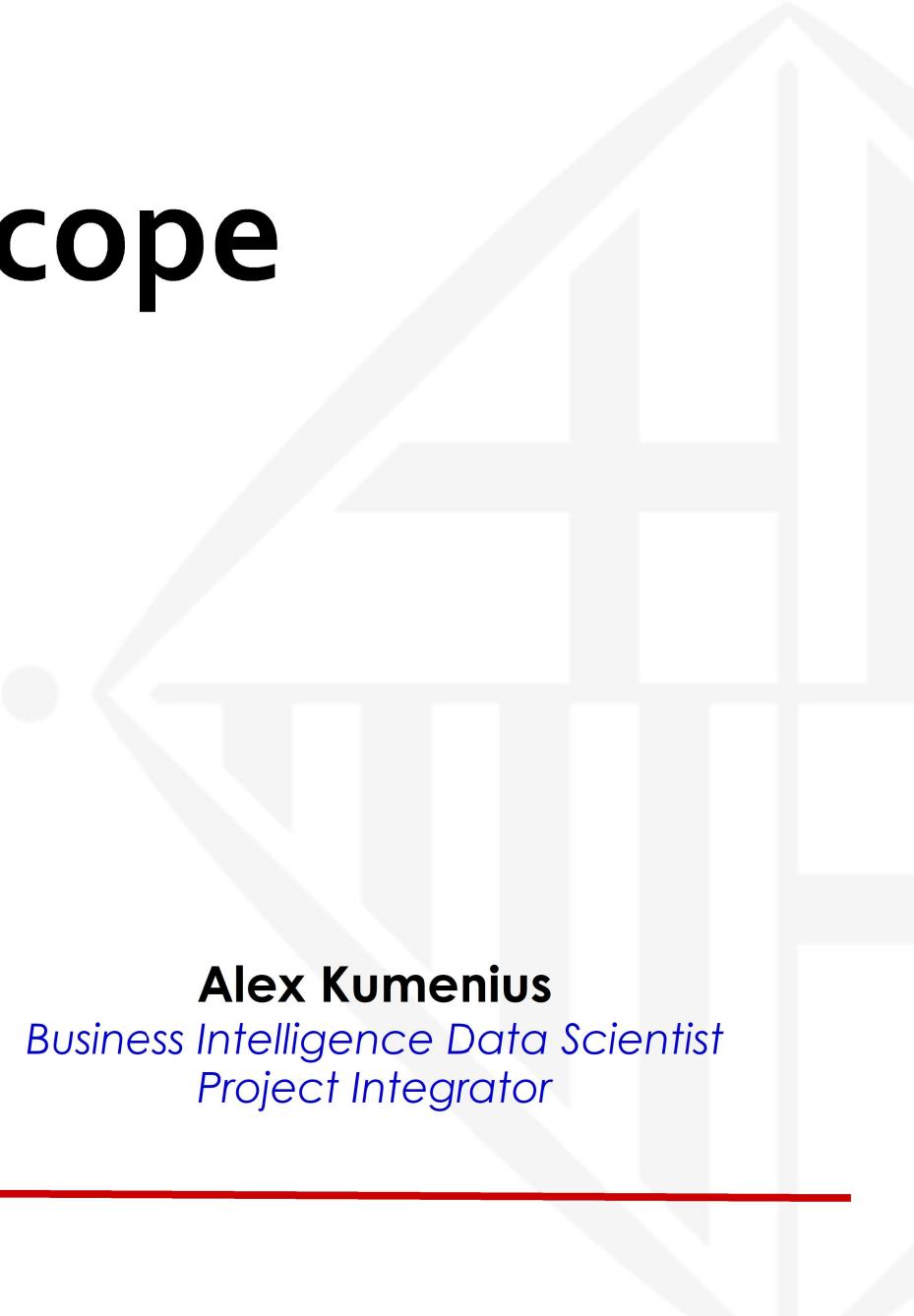




The BigData Scope

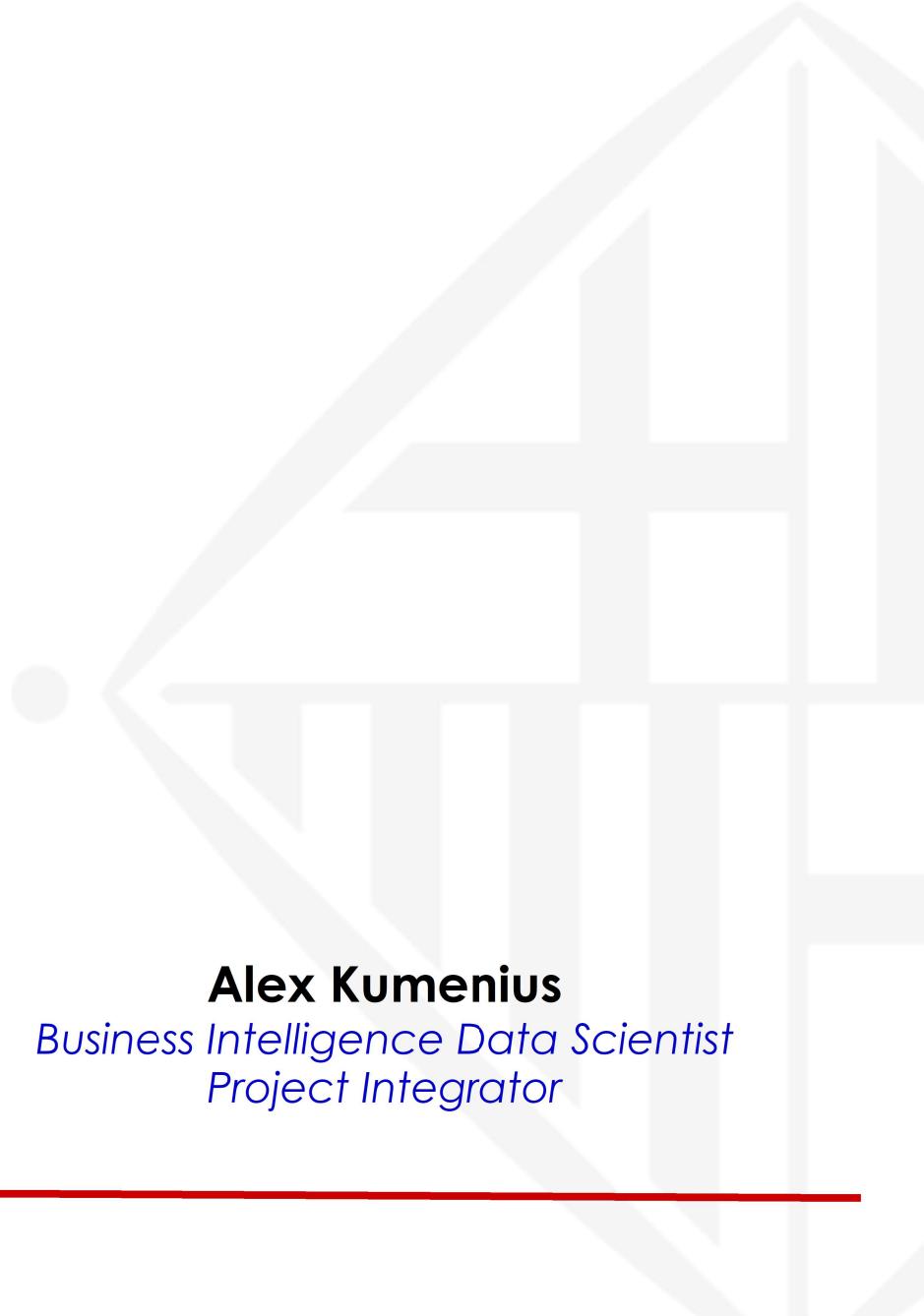


Alex Kumenius
*Business Intelligence Data Scientist
Project Integrator*

4 de Marzo del 2021

01

The BigData Scope



Alex Kumenius

*Business Intelligence Data Scientist
Project Integrator*

4 de Marzo del 2021



“Optimizar, digitalizar un proceso empresarial requiere más esfuerzo que simplemente hacer una compra de software”

¿ Que es Big Data ?

