



**Office of the President
Institutional Research and
Academic Planning (IRAP)**

UCDW Degree User Guide

THIS PAGE HAS BEEN INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

1. INTRODUCTION.....	7
1.1 Overview of Degree	7
1.2 Student Privacy	7
1.3 How Degree Data in UCDW Is Stored?.....	7
1.4 Important points to know when using UCDW-Degree	9
2. Accessing UCDW-Degree	9
3. UCDW-Degree Tables Guide	10
3.1 UCDW Degree Star	10
3.2 Details on Fact and Dimension Tables in UCDW-Degree.....	11
3.2.1 DEGREE_AWARDED_F.....	11
3.2.2 STUDENT_D.....	11
3.2.3 CAMPUS_LOCATION_D.....	12
3.2.4 STUDENT_LEVEL_D	12
3.2.5 CAMPUS_COLLEGE_MAJOR_CODE_D	12
3.2.6 AGE_BAND_D.....	12
3.2.7 ACADEMIC_TERM_D.....	13
3.2.8 ACADEMIC_DEGREE_D	13
3.2.9 ATP_SCHOOL_D	13
3.3 DataMart in UCDW-Degree	14
3.4 MQTS in UCDW Student Enrollment.....	14
3.4.1 DEGREE_HEAD_COUNT_M	14
4. Sample Queries	16
5. Terms and Abbreviations	16

TABLE OF TABLES

Table 1: Glossary 16

TABLE OF FIGURES

Figure 1: Three-Tier UCDW architecture 8
Figure 2: Current UCDW Infrastructure 9
Figure 3: Fact and dimension tables in UCDW Degree Star..... 10
Figure 4: Degree Fact Row Counts 15
Figure 5: Degree Head Count in MQT 16

THIS PAGE HAS BEEN INTENTIONALLY LEFT BLANK

1. INTRODUCTION

1.1 OVERVIEW OF DEGREE

UCDW-Degree data warehouse stores degrees conferred data. UC campuses supply data for all degree granted to UC students. For students who receive multiple degrees, all degree records will be stored. UCDW-Degree provides system-wide reporting data for degrees conferred across all 10 UC campuses. The data in UCDW-Degree is dependent on the submission of degree data by UC campuses' twice a year – Fall, Winter and Spring degrees file for the academic year is submitted around September 15th of the academic year & Summer degrees file is submitted around December 15th of the academic year.

1.2 STUDENT PRIVACY

Personally identifiable information for education records includes direct identifiers, such as a student's name or identification number, indirect identifiers, such as a student's date of birth, or other information which can be used to distinguish or trace an individual's identity either directly or indirectly through linkages with other student information.

Please remember that student data is covered under UC's privacy policy and FERPA. UCDW users will be required to adhere to all Federal, State and local statutes, regulations, and other requirements pertaining to the security, confidentiality, and privacy of data provided by UCOP including, but not limited to, the Family Educational Rights and Privacy Act (FERPA), the California Information Practices Act, and the Privacy Act of 1974, as amended.

Proper security rules will be enforced and access will be restricted to only qualified users. Permissions will be appropriately assigned based on user-access groups.

1.3 HOW DEGREE DATA IN UCDW IS STORED?

All 10 UC campuses submit degree data to UCOP according to the file specifications and the file submission deadline.

File specifications and file submission deadlines for degree data are available in this [link](#).

UCOP UCDW infrastructure stores information in three layers - stage, base and business intelligence (BI). All 10 campuses' degree related data are first stored in staging layer which is a parking lot for all the campus degree data. Staging data is transformed if necessary and stored in base layer. Staging and base data are validated and certified by data stewards in IRAP and then stored in BI layer. Canned reports for business use are generated from BI layer.

