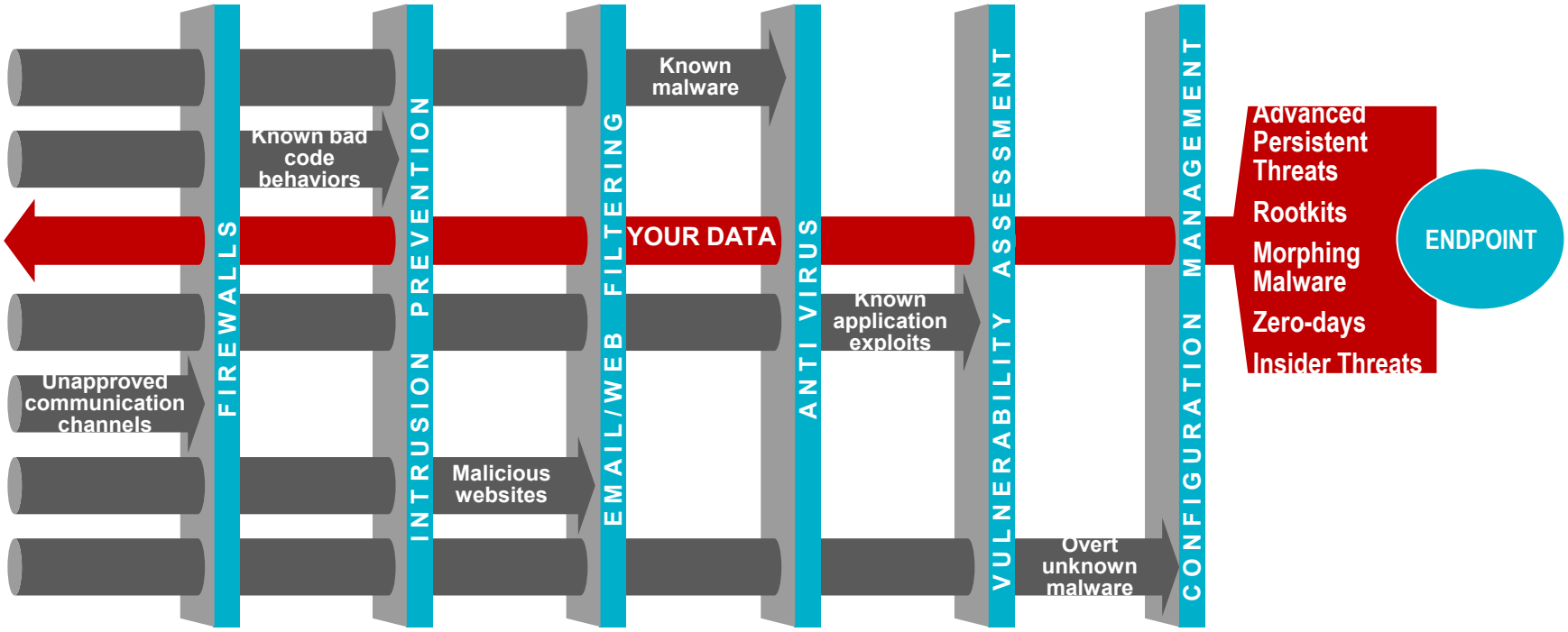
BIGDATA. CAMBIO DE PARADIGMA

SEMANA NAVAL DE LA ARMADA

JORNADAS TECNOLÓGICAS

23 Septiembre 2014

indra



El modelo tradicional de ciberseguridad está llegando al final de su ciclo por la proliferación masiva de dispositivos inteligentes. Según Gartner en el 2020 se alcanzará la cifra de 26mil millones de dispositivos conectados a internet.

Actualmente más del 50% de las amenazas permanecen sin descubrirse durante meses.

Es en este nuevo escenario, donde el tamaño "SÍ" importa ya que la información generada relativa a eventos es directamente proporcional a los dispositivos conectados, se genera un problema de procesamiento para las unidades de Ciberseguridad de las Organizaciones.

En paralelo a estas nuevas reglas de juego en el mundo de la Ciberseguridad, se empiezan a aplicar técnicas de "BigData" que surgen motivadas de la necesidad de procesar cantidades masivas de datos de forma óptima (en términos de tiempo de respuesta y facilidad de crecimiento)

Business Analytics es una práctica extendida de aprovechamiento de los datos para **generar conocimiento útil** a las Organizaciones que escuchan comportamientos, los analizan y actúan de manera analítica.

