



3º Congreso Prevención de Riesgos Laborales Personas, Tecnologías y Tendencias

Bilbao, 6 de Octubre de 2016

Partners

AENOR





Big Data, 10 años después

Carlos Hernández Paradela
Analytics Senior Manager

Accenture

Big Data



Application-Controlled Demand Paging for Out-of-Core Visualization

Michael Cox and David Ellsworth¹
Report NAS-97-010, July 1997

NASA Ames Research Center
MS T27A-2
Moffett Field, CA 94035-1000

Julio 1997

Los investigadores de la NASA Michael Cox y David Ellsworth utilizaron el término Big Data por primera vez al prever las dificultades que tendrían los sistemas informáticos de la época para manejar ritmo de crecimiento del volumen de datos

Febrero 2001

Doug Laney, analista de Gartner, publicó un artículo en donde hacía mención a las 3 Vs (Volume, Velocity, y Variety).

Abril 2006

El proyecto Apache Hadoop libera su primera versión (0.1.0), que se convertiría en las bases del desarrollo de toda la tecnología Big Data actual.

Septiembre 2016



Big Data

Big Data no es una tecnología concreta, sino una situación en la que las **necesidades de almacenamiento y procesamiento** de información estructurada y/o no estructurada **sobrepasan las capacidades** de las tecnologías tradicionales.

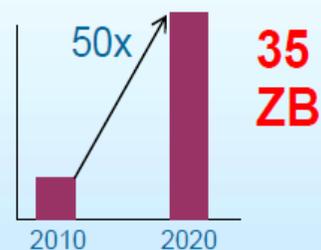


*Sistemas y recursos de información de alto volumen, alta velocidad y alta variedad que requieren nuevas y eficientes maneras de procesar la información **

Nuevas tecnologías para la explotación de datos con alto rendimiento y eficiencia en costes

* Definición de Gartner: <http://www.gartner.com/it-glossary/big-data/>

Procesamiento eficiente en costes de los **volúmenes** crecientes



Respuesta a la **velocidad** creciente

(En generación y analítica)



Tiempo de respuesta *near real-time*

Analizar conjuntamente la mayor **variedad**



El 80% de los datos son no estructurados

Big Data

Las plataformas Big Data están permitiendo a las compañías **integrar y almacenar distintas fuentes de datos**



Nuevos datos internos

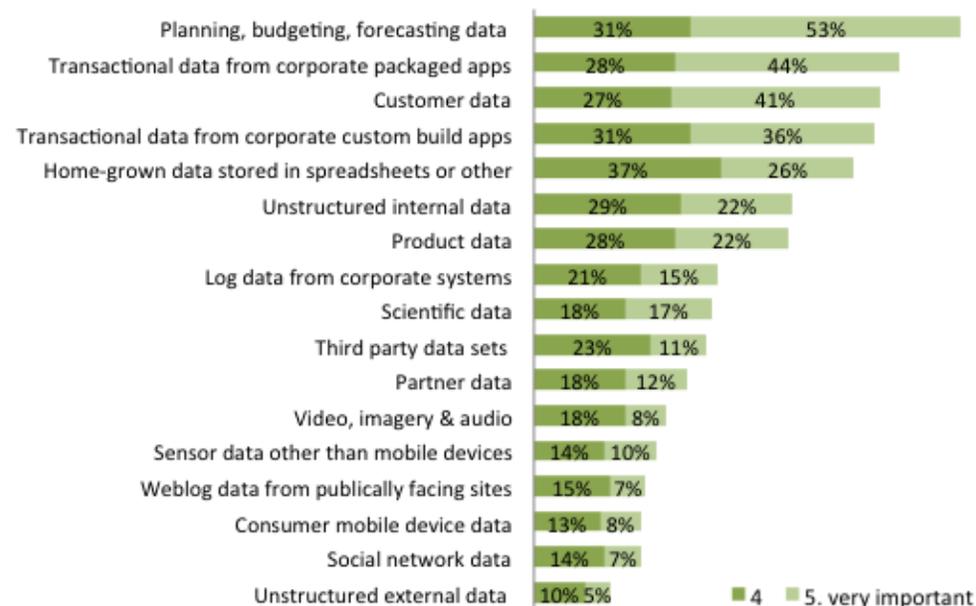
- Webs propias
- Información de cliente
- Sensores
- Smart meeters

