TurnerProject
Data Warehouse Architecture
Project Defintion & Plan

Prepared by:

Bill Turner
Data Warehouse Architect, DBA

851 Mountaintop Drive
Collegeville, PA 19426
484.686.4194
bill.turner@SkippackRidge.com

GETTING STARTED

Introduction

The purpose of this document is to layout the foundation for the analysis, design, and delivery of the **TurnerPDP TurnerProject Data Warehouse / Data Mart**. It also allows for the provision of post-project documentation.

First, I'd like to emphasize that we have targeted the deployment of a scaled down central Data Warehouse which will feed several smaller Data Marts. The central Data Warehouse utilizes a daily extract, transform, and load process (ETL) which aggregates data at the atomic transaction level from a limited set of data sources - primarily the CMS, OLTP, and QA OLTP DBMS's. Subsequently, this daily aggregation of data is input to "primer" routines which apply additional aggregate algorithms at weekly, monthly, quarterly, and annual measures of time. These re-aggregated values are housed in Data Marts and are available for end-user reporting.

The central Data Warehouse database design is comprised of a star schema structure representing the TurnerProject system's Business Reporting requirement needs. Metadata, which is a necessity for large scale Data Warehouses is being utilized in a limited capacity. Due to the fact that we have a single point of report preparation and delivery, our metadata repository is "light" in nature. Stove pipes are typically not a problem w/ our current design and deployment.

Utilizing custom developed SQL stored procedures, the Data Marts are populated w/ summary data. We could have bypassed the development of the central Data Warehouse and populated the Data Marts directly due to the fact that we've integrated a design which includes both conforming dimensions –and-unique atomic data across the Data Marts and Subject Areas.

Understand that as the DW project rolled out, we were able to capture approximately 50% of total reporting requirements and needs. This is typical in any new OLTP / Data Warehouse environment. Until the end users begin to use the system and become familiar w/ its offerings, the complete set of Reports will remain outstanding.

We are evaluating the use of a data warehouse "staging" database. This is optional and may be bypassed if determined that we'll be able to populate the central data warehouse directly from temporary storage housing data sourced from the TurnerProject OLTP DBMS's.

Regarding the capturing of historical data in the TurnerProject Data Warehouse, we have selected the use of Type 2 slowly changing dimensions allowing for the reporting (slicing and dicing) of any historical activity. This architecture is detailed in a later section of this document.

Project Roles Checklist

Front Office	
Business Sponsor	Sharon Smith - TurnerProject
I/T Sponsor	Al Smith
Business Drivers or Steering Committee	DW "Core" Team
Coaches	
Project Manager	Sheldon Smith
Technical Project Manager	Laurie Smith
Regular Lineup: Core Project Team	
Business Systems Analyst	Mark Smith
Data Modeler	Bill Turner
D/W Database Administrator (DBA)	Bill Turner
Data Staging System Designer	Bill Turner
End User Application Developers	Mark Smith, Frank Smith
Data Warehouse Educator	Bill Turner
Special Teams	
Technical/Security Architect	Bill Turner
Technical Support Specialists	Bill Turner, Mauro Smith
Data Staging Programmers	Bill Turner
Data Steward	Bill Turner
D/W Quality Assurance Analyst	Frank Smith
Fans	
Business Users by Group/Function	DW Core Team (point person)

Project Scope

Project Background
The TurnerProject Data Warehouse project is a sub-project nested w/in the overall CMS reengineering effort. The TurnerPDP TurnerProject Development Group has been selected as the prime developer responsible for supplying TurnerPDP w/ the revised, web-based software along w/ custom enhancements identified by the TurnerPDP CORE team (via the FSDS ~ functional system design specifications)

The TurnerProject Client Reporting group is responsible for phasing in the the back-end Client Reporting architecture ~ utilizing current technology and Data Warehouse / Data Mart design techniques. The full range of Data Warehouse tasks will be executed by the TurnerProject Client Reporting team.

Scope Definition
Based on requirements gathered from the business, the TurnerProject Client Reporting team has prepared a project plan w/ a finite list of deliverables. These deliverables include:

- Project Planning
- Business Requirements Definition (sourced from the Business Report Team led by Mark Smith)
- Technical Architecture Design
- Product Selection & Installation
- Interim Phased Development
- Dimensional Data Modeling
- Physical Design
- Data Staging Design/Development
- Deployment
- Maintenance & Growth

Phased Approach
- Phase 1 - Design & Production of **OLTP Standard** Reports
- Exclusions - CMS Standard & Enhanced Reports, QA Standard and Enhanced Reports, Data Warehouse & Data Mart architecture

- Phase 2 - Design & Production of **CMS Standard** Reports
- Exclusions - QA Standard and QA Enhanced Reports, Data Warehouse & Data Mart architecture

- Phase 3 - Design & Production of **QA** Reports
- Exclusions - Data Warehouse & Data Mart architecture

- Phase 4 - **Report Automation**
- Exclusions - Data Warehouse & Data Mart architecture

- Phase 5 - **Static Portal**
- Exclusions - Data Warehouse & Data Mart architecture

- Phase 6 - Build **DW for Standard Client Reports**
- Exclusions – n/a

- Phase 7 - Build **DW for Supplemental Client Reports**
- Exclusions – n/a

Project Risks and Risk Reduction Plan
- Hardware and Software Availability
- Business Requirements Document including a complete list of By's and Measures –and- sample report layouts / requirements
- Utilizing OLTP as direct data source for Interim Result Set Build (Phases 1 – 5)

Deliverable's Acceptance

User Acceptance

Project: TurnerProject Client Reporting & Data Warehouse

Executive Sponsor: Sharon Smith, - TurnerProject

Business Driver:

Project Manager Sheldon Smith

Deliverable Description:

Accepted:

Approval Signature	Project Manager Signature
Approval Name	Project Manager Name
Date	Date

Project Team Kickoff Meeting Agenda

Project Introduction — Business Executive Sponsor ~ Sharon Stein
- Goals and Objectives
- Project Scope

Data Warehousing Overview — Project Manager ~ Bill Turner

Team Intro ~ Roles & Responsibilities — Project Manager ~ Scott Smith

- Mark Smith — Client Reporting
- Bill Turner — Data Warehouse Architect / DBA
- Mark Smith — Team Leader ~ Report Requirements
- Frank Smith — Report Developer
- Mark Smith — Team Leader ~ Report Development
  - Frank Smith — Report Development
  - Tara Smith — Report Development

Project Management — Project Manager ~ Sheldon Smith, Laurie Lindsey
- Project Administration Tools
- Project Milestones
- DRAFT Project Plan

Questions and Answers

Next Steps

Status Meeting Agenda

- **Review Project Plan**
  - Review completed tasks and set schedule flag (behind, on, ahead)
  - Review milestones completed and pending
  - Review status of major deliverables
  - Task assignments for the next period

- **Review Issues & Follow-up**
  - Review issues resolved (resolution, who, when, move to closed)
  - Review new issues (determine steps to resolve, responsible party, priority, date to be resolved by)
  - Review open issues and determine if a change in status is needed

- **Review Change Requests**
  - Review change requests closed since last meeting
  - Review new change requests (determine responsible party for analysis, impact analysis, priority)
  - Review open change requests to determine if a change in status is needed

- **Announcements and General Comments**

Data Warehouse Status Report (sample)

To:

From:

CC:

Period:

Work Accomplished During Period Ending mm/dd/yyyy

-

Work Planned Through Next Period Ending mm/dd/yyyy

-

Open Issues / Change Control

-

Project Issue Log

The following is a summary of the open items that require further investigation, confirmation or resolution. These items have been identified during the Business Dimensional Modeling process. In order to more easily focus on open issues, you may want to shade the closed issues.

Issue #	Task / Topic	Issue	Id Date	Resp	Date Closed	Status	Priority	Rptd By

Change Control Log

The following is a summary of requests that will impact the overall project. These requests may cause one of the following to occur: Change in project scope, increase overall project delivery time or affect overall project cost.

Change #	Change Request	Req. By	Date Req.	Priority	Resp.	Est. Effort	Est. Cost	Date Closed	Status

Project Change Request

Project Change # : _____

Change Request Name: _____

Requested By: _____

Date Requested: _____

Assigned To: _____

Change Request Priority: _____
(Critical, Major, Minor)

Change Request Status: _____
(Open, Deferred, Closed)

Estimated Effort and Cost: _____

Approved By: _____

Date Closed: _____

Change Request Description

Change Request Details

REQUIREMENTS DEFINITION

Interview Preparation Checklist

Identify the Interview Team

- ❑ Lead interviewer
- ❑ Scribe
- ❑ Observers

Conduct the Pre-interview Research

- ❑ Study annual report
- ❑ Review sponsor's internal business plan
- ❑ Peruse external marketing literature
- ❑ Explore external and internal web sites
- ❑ Seek out competitors' web sites
- ❑ Understand previous data warehousing and/or related attempts

Review Interviewee List

Business Perspective

- ❑ Horizontal – cross functional perspective represented
- ❑ Vertical
  - ❑ Executive business management
  - ❑ Middle management represented
  - ❑ Cross section of business analysts

IS Data Audit Perspective

- ❑ Core operational systems
- ❑ Candidate source systems
- ❑ Database administrators
- ❑ Data modelers
- ❑ Is liaisons to the user community
- ❑ Senior IS management

Develop the Interview Questionnaires

- ❑ Business executive questionnaire
- ❑ Business manager and analyst questionnaire
- ❑ IS data audit questionnaire

Schedule the Interviews

- ❑ Review logistics of interview schedule
  - ❑ Number of interviews per day
  - ❑ Location of interviews
  - ❑ Breaks between interviews
  - ❑ Interview durations
  - ❑ Interview sequencing

- ❑ Review interview group composition by job function
- ❑ Review interview group composition by organization levels
- ❑ Review IS interview composition – focus on one source system

Communicate During User Kick-Off Meeting

- Why
  - Overview of this project
  - Rationale
  - Importance of data warehouse project
  - Level of commitment to the data warehouse project
- How / When
  - Project team roles and responsibilities
  - High level plan
- Expected involvement of users – time commitment
  - Interviews
  - Requirements validation
  - Checkpoint reviews on data model
  - Checkpoint reviews on end user application specifications
  - Education

Communicate with Pre-Interview Letter

- Overall project objectives
- Objective of the meetings with business users
  - Understand job responsibilities
  - Business objectives
  - Information and analyses for their job
- Interview time, duration, location
- Ask to bring copies of frequently used reports or spreadsheet analyses

Interview Attendee Invitation

Dear ATTENDEE,

Thank you for participating in user meetings for the PROJECT NAME data warehouse project. As a reminder, the PROJECT NAME project is focused on …

The objective of the user meetings is to better understand your area's business goals and priorities which translate into data and analyses needs. Your insight during these meetings is crucial to defining the requirements for PROJECT NAME.

Specifically, project team members intend to discuss the following topics during their meeting with you:

- **Responsibilities**
  Individual and departmental responsibilities
- **Business Objectives and Issues**
  Business metrics, industry and competitive trends, opportunities and obstacles
- **Analyses and Data Requirements**
  Key reports and analyses, frequencies and current limitations
- **Project Success Criteria**

Please **bring copies of the analyses** you are currently performing and/or requesting.

ATTENDEE, thanks in advance for your participation. The project team looks forward to meeting you on DATE at TIME in MEETING ROOM. Please call me if you have any questions in the meantime.

Sincerely,

Executive Sponsor or Joint Project Managers

CC:

Business Executive Questionnaire

A. INTRODUCTION

- Discuss data warehouse project objectives and overall status.
- Discuss interview goals (e.g., focus on business requirements, talk about what you do, what you want to be doing and why) and interview flow.
- Introduce interview team and roles.
- Confirm time available.
- Describe next steps following interview.

B. RESPONSIBILITIES

- Describe your organization and its relationship to the rest of the company.
- What are your primary responsibilities?

C. BUSINESS OBJECTIVES AND ISSUES

- What are the objectives of your organization? What are you trying to accomplish? What are your top priority business goals?
- What are your success metrics? How do you know you're doing well? How often do you measure key success factors?
- What functions and departments within the organization are most crucial to ensuring that these key success factors are achieved? What role do they play? How do they work together to ensure success?
- What are the key business issues you face today? What prevents you from meeting your business objectives? What's the impact on the organization?
- How do you identify problems/exceptions or know you're headed for trouble?
- What do you see as opportunities for additional profit that are not being addressed today?
- Where do you stand compared to your competition in the use of information technology?
- Are you able to respond quickly to market conditions and assure productivity of your staff?

D. ANALYSES REQUIREMENTS

- What role does data analysis play in decisions that you and other managers make to run the business?
- What key information is required to make or support the decisions you make in the process of achieving your goals and overcoming obstacles? How do you get this information today?
- Is there other information which is not available to you today that you believe would have significant impact on helping meet your goals?
- Which reports do you currently use? What data on the report is important? How do you use the information? If the report were dynamic, what would the report do differently?
- What analytic capabilities would you like to have?
- What opportunities exist to dramatically improve your business based on improved access to information? What's the financial impact?

E. WRAP-UP

- Summarize findings heard.
- What must this project accomplish to be deemed successful? Criteria must be measurable.
- Thank participant.
- Describe next steps (e.g., draft interview write-ups available within week) and upcoming opportunities for user involvement.

Business Manager or Analyst Questionnaire

A. INTRODUCTION

- Discuss data warehouse project objectives and overall status.
- Discuss interview goals (e.g., focus on business requirements, talk about what you do, what you want to be doing and why) and interview flow.
- Introduce interview team and roles
- Confirm time available.
- Describe next steps following interview.

B. RESPONSIBILITIES

- Describe your organization and its relationship to the rest of the company.
- What are your primary responsibilities?

C. BUSINESS OBJECTIVES AND ISSUES

- What are the objectives of your organization? What are you trying to accomplish? What are your top priority business goals?
- What are your success metrics? How do you know you're doing well? How often do you measure key success factors?
- What are the key business issues you face today? What prevents you from meeting your business objectives? What's the impact on the organization?
- How do you identify problems/exceptions or know you're headed for trouble?
- Describe your products (or other key business dimension such as customer, vendor, etc.). How do you distinguish between products? Natural way you categorize products? How would you narrow a list of thousands of products?
- How often do these categorizations change? What should happen with your business analysis following a change?

D. ANALYSES REQUIREMENTS

- What type of routine analysis do you currently perform? What data is used? How do you currently get the data? What do you do with the information once you get it?
- What analysis would you like to perform? Are there potential improvements to your current method/process?
- What type of on-the-fly analysis do you typically perform? Who requests ad hoc analysis? What do they do with the analysis? Do you have time to ask the follow-up questions?
- Which reports do you currently use? What data on the report is important? How do you use the information? If the report were dynamic, what would the report do differently?
- What analytic capabilities would you like to have?
- Are there specific bottlenecks to getting at information?
- How much historical information is required?
- What opportunities exist to dramatically improve your business based on improved access to information? What's the financial impact?

E. WRAP-UP

- Summarize findings heard.
- What must this project accomplish to be deemed successful? Criteria must be measurable.
- Thank participants.
- Describe next steps (e.g., draft interview write-ups available within week) and upcoming opportunities for user involvement.

Information Systems Data Audit Questionnaire

A. INTRODUCTION

- Discuss data warehouse project objectives and overall status.
- Discuss interview goals and interview flow. Introduce interview team and roles.
- Confirm time available. Describe next steps following interview.

B. RESPONSIBILITIES

- Describe your organization and its relationship to the rest of the company.
- What are your primary responsibilities? Role on the DW initiative?
- What business groups do you support? Other IS resources supporting same business user?

C. USER SUPPORT / ANALYSES AND DATA REQUIREMENTS

- What is the current process used to disseminate information?
- What tools are used to access/analyze information today? Who uses them?
- Are you asked to perform routine analyses? Do you create standardized reports?
- Describe typical ad hoc requests. How long does it take to fulfill these requests?
- Who are the most frequent requesters of analysis and/or data?
- What is the technical and analytical sophistication of the users?
- Describe the current user support mechanism - centralized vs. experts located in user departments?
- What is the biggest bottleneck/issues with current data access process?
- Is there a backlog of user requests for analysis, reports and/or data?

D. DATA AVAILABILITY AND QUALITY

- Which source systems are used for frequently-requested information?
  - How do production systems relate to each other? Which systems feed others?
  - What is the granularity?
  - How often is the data updated? Availability following update?
  - How much history is available?
  - What is an estimated size of this data (preliminary # of rows)?
- What are the known data gotchas in current source systems?
  - Which fields are not populated (e.g., not required and/or validated at input)?
  - Are there dual-purpose fields depending on context?
  - What is the availability of decodes? Are they buried in reporting programs?
- What master files do you have? Describe the maintenance of these master files.
  - Do you currently have common source files?
  - Who maintains the source files?
  - How are keys maintained? Are keys reassigned?
  - What is the cardinality (# distinct values)?
- Frequency of hierarchy changes within key business dimensions (product, vendor, facility…)
  - How are changes captured?

E. WRAP-UP

- What else should we know about your organization and/or systems?
- What must this project accomplish to be deemed successful? Criteria must be measurable.
- Thank participants. Describe next steps (e.g., draft interview write-ups available within week) and upcoming opportunities for involvement.

