BIG DATA

SECTION D

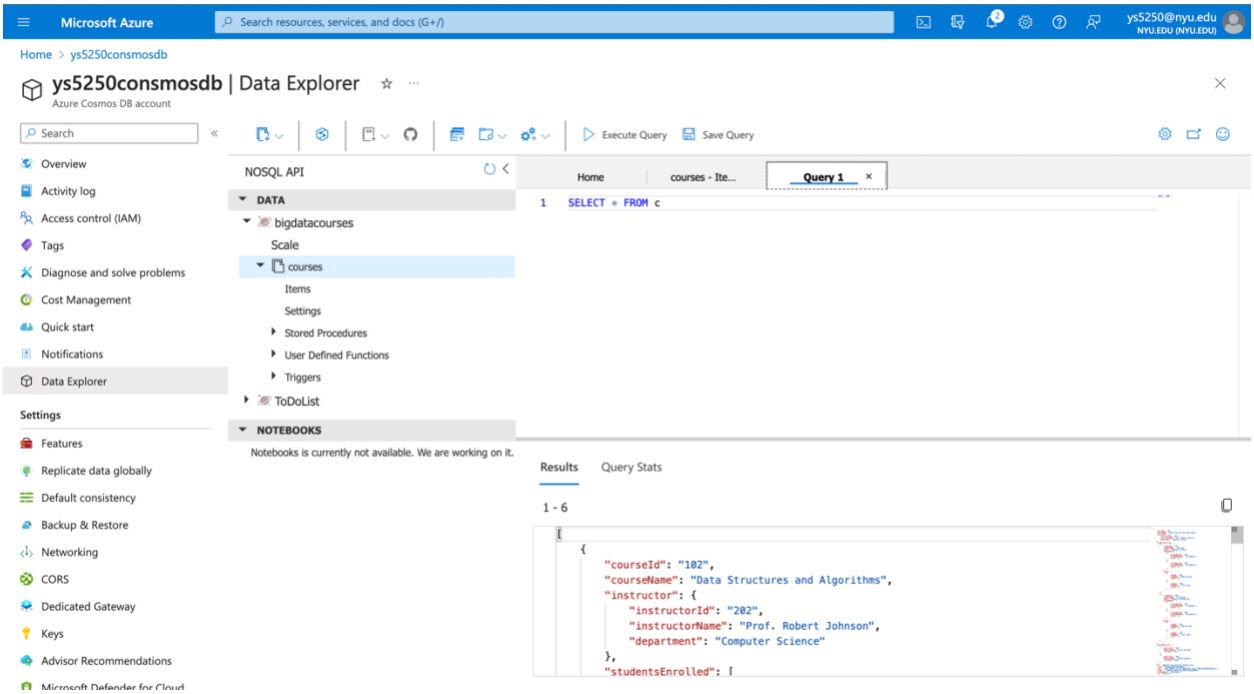
Fall’2023

Yogya Sharma

ys5250

GHW#3

Created database:




```
[
  {
    "courseId": "102",
    "courseName": "Data Structures and Algorithms",
    "instructor": {
      "instructorId": "202",
      "instructorName": "Prof. Robert Johnson",
      "department": "Computer Science"
    },
    "studentsEnrolled": [
      {
        "studentId": "301",
        "studentName": "John Doe",
        "major": "Computer Science",
        "assignments": [
          {
            "assignmentId": "405",
            "assignmentName": "Assignment 1",
            "score": 95
          },
          {
            "assignmentId": "406",
            "assignmentName": "Assignment 2",
            "score": 88
          }
        ],
        "exams": [
```

```
            {
                "examId": "505",
                "examName": "Midterm Exam",
                "score": 92
            },
            {
                "examId": "506",
                "examName": "Final Exam",
                "score": 85
            }
        ]
    },
    {
        "studentId": "303",
        "studentName": "Ella Brown",
        "major": "Computer Engineering",
        "assignments": [
            {
                "assignmentId": "407",
                "assignmentName": "Assignment 1",
                "score": 88
            },
            {
                "assignmentId": "408",
                "assignmentName": "Assignment 2",
                "score": 92
            }
        ],
        "exams": [
            {
                "examId": "507",
                "examName": "Midterm Exam",
                "score": 90
            },
            {
                "examId": "508",
                "examName": "Final Exam",
                "score": 94
            }
        ]
    }
],
"courseMaterials": [
    {
        "materialId": "603",
        "materialName": "Course Syllabus",
        "fileType": "PDF"
    },
    {
        "materialId": "604",
        "materialName": "Coding Examples",
        "fileType": "Code"
    }
],
```

```
    "id": "4a873f83-7bfb-451b-8512-d64323c1f441",
    "_rid": "TTNHAJekiNIBAAAAAAAAAA==",
    "_self": "dbs/TTNHAA==/colls/TTNHAJekiNI=/docs/TTNHAJekiNIBAAAAAAAAAA==/",
    "_etag": "\"080091be-0000-0700-0000-6529a1ab0000\"",
    "_attachments": "attachments/",
    "_ts": 1697227179
},
{
    "courseId": "103",
    "courseName": "Big Data",
    "instructor": {
        "instructorId": "203",
        "instructorName": "Prof. Alice Wright",
        "department": "Computer Science"
    },
    "studentsEnrolled": [
        {
            "studentId": "303",
            "studentName": "Peter",
            "major": "Computer Science",
            "assignments": [
                {
                    "assignmentId": "406",
                    "assignmentName": "Assignment 1",
                    "score": 90
                },
                {
                    "assignmentId": "407",
                    "assignmentName": "Assignment 2",
                    "score": 99
                }
            ],
            "exams": [
                {