"BigData" conjuntos de datos ([datasets](#)) demasiado grandes para [procesar](#) o [almacenar](#) con herramientas tradicionales.

[Datasets](#) crecen exponencialmente y varían sustancialmente de una organización a otra.

Big Data

Requisitos básicos de BigData, son los mismos requisitos que cualquier conjunto de dato de diferentes tamaños.

Sin embargo,

Desafíos importantes al diseñar soluciones BigData :

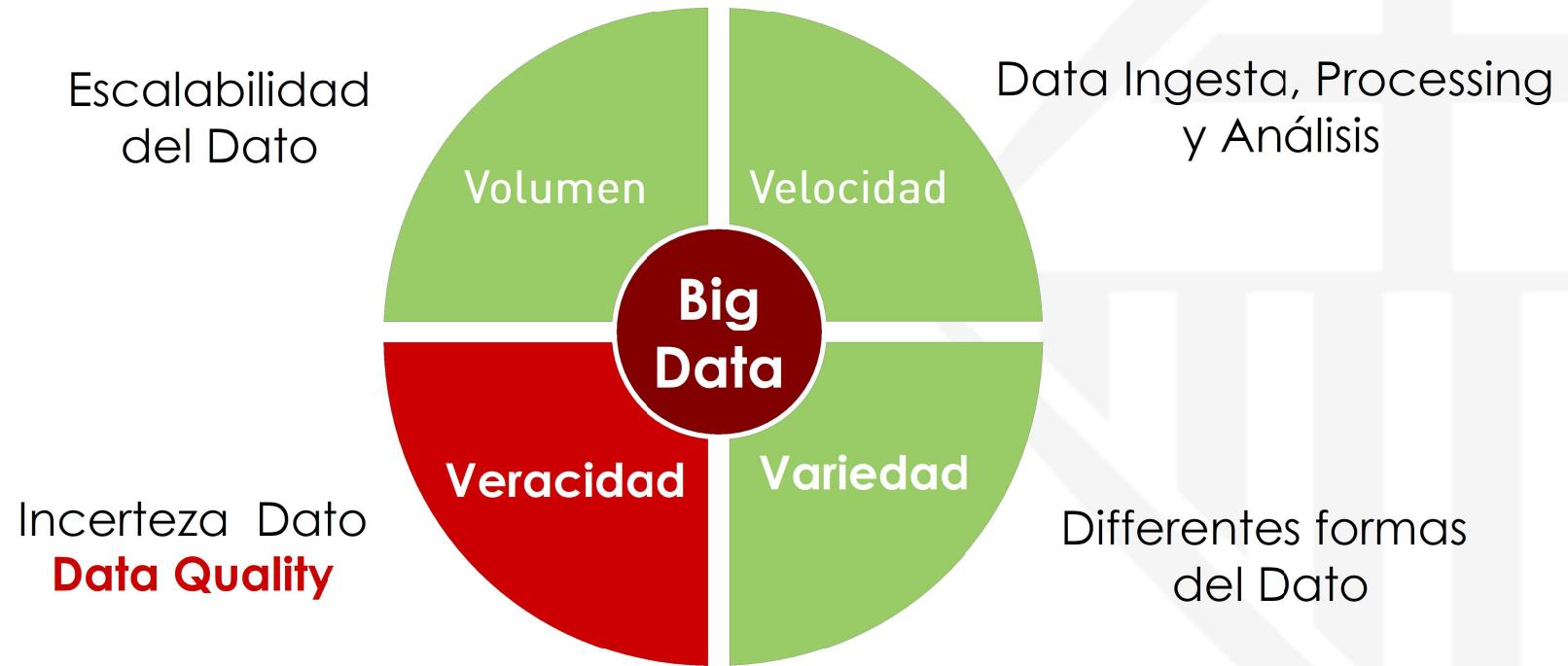
- El crecimiento masivo de datos,
- la velocidad de ingesta y procesamiento,
- y las características de los datos que deben tratarse en cada etapa del proceso.

Objetivo de los sistemas de BigData es obtener información de grandes volúmenes de **datos HETEROGÉNEOS**.

Imposibles con métodos convencionales.

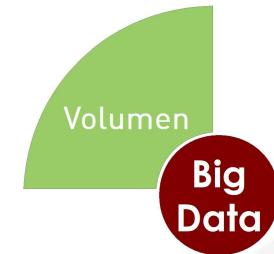
Tres Vs de Big Data

En 2001 Gartner presentó las "**Tres V de BigData**". Describe características que diferencian BigData de otros procesos de datos:



Three Vs of big data

Volumen



La gran cantidad de información procesada, define la arquitectura de los sistemas de BigData.

Datasets tienen órdenes de magnitud más grandes que los conjuntos de datos tradicionales...

... lo que exige una mayor reflexión en cada etapa del ciclo del procesamiento y almacenamiento.



Ajuntament
de Barcelona

NATIONAL
GEOGRAPHIC
MAGAZINE





JAN
2018

DIGITAL AROUND THE WORLD IN 2018

KEY STATISTICAL INDICATORS FOR THE WORLD'S INTERNET, MOBILE, AND SOCIAL MEDIA USERS

TOTAL
POPULATION



we
are
social

7.593
BILLION

URBANISATION:
55%

INTERNET
USERS



4.021
BILLION

PENETRATION:
53%

ACTIVE SOCIAL
MEDIA USERS



we
are
social

3.196
BILLION

PENETRATION:
42%

UNIQUE
MOBILE USERS



5.135
BILLION

PENETRATION:
68%

ACTIVE MOBILE
SOCIAL USERS



PENETRATION:
39%

SOURCES: POPULATION: UNITED NATIONS; U.S. CENSUS BUREAU; INTERNET: INTERNETWORLDSTATS; ITU; EUROSTAT; INTERNETLIVESTATS; CIA WORLD FACTBOOK; MIDEASTMEDIA.ORG; FACEBOOK; GOVERNMENT OFFICIALS; REGULATORY AUTHORITIES; REPUTABLE MEDIA; SOCIAL MEDIA AND MOBILE SOCIAL MEDIA: FACEBOOK; TENCENT; VKONTAKTE; KAKAO; NAVER; DING; TECHRASA; SIMILARWEB; KEPIOS ANALYSIS; MOBILE: GSMA INTELLIGENCE; GOOGLE; ERICSSON; KEPIOS ANALYSIS. NOTE: PENETRATION FIGURES ARE FOR TOTAL POPULATION (ALL AGES).

