

DWH281 Project

Glory[600646] | Tokelo[601009] | Moloko[601352] | Yandile [601553]

Requirements gathering

- End users of the data warehouse that is to be designed include the staff members, lecturers as well as the Dean.
 - The data warehouse shall be used to capture the progress of students' efforts in final class tests per semester at Belgium Campus iTiversity.
 - The Dean of this private institution shall gather information of the average performance of students in class tests within a particular task by initiating interviews with lecturers regarding test performance, lecture attendance and assessment completion.
 - Questions asked in the data warehouse to the end users should capture information about the number of students that have achieved well in exams, compared to assignments and weekly tests, percentage of students that qualified for mid-term and final exams, comparison of the average grades of different groups of exams within the same module, correlation between accomplishment of success and type of high school the student graduated from, e.g., IEB or Cambridge, questions with more correct answers and difficulty of questions.as well as the memorandum.
 - The memo is created for the teaching staff to provide to students for finding mistakes made in tests, exams and or assignments to learn lessons and get them captured for future mid-term and final exams. This can improve the average overall student performance.
 - Recommended textbooks are provided for students to prepare for tests, assignments, and exams, especially mid-terms and final exams.
 - Daily average work is set to qualify students for mid-term and final exams.
 - Daily average work composes of the average mark of the final summative test(Testing of all work covered in the whole module), total number of weekly tests, total number of assignments and/or projects covered in a module, as well as lecture attendance
 - Students do not qualify for mid-term exams if the students lecture attendance is below 85% OR the student achieves less than 50% for their daily average work.
 - Daily Average work counts 30% of final mark, whereas mid-term and final exams count 70% of the final marks.
-

Definition of Users' Requests

Establishing user requirements plays a crucial role in the formulation of a data warehouse design. This entails comprehending the information requirements of end users and the nature of queries they intend to execute within the data warehouse. In the context of this initiative, the primary end users are the teaching staff and other interested members of the staff at

Belgium Campus. By delving into and specifying these user requests, a tailored data warehouse can be crafted to efficiently meet the needs of end users. This, in turn, empowers them to derive insights from data, facilitating informed decision-making and enhancing the teaching and learning experience at Belgium Campus.

These requests may include performance analysis, attendance and assessment impact, student performance trends, course performance metrics and exam question analysis.

Here's how one can group it:

Performance Analysis	Users might want to analyse students' performance across various dimensions, such as courses, exam types, and time.	"Show me the average scores of students in a specific course or across all courses." "How many students passed a particular course or all courses in a semester?" "Which courses have the highest or lowest pass rates?"
Attendance and Assessment Impact	Users might be interested in understanding the relationship between lecture attendance, weekly assessment scores, and exam performance.	"How does lecture attendance correlate with students' performance in exams and assessments?" "How do weekly assessment scores compare to mid-term and final exam scores?"
Student Performance Trends:	Users might want to identify trends in student performance across different courses or semesters.	"Are there any noticeable trends in student performance across different courses or semesters?" "Which students consistently perform well or struggle in their courses?"
Course Performance Metrics:	Users might want to compare course performance metrics between different academic years or semesters.	Which courses have the highest or lowest average weekly assessment scores?" "Which courses have the highest or lowest lecture attendance rates?"

Exam Question Analysis:	Users might want to analyse the exam questions to improve the quality of assessments.	"What is the correctness of the answers to the questions on the exams?" "Which questions have the majority of correct answers, and which are most difficult to answer?"
--------------------------------	---	--

Our Data Warehouse consists of six dimension tables and three fact tables designed to capture various aspects of student performance, exam details, and course assessments.

Defining the Fact Tables

CoursePerformance_Fact: This fact table captures student performance metrics
StudentID(Foreign Key) CourseID(Foreign Key) Semester ExamScores Attendance WeeklyAssessmentScores TotalScores, Pass_Fail_Status Date ExamID(Foreign Key)

ExamResults_Fact: This fact table stores exam-related information.
StudentID(Foreign Key) ExamID(Foreign Key) ExamType ExamScores CorrectAnswers DifficultyLevel AverageScores Pass_Fail_Status QuestionID(Foreign Key)

AssessmentResult_Fact: This fact table contains course assessment metrics
StudentID(Foreign Key)

AssessmentID(Foreign Key) AssessmentType AssessmentScores AverageScores Pass_Fail_Status
--

Defining The Dimensional Tables

Student_Dim

StudentID (Primary Key) Name HighSchoolTypeID(Foreign Key)
--

Course_Dim

CourseID (Primary Key) Name Credits Lecture_ID(Foreign Key)
--

Exam_Dim

ExamID (Primary Key) ExamType Date
--

Assessment_Dim

AssessmentID (Primary Key) AssessmentType Date
--

Date_Dim

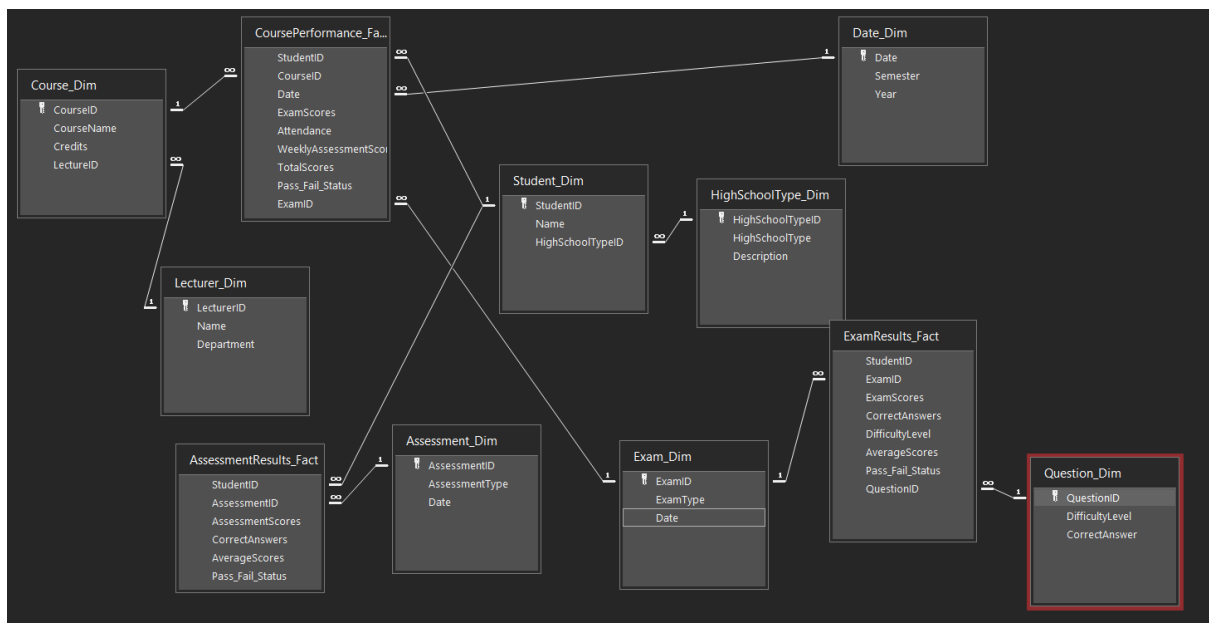
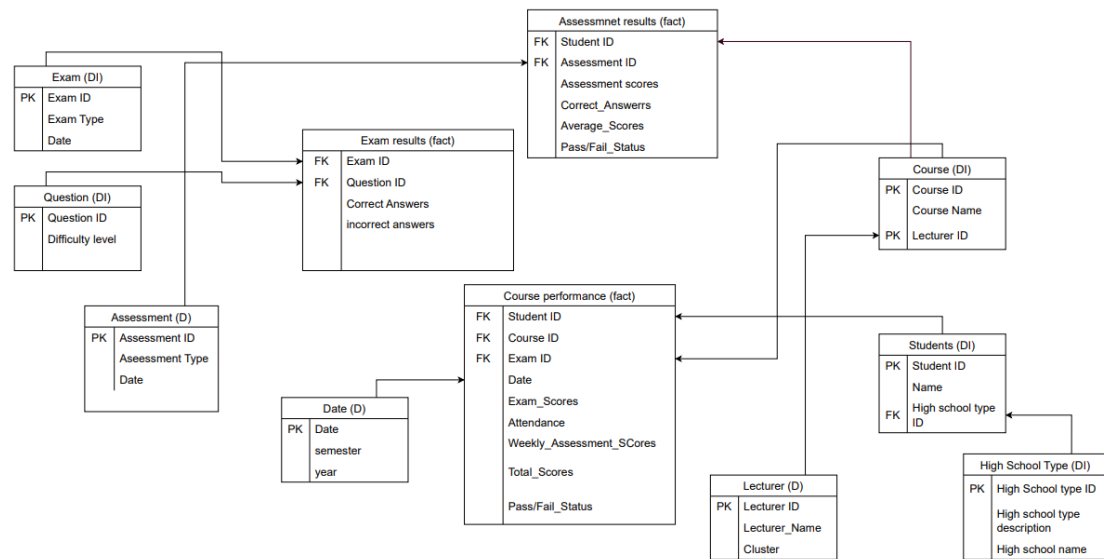
Date (Primary Key) Semester Year
--

