

Real-Time and Big Data Challenge Data Management Best Practices

Philip Russom

TDWI Research Director for Data Management

May 1, 2012

TDWI RESEARCH



Sponsor



Speakers



Philip Russom
TDWI Research Director,
Data Management



Marie Goodell
Sr. Director for Solution
Marketing, SAP

Today's Agenda

- The Mega Trends
 - *Competing trends*
 - *Real-time versus big data, complicated by interoperability*
 - *May lead to compromises in data management best practices*
- Data management strategies without compromise
 - *Moore's law, selecting the right data management discipline, new data management architectures and options, integrated tool platforms, using more real-time data management functions*
- Recommendations



The Mega Trends

*These are today the most influential trends in data-driven IT disciplines, e.g.:
Business intelligence, data warehousing, data management, analytics...*

- Real-Time Operation
- Scaling Up/Out to Big Data
- Interoperability among IT Systems
- Do More With Less

Big Data versus Real-Time Operation

With Interoperability Caught in the Middle



Big Data



erability
Interop-

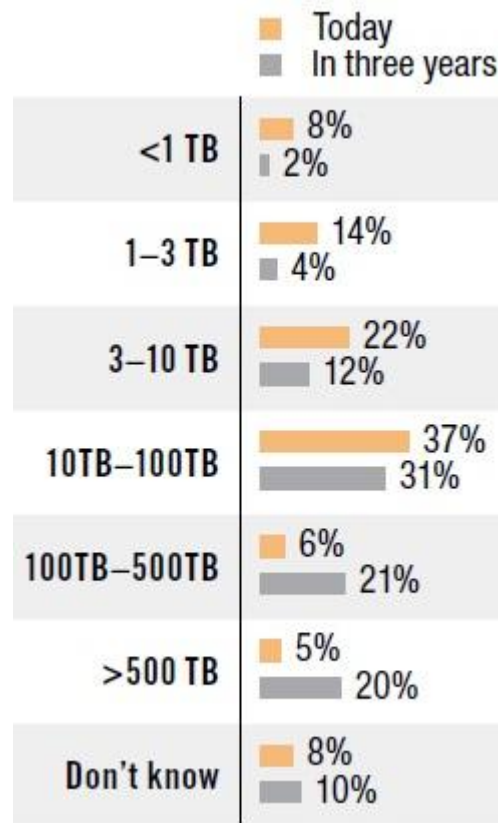
Real-Time Operation



“Do More With Less” – All you need is regulation rope & matching uniforms!

Big Data Volume: TBs today, PBs soon

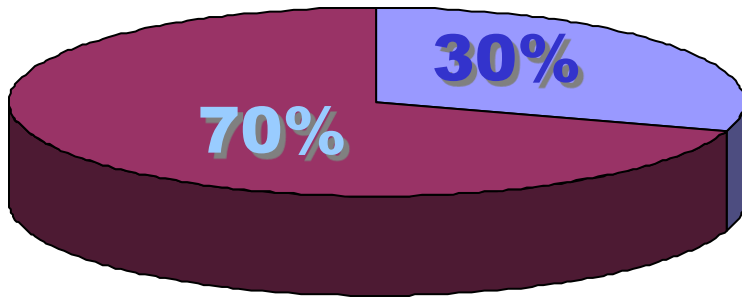
What's the approximate total data volume that your organization manages ONLY for analytics, both today and in three years?





- Users conduct analytics with ever-larger datasets.
- Small analytic datasets will become less common, as they grow into large ones.
- A third of surveyed organizations (37%) have broken 10Tb barrier.
- Very large analytic datasets will become much more common
- Soon we'll measure big data in petabytes, not terabytes.

Big Data isn't the problem it used to be.

In your organization is big data considered mostly a problem or mostly an opportunity?



-  **Problem** – because it's hard to manage from a technical viewpoint
-  **Opportunity** – because it yields detailed analytics for business advantage

Source TDWI. Survey of 325 respondents, June 2011

- Only 30% of survey respondents consider big data a problem.
 - *Oddly enough, big data was a serious problem just a few years ago.*
 - *Storage and CPUs developed greater capacity, speed, intelligence*
 - They also fell in price.
 - *New database mgt systems arrived, designed for big data analytics.*
 - Data warehouse appliances and analytic DBMSs are scalable and relatively affordable.
- The vast majority (70%) considers big data an opportunity.
 - *The recent economic recession forced deep changes in most businesses.*
 - *Big data analytics reveals change's root cause, so you can stop it or leverage it.*

Why is Real-Time a Tech Challenge?

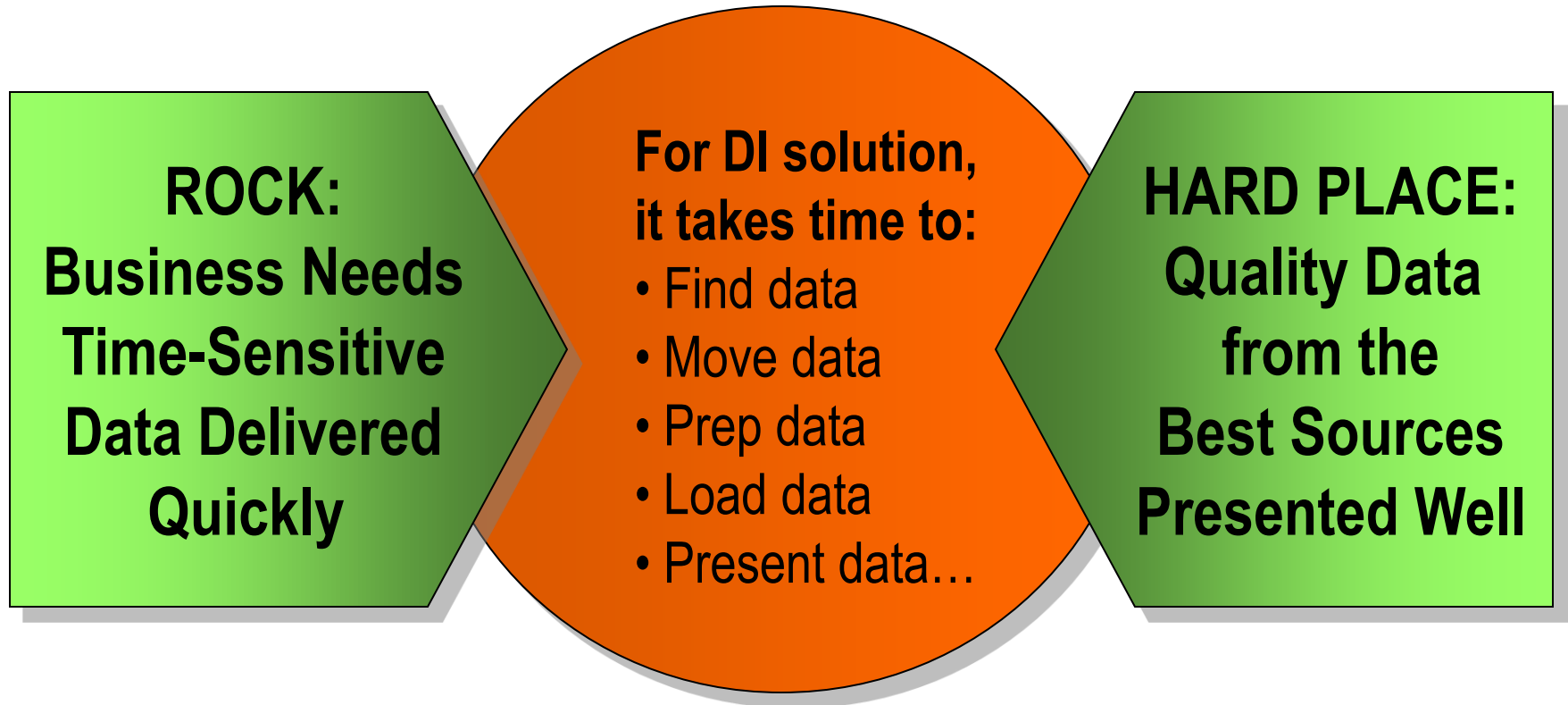
Because it takes time to...

