intel Security

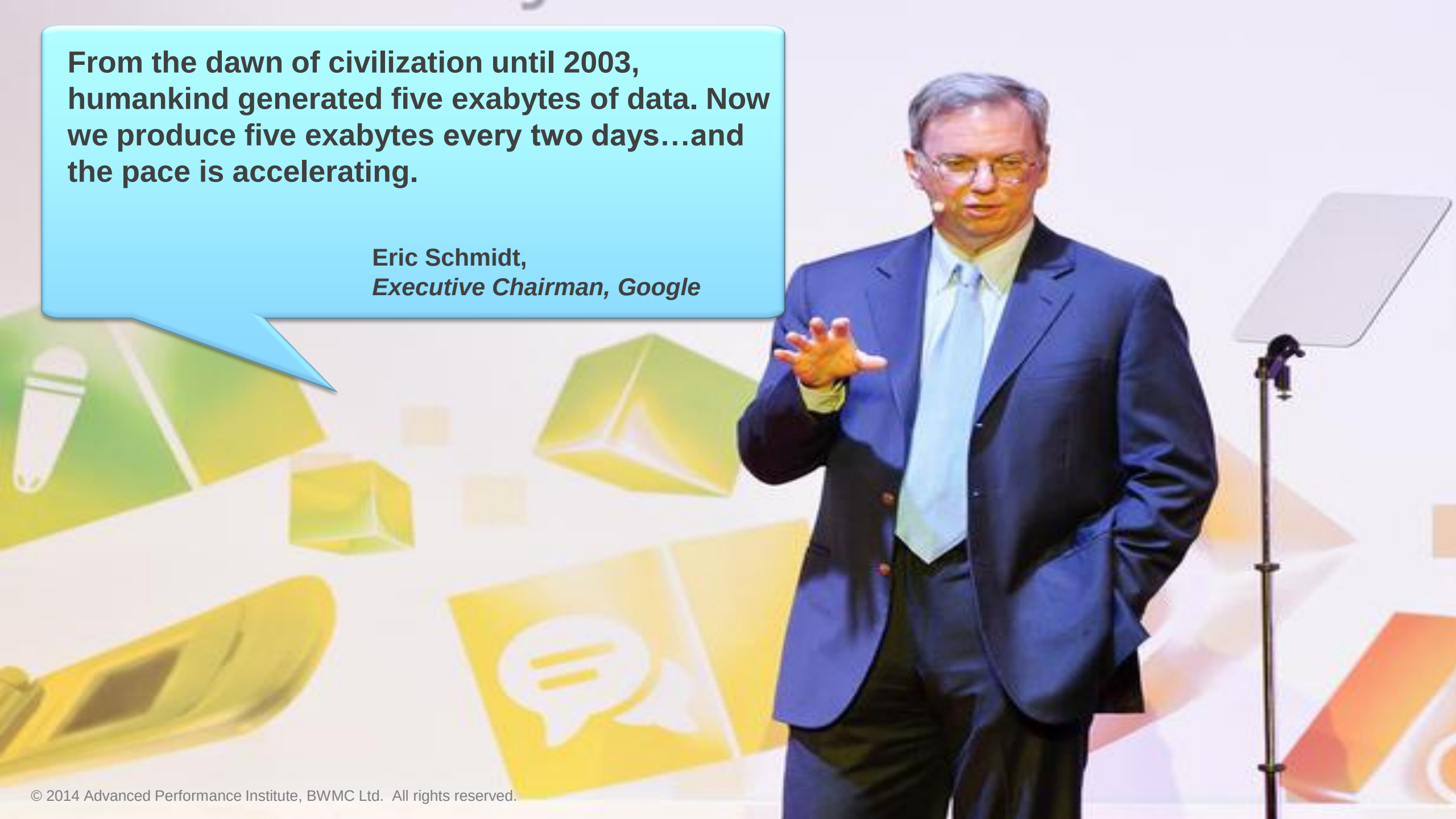
Argentina Software Design Center

Noviembre – 2014

Gabriel Infante-Lopez | Software Architect/Big Data Expert

From the dawn of civilization until 2003, humankind generated five exabytes of data. Now we produce five exabytes every two days…and the pace is accelerating.

