



# Data Science | 2015-2016 Assessment Report

1. Please give a brief overview of the assessment data you collected this year.

MSDS Assessment committee decided to assess the program learning goal (PLG C) for the year 2015-2016. The assessment was carried out in two courses: DS 715 and DS 730 in Spring 2016 using embedded class assessments. We assessed the following two traits of PLG C: (i) Students will be able to choose and apply tools and methodologies to solve data science tasks (DS 730), and (ii) Students will be able to assess the model used to solve data science tasks (DS 715)

## 1. Summary of the assessment:

PLG	Trait	Number of Students	Course	Method of Assessment	Conducted by
C	Students will be able to choose and apply tools and methodologies to solve data science tasks	6	DS730 Big Data: High Performance Computing	Final Project	Prof. Erik Krohn (UW Oshkosh)
	Students will be able to assess the model used to solve data science tasks	21	DS715 Data Warehousing	Two assignments (4 and 5)	Prof. Robert Dolinger (UW Stevens Point)

### 1.1 Outcome being assessed: Students will be able to choose and apply tools and methodologies to solve data science tasks.

**Course Number and Title:** DS730 Big Data: High Performance Computing

**Semester and Year:** Spring 2016

**Method of Assessment:** Final Project

**Examples:** The final project is a large project that requires the students to solve many different data science tasks. In most of the problems, the task at hand was to answer a question based on the supplied data. The data given was a weather history file and they had to answer certain questions such as: “Which hour of the day was the hottest?” or “How many years did it rain more in February than it did in March?” The students learned the following tools for solving these types of problems: Java Threads, Python MapReduce, Hadoop MapReduce, Pig and Hive. The students were to solve each problem using the best software for the problem.

**Results:** Of the 6 students in the course, 3 received at least an 80%, the remaining achieved a grade of, on average, 61%. The detailed breakup is shown below.

Percentage Points	Number of students
0 – 20%	0
21- 40%	0
41 – 60%	2
61 – 80%	2
81 – 100%	2

**Possible Improvements:** This is a tough one to assess because the choice of software can be subjective. This is especially true when it comes to choosing Pig vs Hive. There is no clear winner on when to choose one over the other. Therefore, a good 75% of the assessment is subjective. To keep it objective, most of the assessment came from: did they solve the problem correctly. If yes, then they probably made a good choice. If no, then they likely made a poor choice. A possible improvement would be to create problems in such a way that the obvious choice is to use, for example, Pig.

## 1.2 Outcome being assessed: Students will be able to assess the model used to solve data science tasks.

**Course Number and Title:** DS715 Data Warehousing

**Semester and Year:** Spring 2016

**Method of Assessment:** Assignment#4 and Assignment#5.

**Examples:** Students use OLAP to solve a predefined set of queries on a given data warehouse model featuring one fact table and several dimension tables with hierarchies.

In Assignment#4 students solve the queries by using basic SQL on a multidimensional database provided as a cube stored in a relational table. At this time students are expected to score well on this task which is rather trivial assuming that SQL is already a known quantity.

In Assignment#5 students are provided with a multidimensional data cube created via SQL Server Analysis Service (SSAS) and are required to solve the same queries by using the Multidimensional Expressions (MDX) Query Language. This task is slightly more difficult as students are instructed and expected to pick-up on the MDX syntax as they go by using the tools provided in SSAS. In this effort they rely on their SQL skills, understanding of OLAP, and knowledge of the model and the results it should produce from Assignment#4.

**Results:** Together the students scored between 80 to 100% in both the assignments. The detailed breakup is given as follows:

### **Assignment#4:**

All 22 students scored above 90%.

Lowest score was 92%.

Highest score was: 100%.

Average: 99.45%

**Assignment#5:**

21 students completed the assignment and scored above 80%.

Lowest score was 84%.

Highest score was: 100%.

Average: 94.27%

**Possible Improvements:** Improvements can be made and have been put in place for improving and consolidating basic SQL skills. This predates Assignment#4 and Assignment#5 and is done in Assignment#2 and somewhat in Assignment#3. The extra effort is needed due to the wide range of SQL expertise students come into the Data Science program.

**Instructor Comments:** Students were quite comfortable in picking-up the features of the Multidimensional Expressions (MDX) Query Language once they had a good understanding of the underlying concepts for multidimensional databases and OLAP. Assignment#5 would be much more difficult if not preceded by Assignment#4. Having solved the same OLAP task first in SQL and then in MDX makes things easier and consolidates the OLAP skills.

2. How will you use what you've learned from the data that was collected?

An important task for MSDS assessment committee for 2016-2017 will be to put assessment plan in place to start assessing most of the PLG, and also to take corrective action based on the results from 2015-2016 assessment data.