 **Hootsuite™**  **we
are.
social**



JAN
2018

SPAIN

A SNAPSHOT OF THE COUNTRY'S KEY DIGITAL STATISTICAL INDICATORS



TOTAL
POPULATION



46.38
MILLION

URBANISATION:
80%

INTERNET
USERS



39.42
MILLION

PENETRATION:
85%

ACTIVE SOCIAL
MEDIA USERS



27.00
MILLION

PENETRATION:
58%

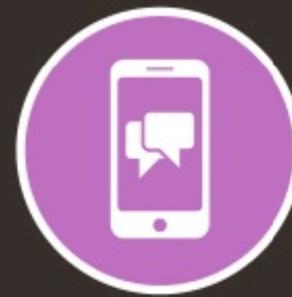
MOBILE
SUBSCRIPTIONS



55.51
MILLION

PENETRATION:
120%

ACTIVE MOBILE
SOCIAL USERS



23.00
MILLION

PENETRATION:
50%

SOURCES: POPULATION: UNITED NATIONS; U.S. CENSUS BUREAU; INTERNET: INTERNETWORLDSTATS; ITU; EUROSTAT; INTERNETLIVESTATS; CIA WORLD FACTBOOK; MIDEASTMEDIA.ORG; FACEBOOK; GOVERNMENT OFFICIALS; REGULATORY AUTHORITIES; REPUTABLE MEDIA; SOCIAL MEDIA AND MOBILE SOCIAL MEDIA: FACEBOOK; TENCENT; VKONTAKTE; KAKAO; NAVER; DING; TECHRASA; SIMILARWEB; KEPIOS ANALYSIS; MOBILE: GSMA INTELLIGENCE; GOOGLE; ERICSSON; KEPIOS ANALYSIS. NOTE: PENETRATION FIGURES ARE FOR TOTAL POPULATION (ALL AGES).

Hootsuite™ **we
are.
social.**

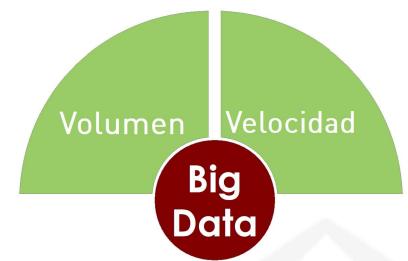


DATA SOURCES – FUENTES DE DATOS



Three Vs of big data

Velocidad



La **Velocidad** con que la información se mueve a través del sistema.

Los datos fluyen al sistema desde múltiples fuentes y se espera que se procesen en tiempo real para obtener información.

Los datos se agregan, transforman, procesan y se analizan constantemente para mantenerse actualizado con la entrada de nueva información en tiempo real.

En este escenario requeriremos de sistemas robustos con componentes de alta disponibilidad para protegerse contra fallos a lo largo del data stream o flujo de datos.

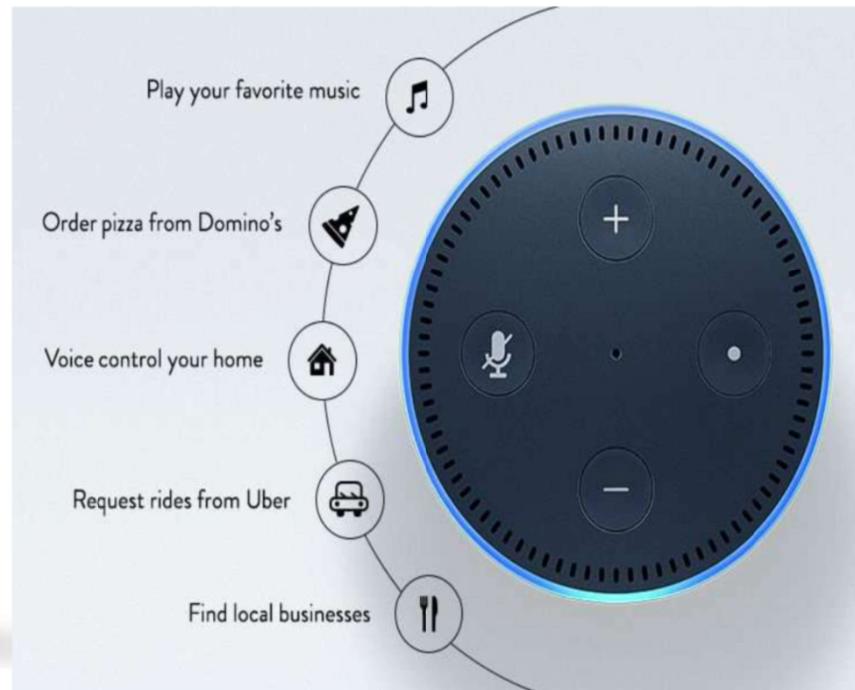
Voice / Speech Recognition

Sample voice commands:

Alexa, turn down the lights. Connect the Echo to any Phillips Hue Smart Bulbs and she can control them through voice commands the same way you can via the Phillips Hue app.

Alexa, turn on my humidifier. Alexa can connect to any WeMo Switch enabled outlet or Wink Home Hub.

Alexa, what's traffic like right now? Alexa can tell you the roads to avoid on your morning commute.