- To move data
 - *Connect to a source system*
 - Typically DI tool connecting to operational app
 - *Runs a query or otherwise identifies data*
 - *Extracts data into a network package*
- To prepare data properly
 - *Join data from multiple sources*
 - *Data quality functions: validate, standardize, dedupe, enhance*
 - *Goal is to deliver clean, compliant, complete, contextual, auditable data*
- To load data
 - *Connect to source, run updates and inserts*
- To present data
 - *Refresh a report/dashboard, alert appropriate user*



Data Integration is caught between A Rock and a Hard Place

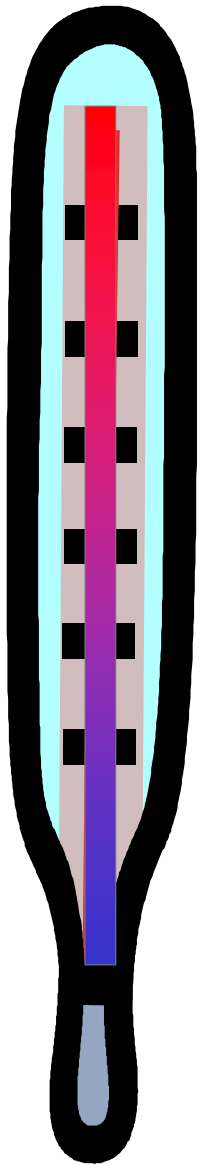
- The two requirements run the risk of cancelling each other out.
- How can you satisfy *both* requirements, without sacrifice?



A COUPLE OF EXCEPTIONS

Sometimes, Junk Data is all the Business Needs

- Fuzzy Metrics
 - *Many metrics and KPIs are approximations*
 - *If a metric is intended to be “fuzzy,” then there’s no point in improving its accuracy*
 - *Likewise, if the metric is purely tactical information, and no strategic decisions are based on it*
- Preliminary Analytic Results
 - *Much of the discovery work business analysts do is not a deliverable & not intended for broad consumption*
 - *However, after an analytic epiphany, analytic findings must be operationalized with quality data, reports, etc.*



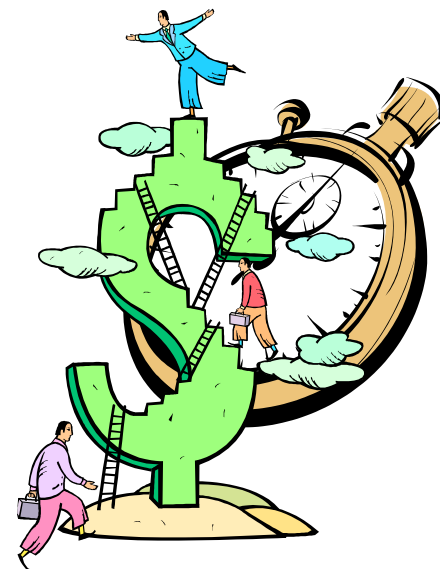
Options for Real-Time Data Management

For users, there are many high-performance options available today.

- Real-Time Data Integration – Next Gen Hot Growth Features:
 - *RT DI, RT DQ, complex event processing (CEP), RT alerts*
- Real-Time Data Quality
 - *Standardize, validate, enhance data before it hits database*
- Real-Time Master Data Management – Next Gen Hot Growth:
 - *Instantiate 360-degree customer view or publish ref data in RT*
- Data Replication – According to Next Gen DI Survey:
 - *46% of users do replication; second only to ETL!*
- Data Federation – According to Next Gen DI Survey:
 - *33% of users surveyed are now using federation. It's arrived.*
- Data Services over an Enterprise Service Bus:
 - *Reach operational application data in RT; 39% w/data services*
- In-Memory Databases
 - *With multi-Tb memory, you can now put Big Data in memory*
- In-Database Analytics
 - *Faster to analyze big data where it's stored, instead of moving it*

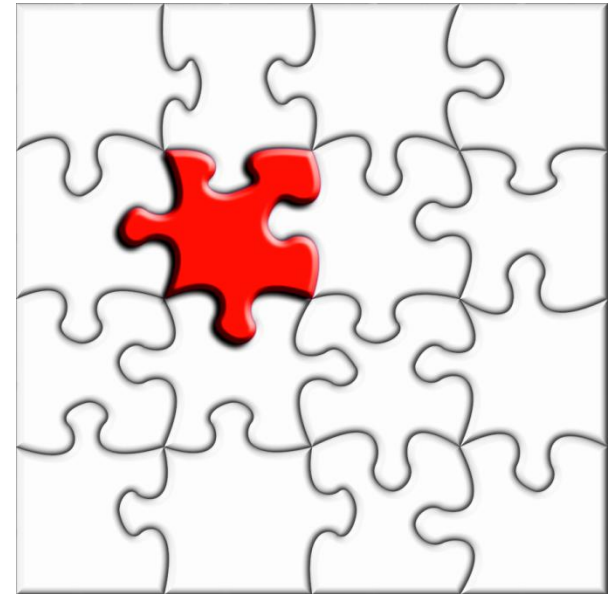
Why Care About Real-Time Data Management?

