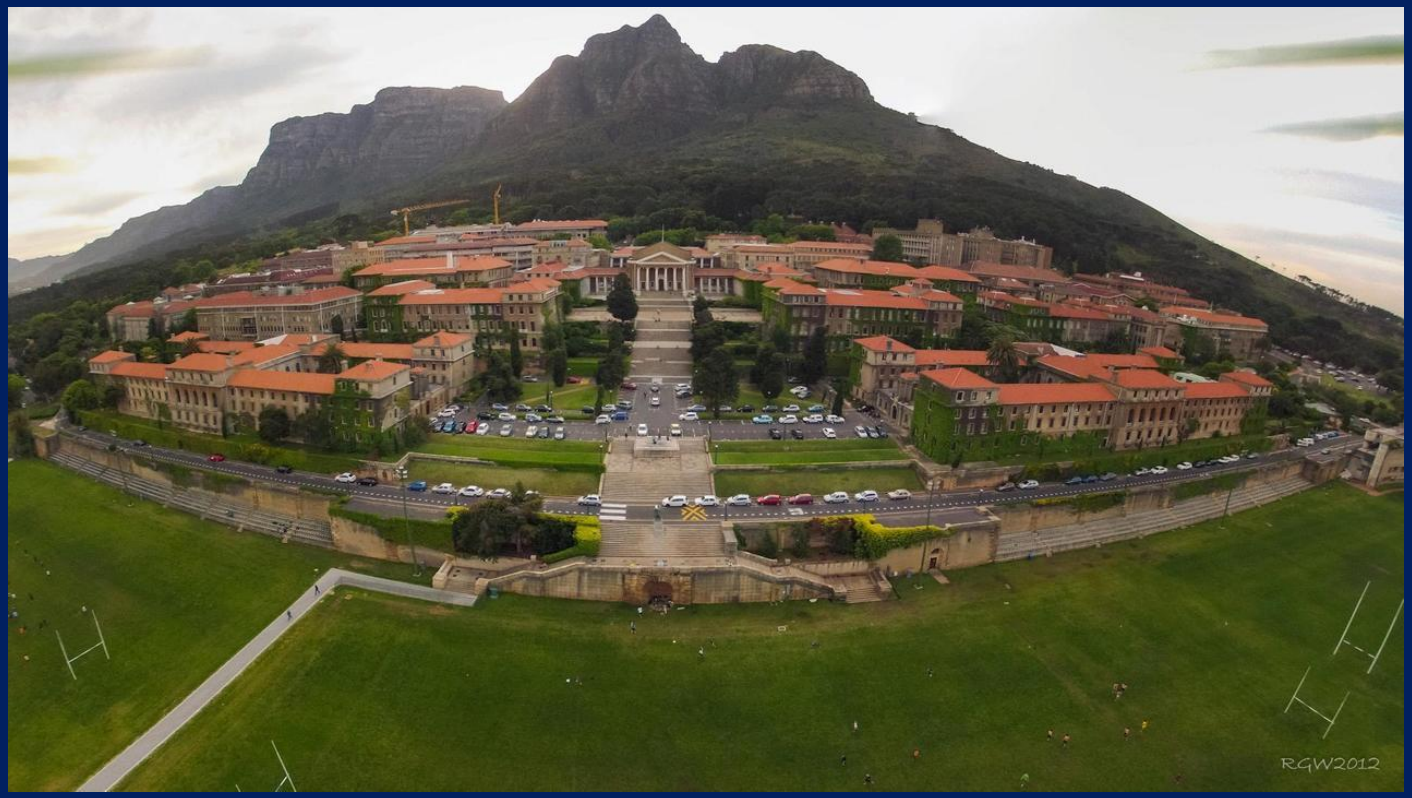
The design of the Student Records Dimensional Data Model and ETL process @ University of Cape Town

Session #7001

July 15, 2015

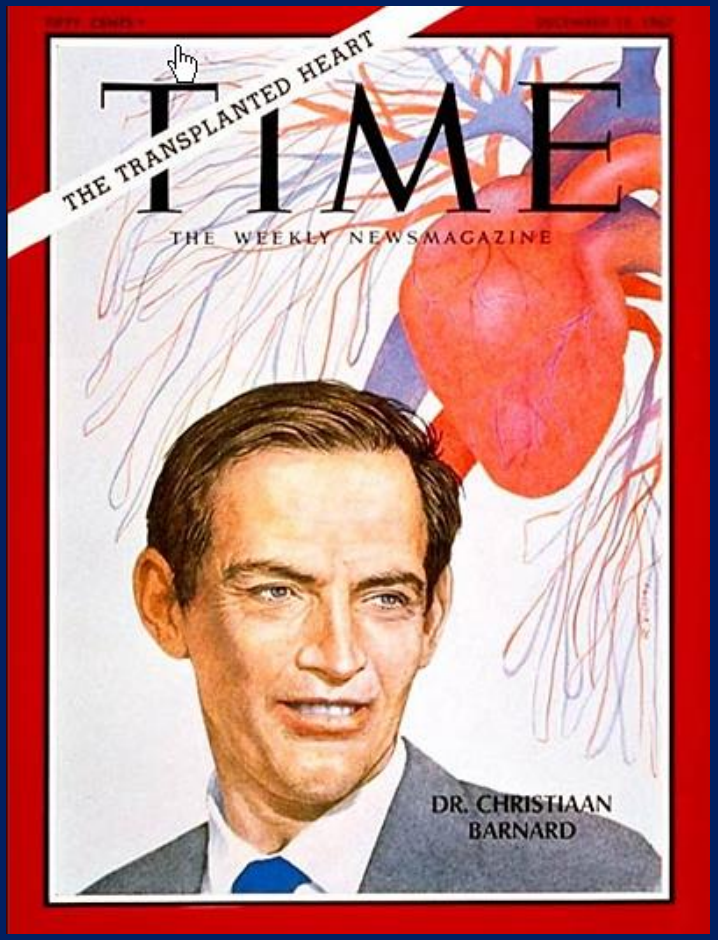


Agenda

- Introduction
- Collecting Requirements
- Dimensional Data Model
- Extract Transform Load (ETL)
- Dashboard
- Questions

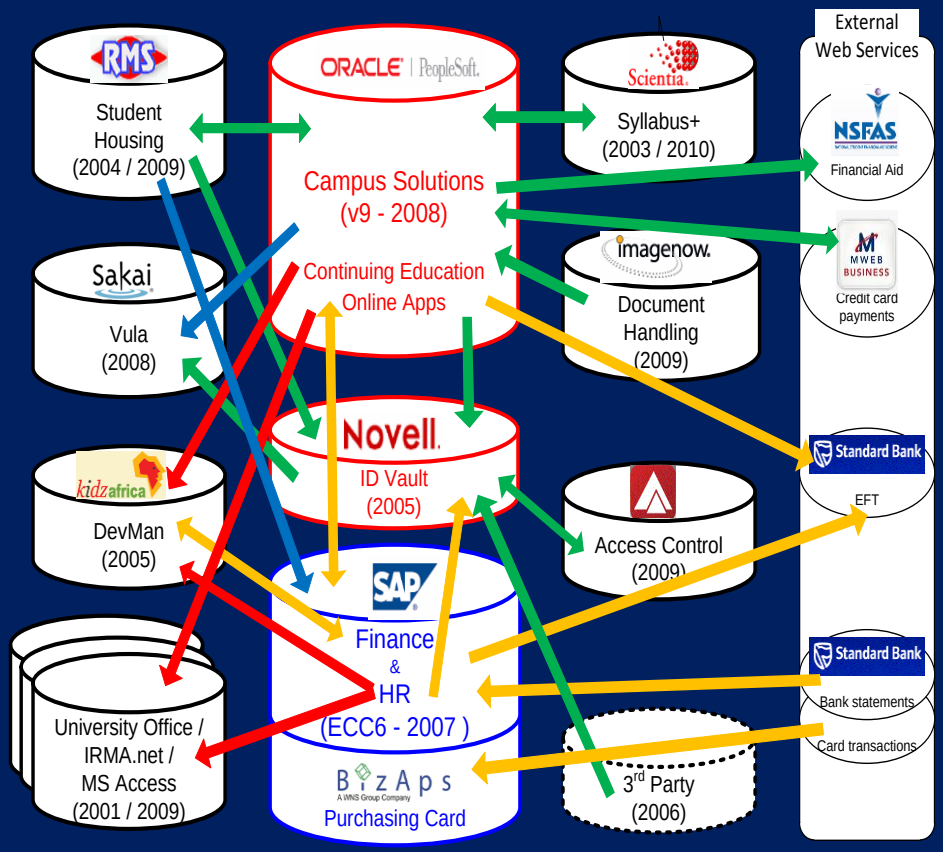
University of Cape Town

- Over 25 000 Students
- Top-rated African University
- Strong Research Focus
- 5 Nobel Prize Winners
- Worlds First Heart Transplant