La **agilidad** con que las decisiones importantes deben ser tomadas, la **complejidad** de los factores sociales, económicos, demográficos y políticos que influyen en cada contexto, y el gigantesco **volumen** de información accesible, justifican la adopción de soluciones BigData & Analytics para anticiparse y obtener ventaja en la decisión.

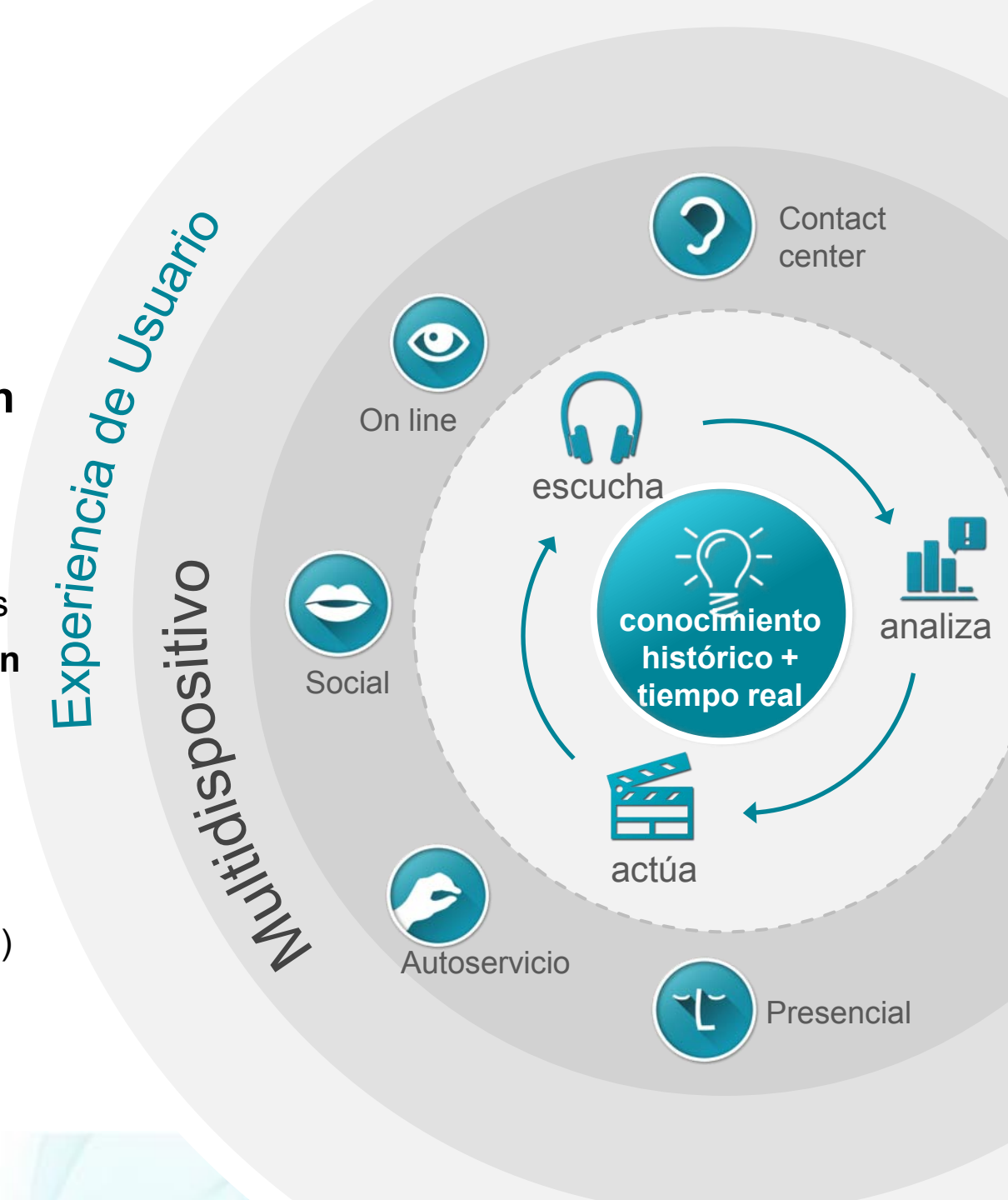
El binomio BigData & Analytics está provocando un cambio de paradigma en la gestión y la explotación de la información ...

Reto Organización

- **Democratización del dato** frente al privativo de expertos
- **Intensificación de la función analítica y colaborativa** "todos participamos"

Reto Gobierno de Datos

- El dato como **activo de la entidad** versus dominio de áreas funcionales (datamarts)
- **Relevancia** de información **externa y no estructurada**



El volumen, la variedad, la velocidad y la complejidad son vectores clave de la transformación en la gestión de los datos...