Interview Summary – Report Requirements

Report: OLTP-Appendix

Interviewed: Title:

Interviewers: Title:

Observers: -none- Title:

Date of Interview:

Background and Business Objectives

Section: **OLTP Categories Visited**

Definition:

- The numbers for Current and YTD equal the number of services provided via OLTP where the category indicated was selected by the user
- Certain Categories of Concern listed in this section of the report include additional categories from our CTI Tree as outlined
  - The service count for Everyday Issues will include the category and topic, International – Life and Travel Abroad
  - The service count for Financial will include the category Your Money
  - The service count for Health & Wellness will include the categories Health and Taking Care of Yourself
  - The service count for Legal will include the category Legal Issues and the category and topic, International – Immigration Issues
  - The service count for Older Adults will include the categories Helping Aging Parents and Midlife and Retirement
  - The service count for Childcare & Parenting will include the category Parenting & Childcare
  - The service count for Work will include the categories For Managers, Managing People, and Your Job & Career plus the category and topic, International – International Relocation. [Peter's comments – I developed the above breakdown—both Product Marketing and Service Delivery must be consulted for accuracy].

Issues:

- None.

Section: **OLTP Top 5 Categories Visited**

Definition:

- A pie chart showing the Top 5 Categories Visited by OLTP Users (same data as the OLTP Categories of Concern).

Issues:

- None.

Section: **OLTP Locator Searches**

Definition:

- A pie chart showing the Top 5 types of Locator Searches completed for OLTP Users
- A Locator Search is defined as the user entering the necessary data and OLTP returning to the user search results.

Issues:

- None.

Section: **OLTP Top 5 Topics Accessed**

Definition:

- A pie chart showing the Top 5 Topics Accessed by OLTP Users.

Issues:

- None.

Section: **OLTP Top 5 Articles Accessed**

Definition:
- A pie chart showing the Top 5 Articles Accessed by OLTP Users.

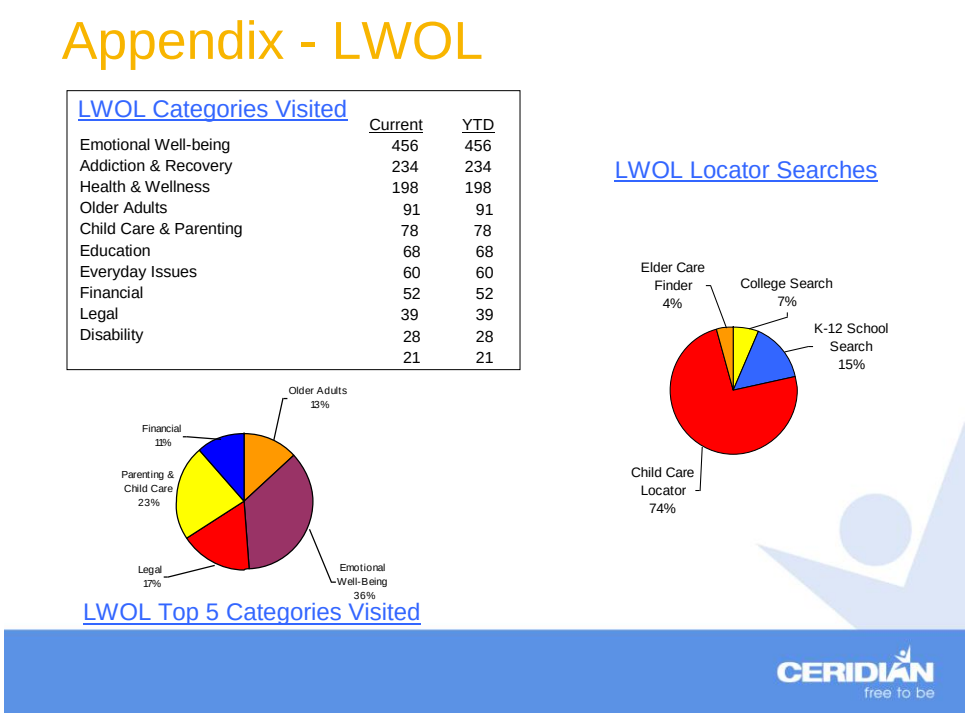
Issues:
- None.

Section: **OLTP Top 5 Materials Ordered**

Definition:
- A pie chart showing the Top 5 Materials Ordered by OLTP Users
- A Materials Ordered will not be counted until the Material has been fulfilled.

Issues:
- Time lag between order and fulfill.

Sample Report Layout:



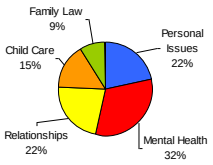
Number of Unique Online Visits	Number of unique logons to the site.	• D/W staging table Prep_Log is populated with a cleansed version of the IIS log. • Temp table #Prep_Log is populated with the Client, Group, & Division ID's plus the aspNetSessionID from Prep_Log. • All duplicate aspNetSessionID combinations beyond the first occurrence (1st occurrence is needed for the report) will be deleted. • Unique Visits are then counted and put into "buckets of time" into the Visitor table which is used for the report.
Categories Visited	Number of unique visits to a category (Note: If someone goes to category and goes to 20 different pages within that category, this counts as 1, not 20. If they go to a category, go to another category and then go back to the first category during	• D/W fact table CTIM_Page is populated from Prep_Log with date, client_id, division_id, group_id, type of transaction - CTIM_Type (C - Category, T - Topic, I - Issue, M - Material, or R - Resource), the actual category_id, topic_id, issue_id, material_id, resource_id, add_date, update_date, and an aggregated count of the hits. • Spanish categories are then inserted into CTIM_Page by looking for a query string '%&lng=2%' and CTIM_Type =

	one unique visit, it counts as 2).	'I' for Issue. • Only rows with a CTIM_Type of C are used for the Category portion of the report along with the OLTP Category table.
Number of Online LifeArticles Accessed	Total number of times that online LifeArticles are accessed.	• See #1 under Category. • CTIM_Page fact rows with a CTIM_Type of 'M' are used for the LifeArticle portion of the OLTP report along with the OLTP Material table. A cde_mat_type_id of 9 (English LifeArticle) is used to determine the English LifeArticle.
LifeArticles Accessed in Spanish	Total number of times that Spanish online LifeArticles are accessed	• See #1 under Category. • CTIM_Page fact rows with a CTIM_Type of 'M' are used for the LifeArticle portion of the OLTP report along with the OLTP Material table. A cde_mat_type_id of 19 (Spanish LifeArtice) is used to determine the Spanish LifeArticles.
Educational Materials Online	Number of materials ordered online. Not total number of orders, but the total number of booklets, recordings, and kits.	• The Educational Material section of the report does not use any Data Warehouse tables, only 3 operational OLTP tables are used. • The 3 tables are Material, Orders, and Order_Detail and the data is based on the Orders Ord_Date.
Number of e-Mails sent to Consultants	Total number of requests for referrals (cc, ec). Total number of requests via e-mail a consultant and plan-a-call.	• D/W fact table called ContactUs is populated from the OLTP operational table formSubmission. with date, client_id, division_id, group_id, cde_frm_type_id (type of referral or e-mail). • Fact table ContactUs is used to produce the Emails section of the report based on the cde_frm_type_id.
Number of Locator Searches	Number of searches submitted for each locator.	• D/W fact table called Page is populated from Prep_Log with date, client_id, division_id, group_id, target_url, and query. • The specific locator is based on the target_url and query.
Newsletter Subscriptn's	Number of active subscriptions (not just from the current period). Measures total # of subscriptions minus the total # of those who have unsubscribed).	• See #1 under Category. • CTIM_Page fact rows with a CTIM_Type of 'R' are used for the Newsletter portion of the OLTP report along with the OLTP Resource table. A cde_res_type_id of 9 (Newsletters) is used to list all newsletters. • Only subscriptions for the current period are on the report. Original specs were modified when we were told there is no way we can determine how many unsubscriptions there are.
Top 10 Topics Accessed	Every time a topic is accessed. Top 10 YTD would show (rather than the top 10 for the month).	• See #1 under Category. • CTIM_Page fact rows with a CTIM_Type of 'T' are used for the Category portion of the OLTP report along with the OLTP Category table. • The Top 10 Category hits are based on YTD numbers.
Top 10 LifeArticles Accessed	Counts every time a LifeArticle is accessed. Top 10 YTD show (rather than the top 10 for the month).	• Based on YTD numbers
Top 10 Materials Ordered	Counts every time a material is ordered. Top 10 YTD would show (rather than the top 10 for the month).	• Based on YTD numbers
Other Features	Number of times the feature is accessed. Note: Same methodology should be used as shown in "Categories visited" above. Calculator uses: total number of times "compute" is hit; B&N, Web Links, Concierge, Premium Concierge, J.I.C.: total	• The first process in the Otherfeatures section is to determine which of the features is client specific. Clients will only see the features they have contracted for. • All self_assessment features are obtained from the OLTP self_assessment table. • The features that are resources are obtained from the CTIM_Page and Resource tables.

	number of times a user clicks on a link to each link which goes to that site. All others: number of times the feature is accessed. This is not page views or hits.	• The features that are materials are obtained from the CTIM_Page and Material tables.
Number by Subscription	Total number of active subscribers by newsletter.	• See Newsletter
My TurnerProject	Number of users who have signed up for MyLW and the number of people who used the site who have a MyLW profile. Includes the number of users who signed up for the period.	• The MyTurnerProject section of the report uses 2 OLTP operational tables: User_Profile and Group_Profile • A user has personalized OLTP when the User_Profile usr_personalization_ind = 1 and the User_Profile usr_reg_time determines when the personalization took place.
Interactive Programs	Number of times a program starts. Note: These should not be showing effective 1/1/03 since they're no longer on the site.	• Still needs to be reported for Off-Cycle clients. • CTIM_Page fact rows with a CTIM_Type of 'R' and a resource_id of either 22, 23, or 24 are used in the Interactive Programs portion of the report along with the Resource table.

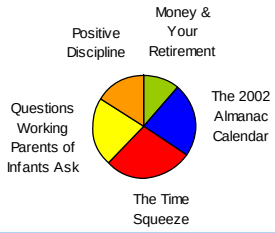
Appendix - LWOL

LWOL Top 5 Topics Accessed



LWOL Top 5 Articles Accessed

LWOL Top 5 Materials Ordered



Report: CMS-Utilization & Services Dashboard

Interviewed:	Nancy Anastas	Title:	Business Analyst
Interviewers:	Bill Turner	Title:	D/W Architect
Observers:	-none-	Title:	
Date of Interview:	01/20/2004		

Background and Business Objectives

Section: **Participant Utilization Summary**

Definition:

- Individuals receiving assessment/consultation services includes both EAP and WL type cases from all categories; does not include management consultations
- Individuals receiving assessment/consultation services may be counted each quarter (current period) but will only be counted once for year-to-date figures
- Annualized participation rate (annual) is the number of individuals receiving assessment/consultation services during the period (annualized) divided by the total number of covered participants.

Issues:

- Naming convention – participant/employee/associate/military-option.

Section: **Activity Summary**

Definition:

- The number of services provided during the period for each activity.

Issues:

- Timing of information – certain activities will be fulfilled within a short period of time (e.g. OLTP visits) other activities will not be fulfilled until 2 to 6 weeks have elapsed (e.g., sessions). Will this be mitigated by only selecting "Closed" cases?

Section: **Category of Concern**

Definition:

- The numbers for Current and YTD equal the number of services provided to the client where the category indicated was selected by the consultant
- Categories of concerns should be reported according to the latest version of the CTI tree, which is slightly different categories for EAP, OS and WL respectively
- Certain Categories of Concern listed in this section of the report include additional categories from our CTI Tree as outlined below:
    - The service count for Everyday Issues will include the category and topic, International – Life and Travel Abroad
    - The service count for Financial will include the category Your Money
    - The service count for Health & Wellness will include the categories Health and Taking Care of Yourself
    - The service count for Legal will include the category Legal Issues and the category and topic, International – Immigration Issues
    - The service count for Older Adults will include the categories Helping Aging Parents and Midlife and Retirement
    - The service count for Childcare & Parenting will include the category Parenting & Childcare
    - The service count for Work will include the categories For Managers, Managing People, and Your Job & Career plus the category and topic, International – International Relocation [Peter's comments – I developed the above breakdown—both Product Marketing and Service Delivery must be consulted for accuracy].

Issues:

- Need to ensure correct linking of the CTI Tree to Categories of Concern.

Section: **Management Consultation Utilization Summary**

Definition:

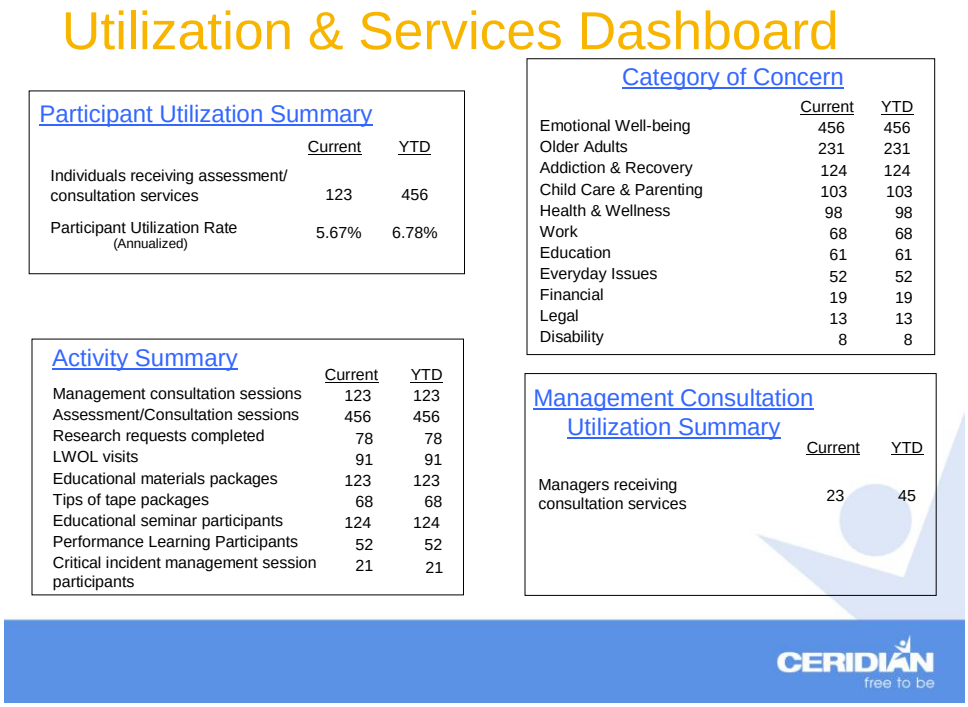
- The number of managers receiving consultation services

- Consultation services are conversations with consultants stemming from a supervisor's, manager's, or HR representative's role in their company; they are not about the supervisor's, manager's, or HR representative's own personal issues
- Managers receiving consultation services may be counted each quarter (current period) but will only be counted once for year-to-date figures.

Issues:

- None.

Sample Report Layout:



Report: CMS-Case Completion Dashboard

Interviewed: Title:

Interviewers: Title:

Observers: -none- Title:

Date of Interview:

Background and Business Objectives

Section: **Avg Number of Sessions Received**

Definition:

- Based on case closed during the period, regardless of when they were opened or had other activity
- Only for EAP and OneSource clients
- Session count includes initial consultant assessment session plus any subsequent telephonic sessions with consultants plus any sessions with EAP affiliates
- Chart showing the average number of sessions received for Emotional Well-being, Addiction & Recovery, and Work Cases only
- Average is calculated by the total number of sessions received divided by the number of individuals receiving the sessions

- The Combined average will be a weighted average based on the frequency of the various cases. ?? Check definition

Issues:
- None.

Section: **Case Resolution**

Definition:
- Based on cases closed year-to-date, regardless of when they were opened or had other activity
- Only for EAP or OneSource clients
- Pie chart showing the percent of cases completed within the EAP/OS and/or Referred to other Resources and the percent of cases Referred into Health Plan
- The sum of the two percentages will equal 100
- Cases completed within EAP/OS includes those cases going to our Affiliates or completed by a TurnerPDP consultant, as long as the case was not referred to a health plan
- Referred to other Resources include alcoholics anonymous, consumer credit counseling, or others
- The cases included in this category include only Emotional Well-being, Addiction and Recovery or Work Cases.

Issues:
- None.