Lecturer_Dim



















LecturerID (Primary Key) LecturerName Department
--

Question_Dim
QuestionID (Primary Key)
DifficultyLevel
CorrectAnswer

Snowflake Schema



Tables and Queries

Tables	
	Assessment_Dim
	AssessmentResults_...
	Course_Dim
	CoursePerformance...
	Date_Dim
	Exam_Dim
	ExamResults_Fact
	HighSchoolType_Dim
	Lecturer_Dim
	Question_Dim
	Student_Dim
Queries	
	CorrectnessPercenta...
	ExamSuccess_HighS...
	Passed_Students
	PASSED_TermAsses...
	PassedByDate
	PercentagePassedM...
	SuccessAttendance...

Defining extraction, transformation and loading process

The ETL (Extract, Transform, Load) process, which involved extracting data from the school operational database into the target database in the data warehouse, comprised the following steps:

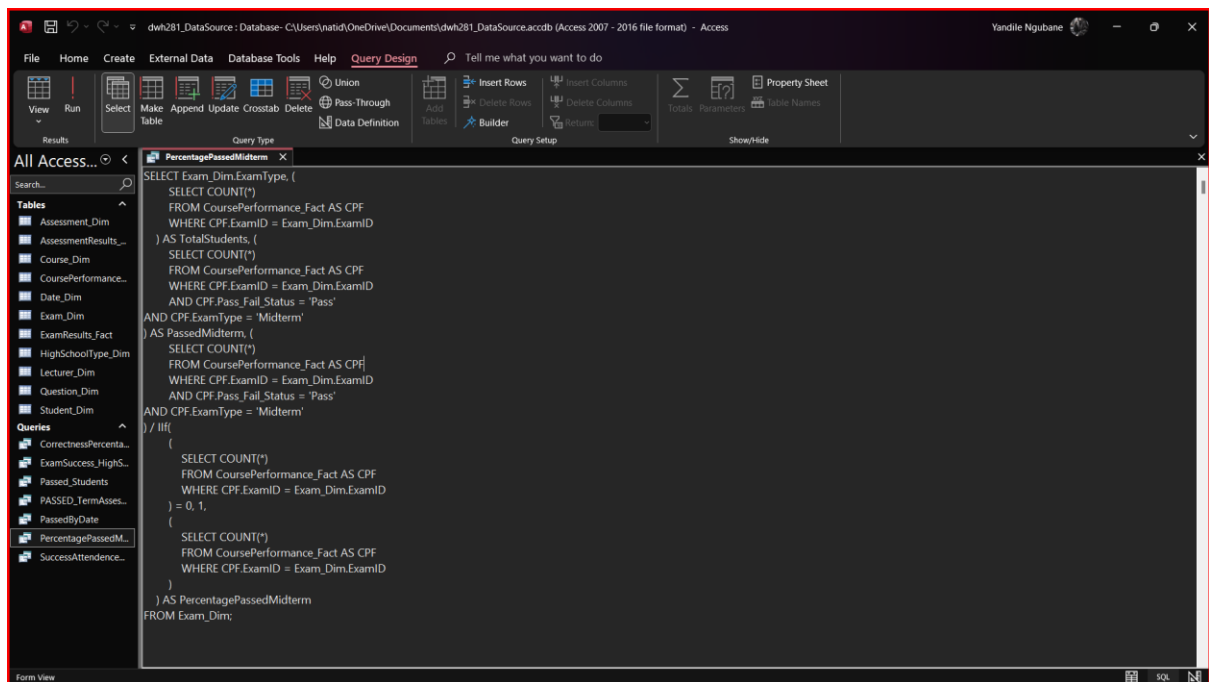
Extracting:

- Connected to the source operational database.
- Retrieved the required data from the source database using SQL queries or other appropriate methods. In this case, the data was extracted from the dimension tables (Student_Dim, Course_Dim, Exam_Dim, Assessment_Dim, Date_Dim, Lecturer_Dim, and Question_Dim).

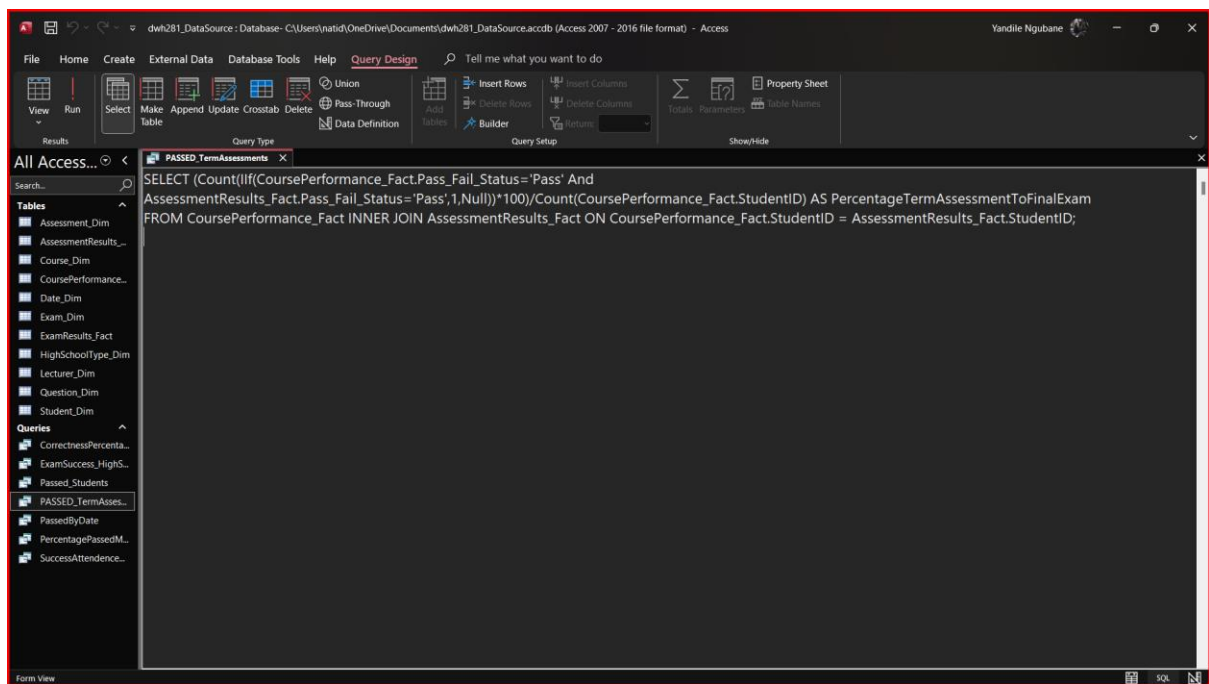
Transforming:

- Cleaned and validated the extracted data to ensure its quality and consistency. This involved handling missing values, removing duplicates, checking data types, and enforcing constraints.