                    "examId": "506",
                    "examName": "Midterm Exam",
                    "score": 90
                },
                {
                    "examId": "507",
                    "examName": "Final Exam",
                    "score": 70
                }
            ]
        },
        {
            "studentId": "304",
            "studentName": "Harry Potter",
            "major": "Computer Engineering",
            "assignments": [
                {
                    "assignmentId": "406",
                    "assignmentName": "Assignment 1",
                    "score": 64
```

```
            },
            {
              "assignmentId": "407",
              "assignmentName": "Assignment 2",
              "score": 88
            }
          ],
          "exams": [
            {
              "examId": "506",
              "examName": "Midterm Exam",
              "score": 56
            },
            {
              "examId": "507",
              "examName": "Final Exam",
              "score": 89
            }
          ]
        }
      ],
      "courseMaterials": [
        {
          "materialId": "604",
          "materialName": "Course Syllabus",
          "fileType": "PDF"
        },
        {
          "materialId": "605",
          "materialName": "Coding Examples",
          "fileType": "Code"
        }
      ],
      "id": "ae1018f1-63e0-4f8e-8162-27163d4b5b8c",
      "_rid": "TTNHAJekiNICAAAAAAAAAA==",
      "_self": "dbs/TTNHAA==/colls/TTNHAJekiNI=/docs/TTNHAJekiNICAAAAAAAAAA==/",
      "_etag": "\"080082fd-0000-0700-0000-6529a3b80000\"",
      "_attachments": "attachments/",
      "_ts": 1697227704
    },
    {
      "courseId": "104",
      "courseName": "DSP",
      "instructor": {
        "instructorId": "204",
        "instructorName": "Prof. Minerva McGonnagal",
        "department": "Computer Science"
      },
      "studentsEnrolled": [
        {
          "studentId": "304",
          "studentName": "Draco Malfoy",
          "major": "Computer Science",
          "assignments": [
```

```
            {
                "assignmentId": "407",
                "assignmentName": "Assignment 1",
                "score": 65
            },
            {
                "assignmentId": "408",
                "assignmentName": "Assignment 2",
                "score": 74
            }
        ],
        "exams": [
            {
                "examId": "507",
                "examName": "Midterm Exam",
                "score": 69
            },
            {
                "examId": "508",
                "examName": "Final Exam",
                "score": 80
            }
        ]
    },
    {
        "studentId": "305",
        "studentName": "Luna Lovegood",
        "major": "Computer Science",
        "assignments": [
            {
                "assignmentId": "407",
                "assignmentName": "Assignment 1",
                "score": 80
            },
            {
                "assignmentId": "408",
                "assignmentName": "Assignment 2",
                "score": 89
            }
        ],
        "exams": [
