

SAP Big Data

El Viaje Instantáneo en el Universo de Información.

Gonzalo Málaga
Digital Enterprise Platform Group

Octubre, 2016



INGENIERÍA
Posgrados

© 2015 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. Please see <http://global12.sap.com/corporate-en/legal/copyright/index.epx> for additional trademark information and notices.

Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors.

National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP SE or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP SE or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

Saber que pasará?

Regresar en el tiempo?

Universos Paralelos?

Viajar instantáneamente?



“Estos motores generan 1Tb de datos por avión por día”

Jeff Immelt – CEO GE



100 millones de usuarios activos
2 mil millones de páginas vistas por día
250 millones de “queries” de búsqueda
20 petabytes

eBay

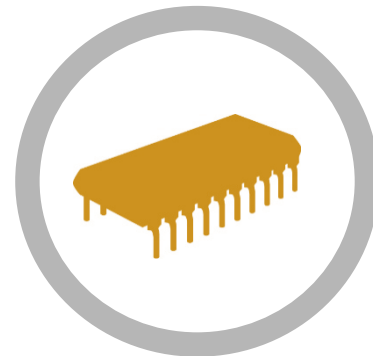
Tendencias que Redefinen la Economía



Hyperconectividad



**Mundo
“Smart”**



**Super
computación**

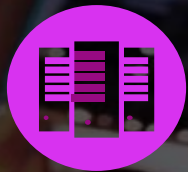


**Cyber
seguridad**



Cloud

Cifras de la Economía Digital



Big Data

90%

De los datos en el mundo han sido generados en los últimos 2 años ¹



Redes Empresariales

40%

de crecimiento en la adopción de redes de negocios y colaboración ²



Cloud

Ya al 2014

51%

De las cargas de procesamiento estaban en cloud ⁵



Mobile

9 mil millones

De Usuarios móviles en el mundo al 2020 ⁴

Internet of Things



212 mil millones

Dispositivos conectados a IoT hacia el 2020 ³

¹ ScienceDaily, May 22, 2013.

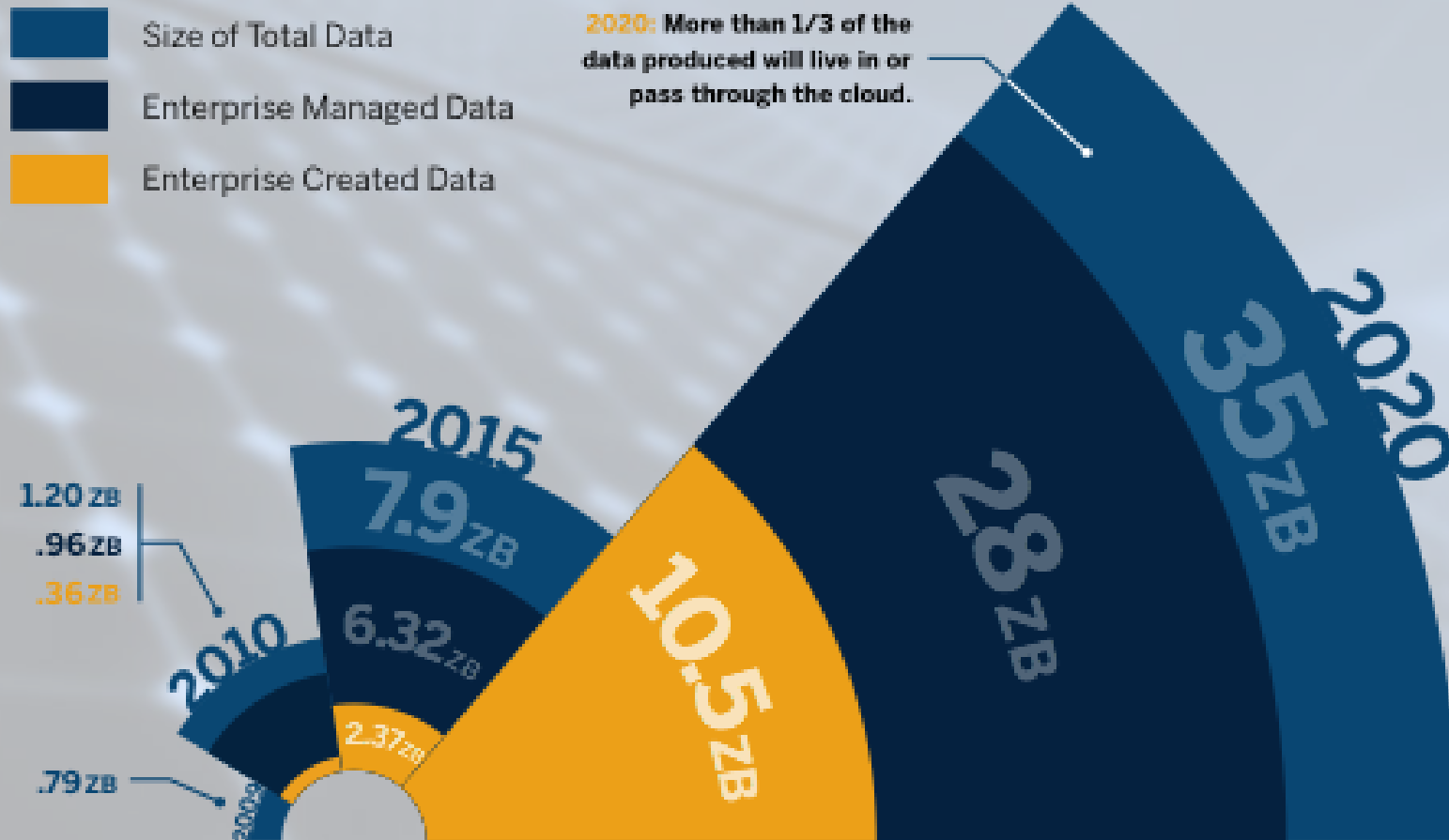
² Technology Adoption Report on Business Networks, Ardent Partners, 2014.

³ Internet of Things (IoT) 2013 to 2020 Market Forecast: Billions of Things, Trillions of Dollars, IDC, 2013.

⁴ Statista, 2014.

⁵ Cisco Global Cloud Index: Forecast and Methodology for 2013–2018, Cisco Systems Inc., November 11, 2014.

