Edmonds Community College

CIS 220 – Summer 2014

The Operational Data Store (ODC) and Corporate Information Factory (CIF)

Last Update: 7/26/2014 6:17:41 PM

The Undecided
Frankie Brooks
James Murray
Paul Darlington
Bill Seling

Table of Contents

Revisions

Summary of Revisions Made to Document			
Version	**Primary Author(s)**	**Description of Version**	**Date**
1.0	Bill Seling	Initial version.	07/23/2014

Table 1. Document revision table.

1. The Main Parts

The Undecided presentation covers a couple of data warehousing-related objects and people: The Operational Data Stores (ODS), the Corporate Information Factory (CIF), Bill Inmon, Claudia Imhoff, lightly introduces how Ralph Kimball's approach to data warehousing compares with the Inmon (and Imhoff) approach.

1.1 A Couple of key definitions to get started…

The Undecided previous presentation (Business Capabilities) provided the team with some difficulties in the definition department—there is no official and/or accepted industry-wide definition. Because data warehousing and any of its related terms can go back as far as the 80s, this week's subjects were no less daunting to define. Yet, we muddled on, because, as IT professionals, that is what teams do.

1.1.1 Operational Data Store (ODS)

Don't worry about knowing what the ODS is for right now. Just read the following three definitions to get a feel for the concept of an 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]

So an ODS is where data comes into, from a variety of sources, and depending on how it's put together, users can peer into it before it gets shipped off to the data warehouse.

1.1.2 Corporate Information Factory (CIF)

Same as above, just let the three following definitions soak into your brain, without expecting to fully understand them, but to know enough to recognize them when they come up later.

- 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 factory4

- 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 2.05

1.2 And some key personnel…

The team apologizes to any persons living or dead, who may have made valuable contributions to the field of data warehousing but go unrecognized because Inmon, Kimball, and Imhoff are clogging up Google's estimated 62,200,000 results.

1.2.1 William (Bill) H. Inmon

Inmon was born July 20, 1945 in San Diego, California. He earned his B.Sc. in Mathematics from Yale University, and his M.S. in Computer Science from New Mexico State University.[7]

The semi-esteemed technology publication, *Computerworld*, in 2007, recognized Inmon as one of "10 IT People Who Mattered in the Past 40 Years," not only as the "Father of the Data Warehouse," but father of the $28 billion industry he helped launch, as well as the author of more than 650 articles and 46 books (and that was seven years ago). Just to add some perspective, Edgar Codd, author of the 12 rules for what makes data relational, is also on that list.[6]

Inmon also authored "DW 2.0," both a book and architecture: DW 2.0 includes everything up to the CIF (which you'll hear more about in a bit) and now includes textual information Inmon calls unstructured data (web, contracts, email, tweets, etc.) in our data warehouse.[5]

1.2.2 Claudia Imhoff

Imhoff received her BS in Education from Duke University, her MS from University of Colorado Boulder, and her PhD in General Biomedical Sciences from University of Tennessee[9]

Imhoff has co-authored at least five books and writes monthly columns and posts for more than 60 technical and business magazines[10]

In a DSSResources.com e-interview in 2006, Imhoff said, "When I first started as a consultant in the late 1980's, decision support systems (DSS) were just beginning to be implemented. The company I worked for was also the employer of Bill Inmon, one of the early promoters and still one of the visionaries in BI. I was fortunate to attend a course given by Bill on this very subject and fell in love with the concept, the architecture, and the challenges associated with such a novel idea."

Imhoff is often credited as a co-creator of the CIF (with Inmon), but I'm never sure which version of the CIF she pitched in on. (Inmon has been working on the CIF since the 80s.) Since her "Corporate Information Factory" book was published in 2001, she must have some relatively early input to today's CIF.[9]

1.2.3 Ralph Kimball[11]

Kimball will likely be the subject of another team's presentation, so our team will give you just enough basics to set the stage, because, to give you a bit of a teaser, these people don't exactly see eye to eye.