            {
                "examId": "507",
                "examName": "Midterm Exam",
                "score": 77
            },
            {
                "examId": "508",
                "examName": "Final Exam",
                "score": 90
            }
        ]
    }
],
```

```
    "courseMaterials": [
      {
        "materialId": "605",
        "materialName": "Course Syllabus",
        "fileType": "PDF"
      },
      {
        "materialId": "606",
        "materialName": "Coding Examples",
        "fileType": "Code"
      }
    ],
    "id": "4ed89504-79b6-4b25-8feb-fda754297870",
    "_rid": "TTNHAJekiNIDAAAAAAAAAA==",
    "_self": "dbs/TTNHAA==/colls/TTNHAJekiNI=/docs/TTNHAJekiNIDAAAAAAAAAA==/",
    "_etag": "\"0900091e-0000-0700-0000-6529a4c60000\"",
    "_attachments": "attachments/",
    "_ts": 1697227974
  },
  {
    "courseId": "105",
    "courseName": "Data Engineering",
    "instructor": {
      "instructorId": "205",
      "instructorName": "Prof. Severus Snape",
      "department": "Computer Science"
    },
    "studentsEnrolled": [
      {
        "studentId": "305",
        "studentName": "Ron Weasley",
        "major": "Computer Engineering",
        "assignments": [
          {
            "assignmentId": "408",
            "assignmentName": "Assignment 1",
            "score": 88
          },
          {
            "assignmentId": "409",
            "assignmentName": "Assignment 2",
            "score": 78
          }
        ],
        "exams": [
          {
            "examId": "508",
            "examName": "Midterm Exam",
            "score": 90
          },
          {
            "examId": "509",
            "examName": "Final Exam",
            "score": 88
```

```
            }
          ]
        },
        {
          "studentId": "306",
          "studentName": "Hermonie Granger",
          "major": "Computer Engineering",
          "assignments": [
            {
              "assignmentId": "408",
              "assignmentName": "Assignment 1",
              "score": 99
            },
            {
              "assignmentId": "409",
              "assignmentName": "Assignment 2",
              "score": 100
            }
          ],
          "exams": [
            {
              "examId": "508",
              "examName": "Midterm Exam",
              "score": 100
            },
            {
              "examId": "509",
              "examName": "Final Exam",
              "score": 100
            }
          ]
        }
      ],
      "courseMaterials": [
        {
          "materialId": "606",
          "materialName": "Course Syllabus",
          "fileType": "PDF"
        },
        {
          "materialId": "609",
          "materialName": "Coding Examples",
          "fileType": "Code"
        }
      ],
      "id": "7827ccf2-c6fe-4cf1-b54f-ee3e3533fbd9",
      "_rid": "TTNHAJekiNIEAAAAAAAAAA==",
      "_self": "dbs/TTNHAA==/colls/TTNHAJekiNI=/docs/TTNHAJekiNIEAAAAAAAAAA==/",
      "_etag": "\"09009038-0000-0700-0000-6529a59e0000\"",
      "_attachments": "attachments/",
      "_ts": 1697228190
    },
    {
      "courseId": "106",
```