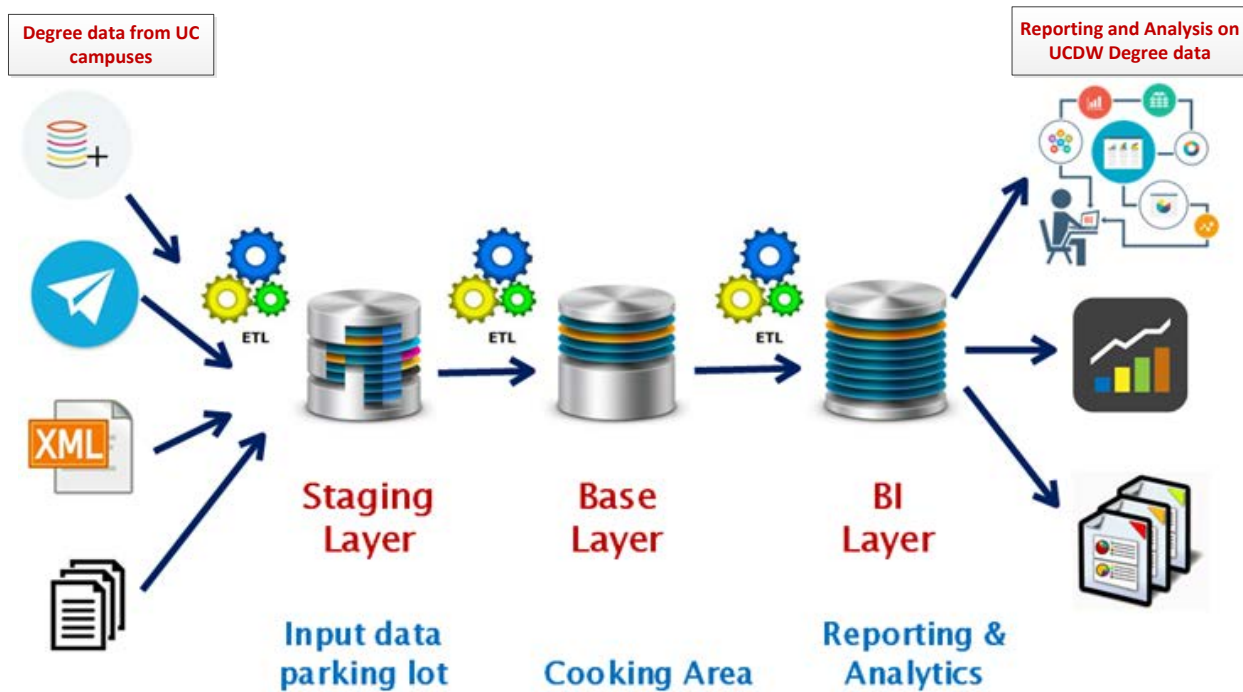


Figure 1: Three-Tier UCDW architecture

For analyses, analytical and reporting purposes, business users use data in BI layer. BI layer data is stored as star schema, one of the styles of dimensional data modelling. Star schema provides important benefits to the user: improved understandability of the data, performance and enhanced reporting capability. Star schema consists of one or more fact tables referencing any number of dimension tables.

When thinking about data and the way it is analyzed, the fact tables provides measurements and the dimension tables provide context. For example, in a data report the dimension tables supply the values for the row and column headings as well as the constraint filters and grouping capabilities. The fact table supplies the measures, or what is being counted.

Refer to sections 3.1 and 3.2 for details on the fact and dimension tables in UCDW-Degree.

Degree Data Mart is a subset of the degree data in UCDW star-schema. Usually data marts are oriented to a specific business need. Data marts are easily accessible by business users, because they don't require queries that are too complicated or resource-consuming.

Refer to section 3.3 for details on the data mart in UCDW-Degree.