Eric Schmidt,
Executive Chairman, Google



Tres Componentes de BigData

- **Minado de Datos**: Algoritmos de Machine Learning o Knowledge Discovery que encuentran datos y relaciones en los datos. e.g., mahout
- **Representación de los datos**: Manera conceptual como los datos serán almacenados en disco o memoria. e.g, modelo relacional en sql, key-halue en HBase.
- **recuperación de los datos:** que expresividad ofrece el motor para recuperar información. SQL, key extraction, traversal languages.

Recommendation Systems

Dept. of Comp. Engg.

What is a Recommmendation System?

Recommendation system is an information filtering technique, which provides users with information, which he/she may be interested in.

Examples:

Intel en Cordoba

asdc
argentina software design center

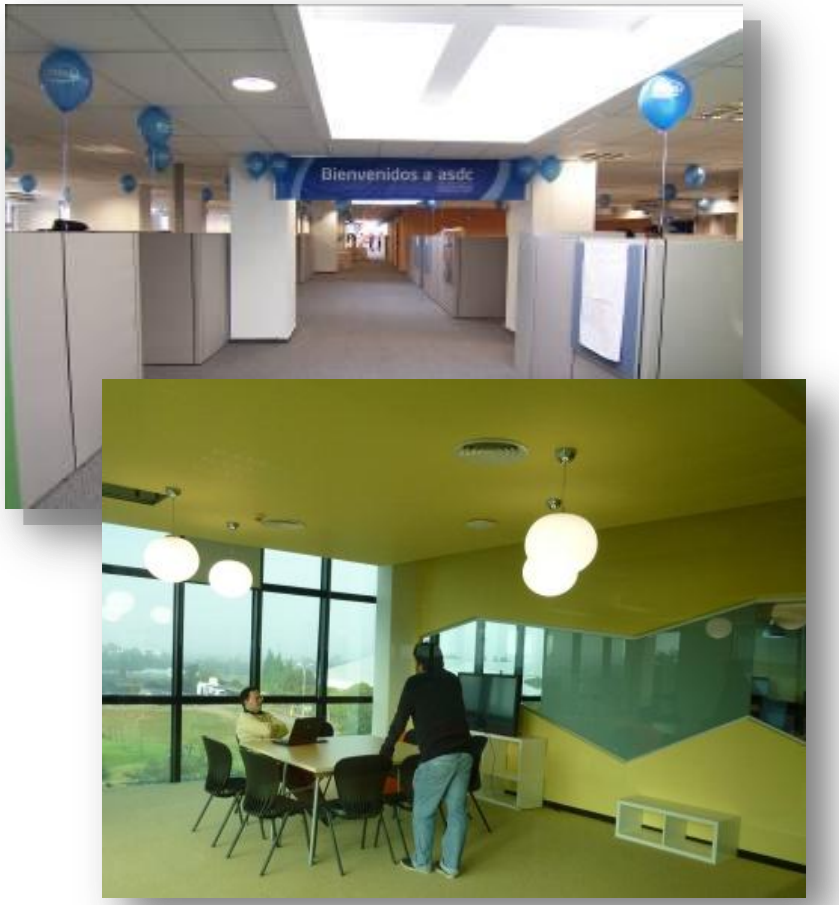
intel Security

Estamos hace 8 años en Cordoba

Varios proyectos relacionados con Cloud y SaaS

- Contribuidores a **OpenStack** (top 5)
- **Intel Update Manager®** deployed in 15M devices
- **Intel AppUp®** +6M clients deployed and 1M active users
- *Context aware client for Lenovo Secure Cloud Access ®*
- **Intel Identity Service®** (500k users)

Recientemente incorporamos 5 proyectos de Intel Security



Elementos en la recomendación

- **Social Network**: El sistema maneja usuarios, y estos tienen “amigos”, es decir una relación entre usuario, con potencialmente un peso.
- **Peliculas**: El sistema maneja peliculas, y los usuarios pueden dar feedback sobre las peliculas. El feedback puede ser escrito y/o a traves de la asignación de una cantidad de estrellas.