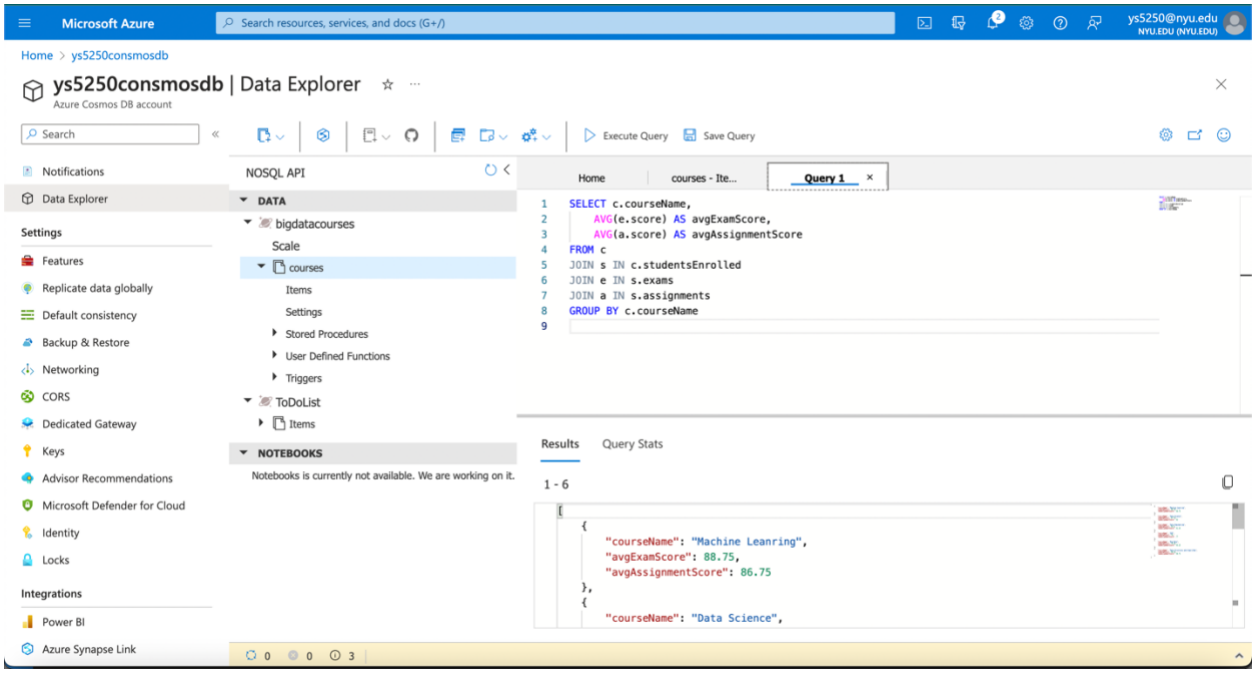
```
"courseName": "Data Science",
"instructor": {
    "instructorId": "206",
    "instructorName": "Prof. Horace Slughorn",
    "department": "Computer Science"
},
"studentsEnrolled": [
    {
        "studentId": "307",
        "studentName": "Neville Longbottom",
        "major": "Computer Science",
        "assignments": [
            {
                "assignmentId": "409",
                "assignmentName": "Assignment 1",
                "score": 78
            },
            {
                "assignmentId": "410",
                "assignmentName": "Assignment 2",
                "score": 83
            }
        ],
        "exams": [
            {
                "examId": "509",
                "examName": "Midterm Exam",
                "score": 88
            },
            {
                "examId": "510",
                "examName": "Final Exam",
                "score": 98
            }
        ]
    },
    {
        "studentId": "308",
        "studentName": "Ginny Weasley",
        "major": "Computer Science",
        "assignments": [
            {
                "assignmentId": "409",
                "assignmentName": "Assignment 1",
                "score": 76
            },
            {
                "assignmentId": "410",
                "assignmentName": "Assignment 2",
                "score": 91
            }
        ],
        "exams": [
            {
```

```
                "examId": "509",
                "examName": "Midterm Exam",
                "score": 89
            },
            {
                "examId": "510",
                "examName": "Final Exam",
                "score": 76
            }
        ]
    }
  ],
  "courseMaterials": [
    {
      "materialId": "607",
      "materialName": "Course Syllabus",
      "fileType": "PDF"
    },
    {
      "materialId": "610",
      "materialName": "Coding Examples",
      "fileType": "Code"
    }
  ],
  "id": "56dd01e1-fa8e-4d9c-aa6f-c3f7baebd13e",
  "_rid": "TTNHAJekiNIFAAAAAAAAAA==",
  "_self": "dbs/TTNHAA==/colls/TTNHAJekiNI=/docs/TTNHAJekiNIFAAAAAAAAAA==/",
  "_etag": "\"0900996f-0000-0700-0000-6529a7720000\"",
  "_attachments": "attachments/",
  "_ts": 1697228658
},
{