Crecimiento de BIG DATA



WHAT IS A ZETTABYTE? 1 ZETTABYTE = 1,000,000,000,000 GIGABYTES

Se Requiere una Mayor Agilidad

SPEED OF BUSINESS
HAS RADICALLY
INCREASED



DATA IS GETTING BIG

MORE DATA

LESS TIME

El **Exito** de BIG DATA Requiere **Cobertura Completa**





SAP HANA



Instant Results



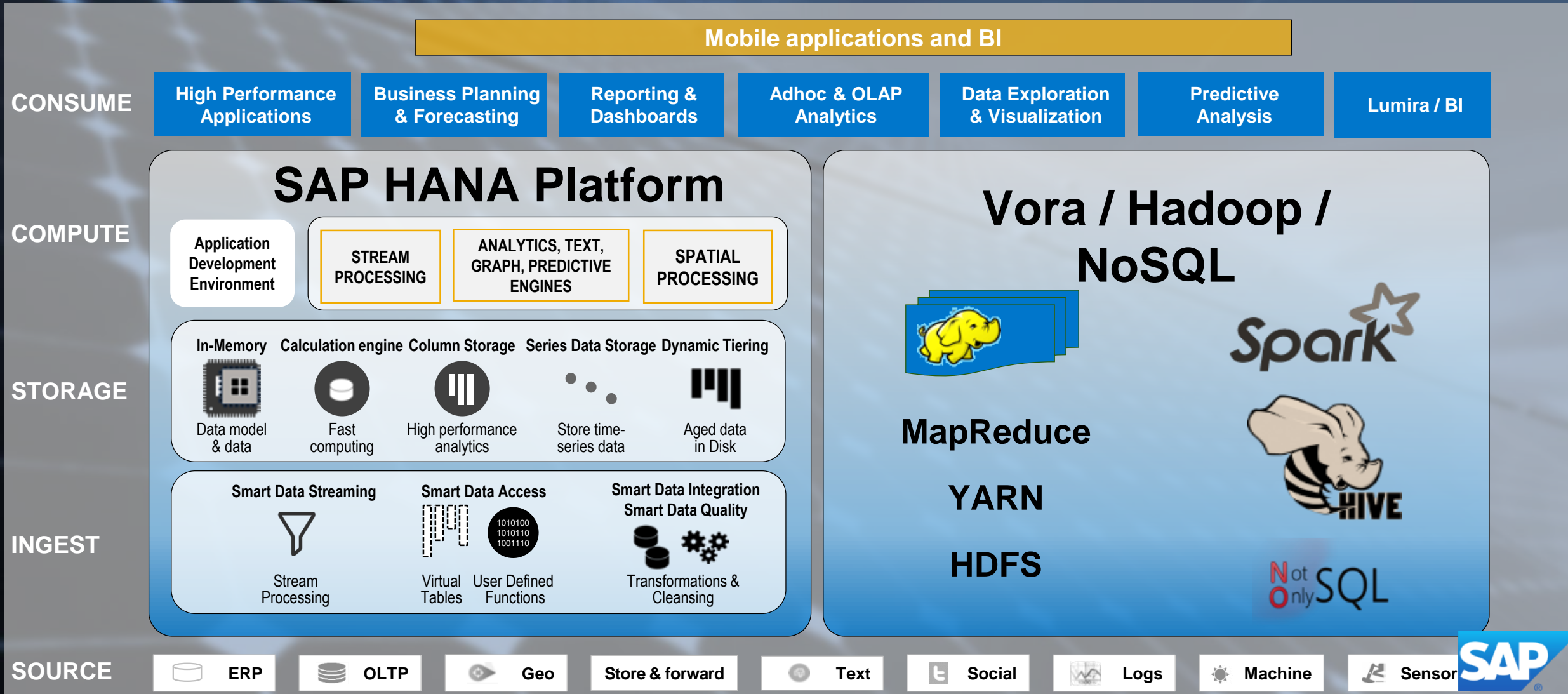
HADOOP



Infinite storage

Plataforma de Big Data que acelera la forma como usted adquiere, analiza y actúa sobre nuevas percepciones.

El enfoque de SAP: Cadena de valor de Big Data de punta a punta

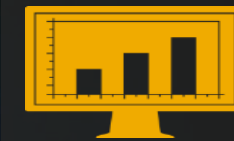


Optimizado para las Personas

Conectando a las personas con las percepciones correctas



Enterprise BI



Agile Visualization



Advanced Analytics



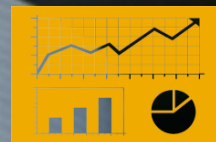
Personalized Insights



Sensing and Responding



Predictive Modeling



Real-time Reporting and Analysis

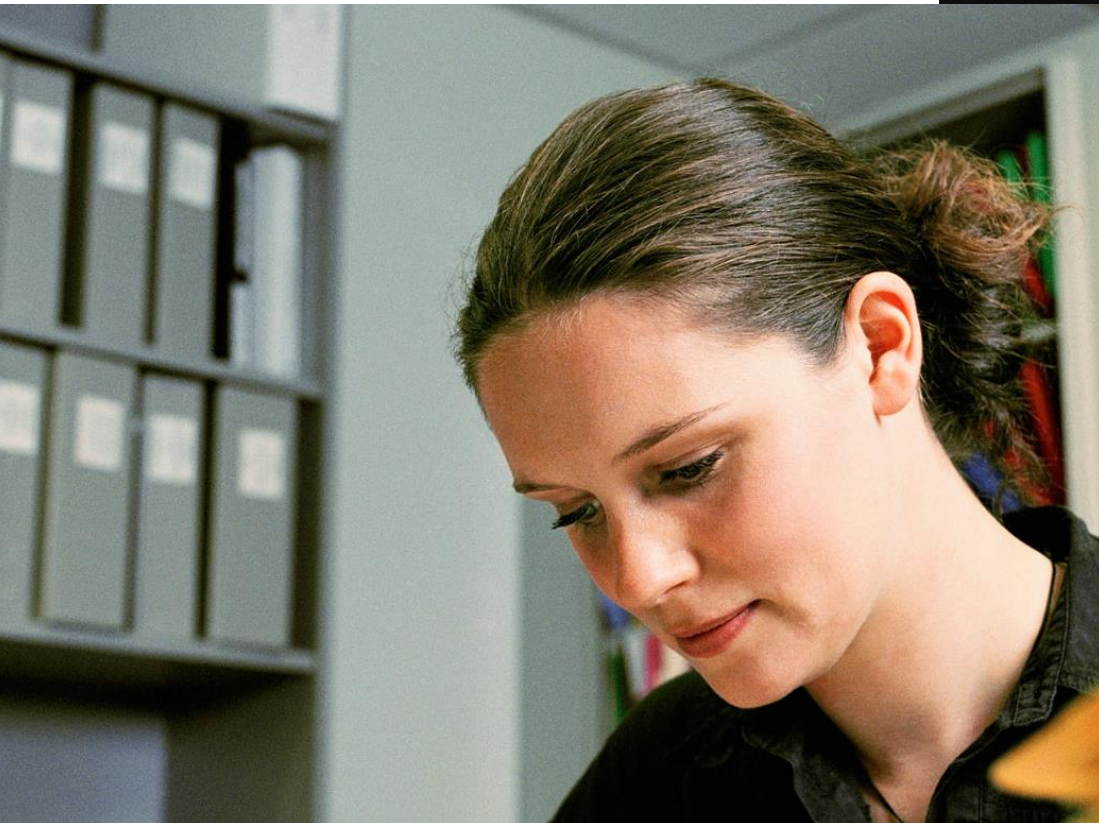


Advanced Planning and Forecasting

ebay

Anticipa el comportamiento del consumidor

Sistema de detección de señales tempranas con base en análisis predictivo



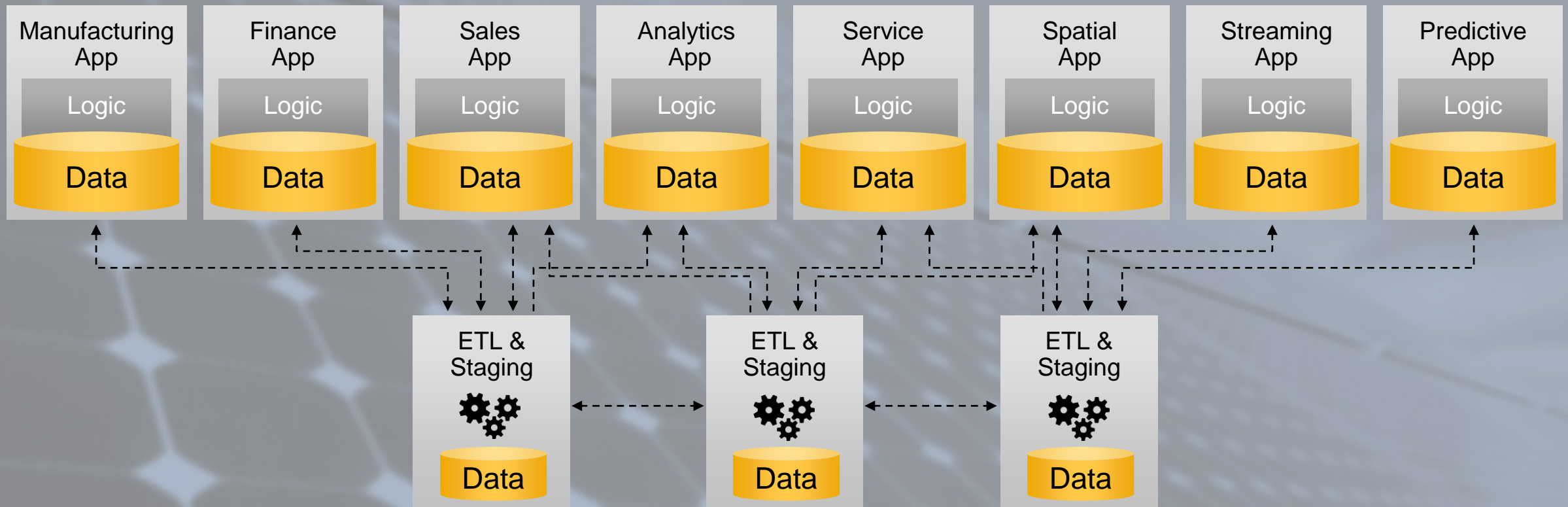
Detección automatizada de señales sobre 100 Pb con un 97% de confianza.

