Evolving toward the Modern Data Warehouse: What Enterprises Must Have Today, and What They’ll Need in the Future

Philip Russom
Senior Research Director for Data Management, TDWI
November 10, 2016
Sponsor

Microsoft
Speakers

Philip Russom
TDWI Research Director, Data Management

Oliver Chiu
Sr. Product Marketing Manager, Big Data and Data Warehousing, Microsoft
Today’s Agenda

• Background on DW Modernization
  – What is it? Why do it?

• Tips for the Modern Data Warehouse
  – Integrate traditional data and big data
  – Tap multi-structured data for new insights.
  – Do advanced analytics for the biz value.
  – Accelerate the biz into near real time.
  – Develop a unified DW architecture.
  – Modernize w/cloud-based tools & platforms.
  – Demand high performance and scalability of all components of a data warehouse architecture.

• Summary and Recommendations
WHAT is Data Warehouse Modernization?

- A wide range of changes in/around DWs
  - To support new tech & biz requirements
  - To embrace big data & analytics
  - To apply updates & upgrades
- Synonyms for “data warehouse modernization”
  - DW augmentation, automation, optimization
- Scenarios for DW Modernization range widely:
  - From simple addition of new subjects areas, sources…
  - To dramatic cases, like rip-and-replace of DW platform
WHY do we do Data Warehouse Modernization?

• So that IT & Data Mgt teams can support new biz goals per data
• When a DW needs to broaden
  – Greater scale, speed, capacity, room for growth
• When DW design becomes outmoded
  – To improve dimensional models, overall architecture
• When DW gets new technical requirements
  – To support real time, virtualization, unstruc’d data
  – More forms of advanced analytics
• When the DW’s platform has outlived its usefulness
  – Rip-and-replace: migrate data from old to new platform
Integrate and leverage traditional enterprise data and new big data.

- The goal of managing big data is to provide biz with analytic insights.
- Leverage big data, don’t just manage it as a cost center.
- Big data comes in many forms, both old and new, traditional and modern.
- Blending traditional enterprise data and modern big data enables:
  - Broader exploratory analytics
  - More complex analytic correlations across events & entities
- ENABLING TECHNOLOGIES for persisting old/new data
  - Relational database management systems (RDBMSs)
  - Big data tools, like Hadoop, Spark, and emerging data lake platforms, including those in the cloud
Make sense of multi-structured data for new and unique business insights.

• Multi-structured types of big data have compelling value
  – Human language text drawn from your website, call center app, and social media fuels sentiment analysis
  – Supply chain documents (XML, JSON, RFID, or CSV formats) show which partners have best reliability, cost, and quality

• ENABLING TECHNOLOGIES for multi-structured data
  – Hadoop and other big data or data lake platforms can capture/manage unstructured data
  – A data lake enables analytics with multi-structured data
  – Natural language processing (NLP) is key to getting biz value from multi-structured data
Implement advanced forms of analytics to enable discovery analytics for big data.

- Biz managers assume that the best route to biz value from big data is through advanced analytics
- Most applications of analytics enable discovery
- Big data can enable new analytic applications
- Big data can extend older analytic applications
- ENABLING TECHNOLOGIES for adv’d analytics
  - RDBMS optimized for SQL-based analytics
  - RDBMS integrated with a big data platform
  - Big data platforms for a wide range of advanced analytics
  - Self service for end users
Empower the business to operate in near real time by delivering data faster.

- **Real time** literally means data is fetched, processed, and delivered at sub-second speed after data is created or altered.
  - However, most real-time data operations take minutes or hours; more aptly called *near real time*.

- Near real time DW for time-sensitive biz processes
  - Operational BI, performance mgt, dashboards…
  - Newer ones – biz monitoring, stream analytics, ecommerce…

- **ENABLING TECHNOLOGIES** for near real time
  - SQL queries that return results in seconds
  - Early ingestion into data lake or big data repository
  - Columnar data stores; in-memory databases
Integrate multiple platforms into a unified DW architecture.

- Diverse big data is subject to diverse processing
  - Which may require multiple platforms for data persistence
  - As they modernize, users diversify their tool/platform portfolio
- Trend: toward data warehouse environment (DWE)
  - Multiple data platforms; mixed platform types; distributed arch.
  - Multiple RDBMS, plus Hadoop; clouds, both on-prem & off
- ENABLING TECHNOLOGIES for unified DW arch
  - Unification for the modern multi-platform DW environment
  - Pre-integration, optimization for components of multi-platform DW
  - Integrating a big data or data lake platform with an RDBMS
AN OVERVIEW OF
The Modern Data Warehouse Architecture

CROSS-SYSTEM INTEGRATION: Data Integration, Interfaces, Metadata; Queries and Views; Unified Architecture

New Data
- Machine Data, Mobile Data, Web Data, Social Media...

Traditional Data
- CRM, SFA, ERP
- Financials, billing, operations, call center, supply chain mgmt...

Many Ingestion Methods
- Data Landing
- Data Staging
- ETL/ELT
- Stream Capture
- Event Processing

Ingestion

Hadoop
- Massive Store of Raw Source Data (Data Lake)
- Set-based Analytics
- Algorithmic Analytics
- Analytic Archive

Many Delivery Methods

Core Warehouse
- Dimensions, cubes, subject areas, time series, metrics...
- Data for reports, dashboards...

Specialized DBMSs
- Based on clouds, appliances, columns, graph, sandboxes, and other specialized analytics

DIVERSE PLATFORMS: Web, Client/Server, Clusters, Racks, Grids, Clouds, Hybrid Combinations
Include cloud-based platforms, as you diversify your modern DW environment.

