The design of the Student Records Dimensional Data Model and ETL process @ University of Cape Town

Session #7001

July 15, 2015

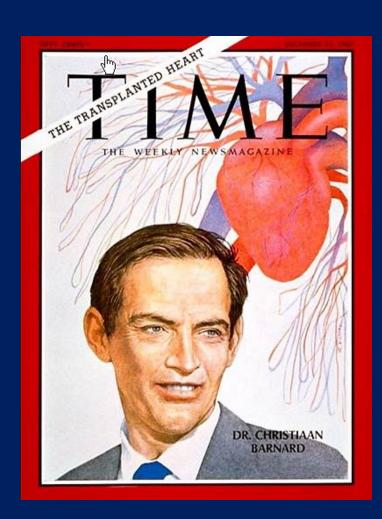# Agenda

- Introduction

- Collecting Requirements

- Dimensional Data Model

- Extract Transform Load (ETL)
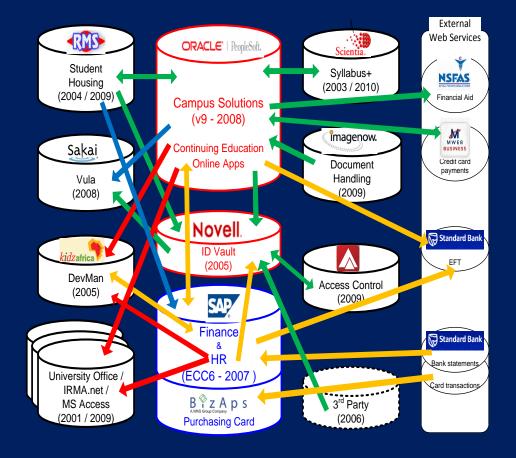
- Dashboard

- Questions

University of Cape Town

- Over 25 000 Students

- Top-rated African University

- Strong Research Focus

- 5 Nobel Prize Winners

- Worlds First Heart Transplant

THE TRANSPLANTED HEART

TIME

THE WEEKLY NEWSMAGAZINE

DR. CHRISTIAAN BARNARD

- Oracle PeopleSoft
  Campus Solutions 9

  - Up to Bundle
    #33 Applied
  - People Tools
    8.53.14

- SAP Finance

- SAP HR

Introduction

- Goal of Data Warehousing BI initiative

  - Easy access to data

  - Slice and dice data from every aspect

  - Make decisions, rather than argue about numbers

Collecting Requirements

- Select one Business Process

- Identify Business Sponsor/ Champion
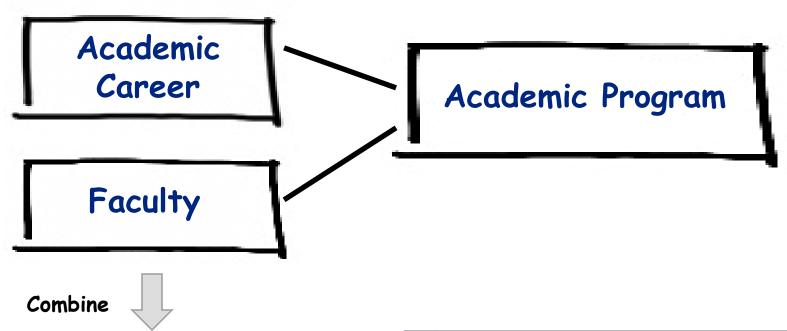
- Focus on user needs