Elementos en la recomendación

- **Social Network**: El sistema maneja usuarios, y estos tienen “amigos”, es decir una relación entre usuario, con potencialmente un peso.
- **Peliculas**: El sistema maneja peliculas, y los usuarios pueden dar feedback sobre las peliculas. El feedback puede ser escrito y/o a traves de la asignación de una cantidad de estrellas.

Que datos se pueden descubrir:

- **Similitud entre Peliculas**: Cuan similares son dos peliculas, aqui puede haber varios tipos de similitud, por director, por los actores, por la trama, etc. cada dimension puede generar una similitud diferente.
- **Similitud entre Usuarios**: Cuando similares son dos usuarios, tambien diferente ejes, dependiendo de que se use. Podra ser por las peliculas que han visto, o por su relacion de amistad
- **Analisis de Sentimiento:** Como se habla en los reviews. Podemos analizar el lenguaje utilizado en cada uno de los reviews y decidir a partir de ahi si el review es positivo o negativo. Mejor aún podemos minar cuan parecido es un review a otro para solo mostrar los mas adecuados.

Como representamos datos (en modelo relacional)

- **Similitud entre Peliculas**: Una tabla de relaciones entre peliculas. La tabla contiene el peso de la relación. Cuantas mas parecidas, menos peso. Distintos tipos de relaciones, en distintas tablas.
- **Similitud entre Usuarios**: Una tabla de relaciones entre usuarios que indique la similitud entre usuarios, puede haber mas de una tabla.
- **Analisis de Sentimiento:** Un atributo extra asociado a los reviews.
- **Red Social:** Una tabla para los usuarios, otra para las peliculas, una para las relaciones entre usuarios.

Que podemos preguntar? (en modelo relacional)

- **Similitud entre Peliculas**: Peliculas similares a una dada. Un acceso a la base, y un ordemiento.
- **Similitud entre Usuarios**: Usualios similares a uno dado, igual que antes.
- **Analisis de Sentimiento:** Peliculas con sentimiento positivo. tambien un acceso a la base.

Posibles recomendaciones:

- Peliculas similares a peliculas que me hayan gustado.
- Peliculas similares a peliculas que hayan visto mis amigos.
- Peliculas que hayan recibido algun review positivo por amigos de mis amigos.
- Amigos similares a mi que sean amigos de mis amigos.
- Personas que sean amigas a mi que hayan escrito reviews similares.
- Peliculas que sean similares en actores y tema a algunas que me hayan gustado.

Problemas:

- Las tablas de relaciones crecen cuadráticamente.
- Las tablas tienen toda la información que necesitamos.
- Las recomendaciones anteriores se pueden resolver con joins
- Cuanto mas compleja la recomendación, mayor cantidad de joins.
- Motores sql y muchos no-sql no soportan una gran cantidad de joins,
- ninguno ofrece una minimización en pesos de cada uno de los joins.

Conclusiones

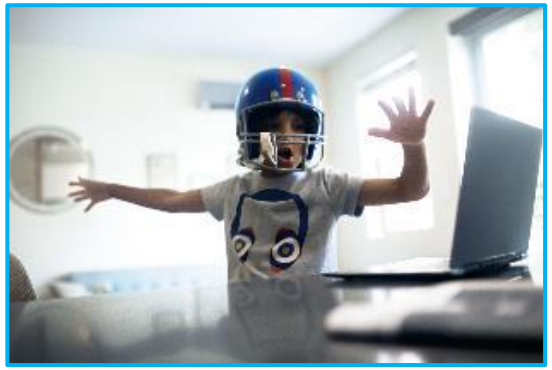
- Aún teniendo la información, no podemos recuperarla.
- Una estrategia de Big Data debe contemplar la representación de los datos, y los caso de negocios que atraparemos.
- Muchas veces, es mas importante la representacion y la recuperación que el minado propiamente dicho.
- Bases de datos de grafos pueden ser una solución al “traversing” de datos.