University of Cape Town ERP Systems

- 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

Faculty

Academic Program

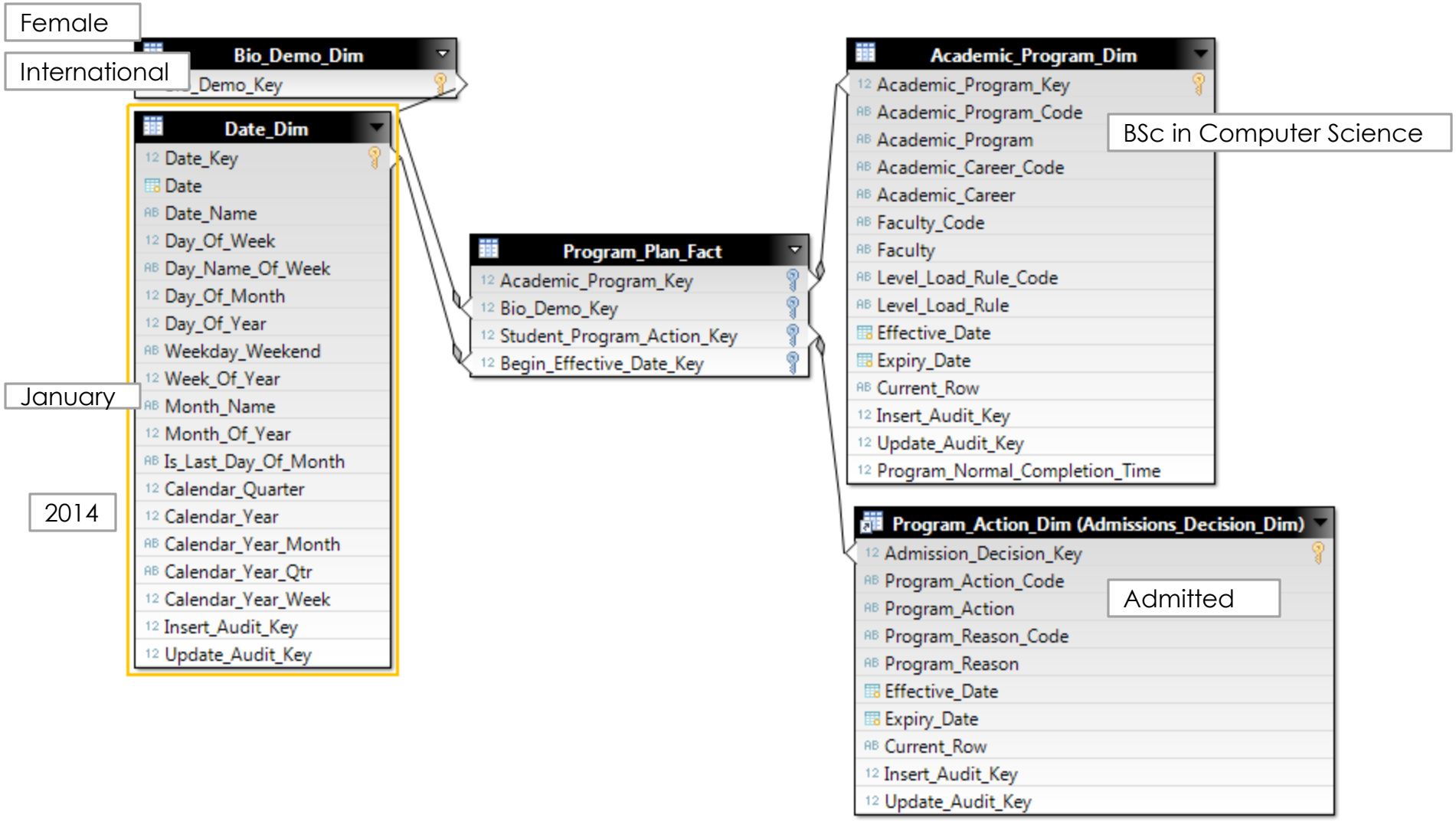
Combine

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



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

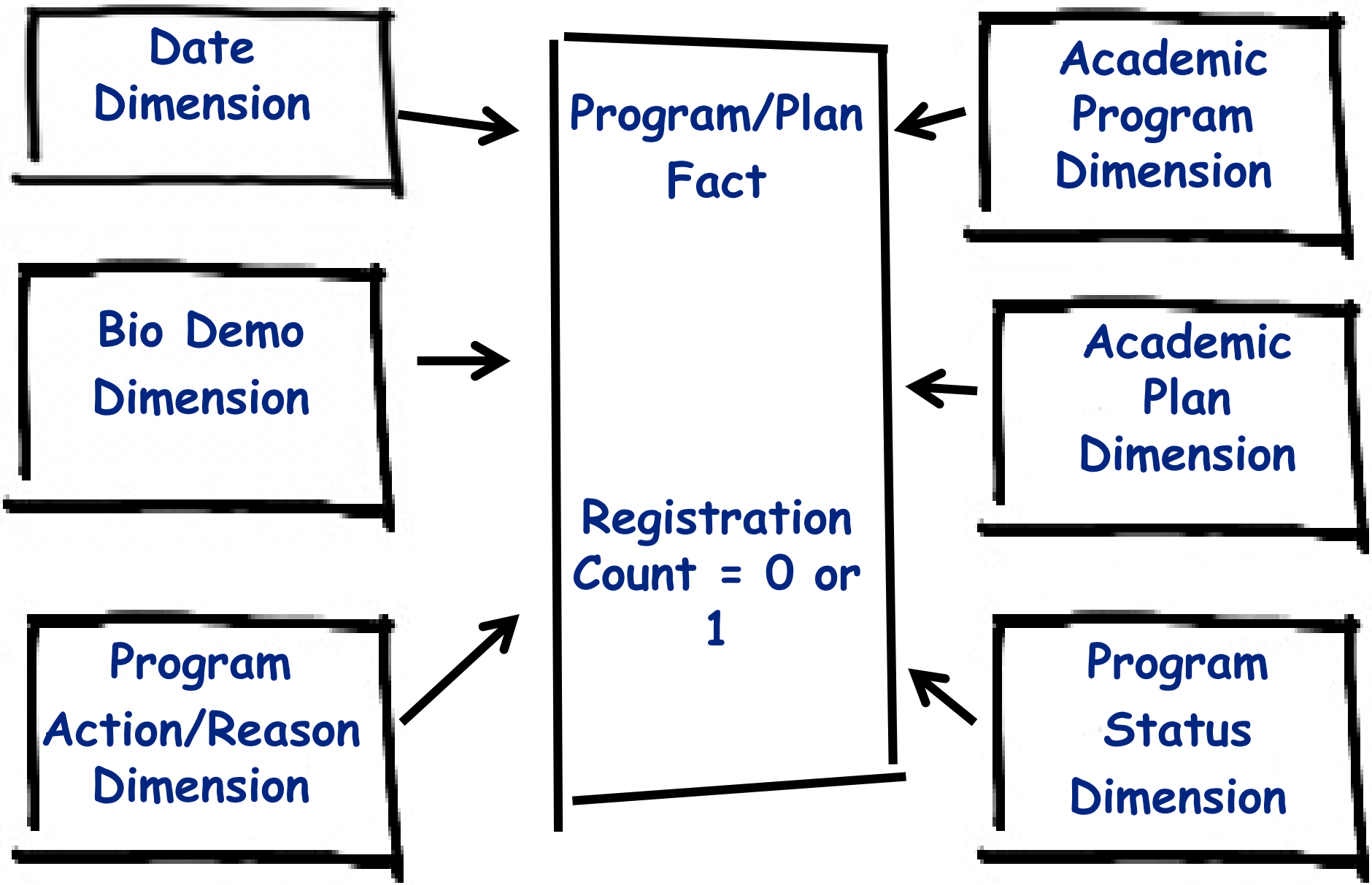
Student accepts offer

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



Program/ Plan Fact Table

SA HEUG

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	

