



NATIONAL ASSOCIATION OF
Community Health Centers



RSM! McGladrey

Data Warehousing Part 2: *Implementation & Issues*

Presented by:

David Hartzband, D.Sc.
Director, Technology Research
RCHN Community Health Foundation

Matthew Grob, CPHIMS, FHIMSS
Director & Practice Leader, Health IT
RSM McGladrey

NACHC FOM-IT Conference
November 11, 2009



Agenda

- Review
- Technical Planning
- Designing the Data Model
- Implementation
- Privacy, Security & Secondary Use
- Non-Technical Issues
- Uses of the Data Warehouse
- Future Developments
- Questions

Review

1. A Data Warehouse is an extract of an organization's data — often drawn from multiple sources — to facilitate analysis, reporting and strategic decision making. It contains only alpha-numeric data, not documents or other types of content.
2. Data Warehouses have different value for patients, providers, populations & organizations
3. Planning must be done with respect to what you want to accomplish
4. Scale implementation both technically & organizationally with respect to what you want to accomplish

Technology is not the Goal!



- Planning a Data Warehouse is **NOT** a technology project
- it must fit into the business & clinical goals of your organization

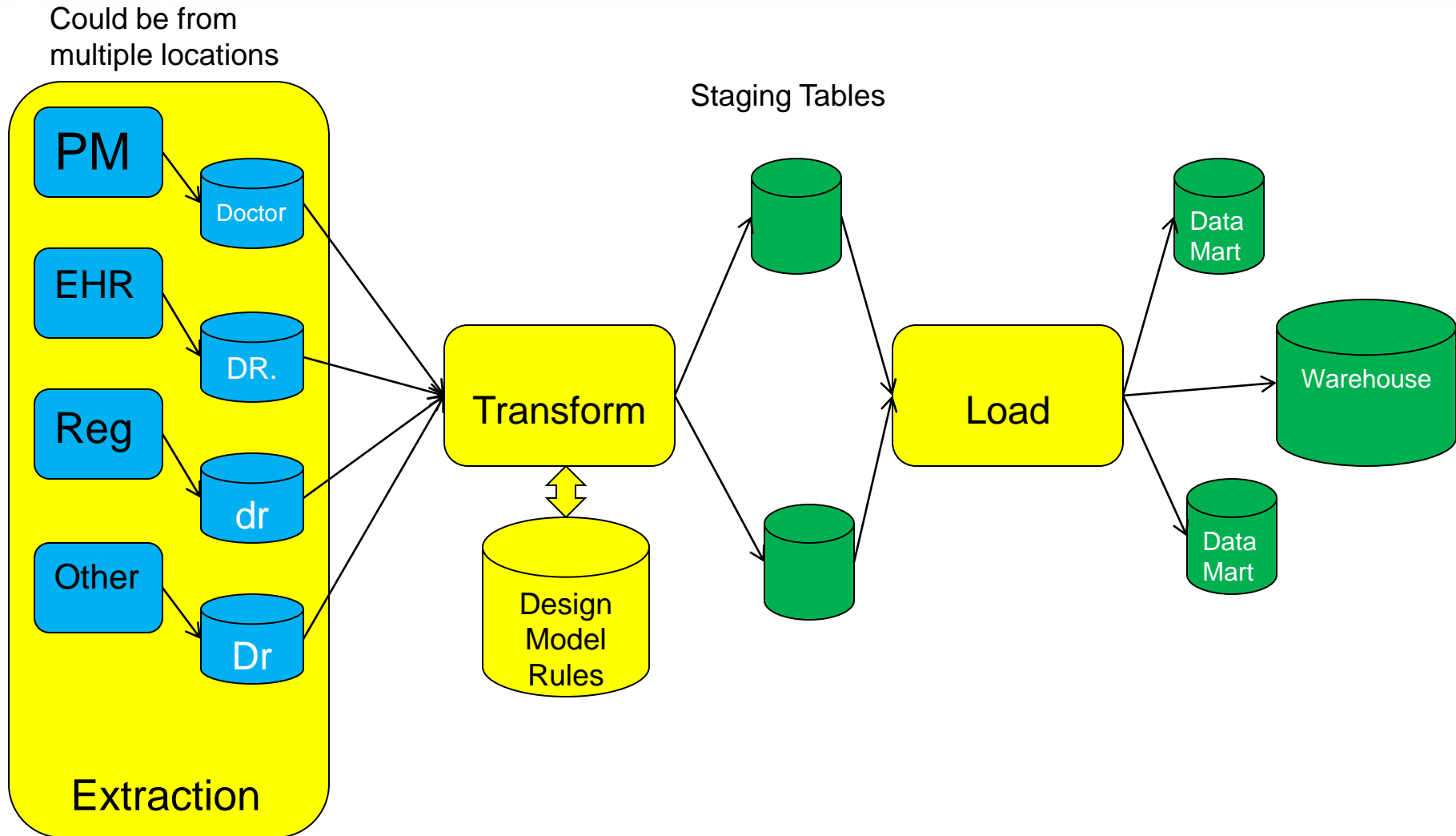
Technical Planning - 2

- Now for the technology:
 - Infrastructure:
 - Network (100MBPS minimum to 1 GBPS) connecting all participating organizations
 - Servers – 1 Ghz processor, 2-4 GB virtual memory, 2 database servers (1 for application data, 1 for the ETL application)
 - Associated scalable disk storage, DW size +30-40% for growth
 - Portfolio of software applications (in addition to the DW/ETL software)
 - HIT apps
 - Commercial relational database
 - Analysis & reporting tools