Permite responder proactivamente a la dinámica del mercado en tiempo real



El problema de las arquitecturas típicas de TI

Silos, demoras, y complejidad limitan la agilidad e innovación del negocio



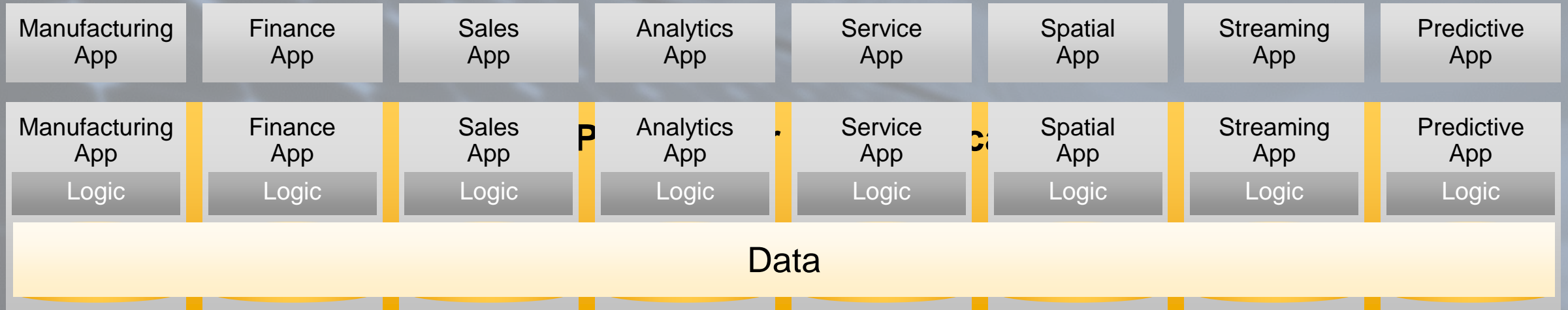
Silos de Aplicaciones
Múltiples Copias de Datos
Procesamiento Batch



Vista parcial del Negocio
Sin Visibilidad en Tiempo Real
Capacidad Limitada para Innovar

La solución: Hacer que todos los datos estén disponibles inmediatamente a todas las aplicaciones

Reducir movimiento y latencia de datos – mejorar agilidad e innovacion del negocio



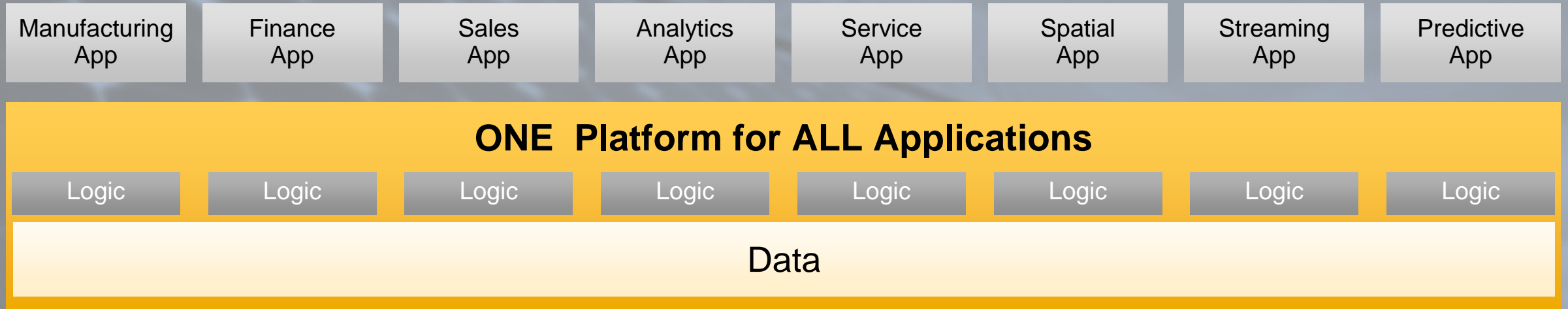
Cargas de aplicaciones unificadas
 Datos Unificados – Unica Copia
 Procesamiento en Tiempo Real



Vista Completa del Negocio
Capacidad para reaccionar en tiempo real
Capacidad para innovar

Una solución solo posible con “in-memory data management”

Solo la gestión de información en memoria habilita el tiempo real a todas las aplicaciones



Sin espera para el acceso a los datos y el procesamiento



Toda la lógica de aplicaciones (OLTP & OLAP) procesada en único sistema



Todos los tipos de datos procesados en un único sistema



Velocidad
Simpleza
Innovación

SAP HANA Platform

La plataforma de gestión de información para todas las aplicaciones

Todos los dispositivos

SAP, ISV and Aplicaciones desarrolladas

SAP HANA PLATFORM

APPLICATION SERVICES

- Web Server
- JavaScript
- Fiori UX
- Graphic Modeler
- Application Lifecycle Management

PROCESSING SERVICES

- Spatial
- Graph
- Predictive
- Search
- Text Analytics
- Streaming Analytics
- Series Data
- Business Functions

INTEGRATION & QUALITY SERVICES

- Data Virtualization
- ELT & Replication
- Data Quality
- Hadoop & Spark Integration
- Remote Data Sync

DATABASE SERVICES

- Columnar OLTP+OLAP
- Multi-Core & Parallelization
- Advanced Compression
- Multi-tenancy
- Multi-Tier Storage
- Data Modeling
- Openness
- Admin & Security
- High Availability & Disaster Recovery

UNA Plataforma Abierta

OLTP + OLAP

UNA Copia de Datos

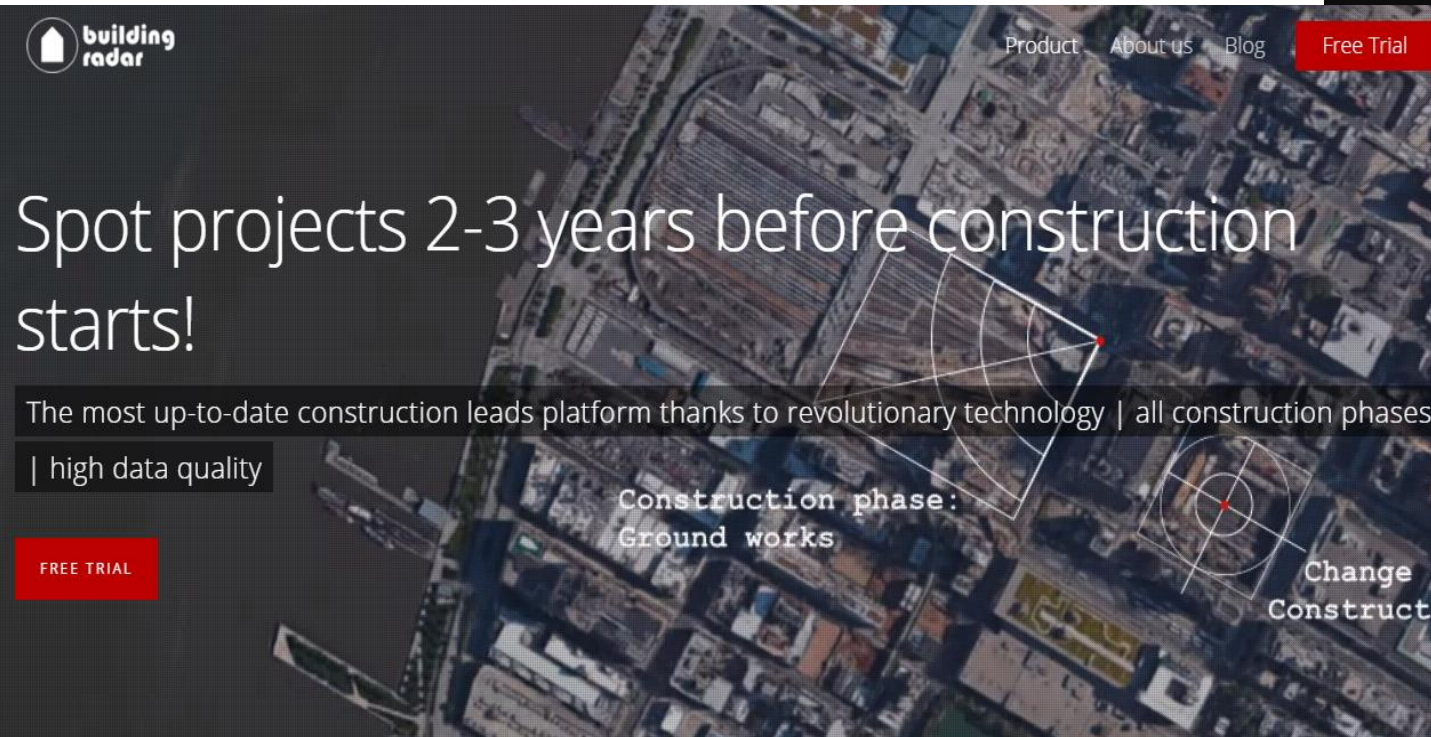
Big Data **Startups** usan SAP



- Unase - startupfocus.saphana.com

Building Radar

Identifica, Valida y Aprende sobre nuevos proyectos de construcción a nivel Global



building radar

Product About us Blog Free Trial

Spot projects 2-3 years before construction starts!

