Operational Data Store & Corporate Information Factory

with bonus comparison of the Ralph Kimball approach to data warehousing

Research Project 2: BUSINESS ARCHITECTURES

Team: The Undecided

Are we having fun yet...?

The Undecided...

Team Leaders
Bill Seling, Frankie Brooks backup

Research Coordinators
Frankie Brooks, James Murray backup

Presentation Coordinators
Paul Darlington, Bill Seling backup

Document Coordinators
James Murray, Paul Darlington backup

© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com

"Had I known Hell was going to be exactly like work, I probably wouldn't have spent as much time there."

I suppose you're wondering why I called you all here today…

Our presentation is on Operational Data Stores (ODS), the Corporate Information Factory (CIF), Bill Inmon, Claudia Imhoff, and what in the world they have to do with Data Warehouses and Data Marts, and how they apply to Ralph Kimball's approach…

Let's get some of the jargon out of the way…
(and more definitions than last week)

Operational Data Store (ODS)

- Real-time availability of analytical data instead of having to deal with delayed access[1]

- A hybrid structure designed to support both operational transaction processing and analytical processing.[2]

- A subject-oriented, integrated, current-valued, volatile collection of detailed data used to support the up-to-the-second collective tactical decision-making process for the enterprise[3]

More jargon...

Corporate Information Factory (CIF)

- First introduced by W. H. Inmon in the early 1980s, the **corporate information factory (CIF)** is the physical embodiment of the notion of an information ecosystem.[3]

- The data warehouse evolved into a full-blown architecture sometimes called the **corporate information factory**[4]

- It started out as Data Warehouse, then it morphed into something called **The Corporate Information Factory** [CIF], and then it morphed from into something called DW 2.0[5]

Let's Meet the Players…

The main cast for Operational Data Store & Corporate Information Factory includes…

William H. Inmon (I call him "Bill")
Father of the Data Warehouse – even though Kimball is older ;-)

Claudia Imhoff

Co-creator of the modern Corporate Information Factory (with Inmon)

Ralph Kimball
Representing the Data Warehouse and Data Mart team (that *other* team)

For our side: William H. Inmon

- Born July 20, 1945 in San Diego, California. B.Sc. in Mathematics from Yale University, and M.S. in Computer Science from New Mexico State University.[7]

- Recognized by Computerworld in 2007 as one of *10 IT People Who Mattered in the Past 40 Years*, not only as Father of the Data Warehouse, but father of the $28 billion industry, and author of more than 650 articles and 46 books (and that was seven years ago).[6]

- Author of *DW 2.0*, a book and architecture: Everything from CIF and now including textual information in our data warehouse, unstructured data (contracts, email, tweets, etc.)[5]

For our side: Claudia Imhoff

- Keeps her birthdate a secret. BS in Education from Duke University; MS from University of Colorado Boulder; and PhD in General Biomedical Sciences from University of Tennessee[9]

- Co-authored 5 books and writes monthly columns (more than 60) for technical and business magazines[10]

- Credited with co-creating the current iteration of the **Corporate Information Factory** architecture with Bill Inmon[9]

For their side: Ralph Kimball[11]

- Born in 1944. PhD from Stanford

- Co-creator of the Xerox Star Workstation, the world's first commercially viable GUI application

- Founder and CEO of Red Brick Systems. They created an extremely fast RDBMS targeted for data warehousing (acquired by Informix and now owned by IBM)

- Authored *The Data Warehouse Lifecycle Toolkit*, introducing the Dimensional Data Model

First, some history...[7]

- In the beginning...There were punch cards, paper tape, and core memory that was wired by hand.

- Then magnetic tape (read sequentially), but the tape could wrinkle, break, or the oxide could fall off or be scratched off and make the tape useless 1200-3600 ft

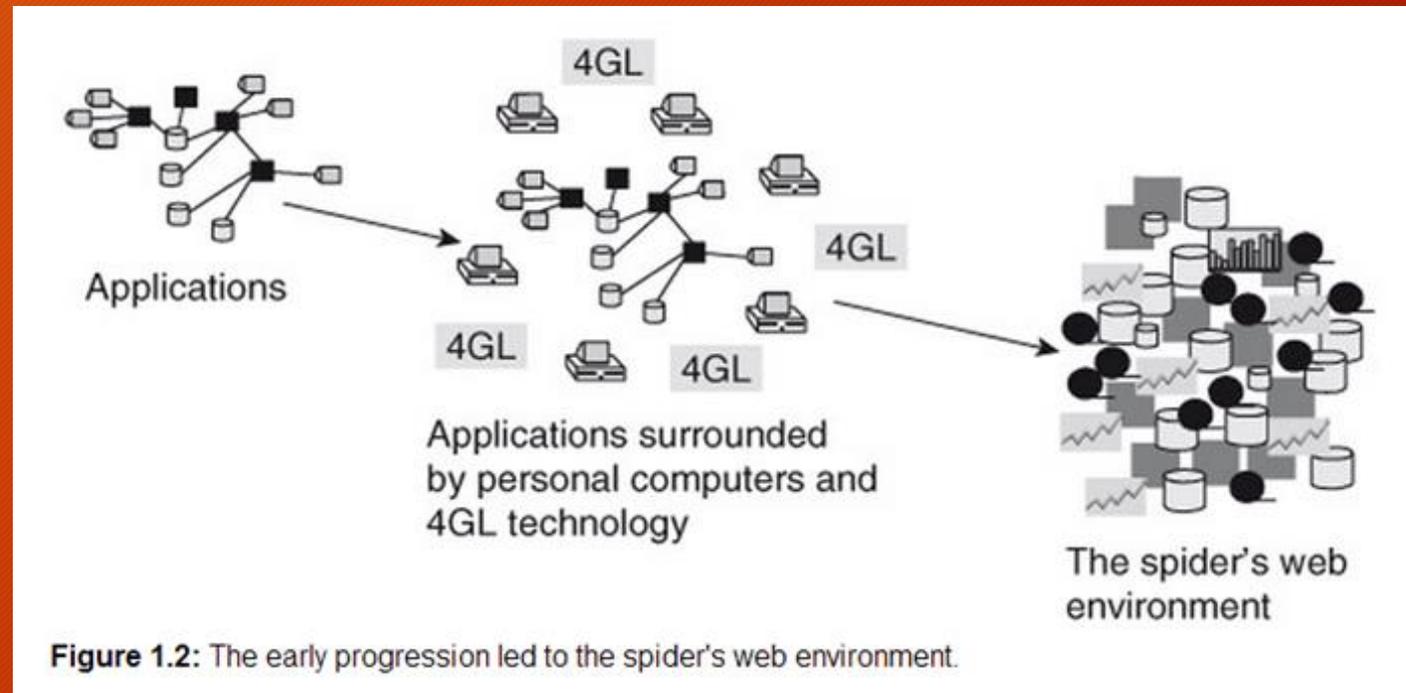- Then, comes disk storage with DBMS software for managing storage on the disk itself