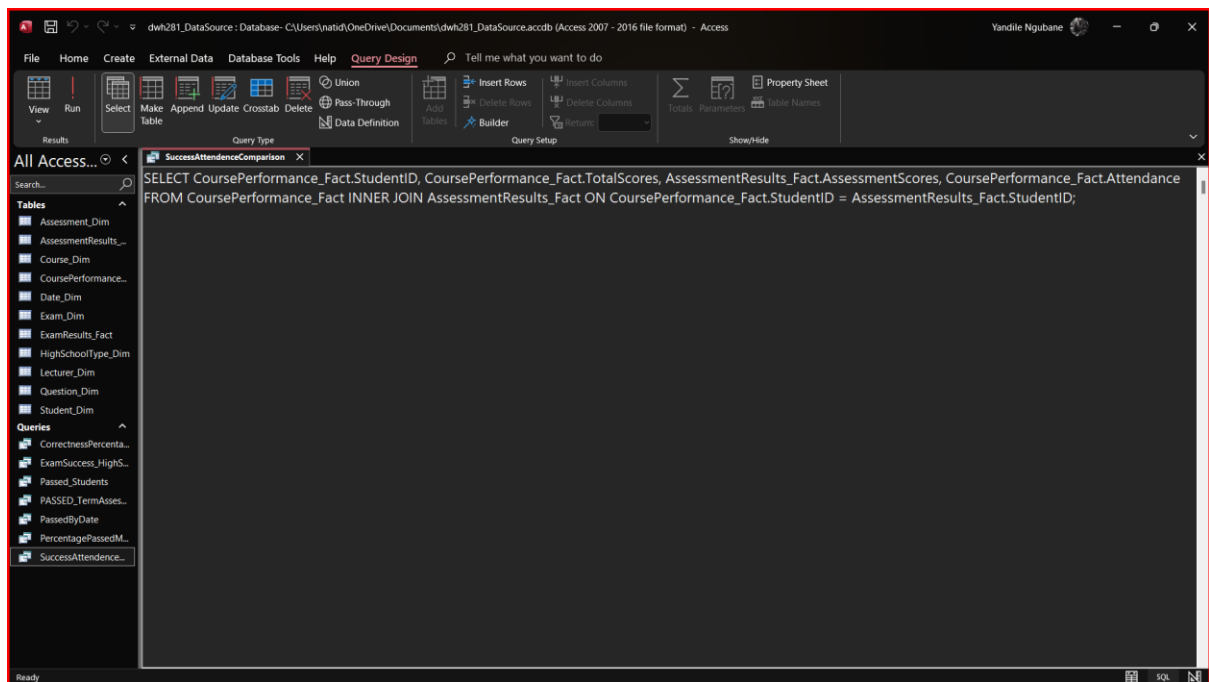
2. Percentage of students having passed the exam by taking the mid-term exams.



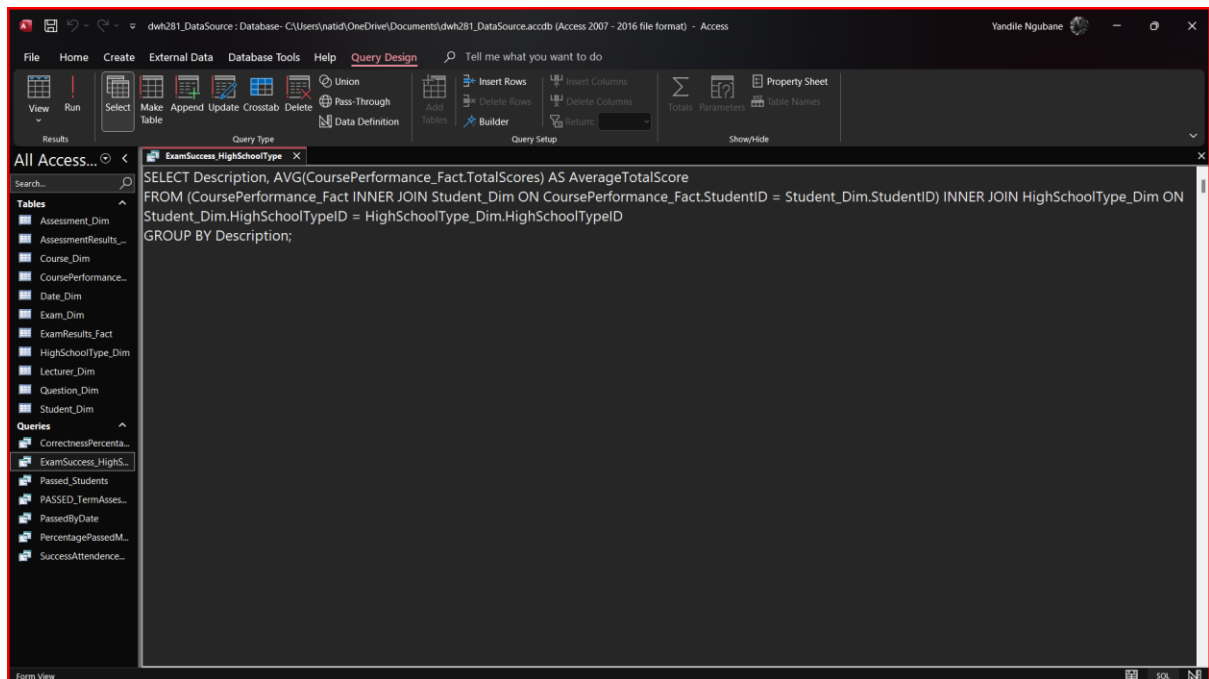
3. Percentage of students whose term assessment scores allow them take final exam.



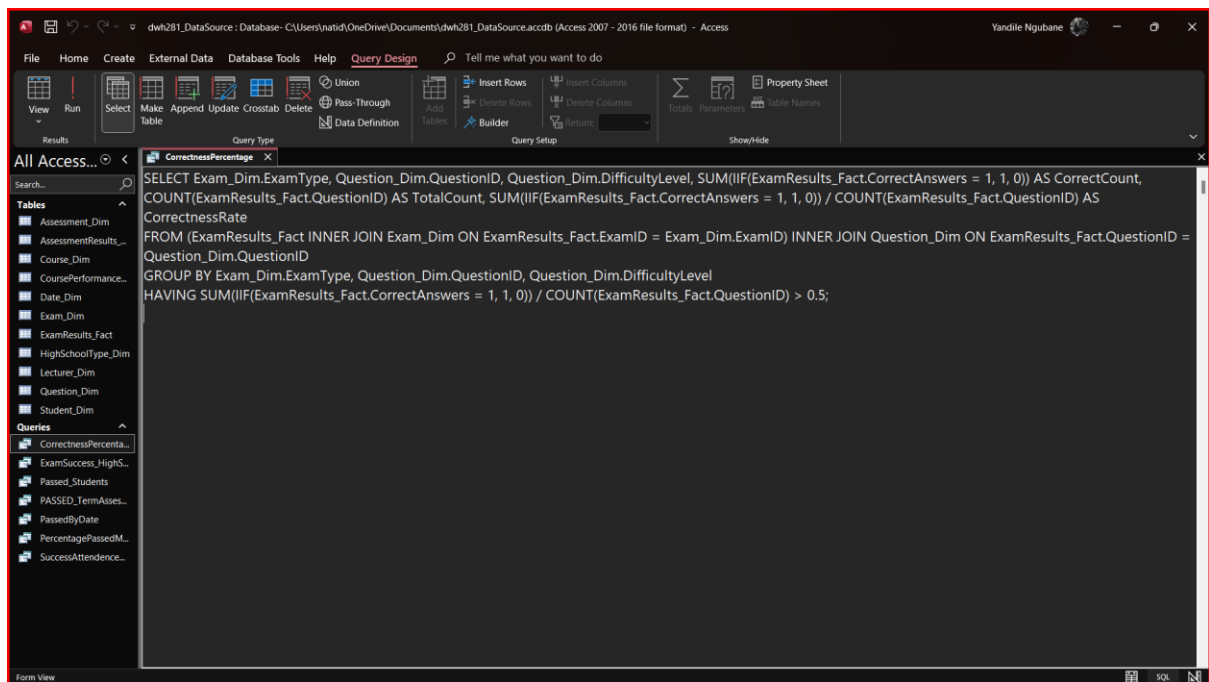
4. The grades that students got compared to their success in term assessment and attendance of the lectures.