The most up-to-date construction leads platform thanks to revolutionary technology | all construction phases | high data quality

FREE TRIAL

Construction phase:
Ground works

Change Construct:

Utiliza algoritmos de búsqueda avanzados, procesamiento de lenguaje, imágenes satelitales y minería de datos

5.5 horas/día menos destinadas a identificación de prospectos potenciales

3M +
Construction projects

50K +
Construction projects updated this month

3000+
Construction projects added this week


200K +
Company profiles



building radar



Procesa 10 TB de datos
satelitales por semana y
relaciona un mapa del
mundo de edad promedio
de 5 días




Construction phase:
Ground works

Utiliza algoritmos de búsqueda para analizar millones de páginas web diariamente e incorporar la información de mercados de bienes inmuebles

Construction phase:
Ground works

Change in
Construction Phase



New construction
project detected

Construction phase:
Ground works

Change in
Construction Phase

Agrega datos en tiempo real, adicionando hasta 7 mil millones de puntos al dia sobre proyectos de construcción.

Despliega la información en gráficos interactivos y mapas de manera ágil para más de 1000 usuarios en forma simultánea.

Los usuarios pueden añadir información socio-demográfica en forma interactiva

Ground Area: 6.22M SF

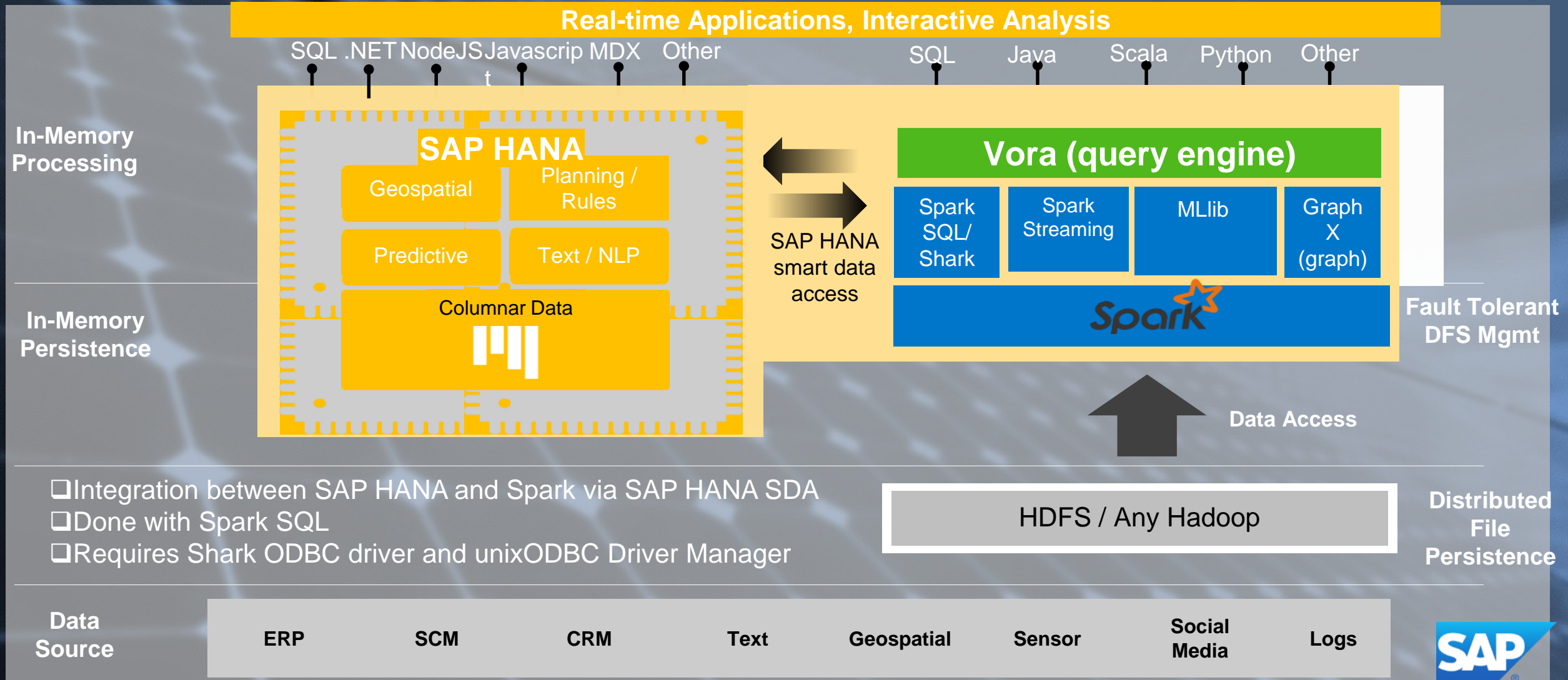
Construction phase:
Ground works

New construction project detected

Change in Construction Phase

SAP HANA VORA

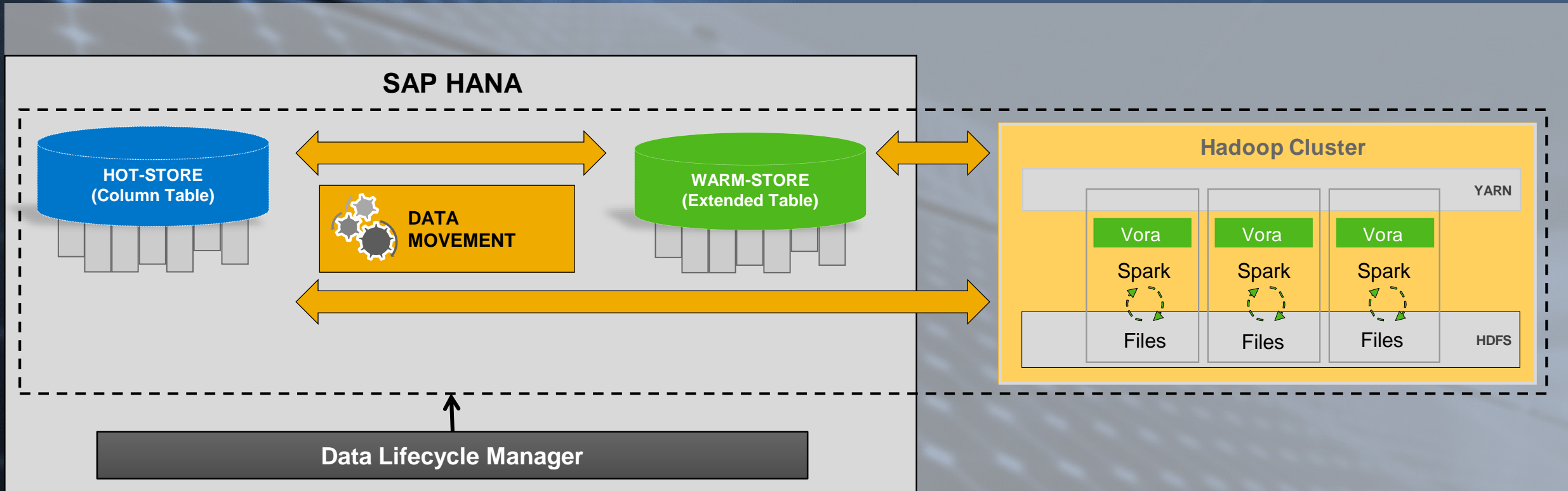
Motor de consultas “In-Memory” distribuido sobre Spark



- Integration between SAP HANA and Spark via SAP HANA SDA
- Done with Spark SQL
- Requires Shark ODBC driver and unixODBC Driver Manager