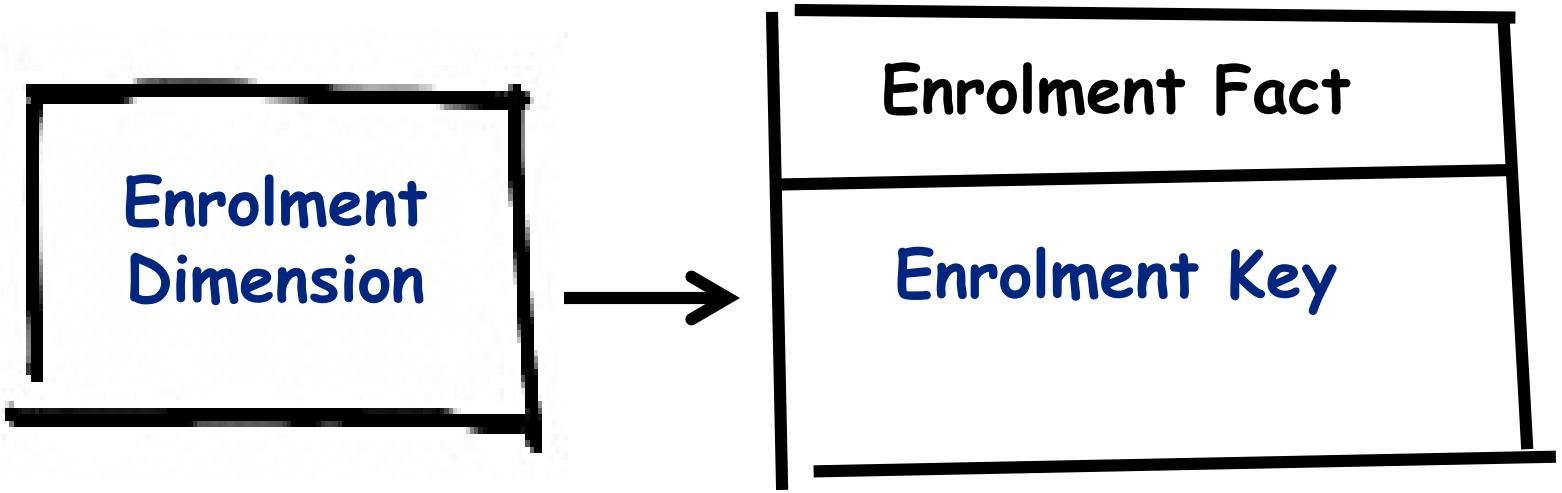
2. Data Change

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

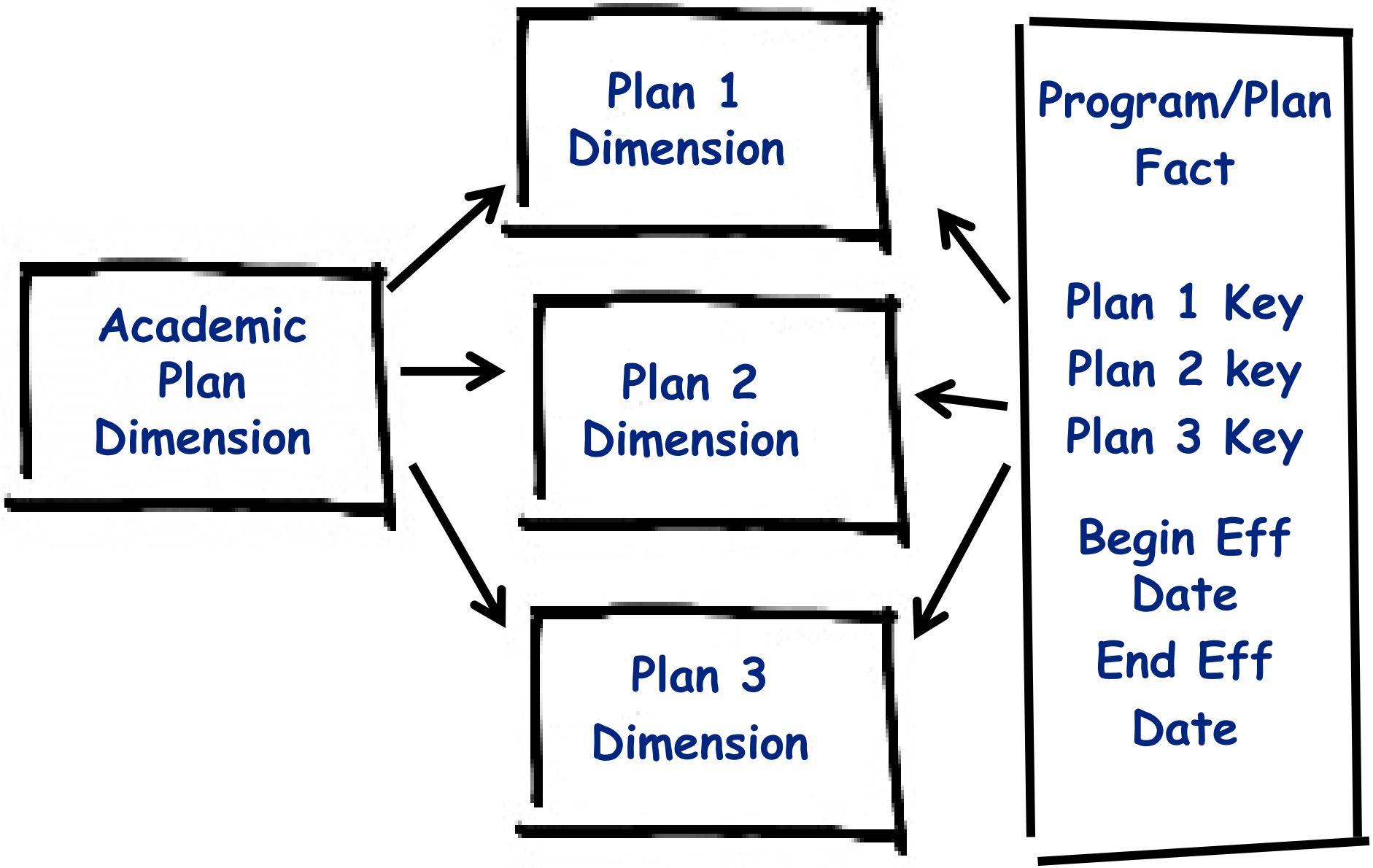
```
SELECT DISTINCT STDNT_ENRL_STATUS
 , ENRL_STATUS_REASON
 , UNITS_ATTEMPTED
 , INCLUDE_IN_GPA
 , EARN_CREDIT
  FROM PS_STDNT_ENRL A
```

41 Rows

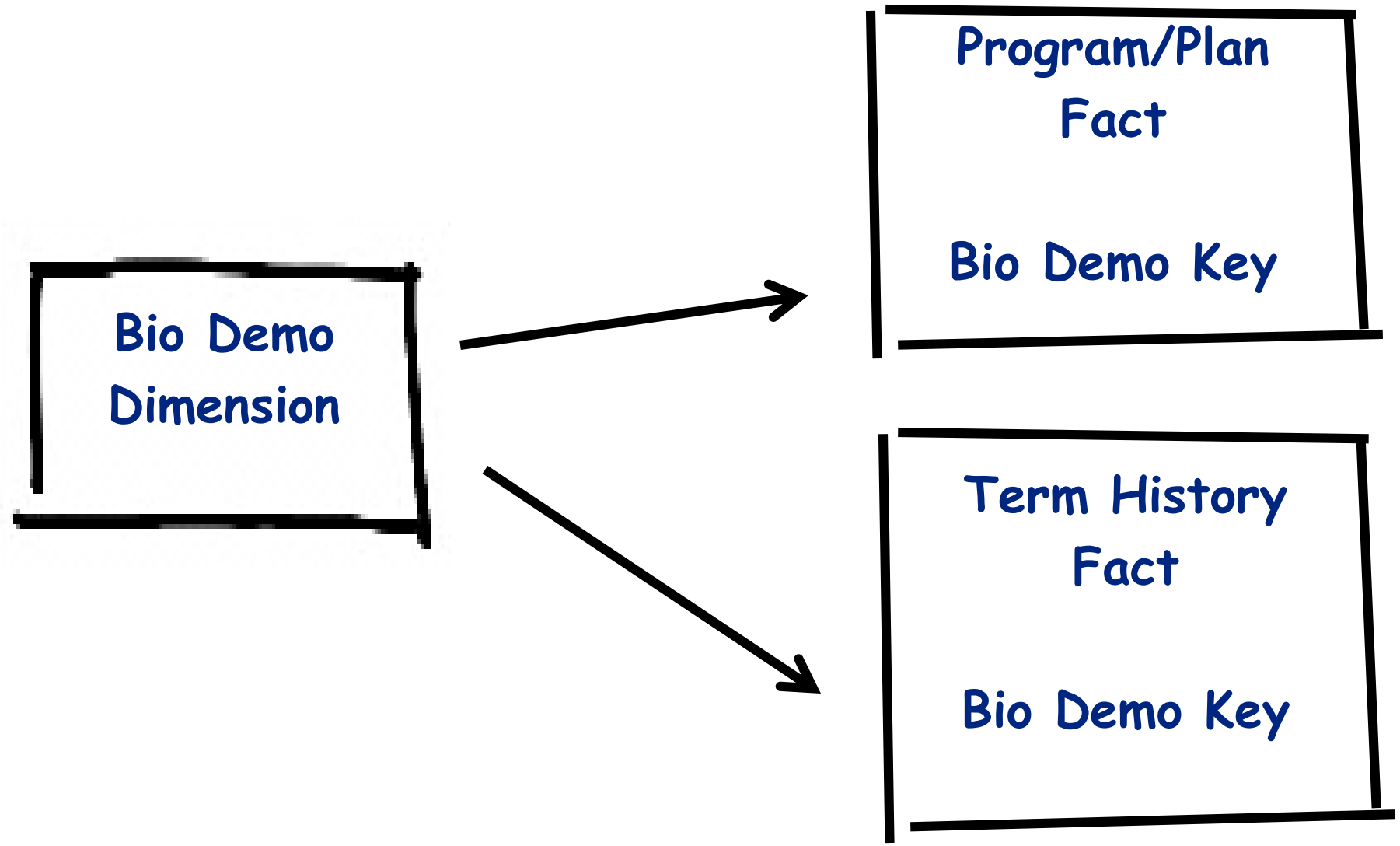
37 seconds to execute

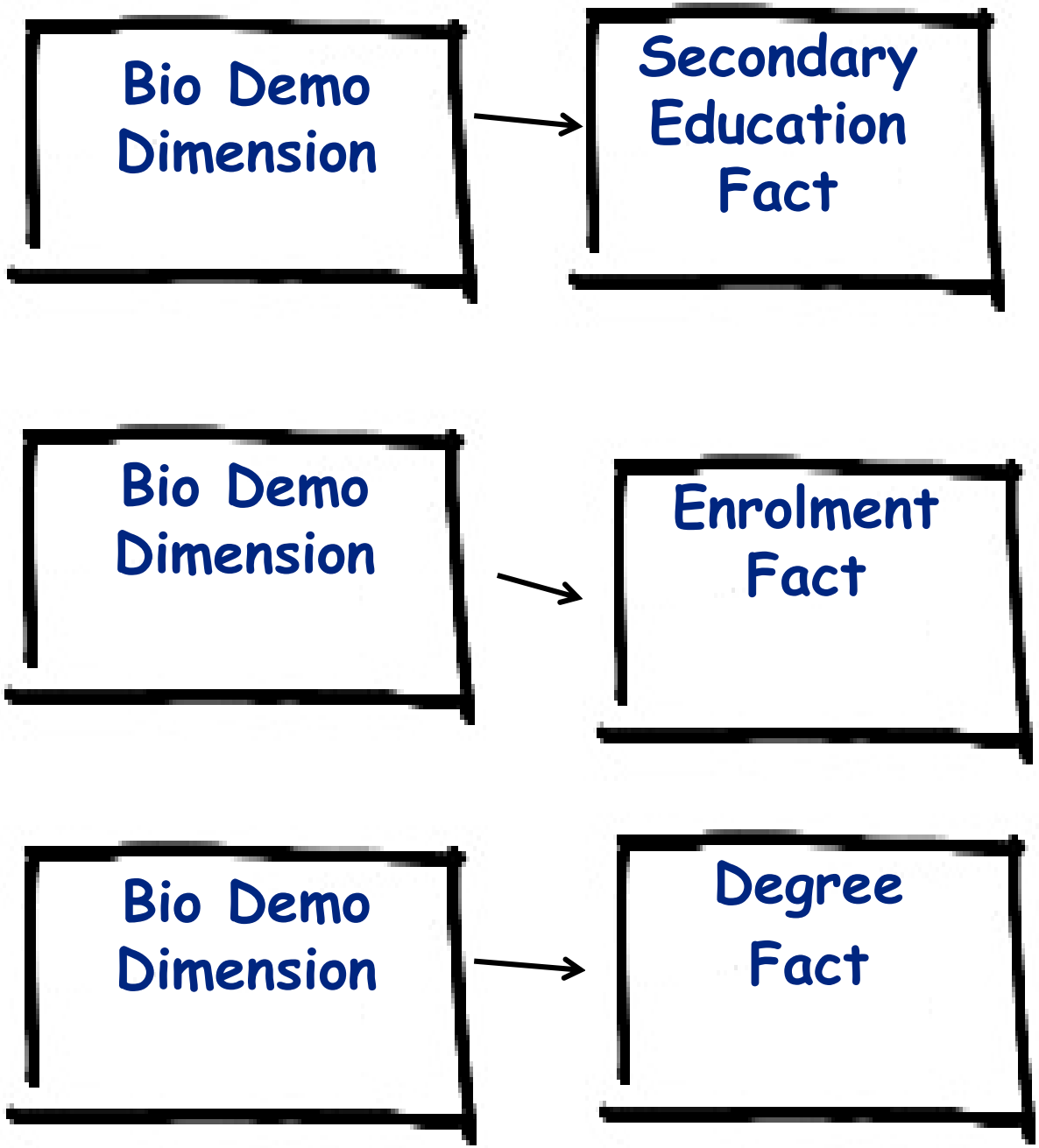
Will get slower as more Enrolment records added