Antiguos datos internos

- Logs
- Emails
- Archivos corporativos
- Documentación
- Vídeo
- Voz

Nuevos datos externos

- Internet (APIs, feeds)
- Social Media
- Movilidad
- Geolocalización



Valoración de las fuentes según su relevancia en la estrategia de negocio

Las fuentes consideradas más relevantes en las estrategias Big Data son los datos corporativos, transaccionales y de clientes

Big Data

Este nuevo contexto supone un reto que requiere adoptar **soluciones y tecnologías basadas en nuevos paradigmas**

1. Analítica avanzada

Análisis de datos para resolver problemas específicos de negocio

A

Analítica estratégica

Insights para la resolución de nuevos problemas o desafíos

B

Plataforma analítica

Optimización de los procesos core y la operativa diaria

C

Transformación de los procesos

Optimización de la cadena de valor gracias a los datos

D

Nuevos modelos basados en datos

Nuevos servicios, productos y publicación de datos

2. Transformación de las organizaciones

Los datos como palanca para la generación de valor, generación de conocimiento y optimización de procesos

Nuevos sistemas de Business Intelligence

Nuevas capacidades y modelos de operación

Big Data

Superando las limitaciones tecnológicas y organizacionales existentes anteriormente

Analítica Actual

...almaceno datos para analizar

- Fuentes de datos internas, limitadas y estructuradas
- Los datos tienen que almacenarse en Datamarts antes de ser analizados
- La mayoría de la analítica es descriptiva o reporting
- La creación de modelos analíticos es un proceso batch que requiere meses de desarrollo
- Los analistas están separados de los responsables y decisiones de negocio