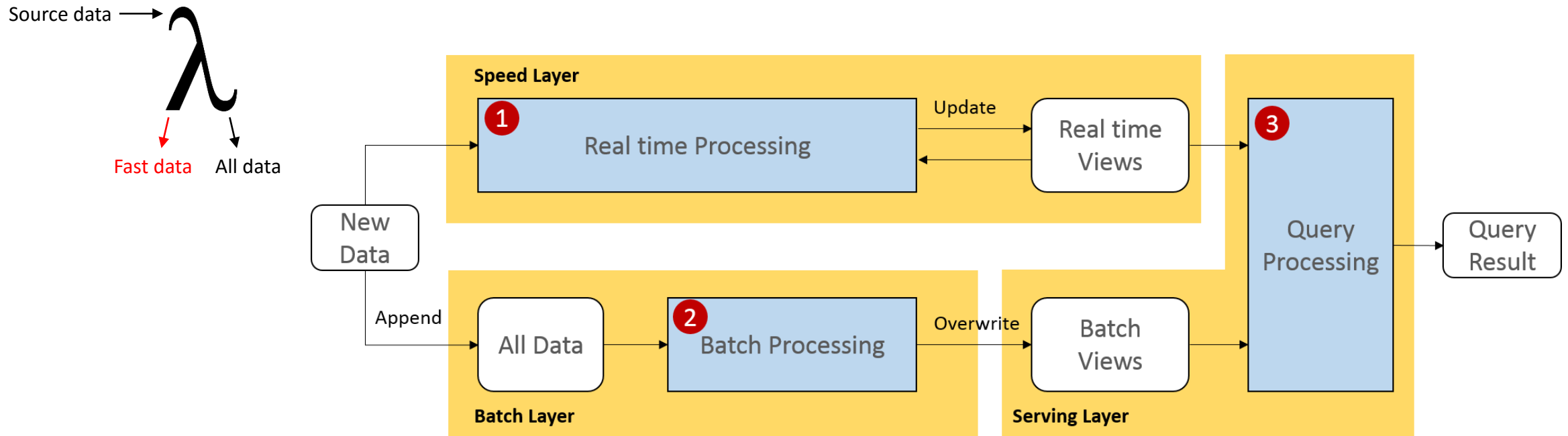
CenterPoint Energy Data Tiering en escala de Petabytes



- DW actual 12TB
- 120TB de Data de Smart Meters

Arquitectura Lambda y SAP HANA

Para todos los datos, SAP HANA actúa como punto central de acceso, modelamiento y administración

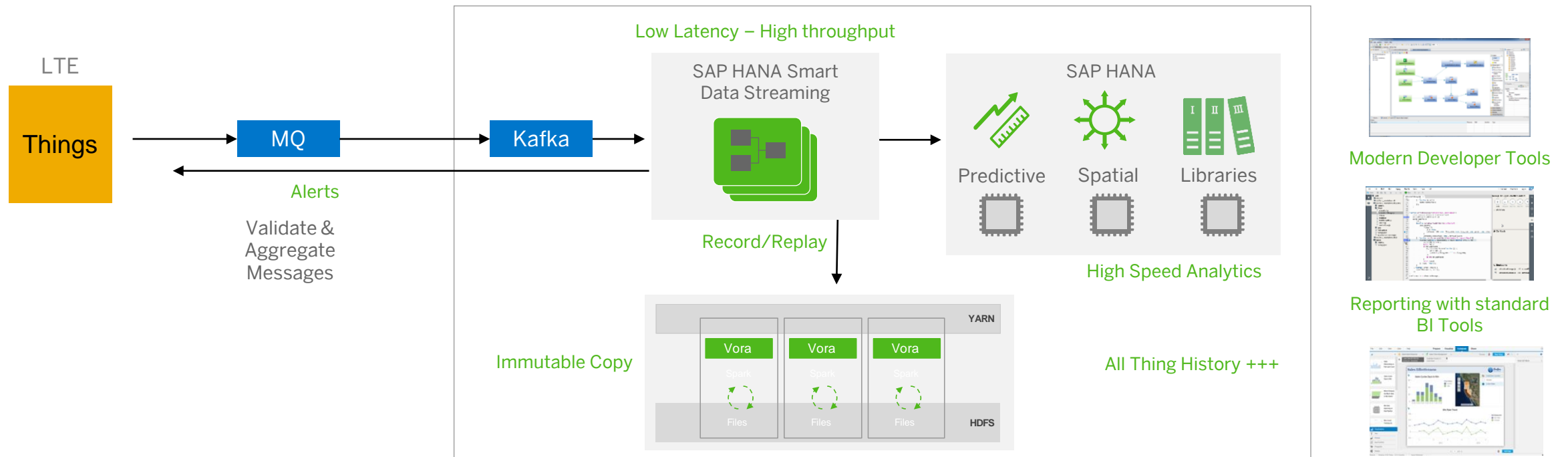


N. Marz, J. Warren: Big Data, Manning 2015, page 19. This book is considered to be the “lambda architecture book”
 But also: <http://radar.oreilly.com/2014/07/questioning-the-lambda-architecture.html>

SAP HANA lo habilita para implementar un arquitectura Lambda completa a partir de componentes pre integrados de la plataforma incluyendo motores de procesamiento y repositorios de datos

Arquitectura Lambda y SAP HANA

Para todos los datos, SAP HANA actua como punto central de acceso, modelamiento y administración

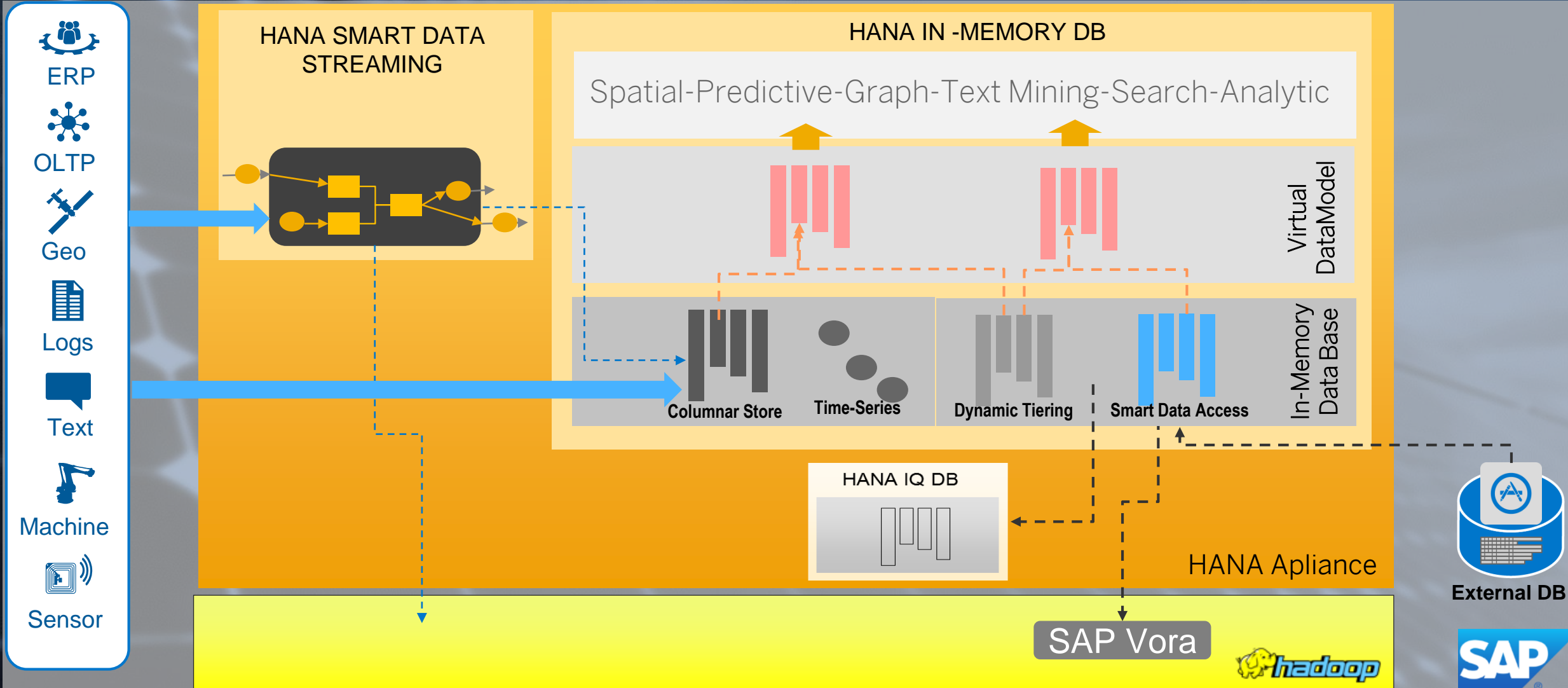


SAP HANA

- Smart Data Streaming gestiona los flujos de información
- High speed analytics habilita analisis avanzado inmediato sobre los datos recientes
- Vora permite utilizar todos los datos para el modelamiento y analisis historico