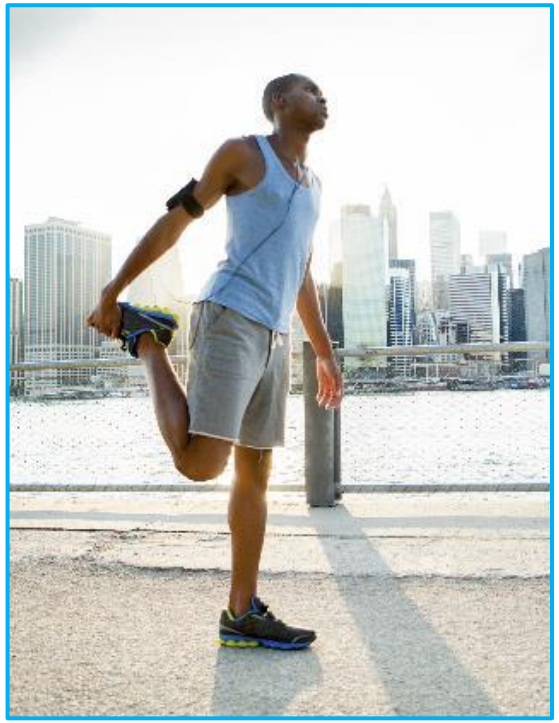
Intel Corporation

La empresa más grande del mundo de Semiconductores

- Principal fabricante mundial de productos de computación, sistema de redes y comunicaciones.
- **170** oficinas en **66** países.
- US$ **52,7** mil millones en ingresos anuales
- Clientes en más de **120** países.
- Cerca de **107.000** empleados , **84,600** roles técnicos**, 10,200** Masters in Science, **5,400** PhD's, **4,000** MBA's
- Una de las **10 marcas más valoradas en el mundo** por 10 años consecutivos.
- Cada año invertimos unos **US$100 millones** en educación en más de **100** países.
- **4 millones** de horas dedicadas al servicio voluntario en nuestras comunidades en la ultima década.

Nuestra Vision



En esta década crearemos y expandiremos la informática Para conectar y enriquecer las vidas de cada persona en la tierra.



Intel® Education

Empowering Youth. Transforming Communities.