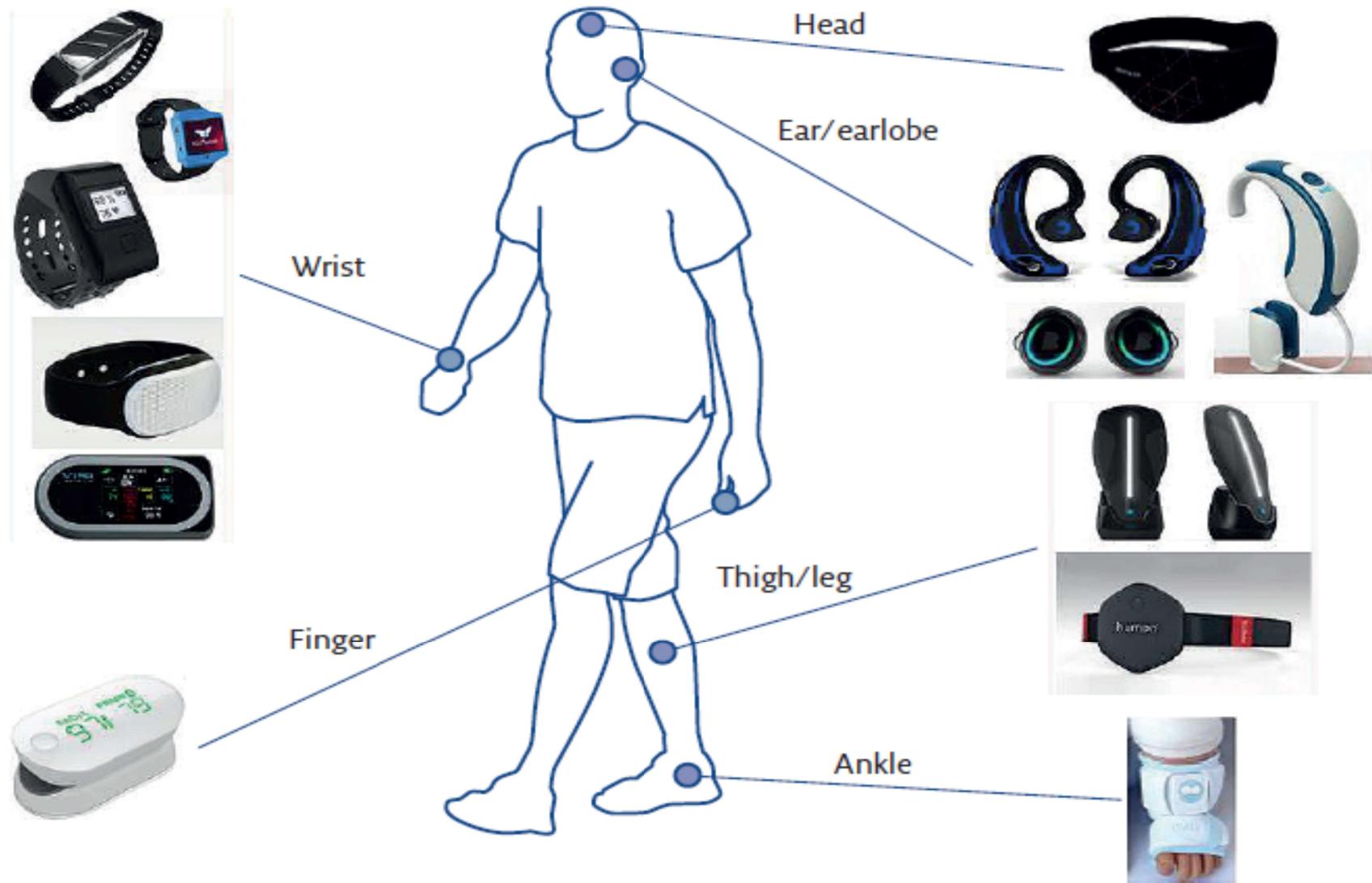


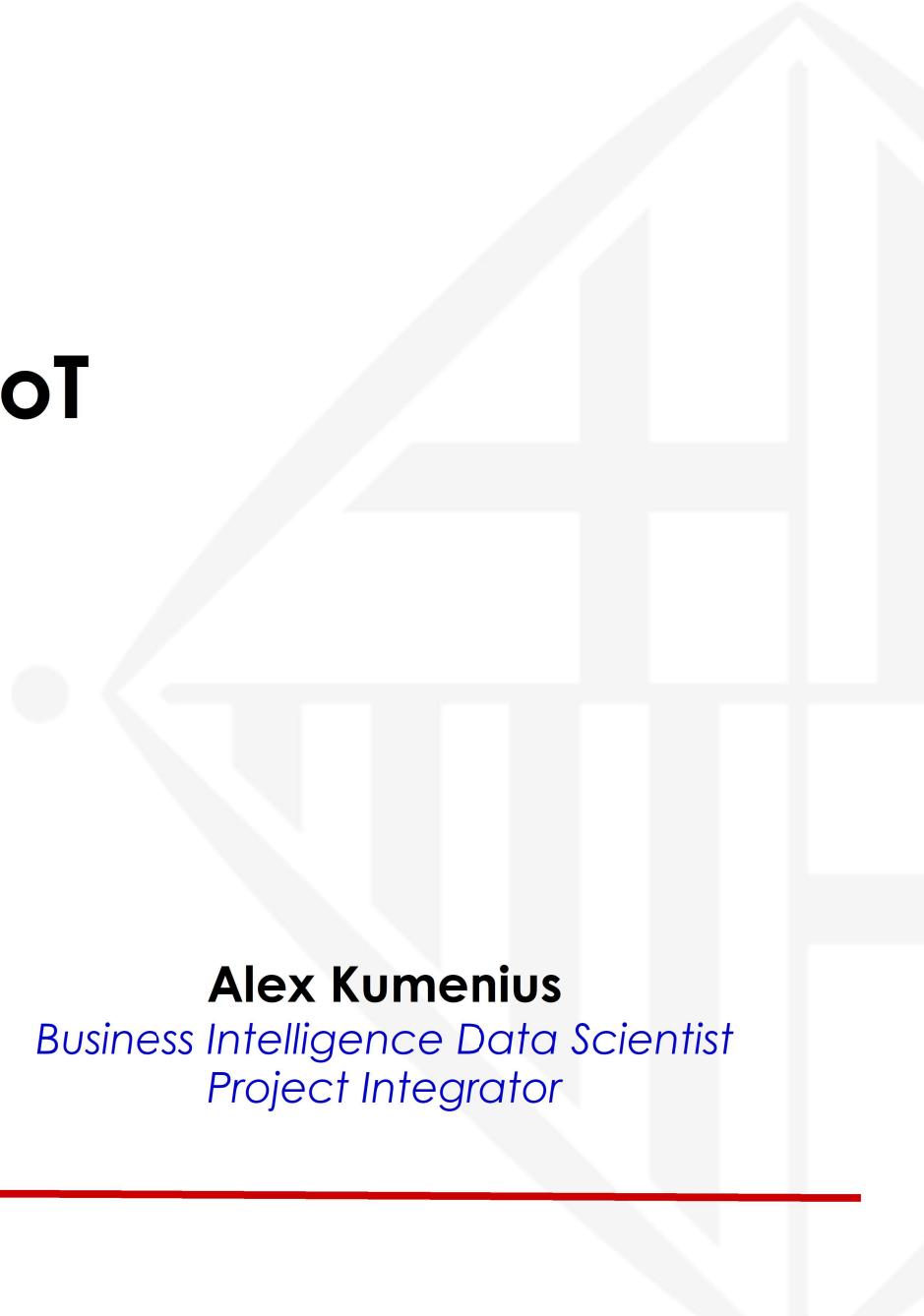
Figure 3 Location on the body of several commercially available wearable devices able to perform pulse oximetry.



Figure 4 Location on the body of several commercially available wearable devices able to estimate ventilation from chest surface motion.

02

Internet de las Cosas – IoT



Alex Kumenius

*Business Intelligence Data Scientist
Project Integrator*

4 de Marzo del 2021



Internet de las cosas, o IoT, Sistema de dispositivos informáticos interconectados.

- Máquinas mecánicas y digitales, objetos, animales o personas que cuentan con identificadores únicos (UID) y
- la capacidad de transferir datos a través de una red **sin** requerir de la interacción entre humanos (human-to-human) o de humano a computadora (human-to-computer).



El objetivo de “Internet de las Cosas” es permitir que los objetos se conecten :

- en cualquier momento,
- en cualquier lugar,
- con cualquier objeto y
- a cualquier persona

a través de la Red y a cualquier servicio.