Figure 1.1: The early progression of systems.
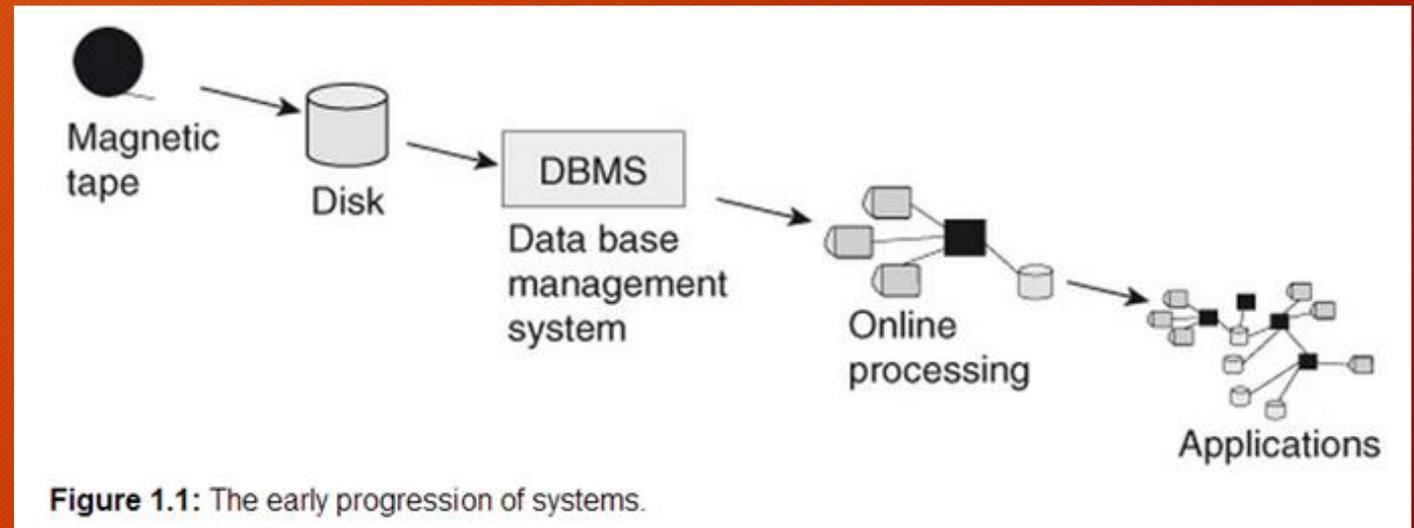
Still on history...[7]

- With direct access and a DBMS, came be online applications: ATMs, bank teller processing, claims processing, airline reservation processing, manufacturing control processing, retail point of sale processing, and many, many more

- Then came PCs and 4GL technology (4th generation programming), corporate users were free to access their own data, but found data was inaccurate, not timely, hard to find, and without documentation

Figure 1.1: The early progression of systems.
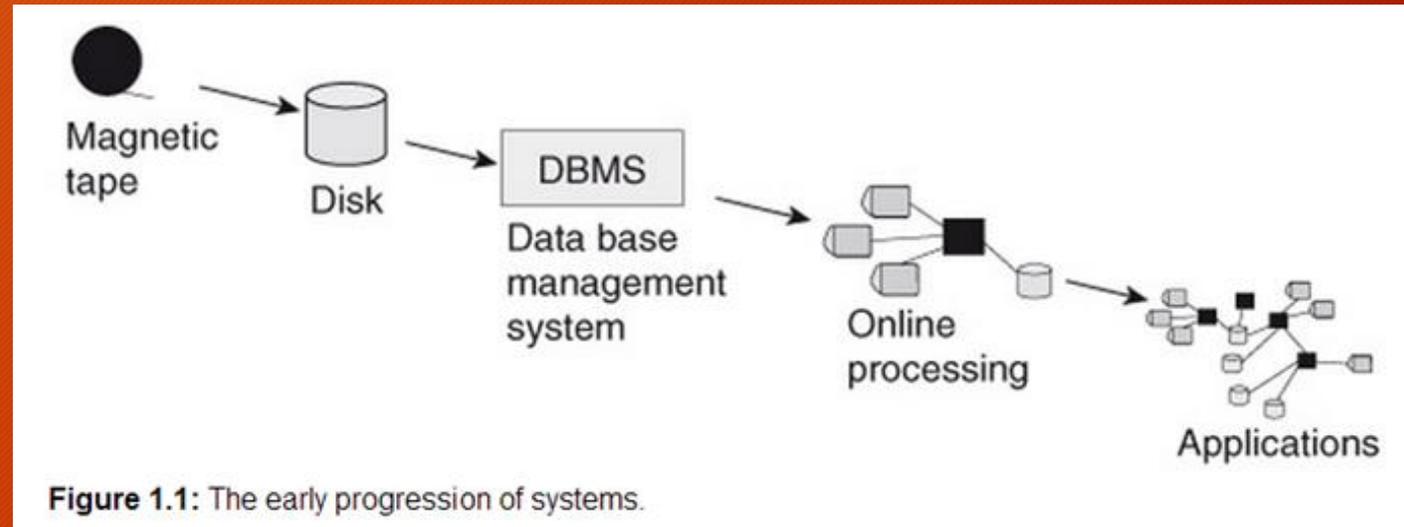
Still on history...[7]

- Inmon: resulted in a big mess, sometimes affectionately called the "spider's web" and a typical corporate IT environment could look like this...

