

# The Logical Data Warehouse

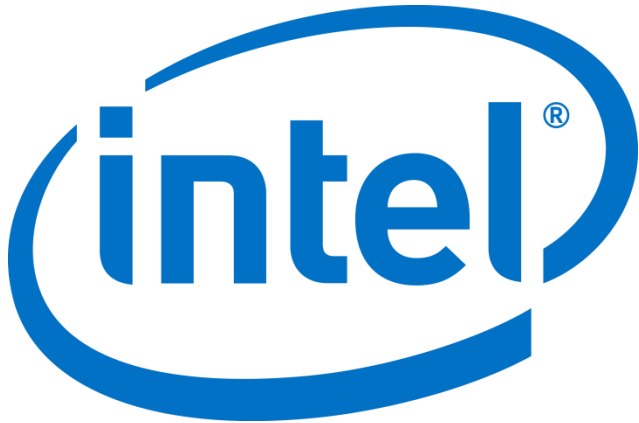
## What It Is and Why You Need It

**Philip Russom**

TDWI Research Director for Data Management

June 24, 2015

# Sponsors



# Speakers



Philip Russom  
TDWI Research Director,  
Data Management



Tom Traubitz  
Senior Director, Product Strategy  
SAP

# Agenda

- The Logical Data Warehouse
  - *Definitions*
  - *Characteristics*
- Enabling Technologies
  - *Virtualization*
  - *In-Memory Functions*
- In-Memory Data Fabric
  - *The Logical Data Warehouse concept applied more broadly*
- Beneficial Use Cases
  - *Real-time & near-time processes*
  - *Biz visibility & situational awareness*
  - *Agility & flexibility, integrating multiple platforms*
- Recommendations



**PLEASE TWEET**  
**@pRussom, @SAP, #TDWI,**  
**#DataFabric, #LogicalDW,**  
**#Analytics, #RealTime**

# Upcoming Points

- There isn't one, single architecture for all data warehouses (DWs)
  - *Each org is different*
- Expect multiple architectures
  - *A well-designed DW has multiple architectural layers*
  - *Architectural approaches get mixed together into hybrids*
  - *A DW architecture interacts with architectures for data integration, reporting, analytics, operational applications, etc.*
- The warehouse is still vital, even central
  - *But it's evolving into a multiple platform environment*
  - *Architecture is more important than ever, but now as a logical design that's deployed over multiple physical platforms*
- Please don't ask me to draw a Reference Architecture for DWs
  - *Given the current diversity, there isn't just one. But I'll describe many.*



# Drivers of Change

## Does your primary enterprise data warehouse have an architectural design?

Yes	79%
No	18%
Don't know	3%

## Is the architecture of your data warehouse environment evolving?

Yes – moderately	54%
Yes – dramatically	22%
No – except with DW updates	22%
Don't know	2%

## What **technical** issues or practices are driving change in your DW architecture?

Advanced analytics	57%
Increasing data volumes	56%
Real-time operations	41%
Business performance mgt	38%
OLAP	30%
Non-relational data	25%
Virtualization of data	23%
Cloud adoption	21%
Streaming data	15%

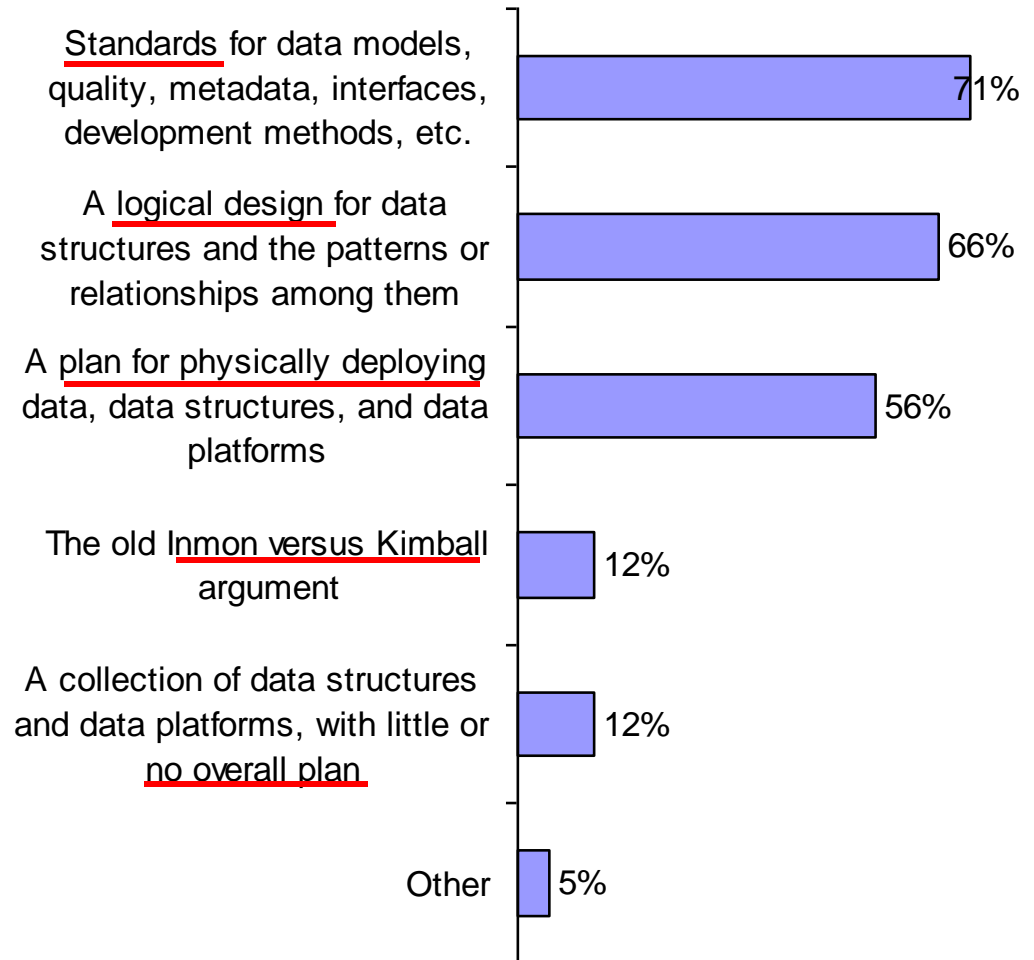
## What **business** issues or practices are driving change in your DW architecture?