Analítica Big Data

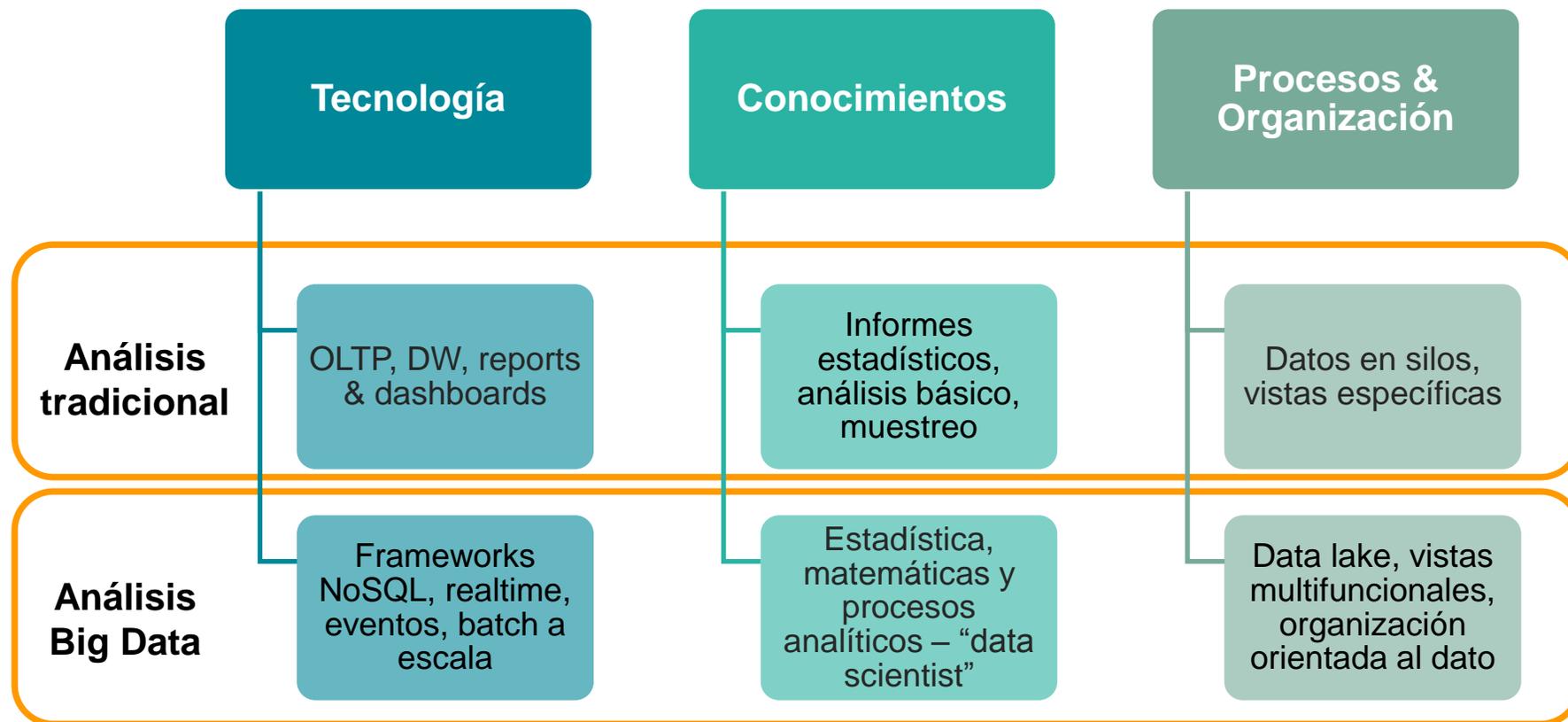
...inteligencia hacia los datos

- Fuentes internas y **externas, ilimitadas** (por bajo coste), estructuradas y **desestructuradas**
- Los datos pueden ser analizados desde la **capa batch** o un Data Warehouse común
- Cuadros de mando e informes **interactivos**, analítica predictiva y prescripción
- El aumento del rendimiento **simplifica** o hace innecesarios los modelos dimensionales, cubos, vistas o la capa semántica
- IT proporciona una **plataforma** y los responsables de negocio **trabajan directamente** con los analistas

Es preciso poder gestionar un ecosistema complejo de datos y tecnologías para descubrir y aportar el valor oculto y diferencial

Big Data

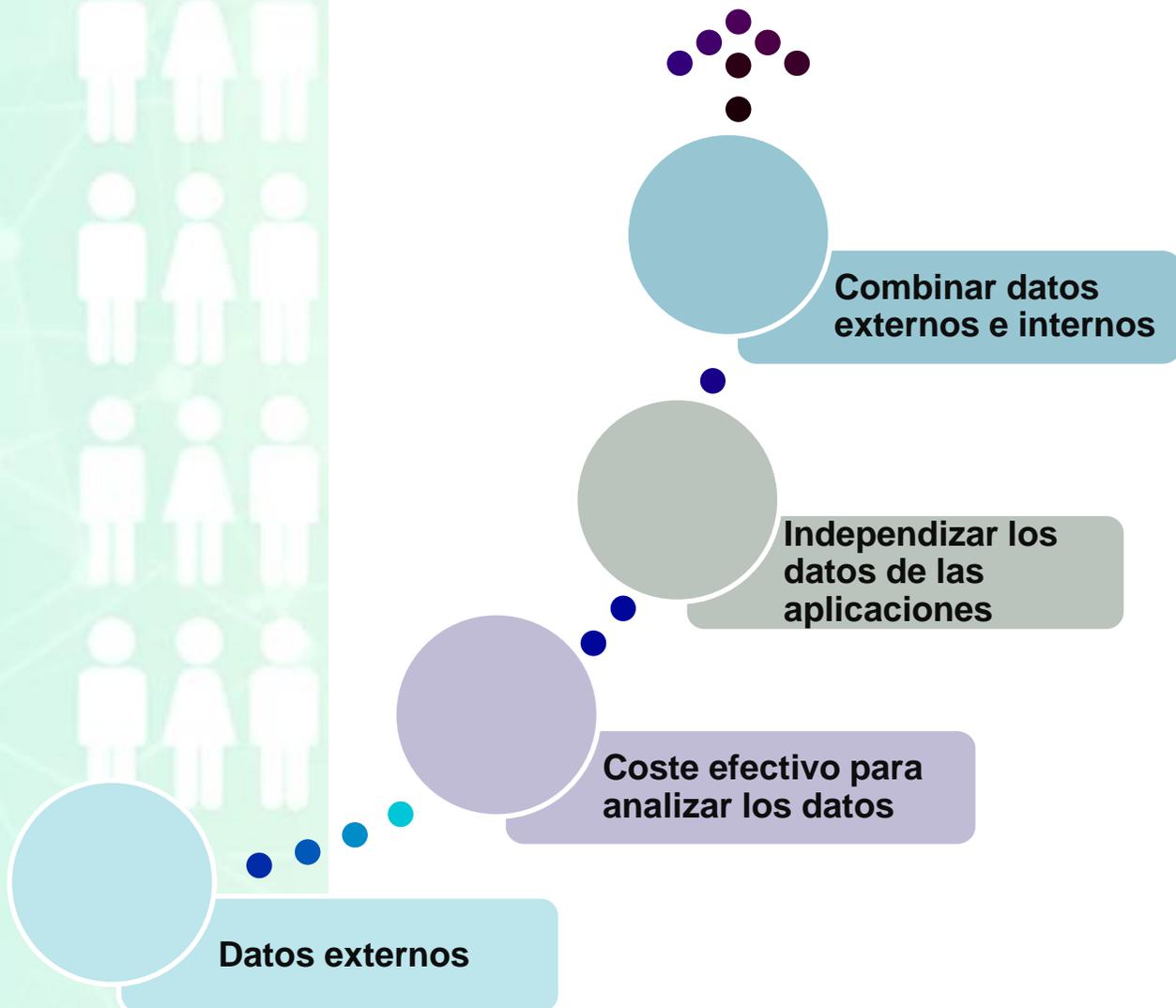
El análisis estructurado está optimizado, no así el no estructurado en cuanto a herramientas, conocimientos y procesos a escala