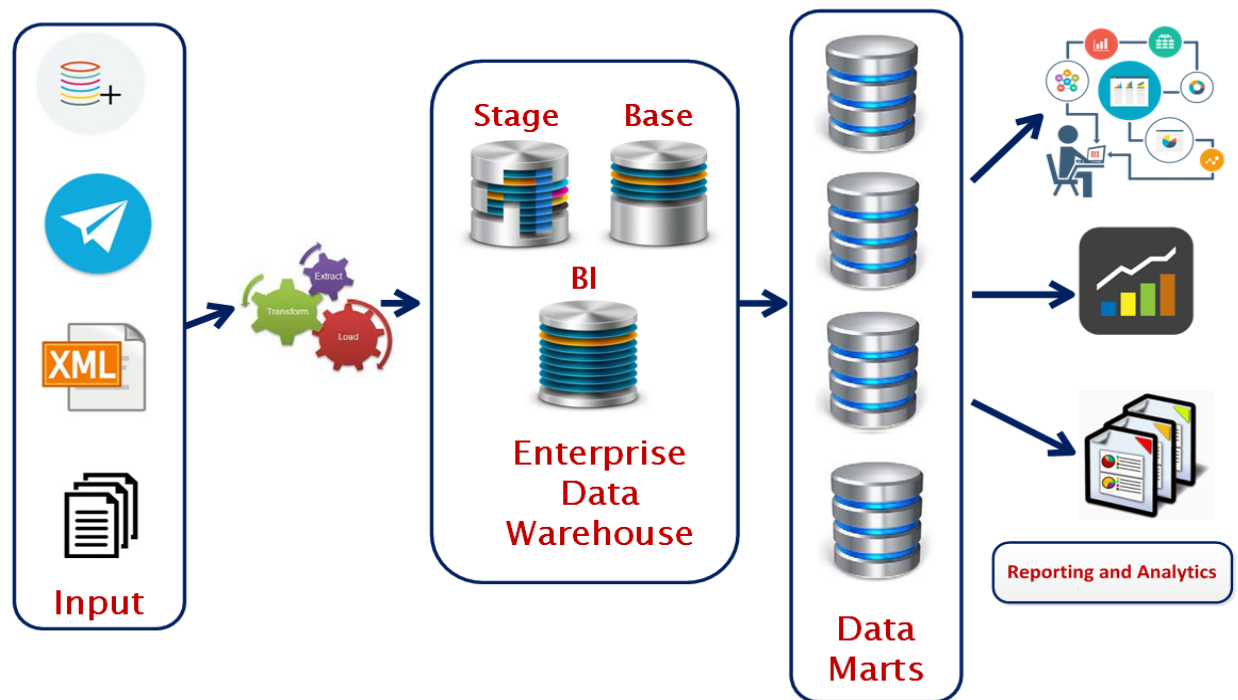


Figure 2: Current UCDW Infrastructure

1.4 IMPORTANT POINTS TO KNOW WHEN USING UCDW-DEGREE

- i. Academic year (fall, winter and spring) degrees and summer degrees are available in Degree star.
- ii. Retroactive degrees supplied by the campuses are also stored in UCDW-Degree star.
- iii. The Degree star and DataMart stores *degree completions*, **not** *degree completers*.
- iv. Refer IRAP Code Library (ICL-25) in JIRA on how to obtain degree completers information from UCDW- Degree.
- v. Do's and Don'ts on querying UCDW is available via this [link](#).

2. ACCESSING UCDW-DEGREE

The following are the different ways to access UCDW-Degree data:

- i. UCDW-Degree data can be accessed using the SQL case tool, Db Visualizer. Download the free version of the software and instructions on how to connect to UCDW through the tool are available in this [link](#).

- ii. Canned and validation reports generated from UCDW-Degree are available in Cognos. Instructions on how to connect to Cognos and access reports, are available under **'UCDW and Cognos'** section on 'How-To's' in 'Support & Training' tab of Data Operations Hub. It can be accessed via this [link](#).
- iii. UCDW-Degree data can be accessed via Tableau. Instructions on how to connect to UCDW through Tableau are available under **'UCDW and Tableau'** section on 'How-To's' in 'Support & Training' tab of Data Operations Hub. It can be accessed via this [link](#).
- iv. UCDW-Degree data can also be accessed using other data analysis and mining tools such as SAS.