- End users, IT professionals, and management frustrated. But it resulted in a movement to an information systems architecture that centered around a **data warehouse**.

Applications

4GL

Applications surrounded by personal computers and 4GL technology

The spider's web environment

Figure 1.2: The early progression led to the spider's web environment.
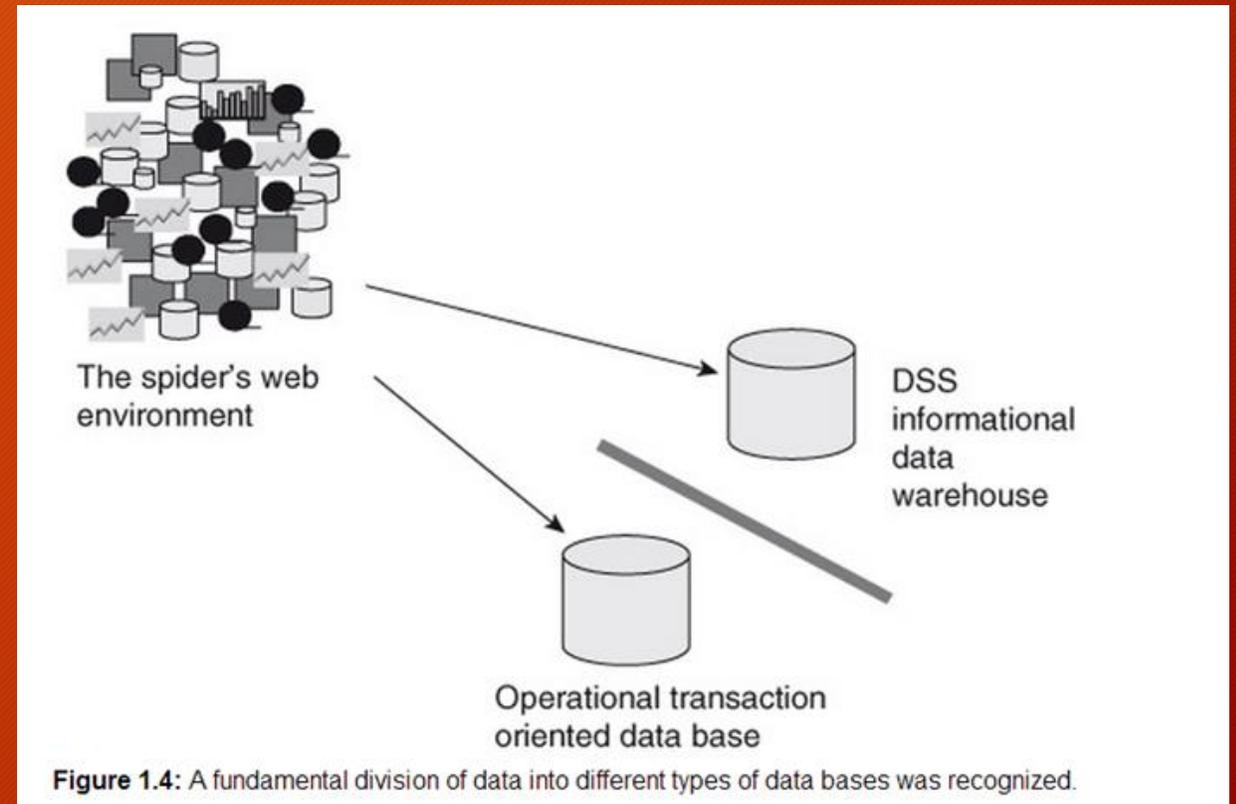
So, what's this data warehouse…?

What is a Data Warehouse?
Inmon defined as being
- subject oriented
- integrated
- nonvolatile
- time variant
- a collection of data in support of management's decision

And even Ralph Kimball agrees with that…!

It's made up of hardware, software and more…

The spider's web environment

DSS informational data warehouse

Operational transaction oriented data base

Figure 1.4: A fundamental division of data into different types of data bases was recognized.

DSS = Decision Support Systems

So, what is the ODS...?[12]

Central point of data integration for business management; delivers a common view of enterprise data.

Legacy systems feeding the DW/BI solution frequently include customer relationship management (CRM) and enterprise resource planning solutions (ERP), and can generate a hug amount of data.