SAP HANA Big Data Platform

BI | Reporting | Ad-hoc Analysis | OLAP | Planning | Unstructured Data Discovery | Predictive | Apps



External DB



Nanjing City: Smart Traffic

Situación del Flujo de Tráfico en Tiempo Real, descubre congestiones y predice estado del trafico



Monitorea Velocidad, Volumen del trafico

Calcula Indice de Desempeño del Trafico cada 5 Minutos

100 millones de registros por día tomados de

- Floating Car Data (FCD)
- Fixed Device Data (FDD)

20:37

Monday, March 3

6°C

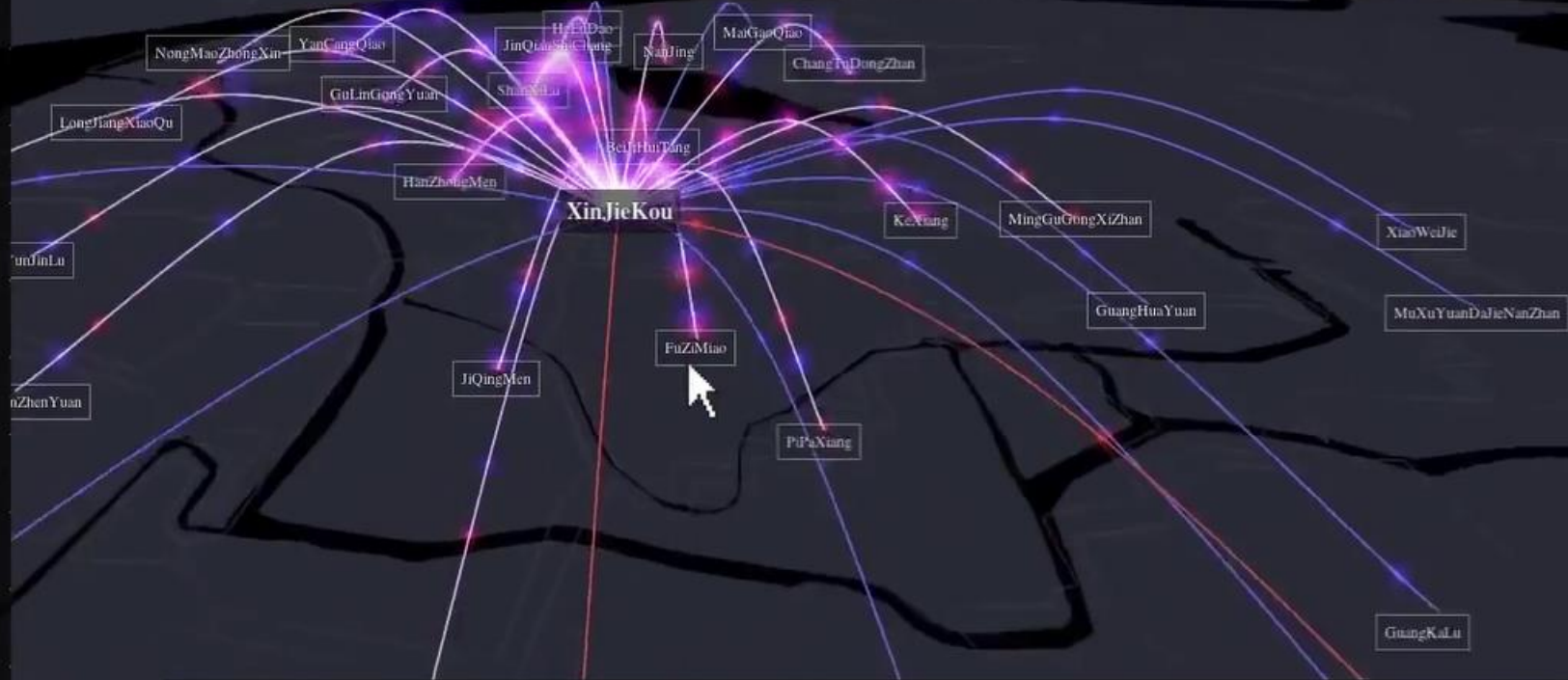
PM2.5 42

3D Curve Key Route Matrix

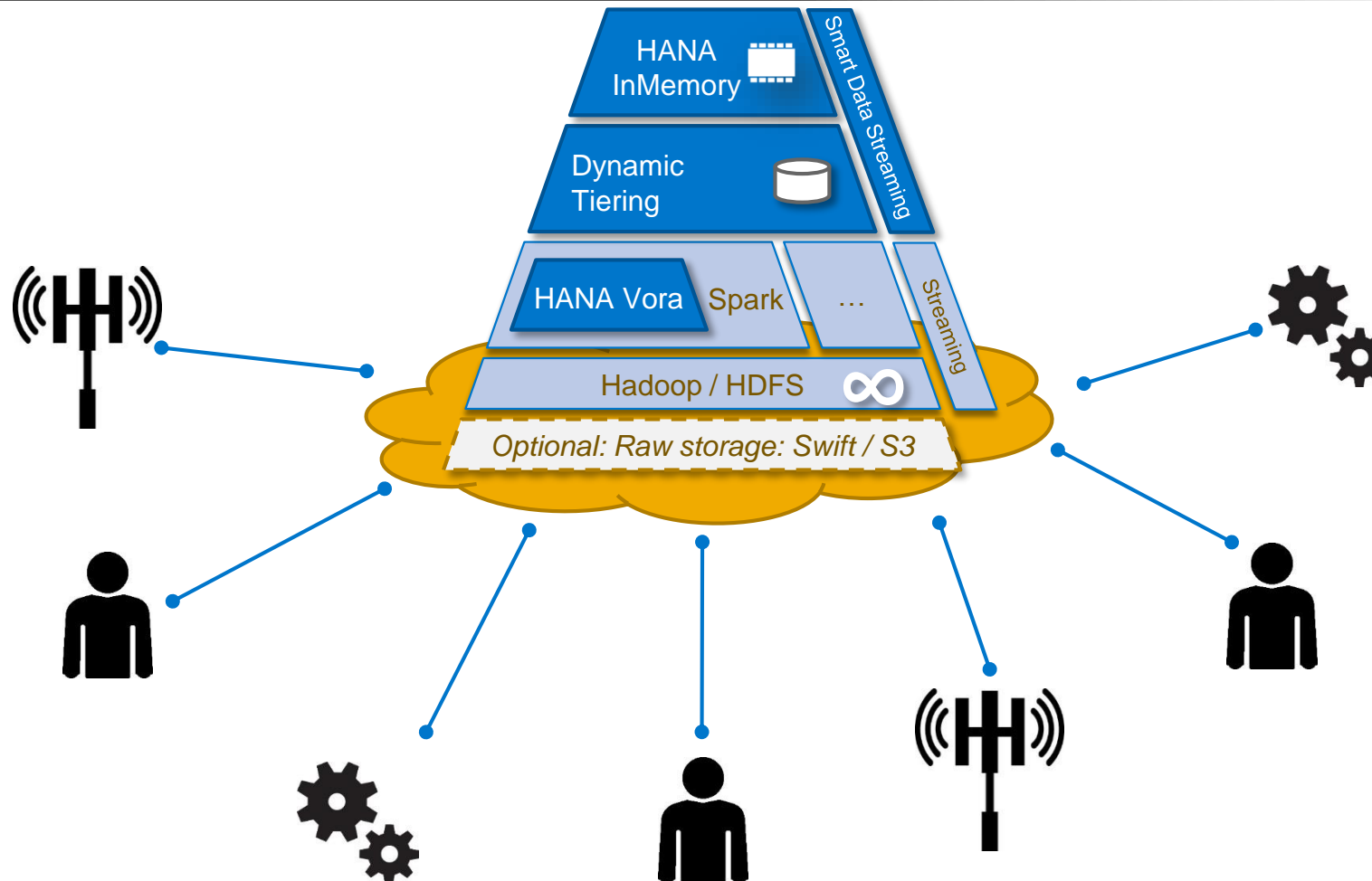
XinJieKou

Area: 3 km²
Avg. Daily Arrivals: 15779
Avg. Daily Departures: 18605

November 29, 2013
Friday



Podemos Siempre Centralizar toda la información?



Donde está la información en el borde?



IoT Gateways (Intel, Dell, Cisco,...)