Kimball was born in 1944. That makes him about a year older than Inmon, yet Inmon is called the *Father of data warehousing* (wouldn't that chap your hide?), and received a PhD from Stanford, so do not underestimate this guy.

Kimball was a co-creator of the Xerox Star Workstation and the world's first commercially viable GUI application—*take **that** Apple!* He also was the founder and CEO of Red Brick Systems, who created an extremely fast RDBMS targeted for data warehousing. It was later acquired by Informix and is now owned by IBM.

Kimball authored *The Data Warehouse Lifecycle Toolkit*, introducing the Dimensional Data Model to the world and set the stage for future debates as which system is better: Kimball or Inmon?

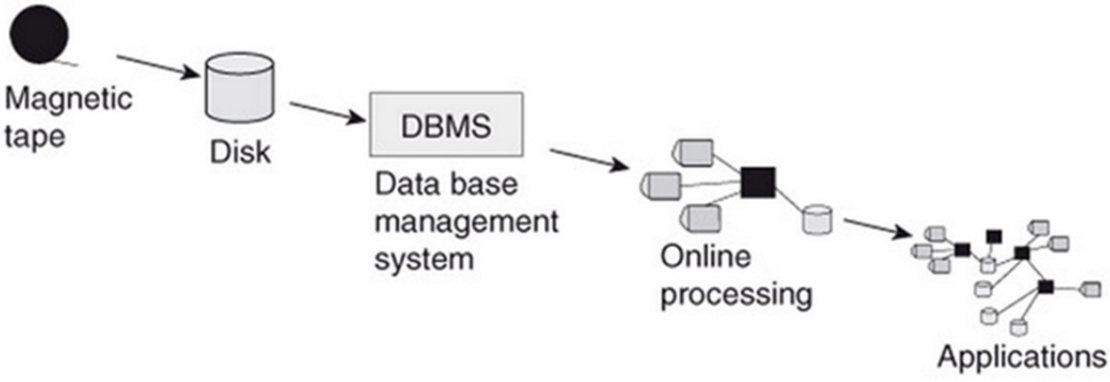
2. First, some history…[7]

In order to better understand our leap to CIFs and ODSs, the team thinks it's necessary to trot out their senior member to tell you about the good ol' days of data processing, before all this newfangled stuff came along. Bill worked for subsidiary of GTE, GTE Data Services (GTEDS), where he experienced the very beginnings of this mess we're in now. Yes, we said it, Bill made this mess we're in; it's his fault.

Before you could integrate, virtualize, stream, or do anything to data, you have to get it **IN** to the computer somehow. Now, for our diagrams, we simply draw a pluffy cloud, but way back in the 80s you had to actually work to get data into a computer before you could anything to it, which wasn't much. There were punch cards, paper tape, and core memory for the computer that was wired by hand. Bill remembers watching his dad program a computer by punching in wire loops. Wild stuff…

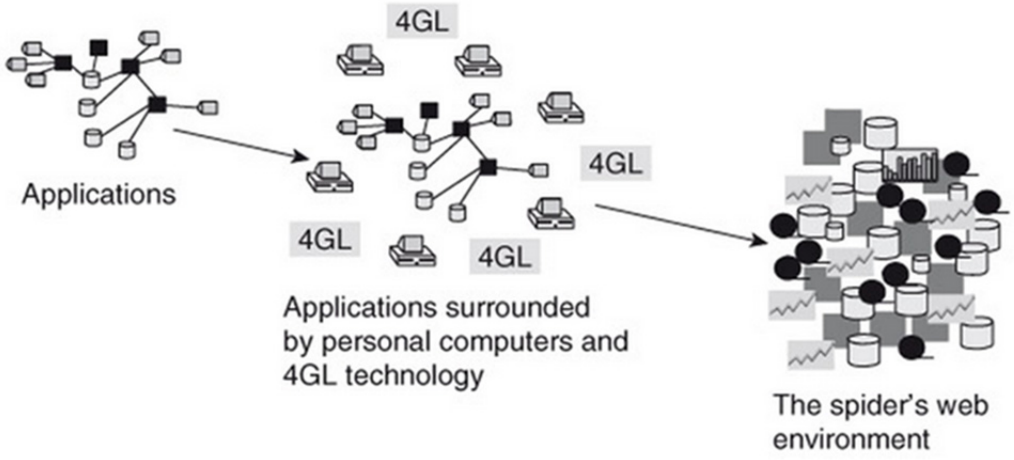
Then things really got wiggy. Magnetic tape (read sequentially) was introduced, but the tape could wrinkle, break, or the oxide could fall off or be scratched off and make the tape useless. Really, tape wasn't new; it had been around since the 50s, but it had been incrementally improved for higher density writing and length—up to 3600 feet. A master file could comprise more than 20 tape volumes and if one tape was bad, the entire set was bad.