  "courseId": "107",
  "courseName": "Machine Leanring",
  "instructor": {
    "instructorId": "207",
    "instructorName": "Prof. Rubeus Hagrid",
    "department": "Computer Science"
  },
  "studentsEnrolled": [
    {
      "studentId": "309",
      "studentName": "Priya Patel",
      "major": "Computer Science",
      "assignments": [
        {
          "assignmentId": "410",
          "assignmentName": "Assignment 1",
          "score": 79
        },
        {
          "assignmentId": "411",
          "assignmentName": "Assignment 2",
          "score": 73
```

```
            }
        ],
        "exams": [
            {
                "examId": "510",
                "examName": "Midterm Exam",
                "score": 89
            },
            {
                "examId": "511",
                "examName": "Final Exam",
                "score": 77
            }
        ]
    },
    {
        "studentId": "310",
        "studentName": "Cho Champbell",
        "major": "Computer Science",
        "assignments": [
            {
                "assignmentId": "410",
                "assignmentName": "Assignment 1",
                "score": 99
            },
            {
                "assignmentId": "411",
                "assignmentName": "Assignment 2",
                "score": 96
            }
        ],
        "exams": [
            {
                "examId": "510",
                "examName": "Midterm Exam",
                "score": 90
            },
            {
                "examId": "511",
                "examName": "Final Exam",
                "score": 99
            }
        ]
    }
],
"courseMaterials": [
    {
        "materialId": "608",
        "materialName": "Course Syllabus",
        "fileType": "PDF"
    },
    {
        "materialId": "611",
        "materialName": "Coding Examples",
```

```
            "fileType": "Code"
          }
        ],
        "id": "f511c9d9-bd6d-4973-bfb6-11c8c277cc24",
        "_rid": "TTNHAJekiNIGAAAAAAAAAA==",
        "_self": "dbs/TTNHAA==/colls/TTNHAJekiNI=/docs/TTNHAJekiNIGAAAAAAAAAA==/",
        "_etag": "\"0900b489-0000-0700-0000-6529a84e0000\"",
        "_attachments": "attachments/",
        "_ts": 1697228878
    }
]
```