Remote servers



Industrial PCs

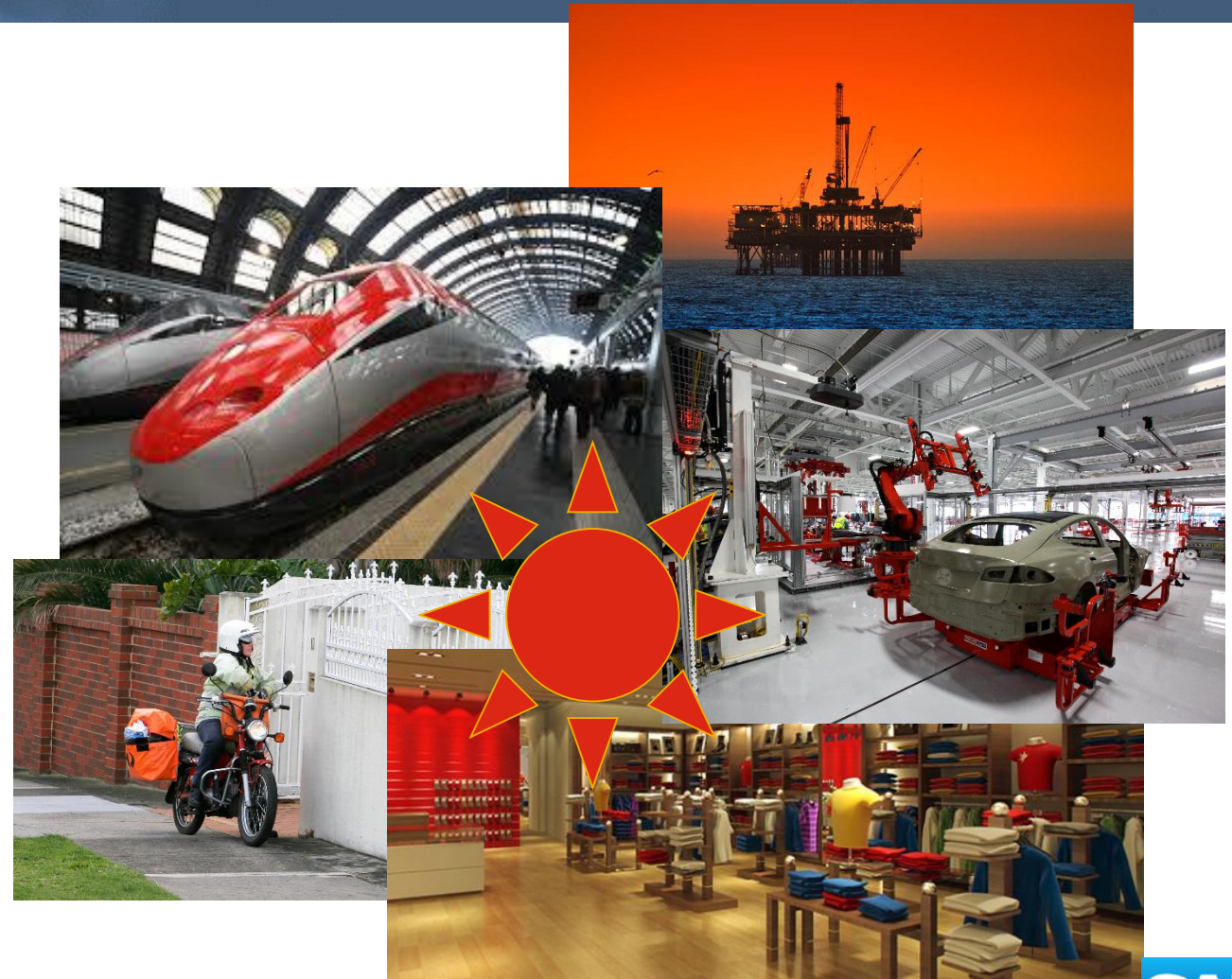


Raspberry Pi, etc

IDC estima:

“ Al 2019, **45%** de los datos creados por IoT se almacenarán... cerca, o al borde de la red * ”

* IDC FutureScope: Worldwide Internet of Things 2016 Predictions, November 2015



Que tan grande son los datos de IoT?

Type of data	Recommended SAP data management solution	Size range	Expressed numerically
Transactional data	SAP HANA – <i>in-memory</i>	Terabytes (TB)	10^{12}
Traditional analytics (OLAP)	SAP IQ or SAP HANA DT (Dynamic Tiering) – <i>disk-based</i>	10s to 100s of Terabytes	10^{13-14}
Big Data	Hadoop with SAP HANA Vora – <i>both in-memory & disk-based</i>	Petabytes (PB)	10^{15}
IoT data	SAP SQL Anywhere <i>distributed as an edge database</i>	Exabytes (EB)	10^{18}

Llegaran a ser tan grandes estos datos?

Ejemplo de un gran Fabricante de Automóviles – Números provistos por el cliente

100 millions de autos (10-años de vida) x 4 mensajes/día x 25-Kilobyte por mensaje

= 10 Terabytes/día = 3½ Petabytes/año x 10 años de data = 3% de 1 Exabyte

Puede crecer a 200 mensajes/día x 50-Kilobyte por mensaje → **3½ Exabytes**

Para 1 solo caso de uso de un cliente

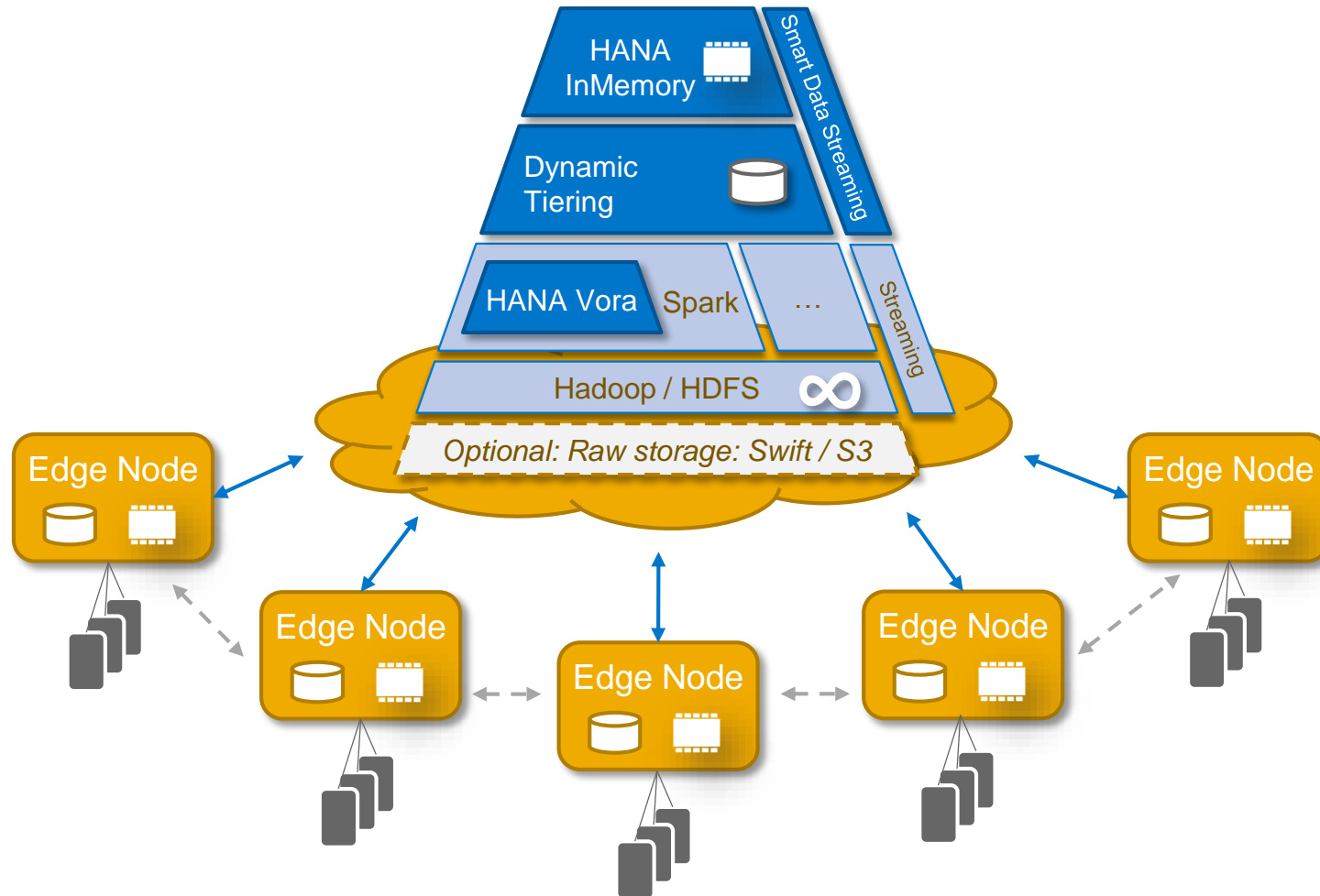
Algunos analistas estiman: 50 mil millones de dispositivos conectados al 2020

Supuesto conservador: Un mensaje promedio de 500-byte por minuto por dispositivo

50 mil millones de dispositivos x ½ million de minutos/año x ½ Kilobyte = **12½ Exabytes/año**

Hacia el 2025 estos números solo crecerán aun más

Será mas eficiente gestionar esta informacion en el borde?



