

Putting a STAR on the Walkway of Higher Education Data Warehousing

Session #25500
March 13, 2008

Alliance 2008 Conference
Las Vegas, Nevada

Building Next Generation Data Marts at Cornell University



Presented by:

- **Jim Singleton – Cornell University**
- **Jeff Christen – Cornell University**
- **Ashley Silverburg – Phytorion, Inc.**
- **Yiorgos Marathias – Phytorion, Inc.**

Cornell University



Cayuga Lake taken from the Cornell University Bell Tower

- Main Campus is in Ithaca, NY
- Founded 1865
- Both a State & Private Institution
- Undergrad enrollment is 13,515
- Graduate enrollment is 5,932
- Faculty members = 2,633
- Staff employees = 11,236



Phytorion

- Full enterprise data warehouses
- Area-specific data marts
- Operational & Strategic content
- Integration of any source systems
- Fully custom approach as well as packaged data marts



PHYTORION

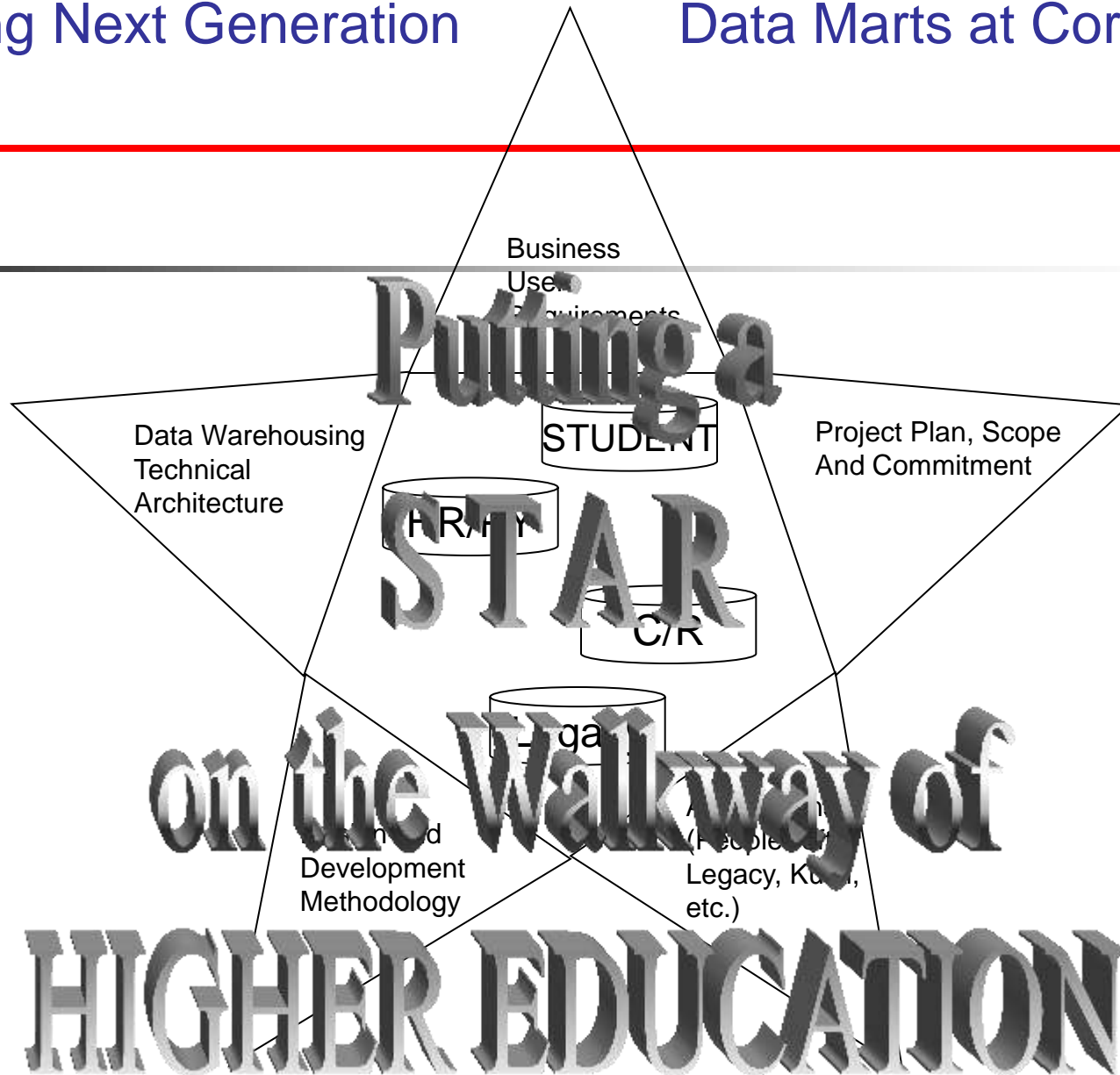
DATA WAREHOUSING BUSINESS INTELLIGENCE

The Cornell Enterprise Data Warehouse



A collection of data that can be defined and shared across the whole University by using common definitions.

began ~ August 2006





Overview

Cornell University has chosen and is implementing a path towards an Enterprise Data Warehousing solution. This strategy involves:

- Using the Kimball Methodology to manage the project lifecycle along with developing Dimensional Models (Star Schemas) for new Data Marts,
- Utilizing the mature infrastructure and resource with Cornell Information Technology,
- Utilizing both Internal Resources and an External Data Warehousing Company, Phytorion for new data marts and when re-engineering existing data stores;
- Delivering data marts in support of new Operational Application roll-outs.



Current EDW Strategy

- Design and build new datamarts and re-engineer existing data-stores using:
 - A "*business questions to be answered*" approach. Focusing on Customer Needs and Data Requirements.
 - Keep requirement sessions open to all development efforts / teams.
 - Meetings
 - Documentation
 - Models
 - Training
 - Data Governance
 - Keep design strategies / implementation open to all development efforts / teams. (for example)
 - Incremental loading
 - Security
 - Metadata Management
- Stay on Schedule, or modify schedule when needed.





Phytorion and Cornell Experience

- Successful Partnership
- Flexible
- Easy to work with
- Focused on Delivery



Dimensional Marts in Production

- CR
- Student
 - Prospect
 - Admissions
 - Financial Aid
 - Campus Community
 - Student Records