5. Correlation between success on exams and different high school types the students attended.



6. Correctness of the answers to the questions on the exams.



Extraction, transformation and loading process scripts

Loading/Transformation Scripts(SQL)

Student Dim

```
INSERT INTO Student_Dim (Name, HighSchoolType)
SELECT Name, HighSchoolType
FROM Student_Source;
```

Course Dim

```
INSERT INTO Course_Dim (CourseName, Credits)
SELECT CourseName, Credits
FROM Course_Source;
```

Exam Dim

```
INSERT INTO Exam_Dim (ExamType, Date)
SELECT ExamType, Date
FROM Exam_Source;
```

Assessment Dim

```
INSERT INTO Assessment_Dim (AssessmentType, Date)
SELECT AssessmentType, Date
FROM Assessment_Source;
```

Date Dim

```
INSERT INTO Date_Dim (Date, Semester, Year)
SELECT Date, Semester, Year
FROM Date_Source;
```

Lecturer Dim

```
INSERT INTO Lecturer_Dim (Name, Department)
SELECT Name, Department
FROM Lecturer_Source;
```

Question Dim

```
INSERT INTO Question_Dim (DifficultyLevel, CorrectAnswer)
SELECT DifficultyLevel, CorrectAnswer
FROM Question_Source;
```

CoursePerformance Fact

```
INSERT INTO CoursePerformance_Fact (StudentID, CourseID, Date,
ExamScores, Attendance, WeeklyAssessmentScores, TotalScores,
Pass_Fail_Status)
SELECT s.StudentID, c.CourseID, cp.Date, cp.ExamScores, cp.Attendance,
cp.WeeklyAssessmentScores, cp.TotalScores, cp.Pass_Fail_Status
FROM CoursePerformance_Source cp
(INNER JOIN Student_Source s ON cp.StudentID = s.StudentID)
```

```
INNER JOIN Course_Source c ON cp.CourseID = c.CourseID;
```

ExamResults Fact

```
INSERT INTO ExamResults_Fact (StudentID, ExamID, ExamScores,  
CorrectAnswers, DifficultyLevel, AverageScores, Pass_Fail_Status)  
SELECT s.StudentID, e.ExamID, er.ExamScores, er.CorrectAnswers,  
er.DifficultyLevel, er.AverageScores, er.Pass_Fail_Status  
FROM ExamResults_Source er  
(INNER JOIN Student_Source s ON er.StudentID = s.StudentID)  
INNER JOIN Exam_Source e ON er.ExamID = e.ExamID;
```

AssessmentResults Fact

```
INSERT INTO AssessmentResults_Fact (StudentID, AssessmentID,  
AssessmentScores, CorrectAnswers, AverageScores, Pass_Fail_Status)  
SELECT Student_Dim.StudentID, Assessment_Dim.AssessmentID,  
AssessmentResults_Facts.AssessmentScores,  
AssessmentResults_Facts.CorrectAnswers,  
AssessmentResults_Facts.AverageScores,  
AssessmentResults_Facts.Pass_Fail_Status  
FROM AssessmentResults_Source ar  
(INNER JOIN Student_Source Student_Dim ON  
AssessmentResults_Facts.StudentID = s.StudentID)  
INNER JOIN Assessment_Source a ON ar.AssessmentID = a.AssessmentID;
```

Loading Scripts(SQL)

The loading SQL Scripts are based on the assumption that they was a middle temporary table that the extracted and transformed data was loaded into.

Date Dim

```
INSERT INTO Date_Dim (Date, Semester, Year)  
SELECT Date, Semester, Year  
FROM Transformed_DateData;
```

Lecturer Dim

```
INSERT INTO Lecturer_Dim (Name, Department)  
SELECT Name, Department  
FROM Transformed_LecturerData;
```

Question Dim

```
INSERT INTO Question_Dim (DifficultyLevel, CorrectAnswer)  
SELECT DifficultyLevel, CorrectAnswer  
FROM Transformed_QuestionData;
```

Course Dim

```
INSERT INTO Course_Dim (CourseName, Credits)  
SELECT CourseName, Credits  
FROM Transformed_CourseData;
```

Assessment Dim

```
INSERT INTO Assessment_Dim (AssessmentType, Date)
SELECT AssessmentType, Date
FROM Transformed_AssessmentData;
```

Exam Dim

```
INSERT INTO Exam_Dim (ExamType, Date)
SELECT ExamType, Date
FROM Transformed_ExamData;
```

Student Dim

```
INSERT INTO Student_Dim (Name, HighSchoolType)
SELECT Name, HighSchoolType
FROM Transformed_StudentData;
```

CoursePerformance Fact

```
INSERT INTO CoursePerformance_Fact (StudentID, CourseID, Date,
ExamScores, Attendance, WeeklyAssessmentScores, TotalScores,
Pass_Fail_Status)
SELECT StudentID, CourseID, Date, ExamScores, Attendance,
WeeklyAssessmentScores, TotalScores, Pass_Fail_Status
FROM Transformed_CoursePerformanceData;
```