Algunas estimaciones:

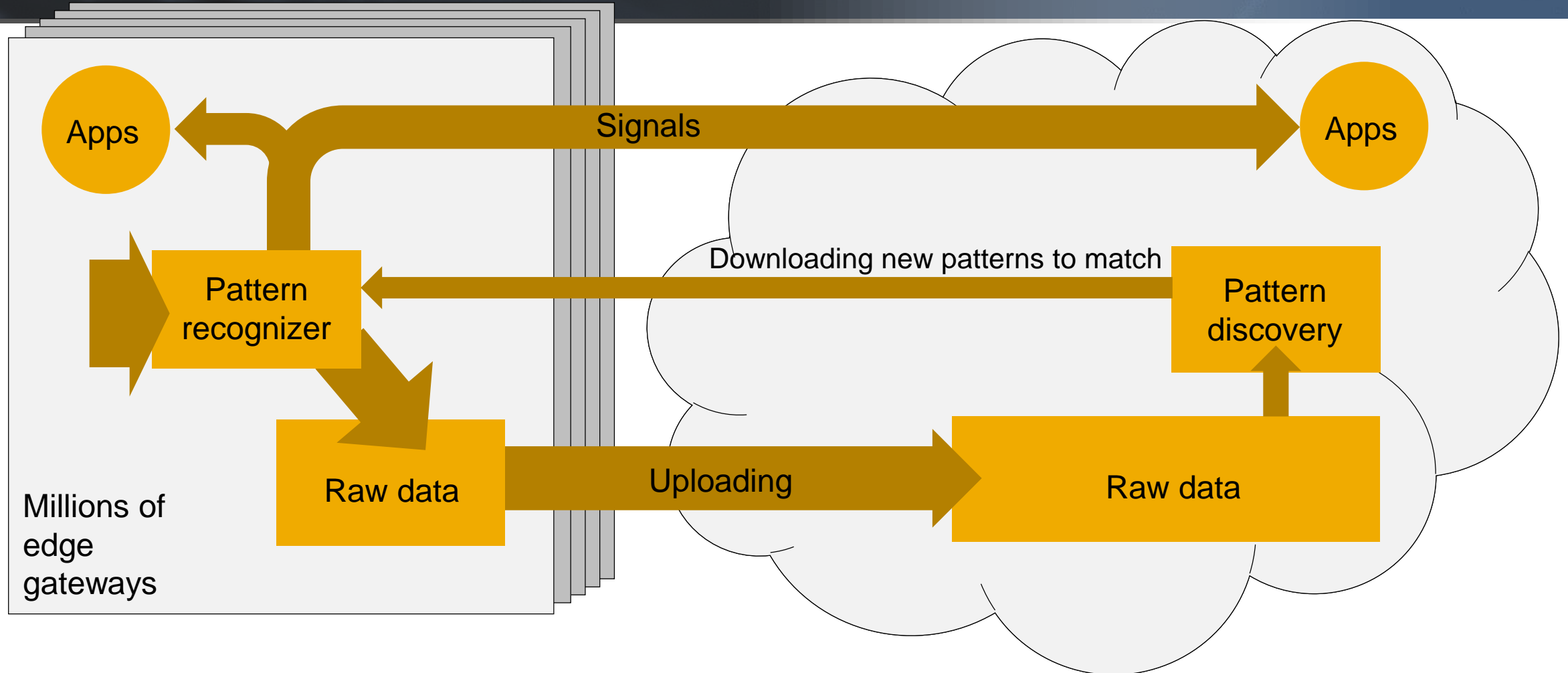
Mover un Exabyte a la nube
1 Exabyte/año = 250 Gigabits/sec,
cada segundo 24/7/52

= 8% de la capacidad promedio DE-CIX o 10% de AMS-IX – Los 2 nodos de intercambio más grandes de Europa*

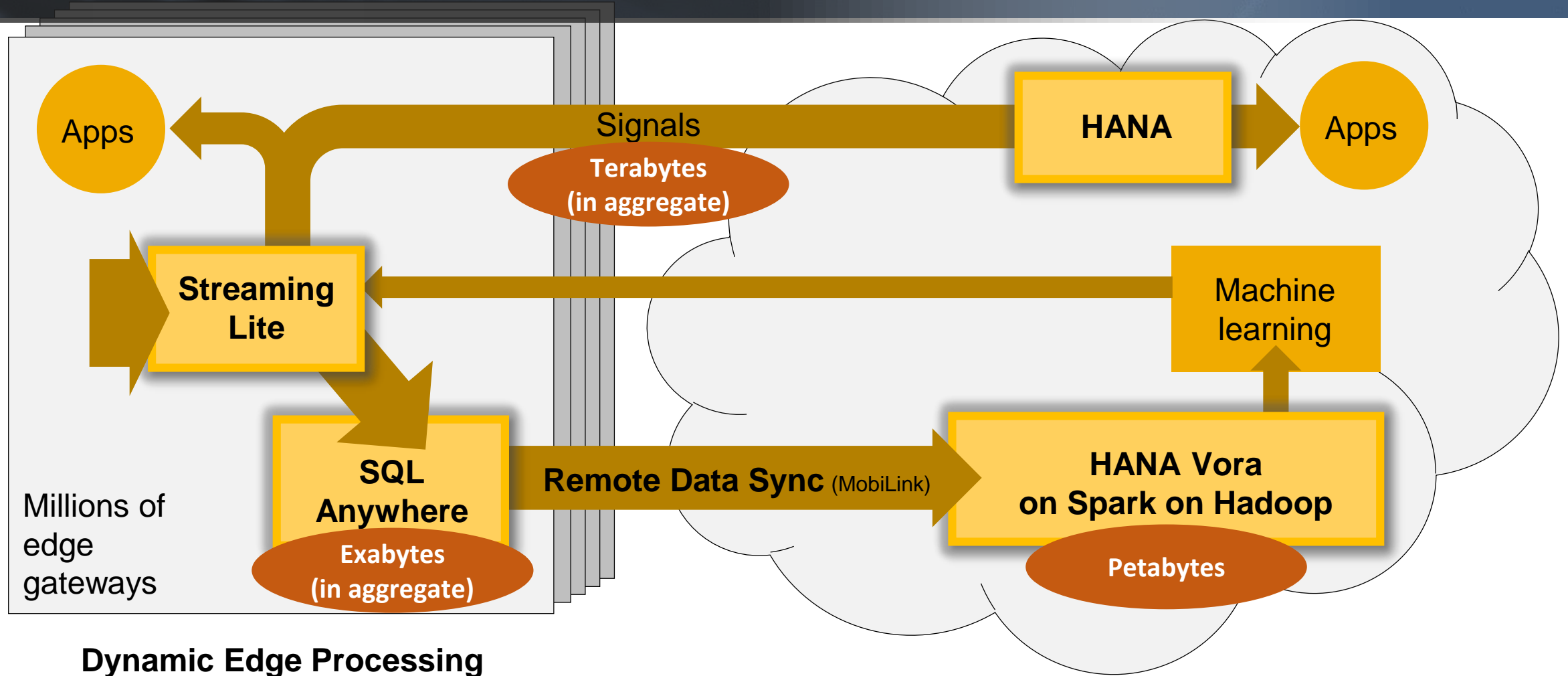
10 Exabytes/año podría consumir toda la Internet en Europa


* en.wikipedia.org/wiki/List_of_Internet_exchange_points_by_size

Se hace necesaria una plataforma distribuida de gestión de información



El reto: automatizar y gobernar centralmente la plataforma distribuida de gestión de información





Haciendo posible el viaje en Universo
de Información

Muchas Gracias!



UNIVERSIDAD AUSTRAL | INGENIERÍA

gonzalo.Malaga@sap.com