Section: **Status of Participant's Concern**

Definition:
- Based on cases closed year-to-date, regardless of when the case was opened or had other activity
- Only for EAP or OneSource clients
- Includes Emotional Well-being, Addiction & Recovery, and Work Cases only
- Pie charts showing the participant's response in 1 of 4 categories (Improved, Resolved, No Change and Worsened)
- The sum of all percentages will equal 100

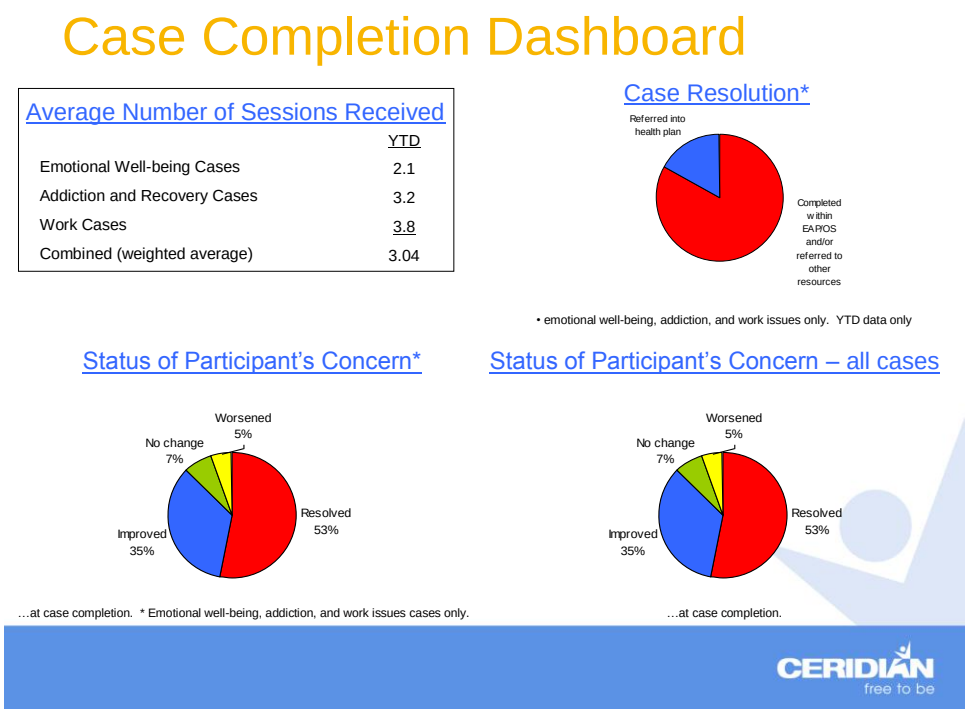
Issues:
- Data not currently captured in CMS – added to CMS Enhancement list.
- Naming convention – participant/employee/associate/military-option.

Section: **Status of Participant's Concern – all cases**

Definition:
- Based on cases closed year-to-date, regardless of when the case was opened or had other activity
- Includes all categories of cases
- Pie charts showing the participant's response in 1 of 4 categories (Improved, Resolved, No Change and Worsened)
- The sum of all percentages will equal 100

Issues:
- Data not currently captured in CMS – added to CMS Enhancement list.
- Naming convention – participant/employee/associate/military-option.

Sample Report Layout:



Report: CMS-Appendix – Assesssment/Consultation Services

Interviewed: Title:

Interviewers: Title:

Observers: -none- Title:

Date of Interview:

Background and Business Objectives

Section: **Impetus for Contacting TurnerProject**

Definition:

- Pie chart outlining the impetus/catalyst for contacting TurnerProject
- Categories include; Self, Co-worker, Mandated, Management Suggested, and Other
- Categories shows as a percent of the total
- The sum of all categories equals 100.

Issues:

- Data not currently captured in CMS – added to CMS Enhancement list.

Section: **Participant Profile (pie chart)**

Definition:

- A chart showing First Time, Repeat, and Total Participants for the period

Issues:

- Naming convention – participant/employee/associate/military-option
- First time users are based on the calendar year—will we be able to "reset a participant" at the start of the year (e.g., when a repeat participant in year 1 uses the service for the first time in year 2—the participant is a First Time Participant for year 2)?
- How will we handle Carry-over Participants?

Section: **EAP Service Modality Selected**

Definition:

- Pie chart outlining which modality the participant selected
- EAP and OneSource (Work Issues, Addiction & Recovery, and Emotional Well-being cases) clients only
- Two categories; EAP Participants choosing telephonic services or EAP Participants choosing face-to-face services
- The sum of all categories equals 100

Issues:

- During the call, do we present modality as a choice? We present, but the choice is always with the user.

Section: **Participant Profile (graph)**

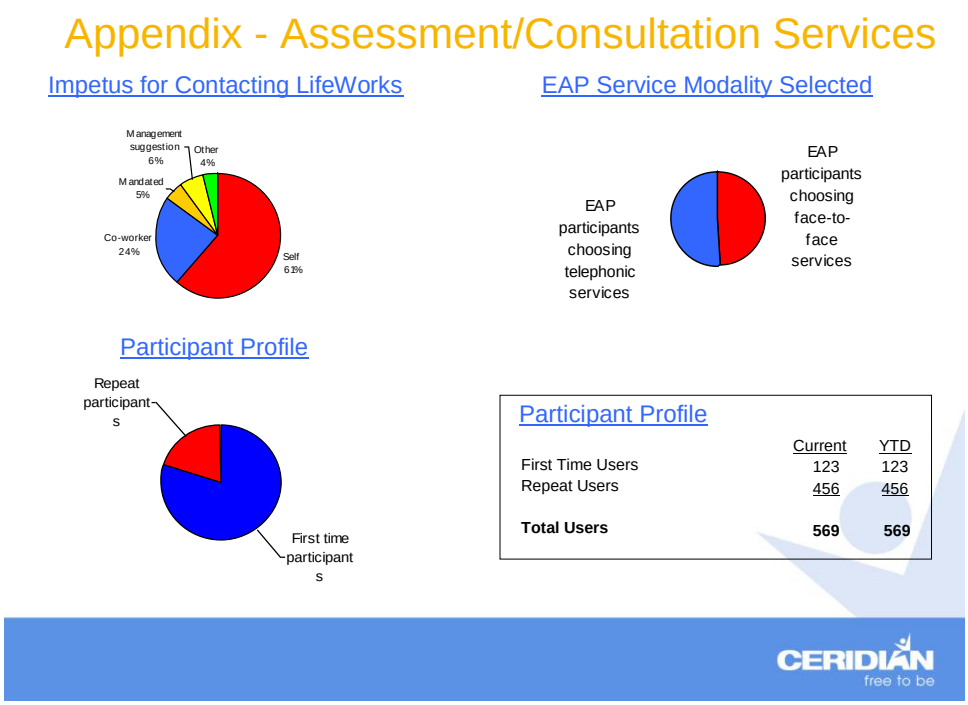
Definition:

- Pie chart outlining First Time and Repeat participants during the reporting period
- EAP and OneSource clients only
- Two categories; First Time Participants and Repeat
- First time participants are a participant using the services for the first time in the calendar year
- Repeat participants are participants with a previously (closed) case in the system in a calendar year)
- The sum of all categories equals 100.

Issues:

- Naming convention – participant/employee/associate/military-option
- First time users are based on the calendar year—will we be able to "reset a participant" at the start of the year (e.g., when a repeat participant in year 1 uses the service for the first time in year 2—the participant is a First Time Participant for year 2)?
- How will we handle Carry-over Participants?

Sample Report Layout:



Report: CMS-Appendix - Demographics

Interviewed: Title:

Interviewers: Title:

Observers: -none- Title:

Date of Interview:

Background and Business Objectives

Section: **Gender**

Definition:
- A pie chart showing the percent of Female users and the percent of Male users compared to the total number of users for the current period.

Issues:
- Current reports shows gender for cases only. Need to specify (in the glossary) that this report is gender for all assessment/consultations (OLTP activity not included).

Section: **Exempt/Non-Exempt**

Definition:
- The pie chart shows the percent of participants receiving a service in the current period and their exempt/non-exempt status.

Issues:
- Naming convention – participant/employee/associate/military-option.
- Clients ask us to capture different items, e.g., Exempt/Non-Exempt or Union/Management, etc. Perhaps we should whatever breakdown exists.

Section: **Relationship to Eligible Employee**

Definition:
- A pie chart showing the participant's relationship to the eligible employee for the current reporting period
- The sum of all categories equals 100.

Issues:
- Relationship to eligible employee categories changing
  - CMS – Employee, Spouse/Family Member, and Other
  - DW – Employee, Spouse or Significant Other, Child, and Other.
- Naming convention – participant/employee/associate/military-option.

Section: **Age**

Definition:
- A pie chart showing the percent of participants falling into the specified age categories for the current period.

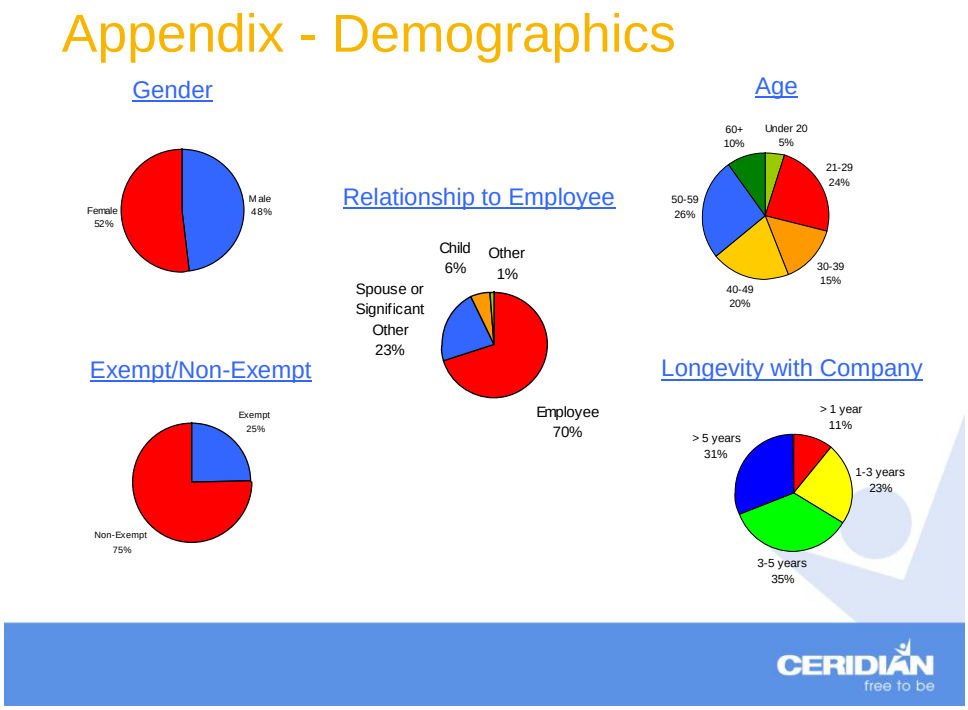
Issues:
- Age range categories are changing
  - CMS – 0-18, 19-30, 31-45, 46-65, and 65+
  - DW – Under 20, 21-29, 30-39, 40-49, 50-59, and 60+.

Section: **Longevity with Company**

Definition:
- A pie chart showing participants longevity with the Company
- Participants include anyone receiving a service during the current period
- Longevity categories include; >1 year, 1-3 years, 3-5 years, and >5 years.

Issues:
- None.

Sample Report Layout:



Report: QA-Performance Dashboard

Interviewed:		Title:
Interviewers:		Title:
Observers:	-none-	Title:
Date of Interview:		

Background and Business Objectives

Section: **Impact on Work Performance**

Definition:

- All ratings on a 5-point scale (strongly agree, agree, neutral, disagree, strongly disagree)
- A high rating is the sum of the top two rankings (strongly agree plus agree)
- A neutral rating is the number of responses in the neutral ranking
- A low rating is the sum of the bottom two ranking (strongly disagree plus disagree)
- Reporting period is the last 4 quarters
- The histogram will display client-specific data is there are 30 or more responses for the reporting period
- If less than 30 client-specific responses, the report will display Book of Business results
- The sum of all rankings (high, neutral and low) for each category (e.g., reduce stress) adds up to 100.

Issues:

- Currently we do not collect data for the following categories; Increased Ability to Focus, Improved Relationship – Coworkers, Improved Relationship – Supervisor.

Section: **Participant's Time Saved**

Definition:

- Data drawn from year-to-date (or the last 2 quarters if the report is the 1st quarter report)

- The number and corresponding percentage of respondent answering the survey question on the amount of time saved as a result of using the service
- Categories include; 1-10 hours, 11-20 hours, 21-30 hours, 31-40 hours, 40+ hours, and No Time
- Sum of all categories equals 100.

Issues:

- Category scale may be changing
- Naming convention – participant/employee/associate/military-option.

Section: **Return on Investment**

Definition:

- Requires information from four sources; client, satisfaction surveys, client utilization, and industry benchmarks
- Client data – number of exempt and non-exempt employees, average annual salary for exempt and non-exempt employees, average annual cost per employee for healthcare benefits, average age of workforce, turnover rate, average number of unscheduled absences per employee per year, industry, and gross revenue
- Satisfaction data – number of "strongly agree" responses to the impact questions (e.g., reduce stress) and the amount of time savings attributed to the service
- Client utilization – number of consultation cases, educational materials, OLTP unique logins, performance learning attendance, and seminar attendance
- Industry benchmarks – employee turnover rate, average replacement costs (exempt and non-exempt), and healthcare costs associated with stress-related illnesses
- Data drawn from year-to-date (or the last 2 quarters if the report is the 1st quarter report)
- The report is a quarterly report.

Issues:

- Requires Account Executives to update information on their client prior to running this report.
- This data could be stored in its current location – Access database. (Note: will not have this information for many clients, since it's currently run only upon request.)

Section: **Participant Satisfaction**

Definition:

- Data drawn from year-to-date (or the last 2 quarters if the report is the 1st quarter report)
- Only one bar will appear for each category (e.g., overall quality) based on the program the client purchases
- The bar represents the sum of strongly agree and agree.

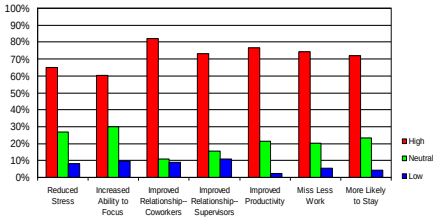
Issues:

- Option - show this graph with only Overall Quality and show it over time (last four quarters)
- Show the client-specific data versus the book of business data when the number of responses is statistically significant
- Naming convention – participant/employee/associate/military-option.

Sample Report Layout:

Performance Dashboard

Impact on Work Performance

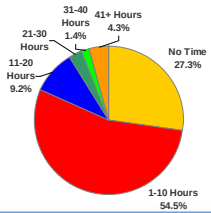


Return on Investment

YTD or Last 2 Quarters

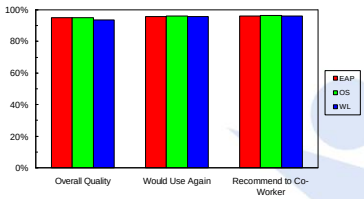
Value of Employee Time Saved	$123
Value of Increased Productivity	$456
Value of Reduced Absenteeism	$789
Employee Replacement Costs Avoided	$123
Healthcare Cost Savings	$456
Disability Cost Savings	$789
Program Fees	($1,234)
Estimated Return on Investment	**3 to 1**

Participant's Time Saved



Participant Satisfaction

YTD or Last 2 Quarters



CERIDIAN free to be

Report: QA-Appendix – Satisfaction Detail

Interviewed: Title:

Interviewers: Title:

Observers: -none- Title:

Date of Interview:

Background and Business Objectives

Section: **Satisfaction Detail – Overall Quality**

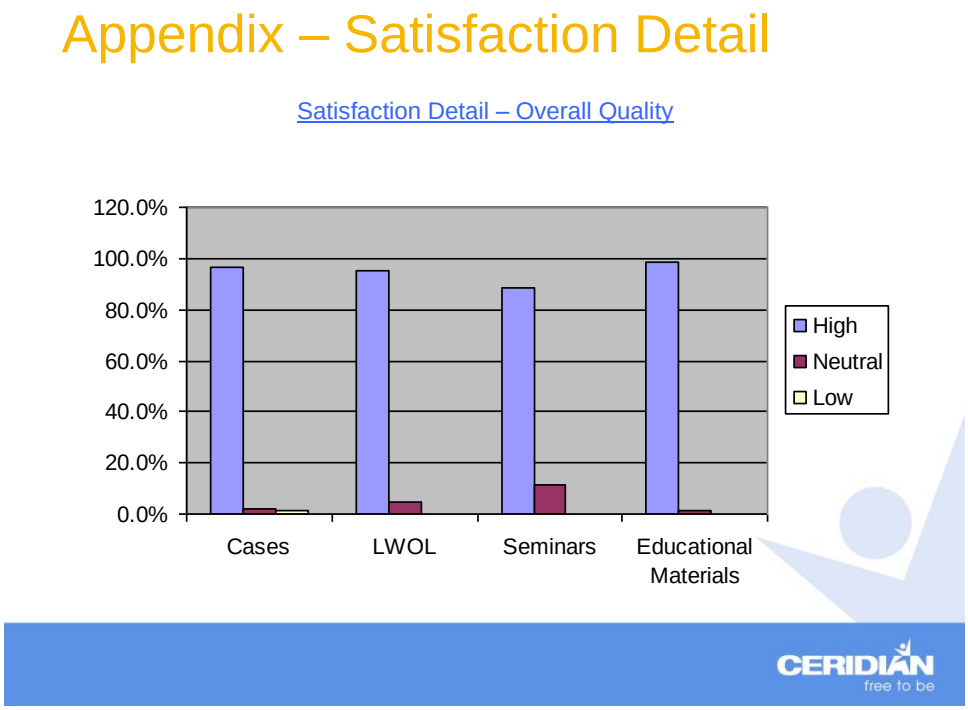
Definition:

- All ratings on a 5-point scale (very satisfies, satisfied, neutral, dissatisfied, very dissatisfied)
- A high rating is the sum of the top two rankings (very satisfied plus satisfied)
- A neutral rating is the number of responses in the neutral ranking
- A low rating is the sum of the bottom two ranking (dissatisfied plus very dissatisfied)
- The histogram will display client-specific data is there are 30 or more responses for the reporting period
- If less than 30 client-specific responses, the report will display Book of Business results
- The sum of all rankings (high, neutral and low) for each category (e.g., cases) adds up to 100.