Competitiveness	45%
Fast-paced business processes	43%
Compliance	29%
Funding	29%
Sponsorship	26%
Reorganizations	25%
Centralizing business control	30%
Departmental power struggles	19%
Mergers and acquisitions	18%

# What do you think data warehouse architecture is?

Select all that apply.

Source: TDWI survey run in late 2013.  
Based on 1197 responses from 538 respondents. 2.2 responses per respondent, on average.



# Logical versus Physical DW Architectures

## And Other Architectural Components that Coexist

Today's  
Focus



- **Logical architecture** – mostly about data models and their relationships, with a focus on how these represent organizational entities and processes
  - **Data standards** – *including standards for data modeling, data quality metrics, interfaces for data integration, programming style, format standards, etc.*
- **Physical architecture** – mostly a plan for deploying data and data structures based on the workload and platform requirements of each
  - **System architecture** – *a topology of hardware servers and software servers, plus the interfaces and networks that tie them together*



# DEFINITIONS OF THE Logical Data Warehouse



- TDWI: A Data Warehouse is user-defined data architecture
  - *The architecture & its design components must be populated by data*
  - *But the data can be physical, virtual, or both*
- Gartner's view: A Logical Data Warehouse depends on virtual tech more than older DWs
  - *From simple federation to object-oriented virtualization, plus virtual views, indices, semantics, server memory...*
- Building out the Logical Layer of your DW is important
  - *Focus on design (not server platforms), agile dev & updates, more real-time options (OpBI), multi-source insights on the fly (analytics)*

# ENABLING TECH for LOGICAL DATA WAREHOUSE

## Data Virtualization

- Purely semantic views of data structures
  - *No physical data, until view is materialized*
- Benefits of data virtualization
  - *Doesn't prep & persist a lot of data on the off chance a user or app might need it*
  - *Collects fresh data, as needed, instead of hoarding stale data*
- Various processing available
  - *Some views are read only*
  - *Others can write data and perform data processing functions (or call them)*
    - E.g., views that represent joins or aggregates
- Virtualization intersects with real time
  - *Most views (but not all of them) execute in real time (or close) when materializing data*



# ENABLING TECH for LOGICAL DATA WAREHOUSE

## In-Memory Data Functions



- Data mgt & processing in server memory
  - *Rarely a DBMS in memory*
  - *Usually a data subset in memory*
- Benefits of in-memory data
  - *Eliminates disk IO, which is traditional bottleneck for data mgt*
  - *Provides high performance for many data-driven applications, including data virtualization*
- Various processing available
  - *Simple table cached in memory*
    - E.g., table of metrics/KPIs for dashboards
  - *Multidimensional data*
    - E.g., cube of sales data for intraday analysis
  - *Analytic models and scores*
    - E.g., rescored intraday to spot/report likely churn

