How to Quickly Integrate Source Marts into a Healthcare Enterprise Data Warehouse

Nationally ranked by *U.S. News & World Report* 16 years in a row, Indiana University Health (IU Health) is dedicated to delivering the highest-quality, leading-edge care to its patients. The academic medical center offers a full range of specialty and primary care services for children and adults, with particular expertise in transplant, pediatrics, orthopedics, neuroscience, cardiovascular and cancer care.

IU Health is also one of the nation’s busiest health systems. Its clinical teams have served more than 150,000 inpatient admissions and 2.5 million outpatient visits.

The system consists of:
- 19 medical centers
- More than 100 clinics
- Ancillary services
- 3,541 staffed beds
- 3,707 physicians
- 36,000 total team members

In keeping with its reputation for quality, IU Health is pioneering efforts to improve the care it provides while lowering costs. To implement a clinical and analytic framework that enables such transformation, the health system has partnered with enterprise data warehouse (EDW) vendor Health Catalyst. Health Catalyst’s EDW platform is organizing IU Health’s data from multiple source systems — clinical, supply chain, patient satisfaction and more — into a single source of truth that spans the enterprise and serves as a foundation for data-driven improvement.

**THE CHALLENGE: INTEGRATING SOURCE SYSTEM DATA QUICKLY AND CONSISTENTLY**

Making sure data from source systems is moved quickly, accurately and consistently into the EDW is an important task for IU Health’s Information Systems (IS) department. This central IS department supports IU Health department and facility data needs. The department has been tasked with managing the components of the technical architecture that contribute to the health system’s clinical decision support environment. Central IS’ goal is to increase the value decision-makers get from the health system’s data — and to do it with fewer resources.
A key advantage of our relationship with Health Catalyst is that they seek to empower us in the same way that we seek to empower our internal customers. So often you find vendors who say, ‘This is our proprietary approach, and this is the way it has to be.’ Catalyst delivers consulting and technology, but at the same time they train us to do things on our own. That creates a unique partnership in this industry. In fact, I’ve been in several meetings with our CIO where he references Catalyst as an example of how a partner should be — one that seeks to understand the client rather than imposing its methodology on us.

Robert Parr, Enterprise Architect

Before clinical teams can access aggregated data to drive clinical improvement initiatives, central IS has to incrementally load data from source systems across the medical centers into the EDW. IS leaders needed a way to streamline this process. Specifically, the central IS department was looking for:

- The ability for multiple data architects to work independently on integrating source systems into an enterprise data warehouse while, at the same time, ensuring data fields remained well structured and consistent
- A tool to empower data architects to integrate source systems faster and easier and to provide more timely support to the plethora of business analysts across the enterprise

**THE SOLUTION: HEALTH CATALYST’S SOURCE MART DESIGNER**

Central IS turned to Health Catalyst’s Source Mart Designer application to help its data architects integrate data quickly, dependably and consistently. The Source Mart Designer enables IS staff to rapidly design, develop and deploy unlimited Source Marts. A Source Mart is an important component of Catalyst’s late-binding data architecture. Each Source Mart contains data from a specific source system that has been extracted, minimally transformed and loaded into the data warehousing system. This data is then available for downstream use by advanced analytics applications running on the EDW platform.

The Source Mart Designer significantly automates the process of integrating data from source systems into the EDW’s Source Marts. The application:

- Facilitates the denormalization of data for an EDW to enable data analysis.
- Analyzes the structure of data sources and recommends pragmatic data types and column names that conform to the EDW’s technical standards.
- Captures the mapping of source system tables and columns to Source Mart tables and columns as metadata, which maintains continuity and traceability between the source system and the destination Source Mart.
- Recommends the best data type for the data moving into the data warehouse. Data types are converted from the native source system type to the native types of the enterprise data warehouse to maintain consistency with the source system.

**BENEFITS: QUICK, CONSISTENT AND EFFICIENT DATA INTEGRATION**

75 percent faster design and development

Using Source Mart Designer, IU Health’s IS team has achieved 75 percent faster design and development of its Source Marts.* This time savings frees IS staff to focus on delivering value to its internal customers through analytics content, such as developing specific subject area marts or reports.

The alternative to using Source Mart Designer would be for IS staff to rely on a SQL import/export utility or an extract, transform and load (ETL) tool to move data. This approach requires multiple steps, including significant ETL work and testing each load on a subset of table rows. Once these steps were completed,
Catalyst’s EDW technology has proved to be very, very strong. It is exactly what we hoped it would be. The Source Mart Designer is essentially a frontend to what is an underlying set of Catalyst tools that allow us to have an ETL environment that really works. The ability to move data dependably and have our people trust that it’s the same as what it was in the source system is a big deal. There aren’t a lot of places that have as good of a standardized process for moving data consistently. We’re so far ahead of where we were and where a lot of people are.