1. Fetch the course name and the average exam and average assignment score for all courses.

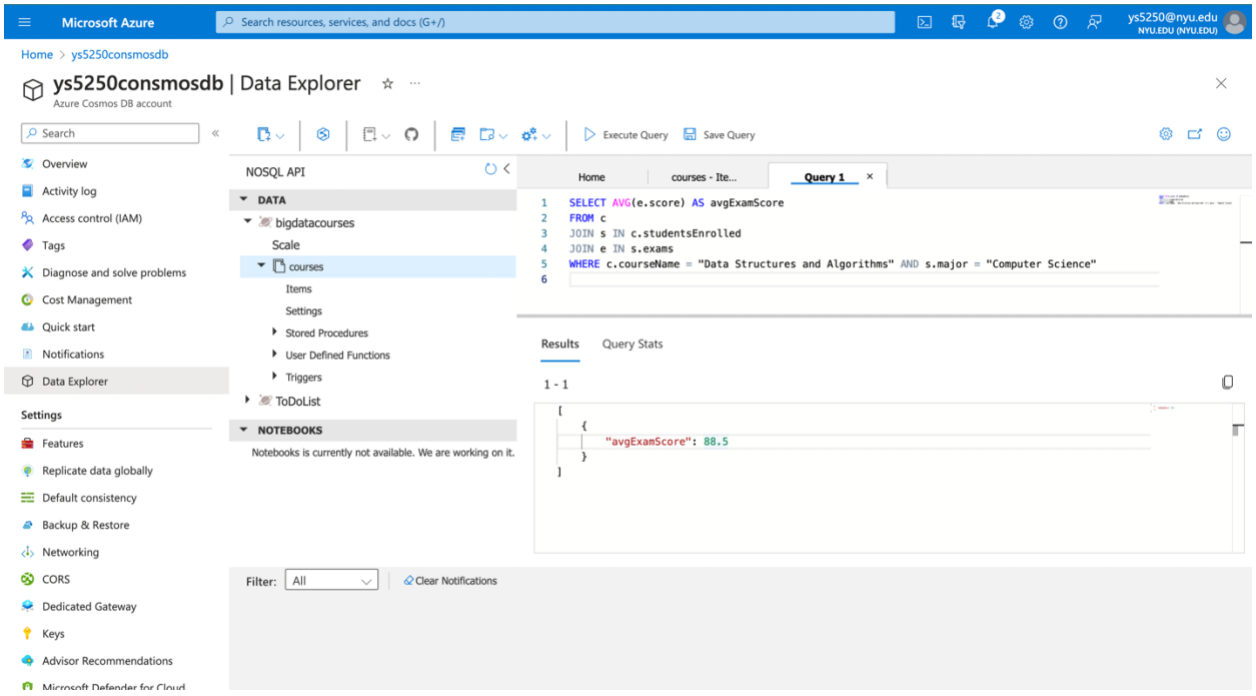


Output:


```
[
  {
    "courseName": "Machine Leanring",
    "avgExamScore": 88.75,
    "avgAssignmentScore": 86.75
  },
  {
    "courseName": "Data Science",
    "avgExamScore": 87.75,
    "avgAssignmentScore": 82
  },
  {
    "courseName": "Data Engineering",
    "avgExamScore": 94.5,
    "avgAssignmentScore": 91.25
  },
  {
    "courseName": "DSP",
    "avgExamScore": 79,
    "avgAssignmentScore": 77
  },
  {
    "courseName": "Big Data",
    "avgExamScore": 76.25,
```

```
    "avgAssignmentScore": 85.25
},
{
    "courseName": "Data Structures and Algorithms",
    "avgExamScore": 90.25,
    "avgAssignmentScore": 90.75
}
  ]
```

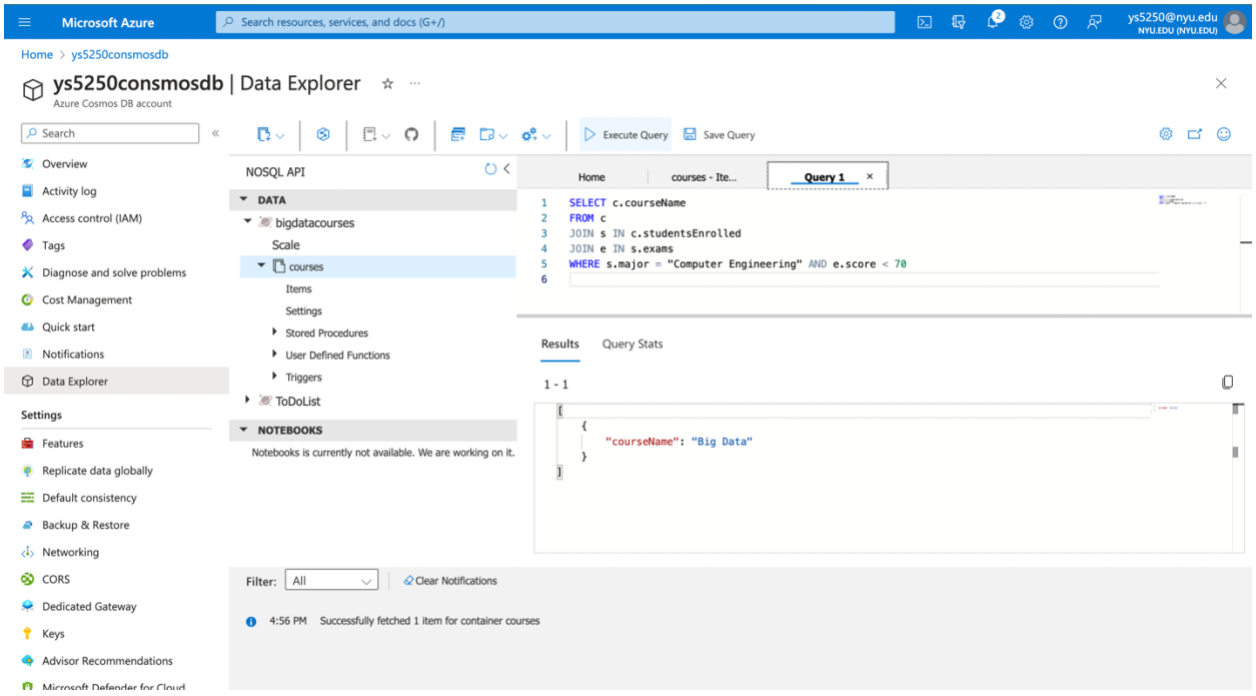
2. Calculate the average exam score for students majoring in "Computer Science" who are enrolled in the "Data Structures and Algorithms" course.