- Focus on KPI's

Dimensional Data Model

- Design Critical

- Get good training

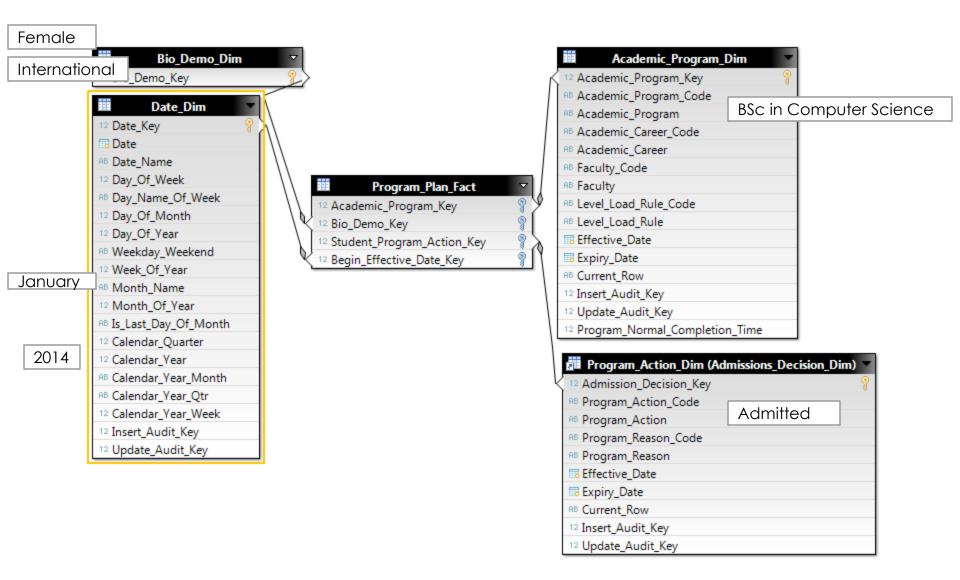- Make everything as simple as possible

Dimensions

Academic Career

Academic Program

Faculty

Combine

Program Dimension

Academic Program
Academic Career
Faculty

	Academic_Program	Academic_Career	Faculty
1	BSc Chem Eng/Cert Theory Acc	Undergraduate	Engineering/ Built Environment
2	BSc Eng Elec & Electron Eng B	Undergraduate	Engineering/ Built Environment
3	BSc Eng Elect & Electron Eng	Undergraduate	Engineering/ Built Environment
4	BSc Eng ElectroMechanical Eng	Undergraduate	Engineering/ Built Environment
5	BSc Eng in Material Eng	Undergraduate	Engineering/ Built Environment
6	BSc Eng in Mechatronics	Undergraduate	Engineering/ Built Environment
7	BSc Engineering (Aspect)	Undergraduate	Engineering/ Built Environment
8	BSc (Honours)	Honours	Science
9	BSc Biology,Earth&Env Sciences	Undergraduate	Science
10	BSc Chem,Molec & Cell Sciences	Undergraduate	Science

Dimension Schema For Student Program/Plan

Bio_Demo_Dim
- Female
- International
- _Demo_Key

Date_Dim
- 12 Date_Key
- 📅 Date
- AB Date_Name
- 12 Day_Of_Week
- AB Day_Name_Of_Week
- 12 Day_Of_Month
- 12 Day_Of_Year
- AB Weekday_Weekend
- 12 Week_Of_Year
- AB Month_Name — January
- 12 Month_Of_Year
- AB Is_Last_Day_Of_Month
- 12 Calendar_Quarter
- 12 Calendar_Year — 2014
- AB Calendar_Year_Month
- AB Calendar_Year_Qtr
- 12 Calendar_Year_Week
- 12 Insert_Audit_Key
- 12 Update_Audit_Key

Program_Plan_Fact
- 12 Academic_Program_Key
- 12 Bio_Demo_Key
- 12 Student_Program_Action_Key
- 12 Begin_Effective_Date_Key

Academic_Program_Dim
- 12 Academic_Program_Key
- AB Academic_Program_Code
- AB Academic_Program — BSc in Computer Science
- AB Academic_Career_Code
- AB Academic_Career
- AB Faculty_Code
- AB Faculty
- AB Level_Load_Rule_Code
- AB Level_Load_Rule
- 📅 Effective_Date
- 📅 Expiry_Date
- AB Current_Row
- 12 Insert_Audit_Key
- 12 Update_Audit_Key
- 12 Program_Normal_Completion_Time