Vendors often build DW solutions to use an ODS to consolidate and facilitate the extract-transform-load (ETL) processes for these systems and parse the information from the ODS into the actual DW.
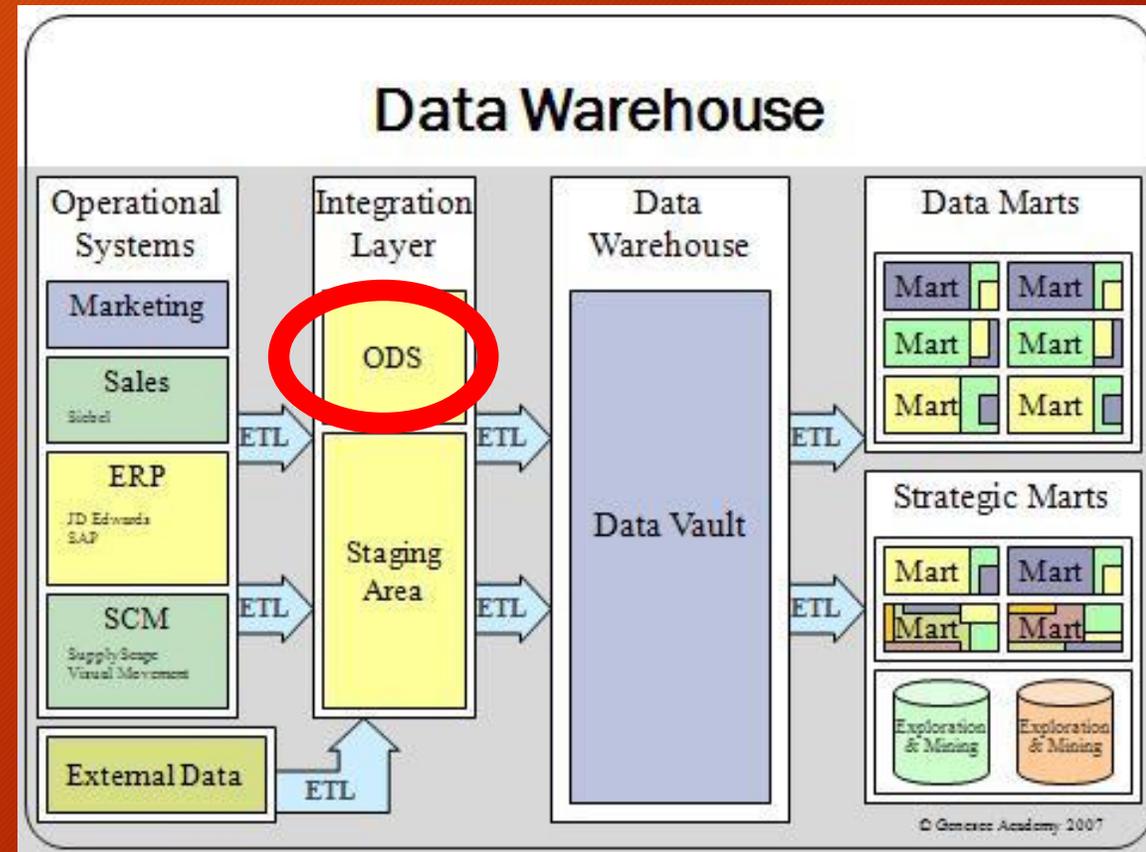
How is an ODS different from a data warehouse...?

DW supports strategic planning & business intelligence decision support. It contains:

- Integrated subject oriented data, e.g. sales data
- Static data, e.g.. data moved into DW should not change in the data warehouse environment
- Historical data, e.g.. data warehouses usually contain several years of historical data
- Aggregated data, e.g.. as it ages, data is summarized, reducing data storage requirements and improving analysis performance

ODS intended to support operational management and monitoring. It contains:

- Integrated subject oriented, e.g.. sales data (similar to data warehouses)
- Volatile data, e.g.. data moved into an ODS will probably change frequently
- Current data, e.g.. an ODS usually contains weeks of data instead of large volumes of historical data
- Detailed data e.g.. as data ages, it is summarized reducing data storage requirements and improving analysis performance.

ODS data is refreshed frequently. It provides a "snap-shot" of the OLTP and legacy systems.

Okay, what's the Corporate Information Factory...?[8]

First introduced by Inmon in early 80s, the CIF comprises:

- The business people generating the transactions
- Applications (in and out of the CIF)
- **The ODS (the CIF contains the ODS)**
- An integration & transformation layer
- Data warehouse
- Data marts
- Internet/intranet
- Metadata, exploration & data mining warehouse, alternative storage
- Decision Support Systems (DSS)

