ellucian LIVE

A first year's experience with EDW

from nothing to production on 4 stars/cubes

Presented by: Clifton Ivy
Purdue University
April 9, 2013
Session ID 2093

Session Rules of Etiquette

- Please hush your cell phone/pager
  (people still have pagers?)
- If you must leave the session early, please do so discretely
- Please avoid side conversations and pizza delivery during the session

Thank you for your cooperation!

Introduction

- Share what we at Purdue have learned as we implemented four EDW stars / cubes
- So you can:
  - learn from our mistakes
  - know more about what it might take to implement EDW
  - learn possibilities of what you can do with EDW
  - for your general education and amusement

Agenda Slide

- Why did we...
- Which stars and cubes did we work on?
- How did we change each star?
- How did we change each cube?
- Why / what / how did we automate?
- What was the “Big Question”? (about “Time” and the run calendar)

ellucian LIVE

Which stars and cubes did we work on?

Priorities, priorities...

Which stars and cubes...

- Started with Snapshot stars
- EDW stars in/out of scope by module
- Sorted “in scope” by interest in the result, who had time to work with it
- Identified first few to work on (Enrollment, Course Registration, Recruiting & Admissions)
- Next models determined by “what next?” (Receivable Customer)

ellucian LIVE

How did we change each star?

Details, details...

How did we change each star...

- Typical Systems Analysis process...
  - Start with the model from the EDW handbook
  - The published EDW metadata provides background for each field
  - Some questions have to be answered from the "...EXTR" pl/sql package
  - Identify "missing" data, and which dimension (or fact) table it goes into

How did we change each star...

How did we change each star...

SGHE Enterprise Data Warehouse
Dimension Table Target: WDT_CUSTOMER

Description	Customer Dimension table provides information for the person or entity with an accounts receivable account. This information includes delinquency status, number of bad cheques and collections assignment indicator. It uses the cleansing rules in the EDW for the sheet and long descriptions.

Target Column	Business Definition	Database Data Type	Source Name	Source Column	Local Source	Local Target	Star Nam
COLLECTION_IND	Indicates if this customer receivable account has been assigned to a collection agency. This will be set to yes when the collection agency count is greater than zero.	VARCHAR2(63)	RECEIVABLE_ACCOUNT	COLLECTION_COUNT	NO	NO	RECEIVABLE CUST
CUSTOMER_KEY	Key for the customer dimension.	NUMBER	ETL GENERATED	NA	NO	NO	RECEIVABLE CUST
DELINQUENCY	Delinquency code assigned to this receivable customer. This maybe blank. Typical VALUES will be 30, 60, 90	VARCHAR2(63)	RECEIVABLE_ACCOUNT	DELINQUENCY	NO	NO	RECEIVABLE CUST

How did we change each star...

```
    CURSOR get_ar_cust_pop(multi_src_in VARCHAR2) IS
      SELECT NVL(RAD.MIF_VALUE, mif_cleanse_value)                    multi_source,
             NVL(RAD.MIF_DESC, mif_cleanse_value)                     multi_source_sd,
             RAD.ZONE_VALUE                                           zone_value,
             RAD.DOMAIN_VALUE                                         domain_value,
             RAD.ACADEMIC_YEAR                                        academic_year,
             RAD.ACADEMIC_PERIOD                                      academic_period,
             RAD.SUB_ACADEMIC_PERIOD                                  sub_academic_period,
             RAD.DETAIL_CODE                                          detail_code,
             RAD.CATEGORY                                             category,
             TRUNC(SYSDATE) - TRUNC(NVL(RAD.BILL_DATE, SYSDATE))      bill_date_aging,
             TRUNC(SYSDATE) - TRUNC(RAD.EFFECTIVE_DATE)               effective_date_aging,
             RA.DELINQUENCY                                           delinquency,
             RA.NSF_COUNTER                                           nsf_count,
             DECODE(SIGN(RA.COLLECTION_COUNT),1,'Y','N')              collection_ind,
             RAD.ACCOUNT_UID                                          customer_uid,
             SUM(RAD.BALANCE)                                         balance,
             SUM(DECODE(RAD.DETAIL_CODE_TYPE,'C',RAD.AMOUNT,0) -
                 DECODE(RAD.DETAIL_CODE_TYPE,'P',RAD.AMOUNT,0))       amount_due,
-- Begin ODS207 fields added (note comma added here - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -,
             RAD.DETAIL_CODE_TYPE                                     detail_code_type,
             SUM(RAD.ACCOUNT_POSTED_AMOUNT)                           account_posted_amount,
             to_char(add_months(rad.posting_date, 6), 'YYYY')         posting_fiscal_year
-- End ODS207 fields added
        FROM MTT_ACCOUNT RA, MTT_ACCOUNT_DETAIL RAD
       WHERE RA.ACCOUNT_UID = RAD.ACCOUNT_UID
         AND NVL(RA.MIF_VALUE, mif_cleanse_value) = multi_src_in
         AND RAD.ACADEMIC_PERIOD = academic_period_in
         AND NVL(RAD.MIF_VALUE, mif_cleanse_value) = multi_src_in
    GROUP BY NVL(RAD.MIF_VALUE, mif_cleanse_value),
             NVL(RAD.MIF_DESC, mif_cleanse_value)
```