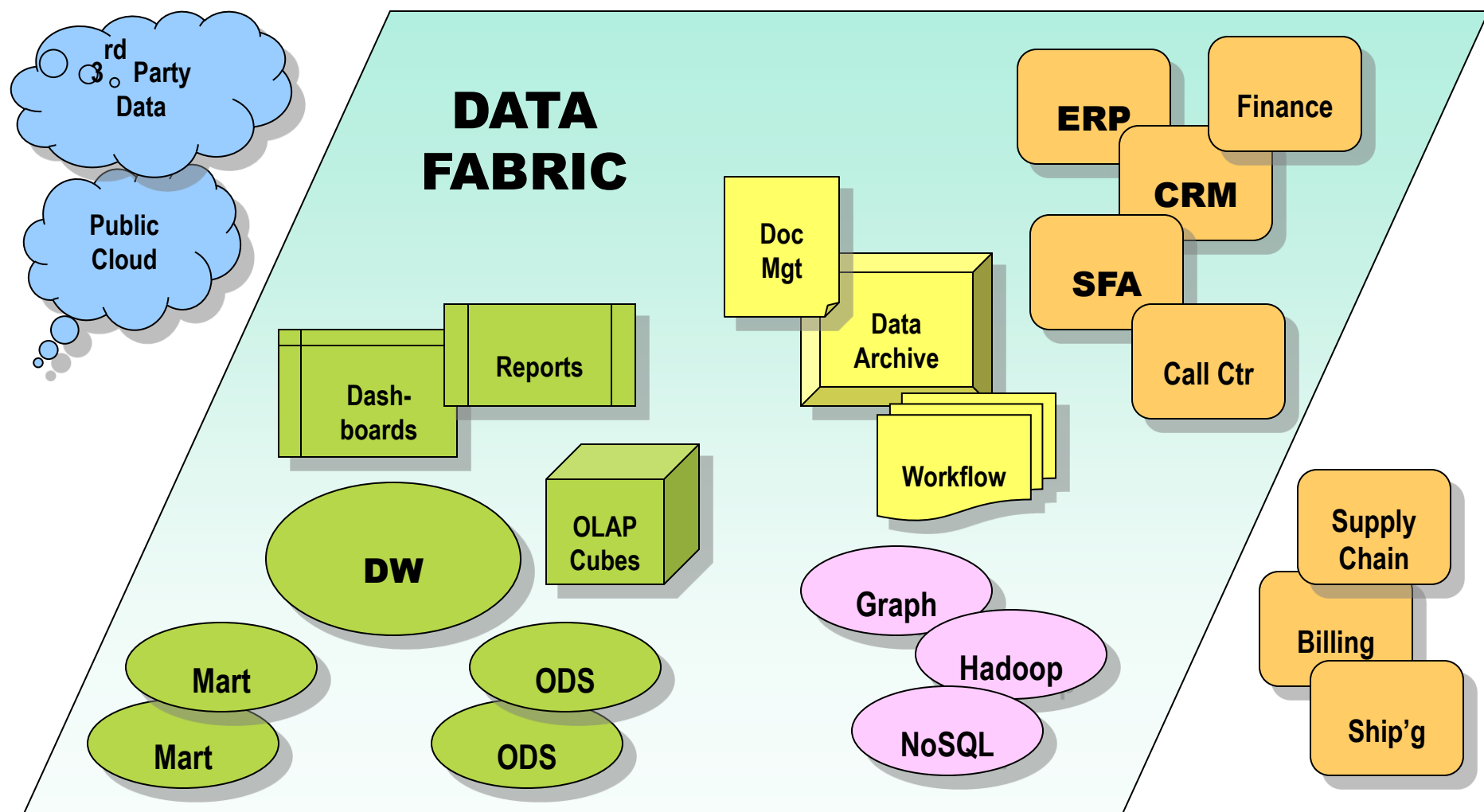
# ENABLING TECH for LOGICAL DATA WAREHOUSE In-Memory Data Fabric



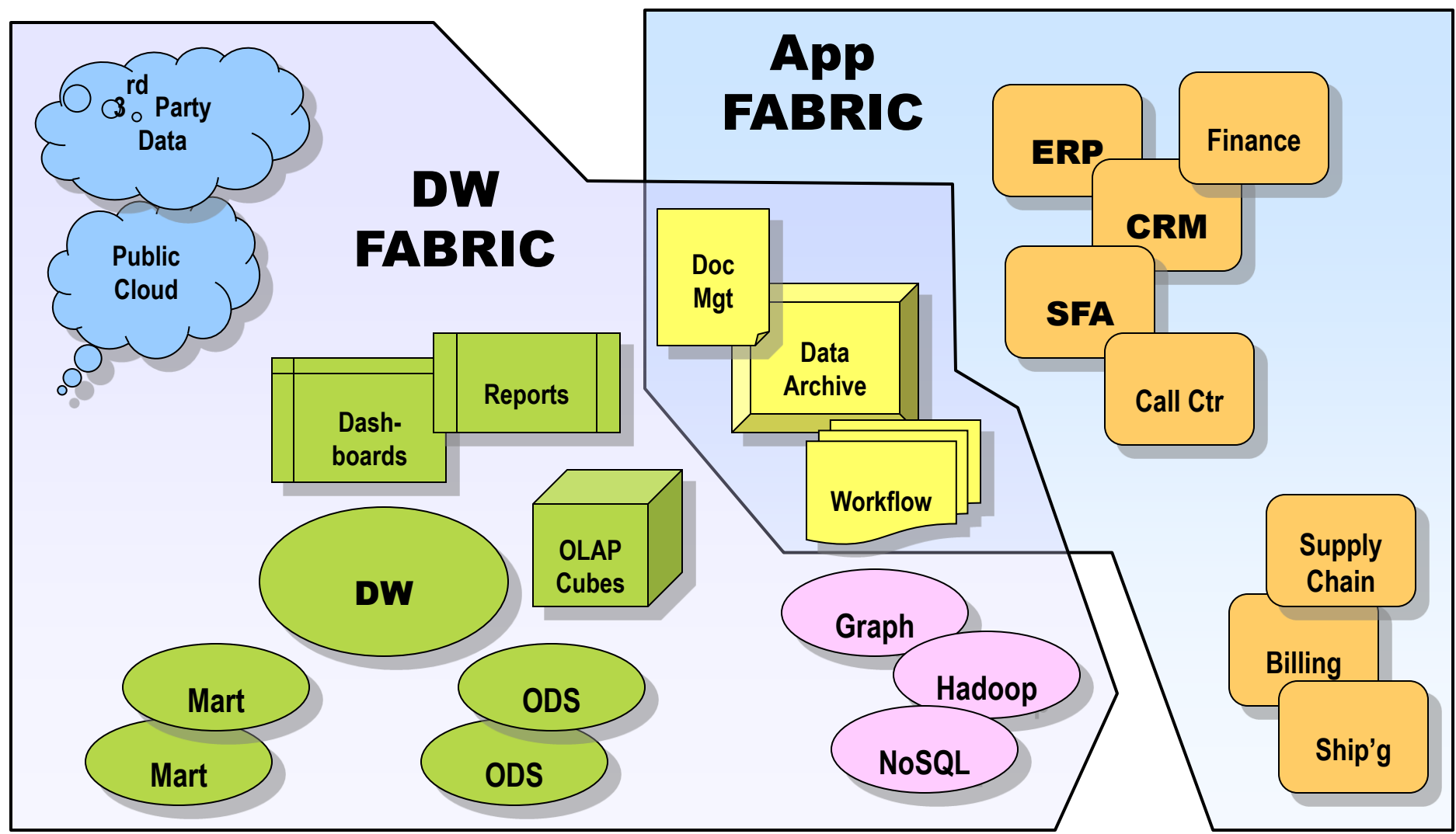
- Data Fabric is a unified view (or collection of views) of data in multiple systems across an enterprise
  - *Plus a simplified (yet diverse & performing) collection of interfaces into such sources and targets*
- The point of a data fabric is to provide:
  - *A fairly comprehensive big picture of enterprise data*
  - *A single layer through which data can be accessed, thereby reducing data redundancy, movement, processing*
  - *A simplified view & mechanism that enables more user types*
- In-Memory Data Fabric (IMDF) is combination of things:
  - *The data fabric, in-memory data functions, and data virtualization discussed earlier, integrated w/usual apps, databases, & data mgt tools*
- Benefits of IMDF
  - *A high-performance form of a data fabric, due to in-memory data functions, parallel processing, direct interfaces, optimization, etc.*
  - *Real-time speed for time-sensitive biz practices, lean data mgt, scalability, embedding analytics in apps, operationalization, etc.*