Both generic and unique to each business

THE CORPORATE INFORMATION FACTORY
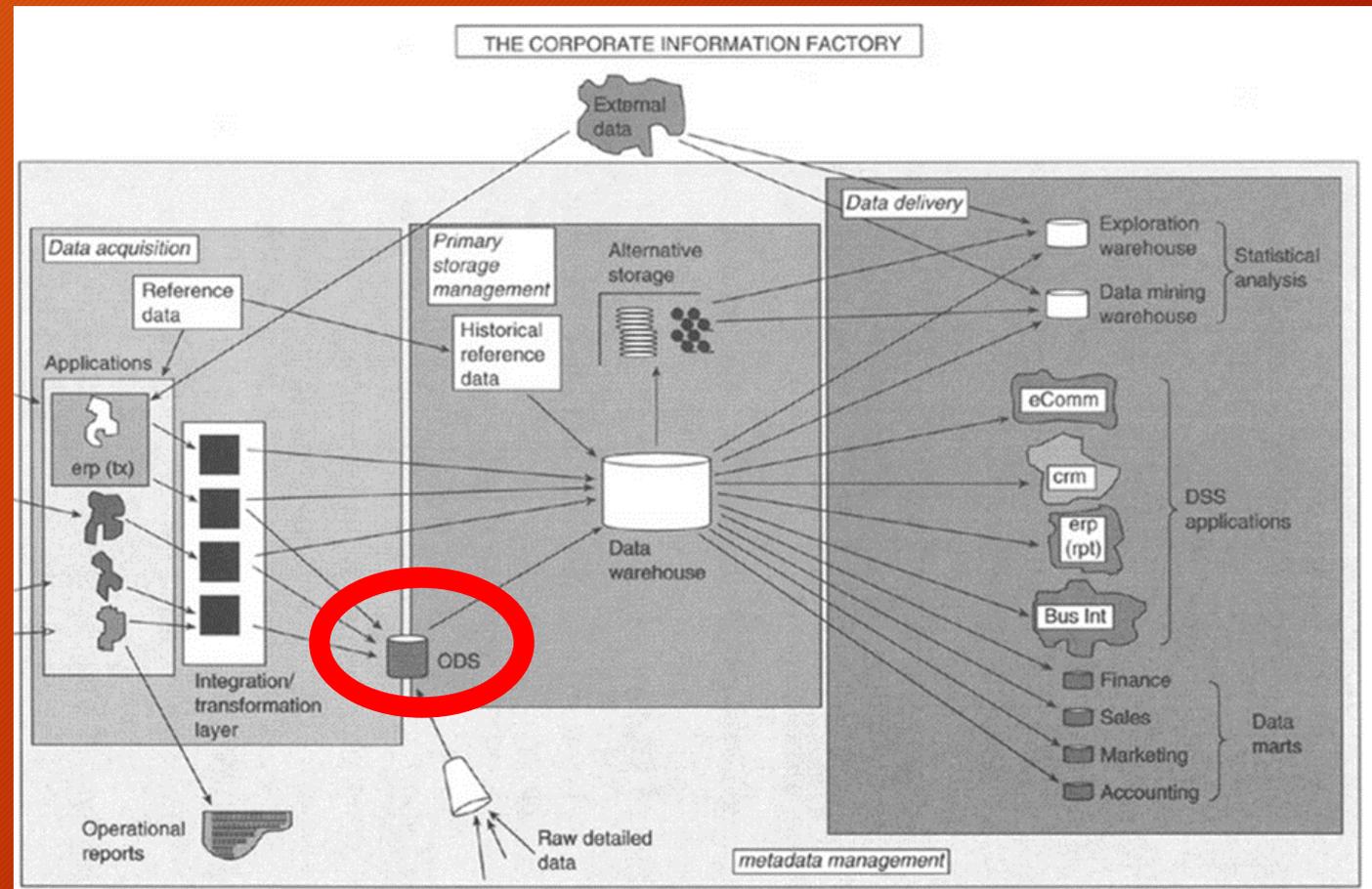
How about these Inmon and Kimball guys; Can't they get along…? (Yes, at first…)[11]

Kimball and Inmon, working separately, arrived at a common set of guidelines (or principles)—at first...

- **Subject Orientation:** Data grouped by subject, rather than author, department, or physical location. Manufacturing data goes together, sales data, promotions data, etc., regardless of where it came from.

- **Data Integration**: Data should be smoothed out to have the same look and feel, even despite getting there from separate applications, departments, etc.,.

- **Nonvolatility**: Data in a data warehouse remains in the warehouse (kind of like Vegas), unlike operational application data which is discarded once the company is finished using it.

- **Time Variant**: All data has a context at a moment in time and a data warehouse preserves that context. All data from 2014 retains its context within 2014.

- **One Version of the Truth**: Data proliferation in 80s and 90s yielded many copies of the same data. A data warehouse contains only one, true gold, standard copy of each data element.

- **Long-Term Investment**: A data warehouse should be flexible enough to absorb changes in the company and the world, and scalable enough to grow with the company. By doing so, a data warehouse can add value to the company for a long time.

# Inmon and Kimball: (VHS v. Beta; Microsoft v. Apple; Chan v. Pete...) [11]

Although sharing a similar set of principles, each used completely different designs for DW.

- Kimball created the Dimensional Data Model.

- AKA the Star Schema, a Dimensional Data Model has a distinct shape (resembles a star).

- Each Dimension table holds all the permutations of a company's single hierarchy (e.g.., geography: city, county, state, region, district, etc.; or time: second, minute, hour, day, fiscal week, payroll week, fiscal quarter, etc.).

Figure 1.2: Dimensional Data Model.
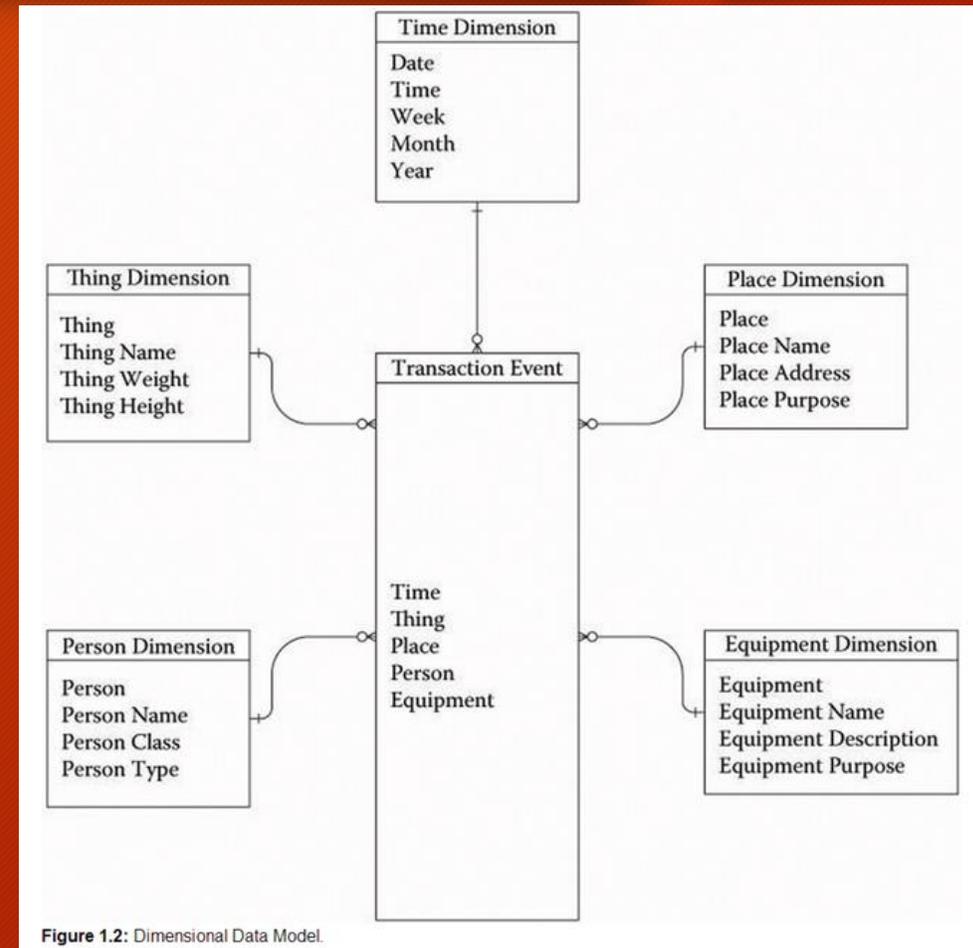
How about these Inmon and Kimball guys; Can't they get along…? (Then, not so much…) [11]