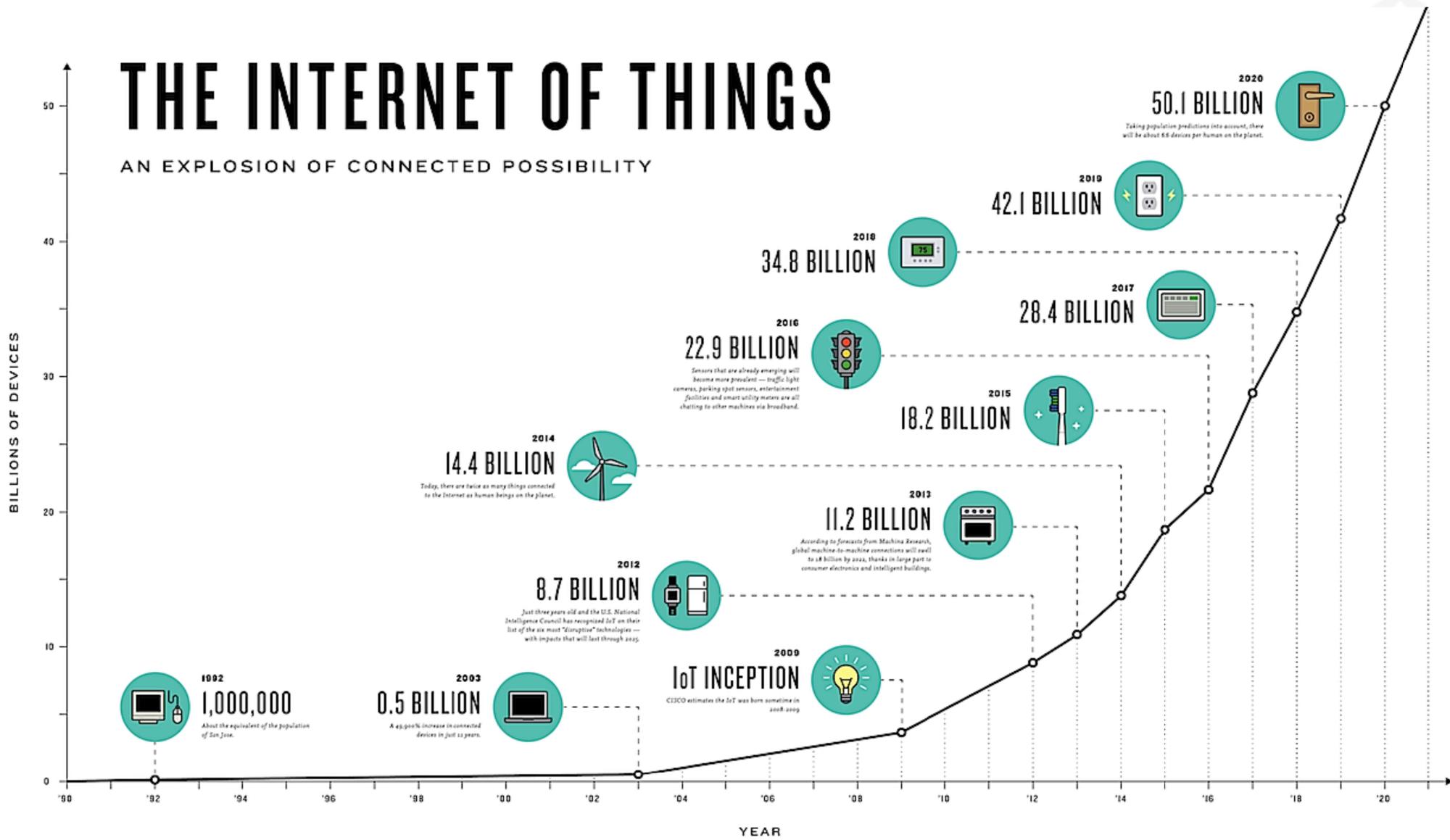
Convergencia IP para comunicar y controlar los dispositivos.





THE INTERNET OF THINGS

AN EXPLOSION OF CONNECTED POSSIBILITY





Internet de las Cosas todavía está madurando, debido a factores que limitan la explotación total de la IoT.

Los factores más relevantes :

- Falta enfoque concreto en la utilización de identificadores únicos (UID).
- Lento desarrollo de arquitecturas de referencia de **IoT**.
- Lento avance en la interoperabilidad semántica: intercambio de información del sensor en entornos heterogéneos.



Los factores más relevantes :

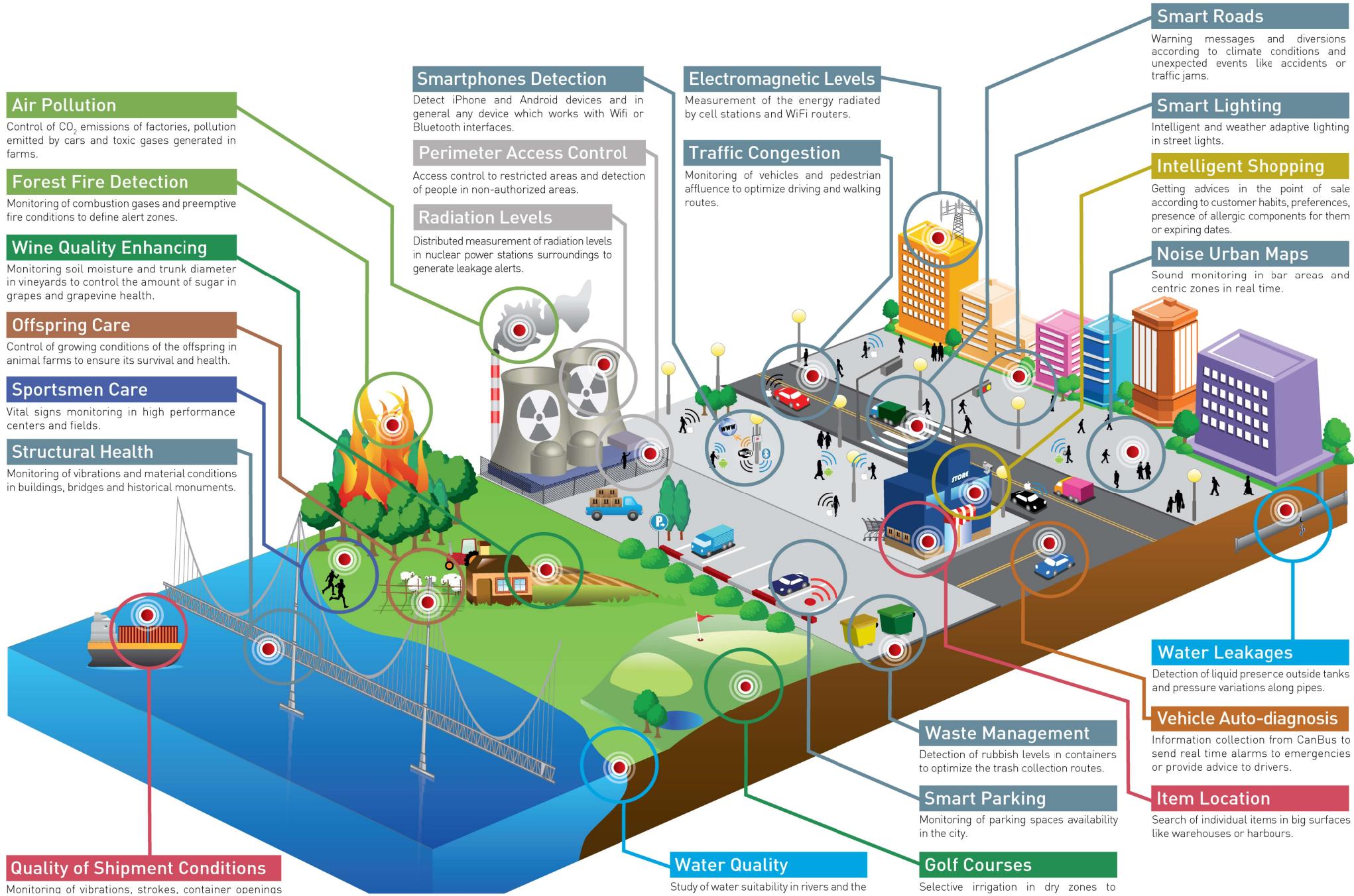
- **Dificultades** desarrollo Marco Legal que **permite proteger** :
 - la **innovación**,
 - la **confianza**,
 - la **propiedad** de los **datos** en el IoT,
 - la **seguridad** y
 - la **privacidad** en un entorno complejo.
- Faltan entornos de prueba a gran escala.
 - experimentar con **redes** de sensores **complejas** y
 - estimular la **innovación** a través de la **reflexión** y la **experiencia**.
- tarifas de roaming para aplicaciones de sensores de gran alcance geográfico
- y la falta de **conectividad de red instantánea** y **confiable**.