3. UCDW-DEGREE TABLES GUIDE

3.1 UCDW DEGREE STAR

The Degree star has Degree Awarded fact table and many dimension tables. These tables are stored in STUD_BI schema of uc_dss database in the production server.

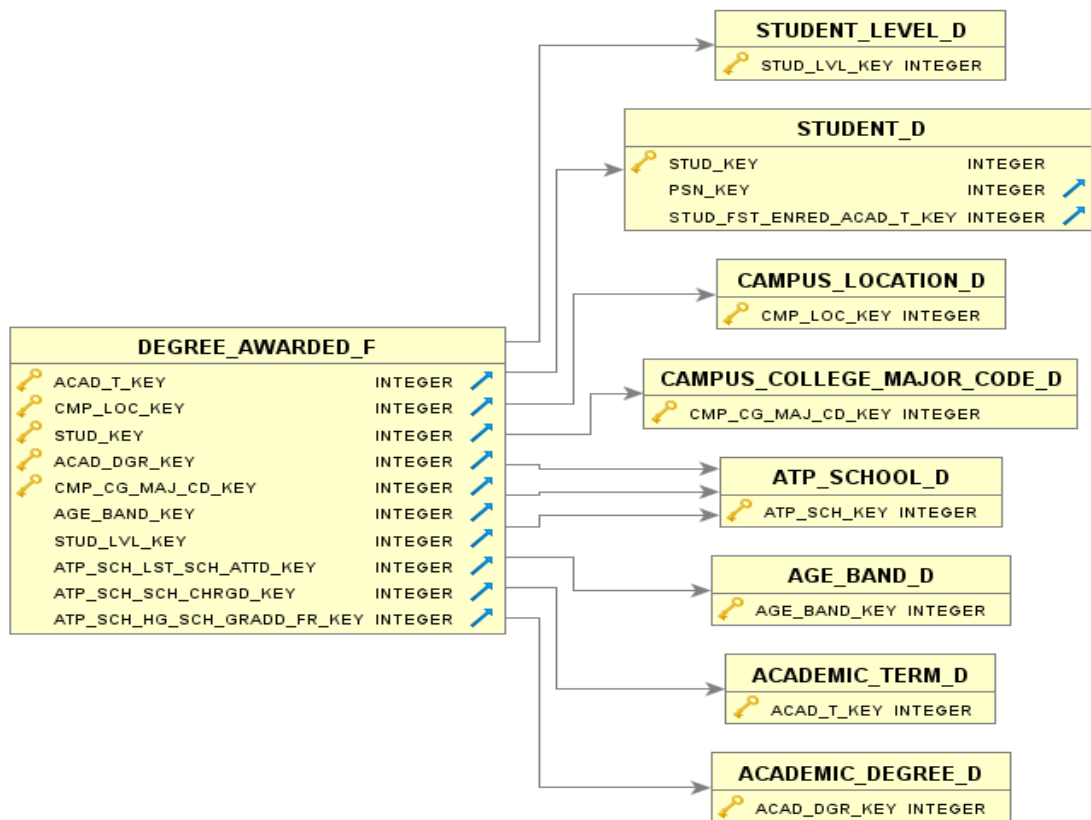


Figure 3: Fact and dimension tables in UCDW Degree Star

Note the fact tables are suffixed with _F and dimension tables are suffixed with _D.

3.2 DETAILS ON FACT AND DIMENSION TABLES IN UCDW-DEGREE

3.2.1 DEGREE_AWARDED_F

Description: Degree Awarded fact table contains the required metrics and other information related to academic degrees conferred by UC campuses.

Grain (lowest level at which data is captured): One row per student per academic term, degree and major combination.

Example 1 – In Fall 2012, UC Berkeley has awarded Jane Smith 3 degrees as follows:

1. Degree code **08-M.A.** and a single major **699-POLITICAL SCIENCE**
2. Degree code **32-C.PHIL.** and a single major **428-HEALTH SERV & POLICY ANA**
3. Degree code **43-PH.D.** and a single major **428-HEALTH SERV & POLICY ANA**

She will have 3 rows in the fact table.