Then, we're not sure exactly how, maybe extraterrestrials, along came direct access storage devices (DASD), specifically disk storage. It had been around since the 50s also (hey…just like Bill—maybe there is something to this extraterrestrial angle), but didn't become "affordable" for smaller companies until the 90s. Boy Howdy, did things take off then. By this time, personal computers were literally invading the workplace (aliens again?) and when affordable hard disk drives (HDD) became available, well, there was no stopping data now, was there? It was out of the barn and never going back.



But the mess was only beginning. With all this direct access to data (and throw in a DBMS), it wasn't long before programmers figured out how to develop online applications. These were used in then innovative ways: ATMs, bank teller processing, claims processing, airline reservation processing, manufacturing control processing, retail point of sale processing, and many, many more. We thought computers would run the world someday. Today, we know that no computer can run for more than a couple of days without a patch, service pack, or upgrade. Take over the world? Please…

And the data mess continued to grow at a staggering rate as PCs proliferated and 4GL technology (4th generation programming), took overtook nearly all corporations. Corporate users were free to access their own data stored out on all these now relatively cheap storage devices (tape, disk—even printouts), but they found the data was inaccurate, it wasn't necessarily timely, it was incredibly hard to find sometimes, and there was no documentation to describe what you were looking at. Today, we think of easy-to-read data, like Word or something. But some of this data was little more than ones and zeroes. If you didn't have the program that was written to consume the data, good luck.



3. Inmon and Kimball Set Out to Save the World

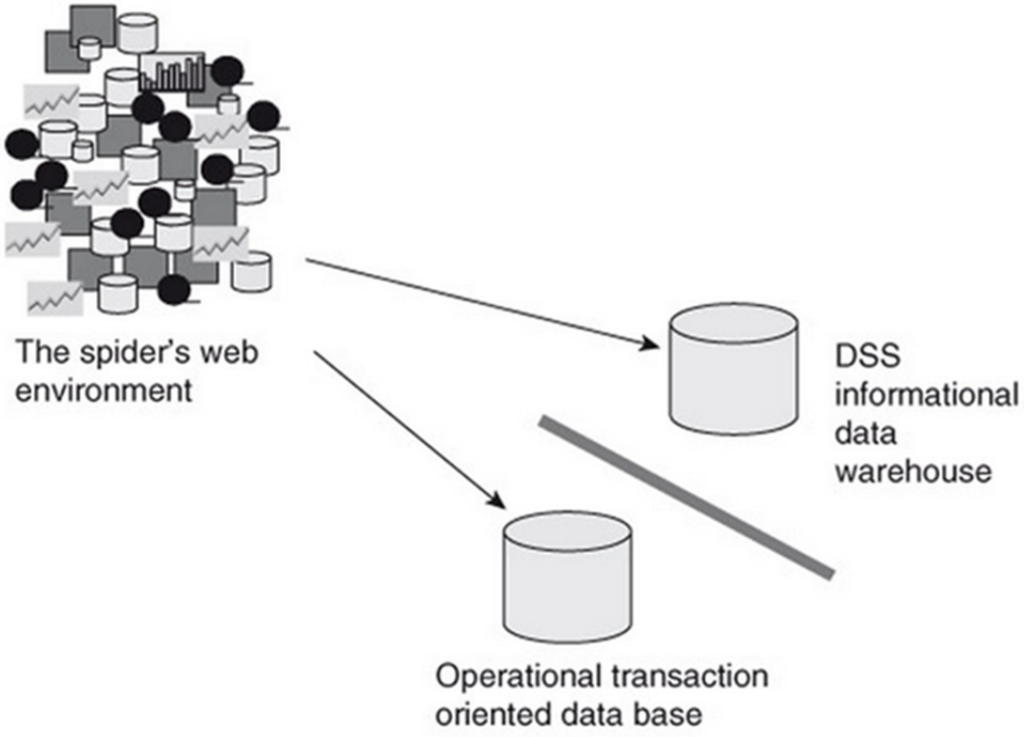
About this same time, we're still stuck in the 80s or so, Inmon and Kimball see this mess, this bunch of data sitting around and they say, "Someone's gotta fix this." As far as I know, neither added "…and get rich trying," but they should have. Oh, I imagine it was rough going in the beginning. Bunches of developers have spilled data all over your company, now how are you going to clean it up? Inmon said the mess was sometimes affectionately called the "spider's web."