- It's a foundation for time-sensitive business practices:
 - *Operational business intelligence*
 - *Just-in-time inventory*
 - *Facility monitoring*
 - *Self-service information portals*
 - *eCommerce recommendations*
 - *Price optimization*
 - *Production yield & workforce mgt in manufacturing*
- Real-Time Analytics is coming
 - *Reporting embraced RT. Just look at Operational BI.*
 - *Analytics, too, will soon take that trip. And with Big Data.*



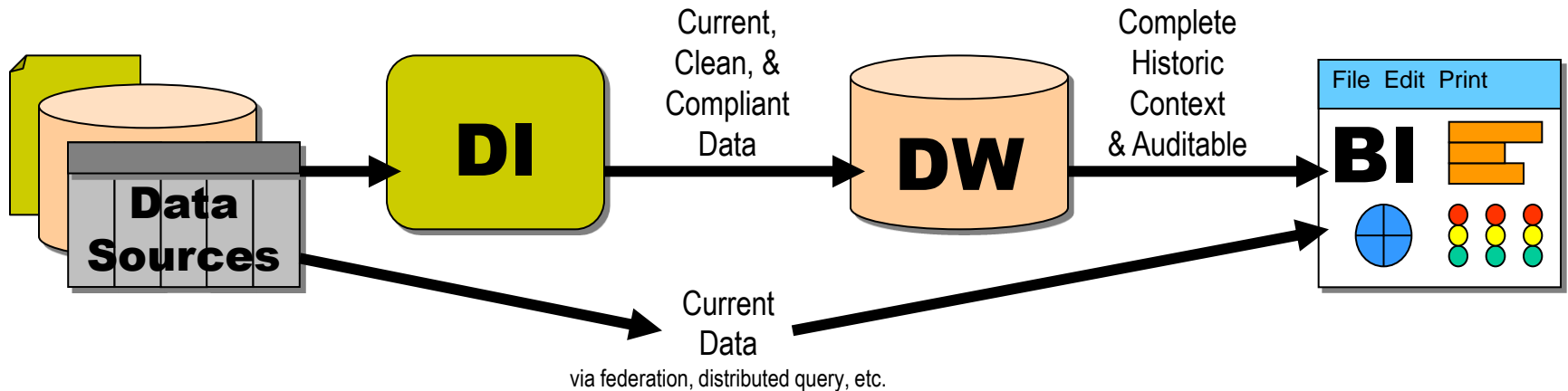
Use Cases for Data Federation

- Data Warehousing
 - *Prototype before you persist*
 - *Augment DW with current data*
 - *Federate DWs and marts*
 - *Migrate data warehouses*
 - *Augment ETL*
- Other
 - *“SWAT” applications*
 - *Operational reporting*
 - *Data services layer in SOA*
 - *Data discovery and modeling*



Source: TDWI

“Do I really need DI or DW for real-time data?”



- Yes. Include data integration in the process, because it:
 - *Transforms complex data into useful and reusable forms*
 - *Improves data's quality, metadata, master data*
 - *Comply with policies for data access, use, security, privacy*
- Yes. Include a data warehouse in the process because it:
 - *Provides a complete historic context to complement real-time data*
 - *Maintains an audit trail*
 - *Records the evolving state of real-time data*
- Omit DI and DW at your peril:
 - *It's possible to get current data by going around DI and DW.*
 - *But the data won't be clean, compliant, complete, contextual, auditable...*

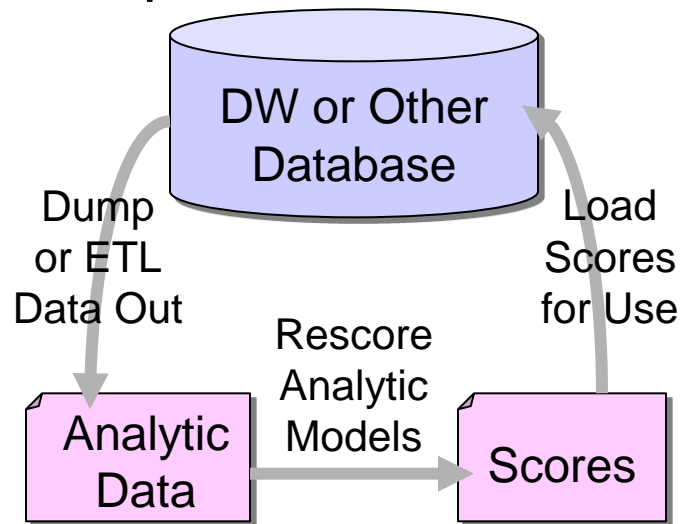
In-Memory Databases as Accelerators

- A use case in OLAP
 - *Put OLAP cube or table(s) in server memory*
 - *Update memory cache as data evolves*
 - *Users can refresh reports or dashboards from memory cache, in real time and on demand*
- A use case in predictive analytics
 - *Put analytic models and data in memory*
 - *Rescore models quickly, so in real time you can*
 - Track changes in consumer behavior
 - Test for fraud and risk
- Related Trends
 - *32-bit to 64-bit upgrades*
 - Many are driven by need for 64-bit's massive addressable memory space
 - *BI & DW Appliances*
 - Many are designed to be an in-memory DB, sometimes called an accelerator
 - *Flash drives and solid-state drives are coming for BI/DW servers*
 - Persistent storage (like disks), but with high speed (like memory)

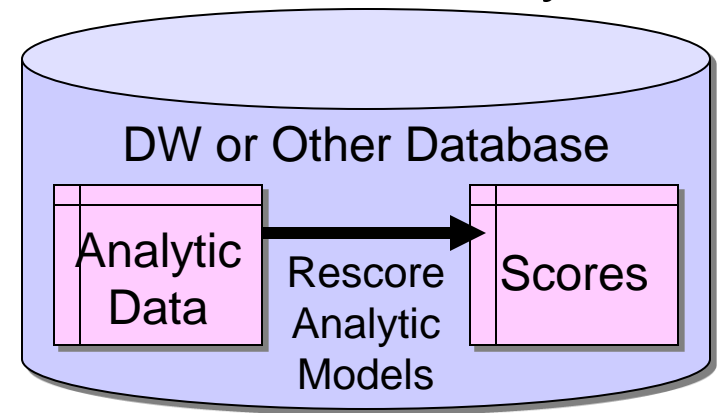


In-Database Analytics Simplifies Data Warehouse Architecture

OLD WAY Dump, Score, and Load



NEW WAY In-Database Analytics

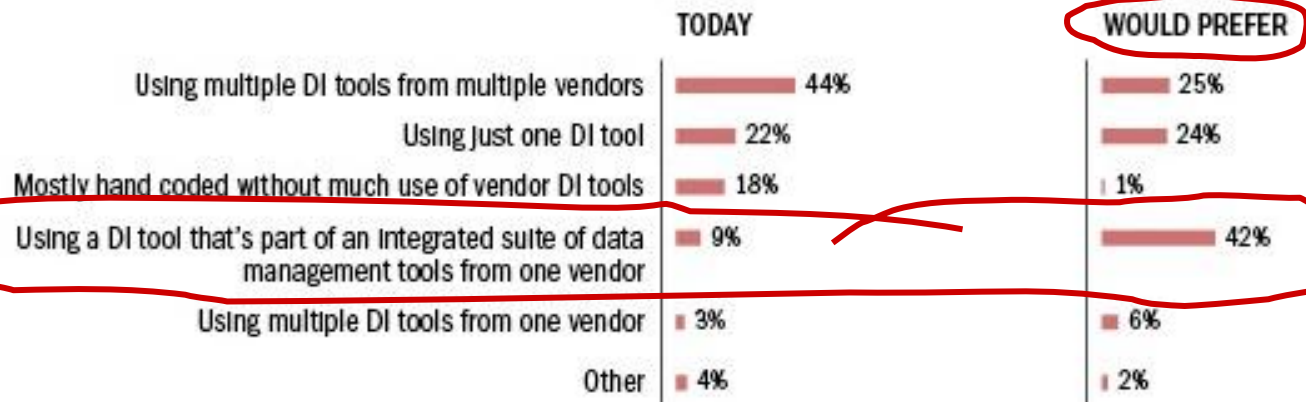


TDWI SURVEY SEZ:

Users would prefer an Integrated DM Platform

- Users with multiple DI tools from multiple vendors will drop from 44% to 25%.
- 9% report using integrated suite of tools one today, yet 42% would prefer one.
- Integrated data management platforms are the future.
 - *Greater reuse of data mgt artifacts, since they're shared across multiple tools*
 - *Enables users to build solutions incorporating multiple data mgt functions*
 - Easier to squeeze data quality functions into more solutions
 - Easy access to multiple real-time data mgt functions

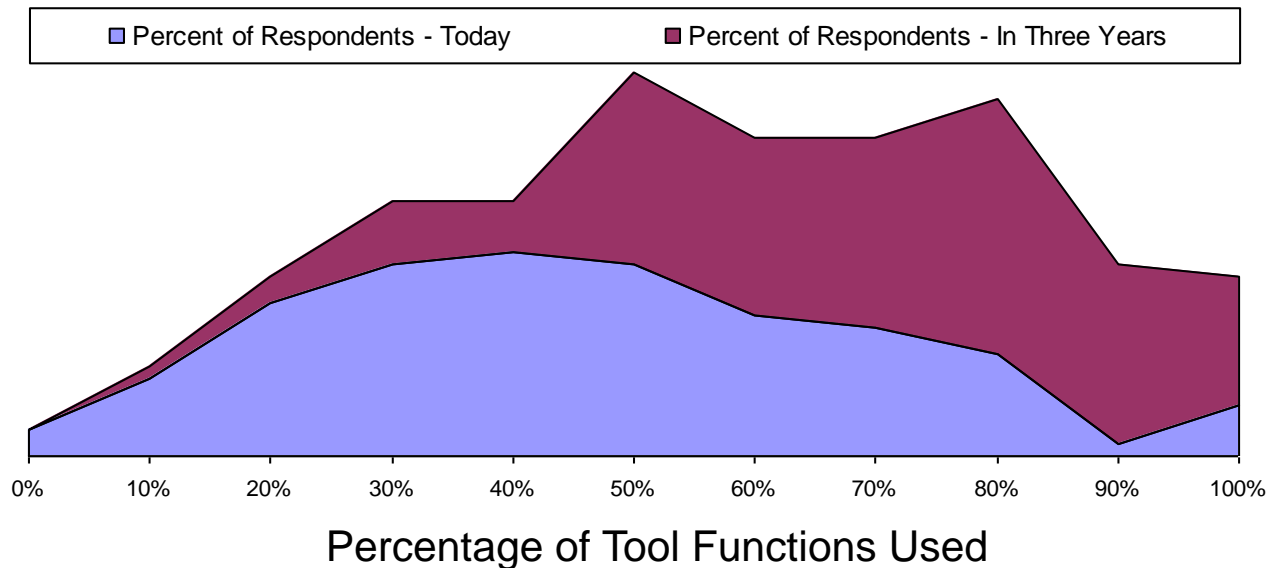
Which of the following best describes your organization's portfolio of DI tools today? For your organization's next generation DI implementation, how would you prefer that the DI portfolio be?



TDWI SURVEY SEZ:

Users will employ more Real-Time DI functions

- Users employ 40% of their DI tool's functions, on average.
- This will jump to 65% of DI tool functions.
- **Real-time DI** functions will see the most growth.



“What approximate percentage of your primary DI tool's functions are you using?”

Miscellaneous Observations & Caveats



- Not all data needs to travel in real time or be analyzed
 - *Work with biz; provide only what they need for operational decisions*
- If you over-prepare analytic data, you can lose nuggets
 - *Some adv'd analytic apps depend on raw, detailed source data, as in Big Data*
 - *Un-processed source takes much more time to process than aggregated data*
- You really do need to mix RT operational data with DW data
 - *RT Op data represents the moment, DW data the historic context*
 - *Without both, you don't have the complete picture*

Recommendations

- Be aware of the Mega Trends for data management
 - *Plan accordingly, esp for real time, big data, interoperability*
- Embrace Big Data
 - *It's not the problem it used to be*
- Expand your use of data management tools & functions
 - *Embrace real-time interfaces, services, buses*
- Rely on key technology enablers
 - *Real-time DI, DQ, MDM, events*
 - *Federation, replication, in-memory DBs*
- Give the business what it needs
 - *Faster, broader data for operational decisions*
 - *In future, more real-time analytics, too*