Program_Action_Dim (Admissions_Decision_Dim)
- 12 Admission_Decision_Key
- AB Program_Action_Code
- AB Program_Action — Admitted
- AB Program_Reason_Code
- AB Program_Reason
- 📅 Effective_Date
- 📅 Expiry_Date
- AB Current_Row
- 12 Insert_Audit_Key
- 12 Update_Audit_Key

Surrogate Key for Dimensions

- Non-meaningful Integer

- Joins dimension and fact tables

- Benefits
  - Isolate the data warehouse from operational changes
  - Improve performance
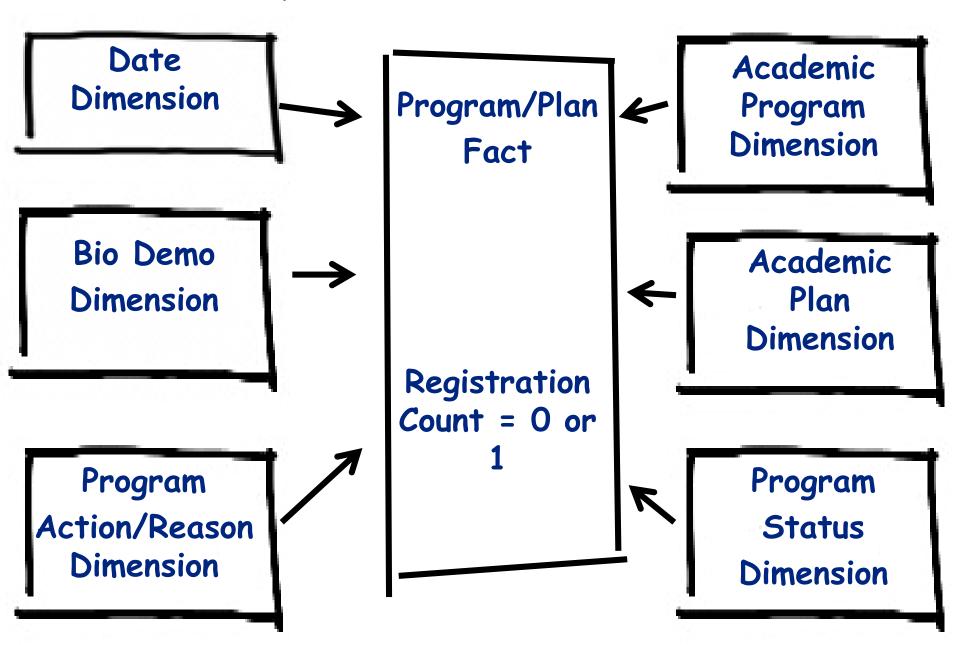 - Supports dimension change tracking

Fact Tables

Business Process

Student
accepts
offer

Translates into →

Program Plan Fact

Bio Demo Key

Academic Program Key

Academic Plan Key

Program Status Key

Effective Date Key

Registration Count
(either 0 or1)

Dimensional Data Model

Date Dimension →	**Program/Plan Fact**	← **Academic Program Dimension**
Bio Demo Dimension →		← **Academic Plan Dimension**
Program Action/Reason Dimension ↗	**Registration Count = 0 or 1**	↖ **Program Status Dimension**

Program/ Plan Fact Table

1. Student Accepts Offer

Begin Effective Date	End Effective Date	Program Status	Program Action	Program Reason
2011-02-14	9999-12-31	Active in Program	Matriculation	

2. Data Change

Begin Effective Date	End Effective Date	Program Status	Program Action	Program Reason
2011-02-14	2013-03-19	Active in Program	Matriculation	
2013-03-19	9999-12-31	Active in Program	Data Change	Change Graduation Term

3. Completes Program

Begin Effective Date	End Effective Date	Program Status	Program Action	Program Reason
2011-02-14	2013-03-19	Active in Program	Matriculation	
2013-03-19	2013-12-04	Active in Program	Data Change	Change Graduation Term
2013-12-04	9999-12-31	Completed Program	Completion of Program	