Role-playing Dimensions



Conformed Dimensions




Bio Demo Dimension
Secondary Education Fact
Bio Demo Dimension
Enrolment Fact
Bio Demo Dimension
Degree Fact

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

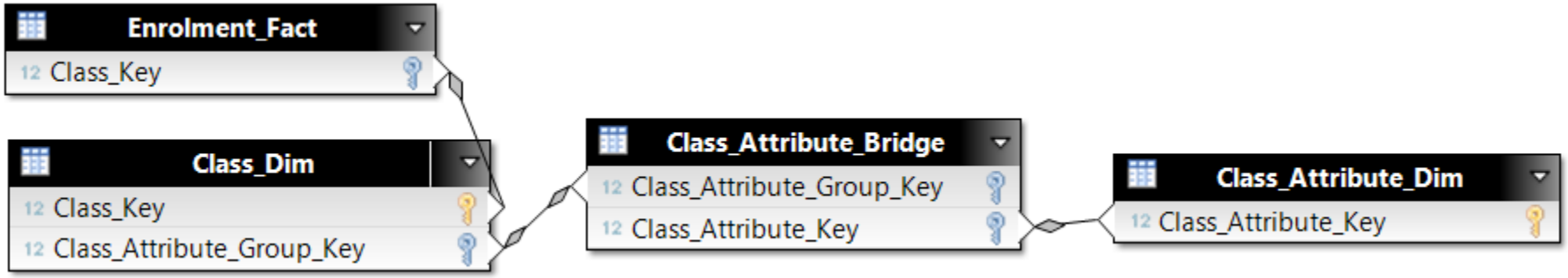
Academic Program Dimension

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



Step 1. Create Class Attribute Dimension

Step 2. Populate the Class Attribute Bridge table

by grouping on

Course Id

Course Offer Number

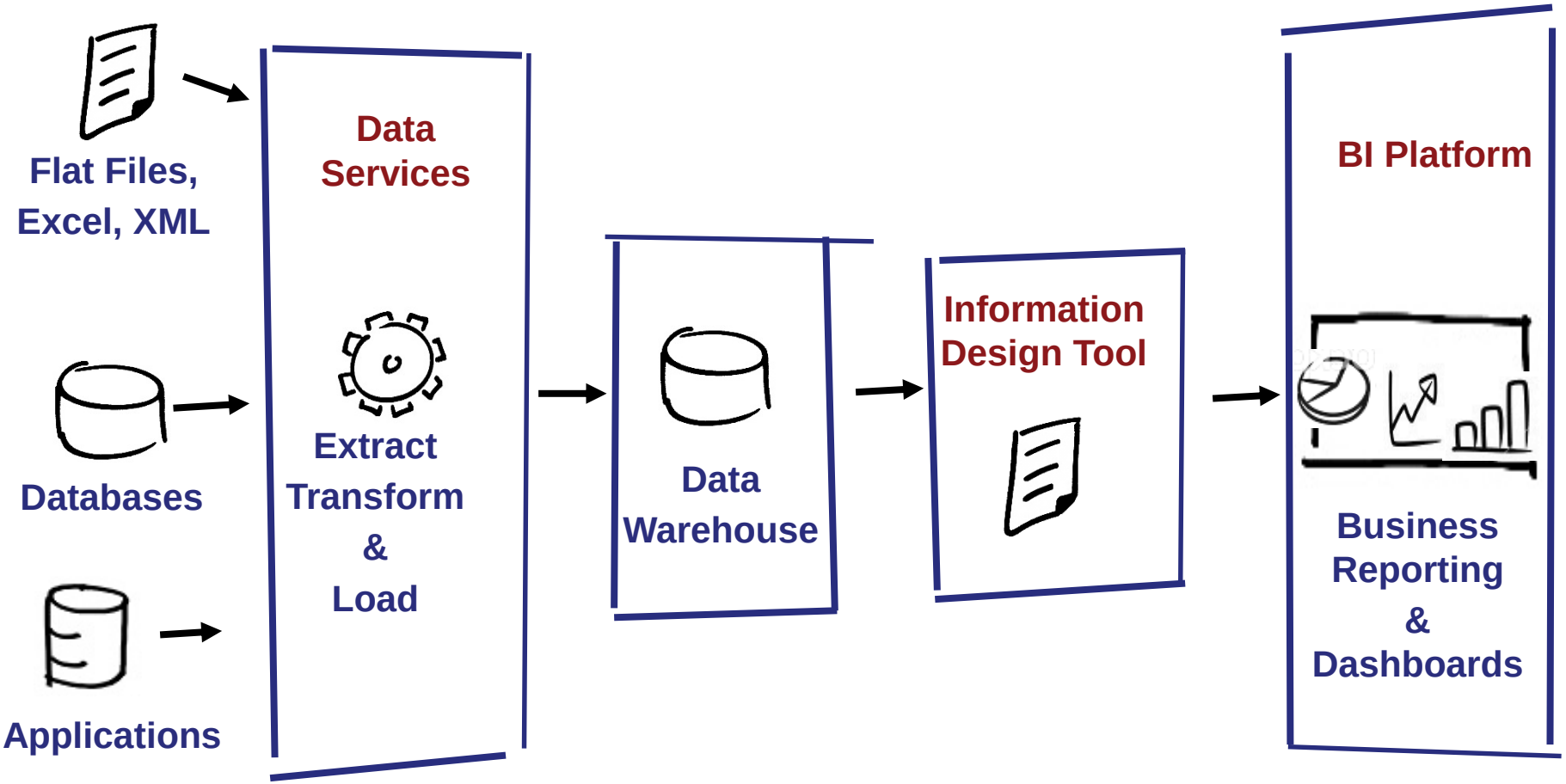
Term Year Code

Session Code

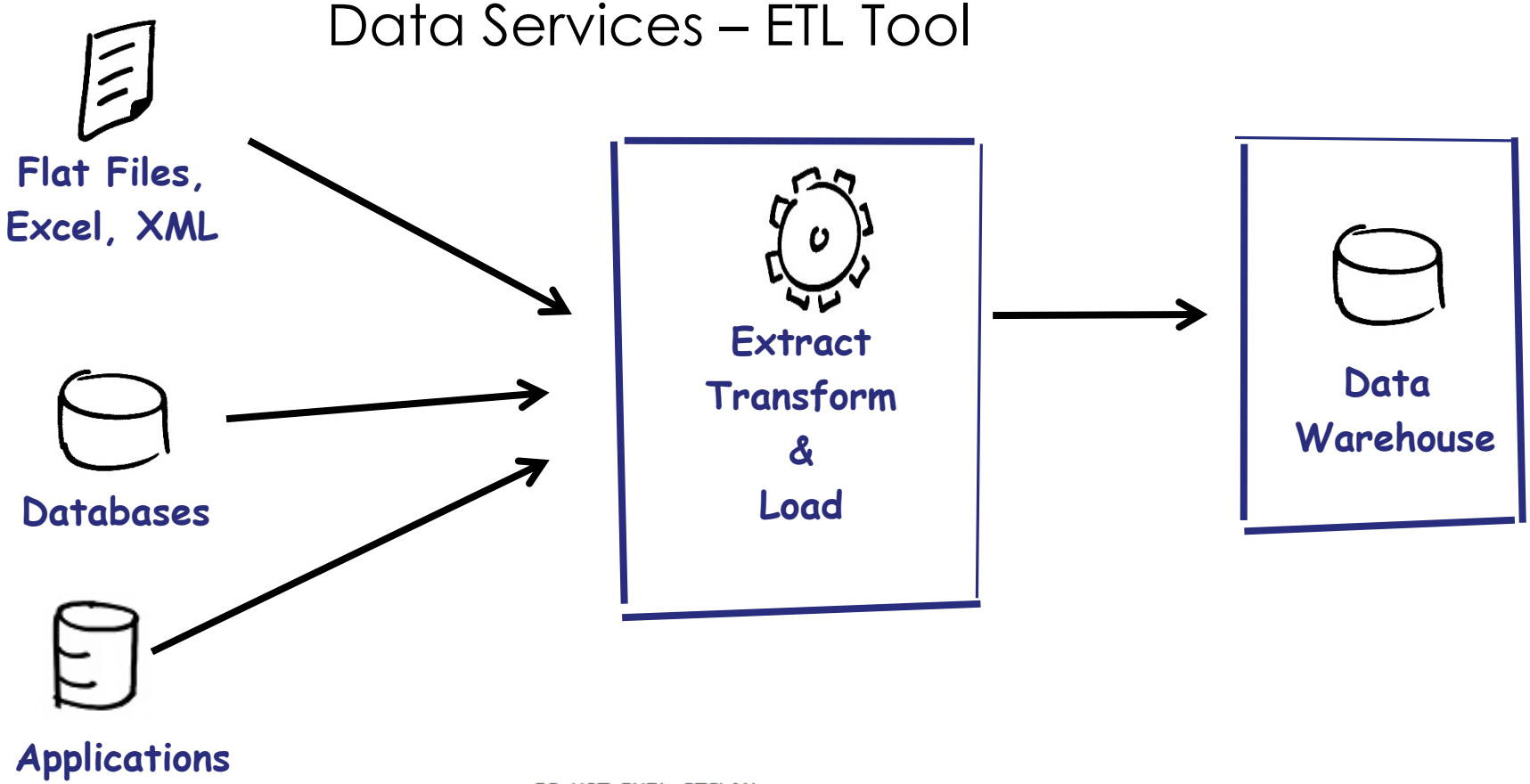
Class Section

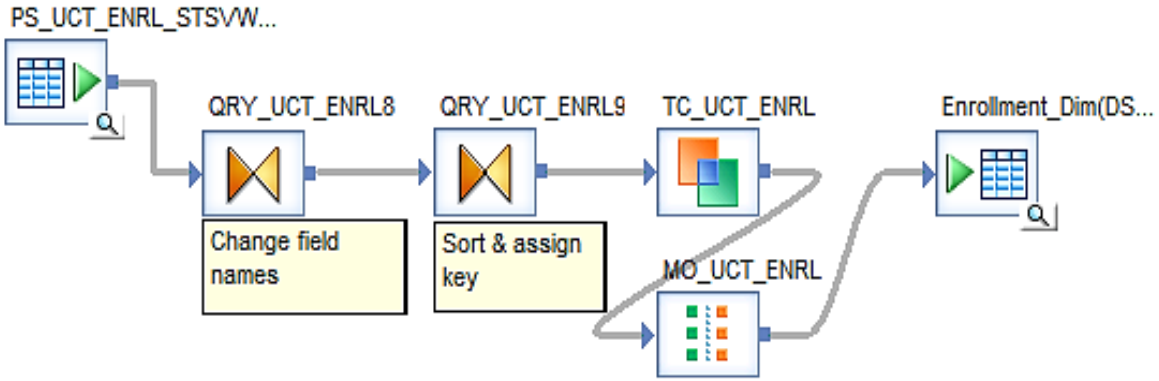
Step 3. Add the Class Attribute Group Key to the Class Dimension

What is Business Objects ?