Issues:

- None.

Sample Report Layout:



Report: QA-Custom Reports - Benchmarking

Interviewed: Title:

Interviewers: Title:

Observers: -none- Title:

Date of Interview:

Background and Business Objectives

Section: **Benchmark Report – Population Utilization**

Definition:

- A graph showing the client's population Utilization percentage compared to other clients of similar size (see size segments in Client Profile below), similar industry (see industry segments in Client Profile below), and our overall Book of Business.

Issues:

- Need to segment EAP, OneSource and WorkLife first.

Section: **Benchmark Report – Services/Participant**

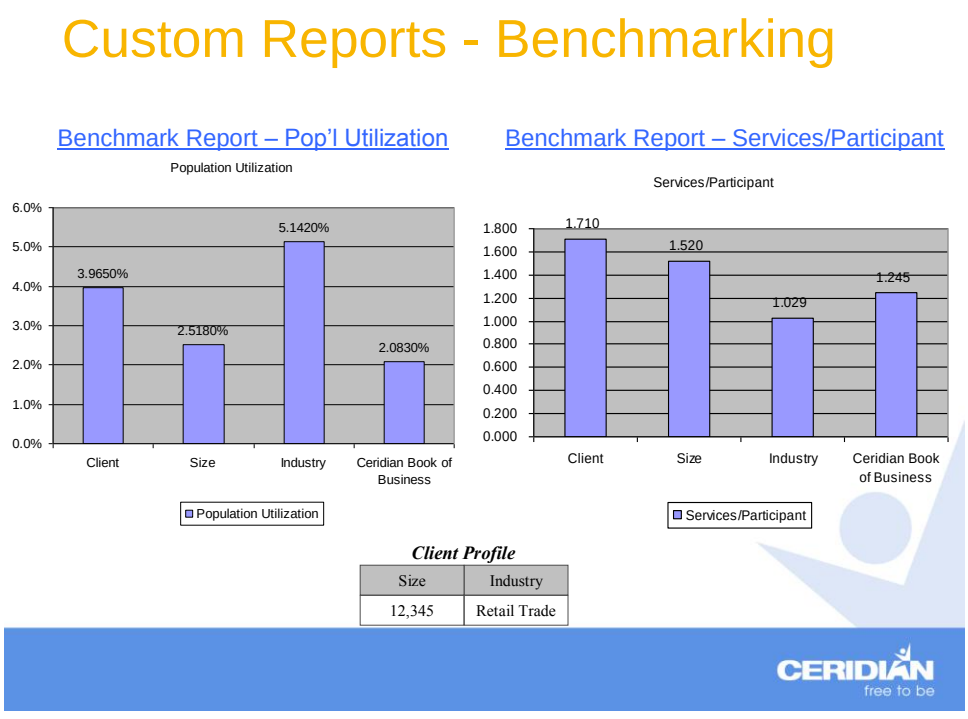
Definition:

- A graph showing the client's services per participant rate compared to clients of similar size (see size segments in Client Profile below), similar industry (see industry segments in Client Profile below), and our overall Book of Business

Issues:

- Need to segment EAP, OneSource and WorkLife first.

Sample Report Layout:



Review "old" CMS Standard Report Requirement (Donahue)

Interviewed:	DW User Community	Title:	Current CMS Users
Interviewers:	Susan Donahue	Title:	Report Development Mgr.
Observers:	-none-	Title:	
Date of Interview:	07/31/2002		

Background and Business Objectives

Susan Donahue forwarded a list of preliminary DW Report Requirements to the Data Warehouse team. Scott Littlefield and Bill Turner reviewed these req's on Mon. 10/14/2002 and derived the initial list of D/W "By's & Measure's". Here is the original list of DW Report Requirments:

Client Reporting – DW Environment

1) Cover Page – Introducing the report
   a) TurnerPDP Logo
   b) Client Name
   c) "TurnerProject Client Report"
   d) Reporting Period for the report
   e) Time Stamp for when report was run
   f) Table of Contents
      i) List each standard report produced for the client (will vary based on profile)
      ii) Provide link to each report element
      iii) Include page number next to each item in contents list

2) Summary Page (everything report for current and YTD)
   a) Current Population = What is in DW on the last day of the reporting period
   b) Average YTD Population
   c) Reporting Year Start Date
   d) Total Service Use/Annualized Rate / For each reporting period
   e) Total Unique Users / New and Repeat /Annualized Rates / For each reporting period / Total SIBs = Total DW Users and can be denominator for "Users" calculations – we need a way to define DW users which differentiates them from "all Users" without making it seem like program deficiency.
   f) Total Mode of Access Uses/Annualized Rate/For each reporting period
   g) Usage Comparisons to by core service level and including only core services:
      i) Industry
      ii) Program
      iii) Client size (within program)
      iv) Client's history – beginning in 2004
   h) Top 10 Issues (to include OLTP and DW and anything else available via CTI)
   i) Top 10 Materials (to include OLTP and DW orders and anything else available via CTI)
   j) Do we need a "parent" client summary page when applicable, what data elements would be included

3) Usage Detail Sections (current period and YTD)
   a) Service Usage
      i) List Summary (i.e. summarize all types of cons –cc, ec, etc. into "cons") Service Usage by type – include number of uses, % to total, and utilization rate for current period and YTD annualized
      ii) List Detail Service for Each Usage type (i.e. how many of each type of cons were there?)
      iii) Note number of CISDs events, Training Units (seminars, pl sessions, visibilities, other)
   b) User Use Detail
      i) # unique users, current and YTD,
      ii) # new users = when a new user record is added
      iii) # repeat users = when a user contacts us more than once during the reporting period for services (how are repeat users derived – every service or every mode of access)
      iv) for each item above calc utilization over population and total SIBs (Total SIBs = total users calc'd)

c) Mode of Access Use Detail
   i) # of accesses by type current and YTD and utilization rate (accesses/pop) and % to total by type of access)
d) Cross-sections of use algorithms (Name??)
   i) Service (which can be attributed to a unique user) over total users (aka SIBs)
   ii) Mode of access over unique/repeat
   iii) Service over mode of access

4) Performance Stats
   a) Get standard guarantees (client specific where applicable – i.e. phone lines) from Jay
   b) Client specific PG reports need to be integrated with the rest of the report

5) Caller Demographics/Client Data Questions – numbers should tie to unique users – would this be just for CASES? If so, do this belong with the Case Summary?
   a) Caller Gender
   b) Caller Relationship (show detail, or just "other")
   c) Caller Employment status (show as actual answer list in DW for each client) – no free form list
   d) How Heard (show as actual answer list in DW for each client) – no free form text (is how heard only for cases? If so belongs in the Case Summary section)
   e) Language
   f) Misc. client data questions (where are these in DW?)
   g) Any other demographics we should capture??

6) Use Coding
   a) Service Use Coding Report – breaks down service use summary level by type by code include population where available
   b) Mode of Access Coding Report – breaks down mode of access use by type by code include population where available
   c) User Use Coding Report – by user type (new, unique, repeat) only available YTD, include population where available
   d) Do we need a summary coding page? What elements would be included?

7) Brief Info/Case Referrals to External Resources, e.g. company benefits (need to note for WL clients EAP refs)

8) Case Summary – See attached document and add the following…
   a) Define Case (TBD)
   b) Only report # of open and recurring cases for open cases
   c) All details come under CLOSED CASES (AE's need to understand that new clients wouldn't show any info until they had closed cases)
      i) Number of Closed Cases
      ii) Cases by Category
      iii) Number of consultations by category
      iv) Number of specialty consultations by category
      v) Number of consultations/case
      vi) Number of categories/case
      vii) Top 10 categories
      viii) Top 10 issues
      ix) Detailed Listing of CTI – Caller identifying Event
      x) Number of research requests/case
      xi) number of referral requests/case
      xii) # materials/case
      xiii) How Heard/Source or Referral – is this the same thing
      xiv) Telephone Time Session Summary
      xv) EAP Session Use – see chart and indicate TELEPHONE SESSIONS vs. FACE TO FACE
      xvi) EAP Case Type
      xvii) Initial Referral Recommendations
      xviii) Final Case Disposition at time of closure
      xix) Company Benefits Treatment Level
      xx) For EAP cases need presenting vs. assessed issue(s)

9) Detail Sections
    (1) Materials by type report
        (a) Need to include some indication when personal action plan is sent – like education reports, or elder care assessment report
        (b) Value added reports (i.e. Consumer Reports)
    (2) Health and Wellness detail report (to be developed)
    (3) Management Line detail report (to be developed)
    (4) Seminars/Visibilities/PL Training (existing report, check format, any new fields needed)
    (5) CISD detail report (specs from Joe, report needs to be developed)
    (6) OLTP – structure is basically as we want it – make sure OLTP workshops and moderated chats are on the report

10) Specialty Product Reports

11) Quality Reports meet with Jay – Input needed from AE's on how quality should be integrated with the rest of the report

12) What else have we forgotten?

Review "old" Hierarchical Reporting (Smith)

Interviewed:	Peter Smith	Title:	DW Report Team Leader
Interviewers:	Scott Littlefield	Title:	Report Development Mgr.
Observers:	-none-	Title:	
Date of Interview:	02/25/2003		

Background and Business Objectives

Peter Smith delivered a list of preliminary DW Report Requirements to the Data Warehouse team on Monday 2/24/2003. Scott Littlefield and Bill Turner reviewed these requirements on Mon. 3/3/2003 and derived an additional list of D/W "By's & Measure's". Here is Peter's list of DW Report Requirments:

DW Client Reporting Data Requirements:

Hierarchical Reporting – identify WHO gets WHAT and WHEN ! We'll need to create a process to build the Hierarchy of DW Entities – by Parent, Client, Group, Division, sub-Division and so on..

DW's **EntityRelation** table identifies Parent-to-Company relationships ONLY!
DW's **EntityProductUnit** table identifies al relationships BEYOND Parent-to-Company (eg. Company-to-Group, Group-to-Division, Parent-to-Division, etc.

EntityProductUnit column definitions include:

- **EntityProductUnitLevelName_ID** – this is NOT the tier # in the relationship; it is the tier NAME of the relationship
- **MemberCountLevelName** – this enables an AE to track the # of employees at various levels (eg. a Company has 1000 employees, an AE can track all 1000 employees at the top level or at lower levels – this column specifies the level to track)
- **UserDefinedData_ID** – this allows TurnerPDP to gen' custom XML to capture Client Data Questions / Answers, or both (eg. How long has the Employee worked for the Company?)
- **EntitySite_ID** – example here would be within Divisions, you can track Sites. If the EntityProductUnitLevelName_ID = Site (a system value) then this column will be used (seldom, if ever, is this being used)

Cover Page – Introducing the report
1. TurnerPDP Logo
2. Client Name
3. Program Name ~ eg. OneSource, Work-Life, or EAP
4. Reporting Period ~ eg. From 11/01/2003 to 11/30/2003
5. Time Stamp ~ eg. Report created on 12/01/2003 at 14:54)
6. Table of Contents ~ Hyperlinks to specific sections of reports and each page contains a page number

Overall Summary Page (everything report for current and YTD)
1. Current Population ~ What is in DW on the last day of the reporting period (sorted by month)

Select count(*) from DIMEligibility where Eligibility_ID_nk between @ClientStart and @Client_End and caldate between @Start_Month and @End_Month

Note: the DIMEligibility table contains the link between Patient (User) and Entity (Client). Multiple Patient rows can exist in the Eligibility table – each one w/ a luEligibiltyRanking_ID associated with it – which prioritizes this Patient's relationship to an Entity

2. Average YTD Population – sum of population for each month of the year divided by the number of months

3. Reporting Year Start Date eg. Reporting Period Begins: 01/01/2003

4. Total Service Use / Use Rate - For each reporting period

5. Total Unique Users / New and Repeat /Annualized Rates / For each reporting period / Total SIBs = Total DW Users and can be denominator for "Users" calculations – we need a way to define DW users which differentiates them from "all Users" without making it seem like program deficiency.

6. Total Mode of Access Uses/Annualized Rate/For each reporting period

7. Usage Comparisons to by core service level and including only core services:
    i. Industry
    ii. Program
    iii. Client size (within program)
    iv. Client's history – beginning in 2004

8. Top 10 Issues (to include OLTP and DW and anything else available via CTI)

9. Top 10 Materials (to include OLTP and DW orders and anything else available via CTI)

10. Do we need a "parent" client summary page when applicable, what data elements would be included

Questions on Requirements Documentation

Prepared by: Bill Turner, Frank DeSefano
Delivered to: Susan Donahue
Date: Thursday 10/17/2002

Regarding Susan Donahue's initial DW Requirements Definition document (see previous pages):

1. Re: 2a & 2b ~ What is the technical definition of **POPULATION**? Is this the equivalent of the CMS "Contract Summary" Client Reports "Total Usage Summary Section"?

   Population is by Client. A review of the Data Stage load job which creates the F_Cust_Contract_History table will detail this build. This job inserts a new row, per client / per month, w/ the value num_employees ~ housed in the CMS Client table.

   eg. American Airlines would have 12 rows (1 per month) for the year 2002 – each containing the number of employees for that month. The num_employees valuye is captured at various levels (parent, company, division) ~ the DS load job reflects the detail.

   Note: If a given month's population (# employees) needs to be changed, the OLTP will not be updated w/ the value as the OLTP contains only the current population. All we need to do in the DW is to UPDATE an existing row's # employee value w/ the new value and add a comment field describing the change. This can be done via a simple SQL UPDATE statement or a front-end Access panel. We'll also want to capture the USER ID responsible for the update along w/ the timestamp.

2. Re: 2c ~ Again, is "Report Year Start Date" the equivalent of the existing CMS "Contract Summary" Client Report "Total Usage Summary Section**"** **Report Year Start Date** value?

   Yes, this is sourced from the AGG_DATE in the profile. (<u>not</u> the START_DATE)

3. Re: 2d ~ what are the calculations/algorithms (monthly/quarterly) for "Annualized Usage" on the Contract Summary page.

   TBD

4. Re: 2g ~ as discussed w/ Scott, we envision this item – Service x Service Type x Time .. as:

Services	Service Type	LOGICAL Model	Physical Model		
C&R	Industry	DIM Tbl = Service_Type	TYPE#	TYPE	DESCRIPTION
C&R Financial	Program	Industry = Financial, Insurance, etc.	1	IND	
Mat	Industry	Program = EAP, WorkLife, 1Source	2	IND	Insurance
Mat	Program	Client Size = Small, Medium, Large	3	PGM	EAP
Mat	Client Size	Client History = TBD ??	4	PGM	WorkLife
			5	PGM	1Source
			6	CLSIZE	Small
			7	CLSIZE	Medium
			8	CLSIZE	Large

   this approach is on target !!

5. Re: 5a-g ~ Explain Client, Customer, User, Member and how they relate~~~ for example:

   Are **<u>all</u>** 'employees' of a given TurnerPDP Customer (Client ~ eg. IBM) bulk loaded into the DW eCura system?

   TBD ~ Scott is organizing a meeting w/ Nancy

Describe a typical transaction whereby a User contacts TurnerPDP for information on a given need…. We need to understand if a mother of an Employee is added as a User, then member – or—is the Mother's child added as the User then Member etc..