ExamResults Fact

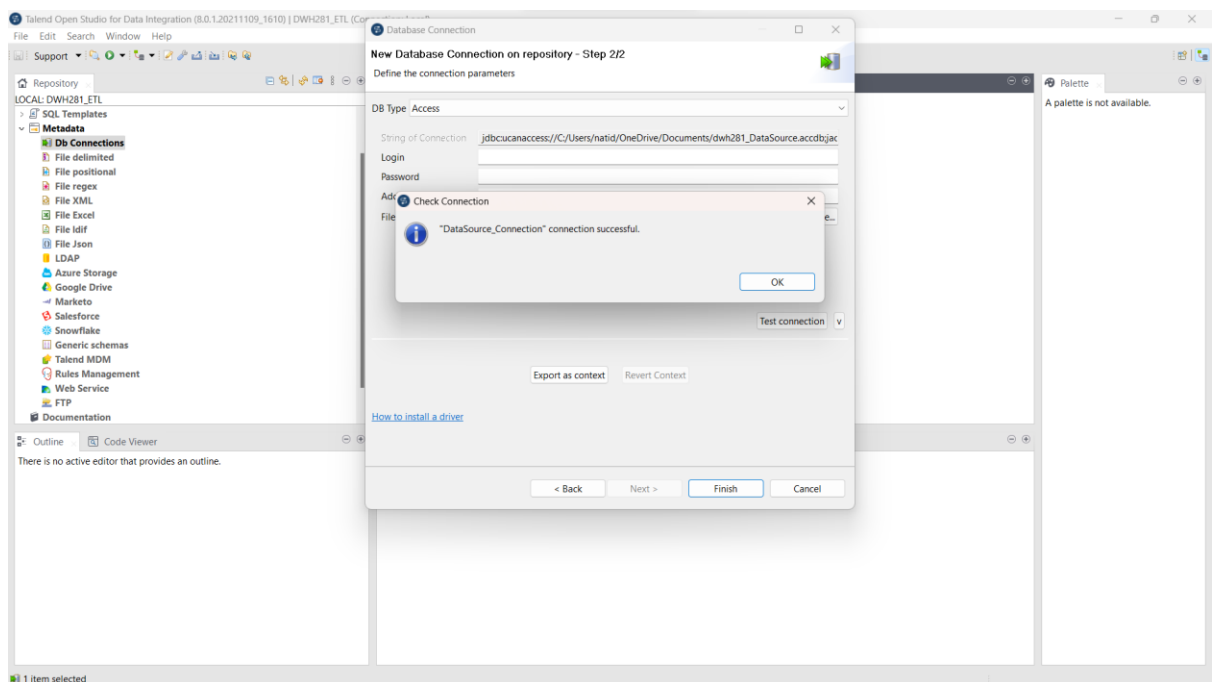
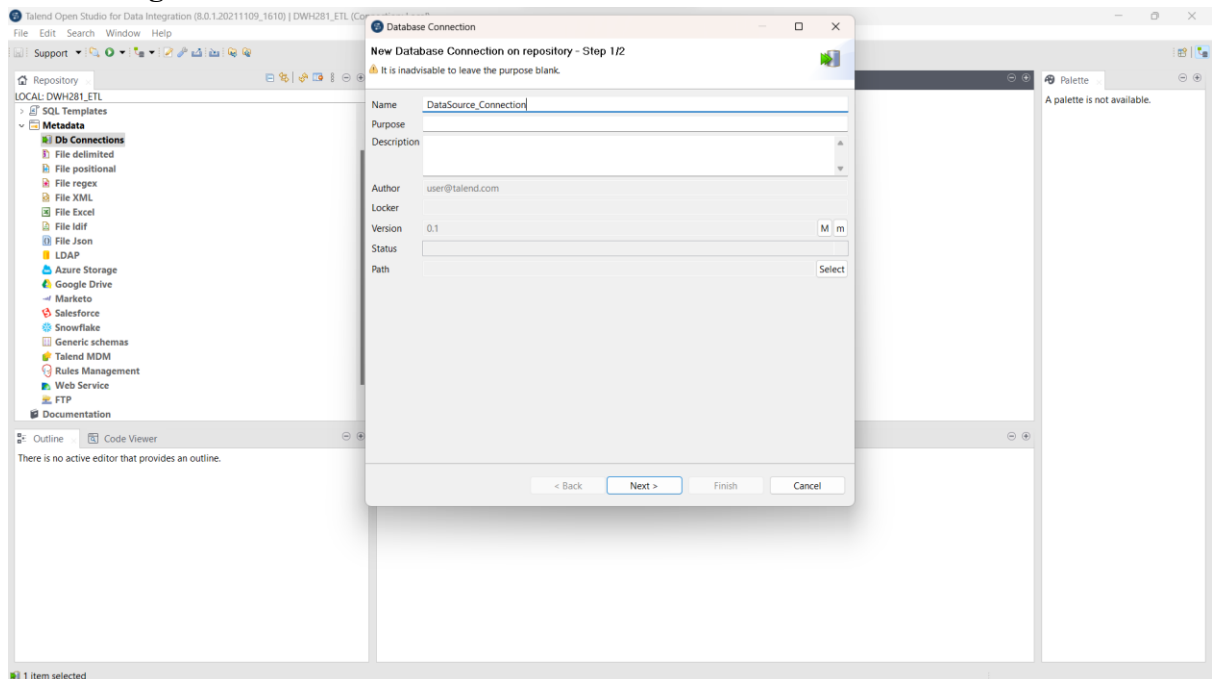
```
INSERT INTO ExamResults_Fact (StudentID, ExamID, ExamScores,
CorrectAnswers, DifficultyLevel, AverageScores, Pass_Fail_Status)
SELECT StudentID, ExamID, ExamScores, CorrectAnswers, DifficultyLevel,
AverageScores, Pass_Fail_Status
FROM Transformed_ExamResultsData;
```

AssessmentResults Fact

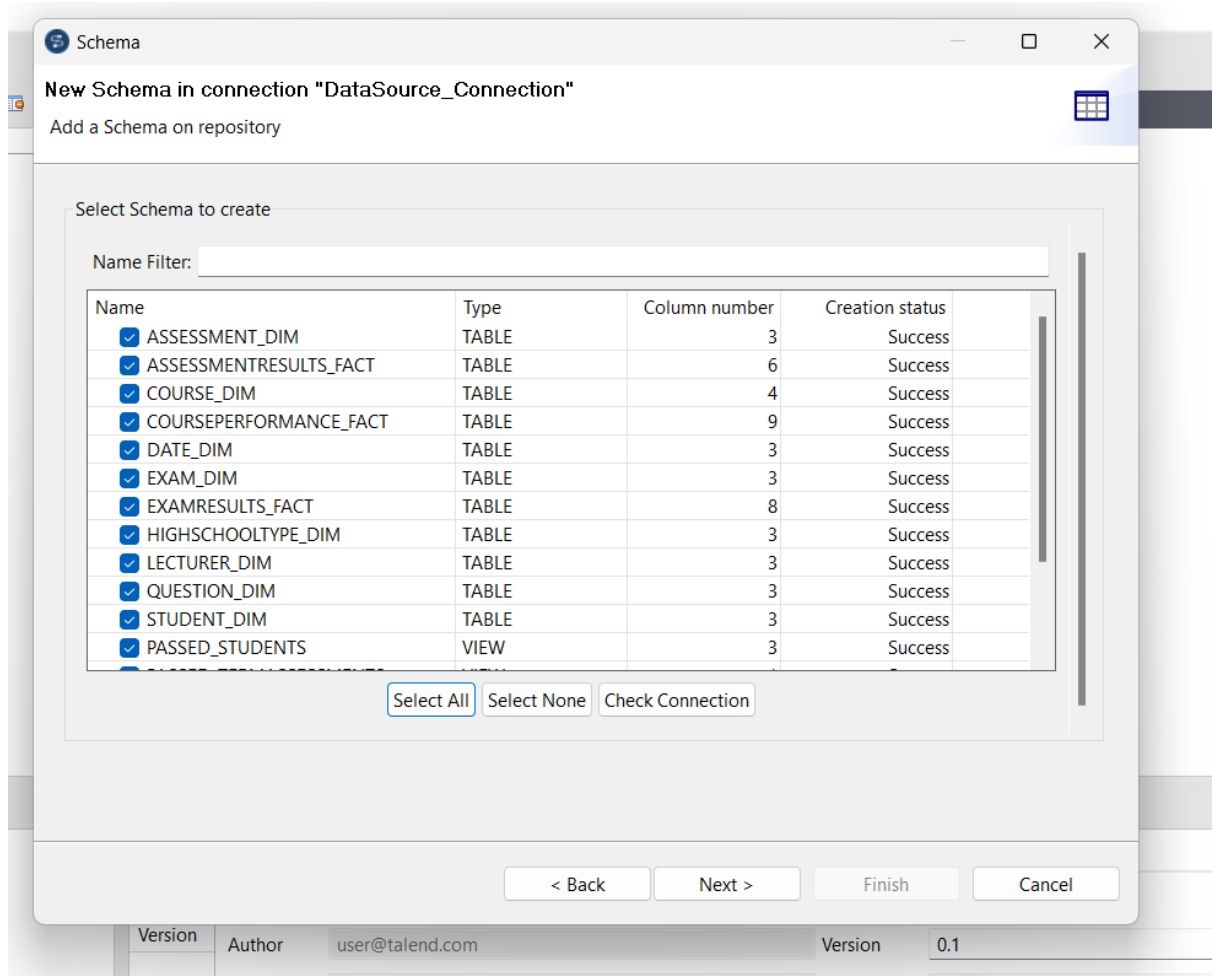
```
INSERT INTO AssessmentResults_Fact (StudentID, AssessmentID,
AssessmentScores, CorrectAnswers, AverageScores, Pass_Fail_Status)
SELECT StudentID, AssessmentID, AssessmentScores, CorrectAnswers,
AverageScores, Pass_Fail_Status FROM Transformed_AssessmentResultsData;
```

