CP-Nexus: A Clinical Data Warehouse at Columbia-Presbyterian Medical Center

Philip Burrowes, Columbia-Presbyterian Medical Center, N.Y., N.Y.
Jason Oliveira, Columbia-Presbyterian Medical Center, N.Y., N.Y.

ABSTRACT

Clinical trialist, healthcare providers, hospital administrators and clinical researchers have a need to access various types of clinical and administrative data. The current multiple sources of clinical and administrative data require extensive programming by information system professionals to extract, integrate and massage. This introduces delays in delivering information to those who need it. There does not exist now at Columbia-Presbyterian Medical Center (CPMC), an integrated data source to facilitate clinical and administrative data analysis. One of the greatest challenges for an organization, especially an academic medical center, is to provide the means for anyone, at every level of the organization, to access the accurate and timely data necessary to support effective decision making, research and process improvement. CP-Nexus is just such an integrated data warehouse that provides a data resource to any application or anyone that needs it. CPMC was able to accomplish this innovative data solution with a combination of the SAS development system, the SYBASE SQL Server™ DBMS, and a robust client/server architecture.

INTRODUCTION

The Columbia-Presbyterian Medical Center (CPMC) is part of the New York and Presbyterian Hospitals Care Network, an integrated delivery network which provides a complete range of services encompassing the continuum of care. The shift in healthcare delivery and reimbursement to a managed care model has created a need to contain costs while maintaining or improving clinical quality. A prime requirement to support this new healthcare business model is complete, accurate, timely and longitudinal patient care data. The current multiple sources of data require extensive programming by decision support systems professionals to extract, integrate and present data. This introduces delays in delivering information to those who need it.

An integrated data warehouse provides the data infrastructure that will drive the next era of healthcare data analysis. Complex multi-dimensional queries crossing financial, administrative and clinical domains will be performed to support broad categories of healthcare analysis. The next era of healthcare analysis includes quality of care evaluation, patterns of variation in service, utilization studies, outcomes analysis and disease management. CPMC has successfully constructed an integrated data warehouse, which we have named the CP-Nexus database. CPMC utilized various ‘best-of-breed’ information system components to accomplish this task. The SAS® System and its various components, namely SAS/ACCESS®, SAS/SHARE® and SAS/CONNECT® were utilized in combination with the Sybase Version 11 SQL Server and a robust Client/Server architecture.

The building of the CP-Nexus data warehouse represents the first phase of an integrated healthcare data warehouse solution. The next phase will involve an application and query tool layer to more easily transform the tens of millions of data rows into actionable information. This data exploitation layer will involve a collection of target subject applications, guided queries, report building, Web enabled data queries and client based analysis tools. The SAS System and its SAS/AF®, SAS/EIS®, SAS/SQL® and SAS/English® modules will play a significant role in designing the data exploitation layer.

BUSINESS IMPERATIVE

An escalation of healthcare expenditures and the general questioning of the value of a healthcare dollar has created a need to contain costs while maintaining or improving clinical outcomes. The shift in healthcare delivery and reimbursement to the managed care model has created a need to more closely track and analyze clinical and financial data related to patient care services. When a fee for each service provided was more pervasive, there were few
if any clinical guidelines for when specific procedures or ancillary tests should be used. Therefore, expensive technologically advanced tests or treatments would be used without assessing if the expected benefits justified the increased costs. As reimbursement schedules began to change to fee-per-episode (i.e. DRGs), and more recently, fee-per-person (i.e. capitation), clinical managers began to focus on total cost per clinical outcome.

There are point of service clinical decision support systems designed to tackle some of the complexities in delivering healthcare and attempt to reduce variation by providing knowledge for benchmarking, preventive actions, clinical pathways and outcomes research. These systems use the most up-to-date clinical evidence to relate tests and treatment to patient outcomes. One such system at The Presbyterian Hospital is the Clinical Event Monitor. The CEM provides on-line alerts in the form of faxes, e-mail and on-line flags to alert clinicians to abnormal test results, drug-drug interactions and sentinel events. The CP-Nexus clinical data warehouse is however designed for more retrospective decision support.

Aggregates of data can be reviewed to identify patterns of behavior in treating patients with specific complaints or diagnoses, adjusting for risk-severity and other factors known to effect outcomes beyond the control of the clinician. A competitive advantage can be gained by continuously improving patient and cost outcomes, using knowledge gained by analyzing quality, resource utilization and other factors of care.

Barry Devlin in his informative book titled Data Warehouse, from Architecture, to Implementation describes the new paradigm of information-based management that CPMC and its Decision Support Systems team was finding an increased need to create. It outlines a data architecture designed on several premises: 1) A single source of information which is fed by the numerous raw data sources which are cleansed and reconciled to ensure its quality and integrity. This reconciled information is the single, ultimate source for information based management. 2) Information is designed and presented in the contexts of business driven needs, not technological expediencies. 3) Information management is automated in terms of its accumulation, extraction and delivery. 4) Information quality and ownership promotes the stewardship of one of the most vital assets of any company; its information.

With business driven and information hungry decision makers on one side and data management complexities on the other, CPMC decided to embark on the creation of a data warehouse to support the next era of healthcare data analysis. One outcome of this initiative is the CP-Nexus clinical data warehouse. A data warehouse is simply: A single, complete, and consistent store of data obtained from a variety of sources and made available to end users in a way they can understand and use in a business context.