6. Re: 8.19 ~ What is Treatment Level x Time (is Treatment Level an attribute of a Service?)

   TBD

7. Re: 8.20 ~ Where is presenting issue vs. assessed issued captured in the eCura DBMS?

   TBD

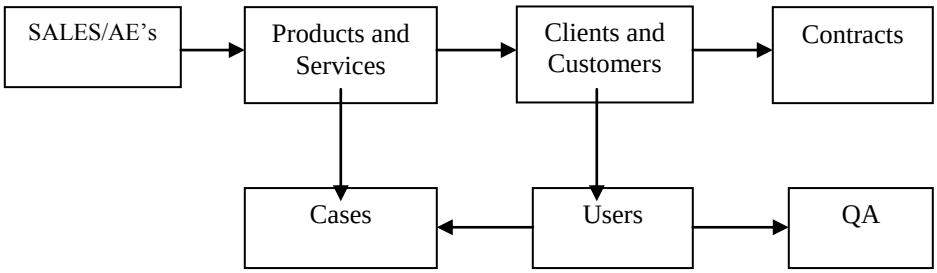
8. Re: 9-12 ~ ask Susan for detail on all these

   TBD

DESIGN / MODELING

Define Data Warehouse "Subject Areas"

REGISTRATION → USERS



- **Sales Rep's / Account Executives "sell" Products & Services "to" Clients / Customers "generating" Contracts…**
- **Users "generate" Cases "requiring" QA**

Registration & User Detail:

Sales/AE's

Products --and their-- Services
- EAP (std) emotional well-being, legal, finance
- WorkLife child care, elder care, every day issues
- OneSource
- OLTP
- EAP (non-std) emotional well-being, concierge <- which is really a WorkLife service)
- WorkLife (non-std)
- OneSource (non-std)

EXTERNAL Data Feeds (sourced/stored from outside of DW):

- NurseLine Direct
- Concierge
- Elder Care
- Managed Care
- ExPat (ex-patriot; eg. an American citizen living abroad requiring assistance)
- Lactation
- International EAP
- Health & Wellness

Clients/Customers
- IBM, Compaq
- 3rd party (eg. Prudential Ins. Offers TurnerProject as part of their service to their customers)

Contracts
- Number of covered lives ~ true number of employees
- Service members (military only)
- Members (eg. non-profit organizations)

Users

- Employee
- Spouse
- Dependants
- Extended Family
- Other

Cases/SIBs

- Demographics
- Services
- Materials
- CDQ
- Work Assignment
    - Provider
    - Research Request
- FollowUp

QA

- Surveys

Important: CMS tracked root information at the CASE level; DW will track root information at the USER level

Dimensional Model Design Checklist

- ❑ No logical or physical tables are reflected in the Dimensional Model. Primary and foreign keys are NOT explicitly included. These are introduced when the Dimensional Model is translated into logical table structures.
- ❑ Each business process will contain one or more fact tables.
- ❑ Every data element from source systems will have one of 3 states: dimension, fact or not used for analysis.
- ❑ Each dimension has one and only one lowest level attribute, called the dimension grain.
- ❑ Most models will have at least one period or time dimension. There may be more than one period dimension. Date and time may be split into 2 separate dimensions.
- ❑ Each attribute can live in one and only one dimension.
- ❑ If there is a single data element that appears to reside in more than one place, this can only occur if the second is a specific instance of the first. Name them uniquely and treat them as separate attributes.
- ❑ A single business attribute can have one or more logical columns associated with it. For example, the product attribute may translate to product code, product short description and product long description.
- ❑ Slowly changing attributes are identified and their treatment documented.
- ❑ Each fact table must have one and only one grain.
- ❑ Dimensions diagrammed for separate fact tables are conformed.
- ❑ Every fact must have a default aggregation rule (e.g., sum, min, max, semi-additive, not additive).
- ❑ Base facts should be included and derived facts should be noted as business measures and documented in the derived fact work sheet. If it is determined that a derived fact should be pre-calculated, then it should be reflected in the Dimensional Model and the logical database design.
- ❑ Numeric data is not by default a fact (e.g., phone numbers, etc.).

Data Source Checklist

- ❑ Identify possible sources
  - ❑ Explore formal data sources
  - ❑ Explore informal data sources
  - ❑ Explore external data sources
  - ❑ Business owner
  - ❑ IS owner
- ❑ Obtain data structure details for each source
  - ❑ File layouts
  - ❑ Data element definitions and descriptions
  - ❑ Data models
- ❑ Obtain data content details for each source
  - ❑ Accuracy
  - ❑ Browse values
- ❑ Considerations when there are multiple sources for similar data
  - ❑ Data accessibility
  - ❑ Future direction of the source
  - ❑ Data accuracy
  - ❑ Project scheduling
- ❑ Considerations for multiple sources are to be integrated
  - ❑ Review prior integration attempts
  - ❑ Understand what has already been integrated
  - ❑ Establish a business priority of the sources

Data Mart Matrix

The data mart matrix shows the relationship between the possible facts and dimensions. A brief description of each fact and dimension follows the matrix.

tomer	Industry	Member	Mode of Access	Product	Service	Category	Topic	Issue	Material	Resource	Consultant	Affiliate	Prov
X			X		X						X		
			X		X	X	X	x	x	X	X		
X					X	X	X	x	x	x	X		
X					x	x	X	x	x	x	x		

Example Subject Area Flow: **USERS** make **CONTACT** creating **SIB's** requiring **FULFILLMENT**

Notes:

- **Customer Level** ~ can be at any level: Group, Parent, Client, Division, Location, etc.
- **Industry Type** ~ financial services, insurance, telecommunications, retail, transportation, etc.
- **Mode's of Access** ~ telephone (inbound/outbound), e-mail, fax, website, attending an event, etc.
- **User's become Member's w/ attributes**: Gender, Relationship, Employment Status, How heard, Language, etc.

Data Mart Dimensions

[Insert Data Mart Dimension Diagram]

Dimension Descriptions

Dimension Name	Dimension Description
Calendar	Contains all of the attributes associated w/ the date & time that an activity occurred
Customer	Represents all Client names at any level
Customer Level	Represents specific Client level (eg. Group, Parent, Client, Division, Location,etc.)
Industry Type	Describes type of industry (eg. financial services, insurance, telecommunications, retail, transportation, etc.
Member	Represents a Client "user" who has contacted TurnerPDP
Mode of Access	Describes how a Member contacted TurnerPDP (eg. ~ telephone (inbound/outbound), e-mail, fax, website, attending an event, etc.
Product	Contains information about products
Product Type	Categorizes the different types of Products
Service	Contains information about services
Service Type	Categorizes the different types of Services
Category	Contains a finite list of all available Categories
Topic	Contains a finite list of all available Topics
Issue	Contains a finite list of all available Issues
Material	Contains a finite list of all available Materials
Resource	Contains a finite list of all available Resources
Consultant	Represents attributes of the person responsible for a given SIB
Affiliate	Represents attributes of a contracted service provider
Provider	Represents attributes of a service provider
Service Call Type	Describes the different types of calls requesting services/products
Service Call Status	Describes the status of a call from a Member requesting a service/product

Dimension Details

XXX Dimension Details

[Insert Dimension Detail Diagram]

XXX Dimension Attribute Descriptions

Attribute Name	Attribute Description	Sample Values

YYY Dimension Details

[Insert Dimension Detail Diagram]

YYY Dimension Attribute Descriptions

Attribute Name	Attribute Description	Sample Values

Building the Physical Dimension (Type 2 – Slowly Changing)

This section describes the steps required to build a physical Data Warehouse Dimension – Type 2 Slowly Changing.

Logon to Data Warehouse server

Fact Details

AAA Fact Details

[Insert Fact Table Diagram]
Includes: Grain, Dimensions and Facts

AAA Base Fact Descriptions

Fact Name	Fact Description	Default Aggregation Rule

BBB Fact Details

[Insert Fact Table Diagram]
Includes: Grain, Dimensions and Facts

BBB Base Fact Descriptions

Fact Name	Fact Description	Default Aggregation Rule

Derived Fact Worksheet

Chg Flag	Fact Group	Derived Fact Name	Derived Fact Description	Type	Agg Rule	Formula	Constraints	Transfor-mations

Logical Table Design

Table Name	Column Name	Data Type	Len	Nulls?	Column Description	PK	PK Order	FK

Data Source Definitions

Source	Business Owner	IS Owner	Platform	Location	Data Source Description

Source to Target Data Map

Target Table	Target Column	Data Type	Len	Target Column Description	Source System	Source Table / File	Source Col / Field	Data Txform Notes

ARCHITECTURE

Architecture Framework

Overview

- **Infrastructure**

Scott Smith	~	Client Reporting Manager
Bill Turner	~	Data Warehouse Architectect / DBA
Mark Smith	~	Report Development Team Leader
Tara Smith	~	Report Team Leader

- **Environment Setup**
- **Main Subject Areas**
- **Slowly Changing Dimensions**
  - Identify DW's Historical Reporting Requirements
  - Design the architecture for processing the Slowly Changing Dimension strategy
  - The 2 'primary' Types of Slowly Changing Dimensions evaluated included:
    - **Type I:** Overwrite the existing data value w/ a new data value thereby losing the ability to track history. Type One dimensions are those in which you want to rewrite history.
    - **Type II:** Create an additional dimension record at the time of the change w/ the new attribute values and thereby segmenting history very precisely between the old description and the new description. Type Two dimensions are those in which you want to keep a record of the old dimension data. (eg's. marital status, region)

      *** We have opted to utilize the **Type II** Slowly Changing Dimension Architecture

TYPE 2 Slowly Changing Dimension e*xample:*

All **DIMENSION** processing takes place first – followed by **FACT** processing.

TIMELINE of ACTIVITY for U.S Army Client=822, Division=16645 (Albany Battalion):

Sep 1, 2004
This US Army Client / Division is added to the Client / Client_Division table. This Division has less than 1000 employees and is categorized as JUMBO_IND = 'N' (no).

(initiates Type 2 dimension SQL INSERT into the D_CLIENT dimension table)

Sep 30, 2004
Total Number of ORDERS are processed for this US Army Client / Division ~ ORDERS = 100.

(initiates SQL INSERT into the F_ORDER fact table)

Oct 31, 2004
Total Number of ORDERS are processed for this US Army Client / Division ~ ORDERS = 200.

(initiates SQL INSERT into the F_ORDER fact table)

Nov 3, 2004
This US Army Client / Division is modified on operational database to reflect number of employees has exceeded 1000 employees ~ JUMBO_IND = 'Y' (yes).

(initiates Type 2 dimension SQL INSERT into the D_CLIENT dimension table)

Nov 30, 2004
Total Number of ORDERS are processed for this US Army Client / Division ~ ORDERS = 40.

(initiates SQL INSERT into the F_ORDER fact table)

Dec 31, 2004
Total Number of ORDERS are processed for this US Army Client / Division ~ ORDERS = 60.

(initiates SQL INSERT into the F_ORDER fact table)

Jan 4, 2005
This US Army Client / Division is modified on operational database to reflect number of employees has fallen below 1000 ~ JUMBO_IND = 'N' (no).

(initiates Type 2 dimension SQL INSERT into the D_CLIENT dimension table)

Jan 31, 2005
Total Number of ORDERS are processed for this US Army Client / Division ~ ORDERS = 50.

(initiates SQL INSERT into the F_ORDER fact table)

DIMENSION table ~ D_CLIENT

1. CLT_SK ← **Surrogate Key** (and Primary Key) on this Dimension table
2. CLT_ID ← **Natural Key** on the Dimension table
3. DIV_ID ← **Natural Key** on the Dimension table
4. JUMBO_IND ← Jumbo Indicator (Y/N) ~ Y=greater than 1000 employess
5. ADD_DATE

D_CLIENT *sample* data:

CLT_SK	CLT_ID	DIV_ID	JUMBO_IND	Add_date
001	822	16645	N (no)	2004-09-01
002	822	16645	Y (yes)	2004-11-03
003	822	16645	N (no)	2004-01-04

FACT table ~ F_ORDER

1. ORDER_SK ← Surrogate Key (and Primary Key) on this Fact table
2. CLT_ID ← **Natural Key** on the Fact table
3. DIV_ID ← **Natural Key** on the Fact table
4. ORDER_QTY
5. ADD_DATE
6. UPDT_DATE
7. ORDER_FK1 ← Foreign Key to related Dimension table's **Surrogate** Key

F_ORDER *sample* data:

ORDER_SK	CLT_ID	DIV_ID	ORDER_QTY	ADD_DATE	UPDT_DATE	ORDER_FK1
410	822	16645	100	2004-09-30	2004-09-30	001
520	822	16645	200	2004-10-31	2004-10-31	001
630	822	16645	40	2004-11-30	2004-11-30	002
740	822	16645	60	2004-12-31	2004-12-31	002
850	822	16645	50	2005-01-31	2005-01-31	003

Each time the nightly ETL process executes, new DIMENSIONS may be added to the D/W.
Each time the nightly ETL process executes, new FACTS may be added to the D/W.

The insertion of DIMENSION records is straight forward. For any given dimension (eg. D_Client) the source operational table is scanned for rows containing a last updated timestamp = today's ETL date. Qualifying rows are then transformed and inserted into the corresponding DIMENSION table.

The insertion of FACT records is more extensive. We'll skip the aggregation logic here (as it varies from fact to fact) but we will address the "link" between a Fact record and it's parent Dimension record.

For TYPE 2 Slowly Changing Dimensions, it is important to match FACTS to their respective DIMENSIONS' point in time value. This is achieved by obtaining the "most current" DIMENSION record surrogate key ~ to be stored w/ the FACT record data.

The following SQL statement will obtain this information for us ~ and thus populate the child (FACT) foreign key linking back to the parent (DIMENSION) surrogate key.

select max(DIMENSION SK) ← the Dimension table's SK
from DIMENSION table name
where FACT table NK = DIM table NK
and DIM table add_date < ETL Run Date ***

The "max" function above will extract the most current SK (prior to this ETL's Run Date) on the DIMENSION table for a given FACT's related dimension.

***** The ETL RUN Date will be 1 day ahead of the activity date. (eg. we process 2/14/2004 data on 2/15/2004.) For ETL re-runs, the ETL run date will ALWAYS be 1 day greater than the activity date)

Here is a sample Report Request Request:

Display the Total Number of Orders for all JUMBO DIVISIONS for the Year 2004

This Query will satisfy this Report Request:

```
select CLT_ID, DIV_ID, sum(ORDER_QTY)
 from     F_ORDER  O,
          D_CLIENT  C
  where
          C.JUMBO_IND = 'Y'
    and   C.CLT_SK = O.ORDER_FK1
    and   O.ADD_DATE between '2004-01-01'  and  '2004-12-31'
group by CLT_ID, DIV_ID
```

→ *will return a Total Order Qty of 100… (order qty's for 11/30/2004 + 12/31/2004)*

- Data Model
- ETL (Extract, Transform, and Load)
- Report Drivers and Primers
- Actuate Web-based Reporting Tool

ETL Design

This section presents an overview of the TurnerProject Data Warehouse Architecture and Report Result Set Generation Architecture. These solutions are built on a Microsoft DTS / Stored Procedures platform. The architectures are built, tested, and in production. ETL packages execute nitely for our TurnerProject Data Warehouse. Report Result Set generation is an automated, on-demand process.

ETL Architectural Features:

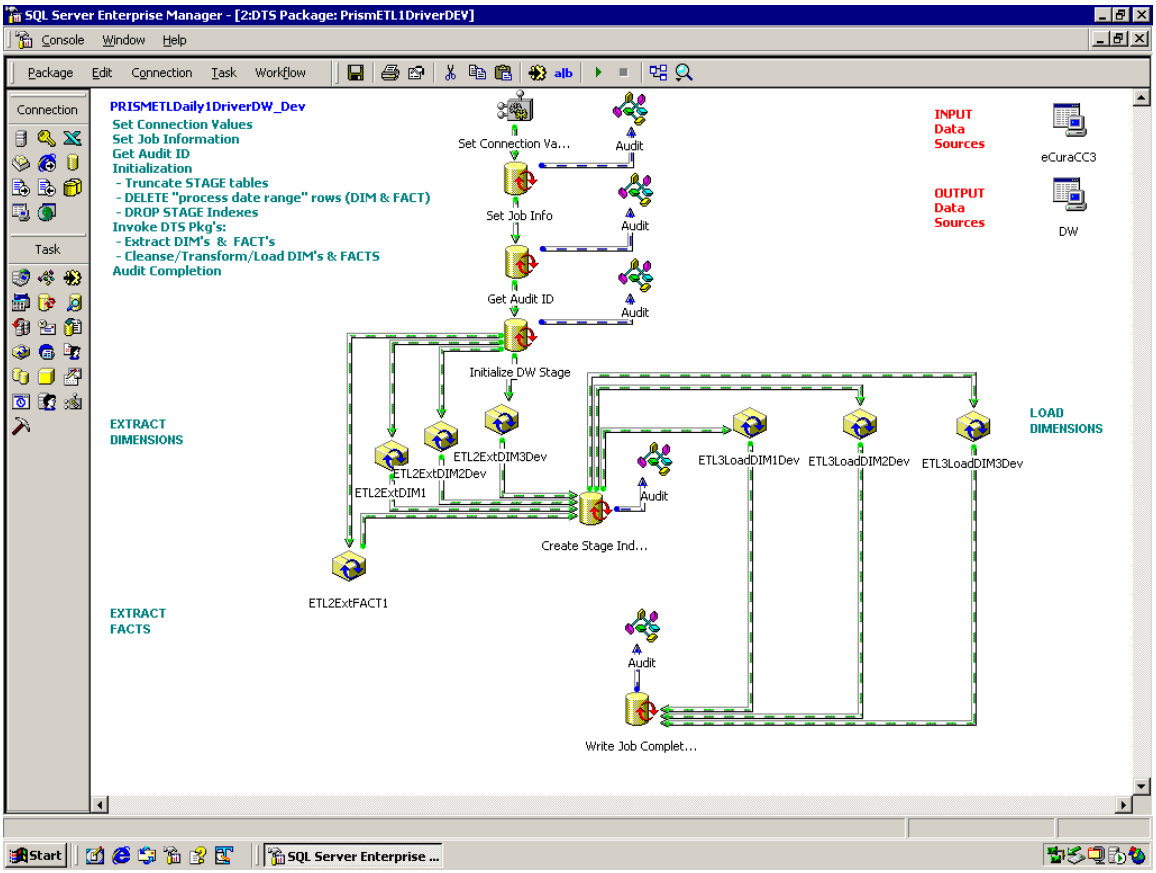
- Dynamic properties tasks - defining connections set thru UDL files easing the porting from Dev, to QA, to Prod
- Nested/layered DTS pkg's allowing for isolated ETL component builds and maintenance (eg. Dimensions extracted to Stage tables for Type 2 SCD interrogation, Facts extracted to Stage tables for transform/aggregation, SQL task loads, etc)
- Full auditing - at both the job and task level - using a common Visual Basic ActiveX script - reusable for all Audits on Exec SQL tasks; all Data Warehouse Fact & Dimension rows are tagged w/ AuditID - marrying each row back to Job & Task responsible for deriving it's existence
- Dynamic restartability for reprocessing any day (or range of days' activity)
- Capture of Type 2 Slowly Changing Dimensions
- Standard Data Warehouse attributes on all Fact & Dimension objects include: Surrogate Key, Primary Key, CreateDate, CreateBy, CreateByUserName, TouchDate, TouchBy, TouchByUserName, AuditID
- Internal Performance Monitor - capturing start/end times for all Stored Procedures and SQL statements w/in Stored Procedures
- On screen Documentation via Text Annotation

We've demo'd the Data Warehouse system/architecture to 2 experienced technical engineers located in our Atlanta Tech Services center - and received positive feedback from them. Most recently, we've installed and configured Microsoft's Reporting Services -- and have successfully imported/converted a few MS-Access based reports.