Enfoque tradicional	Nuevo paradigma
Gigabites a Terabytes	Petabytes a Exabytes
Almacenamiento centralizado	Almacenamiento distribuido
Persistencia de los datos	Volatilidad del dato
Datos estructurados	Cualquier tipo dato
Modelo de datos estable	Esquemas planos
Interrelaciones complejas conocidas	Pocas interrelaciones complejas
Alta calidad de datos	Datos sin tratar (crudos)
Alta transformación fuente de datos	Cualquier fuente de datos
Volumen limitado (agregación)	Cualquier volumen
Análisis Histórico (laboratorio dataming)	Análisis complejos (alertas, realtime machine learning)
Granularidad limitada	Máxima granularidad (no agregados)

…. y en el consumo de los mismos ...

Enfoque Tradicional

Analisis estructurado y repetido

Los usuarios plantean qué quieren consultar

TI estructura los datos para darle respuesta

Ejemplos:

Informes de rentabilidad, credit scoring, venta cruzada, incremental…

Enfoque Big Data

Análisis exploratorio e iterativo

TI proporciona la plataforma para almacenar el máximo posible de datos y que sea el usuario quien descubra información sobre ellos.

Los usuarios exploran qué consultas son las que pudieran ser respondidas

Ejemplos:

Alertas , Comportamientos Anómalos,

Estrategia de Defensa, etc.

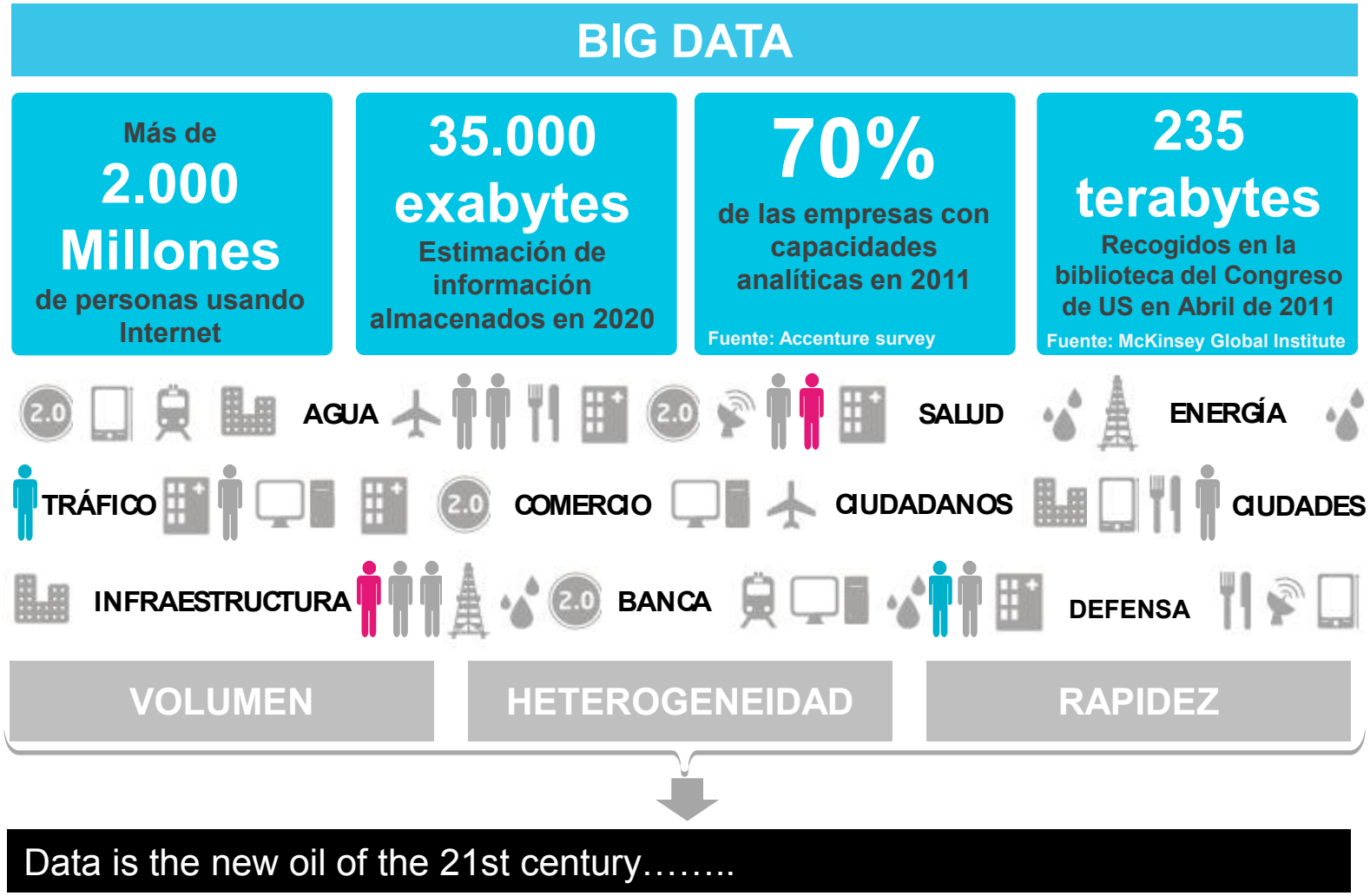
..que obligan a replantearse muchos de los conceptos asociados a los entornos tradicionales….