# Data Fabric – Example 1



# Data Fabric – Example 2



# Use Cases for the Logical Data Warehouse / Data Fabric

## Real-Time Business Practices

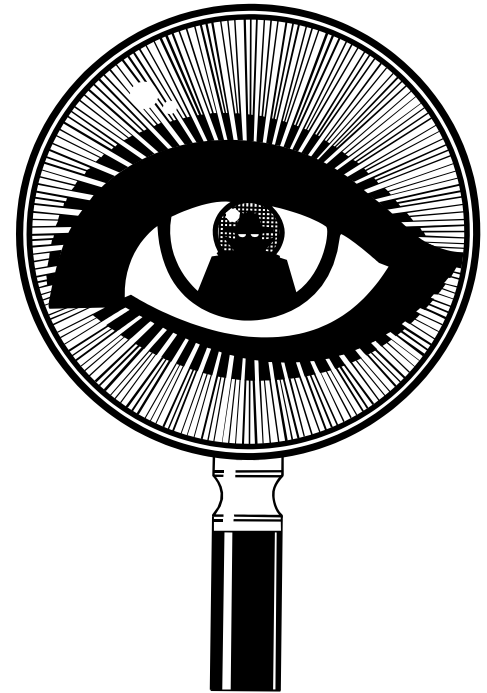
- Real-Time technologies are a foundation for time-sensitive business practices:
  - *Operational business intelligence*
  - *Just-in-time inventory*
  - *Facility monitoring*
  - *Self-service information portals*
  - *eCommerce recommendations*
  - *Production yield & workforce mgt in manufacturing*
- Real-Time Reporting is common
  - *Real-Time Analytics is coming on strong*



# Use Cases for the Logical Data Warehouse / Data Fabric

## Visibility and Awareness

- Visibility = Know and act on the knowledge quickly:
  - *Know and correct SLA or performance problems*
  - *Spot and stop fraud or security breaches*
  - *Feel confident, knowing that “all systems are go”*
- Situational Awareness = See & react accordingly:
  - *See a cluster of street crimes*
    - Deploy squad cars as a deterrent.
  - *See a drop in unit production on manufacturing floor*
    - Bring in more workers and turn on more machinery.
  - *See a product recurring in abandoned shopping carts*
    - Run a promotion to close more sales of that product.
  - *See a social media sentiment or pattern*
    - Direct it or correct it as it evolves.





# Use Cases for the Logical Data Warehouse / Data Fabric

## Integration and Agility

- LDW usually involves advanced forms of data integration
  - *Federation, virtualization*
  - *These are key to unifying multi-platform data ecosystems, especially data warehouse environments*
  - *Move data around less (plus real time, as mentioned earlier)*
- LDW, if used well, enables agile development & upgrades
  - *Developing with virtual views of data can be faster than relocating data physically*
  - *Virtual views can be altered without heavy movement of data to transform and reload data*

# Recommendations

- Recognize that successful data warehouse architectures have integrated logical and physical layers, plus other components.
  - *Determine the business and technical drivers in your organization, and let those determine the evolution of your DW architecture.*
- Note that a data fabric is seldom 100% virtual, in-memory, real time, etc.
  - *Based on your organizational needs, selectively decide which data is best represented virtually, persisted, near real time, in a fabric, etc.*
- In-memory data functions & data virtualization are more viable than ever – so use them!
  - *New level of maturity for speed, reliability, functionality, interoperability*
- Put in-memory functions & data virtualization together in a data fabric
  - *Use in-memory functions for speed and as a point of integration*
  - *Use data virtualization for agile dataset design in development and integration on the fly in deployment*
- A logical layer or data fabric should be an access layer
  - *With interfaces, not just views*
- For success with a logical DW and similar architectures (such as data fabric)
  - *Infuse it with ample data virtualization and in-memory caching and processing.*