Big Data



Algunos de los aspectos tecnológicos predominantes



- Combinar datos sociales y de terceros con información interna mejorando el perfil del cliente
- Conocer el contexto social y circunstancial que rodea al individuo
- El dato externo aumenta su valor respecto al interno

- Múltiples posibilidades para adquirir y almacenar la información
- La analítica a gran escala ya es posible sin necesidad de grandes proyectos
- Soluciones multiplataforma y proveedor

- Flexibilidad y acceso a los datos para el análisis
- Modelado global no focalizado a un área de negocio
- Flexibilidad y agilidad
- Reducción de costes del entorno tecnológico

- Almacenar y procesar datos a una escala nunca contemplada
- Combinar datos de múltiples Data Warehouse en plataformas como Hadoop
- Obtener beneficios operativos o relacionados con sus clientes

Big Data

Fórmula para el éxito del Big Data: Las cuatro T

Tipos



¿Qué hace “Big” a los datos?

Cantidad de Datos

Origen de los Datos

Velocidad de los Datos

Complejidad de los Datos

Tecnologías



¿Qué tecnologías utilizan Big Data?

Plataformas de datos y sistemas gestores de bases de datos

Herramientas avanzadas de análisis

Tecnologías in-memory

Herramientas de visualización

Técnicas



¿Qué técnicas necesito para tener datos significativos?

Data Science

Data Design

Data Discovery

Análisis de Redes Sociales

Gestión de datos maestros

Diseño Visual

Análisis Avanzado

Arquitectura Big Data

Talento



¿Qué talento necesito para convertir Big Data en grandes resultados?