El desarrollo de Internet de las cosas implica que los entornos :

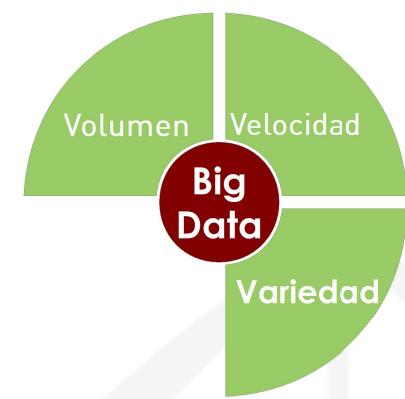
- Las ciudades,
- los edificios,
- los vehículos,
- la ropa,
- los dispositivos portátiles y
- otros objetos.

.. tienen cada vez más información asociada a ellos y/o la capacidad de detectar, comunicarse, conectarse en red y producir nueva información.



Three Vs of big data

Variedad



Los problemas de Big Data son únicos debido a la gran variedad de data sources (fuentes datos) que se procesan y su calidad relativa.

Formatos y tipos de medios pueden variar significativamente. Datos No Estructurados junto con Datos Estructurados.

- Los sistemas tradicionales almacenan datos en estructuras formateadas y organizadas.
- Los sistemas de BigData almacenan datos en estado crudo (RAW).

Tipos de Estructuras de Datos

- Structured Data (Dato Estructurado) Estructura tradicional de almacenamiento de datos :

- estructurados en tablas (*filas y columnas*).
- Bases se Datos Relacionales – RDS
(SQL, MySQL, PostgreSQL, MariaDB, etc..)

- Unstructured Data (Dato NO-Estructurado) Estructuras de almacenamiento de datos que no encajan perfectamente en la estructura tradicional de RDS.

- Bases de Datos NoSQL (Not Only SQL)
(MongoDB, Couchbase, etc).
- Ejemplos de Unstructured Data incluyen :
 - emails,
 - videos,
 - archivos de audio,
 - páginas web, y
 - mensajes de redes sociales.

Structured vs Unstructured Data



Structured Data
tables

Tendency:
20% - Structured y Semi-
Structured,
80% - Unstructured



Unstructured
Data

Text
Images
Video

Semi-Structured
Data

XML
JSON

Amazon Rekognition

Deep learning-based image recognition service
Search, verify, and organize millions of images



Object and Scene
Detection



Facial
Analysis



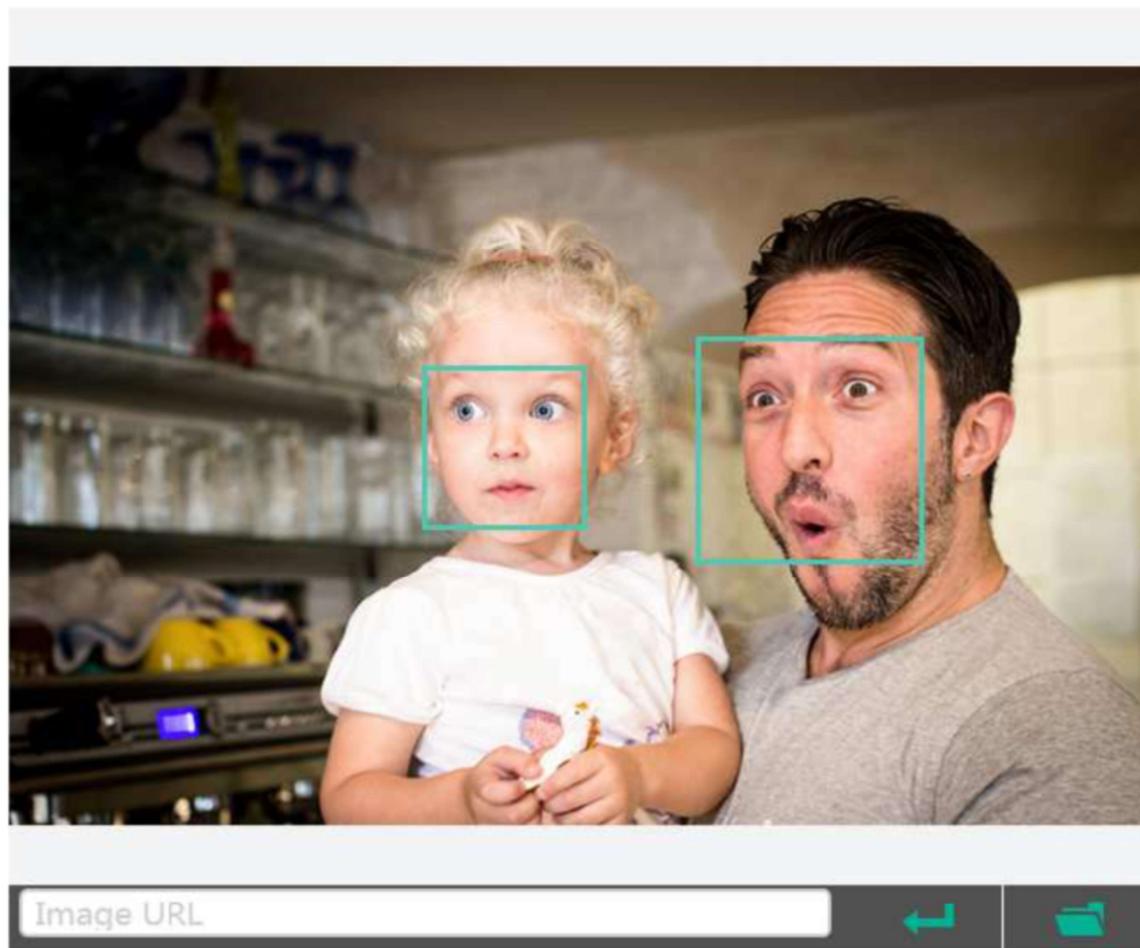
Face
Comparison



Facial
Recognition

Aplicaciones - Unstructured Data

Examples: Amazon Emotions API



Detection Result:

2 faces detected

JSON:

```
[  
  {  
    "faceRectangle": {  
      "left": 479,  
      "top": 189,  
      "width": 160,  
      "height": 160  
    },  
    "scores": {  
      "anger": 0.0000102376043,  
      "contempt": 0.0000217592042,  
      "disgust": 0.000005393408,  
      "fear": 0.00126100751,  
      "happiness": 0.0000120425466,  
      "neutral": 0.0015419157,  
      "sadness": 0.00000278396146,  
      "surprise": 0.9971449  
    }  
  },  
  {  
    "faceRectangle": {  
      "left": 264,  
      "top": 469,  
      "width": 160,  
      "height": 160  
    },  
    "scores": {  
      "anger": 0.0000102376043,  
      "contempt": 0.0000217592042,  
      "disgust": 0.000005393408,  
      "fear": 0.00126100751,  
      "happiness": 0.0000120425466,  
      "neutral": 0.0015419157,  
      "sadness": 0.00000278396146,  
      "surprise": 0.9971449  
    }  
  }]
```



Facial Analytics

Demographic Data

Facial Landmarks

Image Quality

Sentiment Expressed

General Attributes





Facial Analytics

Demographic Data

Male

97.4 %

Facial Landmarks

Happy

96.0 %

Smiling

96.3 %

Image Quality

Beard

65.3 %

Sentiment Expressed

Eyes are open

100 %

Mustache

100 %

General Attributes

Facial Analytics

Facial analysis

Get a complete analysis of facial attributes, including confidence scores. (Your images aren't stored.)



Choose a sample Image



Use your own image

 Upload

or

Type or paste image URL

Go

Done with the demo?

[Download SDKs](#)

▼ Faces | Confidence



looks like a face 99.8%

appears to be female 100%

smiling 99.4%

appears to be happy 93.2%

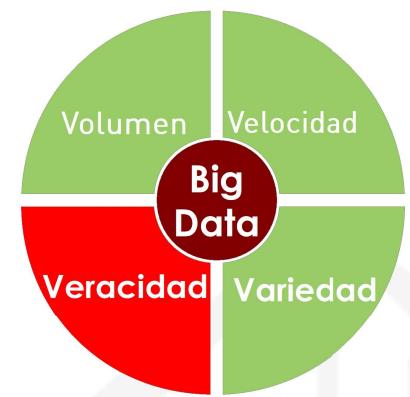
wearing eyeglasses 99.9%

wearing sunglasses 97.6%

[Show more](#)

Three Vs of big data

Veracidad

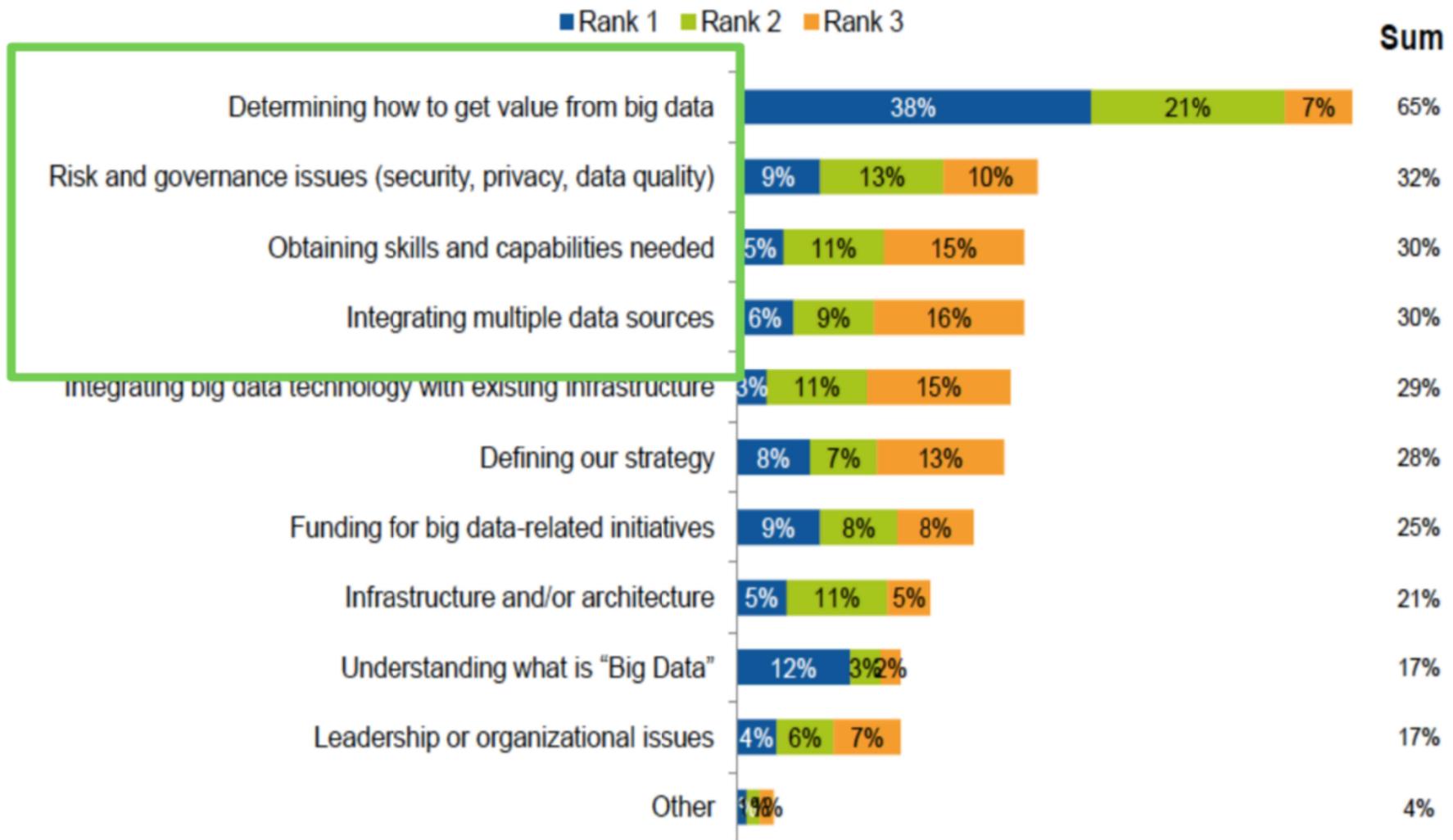


La **Veracidad** de data sources (fuentes de datos) y la complejidad del procesamiento pueden llevar a **comprometer la calidad de los datos** y, en consecuencia, la calidad del análisis resultante.

Valor: El desafío final de BigData es entregar **valor**. A veces, la **extracción de valor real** puede **resultar difícil**.

What are your organization's top 3 hurdles or challenges with big data?

¿Cuáles son los 3 principales obstáculos o desafíos de su organización con big data?