- Clouds are emerging as important platforms and architectural components for modern DW environments.
  - Cloud-based DBMSs, DWs; tools for analytics or data integration
- Include cloud-based functionality in hybrid architectures.
  - Mix of cloud types (public & private), both on-premises & off
- Think of the cloud as a data management strategy.
  - Speed and scale strategy, due to fluid allocation of resources
  - Modernization strategy that quickly addresses new requirements
- ENABLING TECHNOLOGIES for cloud-based DW
  - For scalability on demand, consider an elastic cloud.
  - Demand DW platforms optimized specifically for clouds
Demand high performance and scalability of all components of a DW.

- There’s a need for speed.
- Big data keeps getting bigger.
- The modern DW is highly available.
- ENABLING TECHNOLOGIES for speed & scale
  - Massively parallel processing (MPP) offers scale-out capabilities that symmetrical multiprocessing (SMP) does not.
  - Hadoop and Spark have a strong track record for scaling to massive data volumes
  - Look for platforms that need little or no query optimization or data remodeling
SUMMARY & CONCLUSIONS
Evolving toward the Modern Data Warehouse

1. Deploy a modern data warehouse to integrate and leverage traditional enterprise data and new big data.
3. Implement advanced forms of analytics to enable discovery analytics for big data.
4. Empower the business to operate in near real time by delivering data faster.
5. Integrate multiple platforms into a unified data warehouse architecture.
6. Include cloud-based platforms, as you diversify your modern data warehouse environment.
7. Demand high performance and scalability of all components of a data warehouse.
The report discusses the challenges and solutions for data warehousing in an age of evolving data and business requirements.

Download the free report in a PDF file at: www.tdwi.org
Microsoft’s Vision: Transform data into intelligent action

- All types of data
- Across on-premises and cloud
- Built-in predictive intelligence
- Mission critical

Tailored experiences for multiple audiences

SQL Server 2016 + Cortana Intelligence Suite
How Microsoft’s Modern Data Warehouse aligns with TDWI’s MDW Architecture

New Data
- Machine Data, Mobile Data, Web Data, Social Media...

Traditional Data
- CRM, SFA, ERP
- Financials, billing, operations, call center, supply chain mgt...

Ingestion
- Many Ingestion Methods
- Azure Data Factory
- Azure Data Catalog
- Azure Event Hubs
- SQL Server Master Data Services
- SQL Server Integration Services
- SQL Server Data Quality Services

Hadoop
- Many Delivery Methods
- Azure HDInsight (Hadoop and Spark)
- Azure Data Lake
- Azure Marketplace (Hortonworks, Cloudera, MapR)
- Integration to industry Hadoop distributions (Hortonworks, Cloudera)

Core Warehouse
- Many Delivery Methods
- Azure SQL Data Warehouse
- SQL Server Fast Track Analytics Platform System

Specialized DBMSs
- Azure DocumentDB
- Azure Market Place (Mongo, Cassandra, Couchbase, Riak, etc.)

DIVERSE PLAT FORMS: Web, Client/Server, Clusters, Racks, Grids, Clouds, Hybrid Combinations
SQL Server 2016: Everything built-in

The above graphics were published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from Microsoft. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner’s research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

Consistent experience from on-premises to cloud

TPC-H 10TB non-clustered results as of 04/06/15, 5/04/15, 4/15/14 and 11/25/13, respectively. http://www.tpc.org/tpch/results/tpch_perf_results.asp?resulttype=nocluster

National Institute of Standards and Technology Comprehensive Vulnerability Database update 10/2015
Introducing Cortana Intelligence Suite

Data Sources
- Data Factory
- Data Catalog
- Event Hubs

Apps
- Data
- Sensors and devices

Big Data Stores
- Data Lake Store
- SQL Data Warehouse

Machine Learning and Analytics
- Machine Learning
- Data Lake Analytics
- HDInsight (Hadoop and Spark)
- Stream Analytics

Intelligence
- Cognitive Services
- Bot Framework
- Cortana
- Power BI

Dashboards & Visualizations
- Data Lake
- Analytics

Data → Intelligence → Action

People

Web
Mobile
Bots
Automated Systems
Azure SQL Data Warehouse

Delivers true promise of cloud elasticity to data warehousing

- Separation of compute & storage, **scale compute in seconds**
- Save big with **pause/resume** for batch-based workloads
- Unmatched innovations in security—**Threat Detection**
- **PolyBase** for T-SQL query across Hadoop clusters
- MPP scale to petabytes, geo-redundant w/ **99.9% SLA**
Azure HDInsight
Cloud Big Data: Managed Hadoop-as-a-service

Deploy open source Hadoop as a managed service by Microsoft

- Reliable with an industry leading SLA
- Enterprise-grade security and monitoring
- Productive platform for developers and scientists
- Cost effective cloud scale
- Integration with leading ISV applications
- Easy for administrators to manage
- 63% lower TCO than deploy your own Hadoop on-premises*

*IDC study “The Business Value and TCO Advantage of Apache Hadoop in the Cloud with Microsoft Azure HDInsight”
Azure Data Lake Analytics
An Analytics Job Service to power intelligent action

Process the widest variety of data

Write familiar code (C# + SQL) for single machine

Start in seconds, scale instantly, pay per job

Develop massively parallel programs with simplicity

Debug and optimize your big data programs with ease

Virtualize your analytics

Enterprise-grade security, auditing and support

Drag and drop how much parallelization units to scale out to any amount of data
Call to action

- Download the TDWI Whitepaper
Questions?
Contact Information

If you have further questions or comments:

Philip Russom, TDWI
prussom@tdwi.org

Oliver Chiu, Microsoft
ochiu@microsoft.com