How did we change each star...

```
CURSOR get_general_student(multi_src_in VARCHAR2, person_in NUMBER) IS
  SELECT *
    FROM MST_GENERAL_STUDENT
   WHERE PERSON_UID = person_in
     AND ACADEMIC_PERIOD = academic_period_in
     AND PRIMARY_PROGRAM_IND = 'Y'
     AND NVL(MIF_VALUE, mif_cleanse_value) = multi_src_in;
```

How did we change each star...

```
CURSOR get_ar_cust_pop(multi_src_in VARCHAR2) IS
  SELECT NVL(RAD.MIF_VALUE, mif_cleanse_value)                    multi_source,
         NVL(RAD.MIF_DESC, mif_cleanse_value)                     multi_source_sd,
         RAD.ZONE_VALUE                                           zone_value,
         RAD.DOMAIN_VALUE                                         domain_value,
         RAD.ACADEMIC_YEAR                                        academic_year,
         RAD.ACADEMIC_PERIOD                                      academic_period,
         RAD.SUB_ACADEMIC_PERIOD                                  sub_academic_period,
         RAD.DETAIL_CODE                                          detail_code,
         RAD.CATEGORY                                             category,
         TRUNC(SYSDATE) - TRUNC(NVL(RAD.BILL_DATE, SYSDATE))      bill_date_aging,
         TRUNC(SYSDATE) - TRUNC(RAD.EFFECTIVE_DATE)               effective_date_aging,
         RA.DELINQUENCY                                           delinquency,
         RA.NSF_COUNTER                                           nsf_count,
         DECODE(SIGN(RA.COLLECTION_COUNT),1,'Y','N')              collection_ind,
         RAD.ACCOUNT_UID                                          customer_uid,
         SUM(RAD.BALANCE)                                         balance,
         SUM(DECODE(RAD.DETAIL_CODE_TYPE,'C',RAD.AMOUNT,0) -
           DECODE(RAD.DETAIL_CODE_TYPE,'P',RAD.AMOUNT,0))         amount_due,
-- Begin ODS207 fields added (note comma added here - - - - - - - - - - - - - ^
         RAD.DETAIL_CODE_TYPE                                     detail_code_type,
         SUM(RAD.ACCOUNT_POSTED_AMOUNT)                           account_posted_amount,
         to_char(add_months(rad.posting_date, 6), 'YYYY')         posting_fiscal_year
-- End ODS207 fields added
     FROM MTT_ACCOUNT RA, MTT_ACCOUNT_DETAIL RAD
    WHERE RA.ACCOUNT_UID = RAD.ACCOUNT_UID
      AND NVL(RA.MIF_VALUE, mif_cleanse_value) = multi_src_in
      AND RAD.ACADEMIC_PERIOD = academic_period_in
      AND NVL(RAD.MIF_VALUE, mif_cleanse_value) = multi_src_in
  GROUP BY NVL(RAD.MIF_VALUE, mif_cleanse_value),
         NVL(RAD.MIF_DESC, mif_cleanse_value)
```