# SAP Data Warehouse

## Reinventing the Data Warehouse

June 2015







We are entering into a new era of unprecedented **change** across a multitude of dimensions

**Rising Customer Expectations**

**5 billion**  
people worldwide  
will become  
**middle class**

**A Dramatically Changing Workforce**

**75%**  
of global workforce  
will be  
**Millennials**

**Pressure on Resources**

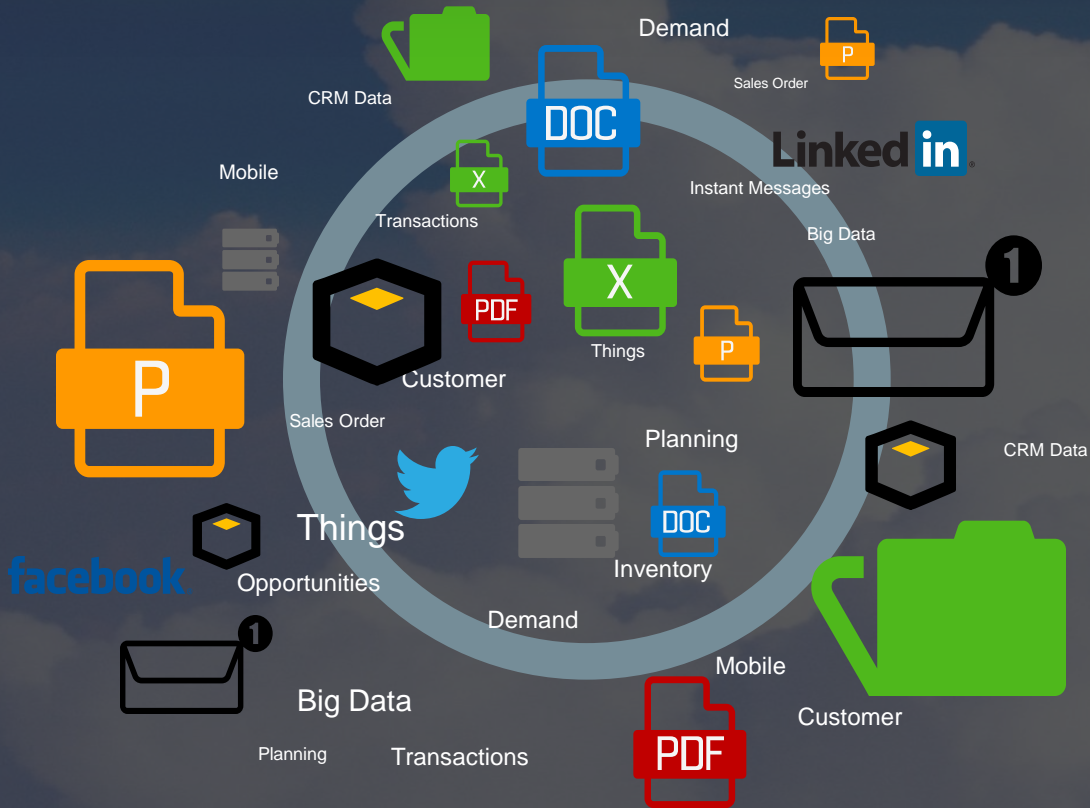
**50%**  
of the world's population  
will live under  
**water shortage**

**Network Effect/Explosion in Structured and Unstructured data**

**1.3 billion**  
**people** on business &  
social networks today

**50 billion**  
connected devices and  
“**internet of things**” by 2030

# Key Trends



# Cloud

Cloud spending will surge by 25%, reaching over \$100 billion. There will be a doubling of cloud data centers.

# Internet of Things

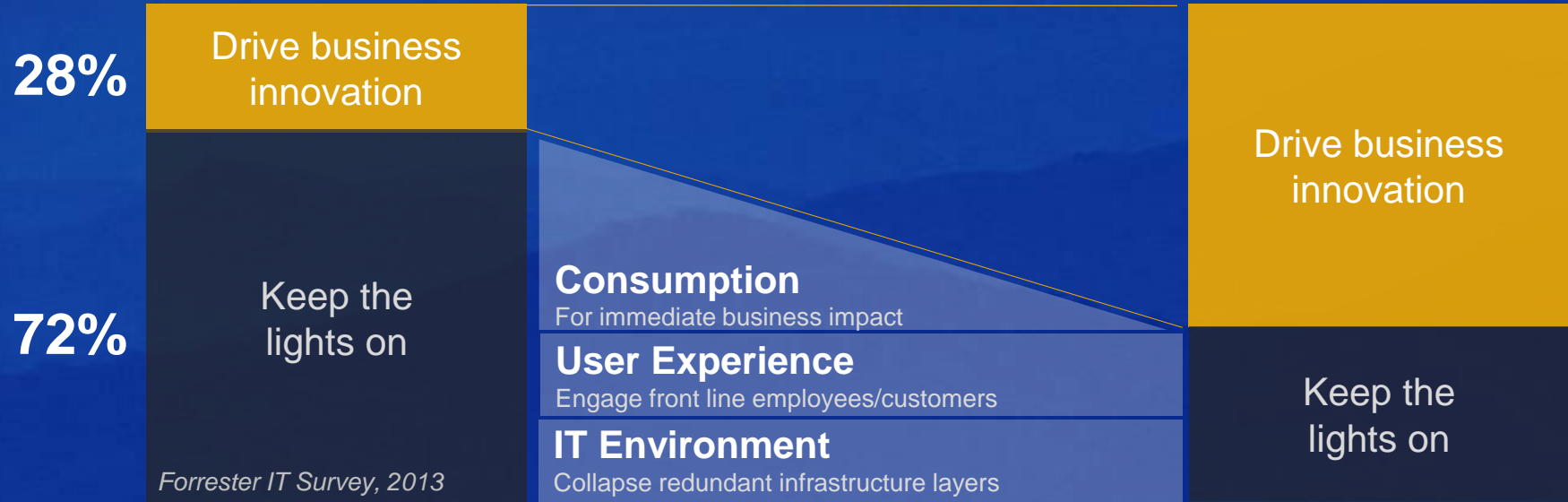
30 billion devices, sensors in 2020 – driving \$8.9 Trillion in revenue. The need for real-time processing and analytics will explode