150M estudiantes aprendiendo con tecnología

10M profesores potenciados con desarrollo profesional

7M estudiantes inscriptos a competiciones de ciencias

4M Horas de voluntariado para la Educación

$100M de Inversión Anual
para mejorar la Educación en
100 países

Intel en Latinoamérica

Costa Rica Site: ATM, FSSC, SMG
2,650 empleados
Gastos Operativos: US$400M

Guadalajara Dev. Center
850 empleados
Gastos Operativos: US$100M

Ventas US$ 4 mil millones
Inversión annual en Ventas y Marketing $300M

Argentina SW Dev. Center: SSG
150 empleados
Gastos Operativos: US$23M

Sources: Consumption Fin, Field Fcst, idc Q1'13, F&A data

Algunos hitos en de la historia de Intel en Córdoba

2003 — Intel se compromete a invertir en Latinoamérica

2004 — Se estudian algunas alternativas

2005 — Se selecciona a Córdoba

2006 — Se firma el acuerdo marco

Nuestra 1era oficina

Se contratan los primeros 13 pioneros

2007 — Nuestra 2da Oficina

2009 — Empleado nro 100

Nuestra 3era oficina

2012 — Empleado nro 200 Nuestra 4ta casa

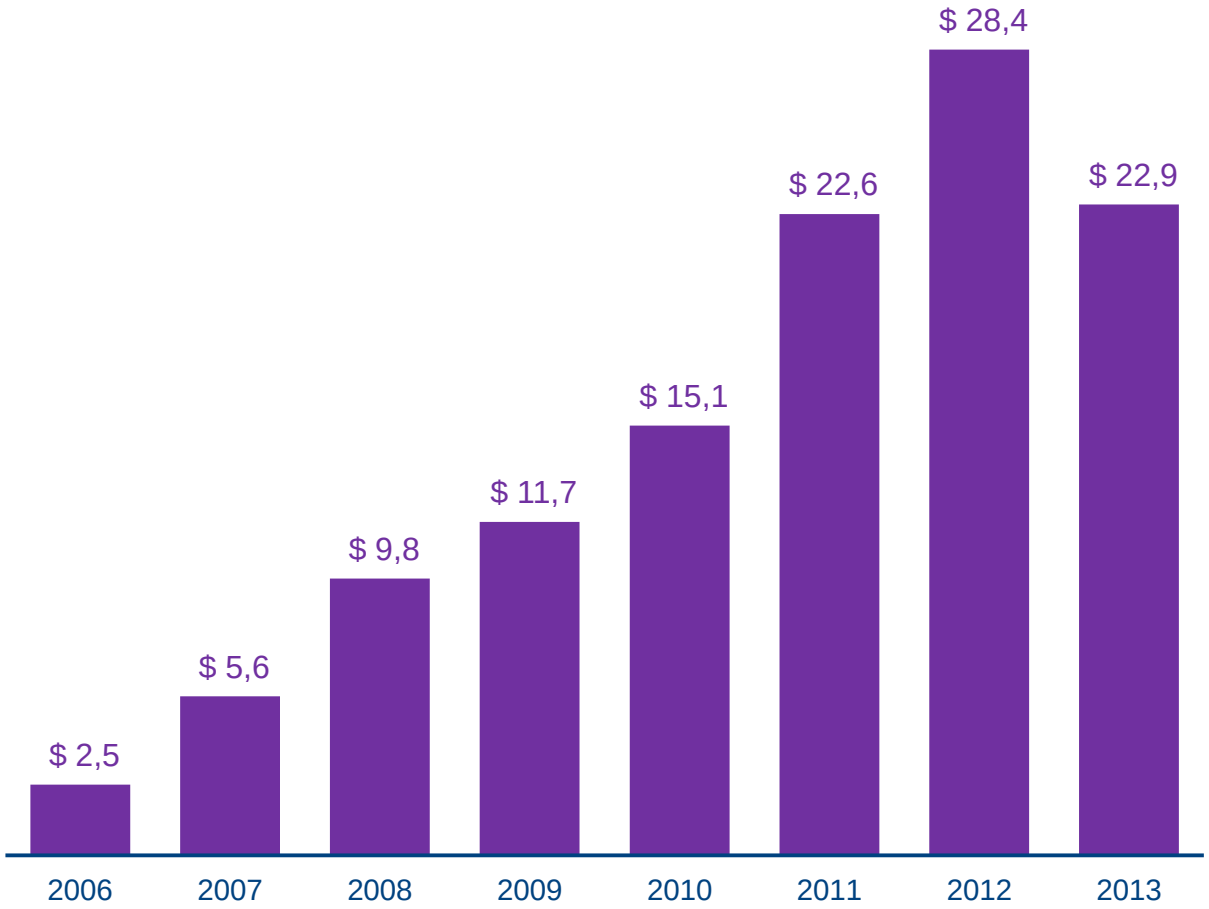
Enmienda al acuerdo marco

Nuestros principales logros

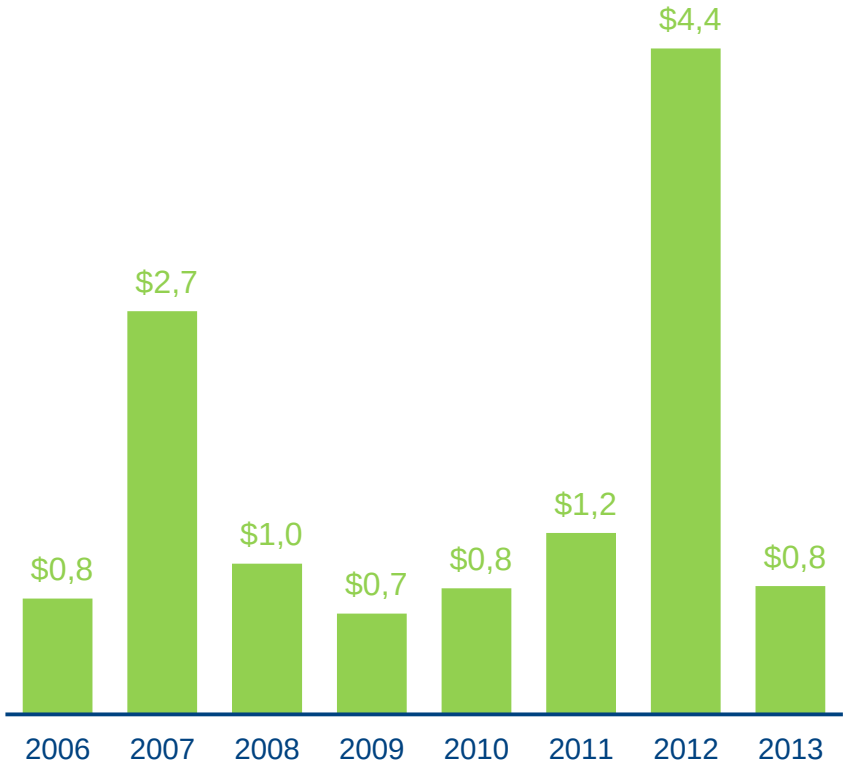
	2006	2007	2008	2009	2010	2011	2012	2013
Productos	XML Security Gateway	First ASDC product: Intel SOA Security Toolkit	Tango Creek Conceived and engineered	AppUp Conceived and engineered First Intel Cluster Ready Certified	AppUp beta @ CES Record time of 100 days Parallel Universe	AppUp expansion and distribution Personal Vacation Assistant Pilot	Mate Creek For McAfee HTML 5 Porting Tool beta launch	Intel® Cloud Services Platform You are your Password For McAfee
Premios y Certificaciones					DFS L3 First IAA #7	#9	#16 Recertificación	

Impacto económico de ASDC en Córdoba

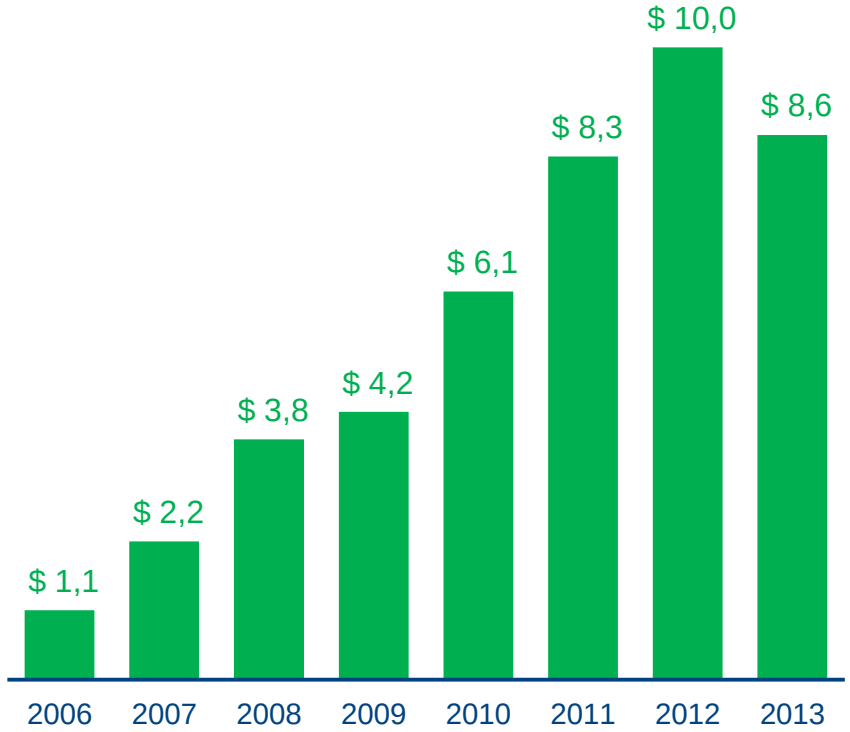
US$ 120 millones acumulados de exportaciones de servicios de software