Session ID 2093 13

How did we change each star...

```
--
  FOR multi_src_rec IN get_mtvparm('EDW EXTRACT PARAMETERS','MULTI_SOURCE_GROUP',multi_source_group_in) LOOP
    FOR ar_cust_rec IN get_ar_cust_pop(multi_src_rec.mtvparm_external_code) LOOP
--
      OPEN get_general_student(multi_src_rec.mtvparm_external_code, ar_cust_rec.customer_uid);
      FETCH get_general_student INTO gen_student_rec;
      CLOSE get_general_student;
--
      -- multi source
      ret_row.multi_source := ar_cust_rec.multi_source;
      ret_row.process_group := ar_cust_rec.zone_value;
      ret_row.administrative_group := ar_cust_rec.domain_value;
      ret_row.msrc_user_attribute_01 := NULL;
      ret_row.msrc_user_attribute_02 := NULL;
      ret_row.msrc_user_attribute_03 := NULL;
      ret_row.msrc_user_attribute_04 := NULL;
      ret_row.msrc_user_attribute_05 := NULL;
      -- time
      ret_row.event_qualifier := '*';
      ret_row.event := event_in;
      ret_row.event_date := event_date_value;
      ret_row.aid_year := NULL;
      ret_row.aid_period := NULL;
      ret_row.academic_year := ar_cust_rec.academic_year;
      ret_row.academic_period := ar_cust_rec.academic_period;
      ret_row.sub_academic_period := ar_cust_rec.sub_academic_period;
-- Begin ODS207 fields added
      --ret_row.fiscal_year := NULL;
      ret_row.fiscal_year := ar_cust_rec.posting_fiscal_year;
-- End ODS207 fields added
```

Session ID 2093 14

How did we change each star...

- What data for Receivable Customer?
  - Student Residency; in Academic_Study user_attribute_02
  - Fiscal Year; in Time, into the fiscal_year field
  - Account Posted Amount; in fact table; into user_measure_01
  - ...

Session ID 2093 15

How did we change each star...

- What data for Receivable Customer?
  - Category Group; into Accounts_Receivable user_attribute_01:

Banner EDW Value	Banner EDW Long Description
05	Tuition
10	Fees
15	Fees Non-Instructional
20	Housing
25	Remissions
30	Loans
35	Scholarships, Awards, and Gifts
40	Federal Gift Aid
45	Fellowships
50	Third Party Sponsors
55	Payments
60	Other

ellucian LIVE

How did we change each star...

- What data for Receivable Customer?
  - Student Cost; into Accounts_Receivable user_attribute_02:

Banner EDW Value	Banner EDW Long Description
10	Net Student Cost
20	Student Housing Cost
30	Student Loans
40	Other

ellucian LIVE

How did we change each star...

- Stars share dimension tables – How do we track which user attributes are in use?
  - Thank you, MTVPARM!

```
select t.mtvparm_internal_code_2,
  decode(t.mtvparm_desc, 'MULTI_SOURCE', null, t.mtvparm_external_code) as User_attribute,
  decode(t.mtvparm_desc, 'MULTI_SOURCE', null, t.mtvparm_internal_code) as Model,
  decode(t.mtvparm_desc, 'MULTI_SOURCE', null, t.mtvparm_desc) Decode_using,
  count(*) Cnt
from ia_admin.mtvparm t
where t.mtvparm_internal_code_group = 'CLEANSING DATA ELEMENTS'
--  and t.mtvparm_internal_code = 'ENROLLMENT'
  and t.mtvparm_external_code like '%USER_ATTR%'
--  and t.mtvparm_desc <> 'MULTI_SOURCE'
group by t.mtvparm_internal_code_2,
  decode(t.mtvparm_desc, 'MULTI_SOURCE', null, t.mtvparm_external_code),
  decode(t.mtvparm_desc, 'MULTI_SOURCE', null, t.mtvparm_internal_code),
  decode(t.mtvparm_desc, 'MULTI_SOURCE', null, t.mtvparm_desc)
order by 1, 2, 3 , 4
```


ellucian LIVE

ellucian LIVE

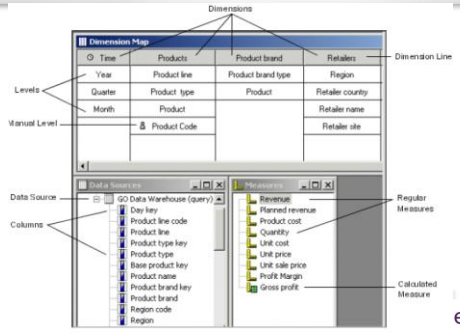
How did we change each cube?

Brick by brick...