# Data Lake

Data volumes will continue to grow to 6 billion petabytes, including unstructured data such as social networking data and low level IoT data. Mining the value from this data is essential



**Complexity** built up over decades limits the ability to innovate; radical **simplification** is needed to unlock the potential.



40% executives **worry that their organizations will not keep pace** with technology change and lose their competitive edge.

– McKinsey study, 2013



# SAP's Data Warehouse enables a revolutionary approach streamlines and simplifies data warehousing

Providing greater speed and scale along with agility for development and efficiency that reduces data movement and data preparation. SAP's complete architecture offers:



Pre-packaged  
or Customize



Flexible  
Architecture



Rapid  
Deployment

# SAP Gives You The Power of Both Custom and Packaged

More Degrees of Freedom

## Customized Data Warehouse

- Usually depends on SQL tools and low-level programming
- Fewer controls on schema updates
- Easier to change

## SAP HANA PLATFORM

Real-time transactions + end-to-end analytics

Extended Application Services

Processing Engine

HANA  
Smart Data  
Streaming

HANA Dynamic  
Tiering

HANA Advanced  
Analytics

Application Function Lib. & Data Models

Integration Services

## Pre-Packaged Data Warehouse

- Controlled schemas, often prepopulated with structure
- Lifecycle management of schemas
- High level languages and less programming
- More prebuilt tools to purpose

## SAP Provides The Best of Both Approaches!

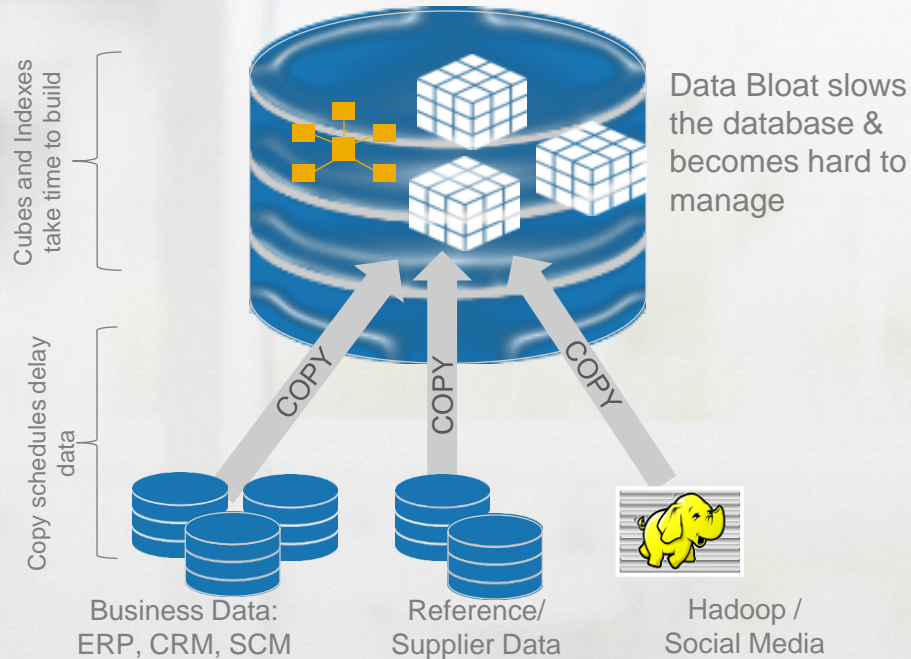
- An integrated architecture that reduces data redundancy while keeping all information at hand
- Utilizes state-of-the-art in-memory techniques that furnish answers in-context, in real time
- Makes more data available at the right time to the right person at the right place in the business process