Select Count(Campus Id) From Program Plan Fact

Where Begin Effective Date <= 2013-12-01 AND

End Effective Date > 2013-12-01 AND

Program Action = 'Data Change'

Program/ Plan Fact Table

1. Student Accepts Offer

Begin Effective Date	End Effective Date	Program Status	Program Action	Program Reason
2011-02-14	9999-12-31	Active in Program	Matriculation	

2. Data Change

Begin Effective Date	End Effective Date	Program Status	Program Action	Program Reason
2011-02-14	2013-03-19	Active in Program	Matriculation	
2013-03-19	9999-12-31	Active in Program	Data Change	Change Graduation Term

3. Completes Program

Begin Effective Date	End Effective Date	Program Status	Program Action	Program Reason
2011-02-14	2013-03-19	Active in Program	Matriculation	
2013-03-19	2013-12-04	Active in Program	Data Change	Change Graduation Term
2013-12-04	9999-12-31	Completed Program	Completion of Program	

Junk Dimensions

Enrolment_Key	Enrolment_Status	Reason	Units_Attempted	Include_In_GPA	Earn_Credit
1499	Enrolled	Enrolled	No	Y	N
1500	Enrolled	Enrolled	Yes	Y	N
1501	Enrolled	Enrolled			Y
1502	Enrolled	Enrolled	In Progress		Y
1503	Enrolled	Enrolled	No		Y
1504	Enrolled	Enrolled	Yes		Y
1505	Enrolled	Enrolled		N	Y
1506	Enrolled	Enrolled	In Progress	N	Y
1507	Enrolled	Enrolled	No	N	Y
1508	Enrolled	Enrolled	Yes	N	Y
1509	Enrolled	Enrolled		Y	Y
1510	Enrolled	Enrolled	In Progress	Y	Y
1511	Enrolled	Enrolled	No	Y	Y
1512	Enrolled	Enrolled	Yes	Y	Y
1513	Enrolled	Enrolled from Wait List			

Junk Dimensions

Cartesian Product

SELECT A. STDNT_ENRL_STATUS
, B. ENRL_STATUS_REASON
, C. UNITS_ATTEMPTED
, D. INCLUDE_IN_GPA
, E. EARN_CREDIT
FROM PS_UCT_ENRL_STS_VW A
, PS_UCT_ENRL_SRSNVW B
, PS_UCT_UNIT_ATM_VW C
, PS_UCT_YES_NO_VW D
, PS_UCT_YES_NO_VW E
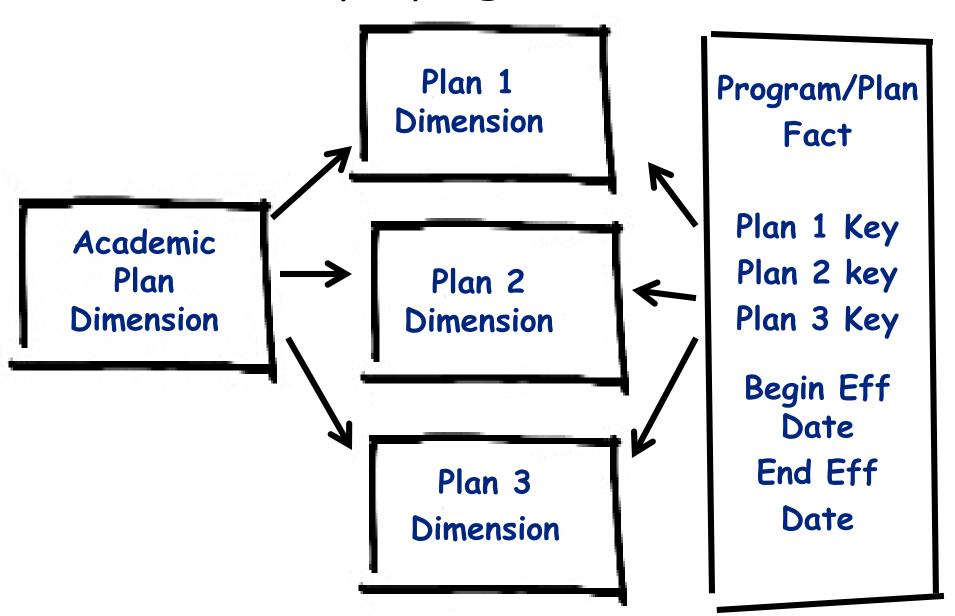
2448 Rows