Initial Load of the Datawarehouse

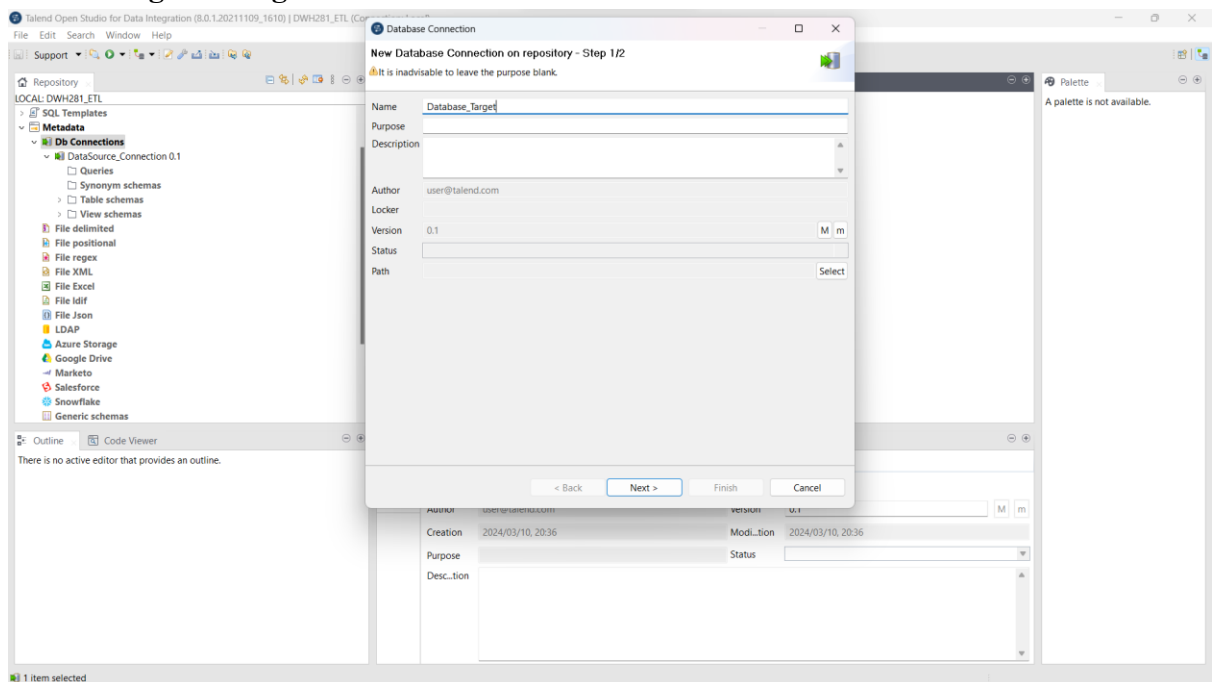
1. Connecting the Data Source Database to extract from:

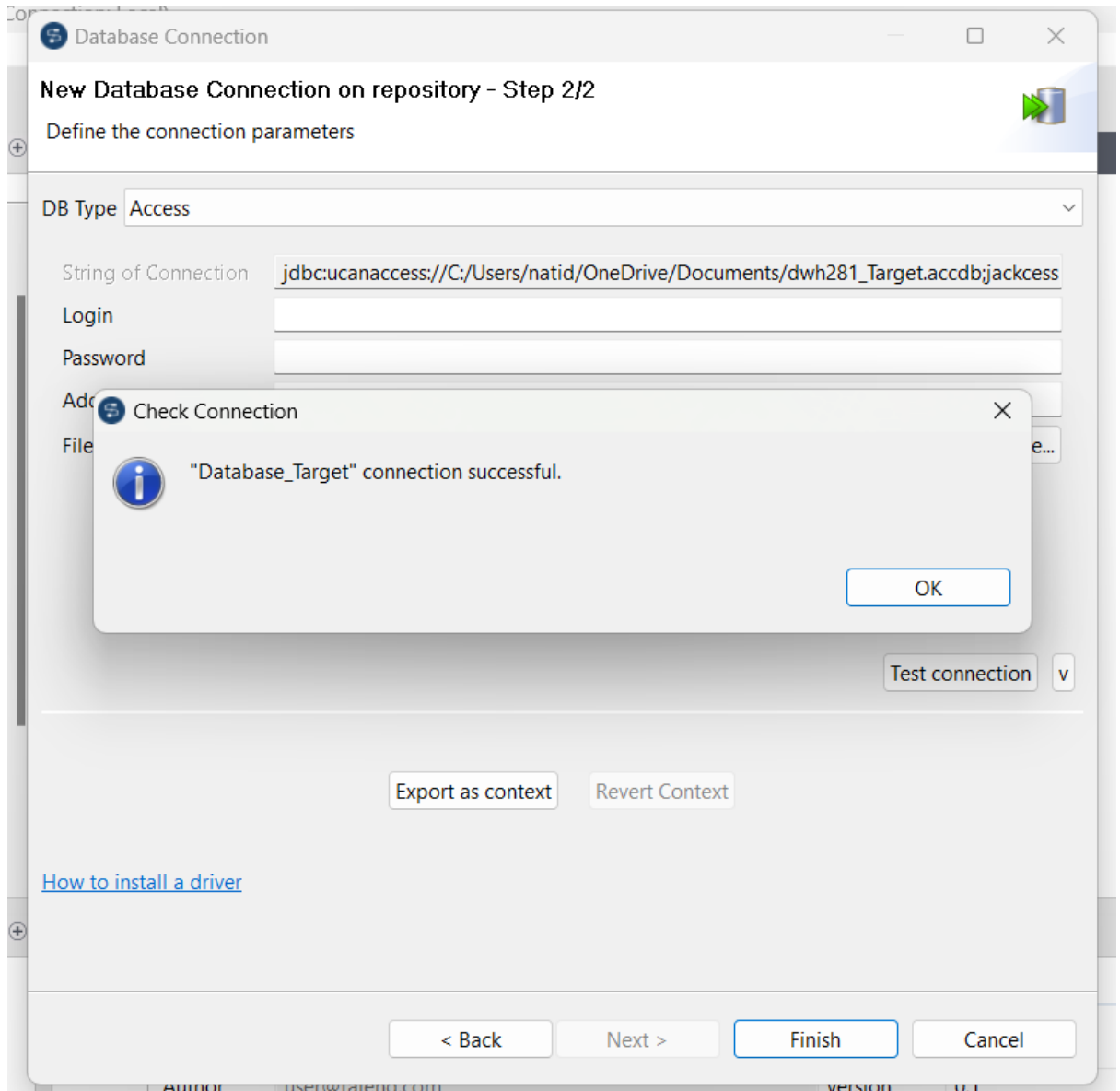


2. Retrieving Schema

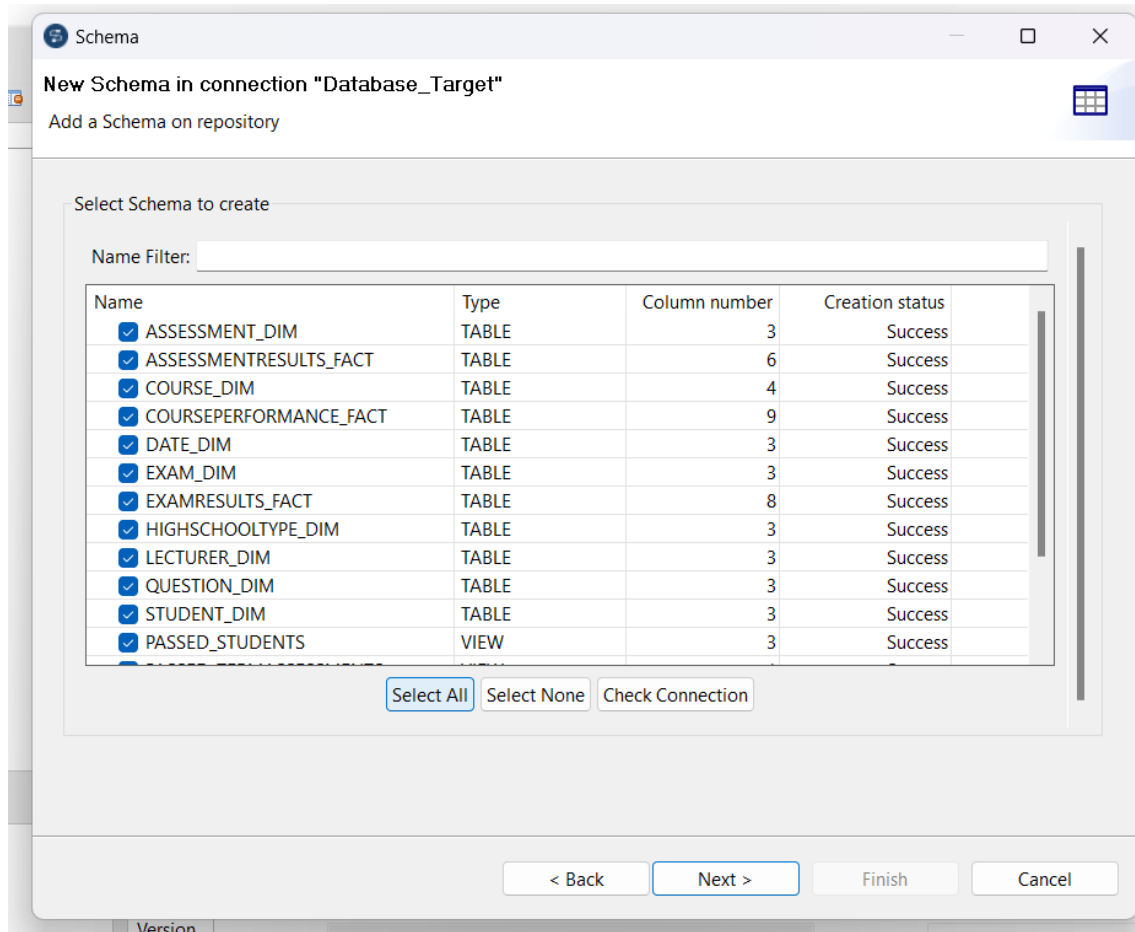


3. Connecting the Target Database





4. Retrieving Schema



Creating Initial Load Job

New job

⚠ It is inadvisable to leave the purpose blank.

Name:

Purpose:

Description:

Author:

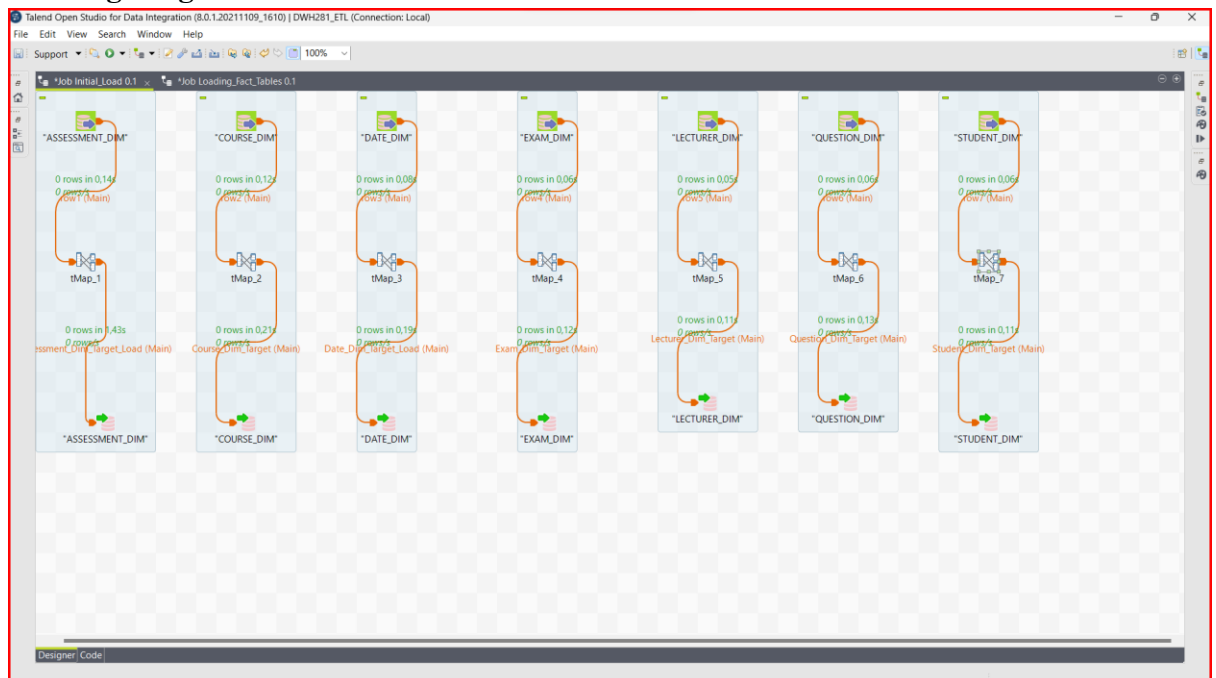
Locker:

Version:

Status:

Path:

1. Loading Target Dimension Tables



2. Loading Fact Tables

New job

It is inadvisable to leave the purpose blank.

Name: Loading_Fact_Tables

Purpose:

Description:

Author: user@talend.com

Locker:

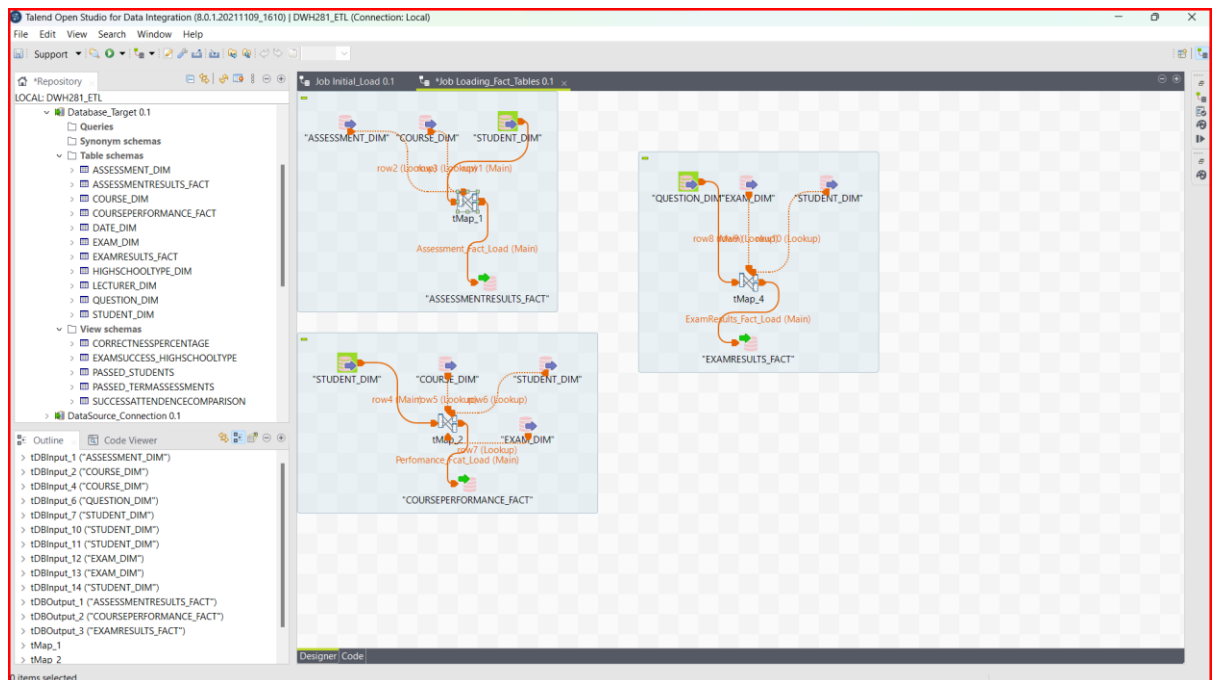
Version: 0.1

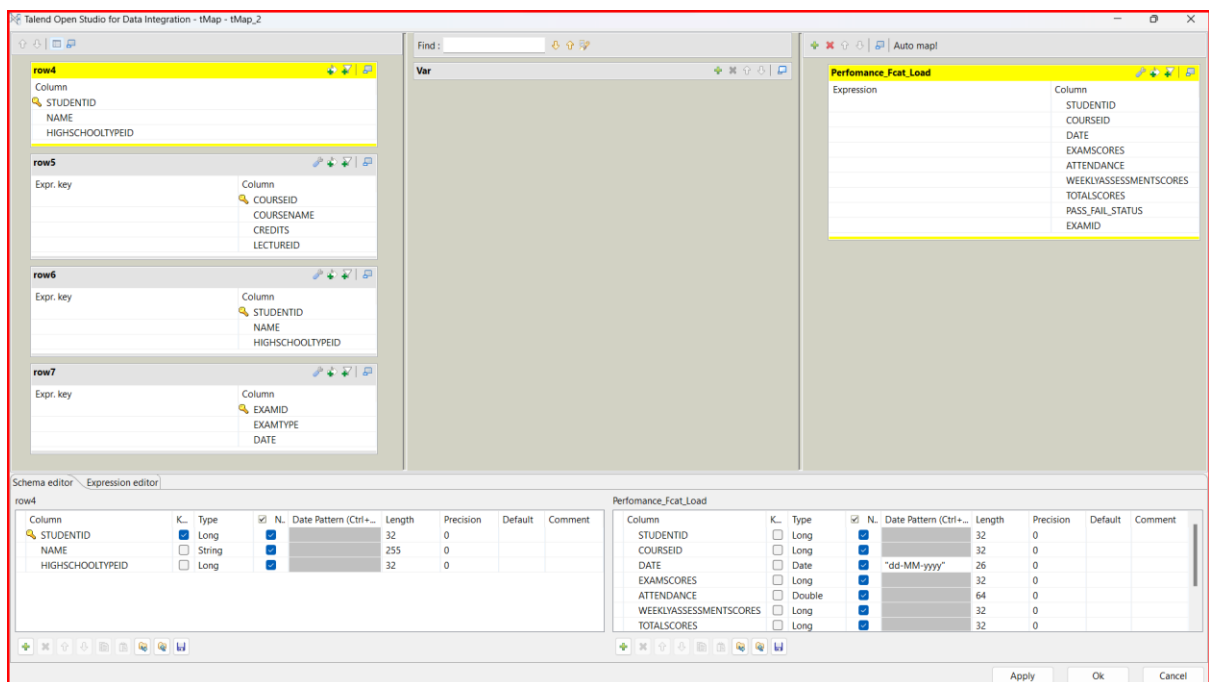
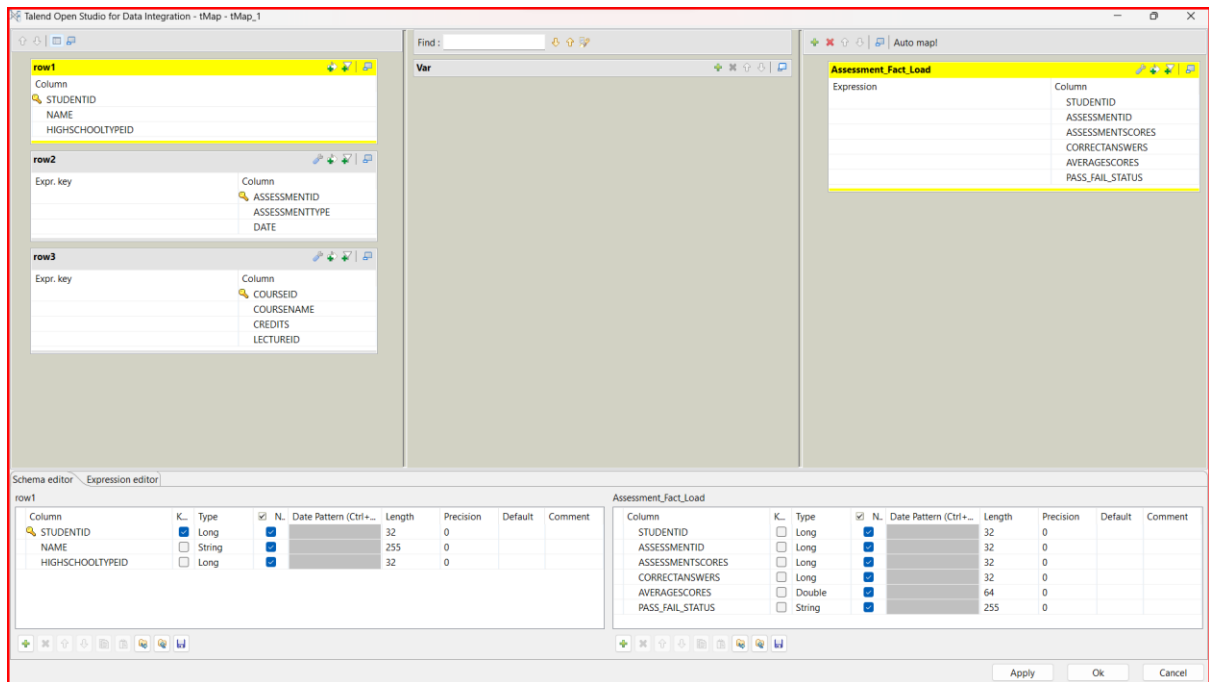
Status:

Path:

Finish Cancel

oad





Talend Open Studio for Data Integration - tMap - tMap_4

row8

Column
QUESTIONID
DIFFICULTYLEVEL
CORRECTANSWER

row9

Expr. key	Column
	EXAMID
	EXAMTYPE
	DATE

row10

Expr. key	Column
	STUDENTID
	NAME
	HIGHSCHOOLTYPEID

Find:

Var

ExamResults_Fact_Load

Expression	Column
	STUDENTID
	EXAMID
	EXAMSCORES
	CORRECTANSWERS
	DIFFICULTYLEVEL
	AVERAGESCORES
	PASS_FAIL_STATUS
	QUESTIONID

Schema editor | Expression editor

row8

Column	K...	Type	<input checked="" type="checkbox"/> N...	Date Pattern (Ctrl+...	Length	Precision	Default	Comment
QUESTIONID		<input checked="" type="checkbox"/> Long	<input checked="" type="checkbox"/>		32	0		
DIFFICULTYLEVEL		<input type="checkbox"/> String	<input checked="" type="checkbox"/>		255	0		
CORRECTANSWER		<input type="checkbox"/> String	<input checked="" type="checkbox"/>		255	0		

ExamResults_Fact_Load

Column	K...	Type	<input checked="" type="checkbox"/> N...	Date Pattern (Ctrl+...	Length	Precision	Default	Comment
STUDENTID		<input type="checkbox"/> Long	<input checked="" type="checkbox"/>		32	0		
EXAMID		<input type="checkbox"/> Long	<input checked="" type="checkbox"/>		32	0		
EXAMSCORES		<input type="checkbox"/> Long	<input checked="" type="checkbox"/>		32	0		
CORRECTANSWERS		<input type="checkbox"/> Long	<input checked="" type="checkbox"/>		32	0		
DIFFICULTYLEVEL		<input type="checkbox"/> String	<input checked="" type="checkbox"/>		255	0		
AVERAGESCORES		<input type="checkbox"/> Double	<input checked="" type="checkbox"/>		64	0		
PASS_FAIL_STATUS		<input type="checkbox"/> String	<input checked="" type="checkbox"/>		255	0		

Apply Ok Cancel