Less Time to Implement

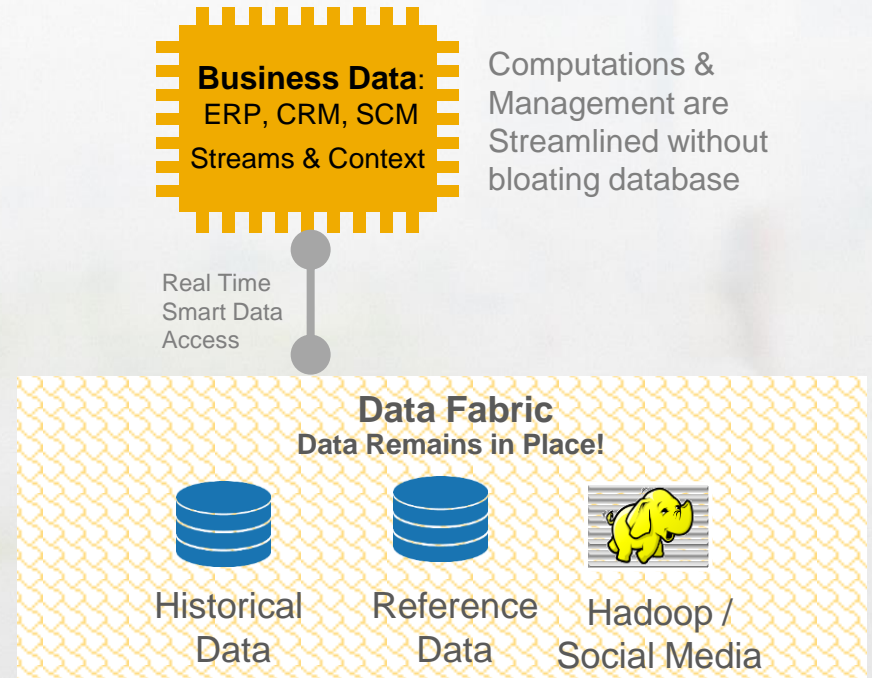


# HANA Flexible Architecture Example: Data Fabric

Traditional Data Warehouses Just Copy Data And Create More And More Copies In Indexes

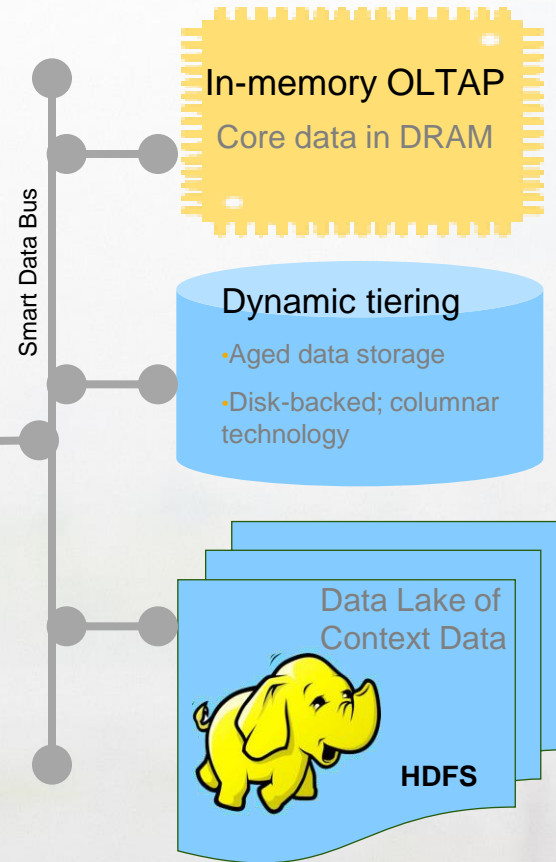
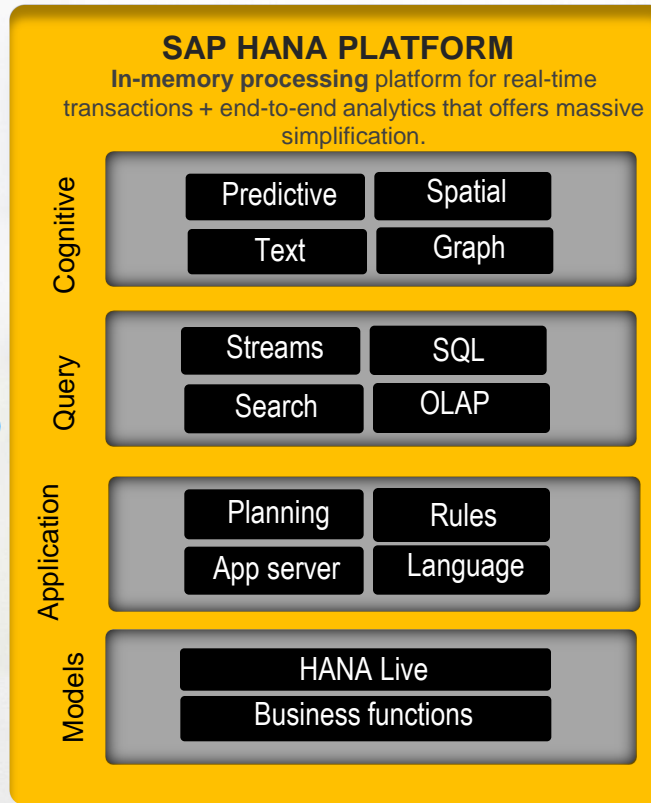


HANA Keeps Critical Data in Memory without Copies or Support Indexes



# HANA Flexible Architecture: Data Lake

Business Network, Real-time Applications, Big  
Data Warehousing, Interactive Analysis,  
Mobile Experience