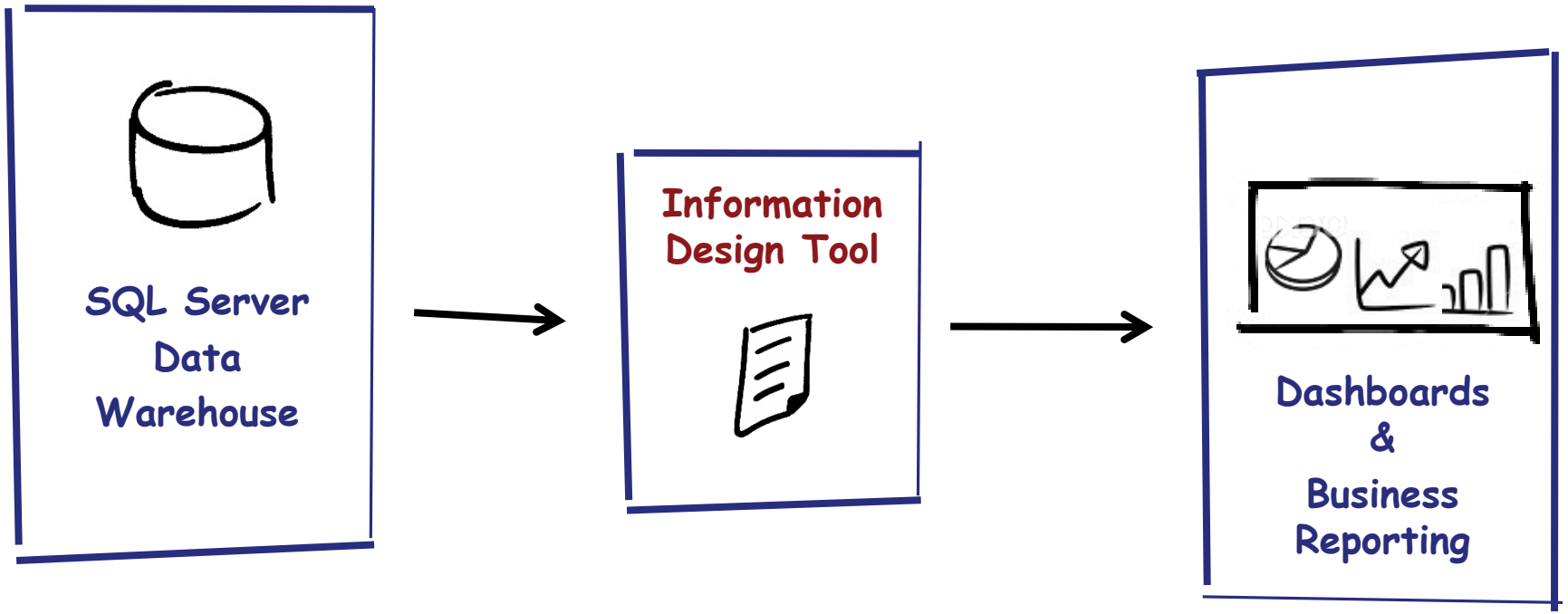


Data Services – ETL Tool

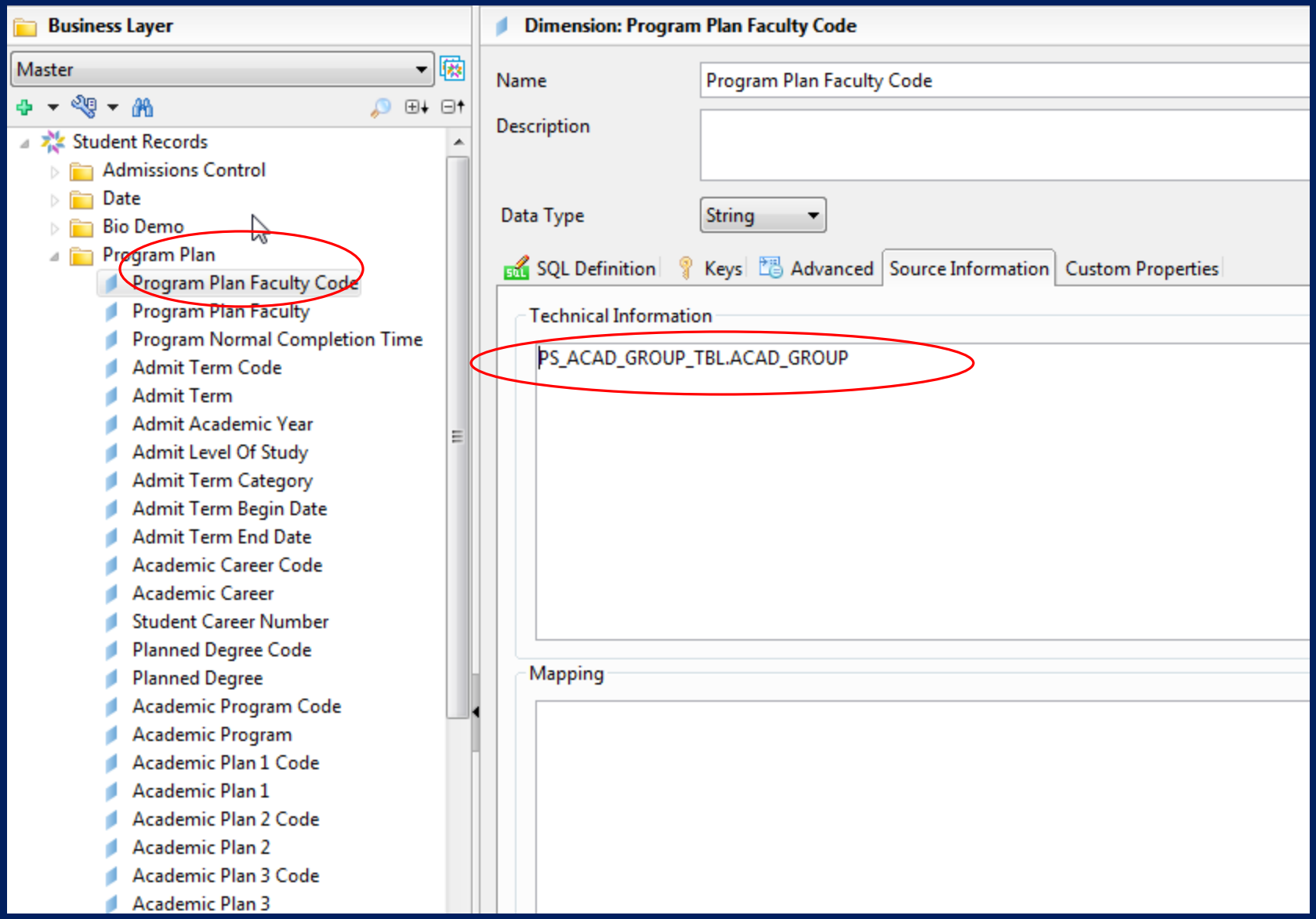




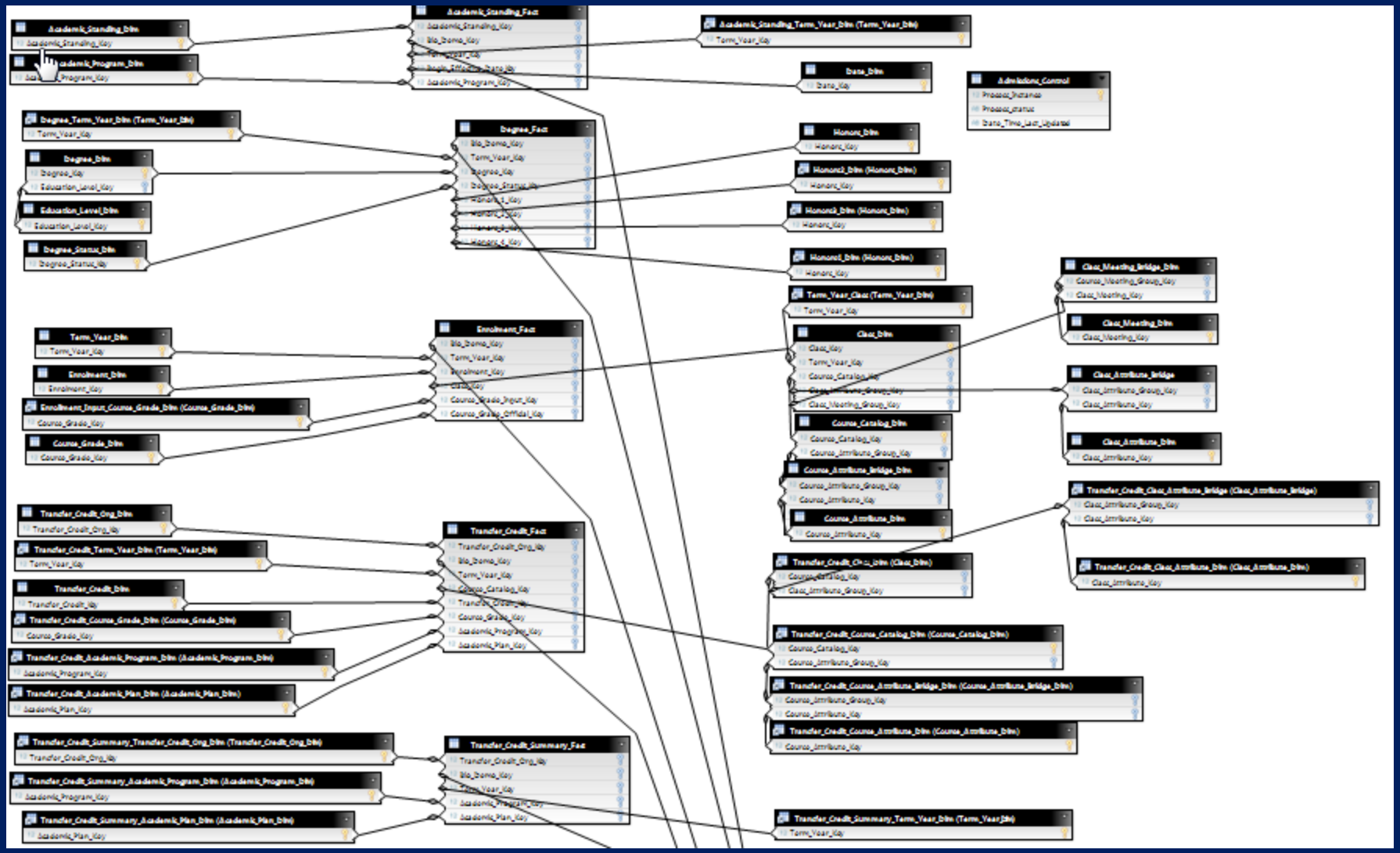
Information Design Tool



Business Layer



Data Foundation – Student Records Data Mart

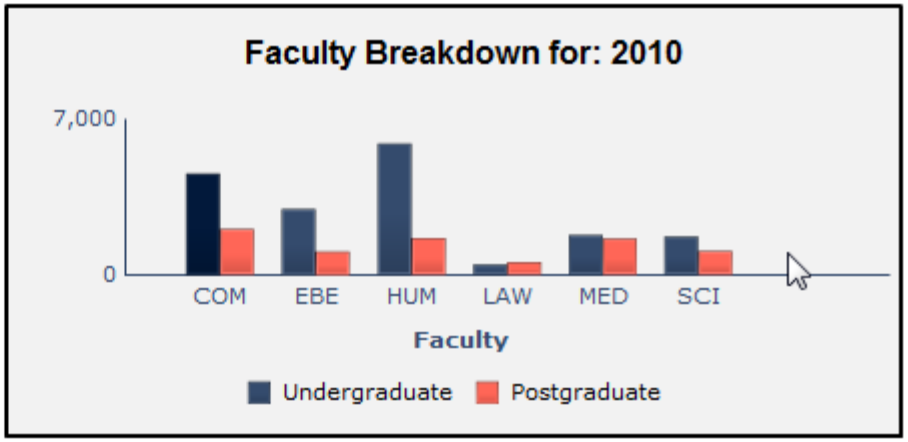


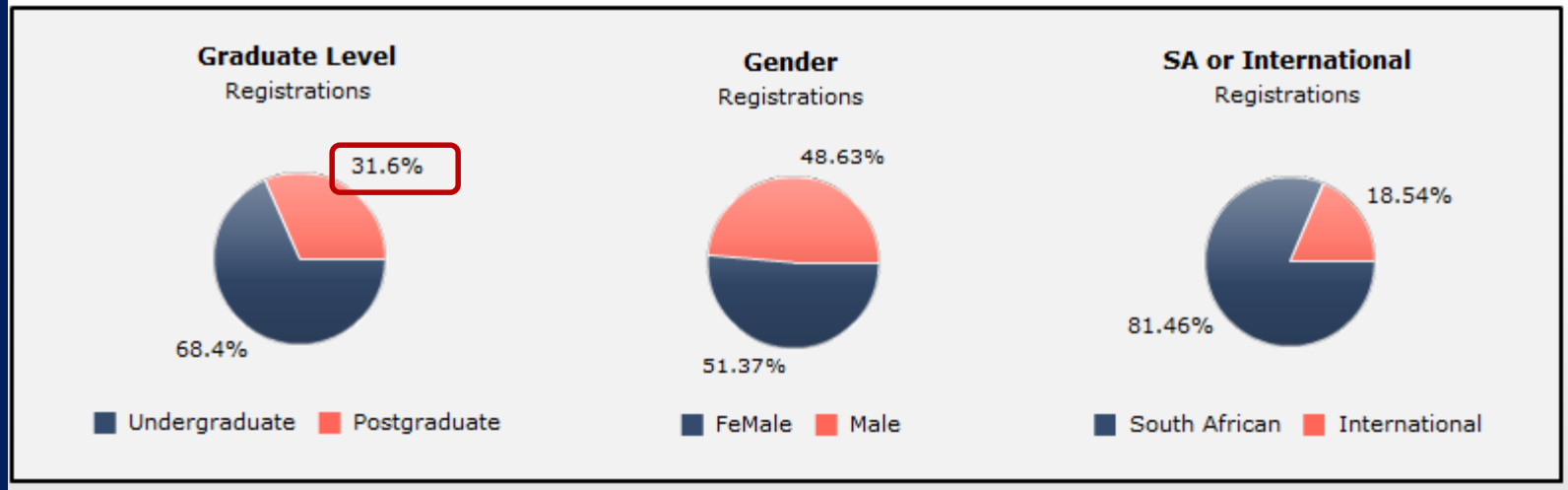
Student Records Dashboard

Student Registration Overview

Undergraduates vs Postgraduates | Population Group | Trend Analysis

Term	Undergraduate	Postgraduate	Total
2009	16 966	7 449	24 415