On the back end of our TurnerProject Data Warehouse, we've built an automated "Results Sets" architecture -- a build routine which primes the full content (SQL Server Data Mart tables) with our Monthly/Quarterly report needs. With the the click of a button on an Access form - the process invokes a SQL stored proc - kicking off a batch DTS package on a remote server in Atlanta - used to run a comprehensive, granular set of SQL Stored Proc's (very easy to maintain; 1 stored proc per report section) processing at both the Client Level and Division w/in Client Level - using common stored procs. (eg. the architecture is built such that the same set of stored proc's can be used for any level of reporting by supplying appropriate variables used as input to predicate data selection).

ETL - Driver

The Extract, Transform, and Load process is invoked by the DRIVER DTS package.



This package performs the following:

- **Sets Connection Values**
  - This Dynamic Properties Task (titled ***Set Connection Values*** above) references the contents of predefined DTS global variables containing pointers to Universal Data Link files (UDL's). The UDL's contain required connection information for source and target SQL Servers and Databases including SQL Server name, Database name, and authentication method (either Windows NT integrated security or specific user name and password)

- **Set Job Information (audited)**
  - This SQL Task is used to dynamically build the ETL parameter characteristics including:
    - JobName – identifies specific ETL jobname
    - JobStatus – status's include: Running, Incomplete, Completed
    - JobStartDate – represents the ETL Job's start day and time
    - PostStartDate – represents the begin date range to process
    - PostEndDate – represents the end date range to process
  - This information is inserted into an SQL Server ETL audit table called AuditJob
  - This allows for provision of ETL rerun's by simply altering the PostStartDate and PostEndDate process date range values; the default for these values is always yesterday – Midnite to Midnite
  - The following SQL INSERT statement drives these parameters:

```
insert into AuditJob
```

```
  (JobName,
   JobStatus,
   JobStartDate,
   PostStartDate,
   PostEndDate)
values
 ('DW DW ETL Daily',
  'Running',
  getdate(),
 -- DATEADD function to process YESTERDAY
  dateadd(ms, 0, convert(varchar, getdate()-1, 101
  dateadd(ms, -3, convert(varchar, getdate(), 101)))
 -- HARDCoded Dates to process variable Date Range
 -- DO NOT set Start Date > 2020 - this will wipe out DIM seed
 -- rows!
 -- '2004-01-22 00:00:00.000',
 -- '2004-01-22 23:59:59.997')
```

- **Get Audit ID (audited)**
  - This SQL Task is invoked to obtain the Audit ID established in the '***Set Job Information'*** task; the Audit ID assigned to all Fact and Dimension rows inserted into the Data Warehouse for a given execution.

  - The following SQL SELECT statement retrieves the Audit ID:

```
select max(AuditID)
 from AuditJob
  where JobName   = 'DW DW ETL Daily'
    and JobStatus = 'Running'
```

- **Initialize Staging Environment (audited)**
  - This SQL Task is invoked to:
    - Truncate all Staging tables (prior to dropping indexes)
    - Drop all Staging Indexes
    - Delete previous versions of DIMension, DIMensionSCD, & FACT rows for this PostStartDate / PostEndDate Date Range
    - The above SQL statements are all dynamic in nature – this allows for the creation/dropping of Data Warehouse Fact tables, Dimension tables, and Indexes without affecting the initialization of the staging environment
    - Sample Initialization stored procedure included here:

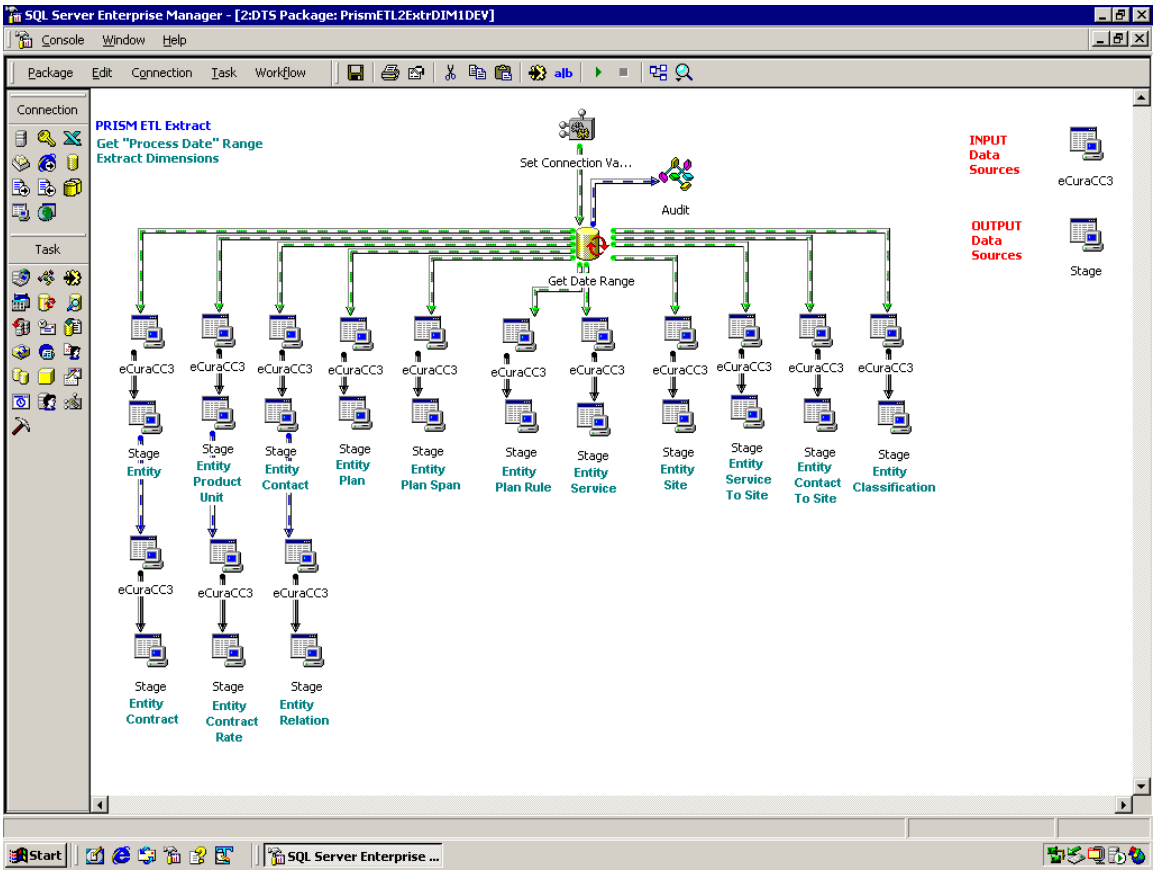
```
--*****************************************************
--TRUNCATE STAGE tables (prior to dropping indexes)  *
--*****************************************************
declare StageTableNameCsr cursor for
 select name from dbo.sysobjects
  where (name like 'STAGE%')
  and OBJECTPROPERTY(id, N'IsUserTable') = 1
open StageTableNameCsr
fetch next from StageTableNameCsr into @table_name
while @@fetch_status = 0
  begin
   set @sql = 'truncate table ' + @table_name
   exec sp_executesql @sql
  fetch next from StageTableNameCsr into @table_name
  end
close StageTableNameCsr
deallocate StageTableNameCsr

--************************************************************
--DROP Stage Indexes                                         *
--************************************************************
declare StageIndexNameCsr cursor for
 select o.name, i.name
  from sysindexes i Join sysobjects o On i.id=o.id
   where o.name like 'Stage%' and ((i.name like 'IX%')
      or (i.name like 'PK%')) and indid > 0
open StageIndexNameCsr
fetch next from StageIndexNameCsr into @table_name, @index_name
while @@fetch_status = 0
```

```
  begin
   set @sql = 'drop index  ' + @table_name + '.' + @index_name
   exec sp_executesql @sql
  fetch next from StageIndexNameCsr into @table_name, @index_name
  end
close StageIndexNameCsr
deallocate StageIndexNameCsr

-- Prime Process Date Range for this ETL execution
select @maxAuditID = (select max(AuditID) from AuditJob)
select @postStartDate = postStartDate, @postEndDate = postEndDate
 from AuditJob where AuditID = @maxAuditID

--************************************************************
-- DELETE previous versions of DIMension, DIMensionSCD, & FACT *
-- rows for this PostStartDate/PostEndDate Date Range          *
-- ************************************************************
--Obtain list of DW D/W "Stage" objects to Audit
declare DWObjectCsr cursor for
 select name from dbo.sysobjects
  where ((name like 'DIM%') or (name like 'FACT%'))
    and (name <> 'DIMCalendar')
    and OBJECTPROPERTY(id, N'IsUserTable') = 1
open DWObjectCsr
fetch next from DWObjectCsr into @table_name
while @@fetch_status = 0
  begin
   set @sql = 'delete from ' + @table_name + '
               where CalDate between ' + ''''
               + CONVERT(VARCHAR, @PostStartDate,121) + ''''
               + ' and ' + ''''
               + CONVERT(VARCHAR, @PostEndDate,121) + ''''
   exec sp_executesql @sql
  fetch next from DWObjectCsr into @table_name
  end
close DWObjectCsr
deallocate DWObjectCsr
```

- **Initiate Asynchronous DTS Extract Packages**
  - Set the number of maximum parallel DTS tasks by right mouse clicking on any "white" area of the DTS Driver package, selecting ***Package Properties***, and setting the value ***'Limit the maximum number of tasks…*** (located at the bottom of the panel)

ETL - Extract

The Extract DTS package is invoked by the DRIVER DTS package.



This package performs the following:

- **Sets Connection Values**
  - This Dynamic Properties Task (titled ***Set Connection Values*** above) references the contents of predefined DTS global variables containing pointers to Universal Data Link files (UDL's). The UDL's contain required connection information for source and target SQL Servers and Databases including SQL Server name, Database name, and authentication method (either Windows NT integrated security or specific user name and password)

- **Get Date Range (audited)**
  - This SQL Task is used to dynamically extract the PostStartDate and PostEndDate parameters previously set in the DTS Driver package's 'Set Job Information' task

  - The values are selected from the AuditJob table and dynamically stored as output DTS Global Variables
    - gvdtPostStartDate
    - gvdtPostEndDate

  - Sample SQL SELECT statement used to extract the date range:

    ```
    select PostStartDate, PostEndDate from AuditJob where AuditID = ?
    ```
    (the '?' listed in the above predicate correlates to the a relative input Global Variable)

- **Data Transformation**
  - Microsoft Data Link Connections are defined representing source (OLTP) and target (D/W staging) SQL Server tables

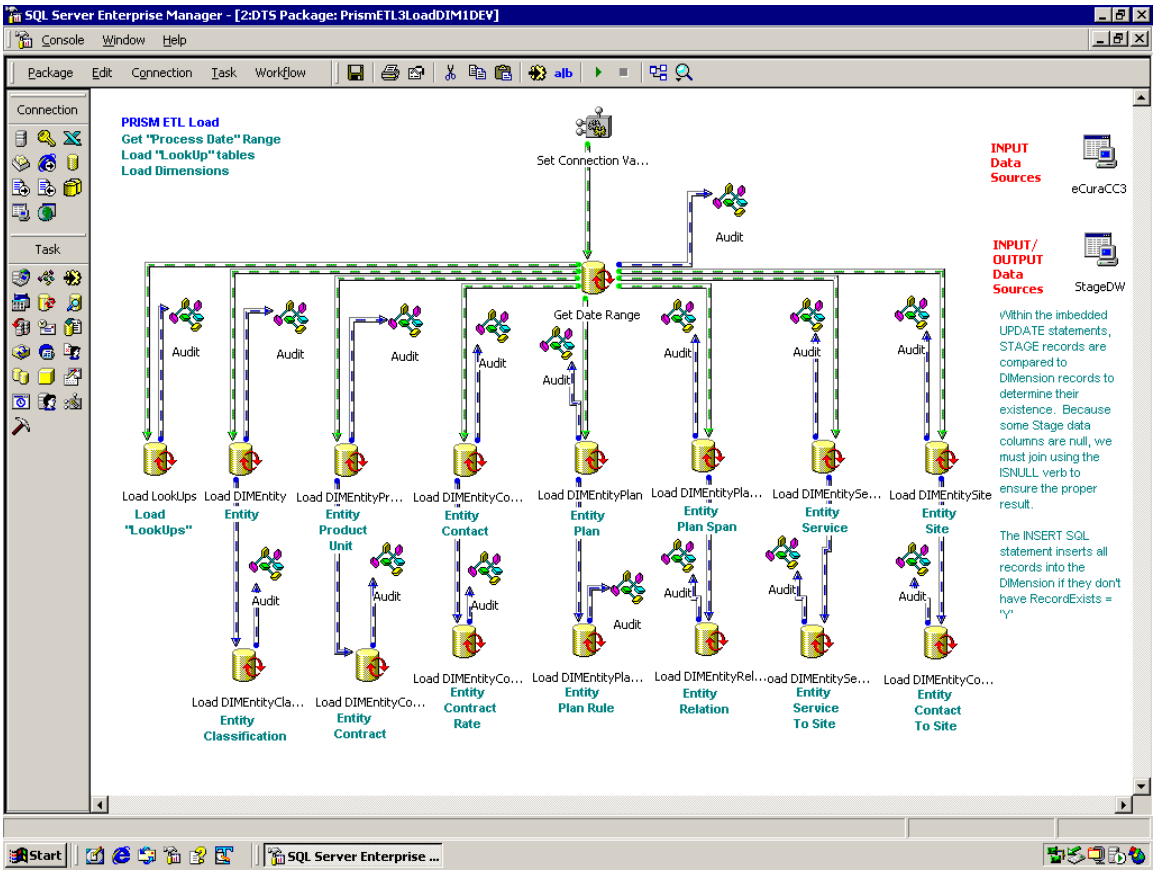
- The Data Link Connections reference predefined UDL's
- Transform Data Tasks are defined between the source and target Data Link Connections; the properties for these tasks are driven by defining SQL SELECT statements
- The SQL SELECT Statements can be multi-table joins used to flatten multi-sourced DIMension table data into a single staging table for a given Dimension
- Sample SQL SELECT used to flatten multi-sourced DIMension data

```
select
 E.Entity_ID,
 E.luEntityPayorType_ID, EPT.Description as EPTDescription,
 E.luEntityType_ID, ET.Description as ETDescription,
 E.luEntitySubType1_ID, EST1.Description as EST1Description,
 E.luEntitySubType2_ID, EST2.Description as EST2Description,
 E.luEntityPreference_ID, EP.Description as EPDescription,
 E.Name,
 E.ShortName,
 E.IncorporationDate,
 E.luTaxStatus_ID, TS.Description as TSDescription,
 E.FirstName,
 E.LastName,
 E.Middle,
 E.SSN,
 E.BirthDate,
 E.luGender_ID, G.Description as GDescription,
 E.OutOfNetwork,
 E.ProductUnitRequired,
 E.luProductType_ID, PT.Description as PTDescription,
 E.luCountry_ID, C.Name as CName,
 E.AccountManager_ID,
 E.URL,
 E.Note,
 E.UserDefinedData_ID,
 E.XMLData,
 E.ExtID,
 E.SystemID,
 dateadd(ss,datediff(ss,GETUTCDATE(),getdate()),E.CreateDate)
  as ECreateDate,
 E.CreateBy,
 CC1.FirstName + ' ' + CC1.LastName as CreateByName,
 dateadd(ss,datediff(ss,GETUTCDATE(),getdate()),E.TouchDate)
  as ETouchDate,
 E.TouchBy,
 CC2.FirstName + ' ' + CC2.LastName as TouchByName
  from entity E
LEFT JOIN luEntityPayorType EPT
 on E.luEntityPayorType_ID = EPT.luEntityPayorType_ID
LEFT JOIN luEntityType ET
 on E.luEntityType_ID = ET.luEntityType_ID
LEFT JOIN luEntitySubType EST1
 on (E.luEntityType_ID = EST1.luEntityType_ID
  and E.luEntitySubType1_ID = EST1.luEntitySubType_ID )
LEFT JOIN luEntitySubType EST2
 on (E.luEntityType_ID = EST2.luEntityType_ID
  and E.luEntitySubType2_ID = EST2.luEntitySubType_ID )
LEFT JOIN luEntityPreference EP
 on E.luEntityPreference_ID = EP.luEntityPreference_ID
LEFT JOIN luTaxStatus TS
 on E.luTaxStatus_ID = TS.luTaxStatus_ID
LEFT JOIN luGender G
 on E.luGender_ID = G.luGender_ID
LEFT JOIN luProductType PT
 on E.luProductType_ID = PT.luProductType_ID
LEFT JOIN luCountry C
 on E.luCountry_ID = C.luCountry_ID
LEFT JOIN ApplicationUser AU1
 on E.CreateBy = AU1.ApplicationUser_ID
```

```
LEFT JOIN ApplicationUser AU2
 on E.TouchBy = AU2.ApplicationUser_ID
LEFT JOIN CompanyContact CC1
 on AU1.CompanyContact_ID = CC1.CompanyContact_ID
LEFT JOIN CompanyContact CC2
 on AU2.CompanyContact_ID = CC2.CompanyContact_ID
 where
  (dateadd(ss,datediff(ss,GETUTCDATE(),getdate()),E.TouchDate)
   between ? and ?)
```

ETL – Transform & Load

The Transform and Load DTS package is invoked by the DRIVER DTS package.