Less than 1 second to execute

Create row as you encounter them in the source data
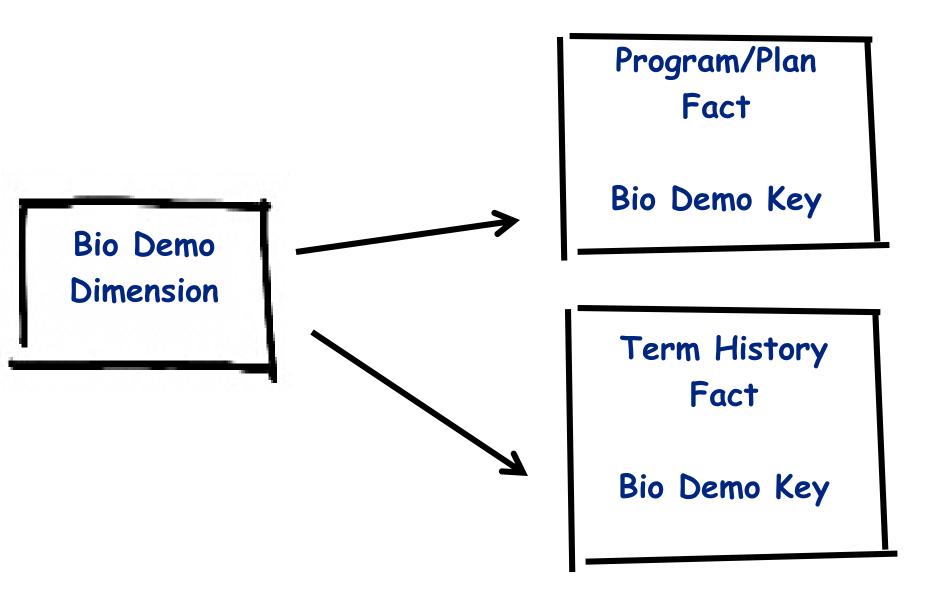
SELECT DISTINCT STDNT_ENRL_STATUS
, ENRL_STATUS_REASON
, UNITS_ATTEMPTED
, INCLUDE_IN_GPA
, EARN_CREDIT
FROM PS_STDNT_ENRL A
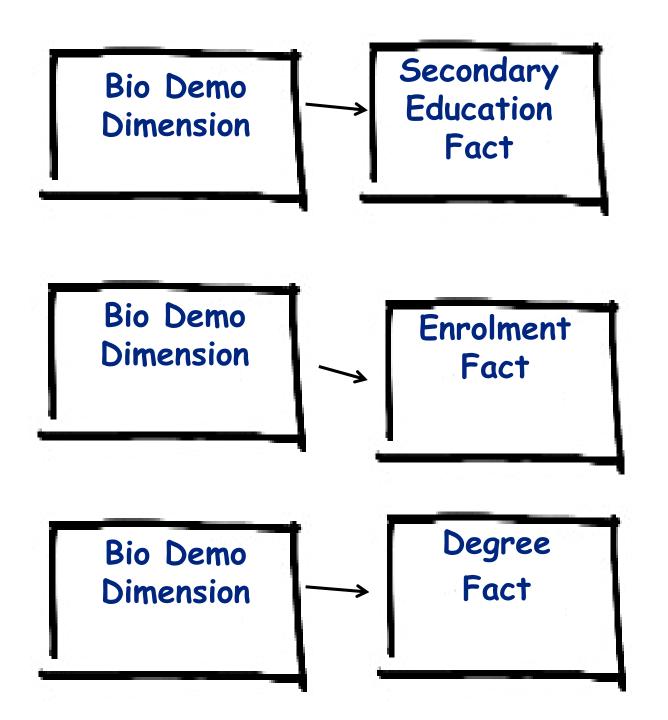
41 Rows