- Inmon preferred the Third Normal Form Data Model

- Third Normal Form allowed the data to have the same flexibility as the company, rather than capture hierarchies and relationships in Dimension tables

- A debate emerged: Which was better?

- Some vendors settled on both and apply the strengths and avoid the weaknesses of both according to the business situation.

Figure 1.3: Third Normal Form Data Model.
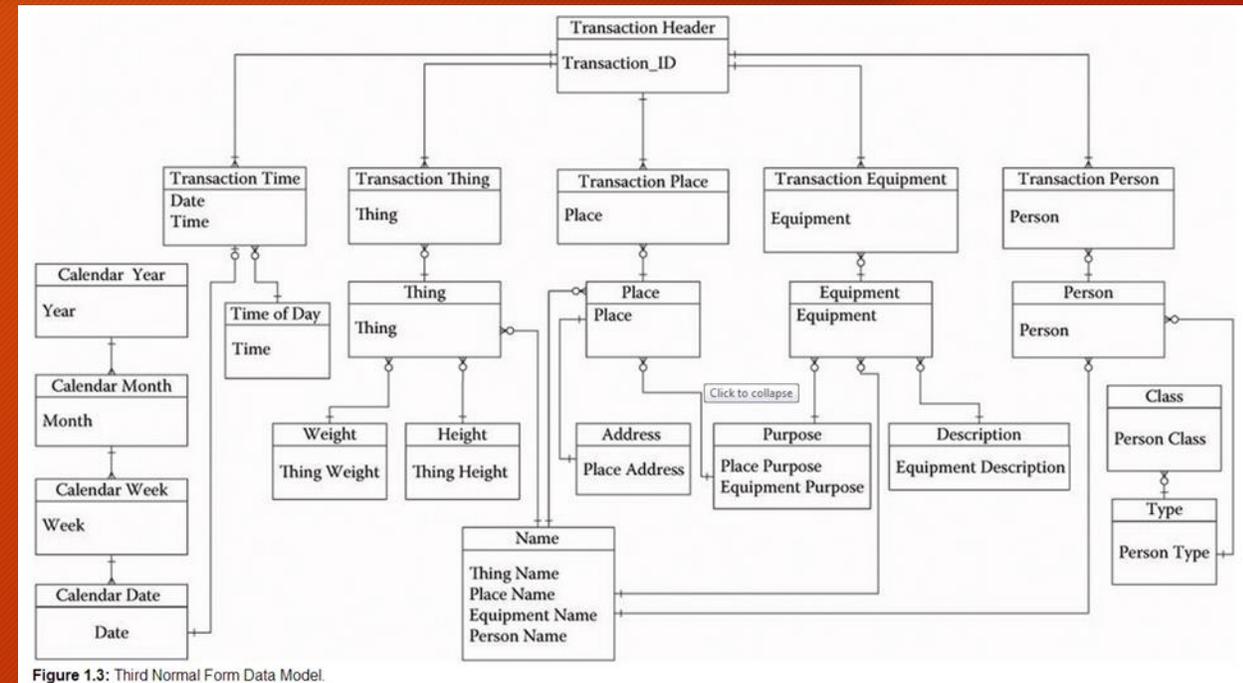
How about these Inmon and Kimball guys; Can't they get along…? (What else can go wrong…?) [14]

Inmon characteristics:

- Top-down approach; "Big-bang" implementation. EDW, then data marts

- Makes use of well-known database development methodologies and tools

- The DW is part of the CIF, the broader corporate information environment

- More dependent on IT professionals; fans say it provides better technical solutions

- Secondary role of business users

Kimball characteristics:

- Bottom-up; build independent data marts, then construct an EDW

- Data mart built for each process

- Multiple data marts to meet all business needs and to create a data warehouse

- Data marts connected via data bus

- Business users play larger role in developing data marts

How about these Inmon and Kimball guys; Can't they get along…? (Some differences)[14]

Complexity

- Inmon - Complex
- Kimball – Simple

Primary focus

- Inmon - IT professionals due to methodology, architecture, and dev approach (spiral); leads to IT ownership
- Kimball - Business users because of simpler development methodology

Tools

- Inmon - ERDs
- Kimball - Dimensional modeling

Data-modeling approach

- Inmon - Subject-oriented or data-driven
- Kimball - Process-oriented; users can participate actively

Data warehouse architecture

- Inmon - EDW serves entire enterprise & feeds departmental data marts/databases
- Kimball - Single-process data arts connected via a data bus and conformed dimensions

Presenting DW 2.0[13]

This is going to make comparisons more difficult, but you have to know where you've been to know where you're going. Inmon and fans think Kimball is about 10 years behind Inmon's thinking.