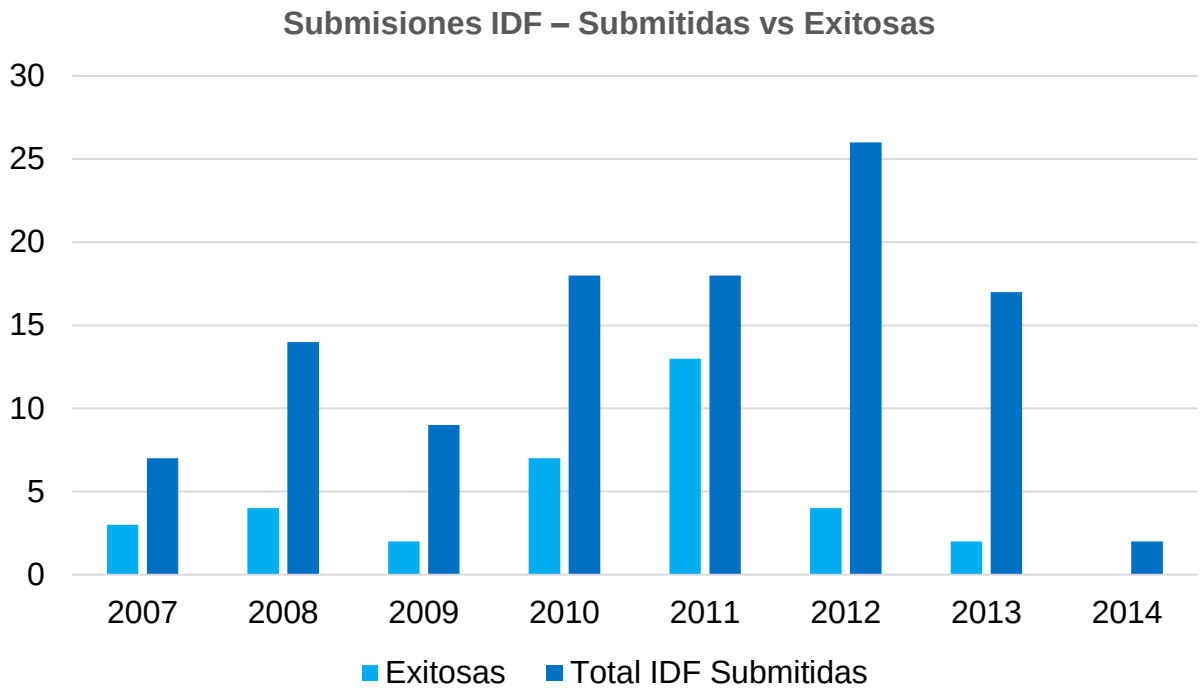
US$12M de inversión acumulada en infraestructura



US$ 45M de inversión acumulada en sueldos



Inventos cordobeses en el Sillicon Valley



111 IDF Submitidos

35 IDF Exitosas presentadas a la USPTO

6 patentes ya otorgadas

1 Trade Secret

6 ingenieros de ASDC

llegaron al top 36

“Most IDF approved in 2011”

intel Security

Emprendedurismo

Desafío Intel: capacitación, viajes al Sillcon Valley para presentarse ante inversores, exposición mediática y premios por USD 100K

Intel Catapult: infraestructura de Intel en nuestras oficinas por 4 meses, asesoramiento técnico y coaching en el desarrollo del negocio, premio de US$ 15.000. 2 emprendedores ganadores. Promocionado por el Ministerio de Industria.

Patrocinio institucional: Endeavor Córdoba y Programa de emprendedores Universidad Siglo XXI

Network: conexión del ministerio de Industria con capitales de

Y por último lo más importante, alentamos a nuestros empleados a ser buenos ciudadanos

Dimos charlas en escuelas sobre privacidad en redes sociales
Reciclamos tapitas y papel para Asociación Hospital Infantil
Ayudamos a la Asociación Surcos Argentinos en la siembra de otoño primavera
Colaboramos en la clasificación de alimentos con la Fundación Banco Alimentos
Dimos charlas en Escuela Sarmiento para inspirar a los niños a estud ingeniería
Recolectamos regalos de navidad para los niños de Chancaní
Y por cada hora de voluntariado, Intel dona US$ 5 a una institución