37 seconds to execute

Will get slower as more Enrolment records added

Role-playing Dimensions

Academic Plan Dimension

Plan 1 Dimension

Plan 2 Dimension

Plan 3 Dimension

Program/Plan Fact

**Plan 1 Key
Plan 2 key
Plan 3 Key**

**Begin Eff Date
End Eff Date**

Conformed Dimensions

Bio Demo Dimension

Program/Plan Fact

Bio Demo Key

Term History Fact

Bio Demo Key

Slowly Changing Dimensions

- Type 1: Overwrite
  - Use only if no value in keeping the old description

Original Row in Academic Program dimension

Academic Program Key	Academic Program Code	Academic Program	Faculty
12345	EB001	BScEng in Chemical Engineering	Engineering/ Built Environment

Update Row in Academic Program dimension

Academic Program Key	Academic Program Code	Academic Program	Faculty
12345	EB001	BScEng in Chemical Engineering	Science

- Advantages
  - Easy and fast to implement
- Disadvantages
  - Lose history of attribute changes

Slowly Changing Dimensions

- Type 2: Add a New Row
  - Use if need to preserve history

Original Row in Academic Program dimension

Academic Program Key	Academic Program Code	Academic Program	Faculty	Effective Date	Expiry Date	Current Row
12345	EB001	BScEng in Chemical Engineering	Engineering/ Built Environment	1900-01-01	9999-12-31	Y

Rows in Academic Program dimension following Faculty re-assignment

Academic Program Key	Academic Program Code	Academic Program	Faculty	Effective Date	Expiry Date	Current Row
12345	EB001	BScEng in Chemical Engineering	Engineering/ Built Environment	1900-01-01	2014-01-31	N
25987	EB001	BScEng in Chemical Engineering	Science	2014-02-01	9999-12-31	Y

- Advantages
  - Preserve history of attribute changes
- Disadvantages
  - ETL is more complex

Slowly Changing Dimensions

○ Type 2: Add a New Row

Academic Program Dimension

Academic Program Key	Academic Program	Faculty	Effective Date	Expiry Date	Current Row
12345	BScEng in Chem Eng	EBE	1900-01-01	9999-12-31	Y
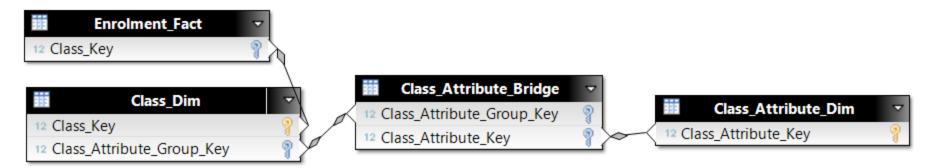
Program Plan Fact

Academic Program Key	Begin Effective Date	Program Action
12345	2011-02-14	Matriculation
12345	2013-03-19	Data Change

Student Completes the program after the program is moved to the Science Faculty

Academic Program Dimension

Academic Program Key	Academic Program	Faculty	Effective Date	Expiry Date	Current Row
12345	BScEng in Chem Eng	EBE	1900-01-01	2014-01-31	N
25987	BScEng in Chem Eng	SCI	2014-02-01	9999-12-31	Y

Program Plan Fact

Academic Program Key	Begin Effective Date	Program Action
12345	2011-02-14	Matriculation
12345	2013-03-19	Data Change
25987	2014-02-10	Completion of Program