- A nivel de **Infraestructuras:**
  - Predominio de entornos escalables, distribuidos, arquitecturas abiertas y de coste contenido: entornos cloud, software libre,…
- A nivel de **Base de Datos:**
  - No es necesario encajar los datos en unas estructuras y esquemas fijos (tablas, filas, columnas).
  - En muchas de las nuevas tecnologías no se crean índices al resultar innecesario
- A nivel de **Procesos de carga:**
  - No es necesario crear agregados, se trabaja siempre al máximo nivel de detalle
  - En muchos casos no es necesario hacer procesos ETL (Extract, Transform, Load) pues se trabaja con los datos allí dónde están o bien se procesan en streaming, según se van generando.
- A nivel de **motores analíticos:**
  - Predominan técnicas avanzadas de procesamiento y análisis estadístico de datos intensivas en computación (machine learning): reglas de asociación, crowdsourcing, algoritmos genéticos (o de evolución genética), lógica difusa…
  - … o bien técnicas de procesamiento de información no estructurada: procesamiento de lenguaje natural…

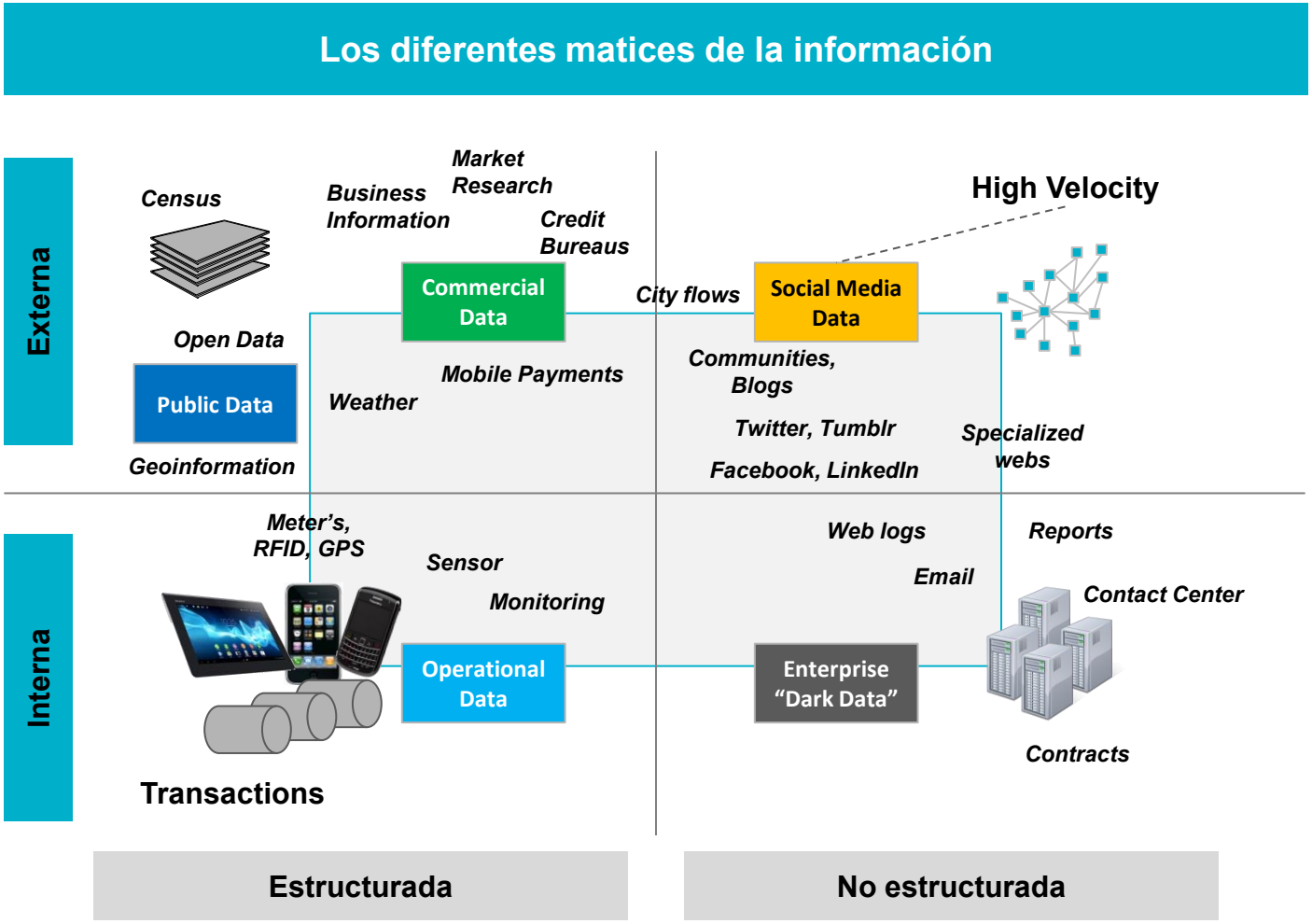
..que obligan a replantearse muchos de los conceptos asociados a los entornos tradicionales….