 - This type of infrastructure & software portfolio usually requires an IT Staff

Data Model Design

- Doctor, DR., DR, Dr., Dr, dr., dr, MD, md all map to Dr.
- A data model defines & maps all the forms that data take in the real world & across different systems
- Data is **E**xtracted from different sources (PM, EHR, registries, etc.) & potentially different organizations
- It is **T**ransformed according to rules defined in the data model
- It's then **L**oaded into the data warehouse
- This is what the **ETL** tool does

Data Model Design - 2



Data Model - 3

- Top-Down design requires that every data item that you want to analyze be defined along every dimension you want to analyze & every relationship it has with other data items
 - Cost: cost-per-item, cost-per-use, cost-per-day, cost-per-week, etc.
 - This is a huge effort & is usually only undertaken by large organizations with deep pockets
- Bottom-up design defines a set of transform & relationship rules with respect to a specific business or clinical problem.
 - A data mart is then loaded to be used for this problem
 - Multiple data marts are defined & loaded for strategic business & clinical problems
 - These designs are analyzed for similarities & data marts are combined where possible
 - This is the approach most often used (unless you are the Government or Goldman Sachs)

Additional Design

- Two more design tasks:
 - Data cleansing
 - Rule set written to eliminate duplicates, obvious errors, ambiguities and/or contradictions
 - Applied to transformed data before Load
 - Query development
 - Linked to Key Performance Indicators
 - May also be linked to regional or Federal reporting needs
 - UDS
 - Epidemiology, Population statistics
 - Other