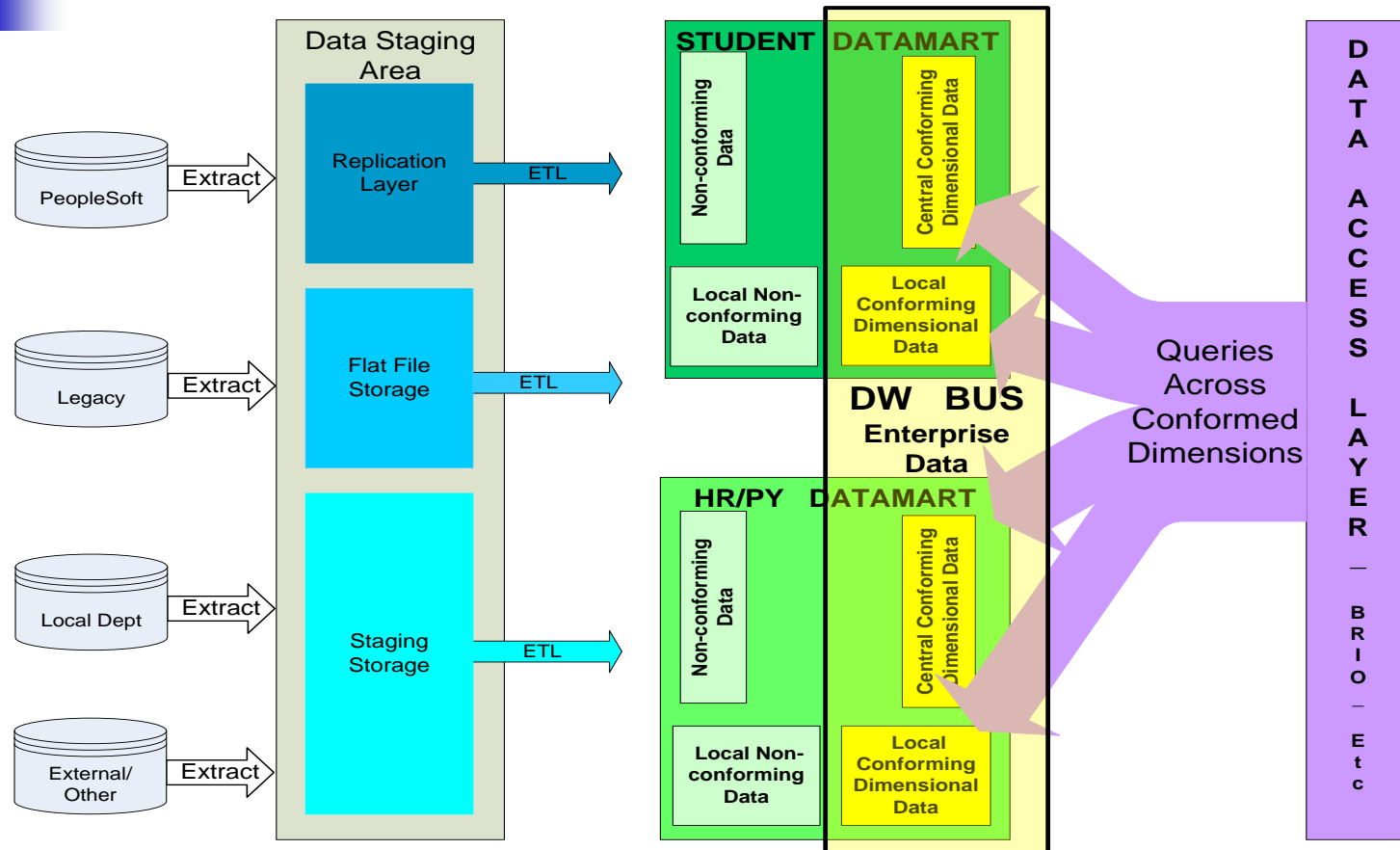


Dimensional Marts in Development

- Student Financials (May 2008)
- Human Resources (July 2008)
- Payroll (Sept 2008)
- Benefits (December 2008)
- Human Resources / Payroll – Non-PS (Jan 2009)
- Kualu / ADW (Accounting Data Warehouse) (2009+)

Riding the Bus at Cornell

Basic Elements of BUS Data Warehouse Proposed for Cornell





Project Plan, Scope and Commitment

- Set High-level Scope and Plan,
- Identify Sub-Project,
- Get Sub-Project Sponsor,
- Set Milestone Dates;
- Assign Project Manager.



Project Management

- Uses the Cornell Project Management Methodology (CPMM).
- Follows the Cornell Project Initiation Process (PIP) for getting an approved University Project / Sub-Project.



Current Warehouse Environment

- Servers: Solaris & Windows
- RDBMS: Oracle 10g R2 using RAC
- ETL Tool:
 - DataManager used for PeopleSoft based marts
 - Legacy marts using PL/SQL, Korn Shell Scripts, etc.



Current Warehouse Environment

continued

- Reporting Tools: Hyperion Brio
 - Brio v8 – Insight & Explorer
 - Custom Web interfaces written in Cold Fusion
- Peoplesoft RDS (Reporting Data Service)
 - HR / Payroll & Contributor Relations
- Phytorion Dimensional Marts
 - Student & new HR / Payroll
- Legacy: GL, Budget, Sponsored Programs



Data Warehouse Technical Architecture

- Cornell Warehouse Environment,
- Consistent and Reliable Source,
- High Data Availability,
- Load Monitoring and Notification,
- High Database Availability,
- System Availability Monitoring,
- Performance Monitoring and Tuning;
- Security Management.



High Data Availability

- A copy of the warehouse data is available to the users at all times.
- Warehouse is fully available during load processes.
- Warehouse administrator may rollback load process in the event of faulty load.



High Data Availability-DM Tools

DMTools is a Data Warehousing infrastructure management tool developed and in use by Cornell University.

- Allows high data availability – 24X7 access
- Repository driven
- Manages loads
- Toolbox written in Oracle PL/SQL (O.S. independent)
- GUI console to manage load related metadata
- Available, as free download, through JA-SIG Clearinghouse www.ja-sig.org



Security Management

- Automation of user & role management
- Peoplesoft Based Marts
 - Role, & Row and Column, and Field Level Security defined in Peoplesoft database
- Non-Peoplesoft Based Marts security mgt.
 - Account creation & removal
 - Role management



Business User Requirements

Requirements are gathered through many sources and ways. Engaging the “Right” Customer is Key.