**CP-NEXUS DESIGN**

*Data Mart versus Data Warehouse*

The authors must apologize at this point for promulgating a common misuse of terminology, that is, a data mart versus a data warehouse. The data warehouse is, as stated, a single reconciled store of data that acts as the ultimate source for information based management. A data mart, however, is a more
targeted application of data whose source is the corporate data warehouse. In this case, CP-Nexus is the targeted application of data from CPMC’s two data warehouses, one administrative and one clinical in nature. CP-Nexus is a subset of financial, administrative and clinical data from the warehouses that is designed to support retrospective clinical research and process improvement activities. We will use the proper terminology from here on to delineate the various data stores implemented at CPMC.

**Hardware & Software Architecture**

The CP-Nexus data mart hardware architecture consists of a Sun® System SPARC 1000E series with four parallel SMP processors. It has sixty-four gigabytes of disk storage to enable redundant arrays of inexpensive disks (RAID). Twenty-four gigabytes have been reserved for the CP-Nexus data mart and the remainder reserved for various clinical department research applications. There are two fourteen gigabyte 8mm tape drives and a QIC-150 tape drive.

Sybase Version 11 SQL Server is being used as the back-end data engine. Sybase was selected for its indexing capabilities, high performance, fast load times and its open architecture allowing for a variety of front-end application options.

The SAS System has been utilized effectively at CPMC for its data exploitation and statistical modules. It is now also being used as an integration tool. The SAS System Version 6.09 is implemented on two IBM® 9021 and IBM 3090 mainframes under MVS. The administrative data warehouse, previously designed several years ago, is implemented under the SAS architecture on the mainframes. The SAS System for Solaris/UNIX is implemented on the SPARC platform and is utilized extensively by the clinical researchers and their data administrators.

**Client/Server Architecture**

CPMC has taken advantage of the robust client/server architecture that exists at the institution for this project. The architecture is built upon a network that comprises well over 30 4/16 MIP Token Ring segments, a FDDI ethernet backbone, terrestrial microwave inter-campus links, T1 inter-campus links and an ATM WAN backbone. We are easily one of the most networked academic medical centers in the country.

CP-Nexus utilizes a distributed data access and distributed function model for client/server computing (see figure 1). The SAS system has allowed us to distribute processing of data extraction, cleansing, transformation and transport utilizing the SAS/ACCESS, SAS/SHARE and SAS/CONNECT modules. The presentation level resides on WINTEL clients using the Solaris/SAS System through a XWindow emulator. Extraction and transformation logic resides on the mainframes using the MVS/SAS System. The two processing resources are integrated using the SAS/CONNECT module. Data access is distributed to both the SPARC and IBM mainframes through the SAS/ACCESS modules. SAS/ACCESS allows us to read data from the SPARC Sybase V11 based CP-Nexus data mart, from the MVS SAS based administrative data warehouse, and from the MVS/DB2 based clinical data warehouse.

![Figure 1: Gartner Group Client/Server Model](image)

**Data Flow**

When a patient arrives at the Hospital, he or she is first registered in a patient management system supplied by American Healthcare’s Eagle 2000 HIS®, which manages ADT, census and billing processes. During the care process, data is captured in a number of other systems such as pharmacy, radiology, operating room and laboratory systems.
The administrative and financial data is warehoused into what we have named the V6 data warehouse. Clinical data is warehoused into the CIS Clinical Data Repository (CDR). The legacy source data resides on a combination of IBM IMS, VSAM and DB2 databases. The V6 data warehouse is a version 6 SAS database and CIS is a DB2 database. Please refer to figure 2 depicting the data flow of decision support activities.

Health Level 7 electronic data interchange messages are used to ‘trigger’ patient accounts with activity that require refreshing in the V6 and CIS data warehouses. The HL7 ‘triggers’ further drive the various data mart population processes, including CP-Nexus. For more information regarding the HL7 data interchange standard, please refer to “A Table Driven SAS Parser for Health Level 7 Healthcare Data Interchange” by Jason Oliveira and Thomas Frenkel in the NESUG 1996 proceedings.

CONCLUSION

One of the greatest challenges of any organization, especially an academic medical center, is to provide the means to allow anyone, at every level of the organization, to access the accurate and timely data necessary to support effective decision making, clinical research and process improvement. An integrated data source, otherwise known as a data warehouse, provides the generalized data resource for any application or anyone who needs it. The Decision Support Systems group at Columbia-Presbyterian Medical Center has leveraged the capabilities of the SAS System, Sybase SQL Server and robust client/server architecture by creating the CP-Nexus clinical data mart. This powerful symbiosis has allowed us to deliver the right business/clinical information, to the right decision maker, on the right computing platform, utilizing the right data exploitation and visualization tools. The data mart has supported cost effectiveness studies for Left Ventricular Assist Devices (LVADs), clinical trials for Syncopy and departmental reviews for numerous clinical indications. These uses of the data only scratch the surface of the possibilities. CPMC hopes to utilize this powerful integrated data source to support disease management, outcomes analysis, JCAHO “ORYX” initiatives and critical pathway development. CP-Nexus will quickly become a strategic tool supporting information based management as we strive to provide clinical value to our patients, and financial value to healthcare purchasers.

Contact

Jason Oliveira
Manager, Columbia-Presbyterian Medical Center
161 Fort Washington Avenue, NY NY 10032
(212) 305-2945
oliveir@phadm1.cis.columbia.edu

Philip Burrowes
Programmer/Analyst, CPMC
(212) 305-3801
burrowe@cpmc3.cpmc.columbia.edu

Homepage
http://www.cpmc.columbia.edu/ais/resources/ic/

SAS, SAS/ACCESS, SAS/AF, SAS/CONNECT, SAS/EIS, SAS/ENGLISH, SAS/SHARE, SAS/SQL software are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.
© indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.