- A nivel de **herramientas de explotación y visualización:**
  - Se requieren nuevas técnicas visuales de representación de datos orientadas a información masiva (heatmaps, nubes de taggs, nodos de redes de individuos, …) alejadas de los tradicionales formatos ofimáticos y visualizaciones tradicionales de BI (dashboards, cuadros de mando gráficos,….):
- A nivel organizativo (áreas corporativas, analíticas, TI):
  - **Orientación analítica** desde la dirección del.
  - **Gobierno del dato** asignando competencias y roles relevantes en la gestión del activo información dentro de cada Organización.
  - Nuevos perfiles como los **Científicos de datos**: expertos en gestión de datos, metodologías científicas, estadística, matemáticas, computación avanzada, investigación, experto en determinados sectores o áreas de actividad
  - **Metodologías de trabajo**: metodologías ágiles de desarrollo, entornos abiertos y colaborativos.

… forzados por la superproducción de información estructurada y no estructurada, histórica y en tiempo real…..



Es necesario razonar qué resultados se buscan, a partir de qué información, para decidir cómo gestionarla correctamente…..



En BigData & Analytics se cubren los procesos clave de generación de conocimiento útil …..

…desde la **captura de los datos en bruto (Escucha)** a partir de cualquier tipo de fuente y origen, pasando por su **procesamiento y análisis (Analiza)** mediante diferentes herramientas y técnicas analíticas, hasta su **puesta en valor mediante acciones concretas (Actúa)** que puedan reportar beneficios al negocio

Captura >Interpreta>

- Identificar fuentes relevantes
- Extraer información
- Almacenar información
- Filtrar lo irrelevante

Categoriza > Entiende >

- Aplicar taxonomías y ontologías (traducir a conceptos)
- Combinar algoritmos convencionales con nuevas técnicas machine learning
- Aprender de forma atendida y desatendida