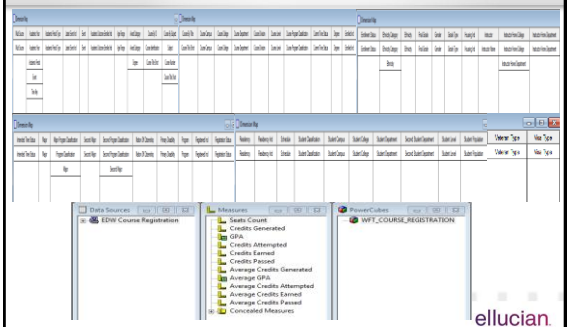
Background on cubes:

- Cube dimension fields mostly come from Star dimension table fields
- Ideal dimension is hierarchical, for drill-down capability
- But many are not (gender, ethnicity...)
- Cube “measures” mostly come from Fact table measure fields
- Measure fields must be additive

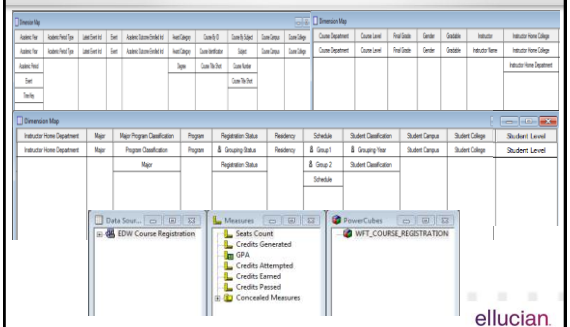
Typical Cube Design Process



Working from Ellucian supplied model...



What did we end up with...

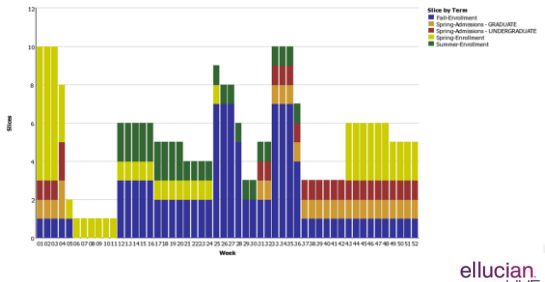


Why, What, and How did we Automate?

Tick tock, tick tock

Why, What, and How did we Automate?

- Enrollment and Admissions Slices – Total 270



Why, What, and How did we Automate?

- Calendar of EDW Events by Star
- PL/SQL procedure to run Star as by the Admin Interface
- Second PL/SQL procedure to run the first, based on the Calendar of Events
- UC4 Jobs to run the second procedure
- Build Cognos cubes in batch
- Emails to report results of the above

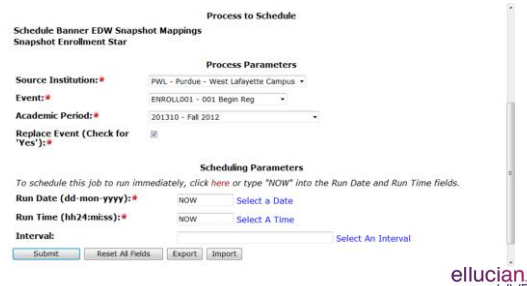
Why, What, and How did we Automate?

Event Calendar

- Determine the business cycle for captures
- Define naming convention for Event Codes
- Create the calendar; review; review; review

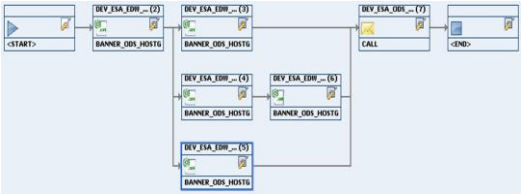
DATE	TERM	CODE	SHORT DESCR	LONG DESCRIPTION	Begin Capturing
8/6/2012	201410	ADM001	001 BEGIN APPL	001 Begin Application Cycle	Weekly Monday morning starting first Monday in August
8/13/2012	201410	ADM008	008 Cycle Wk 01	008 01 Week of Application Cycle	
8/20/2012	201410	ADM015	015 Cycle Wk 02	015 02 Week of Application Cycle	
8/27/2012	201410	ADM022	022 Cycle Wk 03	022 03 Week of Application Cycle	
9/3/2012	201410	ADM029	029 Cycle Wk 04	029 04 Week of Application Cycle	
9/10/2012	201410	ADM036	036 Cycle Wk 05	036 05 Week of Application Cycle	
9/17/2012	201410	ADM043	043 Cycle Wk 06	043 06 Week of Application Cycle	
9/24/2012	201410	ADM050	050 Cycle Wk 07	050 07 Week of Application Cycle	First capture this cycle
10/1/2012	201410	ADM057	057 Cycle Wk 08	057 08 Week of Application Cycle	
10/8/2012	201410	ADM064	064 Cycle Wk 09	064 09 Week of Application Cycle	
10/15/2012	201410	ADM071	071 Cycle Wk 10	071 10 Week of Application Cycle	

Why, What, and How did we Automate?