The end users were not happy; the IT professionals were not happy; management frustrated. "What're we gonna do?" they all whined. Lo, there was much handwringing throughout the land. "Surely other business sectors must have dealt with something like this." Then it occurred to them: What did manufacturers do when they found their factories cluttered with refrigerators? *They built warehouses, that's what they did!* And that's what IT decided to do with its data. A movement started to develop an information systems architecture that centered around a **data warehouse**.

4. So, what's this data warehouse…?[7]

A data warehouse is a software and hardware solution, and it's big business. The nascent industry looked to Inmon, data warehousing's dad, to describe what a data warehouse looks like. Inmon described it as being:

- subject oriented
- integrated
- nonvolatile
- time variant
- a collection of data in support of management's decision

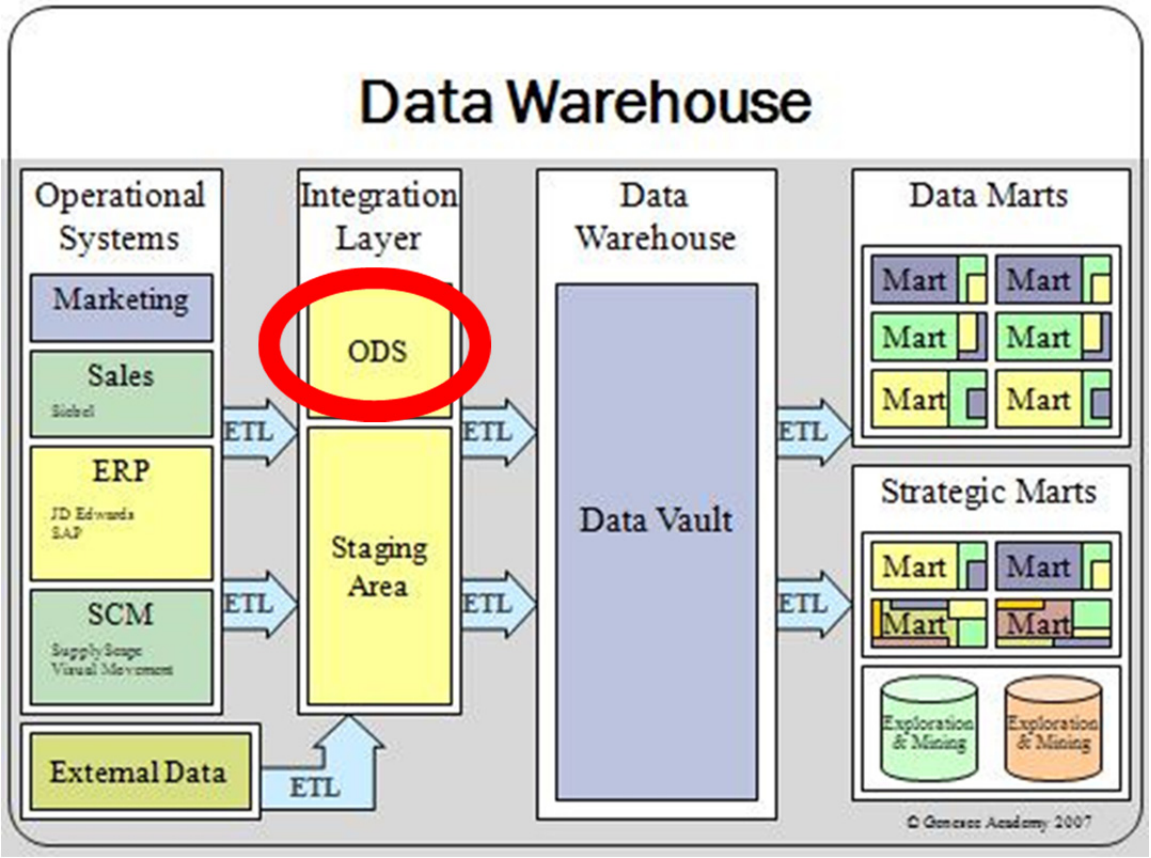


This description is about the only thing universally accepted among the data warehousing community. Even Ralph Kimball uses this definition in his writings.

Okay, we've got enough background to talk about the Operational Data Store and Corporate Information Factory now. Let's get to them before you completely lose interest and head off to play on Facebook or Twitter or some other social media that's been feeding off this data warehousing stuff without you even realizing it.

5. So, what is the ODS…?[12]

The Operational Data Store (ODS) is the central point of data integration for business management inside a data warehouse. The ODS delivers a common view of enterprise data. Vendors often build data warehouse solutions to use an ODS to consolidate and facilitate the extract-transform-load (ETL) processes for customer relationship management (CRM), enterprise resource planning solutions (ERP), and other systems. These systems can generate huge amounts of data and the ODS is the place to integrate it before the data is actually moved into the data warehouse. The ODS is a common for data warehouses designed using the Inmon or Kimball approaches or even a hybrid built on both approaches.



6. How is an ODS different from a data warehouse…?[12]

If the ODS sounds like a data warehouse, you're not alone in your thinking. Several sites and books will gladly tell you how they are different. We may as well, too.