Output:


```
[
  {
    "avgExamScore": 88.5
  }
    ]
```

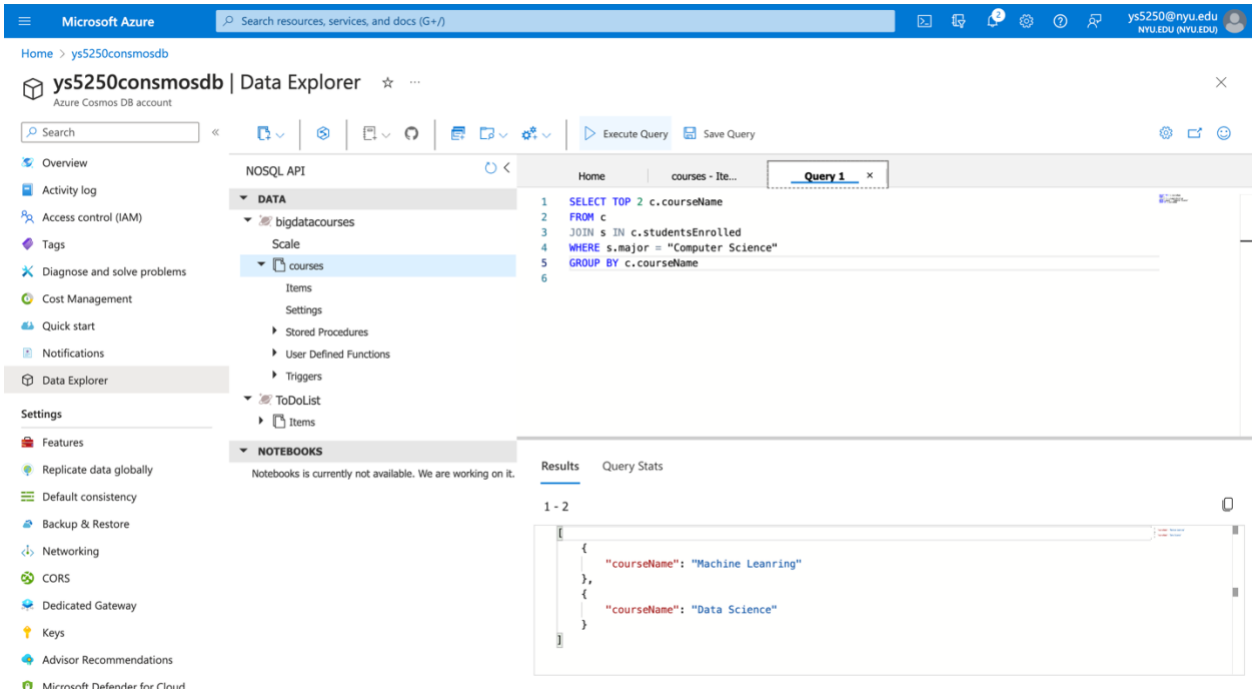
3. Find the courses where at least one student's major is "Computer Engineering" and they scored below 70 on exams.