ERP  
SaaS  
Network  
Text  
Geo  
Sensor  
Social  
Logs

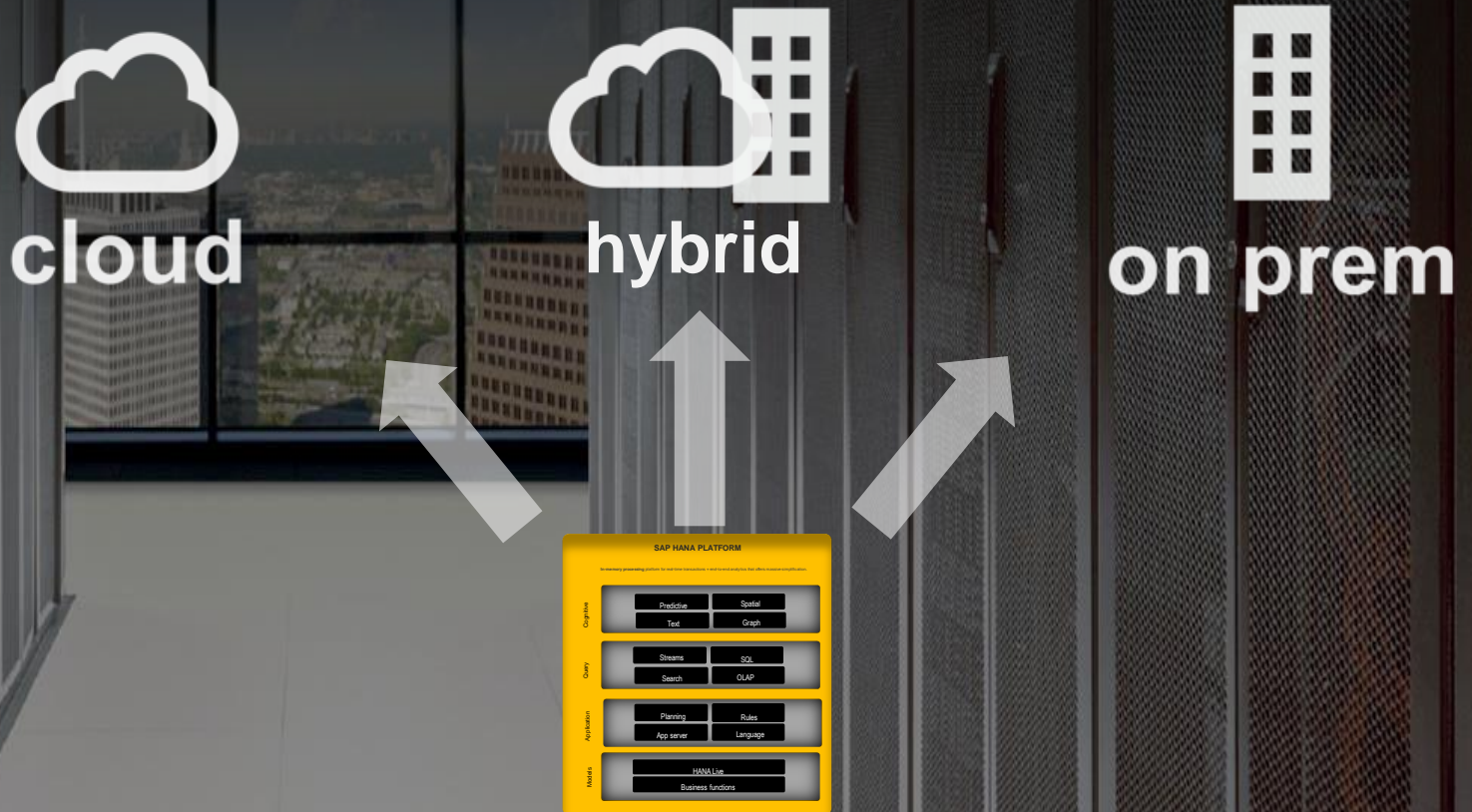
CONSUME

COMPUTE

STORE

INGEST

# Rapid Deployment





# Customer value delivered by SAP Data Warehouse

## Simplified Architecture

- Eliminate or reduce data movement
- Fewer copies of data

## Enterprise Wide Analytics

- Access data across your enterprise
- Unmatched federation of data without centralizing

## Real-time Analytics

- In-memory performance gives answers in seconds, not hours
- Reduced latency means current data is addressed not old data

## Data Lake

- Petabytes of historical data storage
- Advanced analytics for mining non-traditional data
- Extensive Hadoop and no-SQL support

## Internet of Things

- Data management and analytics from device to enterprise
- Streaming analytics

# SAP Data Warehouse Portfolio on Intel

• **SAP Business Warehouse**  
Real-time complex event processing

• **HANA Dynamic Tiering**  
Multi-temperature Tiering

• **HANA Smart Data Streaming**  
Real-time complex event processing

## SAP HANA PLATFORM

Real-time, in-memory database,  
data processing, and application  
platform

• **SAP IQ & NLS**  
Logical Big Data warehousing (OLAP)

• **HANA Advanced Analytics**  
Breakthrough performance at lower cost

• **SAP Data Services**  
All types of data integration

A-z

**Simplify**



**Accelerate**



**Innovate**







# Thank you

<http://hana.sap.com/usecases/data-warehousing.htm>

**Tom Traubitz**  
**[tom.traubitz@sap.com](mailto:tom.traubitz@sap.com)**

# Questions?



# Contact Information

If you have further questions or comments:

Philip Russom, TDWI  
prussom@tdwi.org

Tom Traubitz, SAP  
tom.traubitz@sap.com