Example 2 – In Fall 2012, UC Berkeley has awarded John Doe 3 degrees as follows:

1. Degree code **02-A.B.** and double majors **780-PSYCHOLOGY** and **123-BIOLOGY**
2. Degree code **05-B.S.** and a single major **120-BIOLOGICAL SCIENCES**
3. Degree code **02-A.B.** and a single major **867-SOCIOLOGY**

He will have 4 rows in the fact table.

Primary Key: Primary key of the fact table is a composite key made up of foreign keys to its dimension tables. Combination of ACAD_T_KEY, CMP_LOC_KEY, STUD_KEY, ACAD_DGR_KEY and CMP_CG_MAJ_CD_KEY is primary key to DEGREE_AWARDED_F table.

3.2.2 STUDENT_D

Description: Student dimension contains a record for each student that has enrolled or has been awarded a degree.

Student dimension is a conformed dimension that is shared within Student subject area. Also, this dimension is Slowly Changing Dimension (SCD) type 2 i.e., it contains complete history of changes in different attributes stored for a student.

Primary Key: Primary key of the dimension table is a system generated key to uniquely identify a row in a dimension table. STUD_KEY is the primary key for STUDENT_D table.

Natural Key: This is a type of unique key which is formed of attributes within real data. Combination of Student location campus code (STUD_LOC_CMP_CD) and Student id (STUD_ID) is the natural key in STUDENT_D table.

3.2.3 CAMPUS_LOCATION_D

Description: Campus Location dimension identifies the 10 UC campuses.

Campus Location dimension is a conformed dimension that is shared across multiple subject areas. Also, this dimension is SCD type 2 i.e., it contains complete history of changes in different attributes stored for a campus.

Primary Key: Primary key of the dimension table is a system generated key to uniquely identify a row in a dimension table. CMP_LOC_KEY is the primary key for CAMPUS_LOCATION_D table.

Natural Key: N/A

3.2.4 STUDENT_LEVEL_D

Description: Student level dimension contains various categorizations of academic levels such as Graduate, Undergraduate, and Credential and Certificate.

Student level dimension is a conformed dimension that is shared within Student subject area. Also, this dimension is SCD type 1 i.e., the new information overwrites the original information. No history is kept.

Primary Key: Primary key of the dimension table is a system generated key to uniquely identify a row in a dimension table. STUD_LVL_KEY is the primary key for STUDENT_LEVEL_D table.

Natural Key: N/A

3.2.5 CAMPUS_COLLEGE_MAJOR_CODE_D

Description: This dimension contains information about majors offered by various colleges in each campus within the UC system.

Campus College Major dimension is a conformed dimension that is shared across multiple subject areas. Also, this dimension is SCD type 2 i.e., it contains complete history of changes in different attributes stored in this dimension.

Primary Key: Primary key of the dimension table is a system generated key to uniquely identify a row in a dimension table. CMP_CG_MAJ_CD_KEY is the primary key for CAMPUS_COLLEGE_MAJOR_CODE_D table.

Natural Key: N/A

3.2.6 AGE_BAND_D

Description: This dimension contains the age ranges required for banding students by age.

Age Band dimension is a conformed dimension that is shared across multiple subject areas. Also, this dimension is SCD type 1 i.e., the new information overwrites the original information. No history is kept.

Primary Key: Primary key of the dimension table is a system generated key to uniquely identify a row in a dimension table. AGE_BAND_KEY is the primary key for AGE_BAND_D table.

Natural Key: N/A

3.2.7 ACADEMIC_TERM_D

Description: Academic Term dimension is a rollup of Academic Sub Term master dimension. It identifies academic terms in each calendar and academic year. The grain is one record per term (fall or winter or spring or summer)

Academic Term dimension is a conformed dimension that is shared across multiple subject areas. Also, this dimension is SCD type 1 i.e., the new information overwrites the original information. No history is kept.