Bridge table for Multiple Class Attributes

Enrolment_Fact
| 12 Class_Key |

Class_Dim
| 12 Class_Key |
| 12 Class_Attribute_Group_Key |

Class_Attribute_Bridge
| 12 Class_Attribute_Group_Key |
| 12 Class_Attribute_Key |

Class_Attribute_Dim
| 12 Class_Attribute_Key |

Step 1. Create Class Attribute Dimension

Step 2. Populate the Class Attribute Bridge table
            by grouping on
                        Course Id
                        Course Offer Number
                        Term Year Code
                        Session Code
                        Class Section
Step 3. Add the Class Attribute Group Key to the Class Dimension
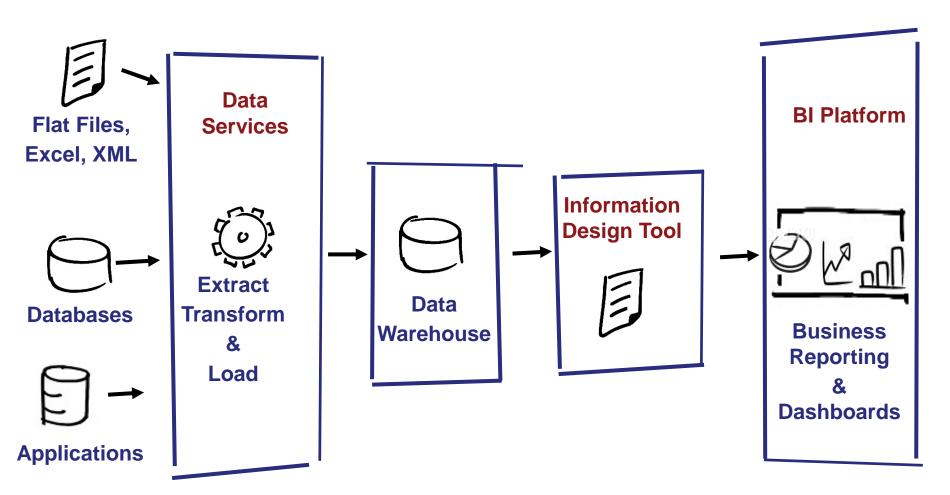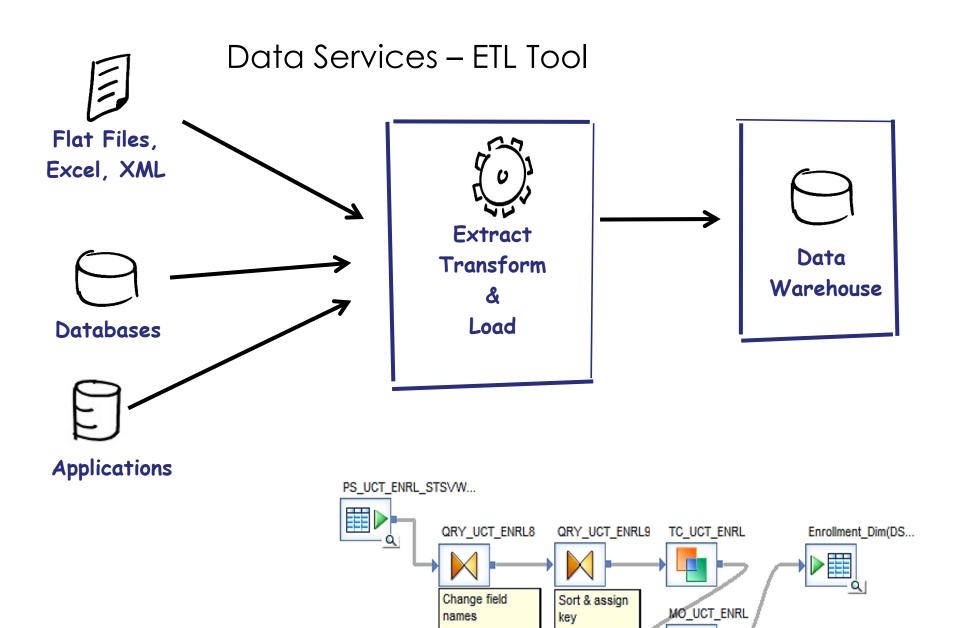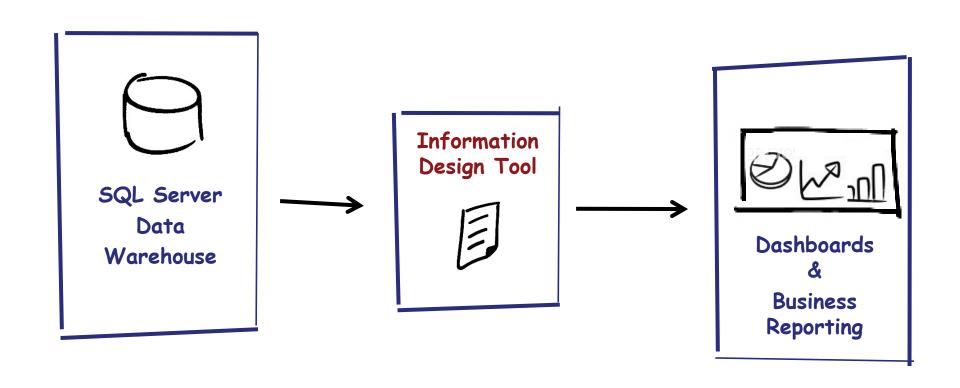
What is Business Objects ?