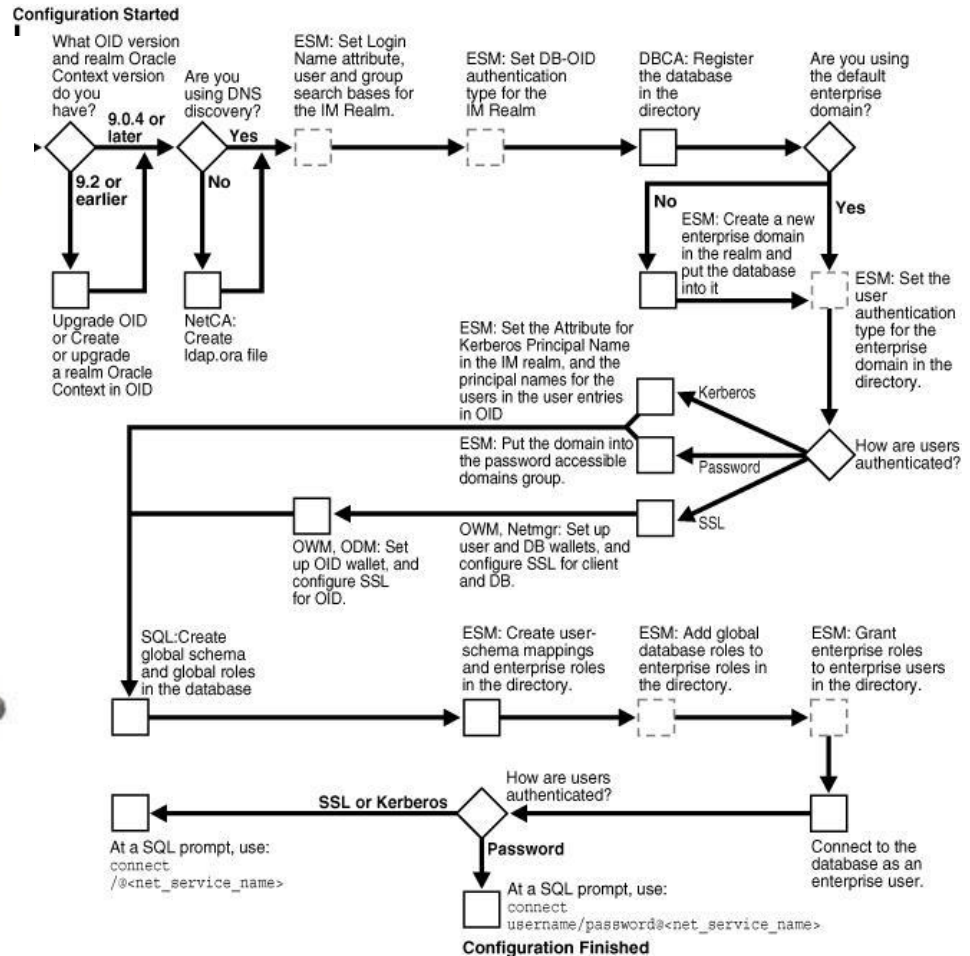
Implementation

- Generally requires selecting a vendor who will:
 - Evaluate infrastructure & software portfolio
 - Assist in data model design & query development
 - Deploy hardware & software (or host)
- HCCN or PCA may provide leverage for single vendor selection across the network or association
 - Already may provide hardware & software portfolio
 - Might also already aggregate data from members
 - Makes many of the technology (& some of the organizational) issues easier to address

Implementation - 2

- Design & deployment of a data warehouse is real work
- Even with a vendor assisting, your staff (not just your IT staff) will have months of work that you will have to plan for
- It may take as much as 2 years to go from your decision to develop a data warehouse until you are routinely producing reports with it

Privacy, Security, Secondary Use



Privacy

- A data warehouse may contain huge amounts of protected health information (PHI) – patient information that cannot be disclosed except under specific circumstances
- Most data is mined & analyzed in aggregated & deidentified form
- Actual PHI can only be shared if ‘business associates agreements’ are in place among all covered entities that participate in the data warehouse
- PHI can be disclosed between covered entities:
 - For treatment purposes
 - If both entities have relationships with the patient
 - For health operations purposes (including quality efforts)
 - To authorized public health entities for a specific public health purpose
 - For research purposes only with patient consent

Security

- The data in a data warehouse is only useful if it can be used, but this requires effective security be deployed
- HIPAA privacy rule requires that administrative, technical & physical security to ensure the privacy of PHI & to reasonably safeguard PHI from intentional or unintentional disclosure
- HIPPA security rule requires that covered entities protect against reasonably anticipated threats to the security of electronic PHI, that their workforces adhere to the Security rule & that their contractors (like HCCNs & HIEs) assure compliance through business associate agreements

Secondary Use

- The large amounts of data in a data warehouse are attractive to many parties for a variety of purposes
- PHI is generally not appropriate for secondary use disclosure unless it is deidentified
- The stewardship of PHI is a primary issue for participants in a data warehouse. Policy on privacy, security & secondary use must be developed before issues come up
- Secondary use must be addressed in the business associates agreements among covered entities & with their contractors – HCCNs, PCAs, HIEs

Culture Matters!



Non-Technical Issues

- Culture?
 - Use of a data warehouse requires substantial commitment to transparency & data sharing across all participants
 - Not all participating organizations may have that commitment
 - Some queries may reveal differences among organizations
 - Differences in adherence to quality criteria
 - Differences in outcomes
 - Etc.

Non-Technical Issues - 2

- Readiness
 - Not all organizations may be at the same level of preparation:
 - Training
 - Use of electronic technology
- Lack of Executive/Board commitment
 - Nothing ensures failure of a project that requires such a large investment than lack of support from the management or Board of a health center, HCCN or PCA
- Lack of agreement on:
 - Appropriate level of transparency
 - Assignment of liability
 - Data stewardship



Current Use

Future Developments



Uses of the Data Warehouse

- Development of new or more productive reports for:
 - Quality evaluation
 - Registry development
 - Epidemiologic & public health efforts
 - Etc.
- Larger scale & novel analysis for
 - Improvement of clinical outcomes
 - Improvement of business results
- Provision of deidentified data for
 - Research
 - Quality comparisons
 - Etc.

Future Developments

- Search
 - Becoming more & more important as even results sets become very large
 - New forms of search including numeric, pattern matching, context based etc.
- Visualization
 - Visual representation of analytic results makes them more understandable
- New organizational forms
 - Virtual &/or actual organizations based on business or clinical similarities revealed by analysis
 - May provide new/different opportunities for improvement of clinical outcomes or business results

Questions



Matthew Grob
(212) 372-1606
matthew.grob@rsmi.com

David Hartzband
(617) 324-6693
dhartzband@rchnfoundation.org