A. MongoDB

1. The Data Model: Query Analysis s

Query	1. Which major has the most students retained after the first year?
Collection	Major, Student, Enrollment
Relationship	• Embed major in student • Link student and enrollment • Duplicate major.name in Enrollment
Indexes	• Major.name
Precompute	• None

Query	2. What course has the highest percent of people with grade W?
Collection	Course, Student, Enrollment
Relationship	• Link everything
Indexes	• Enrollment.Grade
Precompute	• none

Query	3. What's a 4-year graduation rate among the female students, grouped by majors?
Collection	Major, Student, Graduation
Relationship	• Embed major in student • Embed graduation in student
Indexes	• Graduation.year
Precompute	• None
Input	• LateDate

Query	4. Rank all courses in terms of their popularity (cumulated number enrollments).
Collection	Course, Enrollment
Relationship	• Link • Duplicate course.name in enrollment
Indexes	• Course.name in enrollment
Precompute	• none

Query	5. What is the trend of enrollment over the years?
Collection	Enrollment
Relationship	• none

Indexes	• year
Precompute	• none

Query	6. How does the number of full-time students vs part-time students compare for both graduate and undergraduate students?
Collection	Student, Enrollment, Course
Relationship	• link everything • duplicate course.credit in enrollment
Indexes	• default
Precompute	• none
Input	• EnrollmentYear

Query	7. What are the percentages of female vs male students in STEM programs?
Collection	Student, Major
Relationship	• Embed major in student
Indexes	• Major.is_stem
Precompute	• Is_stem
Input	• ListSTEM

Query	8. What is the distribution of students with respect to their race?
Collection	Student
Relationship	• none
Indexes	• student.race
Precompute	• none

Query	9. What is the percentage of enrollment per state in the USA?
Collection	Student, Enrollment, Address
Relationship	• Embed address in student • Link student to enrollment • Duplicate address.state and address.country to enrollment
Indexes	• Enrollment-address.country • Enrollment-address.state
Precompute	• none

Query	10. Which instructor is most "effective" in the long run? You can define "effective" as some sort of complex measure of students' performance in instructors' classes. – Average(80%*score20%*absenceRate)
Collection	Enrollment
Relationship	• Link Instructor and Enrollment • Duplicate Instructor.FN and Instructor.LN in Enrollment
Indexes	• Enrollment.Instructor.ID
Precompute	• Effectiveness = Average(80%*enrollment.score20%*enrollment.absenceRate)

Query	11. What is the most popular course (based on enrollment count) With respect to the country of origin?
Collection	Student, Enrollment, Course, Address
Relationship	• Embed address in student • Link student and course to enrollment • Duplicate address.country and course.name to enrollment
Indexes	• Enrollment-address.country • Enrollment-course.name
Precompute	• none