This package performs the following:

- **Sets Connection Values**
  - This Dynamic Properties Task (titled ***Set Connection Values*** above) references the contents of predefined DTS global variables containing pointers to Universal Data Link files (UDL's). The UDL's contain required connection information for source and target SQL Servers and Databases including SQL Server name, Database name, and authentication method (either Windows NT integrated security or specific user name and password)

- **Get Date Range (audited)**
  - This SQL Task is used to dynamically extract the PostStartDate and PostEndDate parameters previously set in the DTS Driver package's 'Set Job Information' task

  - The values are selected from the AuditJob table and dynamically stored as output DTS Global Variables
    - gvdtPostStartDate
    - gvdtPostEndDate

  - Sample SQL SELECT statement used to extract the date range:

    ```
    select PostStartDate, PostEndDate from AuditJob where AuditID = ?
    ```

    (the '?' listed in the above predicate correlates to the a relative input Global Variable)

- **Load Dimension tables (audited)**
  - This Dynamic Properties Task (titled '***Load DIMxyz'*** above) is used to transform and load a DIMension table; incorporating Type 2 Slowly Changing Dimension logic

- The first component of the extended SQL in this task is to set staged DIMension rows with the AuditID value established in the '***ETL Driver***' DTS package. Sample SQL used to set perform this task:

```
update StageEntity set AuditID = ?
```

- The next component of the extended SQL in this task is to determine the PostStartDate (process start date range value); this is achieved by SELECTing the max(PostStartDate) value stored in the AuditJob table. Sample SQL used to perform this task:

```
declare @caldate datetime   declare @auditid int
set @auditid = ?
select @caldate =
 (select max(PostStartDate) from AuditJob
   where AuditID = @auditid)
```

- The next component of the extended SQL in this task is to set the 'RecordExists' flag to 'Y' for all Staged rows having a matched (existing) Dimension row; this value will be interrogated later. Sample SQL used to perform this task:

```
update StageEntity set RecordExists = 'Y'
 from StageEntity S, DIMEntity D
  where S.Entity_ID_nk = D.Entity_ID_nk
```

- The next component of the extended SQL in this task is to create (INSERT) a history row (slowly changing dimension record) into a history (SCD) table. This is achieved by matching the Staged row's natural key to the Dimension row's natural key. If a match is found, the 'before image' of the Dimension row is INSERTed into the SCD (history) table. Alternatively, an 'active flag' could have been reset to 'I' for this row in the active Dimension table. Sample SQL used to perform this task:

```
insert into DIMEntitySCD
(CalDate,
 Entity_ID_sk,
 Entity_ID_nk,
 luEntityPayorType_ID,
 EntityPayorTypeDescription,
……
 CreateDate,
 CreateBy,
 CreateByUserName,
 TouchDate,
 TouchBy,
 TouchByUserName,
 AuditID)
select
 D.CalDate,
 D.Entity_ID_sk,
 D.Entity_ID_nk,
 D.luEntityPayorType_ID,
 D.EntityPayorTypeDescription,
 …
 D.CreateDate,
 D.CreateBy,
 D.CreateByUserName,
 D.TouchDate,
 D.TouchBy,
 D.TouchByUserName,
 D.AuditID
 from StageEntity S, DIMEntity D
  where S.Entity_ID_nk = D.Entity_ID_nk
```

- The next component of the extended SQL in this task is to UPDATE the active Dimension row w/ the Staged Dimension row's data where a match exists between the staged row's natural key and the Dimension row's natural key. Sample SQL used to perform this task:

```
update DIMEntity set
```

```
 luEntityPayorType_ID          = S.luEntityPayorType_ID,
 EntityPayorTypeDescription    = S.EntityPayorTypeDescription,
…
 CreateDate                    = S.CreateDate,
 CreateBy                      = S.CreateBy,
 CreateByUserName              = S.CreateByUserName,
 TouchDate                     = S.TouchDate,
 TouchBy                       = S.TouchBy,
 TouchByUserName               = S.TouchByUserName,
 AuditID                       = S.AuditID
 from StageEntity S, DIMEntity D
  where
   S.Entity_ID_nk = D.Entity_ID_nk
```

- The next component of the extended SQL in this task is to create (INSERT) a new Dimension row when it is determined that this Dimension does not exist in the Dimension table – this is accomplished by interrogating the RecordExists flag. Sample SQL used to perform this task:

```
insert into DIMEntity
(CalDate,
Entity_ID_nk,
luEntityPayorType_ID,
EntityPayorTypeDescription,
…
CreateDate,
CreateBy,
CreateByUserName,
TouchDate,
TouchBy,
TouchByUserName,
AuditID)
select
 @caldate,
 Entity_ID_nk,
 luEntityPayorType_ID,
 EntityPayorTypeDescription,
 …
 CreateDate,
 CreateBy,
 CreateByUserName,
 TouchDate,
 TouchBy,
 TouchByUserName,
 AuditID
  from StageEntity
   where RecordExists is null
```

ETL – Audit

An ActiveX Script Task – **Audit** - is used to dynamically capture and store DTS Task level information. Details related to a specific SQL Task are INSERTed into the table **AuditTask**. This table contains the following attributes:

- TaskID
- AuditId
- PackageName
- TaskName
- TaskStatus
- TaskStartDate
- TaskEndDate
- RecordsProcessed

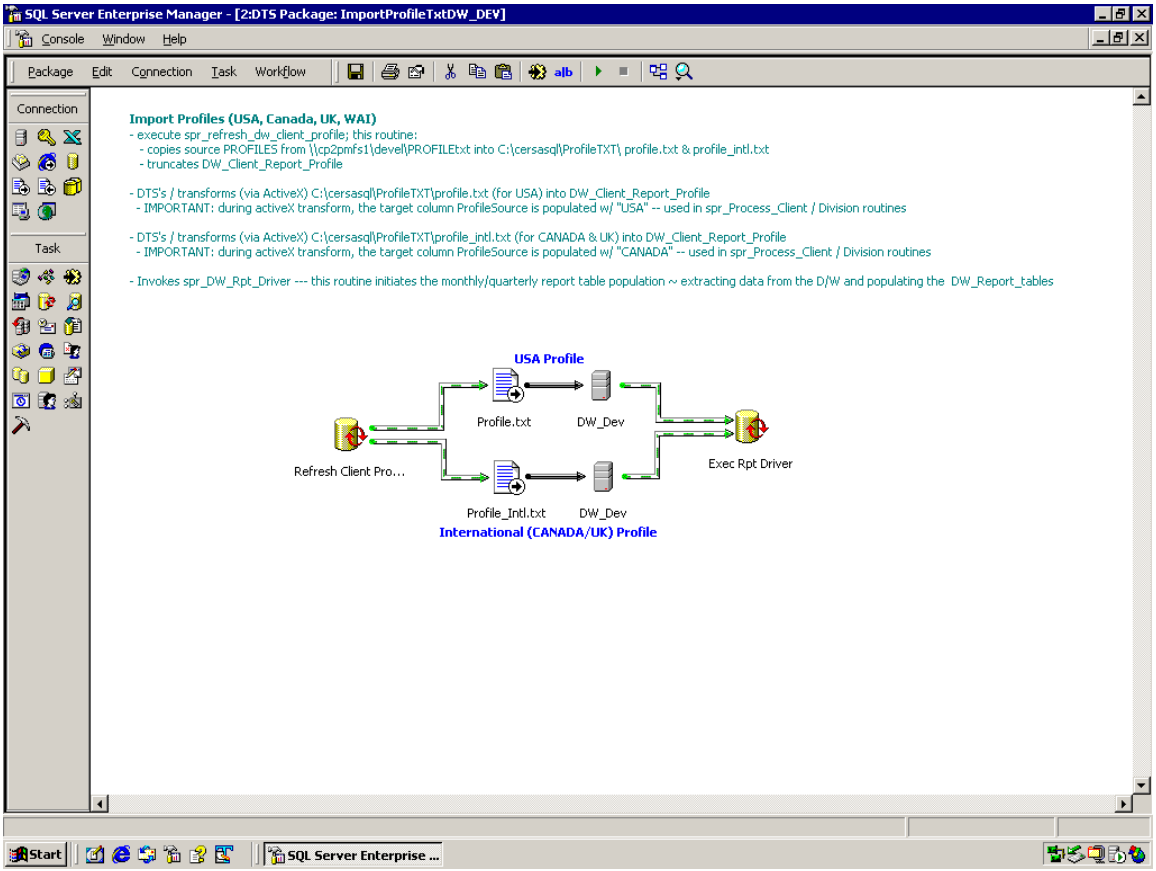
Sample ActiveX script used to capture this AuditTask information:

```
'**********************************************************************
'  Visual Basic ActiveX Script
'**********************************************************************

Function Main()
Dim sStepName
Dim sTaskName
Dim nLastRow
Dim oPackage
Dim oStep
Dim oTask
Dim oProperty
Dim oConn
Dim sSQL
Dim nCntr
Dim vRecordsProcessed

'!!!!!!!!!!!!!!!!!!!!!  Set Step Name !!!!!!!!!!!!!!!!!!!!
sStepName = "DTSStep_DTSExecuteSQLTask_4"

'Get Handle to Current DTS Package
Set oPackage = DTSGlobalVariables.Parent

'Find Step
For nCntr = 1 to oPackage.Steps.Count
  If oPackage.Steps(nCntr).Name = sStepName Then
     Exit For
  End If
Next
Set oStep = oPackage.Steps(nCntr)

'Set Step
sStepName = oStep.Description & " (" & oStep.Name & ")"

'Get Handle to Task
For nCntr = 1 to oPackage.Tasks.Count
  If oPackage.Tasks(nCntr).Name = oStep.TaskName Then
     Exit For
  End If
Next
Set oTask =  oPackage.Tasks(nCntr)

'If the previous task processed records, we can access the
'property, otherwise set to NULL
vRecordsProcessed = "NULL"
For Each oProperty In oTask.Properties
  If oProperty.Name = "RowsComplete" Then
    vRecordsProcessed = oProperty.Value
  End If
```

```
Next

'Build SQL Statement
sSQL = "INSERT INTO AuditTask (AuditID, PackageName, "
sSQL = sSQL & "TaskName, TaskStatus, TaskStartDate, "
sSQL = sSQL & "TaskEndDate, RecordsProcessed) VALUES ("
sSQL = sSQL & DTSGlobalVariables("gviAuditID").Value & ", "
sSQL = sSQL & "'" & oPackage.Name & "', "
sSQL = sSQL & "'" & sStepName & "', 'Successful', "
sSQL = sSQL & "'" & oStep.StartTime & "', '" & oStep.FinishTime
sSQL = sSQL & "'," & vRecordsProcessed & ")"

'Insert Row
Set oCONN = CreateObject("ADODB.Connection")
oConn.Open "File Name=" & _
  DTSGlobalVariables("gvsDW").Value
oConn.Execute sSQL

'Clean up
oConn.Close
Set oConn = Nothing

Main = DTSTaskExecResult_Success
End Function
```

- For each DTS SQL Task to be audited, the sStepName value must be modified in the ActiveX script above. (see **DTSStep_DTSExecuteSQLTask_4** above)
- To accomplish this, follow these guidelines:
    - Right mouse click the Exec SQL Task to be Audited
    - Click **Workflow**, **Workflow Properties**, then **Options**
    - Copy the Name into the sStepName of this tasks ActiveX Audit Task

- **Report Result Set Generation**



SQL Server Security

In order to allow the Reporting Team (role=**DW_REPORTING**) access to XP_CMDSHELL for Report Driver Invocation, perform the following:

- grant execute on the XP_CMDSHELL (in the Master DB) to **DW_Reporting**
- then, determine if a SQL Server Agent Proxy Account exists:

```
EXEC master.dbo.xp_sqlagent_proxy_account N'GET'
```

If no SQL Server Agent Proxy Account exists:

- review the examples/documentation below for setting the SQL Server Agent Proxy Account
- set the SQL Server Agent Proxy Account using the following DDL:

```
EXEC master.dbo.xp_sqlagent_proxy_account N'SET',
        N'CEN', -- agent_domain_name
        N'CERSA', -- agent_username
        N'sql_pwd_here' -- agent password
```

- under Enterprise Manager's **Management** tab:
    - right mouse click SQL Server Agent
    - click Properties then Job System and "un-check" box at bottom - Non-Sysadmin JobStep Proxy Account

Examples: Setting / Re-Setting the SQL Server Agent Proxy Account

A. Retrieve the currently assigned SQL Server Agent proxy account

This example retrieves the account currently assigned for use as the SQL Server Agent proxy account.

```
EXEC master.dbo.xp_sqlagent_proxy_account N'GET'

Domain            Username
-----------------------------------
NETDOMAIN                                   john
```

B. Set the SQL Server Agent proxy account without a password

This example sets the SQL Server Agent proxy account to LONDON\ralph without specifying a password. This example will receive an error that the extended stored procedure cannot log in if the LONDON/ralph account actually has a password.

```
EXEC master.dbo.xp_sqlagent_proxy_account N'SET',
        N'NETDOMAIN', -- agent_domain_name
        N'ralph', -- agent_username
        N'' - agent password
```

C. Set the SQL Server Agent proxy account with a password

This example sets the SQL Server agent proxy account to LONDON\Ralph and specifies a password.

```
EXEC master.dbo.xp_sqlagent_proxy_account N'SET',
        N'NETDOMAIN', -- agent_domain_name
        N'ralph', -- agent_username
        N'RalphPwd', - agent password
```

Metadata Catalog Maintenance Functions and Services

- Information catalog integration/merge (e.g., from the data model to the database to the front-end tool)
- Metadata management (e.g., remove old, unused entries)
- Capture existing metadata (e.g., DDL from mainframe or other sources)
- Manage and display graphical or tabular representations of the metadata catalog contents (the metadata browser)
- Maintain user profiles for application and security use
- Security for the metadata catalog
- Local or centralized metadata catalog support

Data Access Tool Capabilities

Push-Button Access

- Easy generation of sophisticated, interactive, engaging front-end screens
- Simple navigation controls
- Automatic replacement or on-the-fly creation of underlying report contents
- User interface controls for conceptual representations like stoplights (high, medium, low), gauges, and sophisticated charts
- Geographical charts with links to underlying reports
- "Alerters" or exception controls that monitor specified values, ranges, or differences and notify the user when they exceed target levels
- Ability to define and interact with multiple simultaneous connections to multiple data sources
- Standard reports

Ad Hoc Capabilities

- Query formulation
    - Multipass SQL
    - Highlighting
    - Successive constraints
    - Semiadditive summations
    - ANSI SQL 92 support
    - Direct SQL entry
- Analysis and presentation capabilities
    - Basic calculations on the results set
    - Pivot the results
    - Column calculations on pivot results
    - Column and row calculations
    - Sorting
    - Complex formatting
    - Charting and graphs
    - User-changeable variables
- User Interaction
    - Ease of use
    - Metadata access
    - Pick lists
    - Seamless integration with other applications
    - Export to multiple file types, including HTML
    - Embedded queries
- Technical features
    - Multitasking
    - Cancel query
    - Scripting
    - Connectivity
    - Scheduling
    - Metadata driven
    - Software administration
    - Security
    - Querying

Modeling Applications and Data Mining

- Clustering
- Classifying
- Estimating and predicting
- Affinity grouping
- Support for periodic pulls of large files
- Update access to the staging area to return the result of scoring or forecasting runs

Security Checklist

Thrusts of the Data Warehouse Security Program

- ❑ Develop security awareness marketing campaign
- ❑ Gain executive support
- ❑ Create a healthy respect for the security measures in place
- ❑ Set good examples
- ❑ Develop, implement and maintain data warehouse security policies
- ❑ Maintain vigilance
- ❑ Keep a suspicious attitude
- ❑ Continuous renewal of the security environment

Immediate Tactical Measures

- ❑ Install virus checking software everywhere.
- ❑ Keep up-to-date on virus alerts and changes in virus technology.
- ❑ Remove floppy disks drives from your environment.
- ❑ Remove local modems from your environment.
- ❑ Control all software installed on internal machines.
- ❑ Assign passwords that your users must memorize and use.
- ❑ Funnel all Internet access through an Internet proxy server.
- ❑ Monitor and control access to remote sites.
- ❑ Provide clear written guidelines for appropriate Internet use.
- ❑ Install a packet-filtering firewall to restrict access from the outside world to known IP addresses. Install a bastion server to intercept all service requests from the outside world except known service requests from known IP addresses, which you regard as trusted. Isolate the bastion server from the true internal network with a second packet-filtering firewall.
- ❑ Remove all unnecessary services from the bastion server so that if it is breached, there is very little the intruder can do. Follow modern security practices for trimming and isolating the functions available on the bastion server.
- ❑ Implement a program for security education and security appreciation. Target all levels in the organization, including executives.
- ❑ Implement a program for auditing threats to security, such as break-in attempts, failed login attempts, and inappropriate use.
- ❑ Implement a security tracking program that regularly reviews the security privileges of all employees (what information they can see), as well as the security exposures of all information resources (who has access to the data). Make sure that both on-line and backup media are covered by this analysis.
- ❑ Physically secure all servers and all backup media. Inspect and secure all communications facilities and cable vaults. Apply an electrical sweep of all networks and account for all taps and connections.

