrará en los dispositivos inteligentes integrados fuera de las TIC y las industrias de telecomunicaciones.



PREDICCIONES EN TIC 2015

7. Los **centros de datos están experimentando una transformación fundamental** ya que requieren de estas tecnologías en crecimiento. Este cambio conducirá a una mayor consolidación entre proveedores de servidores, almacenamiento, software y redes.



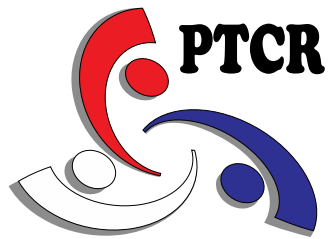
PREDICCIONES EN TICs 2015

América Latina

1. Internet de la Cosas (IoT) 291 millones de objetos conectados 600 millones.

2. Big Data y Analítica (BDA)

- Se calcula inversión de 1.4 mil MDD en América Latina. La línea de negocios se involucrará de manera más profunda en las decisiones de implementación tecnológica con los responsables de Tecnología.



PREDICCIONES EN TICs 2015

3. Transformación de las aplicaciones de negocio

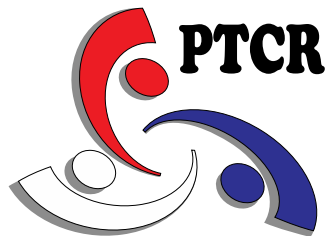
- Provocará modificaciones en la seguridad en middleware y reglas de negocio. Inversión de 2.9 mil MDD en soluciones de seguridad por proyectos de 3ª plataforma.

4. Crecimiento de SaaS

- Importante crecimiento del SaaS del 43%, con importante participación del gerente de la Línea de negocios.

5. Infraestructura y servicios para la Nube

- Será clave para captar negocios de desarrollo de Centros de Datos. AL experimentará un incremento superior al 50% en el mercado de nubes públicas.



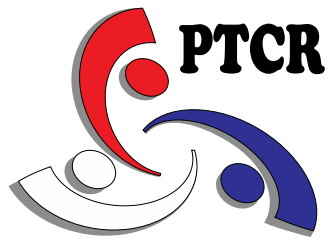
PREDICCIONES EN TICs 2015

6. El SDx para el desarrollo de infraestructura

- El todo definido por software será crucial para la automatización de la infraestructura para las aplicaciones en la 3ª plataforma. El mercado de Administración de Sistemas en AL será de 737 MDD. La demanda de almacenamiento se duplica cada 18 meses, por lo que una consola única de administración será de gran utilidad

7. Redes móviles

- En la región prevalece la brecha entre usuarios que tienen smartphone y que realmente lo pueden utilizar. Sólo **1% cuenta con cobertura 4G**. Los modelos de prepago de Voz y SMS cambiarán a Datos.



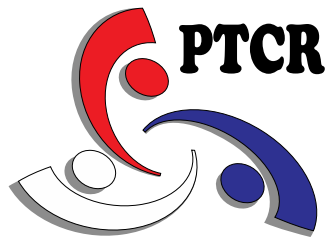
PREDICCIONES EN TICS 2015

8. Evolución de la movilidad

- Las organizaciones cambiarán el modelo de movilidad de empleados a movilidad de procesos y canales. Se estima que habrán más de 100 millones de empleados móviles en AL. Solo el 30% de las empresas cuenta con aplicaciones desarrolladas para ambientes móviles.

9. El rol de China en el entorno tecnológico empresarial

- Las empresas de tecnología chinas evolucionarán de ser proveedores de productos de bajo costo a partners clave en la región. 1 de cada 5 Smartphones serán originarios de China.