Flat Files, Excel, XML

Databases

Applications

Data Services

Extract Transform & Load

Data Warehouse

Information Design Tool

BI Platform

Business Reporting & Dashboards

Data Services – ETL Tool

Flat Files, Excel, XML

Databases

Applications

Extract Transform & Load

Data Warehouse

PS_UCT_ENRL_STSVW...

QRY_UCT_ENRL8

QRY_UCT_ENRL9

TC_UCT_ENRL

Enrollment_Dim(DS...

Change field names

Sort & assign key

MO_UCT_ENRL

Information Design Tool

SQL Server
Data
Warehouse

Information
Design Tool

Dashboards
&
Business
Reporting

Business Layer

Student Records Dashboard

Student Records Dashboard

Student Records Dashboard

Student Records Dashboard

Student Records Dashboard

Student Records Dashboard

Where to from here ?

Project Name	2014				2015										
	Q1	Q2	Q3	Q4	J	F	M	A	M	J	J	A	S	O	N	D
Admissions Data Curation																
Business Objects Enterprise Upgrade 4.1																
Business Objects Enterprise Upgrade 4.1 SP5																
Employment Equity Reporting																
Exit Survey Data Mart and Initial Reporting																
Faculty Reporting (EBE Pilot)																
Financial Aid Data Mart and Initial Reporting																
Incorporate Graduate Research Management into																
Lab Statistics Data Mart and Initial Reporting																
Library Usage Data Mart and Initial reporting																
NBT Data Mart and Initial Reporting																
NSC Results Data Mart and Initial Reporting																
Overtime, Special Leave, Standby and Additional																
Redevelop SAP Finance Data Mart and ETL																
SAP HR Data Mart and Initial reporting																
SAP MyFinances																
Student Housing Data Mart and Initial Reports																
Systems Access Data Mart and Initial Reporting																
Timetabling Data Mart and Venue Planning																
Student Financials Data Dart and Initial Reporting																

Legend:
- Scheduled, start date not reached
- On hold
- In progress and on schedule
- In Progress, minor issues
- Overdue/behind schedule
- Complete

Questions?

Contact

Sean Bester

Systems Division, University of Cape Town

sean.bester@uct.ac.za

+27 21 650 4407

References
The Data Warehouse Toolkit (Third Edition)
The Definitive Guide to Dimensional Modeling
Ralph Kimball and Margy Ross

This presentation and all Alliance presentations are available for download from the Conference site at www. heug.org