"The emphasis for the predicted Stage 4 Kimball model – the need for integration across the enterprise - was the one that Inmon recognized 10 years earlier."
--Inmon in *A Tale of Two Architectures*

Inmon has moved on to include everything that was in the CIF to all a businesses unstructured data and textual data, however that appears: text, emails, tweets, contracts, whatever…
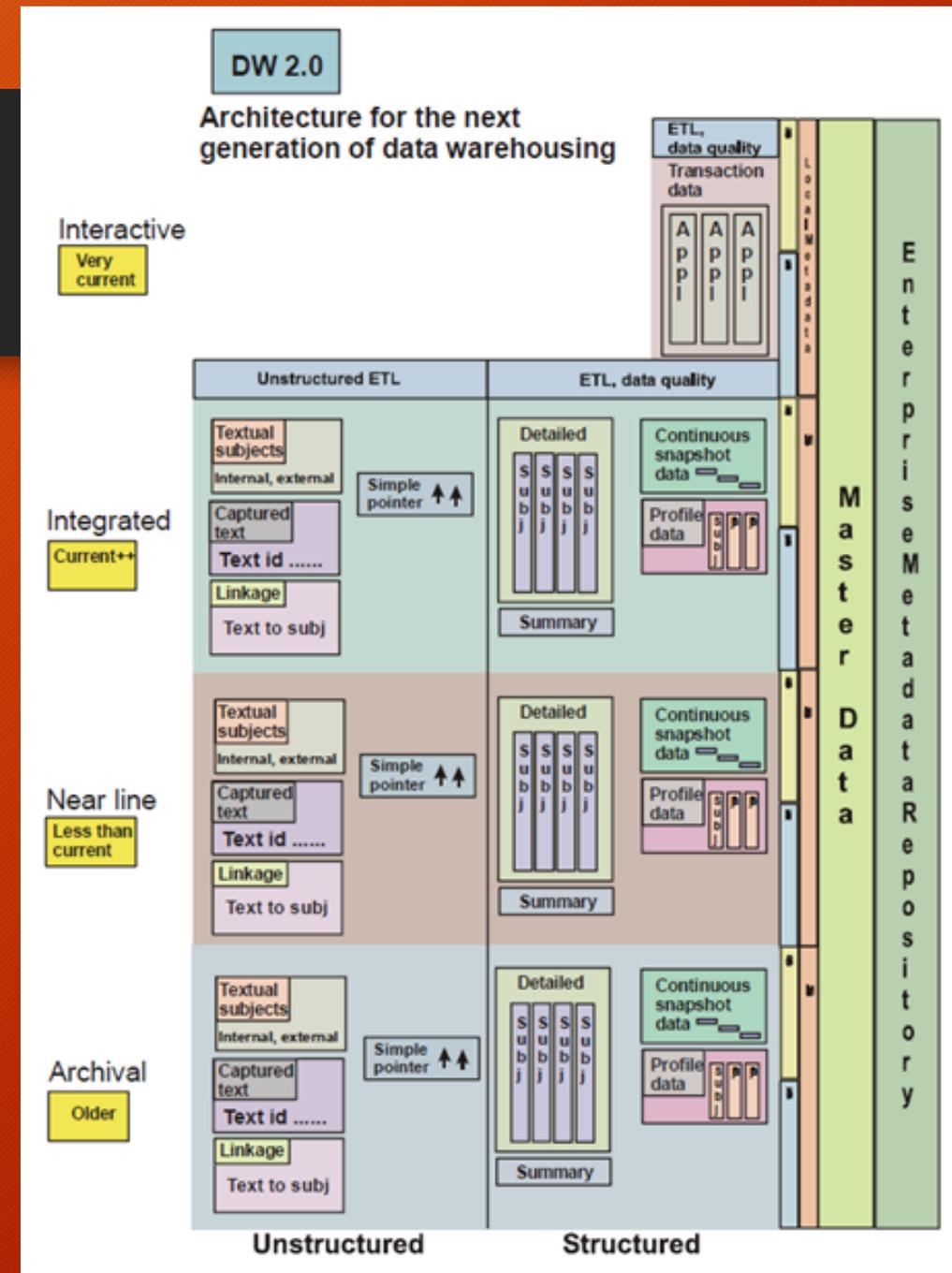
Case Study: WASHINGTON DENTAL SERVICES[15]

- WDS tried to build a Corporate Information Factory (CIF) including an Enterprise Data Warehouse (EDW) and accompanying Data Marts (Inmon's approach)
- Attempted in-house first, but hit many design & maintenance snags; unable to effectively and efficiently load and retrieve data. Also had resource issues, reporting restrictions and space and time limitations, creating a need for expansion and enhancements.
- WDS also needed to aggregate the data for reporting to state and government agencies, select clients, and corporate management. Ended up brining in TEKsystems.
- TEKsystems assembled a 13-person team: an information architect, BI architect, data architect, ETL architect, business architect, four ETL developers, two QA ETL engineers and a dedicated Project Manager (PM).
- TEKsystems successfully planned and executed a strategy for a successful CIF/DW. The solution enables WDS to further define its best DW architecture and data models, aligns the system with key business objectives, and expands the system to accommodate new products, new services and business requirements.