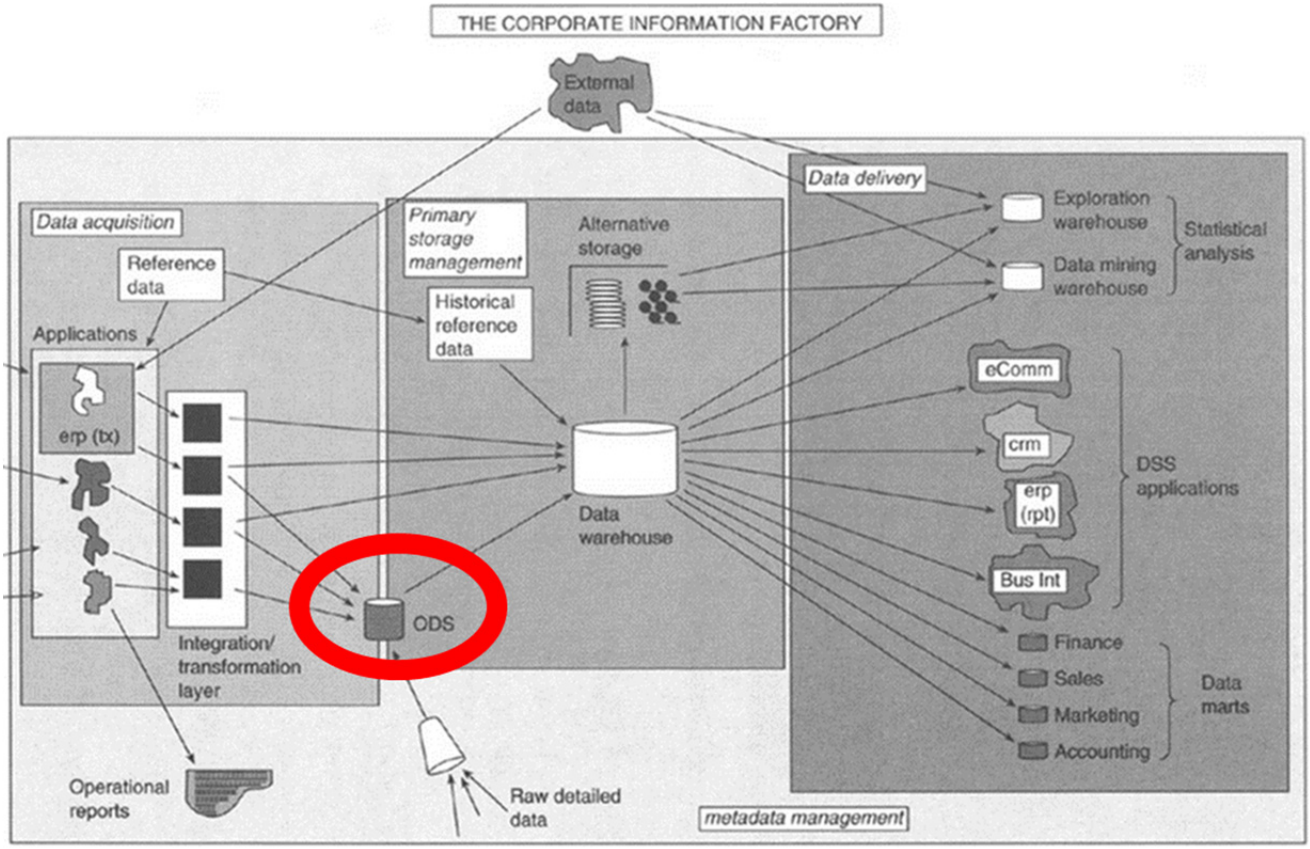
- **A data warehouse supports strategic planning and business intelligence decision support. It contains:**
  - Integrated subject oriented data, sales data, for example
  - Static data. For example, data moved into data warehouse should not change in the data warehouse environment
  - Historical data. Data warehouses usually contain several years of historical data
  - Aggregated data. As data ages, it is summarized to reduce data storage requirements and improve performance for analysis
- **The ODS is intended to support operational management and monitoring. It contains:**
  - Integrated subject oriented. For example, similar to a data warehouse, ales data
  - Volatile data. Data moved into an ODS will probably change frequently
  - Current data. An ODS usually contains weeks of data instead of years of historical data
  - Detailed data. As data ages, it gets summarized, reducing data storage requirements and improving analysis performance
- **The ODS data is refreshed frequently. It provides a "snap-shot" of the online transaction processing (OLTP) and legacy systems.**

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

Inmon first introduced his notion of a corporate information back in early 80s (boy, a lot of stuff happened in the 80s). Inmon's CIF comprises:

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

Experienced data warehouse architects could look at any company's data warehouse and recognize both generic and unique features for each business.



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

Recall that Kimball and Inmon are contemporaries. They were both working to solve the same problem: how to round up all that company data and put it to work. Working separately, they arrived at a common set of guidelines (or principles) for data warehouses.

- 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, despite arriving from separate applications, departments, etc.
- Nonvolatility: (a new word for our lexicon) 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. (Inmon, in recent interview, claims that Kimball is violating that rule. Our team's take is that Inmon's statement requires more checking. They are competing consultants and both looking to get business, after all.)
- 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.

8.1 Similar principles; completely different designs

Although Inmon and Kimball shared a similar set of principles for a data warehouse, each used completely different designs for a data warehouse.