Andrew Sehr, Program Analyst

Data architects would then have to watch and wait as the data copied from one database to the other. Significantly, this load process does not include the important step of denormalization. With the Source Mart Designer, data architects simply determine how they want their columns set up. They can then be assured that the load will occur overnight during the nightly ETL process.

Features of particular benefit to speeding IU Health’s process are:

- Automated upload of common source system fields
- Column name suggestions that can be easily modified
- Ability to determine if a data load is full or incremental based on the volume of data and speed requirements. If you choose an incremental load, you can identify what fields you want to valid as shown in Figure 1 in the row/validation section in the right column.

![Figure 1](image)

**WELL-STRUCTURED, CONSISTENT DATA FIELDS WITHIN THE EDW**

Source Mart Designer has enabled IU Health to maintain well-structured, consistent data fields within the EDW — regardless of how many staff members are involved in the process of moving the data. Specifically, IU Health staff have benefitted from the following features:

- Source Mart Designer automatically imports the source schema, data type, columns and more. On the right side of the user interface, staff can view and edit auto-populated destination information as illustrated in Figure 2. The destination type is calculated based on the original source type. Staff can also quickly identify and select the fields that will be loaded.

- The application largely eliminates typing from the load process. Relying on point-and-click mouse work rather than typing helps reduce errors and inconsistency.

- Source Mart Designer recommends the best data type for the data moving into the data warehouse. Data types are converted from the native source
system type to the native types of the enterprise data warehouse. These conversions maintain consistency with the source system.

The application maintains consistent naming taxonomy guidelines. Extract, transform, load (ETL) processes initiate only after the data architect has reviewed and confirmed taxonomy guidelines and sets the status to active.

Figure 2

AUTONOMY OF DATA ARCHITECTS WHILE ENSURING ENTERPRISE SUPPORTABILITY

Standardizing data loads using the application has enabled data architects to work autonomously while still ensuring the EDW maintains a uniform structure. An easy-to-use graphical interface helps staff maintain enterprise-wide consistency. This consistency helps ensure that the ETL code produced is supportable by the entire team. The application also allows data architects to quickly view how fields were previously named and the usage of those fields within the EDW.

IU Health’s IS leaders cite ease of use as a principal benefit of Source Mart Designer. Online training and tutorials are available for the application and IU Health estimates that a data architect can be fully trained on the application in less than two hours.

IMPROVED ANALYSIS THROUGH THE USE OF METADATA

Many enterprise data models contain no metadata – making it difficult, if not impossible, for a data architect to know where the column originated in the source system. Source Mart Designer captures as metadata the mapping of source system tables and columns to their Source Mart equivalents. Metadata — captured for every field — maintains continuity and traceability between the source system and the destination Source Mart.

This metadata enables analytics to be performed more quickly by providing clinicians with descriptive names. It defines how a source system’s data will look. With this knowledge, analysts who are using the data to drive clinical quality
improvement are better able to work with and query the data. Because metadata includes information about the origin of the data, data architects and analysts can quickly go to the source data itself if necessary. This ability is particularly important when resolving data quality issues.

Empowered IS staff

IU Health’s IS leaders commend Health Catalyst for empowering their team to do more with less while improving quality. Not only has the Catalyst team introduced ideas, processes and tools like the Source Mart Designer to data architects and other IS staff; Catalyst has also worked with IS to incorporate the department’s culture and ideas and to help them own and manage their EDW applications.  

* Calculation based on 5 tables with 20 columns apiece, each containing an estimated 50 million rows

ABOUT HEALTH CATALYST

Based in Salt Lake City, Health Catalyst delivers a proven, Late-Binding™ Data Warehouse platform and analytic applications that actually work in today’s transforming healthcare environment. Health Catalyst data warehouse platforms aggregate and harness more than 3 trillion data points utilized in population health and ACO projects in support of over 22 million unique patients. Health Catalyst platform clients operate 96 hospitals and 1,095 clinics that account for over $77 billion in care delivered annually. Health Catalyst maintains a current KLAS customer satisfaction score of 90/100, received the highest vendor rating in Chilmark’s 2013 Clinical Analytics Market Trends Report, and was selected as a 2013 Gartner Cool Vendor. Health Catalyst was also recognized in 2013 as one of the best places to work by both Modern Healthcare magazine and Utah Business magazine.

Health Catalyst’s platform and applications are being utilized at leading health systems including Allina Health, Indiana University Health, Memorial Hospital at Gulfport, MultiCare Health System, North Memorial Health Care, Providence Health & Services, Stanford Hospital & Clinics, and Texas Children’s Hospital. Health Catalyst investors include CHV Capital (an Indiana University Health Company), HB Ventures, Kaiser Permanente Ventures, Norwest Venture Partners, Partners HealthCare, Sequoia Capital, and Sorenson Capital.

Visit healthcatalyst.com, and follow us on Twitter, LinkedIn, Google+ and Facebook.