Analistas de Datos

Gestores de Datos

Desarrolladores

Estrategas de Negocio

Arquitectos de Sistemas

Ingenieros de Software

Analistas Cuantitativos

Visualization Designers

Big Data



CASOS PRÁCTICOS

Casos prácticos

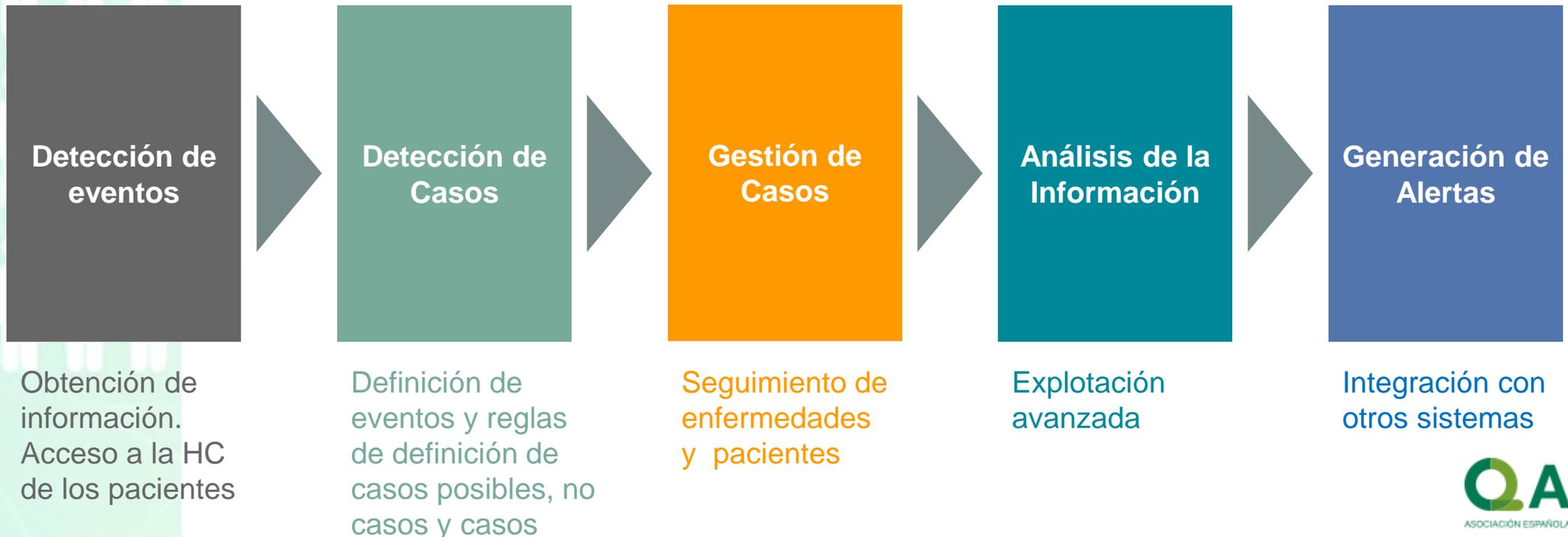
Caso 1 – Servicio de Salud – Vigilancia Epidemiológica

Objetivos y enfoque

- Creación de un sistema que centralice información clínica de los pacientes y que permita la detección temprana de casos medicamente relevantes mediante la aplicación de reglas, analítica avanzada y generación de alertas. Obteniendo evidencia clínica y ayudando al procesos de investigación.

Gestión de la información

- Pasos en el tratamiento de la información para la obtención de datos útiles:

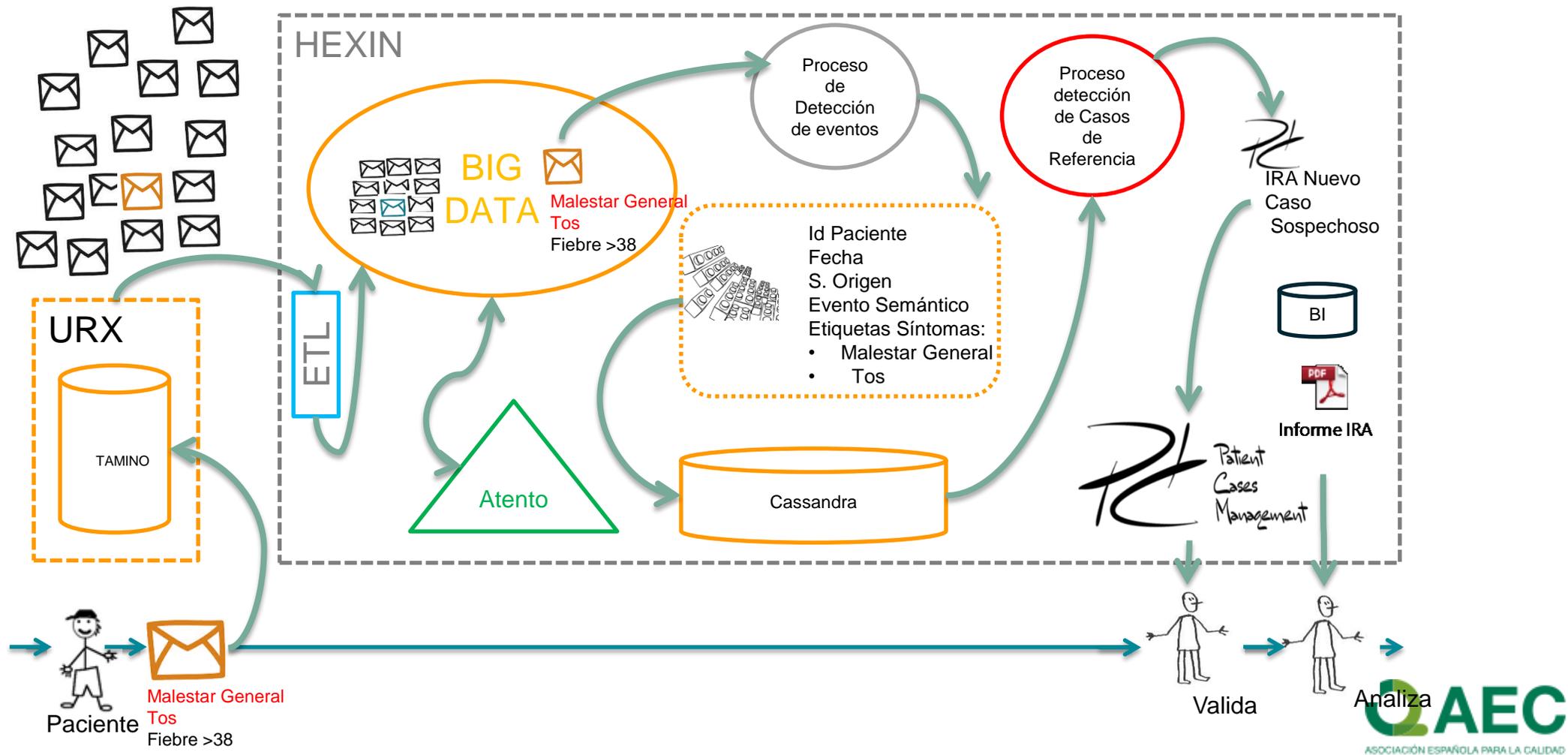


Casos prácticos

Caso 1 – Servicio de Salud – Vigilancia Epidemiológica

Modelo de solución propuesto

- Ejemplo para el caso de uso de sindrómicas:



Big Data

Caso 2 – Análisis en Social Media

Security department monitoring the company's reputation in traditional and social media

- The problem: due to the large volume of data in the Social Media, identifying potential risks is delayed with the consequent danger of not tackling these reputation problems timely.
- Solution: AQUA SM software implementation, which enables real-time analysis of buzz in Social Media

Capacity

Over 10.000 documents processed per day

Alerts

The system prioritizes potentially dangerous content, activates alerts, and send the messages for review within minutes

- High risk alerts are sent with more than 85% of reliability
- Low risk alerts are sent with more than 98% of reliability

Cost saving

Automatic exclusion of harmless content:
Over 50% documents excluded, saving costs of manual revision