- 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.).
- Inmon preferred the Third Normal Form Data Model

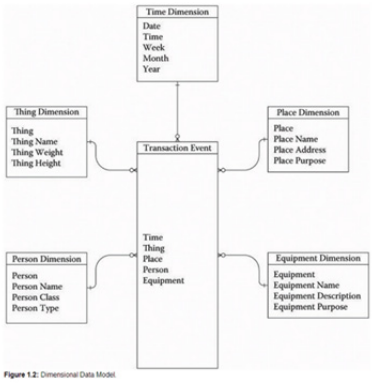
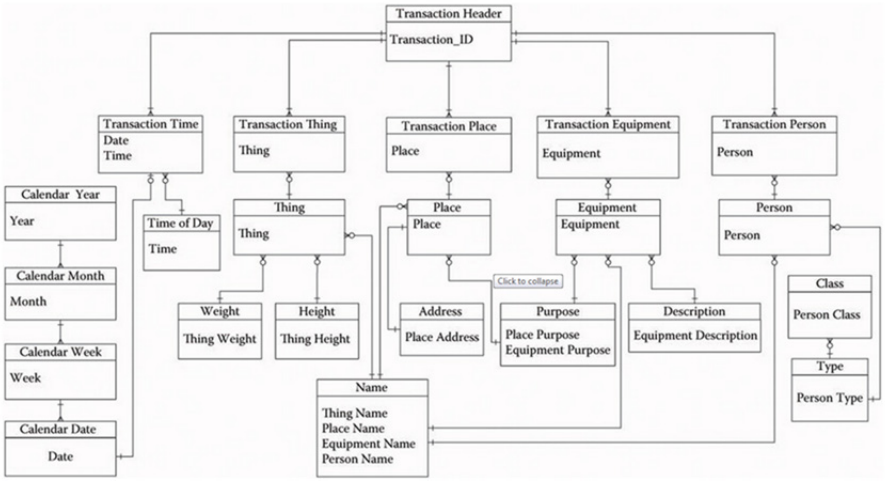


Figure 1.2: Dimensional Data Model

- 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 approaches, apply the strengths, and avoid the weaknesses of both according to the business situation.



8.2 Characteristics of the Inmon Approach[14]

- 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

8.3 Characteristics of the Kimball Approach[14]

- 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

8.4 More differences between the two approaches…[14]

Complexity

- Inmon – Complex
- Kimball – Simple

Primary focus

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

Tools
- Inmon – ERDs
- Kimball – Dimensional modeling

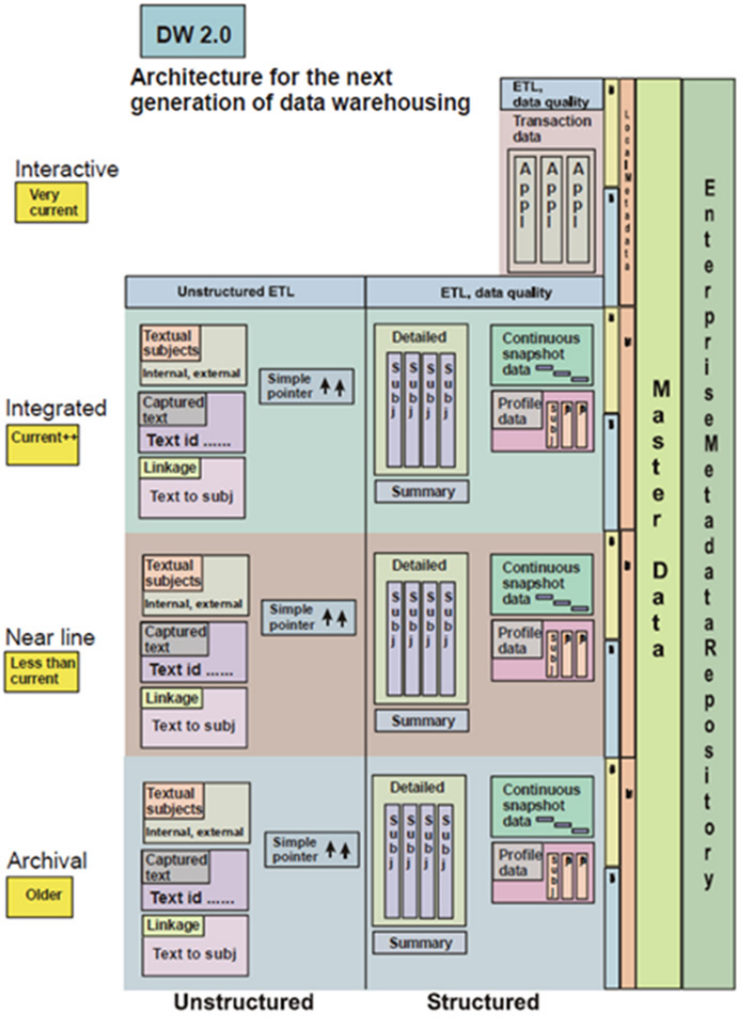
Data-modeling approach
- Inmon – Subject-oriented or data-driven
- Kimball – Process-oriented; users can participate actively