Output:


```
[
  {
     "courseName": "Big Data"
  }
    ]
```

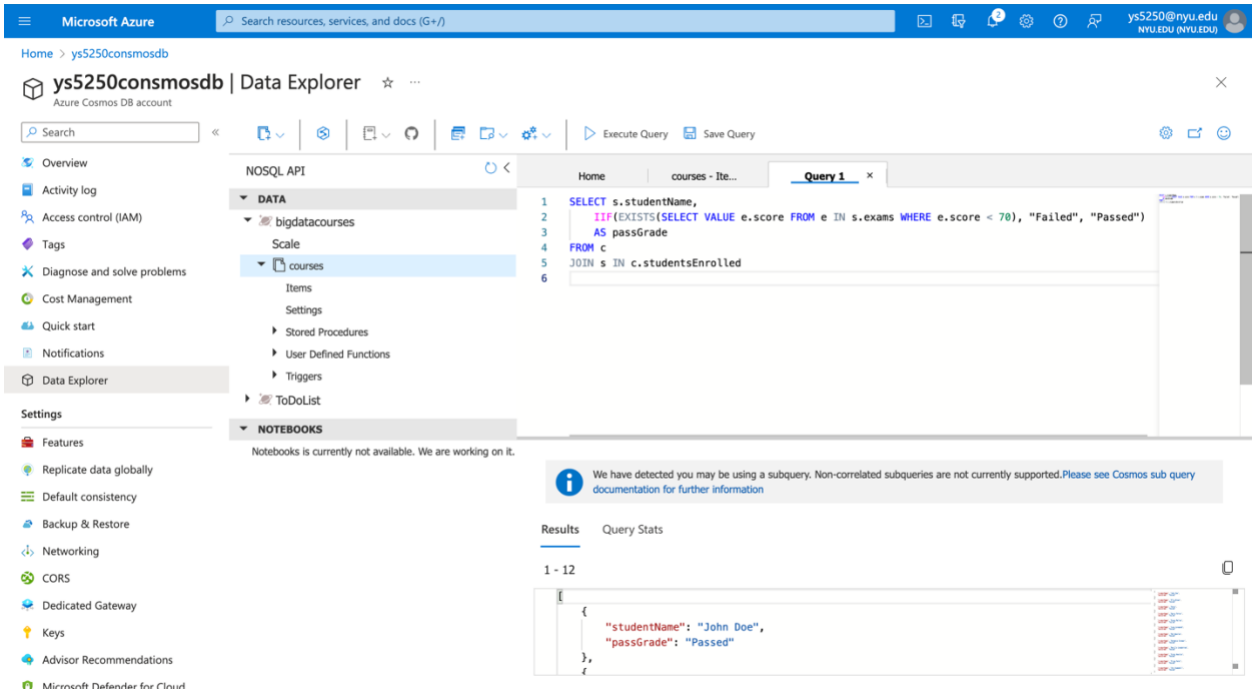
4. Select any two courses where there is at least one student's major that is "Computer Science" (There must be at least three students in different courses, who's major is Computer Science)



Output:


```
[
  {
    "courseName": "Machine Leanring"
  },
  {
    "courseName": "Data Science"
  }
    ]
```

5. Calculate the overall pass grade (if they passed or failed) for each student, considering a passing grade as an exam score (exams) of 70 or higher. (Make sure at least one student fails in your database)



Output:


```
[
  {
    "studentName": "John Doe",
    "passGrade": "Passed"
  },
  {
    "studentName": "Ella Brown",
    "passGrade": "Passed"
  },
  {
    "studentName": "Peter",
    "passGrade": "Passed"
  },
  {
    "studentName": "Harry Potter",
    "passGrade": "Failed"
  },
  {
    "studentName": "Draco Malfoy",
    "passGrade": "Failed"
  },
  {
```

```
        "studentName": "Luna Lovegood",
        "passGrade": "Passed"
    },
    {
        "studentName": "Ron Weasley",
        "passGrade": "Passed"
    },
    {
        "studentName": "Hermonie Granger",
        "passGrade": "Passed"
    },
    {
        "studentName": "Neville Longbottom",
        "passGrade": "Passed"
    },
    {
        "studentName": "Ginny Weasley",
        "passGrade": "Passed"
    },
    {
        "studentName": "Priya Patel",
        "passGrade": "Passed"
    },
    {
        "studentName": "Cho Champbell",
        "passGrade": "Passed"
    }
]
```

Learning:

In this assignment I've gained valuable insights into the world of data modeling and querying using Azure Cosmos DB. The task revolved around creating a structured database for educational courses, instructors, and students, showcasing the importance of sound data modeling practices. I learned how to design a NoSQL database schema that accommodates complex relationships within the data. The assignment further improved my proficiency in crafting Cosmos DB queries to extract meaningful information from the dataset, including calculating averages, filtering data, and assessing pass/fail status based on exam scores. Understanding the data import and export process, involving Azure services like Azure Data Factory, proved invaluable for potential real-world data migration and analysis tasks.