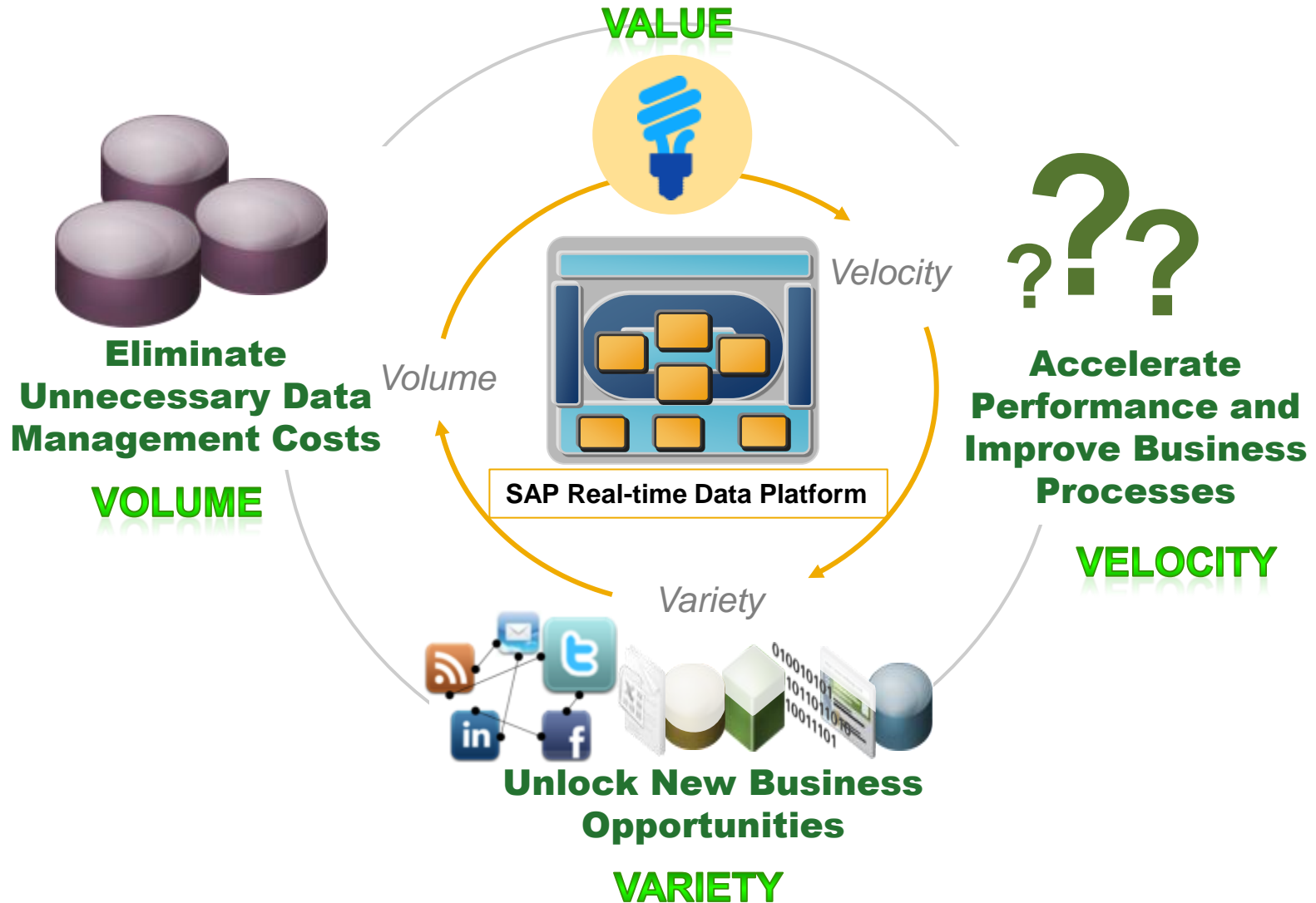


Managing Big Data to Gain Insight Never Seen Before

Marie Goodell

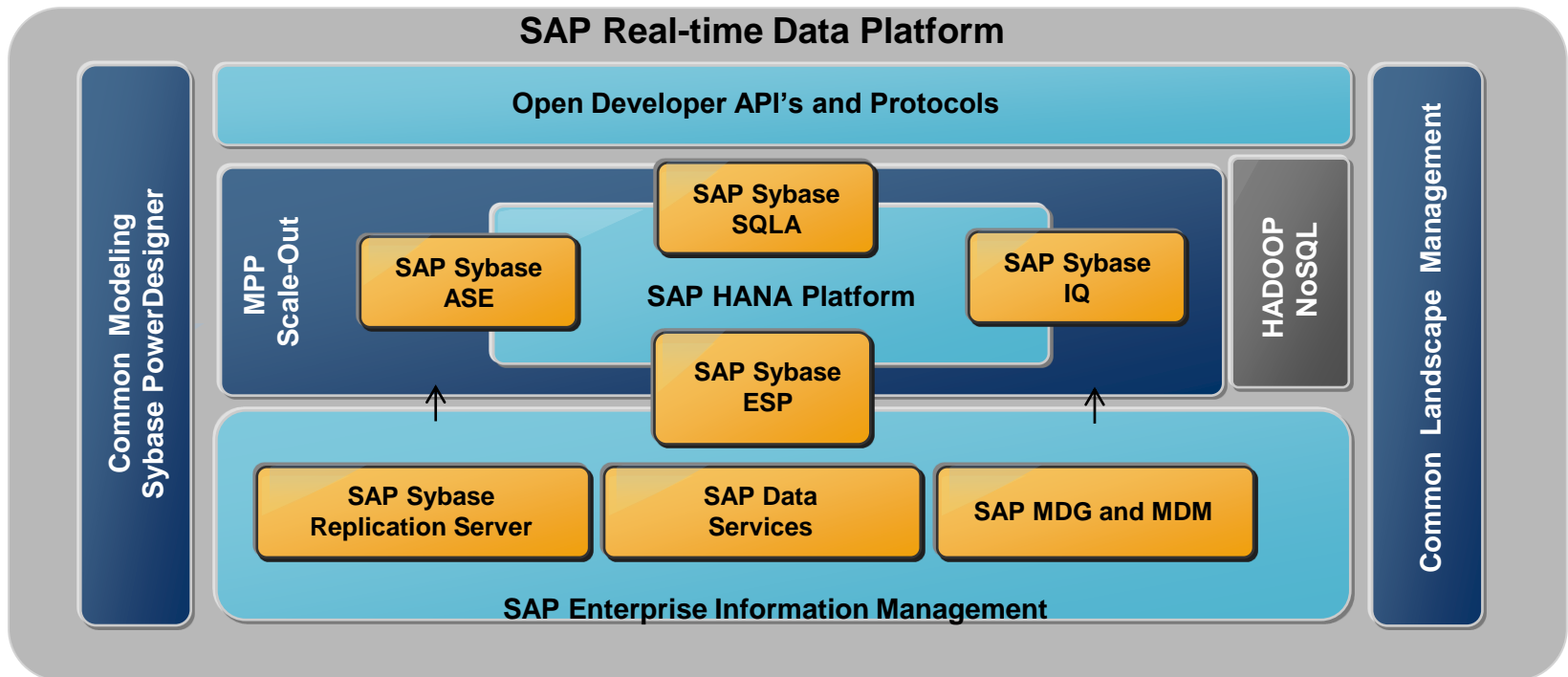
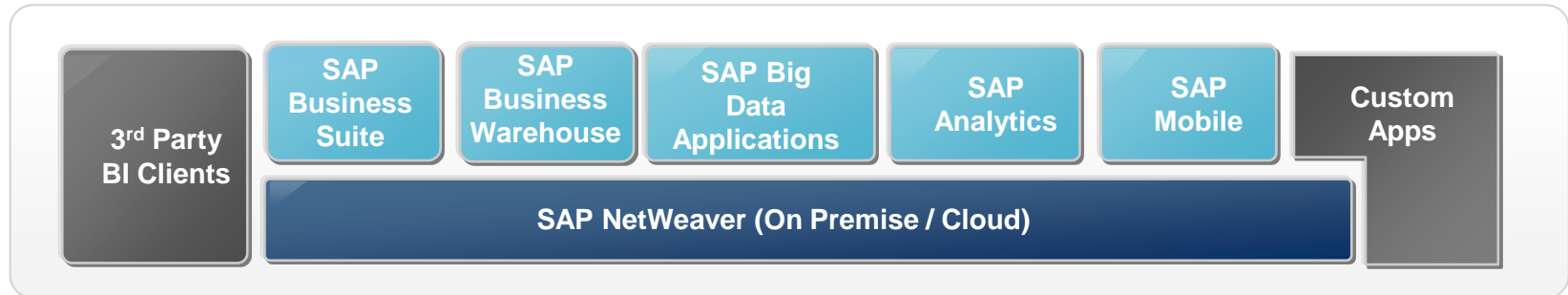
How To Manage Big Data Effectively?