2010	17 446	8 061	25 507
2011	17 331	8 590	25 921
2012	17 633	8 890	26 523
2013	17 407	9 380	26 787
2014	17 207	9 604	26 811



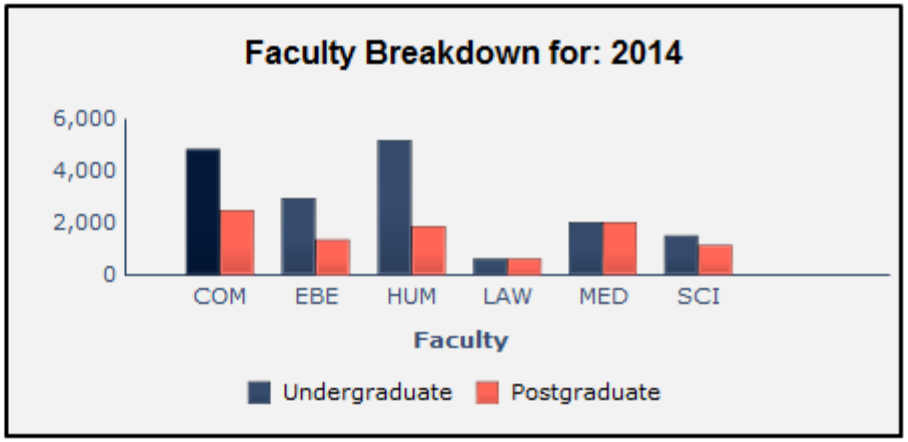


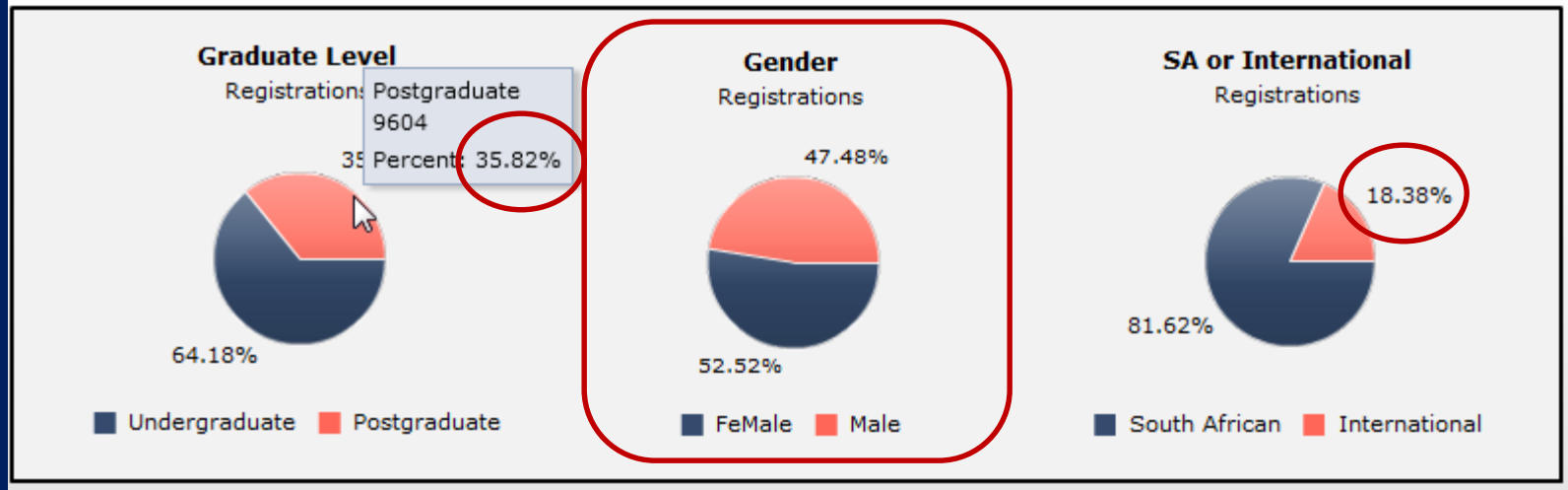
Student Records Dashboard

Student Registration Overview

Undergraduates vs Postgraduates | Population Group | Trend Analysis

Term	Undergraduate	Postgraduate	Total
2009	16 966	7 449	24 415
2010	17 446	8 061	25 507
2011	17 331	8 590	25 921
2012	17 633	8 890	26 523
2013	17 407	9 380	26 787
2014	17 207	9 604	26 811