Source: Gartner

Capacidad de Almacenamiento

Cálculo aproximado de los **recursos** requeridos para almacenar nuestro **DATASETS**.

Es útil saber algunas cosas sobre tu sistema.

- ¿Cuánta memoria está disponible?
- ¿Qué otras aplicaciones están en uso?
- ¿Hay otros usuarios registrados en el mismo sistema?
- ¿Qué sistema operativo?
- ¿Es el OS 32 o 64 bit?

Con este análisis, podremos establecer nuestras necesidades actuales y futuras de almacenamiento de datos. Y por tanto decidir que infraestructura de hardware necesitaremos.

03

Business Intelligence, Data Scientist and Data Wrangling



Alex Kumenius

*Business Intelligence Data Scientist
Project Integrator*

4 de Marzo del 2021

business intelligence



Barcelona
Activa

¿ Que es Inteligencia Empresarial (BI) ?

La inteligencia empresarial (BI) es un término general que incluye :

- las aplicaciones,
- la infraestructura,
- las herramientas,
- y prácticas

Acceden y analizan la información para mejorar y optimizar las decisiones y el rendimiento empresarial.



Funciones Analista Inteligencia Empresarial (BI)

Su objetivo :

- Interpretación de datos masivos - BigData.
- Identificar nuevas oportunidades e implementa una estrategia efectiva basada en los resultados analizados.
- Proporcionar una ventaja competitiva de mercado.

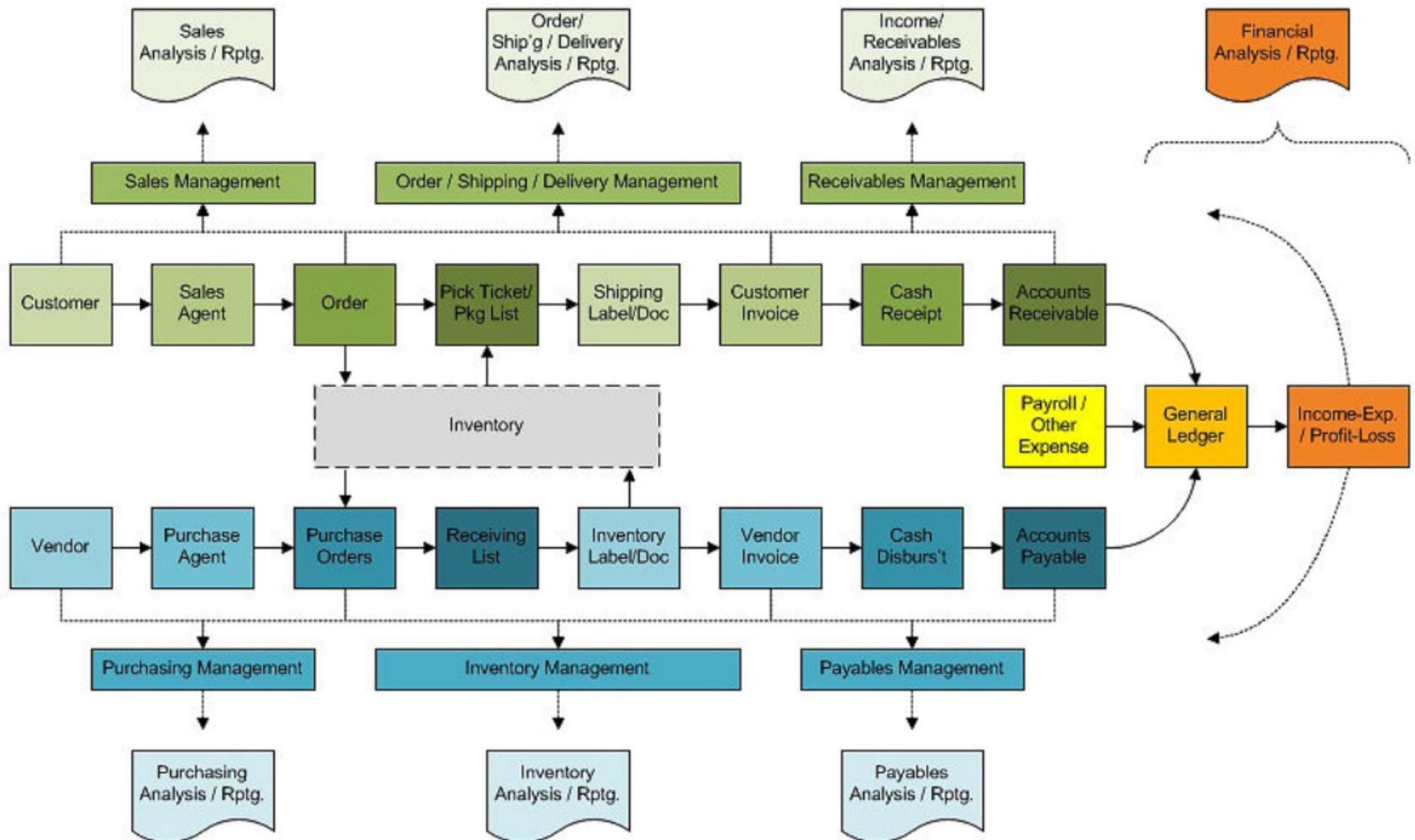
Los analistas de BI tienen excelentes habilidades de comunicación, lo que les permite tener interacciones significativas con todos los miembros de la jerarquía de una organización.

Aprovechan su conocimiento de la industria para recomendar ideas que pueden aplicarse de inmediato.

Combinar, external data (*datos sectoriales del mercado*) e internal data (*financiera, contable, operaciones, etc..*) proporciona una nueva "**inteligencia singular**" específica de la empresa, que no se puede derivar de ningún otro conjunto de datos (datasets).

Inteligencia Empresarial

Conocimiento del Negocio – Work Flow





Swot Analysis



Principales Funciones Inteligencia Empresarial (BI)

Reporting

Analíticas

Minería de datos

Procesos Minería

Procesamiento complejo de eventos

Evaluación comparativa

Extracción de textos

Analysis Predictivo

Analysis Prescriptivo

Procesamiento analítico Online

Gestión del rendimiento empresarial

La **Inteligencia Empresarial** capacita a las organizaciones para obtener información sobre :

- nuevos mercados,
- evaluar la demanda de un producto
- la idoneidad de los productos y servicios para diferentes segmentos del mercado
- evaluar el impacto de los esfuerzos de marketing.

Data Scientist

Un científico del dato es alguien “que es mejor en estadística que cualquier programador, y mejor programador que cualquier estadista”

(Josh Wills built and led data science teams at Cloudera and Google).

Data Scientist

Un Data Scientist es un profesional que ante enormes bases de datos, aplica sobre ellas sus conocimientos:

- Programación,
- Matemáticas
- y Estadística

para recopilar, extraer y procesar la información relevante que contienen las diferentes fuentes de datos.

EL DATO CRITICO

Prescriptive Analytics

La **Analítica Prescriptiva (PI)** se extiende más allá de la **analítica predictiva** al especificar tanto las **acciones** necesarias para lograr los resultados previstos como los efectos inter-relacionados de cada decisión.