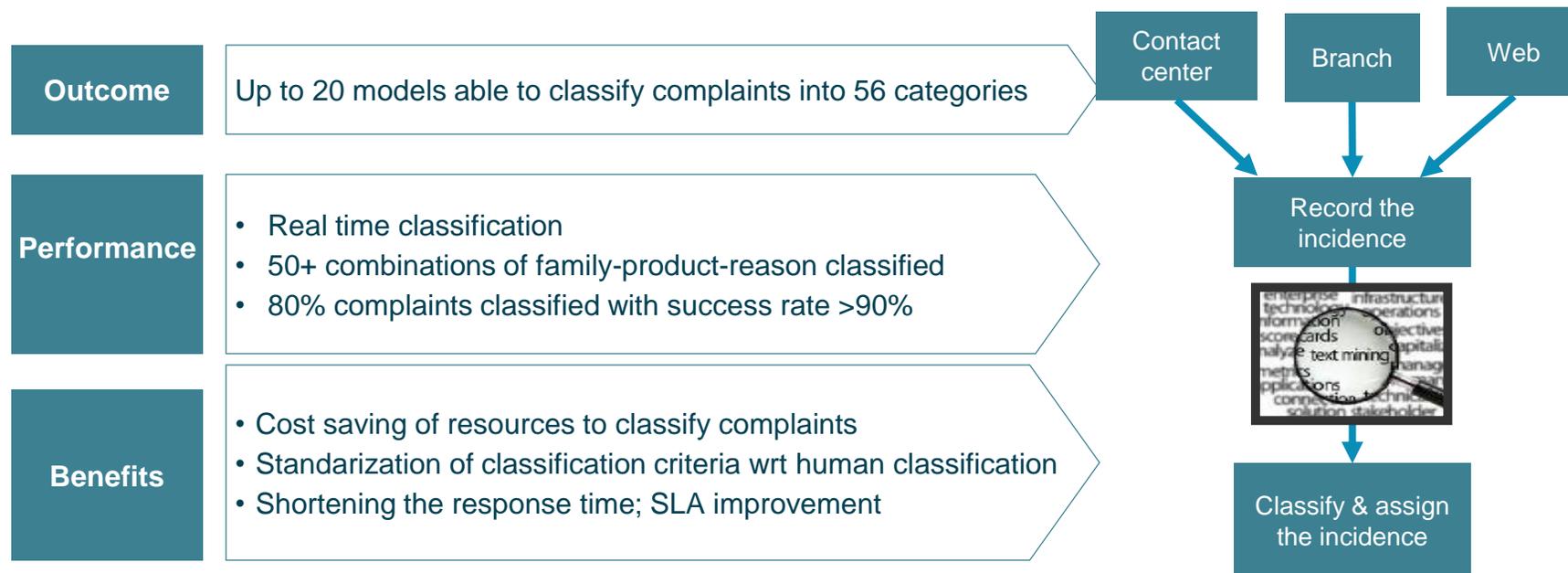
Early tackling
of reputational
problems

Big Data

Caso 3 – Análisis en Social Media

Leading bank – corporate tool for complaints management: multichannel and multiproduct

- The problem: Early assignment of resolution group of incidents, to ensure customer satisfaction
- The solution: Based on the description of the incidence (unstructured text) the incident must be classified into the following categories: family-product-reason, forwarding **the incidence to the right resolution group**



Big Data

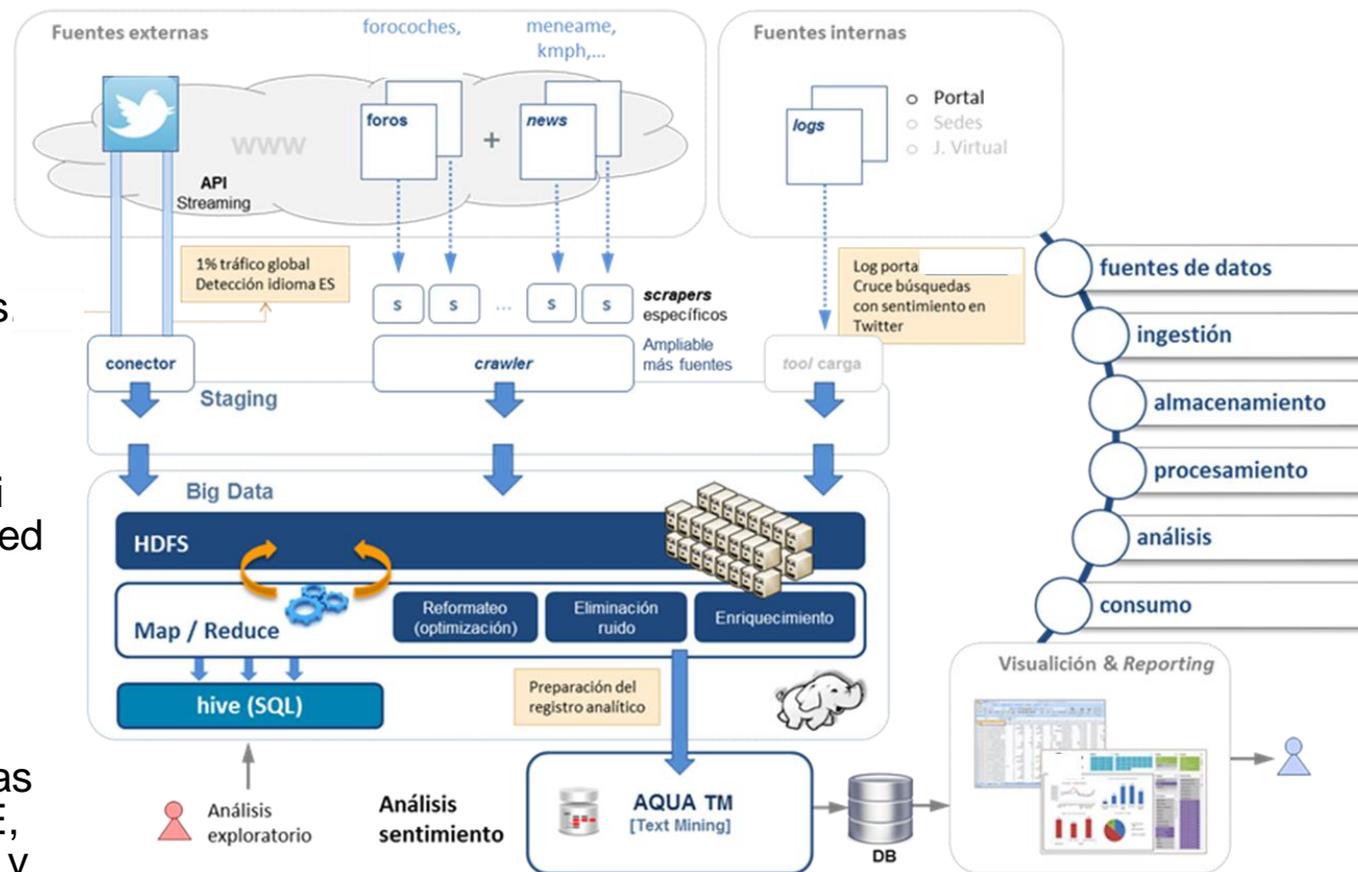
Caso 4 – Impacto en campañas de prevención en tiempo casi real