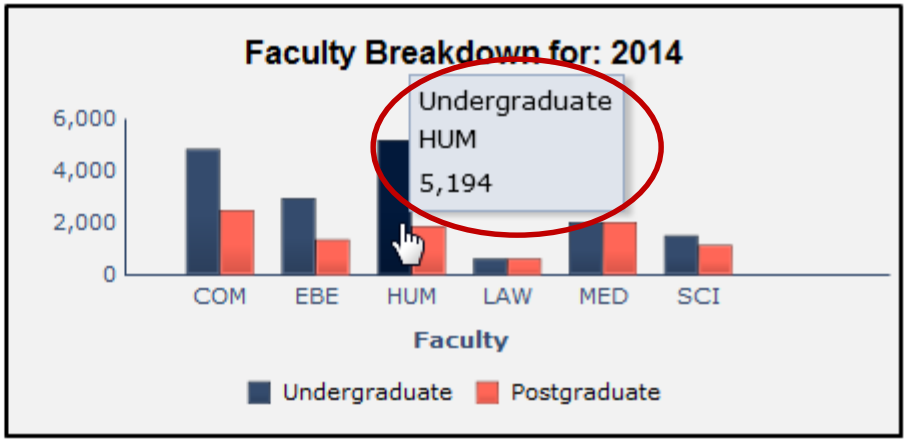


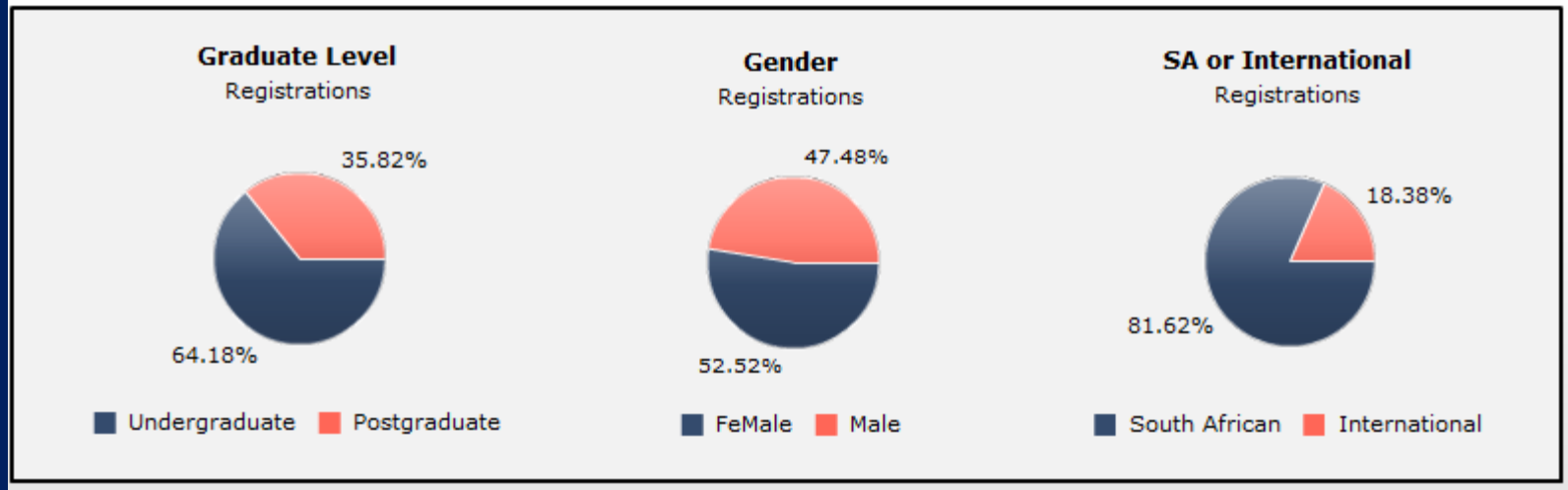
Student Records Dashboard

Student Registration Overview

Undergraduates vs Postgraduates | Population Group | Trend Analysis

Term	Undergraduate	Postgraduate	Total
2009	16 966	7 449	24 415
2010	17 446	8 061	25 507
2011	17 331	8 590	25 921
2012	17 633	8 890	26 523
2013	17 407	9 380	26 787
2014	17 207	9 604	26 811