Examples include:

- Interviews
- Reports
- Existing Applications / Documentation



Business User Requirements

Requirements are gathered through many sources and ways. Engaging the right Customer is Key.

Examples include:

- Interviews
- Reports
- Existing Applications / Documentation



Business Areas

- Undergraduate and Graduate Admissions
- Institutional Research and Planning
- Student Services
- Bursar
- Continuing Education
- Accounting
- Financial Aid



Colleges We Met With

- Engineering
- Law
- Business
- Natural Resources
- Veterinary
- Hotel Management
- Library Science



What We Deliver

- Prospects and Admissions
 - Application
 - Athletic Participation
 - Academic Interests
 - External Academic Data
 - Honors and Awards
 - Prior Work Experience
 - Prospect Data
 - Test Scores



What We Deliver

- Student Records
 - Advisors
 - Courses and Classes
 - Enrollment and Grades
 - Enrollment Appointments
 - Milestones
 - Programs and Plans
 - Student Degrees
 - Student Groups
 - Transfers and Test Credits



What We Deliver

- Student Financials
 - Accounting Line
 - Credit History
 - General Ledger
 - Financial Items
 - Student Accounts
 - Tax Data



What We Deliver

- Campus Community
 - Checklists
 - Comments
 - Communications
 - Bio Demo Data
 - External Organizations
 - Events and Meetings
 - Service Indicators



What We Deliver

- Financial Aid
 - Financial Aid Applications
 - ISIR
 - Awards and Disbursements
 - Item Types
 - Student Budgets
 - Promissory Notes
 - Loans
 - Pell Grants

PERSON_DATED_DATA_DM
PERSON_DATED_DATA_SID

PRIM_PROG_PLAN1_DM
PRIM_PROG_PLAN1_SID

PRIM_PROG_PLAN2_DM
PRIM_PROG_PLAN2_SID

ALT_PROG_PLAN1_DM
ALT_PROG_PLAN1_DM

ALT_PROG_PLAN2_DM
ALT_PROG_PLAN2_DM

LAST_SCHOOL_ATTENDED_DM
LAST_SCHOOL_ATTENDED_SID

APPLICATION_ATTRIBUTES_DM
APPLICATION_ATTRIBUTES_SID

APPLICATION_FC
PERSON_DATED_DATA_SID
PRIM_PROG_PLAN1_SID
PRIM_PROG_PLAN2_SID
...
LAST_SCHOOL_GPA
LAST_SCHOOL_CLASS_RANK
LAST_SCHOOL_CLASS_SIZE
ACADEMIC_INDEX
APPLICATION_COUNT_FIELD
COMPLETED_APP_COUNT_FIELD
WITHDRAWN_COUNT_FIELD
ADMITTED_COUNT_FIELD
ADMITTED_PRIM_COUNT_FIELD
ADMITTED_ALT_COUNT_FIELD
ENTERING_PRIM_COUNT_FIELD
ENTERING_ALT_COUNT_FIELD
ENTERING_COUNT_FIELD
DENIED_PRIM_COUNT_FIELD
DENIED_ALT_COUNT_FIELD
DENIED_COUNT_FIELD
RUN_DATE
CURRENT_FLAG
SNAPSHOT_FLAG

RECRUITER_DM
RECRUITER_SID

ADMIT_TERM_DM
ADMIT_TERM_SID

REGION_DM
REGION_SID

SNAPSHOT_DATE_DM
SNAPSHOT_DATE

SPECIAL_APP_INDICATORS_DM
SPECIAL_APP_INDICATORS_SID

PRI_PROG_ACTN_STAT_RSN_DM
PRI_PROG_ACTN_STAT_RSN_SID

ALT_PROG_ACTN_STAT_RSN_DM
ALT_PROG_ACTN_STAT_RSN_SID



Design and Development Methodology

- Data Modeling – Designing a STAR,
- ETL Delivery Techniques;
- Data Delivery - Brio.