Reto

- Solución que permita adquirir y analizar en tiempo real datos externos heterogéneos y medir la efectividad de las diferentes campañas de sensibilización dirigidas a conductores y diferentes colectivos profesionales

Resultado

- El organismo ya dispone de un sistema que le permite evaluar casi en tiempo real su relevancia en la red social Twitter y medir impacto en la misma de las diferentes campañas lanzadas.
- Algunas de las tecnologías utilizadas fueron: Hadoop, MapReduce, HIVE, Twitter Streaming API, PowerPivot, y Aqua TM para modelos avanzados de minería de texto.



Big Data

Varios en sector Salud

| | Multi-crónicos Analytics predictivo | Alarmas multinivel | Fraude farmacéutico |
|----------------------|---|--|--|
| Reto de negocio | <p>Reducir el consumo de recursos sanitarios de los pacientes multi-crónicos.</p> <p>Reducir camas hospitalarias y mejorar el equilibrio entre actividad hospitalaria programada / no programada.</p> | <p>Mejorar las capacidades de diagnóstico de los profesionales para evitar problemas de salud derivados de errores.</p> <p>Detectar lo antes posible situaciones de riesgo, con el objetivo de abordarles en etapas tempranas.</p> | <p>Control del gasto farmacéutico desde el punto de vista de inspección y reducción del fraude farmacéutico.</p> <p>Identificar posibles casos de fraude farmacéutico o malas prácticas de prescripción y dispensación, auditando el ciclo del medicamento.</p> |
| Enfoque de actuación | <p>Predecir con un año de antelación, qué pacientes tienen un mayor riesgo de generar hospitalizaciones no planificadas pero evitables.</p> <p>Aplicación de programas de gestión del caso sobre estos pacientes.</p> | <p>Utilizar la información disponible en la HCE y la generada por los SSII del hospital para mediante métodos avanzados de análisis de datos generar alarmas relevantes para los profesionales.</p> | <p>Rediseño de los procesos de facturación, adaptados a la detección del fraude farmacéutico.</p> <p>Establecimiento de indicadores y controles específicos para mejorar en los procesos de auditoría (inspección), de las dispensaciones y prescripciones.</p> |
| Beneficios / ahorros | <p>Mejora de la capacidad de predicción.</p> <p>Reducción de hospitalizaciones no planificadas pero evitables en un 80%</p> | <p>Reducción de costes por atención de problemas de salud evitables.</p> <p>Mejora de la calidad asistencial.</p> | <p>Identificación de comportamientos o prácticas fraudulentas.</p> <p>Control del gasto farmacéutico desde la auditoría y control.</p> |