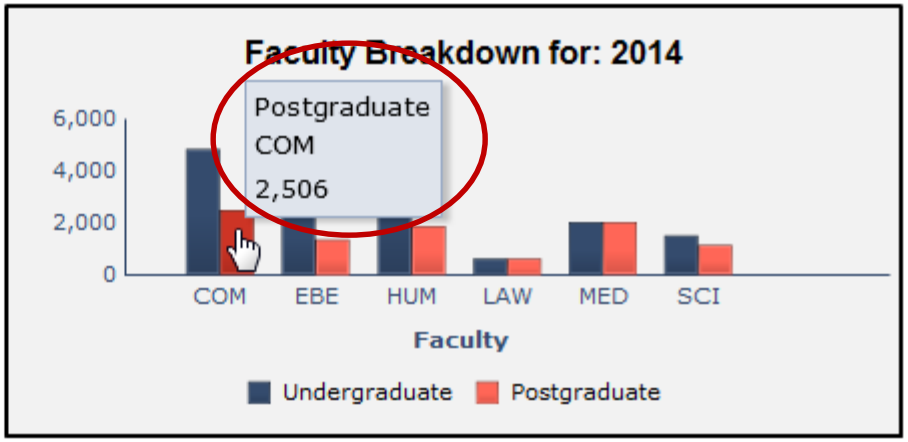


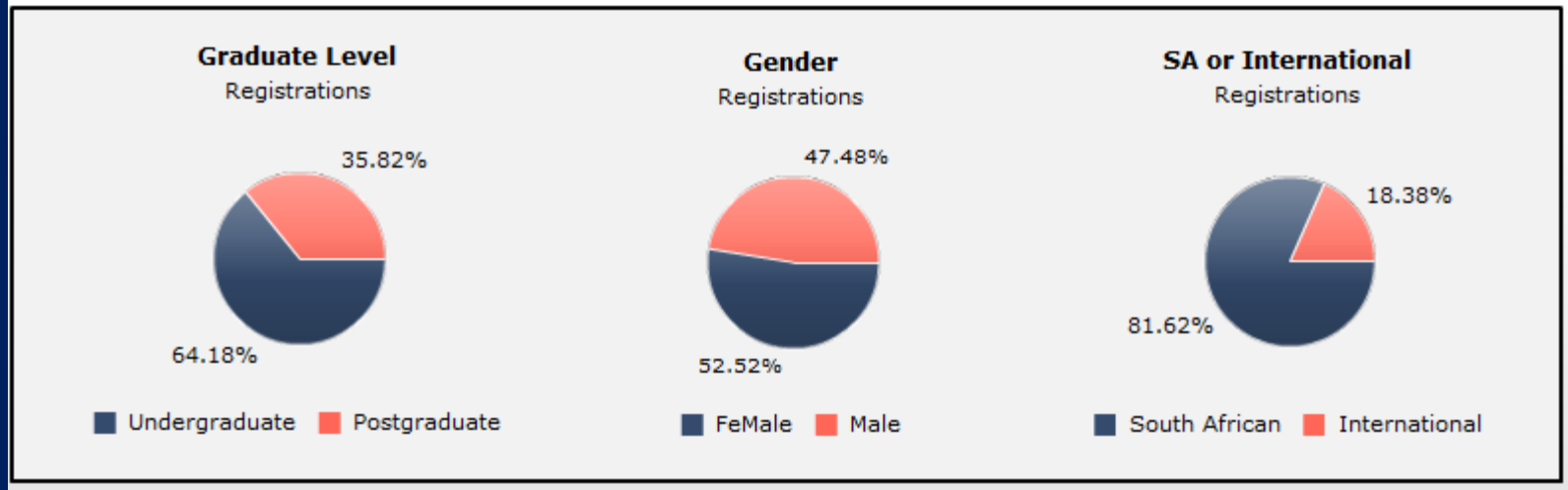
Student Records Dashboard

Student Registration Overview

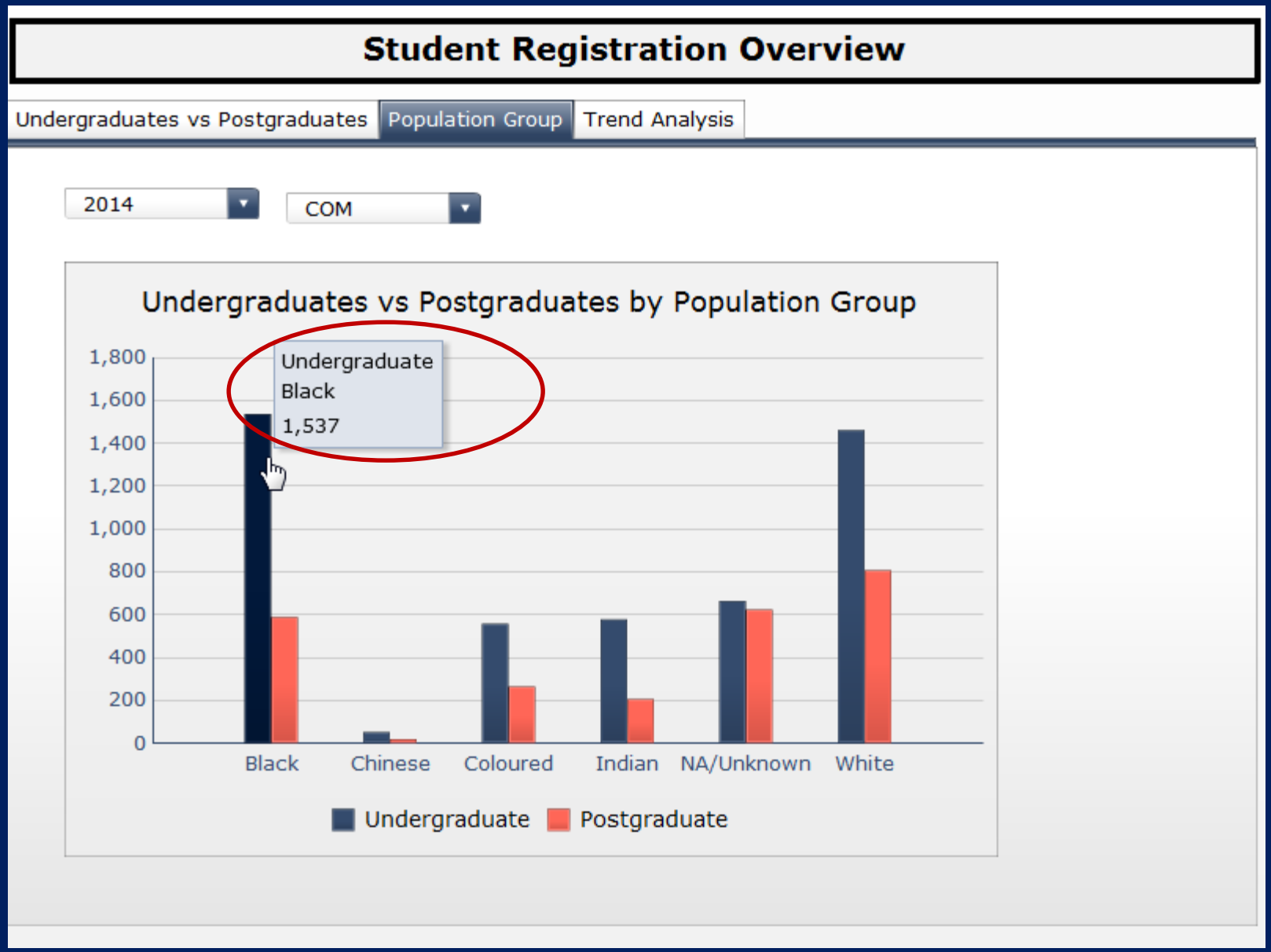
Undergraduates vs Postgraduates | Population Group | Trend Analysis

Term	Undergraduate	Postgraduate	Total
2009	16 966	7 449	24 415
2010	17 446	8 061	25 507
2011	17 331	8 590	25 921
2012	17 633	8 890	26 523
2013	17 407	9 380	26 787
2014	17 207	9 604	26 811





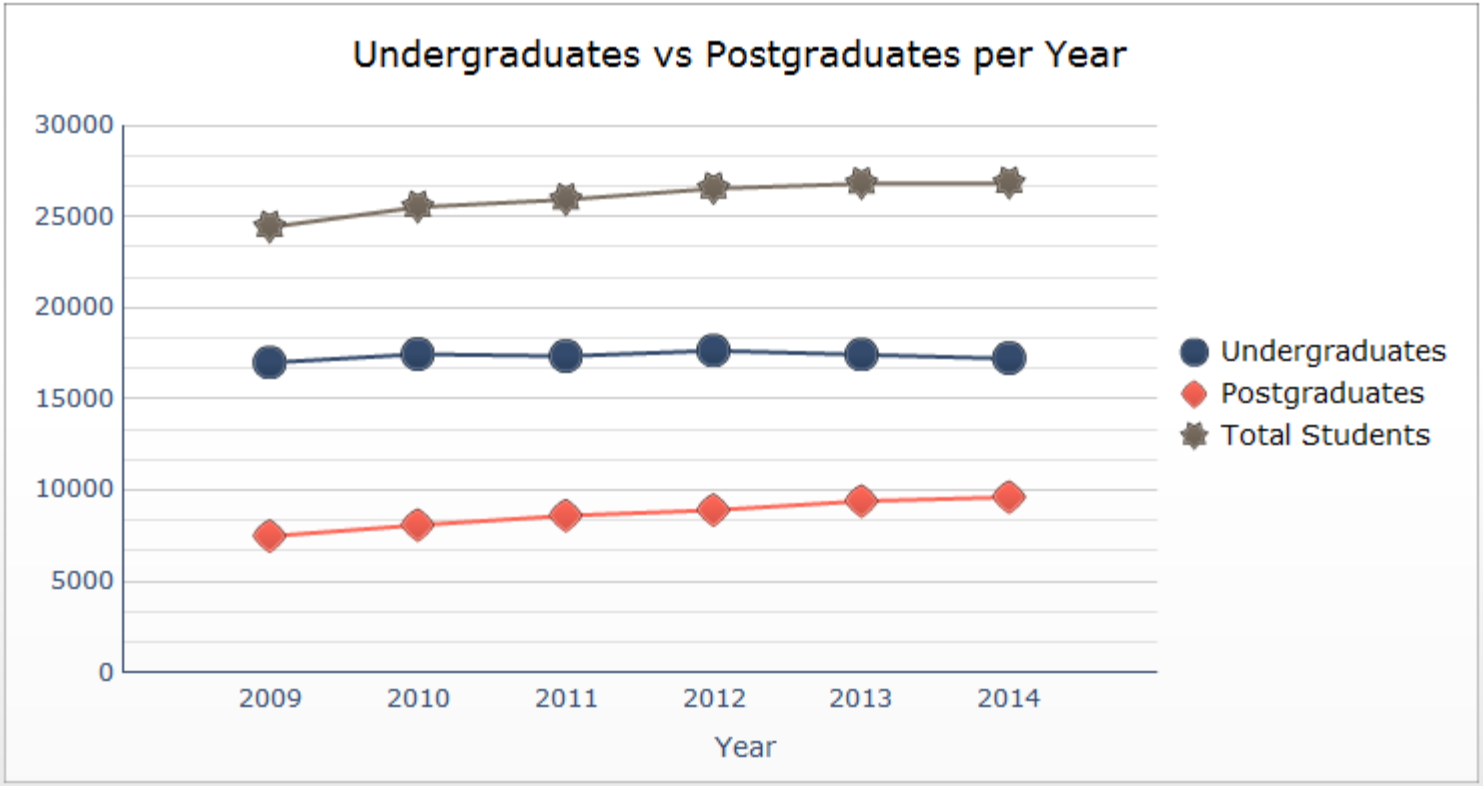
Student Records Dashboard



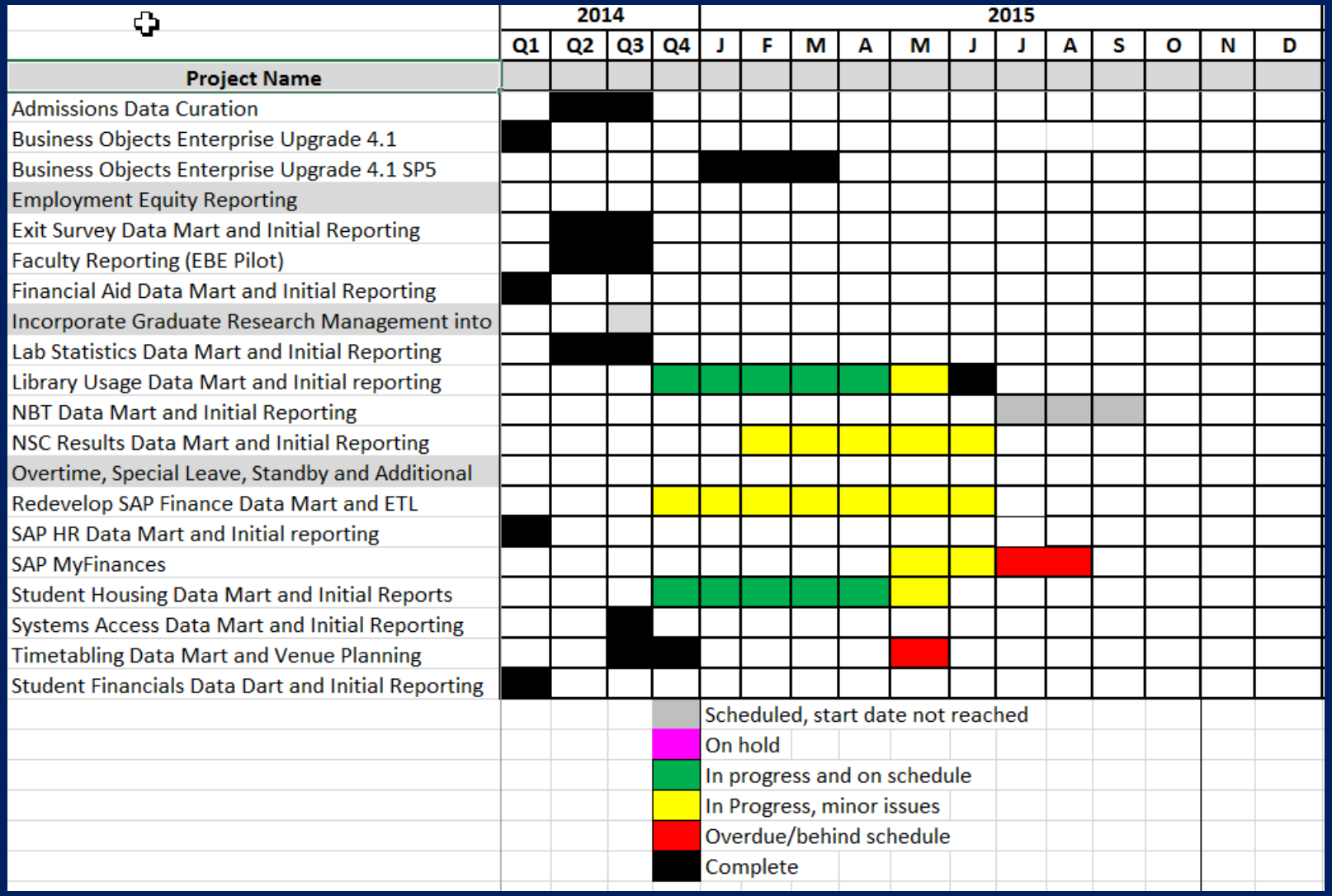
Student Records Dashboard

Student Registration Overview

Undergraduates vs Postgraduates | Population Group | Trend Analysis



Where to from here ?



Questions?

Contact

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