Alerta> Responde

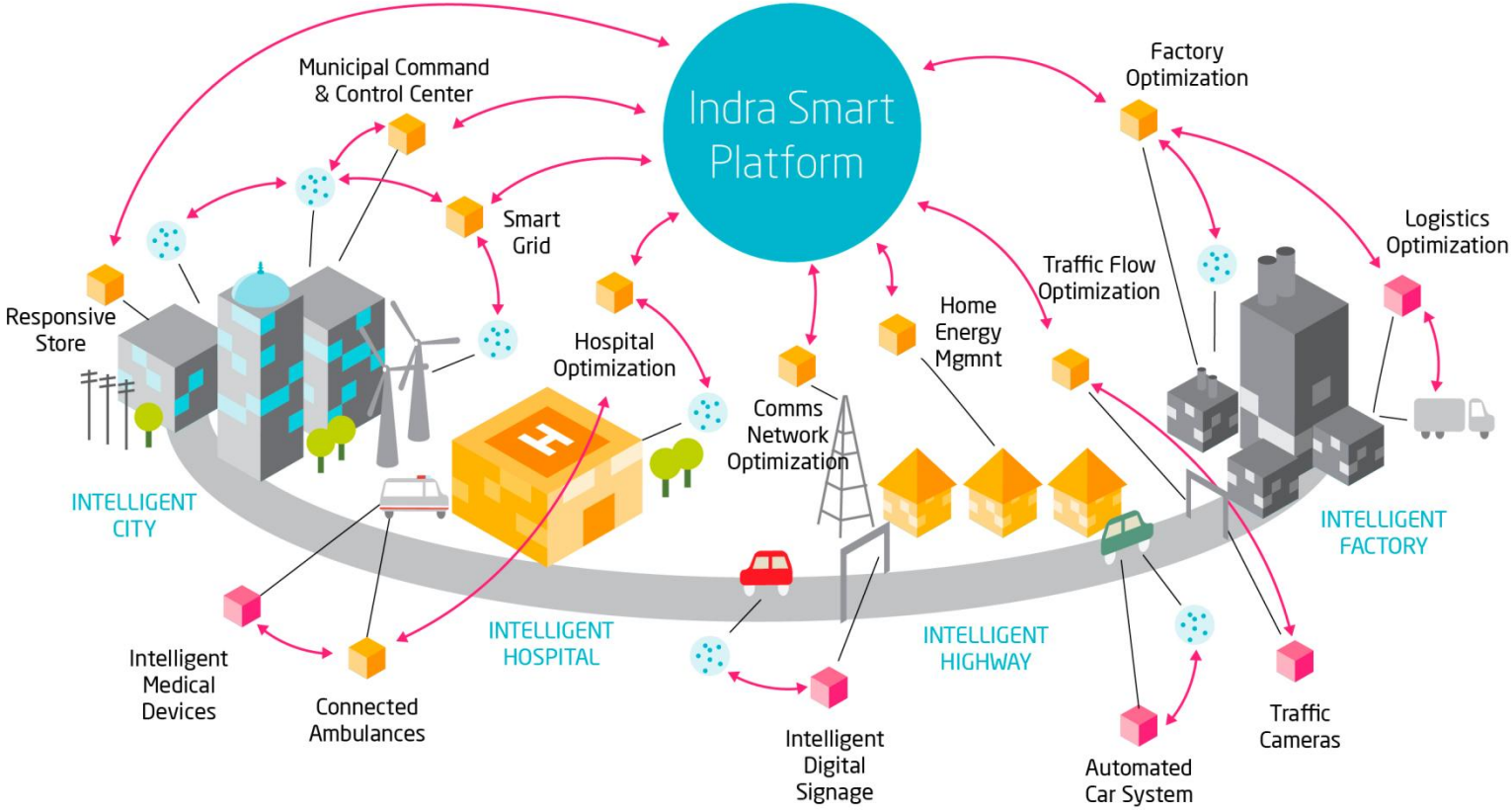
- Actuar con antelación y precisión
- Automatizar y personalizar respuestas
- Atención directa onmicanal

El Big Data está impactando profundamente en la evolución de los Sistemas de Ciberdefensa. Se está evolucionando desde las aproximaciones basadas en patrones y firmas a análisis en tiempo real del tráfico de red o logs de servidores entre otras fuentes de información

Ejemplo de Resultados

- Detección de tráfico entrante inusual que enmascara una exfiltración de datos
- Distribución atípica de ocurrencias de eventos: Por geografías, Por dimensión temporal, Por tipo de host
- Comportamientos anómalos:
  - Host que se conecta a varios servidores web
  - Consultas de páginas a intervalos regulares o siempre con el mismo tamaño
  - Etc.
- Comportamientos estables en el tiempo
  - Pueden implicar malware escondido si no están justificados
- Identificación de ataques de ingeniería social con identificación del patrón de ocurrencia y búsqueda de incidencias similares en los logs de servidores y tráfico de red
- Etc.

Comprobamos una demanda creciente de soluciones basadas en movilidad, cloud computing, social business, internet of things… que requieren de plataformas interoperables sobre las que desarrollar servicios SMART basados en conocimiento analítico



https://www.youtube.com/watch?v=tNIKZo12UrU

CIBERSEGURIDAD

NUEVOS RIESGOS, NUEVOS MODELOS DE PROTECCIÓN

HACIENDO DEL CIBERESPACIO UN LUGAR SEGURO

Muchas Gracias por su Atención y su Paciencia ……

indra