Strategic Security Measures

- ❑ Commit to an access token approach to replace all use of typed passwords (e.g., smart cards or biometric scanning) both internally and in the field. Include all contractors and industrial partners.
- ❑ Assign a public/private key combination to every end user to use as the basis for secure authentication. This pair of keys is probably coupled to the access token specified in the preceding paragraph.
- ❑ Commit to a secure tunneling approach for remote access by trusted individuals.
- ❑ Centralize all authentication and access control through a directory server based on the LDAP protocol. Require all users to funnel through the directory server whether they are internal user or external users. Administer all security from this one central point. Do not allow direct access to a database or application server by anyone.
- ❑ Require all software downloads to be based on signed certificates. Actively administer the list of trusted software vendors whose software you will accept.

Physical Database Design

Table / Column name	Data Type	Permit nulls?	Prim. Key	Comment

Index Plan

Table / Index Name	Index Type	Unique?	Columns	Location	Justification

Data Staging Checklist

Preliminary Work

- ❑ Set up a header format and comment fields for your code
- ❑ Hold structured design reviews early enough to allow changes
- ❑ Write clean, well-commented code
- ❑ Enforce naming standards
- ❑ Use the code library and management system
- ❑ Test everything—both unit testing and system testing
- ❑ Document everything—hopefully in the information catalog

Step 1. High-Level Plan

- ❑ Create a very high-level, one-page schematic of the source-to-target flow
- ❑ Identify starting and ending points
- ❑ Label known data sources
- ❑ Include placeholders for sources yet to be determined
- ❑ Label targets
- ❑ Include notes about known gotchas

Step 2. Data Staging Tools

- ❑ Test, choose, and implement a data staging tool

Step 3. Detailed Plan

- ❑ Drill down by target table, graphically sketching any complex data restructuring or transformations
- ❑ Graphically illustrate the surrogate-key generation process
- ❑ Develop a preliminary job sequencing

Step 4. Populate a Simple Dimension Table

- ❑ Static dimension extract
- ❑ Creating and moving the result set
    - ❑ Data compression
    - ❑ Data encryption
- ❑ Static dimension transformation
- ❑ Simple data transformations
- ❑ Surrogate key assignment
- ❑ Combining from separate sources
- ❑ Validating one-to-one and one-to-many relationships
- ❑ Load
    - ❑ Bulk loader
        - ❑ Turn off logging
        - ❑ Pre-sort the file
        - ❑ Transform with caution
        - ❑ Aggregations
        - ❑ Use the bulk loader to perform "within-database" inserts
- ❑ Truncate target table before full refresh
- ❑ Index management
    - ❑ Drop and re-index
    - ❑ Keep indexes in place
- ❑ Maintaining dimension tables
    - ❑ Warehouse-based maintenance
    - ❑ Source system–based maintenance

Step 5. Implement Dimension Change Logic

- Use surrogate keys
- Dimension table extracts
  - Copy entire current master file
  - Pull only changed rows – source system change flag
- Processing slowly changing dimensions
  - Type 1: Overwrite
  - Type 2: Create a new dimension record
  - Type 3: Push down the changed value into an "old" attribute field
- Dimension table transformation and load

Step 6. Populate Remaining Dimensions

- Repeat steps 4 & 5 for each remaining dimension

Step 7. Historical Load of Atomic-Level Facts

- Historic fact table extracts
- Capture audit statistics
- Fact table processing
  - Fact table surrogate key lookup
- Ensure proper handling of nulls
- Improving fact table content
- Data restructuring
- Data mining transformations
    - Flag normal, abnormal, out of bounds, or impossible facts
    - Recognize random or noise values from context and mask out
    - Apply a uniform treatment to null values
    - Flag fact records with changed status
    - Classify an individual record by one of its aggregates
    - Divide data into training, test, and evaluation sets
    - Add computed fields as inputs or targets
    - Map continuous values into ranges
    - Normalize values between 0 and 1
    - Convert from textual to numeric or numeral category
    - Emphasize the unusual case abnormally to drive recognition

Step 8. Incremental Fact Table Staging

- Incremental fact table extracts
    - New transactions
    - Updated transactions
    - Database logs
    - Replication
- Incremental fact table load
- Speeding up the load cycle
    - More frequent loading
    - Partitioned files and indexes
    - Parallel processing
        - Multiple load steps
        - Parallel execution
        - Parallel databases
        - Parallel tables

Step 9. Aggregate Table and MOLAP Loads

- ❑ Build aggregates
- ❑ Maintain aggregates
- ❑ Prepare MOLAP loads

Step 10. Warehouse Operation and Automation

- ❑ Typical operational functions
  - ❑ Job definition—flow and dependency
  - ❑ Job scheduling—time and event based
  - ❑ Monitoring
  - ❑ Logging
  - ❑ Exception handling
  - ❑ Error handling
  - ❑ Notification
- ❑ Determine job control approach
- ❑ Record extract metadata
- ❑ Record operations metadata
- ❑ Ensure data quality
- ❑ Set up archiving in the data staging area
- ❑ Develop disk space management procedures

Typical Job Schedule

- ❑ Extract dimensions—write out metadata
- ❑ Extract facts—write out metadata
- ❑ Process dimensions
  - ❑ Surrogate key/slowly changing processing/key lookup, etc.
  - ❑ Data quality checks—write out metadata
- ❑ Process facts
  - ❑ Surrogate key lookup—RI check—write out failed records
  - ❑ Data transformations
- ❑ Process aggregates
- ❑ Load dimensions into base level warehouse (dimensions first if RI is enforced)
- ❑ Load facts
- ❑ Load aggregates
- ❑ Review load process—validate load against metadata
- ❑ Change pointers or switch instance for high uptime (24 x 7), or parallel load warehouses
- ❑ Extract and load (or notify) downstream data marts (and other systems)
- ❑ Change metadata as needed (e.g., Period table attributes—current month)
- ❑ Write job metadata
- ❑ Review job logs, verify successful load cycle

Data Validation Checklist

Search for Common Data Problems

- ❑ Inconsistent or incorrect use of codes and special characters
- ❑ Single field used for unofficial or undocumented purposes
- ❑ Overloaded codes
- ❑ Evolving data
- ❑ Missing, incorrect, or duplicate values

Ensure Proper Name and Address Handling

- ❑ Name and address split into individual components
- ❑ Individual components cleaned and corrected
- ❑ All appropriate components completed
- ❑ Duplicates eliminated
- ❑ Cleaned data fed back into source systems
- ❑ Individuals grouped into households

Improving the Data

- ❑ Search out the highest quality source system
- ❑ Examine the source to see how good/bad it is. Our favorite, and rather low-tech, approach is to perform a frequency count on each attribute to identify variations in spellings.
- ❑ Correct variations in spelling, manually or preferably using a tool
- ❑ Raise problems with Steering Committee
- ❑ Fix problems at the source if at all possible
- ❑ Fix some problems during data staging
- ❑ If can't correct the data, be prepared to discuss where the data came from and why it looks the way it does
- ❑ Use data cleansing tools against the data, and use trusted sources for correct values like address
- ❑ Work with the source system owners to help them institute regular examination and cleansing of the source systems
- ❑ Make the source system organizationally responsible for a clean extract

Data Quality Assurance

- ❑ Define standards of acceptable data quality
- ❑ Accuracy - documented audit trail that explains any differences between data warehouse and system of record

  Basic Data Staging Audits
  - ❑ Correct number of rows processed
  - ❑ Referential integrity checking
  - ❑ Cross-footing
    - ❑ Set up series of queries against source system at different levels – compare to equivalent query against the data warehouse
    - ❑ Automate cross footing process
  - ❑ Manual examination
    - ❑ Look for numbers beyond acceptable ranges
    - ❑ Create set of 'reasonableness' data checks
  - ❑ Data staging process validation
    - ❑ Ensure that the process is sound

Data Access Audits

- ❑ Check report logic / calculations
- ❑ Confirm data access tool metadata set up properly
- ❑ Review data content
  - ❑ Meaningless descriptions
  - ❑ Duplicate dimension information
  - ❑ Incorrect dimensional relationships
  - ❑ Data not balancing

SQL Server Database Utilities

Backup Schedule & Retention Boundary

Recommended production SQL Server database backup schedule and retention boundaries.

Schedules

- 10 daily full image copy's (taken weeknites M-F @ 12am)
- 5 weekly full image copy's (taken Saturday nites @ 2am)
- 2 monthly full image copy's (taken 1st of the month @ 4am)
- 1 annual full image copy (taken 1st of the year @ 6am)

Rules

- Only the most recent daily full image copy is required to reside on disk. All other copies may be archived to tape.
- Log files will continue to be taken hourly between 7am and 11pm. Log files are good for 1 day only and will reside on disk - each hour is appended onto the backend of the previous.
- We'll plan on staggering the daily, weekly, monthly, and annual backup times allowing for multiple backups in one night. Example, Thursday Jan 1, 2004 would kick off the daily at Midnight, (the weekly (if this were a Saturday) at 2am, the Monthly at 4am, and the Annual at 6am. The oldest daily copy expires every 10th business day, the weekly copy expires every 5th week, the monthly copy expires

This recommendation is predicated on the individual business application team's approval. It can be easily customized.

Ancillary Database Utilities

The following SQL Server stored procedures are invoked from a driver stored procedure called **spr_whole_system_backup**. Listed below are their names and a brief desctiption of each. Note that a DECLARE CURSOR statement is contained w/in the driver stored procedure rendering only active databases to be operated on ~ therefore, as DB's are created and dropped there is no need to modify any of these utility jobs.

spr_backup_all_db	performs full database backups on all active databases
spr_reindex_all_db	performs a comprehensive reindexing of all active databases
spr_updateusage_all_db	reports and corrects inaccuracies in the sysindexes table
spr_updatestats_all_db	updates information pertaining to distribution of key values for one or more stats groups
spr_checkdb_all_db	checks allocations and the structural integrity of all objects within a database
spr_check_catalog_all_db	checks for inconsistencies in and between system table's on all active databases

REPORT DEVELOPMENT

Application Development Checklist

Develop End User Application Standards

- ❑ Standards for naming of queries, reports, filters, templates – all objects within the data access environment.
- ❑ Application location standards - The applications should be saved in a master location where they can easily be accessed by the users. The applications should also be read-only so that they can't be inadvertently changed.
- ❑ Output report standards - The report should have a consistent look and feel. It should include:
    - ❑ Page orientation - Landscape or portrait
    - ❑ Report header - Name of the report, parameters used, and center justified
    - ❑ Report body - Column/row layout of data
    - ❑ Data justification - Right justified for numbers, left justified for text
    - ❑ Data precision - Dependent on data
    - ❑ Column and row heading format
    - ❑ Formatting of totals or subtotal breakout rows
    - ❑ Header and Footer - The following items should be found somewhere in the header or footer:
        - ❑ Report notes (i.e., The market share calculation excluded market X)
        - ❑ Page numbering
        - ❑ Run time and date
        - ❑ Data source
        - ❑ Confidentiality statement
- ❑ Tool specific standards - Standards should be set up based on the characteristics of the particular tool used. For example, if the applications were built using Visual Basic, some programming standards should be defined such as the naming convention of variables and modules, or standard procedures that should be included in every application. If using a third-party decision support tools, then some standards might be where to store user defined calculations, how to name them, size of the user input dialog, etc.

End User Application Documentation Standards

- ❑ Document revision date - The date that the document was created/revised.
- ❑ Application summary - This should include the name of the application, where it can be found, a brief purpose of the application, and the output(s) created by the application.
- ❑ Report output example - Show an example of the report so the user can see the content and format.
- ❑ User selections/dialog - This should specify the parameters inputted by the user to run the report, such as period, geography, etc.
- ❑ Maintenance/upgrades - Identify the scope of the application (division specific, department specific, et. al.) and any year end considerations to keep the application working year to year.
- ❑ Application notes - Document any issues specific to the application that could be interpreted as bugs.
- ❑ Release information - Outline information such as the release date of the application, version number, name of developer, and a brief description (Initial release, Fix share calculation, et. al.).

End User Application Testing

- Ensure all applications adhere to application standards.
- Verify user input.
- Verify data and calculations.
- Conduct tests with live data.
- Have data from existing reports to compare against, if possible.
- Have users check the data in the report to see if the numbers are reasonable.
- Test the application in the system environment that it will be used in.
- Test different business scenarios. Choose different user input combinations to insure that all business cases are met.

User test and sign off

- Review the application with users.
    - Where to find it
    - How to run it
    - Review any special considerations
- Provide opportunity for user to run several different business scenarios on their own.
- Obtain formal user acceptance.

End User Application Template Definition

Template Name:

Description / Purpose:

Frequency:

User Inputs:

Default Constraints:

Calculations:

Notes:

End User Application Template Layout

Template Title
Title Details
Constraints

(Table below to contain report body)

	xxxxxx				xxxxxx				xxxxxx			
Label	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12
A	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9
B	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9
C	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9
D	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9
E	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9
F	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9
G	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9
etc.	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9	999,999	99.9	99.9	9.9

DEPLOYMENT

Desktop Installation Readiness Checklist

- ❑ Determine client configuration requirements to support end user data access software, including hardware configurations, ODBC connections, intranet and Internet connectivity, and so on.
- ❑ Determine LAN addresses for the identified target users if you are not already using dynamically assigned LAN addresses.
- ❑ Conduct a physical audit of the technology currently installed on these users' desks and compare it to the stated client configuration requirements.
- ❑ Complete contract and procurement process to acquire any necessary client hardware, software, and/or upgrades.
- ❑ Estimate lead time required to acquire hardware, software and/or upgrades.
- ❑ Acquire user logons and security approval as necessary for network and database access.
- ❑ Estimate lead time required to acquire user logins and security authorizations.
- ❑ Ensure security maintenance procedures are in place (e.g., force changes in passwords with specified frequency).
- ❑ Test installation procedures on a variety of machines. These procedures can be refined via the alpha and beta release processes discussed later in this chapter.
- ❑ Schedule the installation with the users to align with their data warehouse education.
- ❑ Install the hardware and/or software and complete installation testing. It is important to fully test each installation to verify the user's existing system has not been adversely impacted and to ensure that the appropriate connectivities have been established or retained.

Deployment Readiness Checklist

Desktop installation

- ❑ Technology in place for business end user access to data warehouse
- ❑ User logons and security authorizations obtained

Data Quality Verification And Reconciliation

- ❑ Data quality assurance testing performed
- ❑ Inconsistencies with historically reported investigated, resolved and documented
- ❑ Beta team business representative signed-off on data legibility, completeness, and quality

End User Applications

- ❑ End user applications developed and tested
- ❑ Beta team business representative sign-off on application template quality and business relevance

End User Education

- ❑ Introductory business end user education materials developed on data content, application templates, and end user tool usage
- ❑ Beta team business representatives sign-off on introductory user education offering
- ❑ Education delivery logistics (e.g., venue, projection capabilities, user PCs with necessary data and application access, education materials duplication, etc.) handled
- ❑ Production end users registered for education with appropriate approval from their managers

End User Support

- ❑ Support organization in place and thoroughly trained
- ❑ Support communication, bug report, and change request tracking procedures tested during beta period