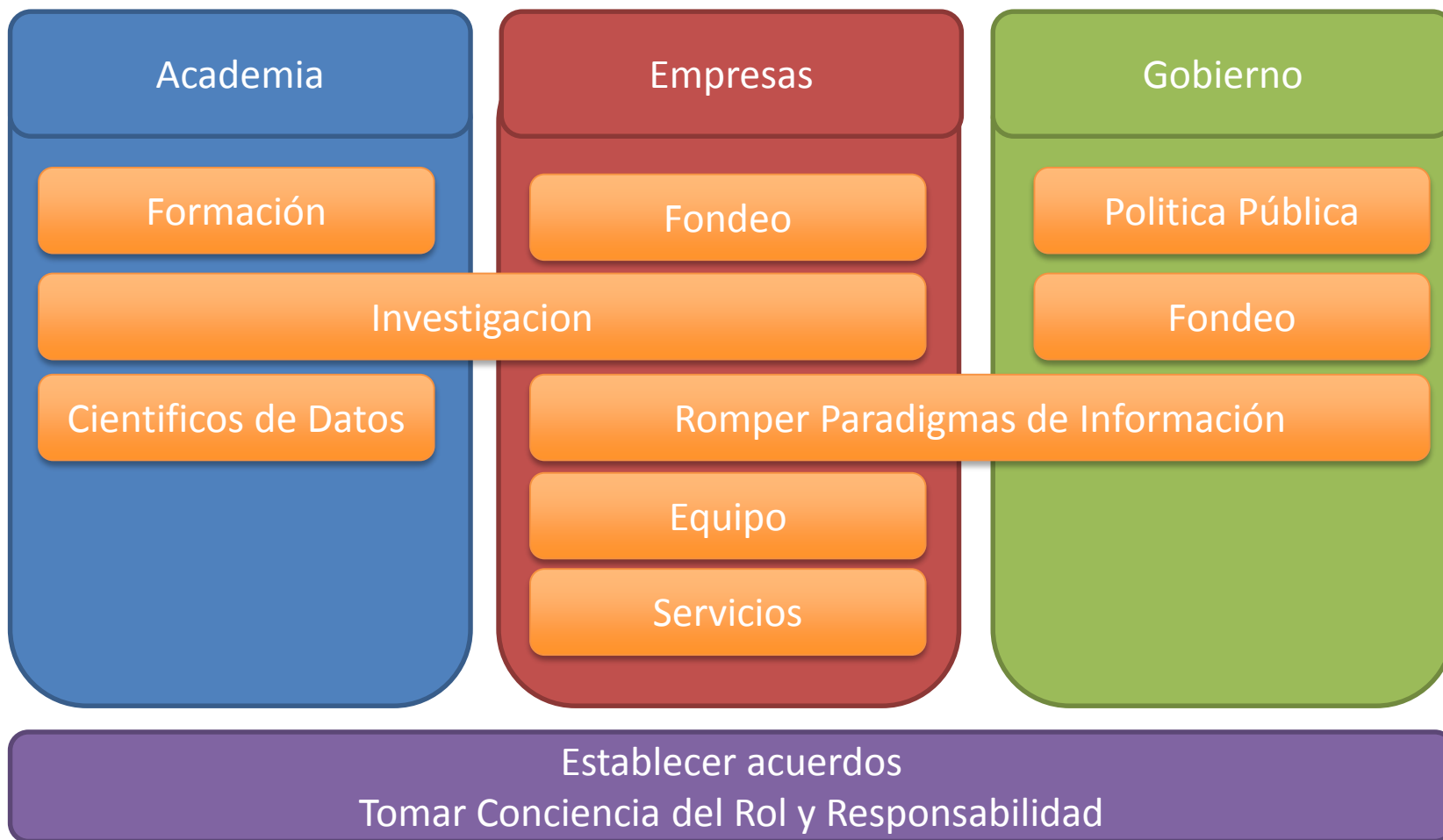


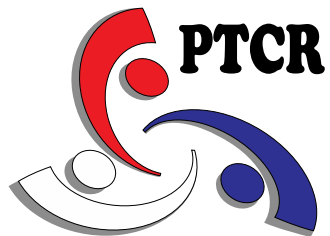
PREDICCIONES EN TICS 2015

10. Enfoque integral de seguridad

- Una visión holística integrará elements de movilidad, Nube, adminitración y analítica. Para 2015 el mercado de servicios profesionales de seguridad para AL se estima en 1.1 mil MDD, para 2018 aumentaría a 1.4 mil MDD
- Todos estos factores hacen evidente que una mayor competencia y productividad serán un imperativo para el crecimiento en América Latina.
- El 2015 se vislumbra como el año de la reestructuración. Los proveedores de servicios deberán ser capaces de distinguir entre crisis y oportunidad.

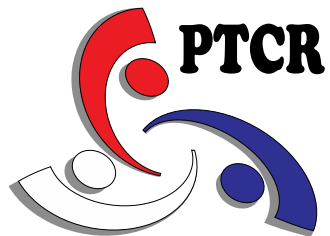
Los desafíos actuales...





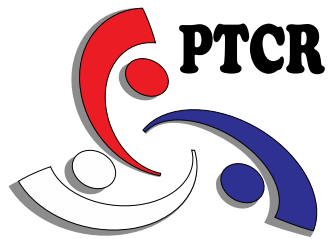
DESAFIOS POR SECTOR

- Academia:
 - Desarrollar Científicos de Datos
 - El Big Data requiere profesionales más integrales (interdisciplinales)
 - Investigación debe estar enfocada a desarrollar modelos más eficientes



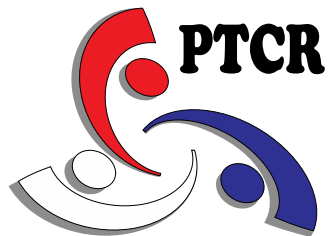
DESAFIOS POR SECTOR

- Industria:
 - Generar equipos y modelos de negocio que permitan el BI
 - Romper los paradigmas del análisis traccional como única forma de ver los datos.



DESAFIOS POR SECTOR

- Gobierno:
 - La legislación en nuestros países se está quedando muy resagada con respecto al avance de las tecnologías de información.
 - Dotar de fondos para la investigación e innovación.



DESAFIOS POR SECTOR

- Sociedad:
 - Llegar a un acuerdo social sobre lo que es privacidad y uso apropiado de la información.
 - Tomar “conciencia” de su gran responsabilidad en el proceso.
 - Tomar un rol activo.

The image features a stylized logo composed of several curved, ribbon-like shapes and circles. A large red shape curves from the top left towards the center, with a smaller red circle positioned below it. A blue shape curves from the center towards the bottom right, with a smaller blue circle positioned below it. At the bottom left, there is a white shape that appears to be a stylized 'E' or a similar letter, with a smaller white circle above it. The shapes have a slight 3D effect with a grey shadow underneath. The text 'PTCR' is positioned in the upper right area, and 'ESPACIO ABIERTO' is in the lower left area.

PTCR

ESPACIO ABIERTO