Primary Key: Primary key of the dimension table is a system generated key to uniquely identify a row in a dimension table. ACAD_T_KEY is the primary key for ACADEMIC_TERM_D table.

Natural Key: N/A

3.2.8 ACADEMIC_DEGREE_D

Description: Academic degree dimension contains all types of degrees conferred by UC campuses.

Academic Degree dimension is a conformed dimension that is shared across multiple subject areas. Also, this dimension is SCD Type 2 i.e., it contains complete history of changes in different attributes stored in this dimension.

Primary Key: Primary key of the dimension table is a system generated key to uniquely identify a row in a dimension table. ACAD_DGR_KEY is the primary key for ACADEMIC_DEGREE_D table.

Natural Key: N/A

3.2.9 ATP_SCHOOL_D

Description: ATP School dimension identifies US, non-US high schools and other academic institutions. It is an exact replica of the master ATP school master conforming dimension.

ATP School dimension is a conformed dimension that is shared across multiple subject areas. Also, this dimension is SCD type 2 i.e., it contains complete history of changes in different attributes stored in this dimension.

Primary Key: Primary key of the dimension table is a system generated key to uniquely identify a row in a dimension table. ATP_SCH_KEY is the primary key for ATP_SCHOOL_D table.

Natural Key: N/A

3.3 DATAMART IN UCDW-DEGREE

Degree Data Mart is a subset of data generated from Degree Star Schema. DEGREE_DM is in IRAP_BI schema in uc_dss database in dwp2 server.

Grain (lowest level at which data is captured): One record per student per academic term, degree and major combination for a given academic year.

Primary Key: Primary key of the data mart is used to uniquely identify a row in a data mart table. Combination of ACADEMIC_YR, CALENDAR_YR, TERM_CD, TERM_NAME, CAMPUS_CD, STUD_ID, DEG_AWARD_CD and DEG_MAJOR_CD is the primary key for DEGREE_DM table.

3.4 MQTS IN UCDW STUDENT ENROLLMENT

A Materialized query table (MQT) is a table whose definition is based upon the result of a query. The data that is contained in an MQT is derived from one or more tables on which the materialized query table definition is based.

MQTs are materialized views. Both views and MQTs are defined on the basis of a query. The query on which a view is based is run whenever the view is referenced; however, an MQT actually stores the query results as data, and you can work with the data that is in the MQT instead of the data that is in the underlying tables. UCDW-Degree currently has one MQT: DEGREE_HEAD_COUNT_M.

3.4.1 DEGREE_HEAD_COUNT_M

DEGREE_HEAD_COUNT_M provides the correct degree counts regardless of the number of majors awarded to each student. This MQT also has DGR_AWRDED_MULTI_MAJOR_FL column that shows whether the student has multiple majors.

Examples of Degree Head Count MQT Grain:

Example 1 – In Fall 2012, UC Berkeley has awarded Jane Smith 3 degrees as follows:

4. Degree code **08-M.A.** and a single major **699-POLITICAL SCIENCE**
5. Degree code **32-C.PHIL.** and a single major **428-HEALTH SERV & POLICY ANA**
6. Degree code **43-PH.D.** and a single major **428-HEALTH SERV & POLICY ANA**

She will have 3 rows in the fact table.

Example 2 – In Fall 2012, UC Berkeley has awarded John Doe 3 degrees as follows:

4. Degree code **02-A.B.** and double majors **780-PSYCHOLOGY** and **123-BIOLOGY**
5. Degree code **05-B.S.** and a single major **120-BIOLOGICAL SCIENCES**
6. Degree code **02-A.B.** and a single major **867-SOCIOLOGY**

He will have 3 rows in the fact table (double majors are lost).

Look at the below example to better understand the difference between DEGREE_AWARDED_F and DEGREE_HEAD_COUNT_M tables.