Why, What, and How did we Automate?

- One JobPlan in UC4
- LOAD_CLEANSING to pick up new codes
- Stack Jobs for each Star



- Run the notification email to summarize what was run

Why, What, and How did we Automate?

- Can run in UC4 or as a ".bat" file

```
:SET &DATE# = SYS_DATE("YYYYMMDD")
:SET &TIME# = SYS_TIME("HHMMSS","TZ.GMT")
:SET &TIMESTAMP# = '&DATE#_&TIME#'
:PRINT &TIMESTAMP#
:SET &CLNT# = SYS_ACT_CLIENT()

"D:\cog10_Transformer\bin\cogtr.exe" $
-dLogFileDirectory="\\itifs02\Cognos\Cubes\logfiles\uc4_&CLNT#" $
-dLogFileName="&CUBENAME#_&TIMESTAMP#.log" $
-dCubeSaveDirectory="\\itifs02\Cognos\Cubes\cubefiles\uc4_&CLNT#" $
-lFRED -c -g $
-m"\\itifs02\Cognos\Cubes\modelfiles\uc4_&CLNT#\&CUBENAME#.mdl" $
-n1 -s
```

Why, What, and How did we Automate?

- Typical daily email:

```
From: itap-uc4@lists.purdue.edu
Sent: Friday, September 28, 2012 3:20 AM
To:   Ivy, Clifton (and etc, etc, etc)
Subject: WL PRODODS PRD_ESA_EDW_STAR_RUN1_JOBP has run;
    No errors
#16825 OK; EDW cleansing:;
#16826 OK; LOAD_EDW_INST_OPERATIONAL_STAR;
#16827 OK;
    EDWstar:LOAD_EDW_COURSE_REGISTRATION/SWK06D5/201310;
#16828 OK; EDWstar:LOAD_EDW_ENROLLMENT/SWK06D5/201310;
LOAD_EDW_RECRUITING_ADMISSIONS - nothing to run
    09/28/2012;
```


ellucian LIVE

ellucian LIVE

Time, Events, and the Big Question

Sometimes, you can

go back home again...

Time, Events, and the Big Question

- In April, 2012, we found an event naming problem
- For Enrollment and Course Registration
- Events named ENROLL001, 2, 3... from first day of registration
- But Fall has more lead time than Spring

ellucian LIVE

Time, Events, and the Big Question

Questions:

- What is a better way to name events?
- Can we fix Purdue's EDW calendar table?
- When can we do this?
- And the **Big Question**:

Can we fix the already-captured Events?

Or are we doomed to live with this?

Time, Events, and the Big Question

Discoveries:

- Time is a "sort-of" shared dimension
- Columns include System_Load_Process
- It is which Star created this Time record
- Table is shared – rows are not shared
- Can update Enrollment and Course Registration "Time" rows without messing up all of EDW

Time, Events, and the Big Question

Action:

- Create table of before / after codes and descriptions
- Back up WDT_TIME, Purdue's calendar table
- Use before/after table to update Purdue's EDW run Calendar table
- Then used it to update WDT_TIME

Time, Events, and the Big Question

Remember:

- Test, Test, Test
- Don't forget MGRCVAL, MGRCDES
- To PRODODS in June, 2012
- Celebrate!

Summary

- Vanilla? Or Neopolitan? Or Moose Tracks?
- Prioritize Stars; work your list
- Analyze star data and your needs; customize as needed (carefully!)
- Ditto for the cubes
- Event naming scheme is important
- Be perfect or prepare to punt

Questions & Answers

- Well, to tell you the truth, in all this excitement I kind of lost track myself...

Thank You!

Clifton Ivy
clif@purdue.edu

Please complete the online session evaluation form
Session ID 2093