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

9. And just when you thought it was safe to go into the water…[13]

Remember our earlier history lesson, about how we ended up with data warehouses? It required evolutionary and incremental movements to get where we are today and things still aren't settled. You can't go to Costco and buy a data warehouse, (but that would be a fun, wouldn't it?) yet. Around 2006, Inmon, a prolific writer, began firing shots across everyone's bow with a white paper declaring that the CIF needed more, and data warehousing itself needed to catch up to the latest data generating technologies, especially textual, unstructured data.

This includes things like contract, text messages, emails, etc. Inmon and others began suggest that Kimball's thinking was about ten years behind his approach (remember: competing consultants). Then in 2010, Inmon published his book *DW 2.0 the architecture for the next generation of data warehousing*. Kimball and Inmon both published books more than 20 years ago that are still being studied, so it may take a while for Inmon's latest text to catch on, but it would seem a safe strategy to buy his book and see what he's got to say about it.



10. Case Study: WASHINGTON DENTAL SERVICES (WDS)[15]

This case study was excerpted from an executive overview published by TEKsystems.

WDS tried to build a Corporate Information Factory (CIF) including an Enterprise Data Warehouse (EDW) and accompanying Data Marts (Inmon's approach). Their first attempt used in-house staff, but they ran into design and maintenance snags. They were unable to effectively and efficiently load and retrieve data; they had problems with resource issues, reporting restrictions, and space and time limitations. All the problems forced them to add extensions and enhancements to what they had—and they were not happy with what they had. (Recall that in earlier comparisons of the Kimball and Inmon approaches, we learned that Inmon's CIF approach required a lot of IT resources and is more complicated to implement than Kimball's approach. The team isn't implying anything with this statement, other than acknowledging that it takes pretty smart cookies to implement a data warehouse, and Inmon's approach may require smart cookies who have done this before.)

WDS also learned needed to aggregate their data for reporting to state and government agencies, select clients, and their own corporate management. They threw in the towel for the in-house project and brought 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 successfully creating the CIF/DW. Their solution allows 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. According to TEKsystems, WDS is happy with their solution.

Appendix A – Glossary

Item	Explanation
Operational Systems	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.)
Data Warehouse	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	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.
Data Mart	Consists of 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.
Operational Data Store (ODS)	A subject-oriented, integrated, current, volatile collection of data used to support tactical decision-making processes for the enterprise. It is the central point of data integration for business management, delivering a common view of enterprise data.
Management Meta Data	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.
Exploration Warehouse	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.
Data Mining Warehouse	An environment created allowing analysts to test hypotheses, assertions, and assumptions developed in the exploration warehouse. Specialized data mining tools containing intelligent agents are used to perform these tasks.
Activities	The events captured by the enterprise legacy and/or ERP systems as well as external transactions such as Internet interactions.

External Data	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.

11. Appendix B – Bibliography

The team used the following sources for the presentation and the 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.
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

Power, Dan. "Claudia Imhoff eInterview." *Claudia Imhoff eInterview by Dan Power*. DSSResources.com, 10 Dec. 2006. Web. 24 July 2014. <http://dssresources.com/interviews/imhoff/imhoff2282007.html>.