Example:

Find Degrees Conferred count by term across all campuses using DEGREE_AWARDED_F fact for academic year 2012.

Sample SQL:

```
select
    t.ACAD_T_ACAD_YR,
    t.ACAD_T_CD,
    t.ACAD_T_NAM,
    COUNT (*) AS HEAD_COUNT
from STUD_BI.DEGREE_AWARDED_F f
inner join STUD_BI.ACADEMIC_TERM_D t
    on f.ACAD_T_KEY = t.ACAD_T_KEY
where ACAD_T_ACAD_YR = 2012
and ACAD_T_CD = '2'
group by
    t.ACAD_T_ACAD_YR,
    t.ACAD_T_CD,
    t.ACAD_T_NAM
order by
    t.ACAD_T_ACAD_YR,
    t.ACAD_T_CD,
    t.ACAD_T_NAM
for read only with ur;
```

ACAD_T_ACAD_YR	ACAD_T_CD	ACAD_T_NAM	HEAD_COUNT
2012	2	Fall	10647

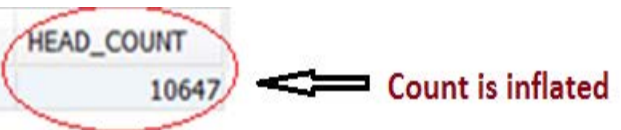


Figure 4: Degree Fact Row Counts

Modify the previous SQL to join to DEGREE_HEAD_COUNT_M MQT instead of the DEGREE_AWARDED_F fact table to see a degrees conferred count by term across all campuses for academic year 2012.

```
select
    t.ACAD_T_ACAD_YR,
    t.ACAD_T_CD,
    t.ACAD_T_NAM,
    COUNT (*) AS HEAD_COUNT
from STUD_BI.DEGREE_HEAD_COUNT_M f
inner join STUD_BI.ACADEMIC_TERM_D t
    on f.ACAD_T_KEY = t.ACAD_T_KEY
where ACAD_T_ACAD_YR = 2012
and ACAD_T_CD = '2'
group by
    t.ACAD_T_ACAD_YR,
    t.ACAD_T_CD,
    t.ACAD_T_NAM
order by
    t.ACAD_T_ACAD_YR,
```

```
t.ACAD_T_CD,
t.ACAD_T_NAM
for read only with ur;
```

You should see the counts below:

ACAD_T_ACAD_YR	ACAD_T_CD	ACAD_T_NAM	HEAD_COUNT
2012	2	Fall	10108

Figure 5: Degree Head Count in MQT



Did you notice that the counts are lower when you use DEGREE_HEAD_COUNT_M MQT instead of DEGREE_AWARDED_F fact table?



Double major students have two rows in the fact table and triple major students have 3 rows in the fact table. The counts on Degree awarded fact table are slightly higher because double major students are counted twice and triple major students are counted three times.

Always use the DEGREE_HEAD_COUNT_M instead of DEGREE_AWARDED_F if you want to count a student only once regardless of their multiple majors.

4. SAMPLE QUERIES

Sample queries for UCDW-Degree are available in JIRA-IRAP Code Library Project. Use this link to access JIRA-IRAP Code Library:

<https://ucidwrequirements.atlassian.net/issues/?jql=project%20%3D%20ICL%20AND%20resolution%20%3D%20Unresolved%20ORDER%20BY%20key%20ASC%2C%20priority%20DESC%2C%20updated%20DESC>

5. TERMS AND ABBREVIATIONS

#	ACRONYM	MEANING
1	UCDW	University of California Data Warehouse
2	MQT	Materialized Query Table
3	FERPA	Family Educational Rights and Privacy Act
4	UC	University of California
5	UCOP	University of California Office of the President
6	MQT	Materialized Query Table
7	BI	Business Intelligence
8	EOT	End of Term
9	3WK	Third Week
10	SAS	Statistical Analysis System
11	SCD	Slowly Changing Dimension
12	SQL	Structured Query Language

Table 1: Glossary