...With a real-time data platform



SAP Real-Time Data Platform

A flexible deployment environment for managing big data



SAP Data Management Portfolio

Flexible Storage & Processing Options for Your Business Needs



Embedded

Mobile and Embedded Database

Sybase SQL Anywhere

- Leading mobile and embedded database
- True zero administration; Best store and forward synchronization
- Optimized for extended enterprise, ISVs for cloud and device based apps



Transactions

Transactional Database, Best TCO

Sybase ASE

- Database for extreme transaction applications with best TCO
- High reliability and low maintenance
- Optimized for Purpose Built Apps, Business All-in-One, SAP ERP



Open EDW

Open Analytical Database, Best TCO

Sybase IQ

- Smart "Big Data" Store with best TCO
- High performance, low price / performance
- Optimized for: Open Enterprise Data Warehouse

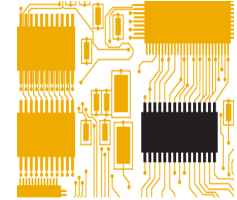


Continuous Intelligence

Insight into Streaming Data

Sybase ESP

- #1 Complex Event Processing (CEP) platform
- Faster, simpler way to analyze and act on event streams;
- Optimized to support real-time data warehouse; sense & response application



Real-time Business

Next-Gen Platform for Real-time Business

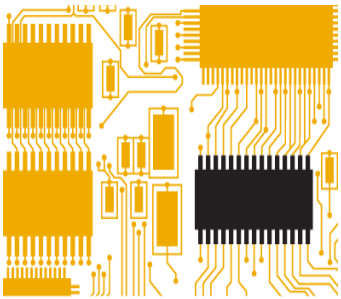
SAP HANA

- Transactional, analytical and application logic processing in ONE DBMS
- Best real-time data mgmt – eliminate redundant layers, latency, modeling (OLAP on OLTP model)
- Optimized for BW on HANA, HANA as data mart, HANA apps

SAP HANA Data Platform

Ideal for real-time business

1. Revolutionary in-memory platform



- Real-time analytics on detailed data on the fly
- In-memory calculations
- Real-time replication to eliminate data latency
- No aggregates, tuning of data for

2. Empowers you to interrogate data



- Wizard-driven data modeling for business
- Fast & easy creation of ad-hoc views
- Optimized for SAP BusinessObjects BI
- Open platform for other clients

3. Powerful predictive analytics



- Embedded data mining algorithms for predictive analytics
- Bring decision support capabilities to the business users through simplified experience and pre-built scenarios

4. High data quality



- Real-time replication
- Faster integration with optimized ETL
- Tightly connected with EIM data quality capabilities

SAP EIM Portfolio for Big Data

Ensure complete and accurate information



**Collaborative
IT / business**

**Design your
information
architecture**

**Sybase
PowerDesigner**

- Data modeling
- Information modeling
- Enterprise architecture modeling
- Common metadata repository

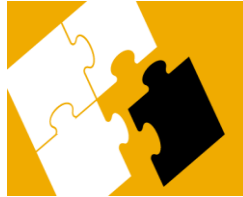


Available

**Replicate data
rapidly**

**Sybase
Replication
Server**

- Move & synchronize data
- Disaster recovery
- High availability
- Load balancing
- Sybase, Oracle, IBM, Microsoft



Complete

**Integrate &
process data from
a variety of stores**

**SAP
Data Services**

- Rapid ETL
- Read & load Hadoop data
- Process text data to unlock meaning from unstructured documents



Accurate

**Cleanse &
enhance data**

**SAP Data Services
–Data Quality Mgmt**

- Assess, correct, complete data
- Validate & enrich data
- Quality coverage for 230 countries
- Geo-location / address cleansing



Single view

**Ensure master
data quality**

**SAP Enterprise
Master Data Mgmt**

- Manage master data created in business apps
- Manage master data across distributed systems
- Product, material, supplier, customer
- Global data sync



Trustworthy

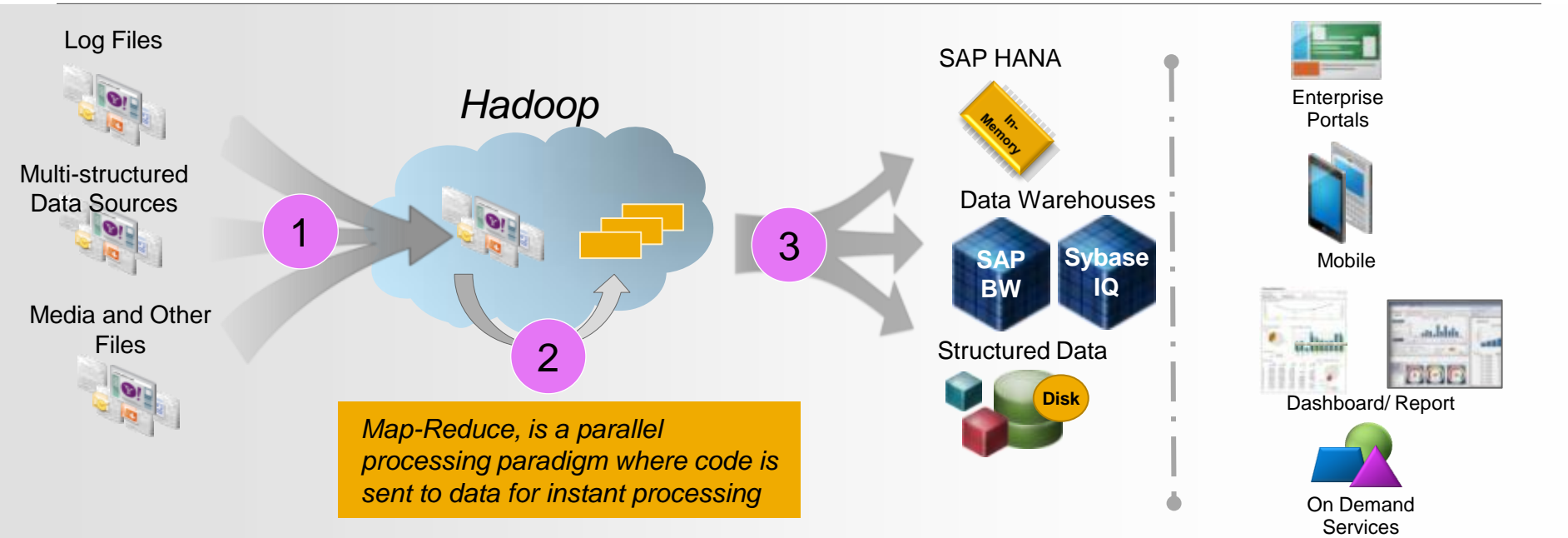
**Profile and
measure data
quality**

**SAP Information
Steward**

- Graphical UI for business users
- Data quality scorecards
- Data lineage
- Change management impact analysis
- Common metadata repository

SAP Data Services – Hadoop Connector

Connect unstructured & structured data for greater insight



1 Collect & Store

- Files are stored in their native format incurring no transformational costs.
- Built in fault-tolerance.
- Commodity hardware and software solution makes *Hadoop* scale cost effectively.