Glossary

- **Operational Systems** are the internal and external core systems that support the day-to-day business operations. They are accessed through application program interfaces (APIs) and are the source of data for the data warehouse and operational data store. (Encompasses all operational systems including ERP, relational and legacy.)
- **The Data Warehouse** is a subject-oriented, integrated, time-variant, non-volatile collection of data used to support the strategic decision-making process for the enterprise. It is the central point of data integration for business intelligence and is the source of data for the data marts, delivering a common view of enterprise data.
- **Primary Storage Management** consists of the processes that manage data within and across the data warehouse and operational Data store. It includes processes for backup and recovery, partitioning, summarization, aggregation, and archival and retrieval of data to and from alternative storage.
- **Alternative Storage** is the set of devices used to cost-effectively store data warehouse and exploration warehouse data that is needed but not frequently accessed. These devices are less expensive than disks and still provide adequate performance when the data is needed.
- **The Data Mart** is customized and/or summarized data derived from the data warehouse and tailored to support the specific analytical requirements of a business unit or function. It utilizes a common enterprise view of strategic data and provides business units more flexibility, control and responsibility. The data mart may or may not be on the same server or location as the data warehouse.
- **The Operational Data Store (ODS)** is a subject-oriented, integrated, current, volatile collection of data used to support the tactical decision-making process for the enterprise. It is the central point of data integration for business management, delivering a common view of enterprise data.
- **Management Meta Data** is the process for managing information needed to promote data legibility, use and administration. Contents are described in terms of data about data, activity and knowledge**.**
- **The Exploration Warehouse** is a DSS architectural structure whose purpose is to provide a safe haven for exploratory and ad hoc processing. An exploration warehouse utilizes data compression to provide fast response times with the ability to access the entire database.
- **The Data Mining Warehouse** is an environment created so analysts may test their hypotheses, assertions and assumptions developed in the exploration warehouse. Specialized data mining tools containing intelligent agents are used to perform these tasks.
- **Activities** are the events captured by the enterprise legacy and/or ERP systems as well as external transactions such as Internet interactions.
- **External Data** is any data outside the normal data collected through an enterprise's internal applications. There can be any number of sources of external data such as demographic, credit, competitor and financial information. Generally, external data is purchased by the enterprise from a vendor of such information.

Operational Data Store & Corporate Information Factory Bibliography

The following sources were used for this presentation and associated research paper.

1. Simon, Alan R.. "Chapter 3 - Have it Your Way—The Structure of a Data Warehouse." *Data warehousing for Dummies*. Foster City, CA: IDG Books Worldwide, 1997. . Print.

2. Inmon, William H.. "Chapter 10 - The Data Warehouse and the Web." *Building the data warehouse*. 4th ed. Indianapolis, Ind.: Wiley, 2005. . Print.

3. Inmon, William H., and Claudia Imhoff. "Chapter 1 - Creating an Information Ecosystem." *Corporate information factory*. 2nd ed. New York: John Wiley & Sons, 2000. Print.

4. Inmon, William H., Derek Strauss, and Genia Neushloss. "Chapter 1 - A Brief History of Data Warehousing and First-Generation Data Warehouses." *DW 2.0 the architecture for the next generation of data warehousing*. Amsterdam: Morgan Kaufmann/Elsevier, 2008. . Print.

5. Brunelli, Mark. "Bill Inmon: Kimball methodology ignores the value of textual data." *September 2011*. SearchDataManagement.com, 13 Sept. 2011. Web. 20 July 2014. <http://searchdatamanagement.techtarget.com/archive/2011/9>.

6. Hall, Mark. "The 10 IT People Who Mattered in the Past 40 Years (but You May Not Know Why)." *Computerworld*. Computerworld, 9 July 2007. Web. 20 July 2014. <http://www.computerworld.com/s/article/295941/The_10_IT_People_Who_Mattered_in_the_Past_40_Years_but_You_May_Not_Know_Why_?pageNumber=2>.

7. "Bill Inmon." *Wikipedia*. Wikimedia Foundation, 19 June 2014. Web. 15 July 2014. <http://en.wikipedia.org/wiki/Bill_Inmon>.

8. Inmon, William H., Derek Strauss, and Genia Neushloss. *DW 2.0 the architecture for the next generation of data warehousing*. Amsterdam: Morgan Kaufmann/Elsevier, 2008. Print.

Operational Data Store & Corporate Information Factory Bibliography, Page 2

Continued from previous page...

9. "Claudia Imhoff." *Wikipedia*. Wikimedia Foundation, 13 July 2014. Web. 15 July 2014. <http://en.wikipedia.org/wiki/Claudia_Imhoff>.

10. "Book Claudia Imhoff, Speaker at Speakers.com for Corporate Events." *Speakers.com*. N.p., n.d. Web. 20 July 2014. <http://www.speakers.com/Claudia-Imhoff-speaker-biography>.

11. Silvers, Fon. "The Big Picture: An Introduction to Data Warehousing." *Building and maintaining a data warehouse*. Boca Raton: CRC Press, 2008. . Print.

12. "Data warehouse." *Wikipedia*. Wikimedia Foundation, 17 July 2014. Web. 20 July 2014. <http://en.wikipedia.org/wiki/Data_warehouse>.

13. "A TALE OF TWO ARCHITECTURES." . N.p., 13 June 2010. Web. 20 July 2014. <http://inmoncif.com/registration/whitepapers/A%20TALE%20OF%20TWO%20ARCHITECTURES.pdf>.

14. Khan, Arshad. "Chapter 25: Inmon Versus Kimball." *Business intelligence and data warehousing simplified 500 questions, answers, and tips*. Dulles, Va.: Mercury Learning and Information, 2012. . Print.

15. "Building a Corporate Information Factory for Washington Dental Services." . TEKsystems, n.d. Web. 21 July 2014. <http://www.teksystems.com/resources/case-studies/healthcare/data-services-for-wds>.

Other sources used for this project:

"What is Operational Data Store (ODS) | Online Learning." Online Learning. N.p., 15 Dec. 2007. Web. 17 July 2014. http://www.learn.geekinterview.com/data-warehouse/dw-basics/what-is-operational-data-store-ods.html

Bowman, David. Operational Data Store. 2009. http://www.information-management-architect.com/data-warehouse-architecture.html July 2014.

Imhoff, Claudia. "To V or Not To V-Business Intelligence Gets Virtual!." To V or Not To V. Dashboard Insight, 15 Dec. 2008. Web. 15 July 2014. http://www.dashboardinsight.com/articles/new-concepts-in-business-intelligence/to-v-or-not-to-v.aspx