Data Modeling - Designing a STAR

For Dimensional Data Modeling Activity:

- Decision Matrix
- Bus Matrix
- Object Definition Document (Dimensions / Facts)
- Source Data Model

Decision Matrix (Applicant Example)

Cornell University -- Applicant Decision Matrix

Objectives	Decisions	Type of Info	Grad/Pr	CESS	Colleges	Questions	Associated To-Be(s)	Readiness	Description	In Scope/Out Scope	Priority
	Waive Requirements	List				Which international applicants should have TOEFL waived	Create Applicants	now			
	Determine if we are	Count/%	X		X	Application Funnel: Application	Create Applicants	now	Note: multiple business units		
		Count/List/%	X		X	What is the ethnicity breakdown of the applicant pool?	Create Applicants (all)	now			
		Count/List	X		X	What is the international applicant breakdown (based on visa type, citizenship) by region?	Create Applicants (all careers)	now			
		Count/List	X		X	What are the admissions decisions for all applicants?	Create Applicants (all careers)	now			

Bus Matrix (Partial Example)

Functional Area	Business Process	Dimensions																							
		Date	Constituent	Inv Mbr Org	Role / Title	Session	Designation Brg	Account Tbl	Motivation Cd	Dept Cd	Status Checklist	Status Checklist Item	Employee	Job	Position	Term	Student	Course	Grad. Appointment	Academic Org	Account	Financial Org	Accounting Period	Grant	Funding Agency
Alumni Affairs & Development	Campaign & Proj. Management																								
	Monetary	X	X	X																	X	X			
	Workflow	X	X	X																					
	Goals	X	X	X																					
	Membership Management																								
	Membership Dues	X	X	X	X				X																
	Goals	X	X	X																					
	Gift & Pledge Management																								
	Recognition	X	X	X		X																			
	Designation	X	X	X	X	X	X	X	X							X									
	Tributes	X	X	X		X																			
	Matching Gifts	X	X	X		X	X																		
	Other																								
	Volunteer Participation	X	X	X	X																				
GL Account Reconciliation	X	X	X		X	X															X	X			
Pledge Write-Off Process	X	X	X	X	X	X	X																		
HRP	Employment Management																								
	Employment Transactions	X										X	X	X		X					X				
	Position Management	X										X	X												
Student	Enrollment Management																								
	Registration																X								
	Appoint Grad. Students																								
Finance	Course Management	X															X								
	Expense Management																								
	Expense Reporting	X																			X	X	X		
Research Administration	Grant Management																								
	Submission	X										X	X								X	X	X	X	X

Dimension Documentation

This dimension describes ethnic background. It includes descriptions for multi-ethnicity (both under-represented and non-under-represented). History is not maintained on this dimension.

Table Type : Dimension

Security:

Shared Dimensions

Technical Notes:

Add the following static values to the dimension: MLT_URM, Mult URM, Multi-ethnic Underrepresented Minority
MLT_NURM, Mult NonUR, Multi-ethnic Not Underrepresented Minority

Change History 30 July 2006 – Created

ETHNICITY_D

M

Attribute	SourceTable	SourceColumn	Sourcing Instructions	Description
ETHNICITY_SID			Populated by the ETL	
ETHNICITY	PS_ETHNIC_GRP_TBL	ETHNIC_GRP_CD	IGNORE SETID KEY, MAX(EFFDT), EFF_STATUS = 'A'. See technical notes above.	8-character code for the ethnicity
ETHNICITY_SDESCR	PS_ETHNIC_GRP_TBL	DESCRSHORT		10-character description of ethnicity
ETHNICITY_LDESCR	PS_ETHNIC_GRP_TBL	DESCR50		50-character description of ethnicity



ETL / Delivery Techniques

- Development / Migration Strategy
- Metadata used for:
 - Object Definition Document(s)
 - ETL Code/Build (DataManager)
 - Business Metadata Definitions
 - Unit Testing



Data Delivery - Brio

- Standard reports
- Dashboard
- Ad-hoc Reports

Brio Models:

- Single Fact
- Non-fact
- Galaxies and Constellations (monster model)



Questions ?

Thank You:

- **Jim Singleton – Cornell University**
js537@cornell.edu
- **Jeff Christen – Cornell University**
jrc42@cornell.edu
- **Ashley Silverburg – Phytorion, Inc.**
ashley.silverburg@Phytorion.com
- **Yiorgos Marathias – Phytorion, Inc.**
Yiorgos.Marathias@Phytorion.com

Cornell University Bell Tower