2 Analyze & Process

- Prepare the data to enable the class of problems to be solved.
- Problems like searching, counting, pattern detection lend themselves well to Map Reduce paradigm

3 Integrate & Consume

- Flexibility and change management ; security and access control
- Administrative and operational skill set for business continuity
- Enterprise support and Integration into enterprise infrastructure

SAP Data Services – Text Data Processing

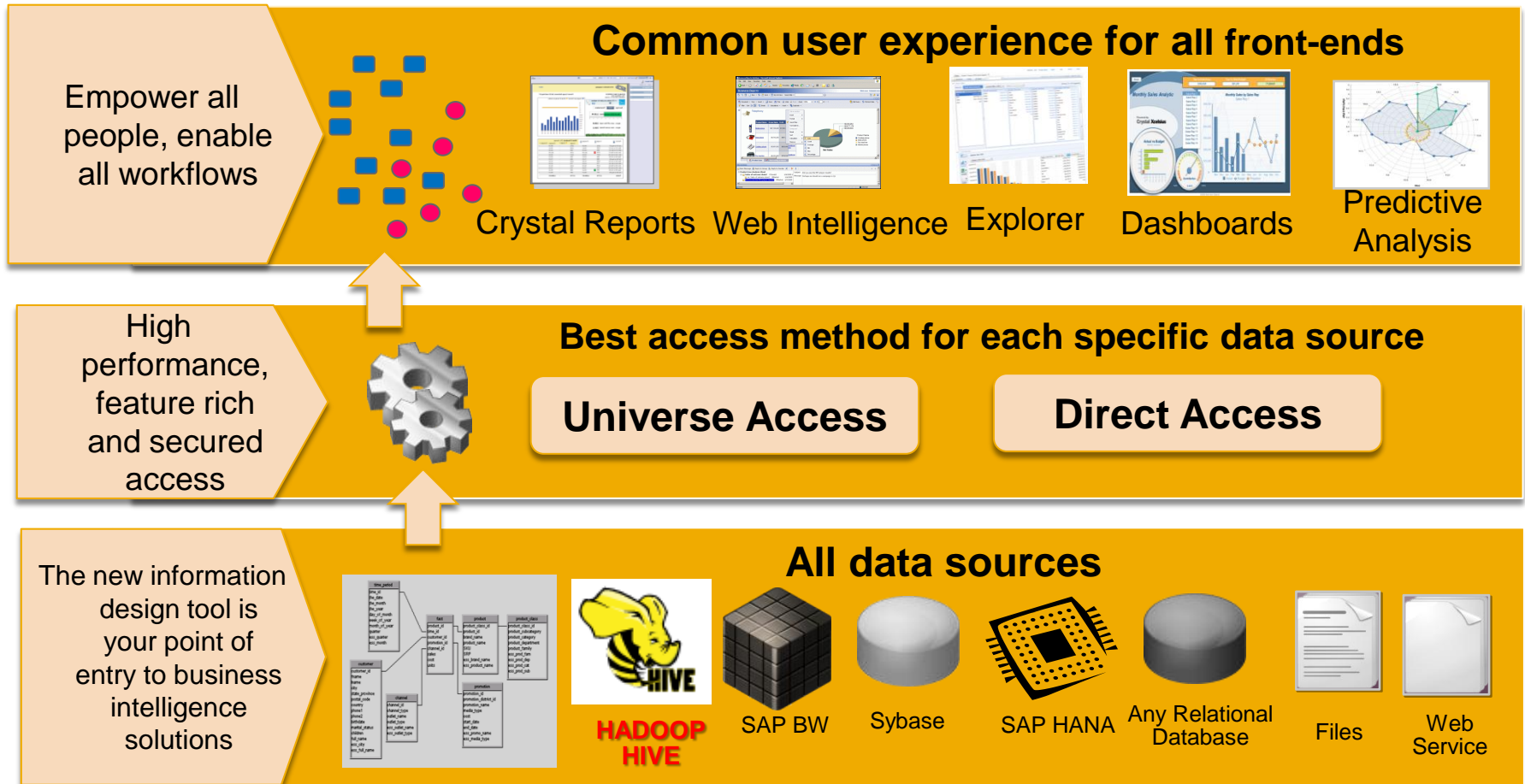
Process text data for valuable insight

Steven Paul Jobs (born February 24, 1955) is an American business magnate and inventor. He is well known for being the co-founder and chief executive officer of Apple. Jobs also previously served as chief executive of Pixar Animation Studios; he became a member of the board of The Walt Disney Company in 2006, following the acquisition of Pixar by Disney.

Steven Paul Jobs; Jobs	PERSON
February 24, 1955; 2006	DATE
co-founder; chief executive officer; chief executive; member of the board	TITLE
Apple; Pixar Animation Studios; Pixar; The Walt Disney Company; Disney	ORGANIZATION_COMMERCIAL
American business magnate; inventor; acquisition	NOUN_GROUP or Concept
Jobs also previously served as chief executive of Pixar...	Executive Job Change
Acquisition of Pixar by Disney	Merger and Acquisition

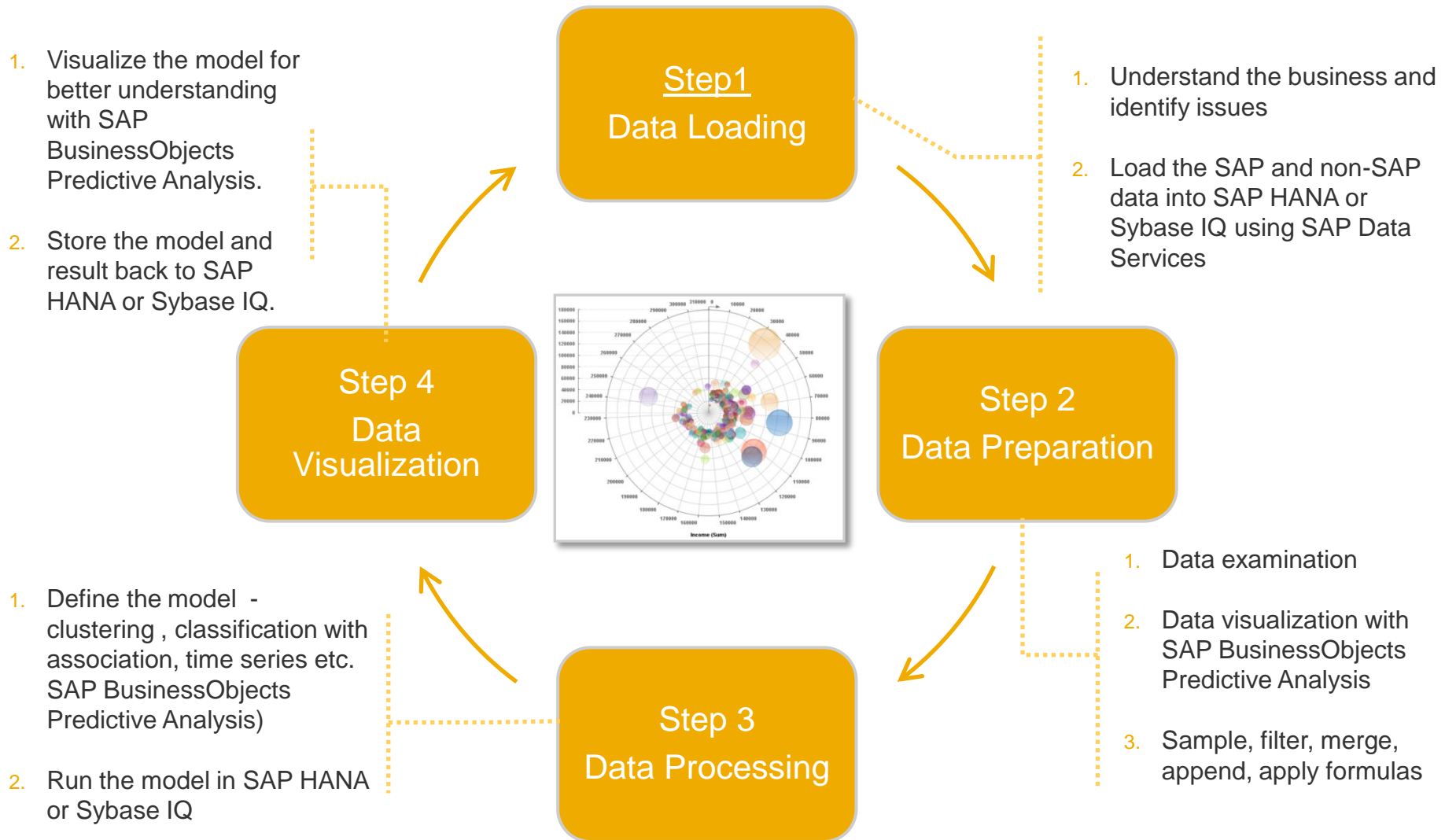
SAP BusinessObjects Business Intelligence platform

Access, analyze and share information from big data



SAP BusinessObjects Predictive Analysis

Unlock the power of big data

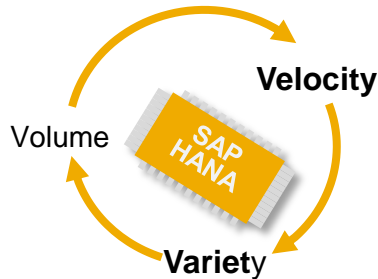


Flexible Real-time Data Platform

Based on business needs, speed, and cost

SAP HANA for Real-time Business

- Built for extreme transactional, analytical and application logic processing in ONE database
- Best real-time data management - eliminate redundant layers, latency and modeling (OLAP on OLTP model)

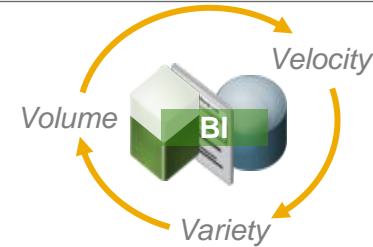


Sybase ESP for Streaming Events

- CEP for monitoring & processing streaming events in real time
- Best for immediate action

Sybase ASE for Extreme Transactional Database

- Built for transaction applications
- Simplest maintenance
- Best TCO for SAP application + database

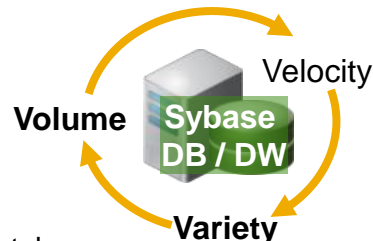


*Business
Analytic*

*In-Memory
Innovation*

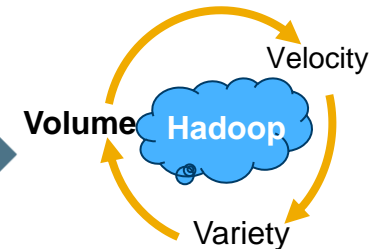
*Cost-effective
Distributed FS*

*Transactional
Management*



Apache Hadoop for Distributed file system based storage and processing:

- Open source
- Emerging storage and processing model for unstructured data



Sybase IQ for Open Analytical Database

- Built for data warehousing
- Operate with any technologies (e.g. HW, Hadoop)
- Best cost/performance in market



THANK YOU

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG. The information contained herein may be changed without prior notice.

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

Microsoft, Windows, Excel, Outlook, and PowerPoint are registered trademarks of Microsoft Corporation.

IBM, DB2, DB2 Universal Database, System i, System i5, System p, System p5, System x, System z, System z10, System z9, z10, z9, iSeries, pSeries, xSeries, zSeries, eServer, z/VM, z/OS, i5/OS, S/390, OS/390, OS/400, AS/400, S/390 Parallel Enterprise Server, PowerVM, Power Architecture, POWER6+, POWER6, POWER5+, POWER5, POWER, OpenPower, PowerPC, BatchPipes, BladeCenter, System Storage, GPFS, HACMP, RETAIN, DB2 Connect, RACF, Redbooks, OS/2, Parallel Sysplex, MVS/ESA, AIX, Intelligent Miner, WebSphere, Netfinity, Tivoli and Informix are trademarks or registered trademarks of IBM Corporation.

Linux is the registered trademark of Linus Torvalds in the U.S. and other countries.

Adobe, the Adobe logo, Acrobat, PostScript, and Reader are either trademarks or registered trademarks of Adobe Systems Incorporated in the United States and/or other countries.

Oracle is a registered trademark of Oracle Corporation.

UNIX, X/Open, OSF/1, and Motif are registered trademarks of the Open Group.

Citrix, ICA, Program Neighborhood, MetaFrame, WinFrame, VideoFrame, and MultiWin are trademarks or registered trademarks of Citrix Systems, Inc.

HTML, XML, XHTML and W3C are trademarks or registered trademarks of W3C®, World Wide Web Consortium, Massachusetts Institute of Technology.

Java is a registered trademark of Sun Microsystems, Inc.

JavaScript is a registered trademark of Sun Microsystems, Inc., used under license for technology invented and implemented by Netscape.

SAP, R/3, SAP NetWeaver, Duet, PartnerEdge, ByDesign, SAP BusinessObjects Explorer, StreamWork, and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and other countries.

Business Objects and the Business Objects logo, BusinessObjects, Crystal Reports, Crystal Decisions, Web Intelligence, Xcelsius, and other Business Objects products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Business Objects Software Ltd. Business Objects is an SAP company.

Sybase and Adaptive Server, iAnywhere, Sybase 365, SQL Anywhere, and other Sybase products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of Sybase, Inc. Sybase is an SAP company.

All other product and service names mentioned are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

The information in this document is proprietary to SAP. No part of this document may be reproduced, copied, or transmitted in any form or for any purpose without the express prior written permission of SAP AG.

This document is a preliminary version and not subject to your license agreement or any other agreement with SAP. This document contains only intended strategies, developments, and functionalities of the SAP® product and is not intended to be binding upon SAP to any particular course of business, product strategy, and/or development. Please note that this document is subject to change and may be changed by SAP at any time without notice.

SAP assumes no responsibility for errors or omissions in this document. SAP does not warrant the accuracy or completeness of the information, text, graphics, links, or other items contained within this material. This document is provided without a warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or non-infringement.

SAP shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials. This limitation shall not apply in cases of intent or gross negligence.

The statutory liability for personal injury and defective products is not affected. SAP has no control over the information that you may access through the use of hot links contained in these materials and does not endorse your use of third-party Web pages nor provide any warranty whatsoever relating to third-party Web pages.

Questions?



Contact Information

If you have further questions or comments:

Philip Russom, TDWI
prussom@tdwi.org

Marie Goodell, SAP
marie.goodell@sap.com