Team4-DummyHead-Database Design

1. Main Tables

Our team chooses Table User, Student, Teacher, Reservation, Message, Court, Sport, Hobby as the main tables.

2.DDL Commands for Tables

```
CREATE TABLE User (
Userld VARCHAR(50) PRIMARY KEY,
FirstName VARCHAR(20) NOT NULL,
LastName VARCHAR(20),
Gender VARCHAR(20),
Age REAL,
Email VARCHAR(50) NOT NULL,
Password VARCHAR(100) NOT NULL,
Phone VARCHAR(20),
Location VARCHAR(200),
Type ENUM('S', 'T')
);
CREATE TABLE Student (
Studentid VARCHAR(50) REFERENCES User ON DELETE CASCADE,
Major VARCHAR(20),
PRIMARY KEY (StudentId)
);
CREATE TABLE Teacher (
Teacherld VARCHAR(50) REFERENCES User ON DELETE CASCADE,
Rating REAL,
YearofTeaching REAL NOT NULL,
```

```
PRIMARY KEY (TeacherId)
);
CREATE TABLE Sport(
SportId VARCHAR(50) PRIMARY KEY,
SportName VARCHAR(50) NOT NULL,
SportIntroduction VARCHAR(1000) NOT NULL,
SportImage VARCHAR(500)
);
CREATE TABLE Course (
Courseld VARCHAR(50) PRIMARY KEY,
Name VARCHAR(20) NOT NULL,
Location VARCHAR(200),
StartDate DATE NOT NULL,
EndDate DATE NOT NULL,
CourseType ENUM('Online', 'In-person'),
SportId VARCHAR(50) REFERENCES Sport ON DELETE CASCADE,
TeacherId VARCHAR(50) REFERENCES Teacher ON DELETE CASCADE
);
CREATE TABLE Message(
Messageld VARCHAR(50),
LaunchTime DATETIME,
Title VARCHAR(50) NOT NULL,
Content VARCHAR(2000) NOT NULL,
StudentId VARCHAR(50) REFERENCES Student ON DELETE CASCADE,
PRIMARY KEY (Messageld, StudentId)
);
CREATE TABLE Court(
Courtld VARCHAR(50) PRIMARY KEY,
Name VARCHAR(50),
Location VARCHAR(200) NOT NULL
SportId VARCHAR(50) REFERENCES Sport ON DELETE CASCADE
);
CREATE TABLE Reservation(
ReservationId VARCHAR(50) PRIMARY KEY,
```

CourtId VARCHAR(50) REFERENCES Court ON DELETE CASCADE,

```
StudentId VARCHAR(50) REFERENCES Student ON DELETE CASCADE,
BeginTime DATETIME NOT NULL,
EndTime DATETIME NOT NULL
);
CREATE TABLE Appointment (
AppointmentId VARCHAR(50) PRIMARY KEY,
StudentId VARCHAR(50) REFERENCES Student ON DELETE CASCADE,
TeacherId VARCHAR(50) REFERENCES Teacher ON DELETE CASCADE,
Link VARCHAR(200),
ReservationId VARCHAR(50) REFERENCES Reservation ON DELETE CASCADE,
AppointmentType ENUM('Online', 'In-person'),
Comment VARCHAR(2000)
);
CREATE TABLE Enrollment (
StudentId VARCHAR(50) REFERENCES Student ON DELETE CASCADE,
Courseld VARCHAR(50) REFERENCES Course ON DELETE CASCADE,
PRIMARY KEY (StudentId, CourseId)
);
CREATE TABLE Rate (
StudentId VARCHAR(50) REFERENCES Student ON DELETE CASCADE,
Courseld VARCHAR(50) REFERENCES Course ON DELETE CASCADE,
Rating REAL,
Review VARCHAR(2000),
PRIMARY KEY (StudentId, CourseId)
);
CREATE TABLE Hobby (
StudentId VARCHAR(50) REFERENCES Student ON DELETE CASCADE,
SportId VARCHAR(50) REFERENCES Sport ON DELETE CASCADE,
Year REAL.
PRIMARY KEY (StudentId, SportId)
);
```

3. Screenshots of Data Counts in Main Tables

```
mysql> SELECT COUNT(*) FROM User;
| COUNT(*) |
3000 |
1 row in set (0.01 sec)
mysql> SELECT COUNT(*) FROM Student;
[+----+
| COUNT(*) |
| 1500 |
1 row in set (0.00 sec)
mysql> SELECT COUNT(*) FROM Teacher;
| COUNT(*) |
| 1500 |
[+----+
1 row in set (0.00 sec)
mysql> SELECT COUNT(*) FROM Hobby;
+----+
| COUNT(*) |
| 1007 |
[+----+
1 row in set (0.00 sec)
mysql> SELECT COUNT(*) FROM Reservation;
| COUNT(*) |
    2000 |
1 row in set (0.00 sec)
mysql> SELECT COUNT(*) FROM Message;
+----+
| COUNT(*) |
+----+
2000 |
[+----+
1 row in set (0.00 sec)
mysql> SELECT COUNT(*) FROM Sport;
| COUNT(*) |
| 46 |
1 row in set (0.00 sec)
mysql> SELECT COUNT(*) FROM Court;
| COUNT(*) |
38 |
1 row in set (0.00 sec)
```

4. Advanced SQL Queries

(1) Find top 15 users who share most number of sport hobbies with the user whose id is 1039.

```
SELECT h.StudentId, u.FirstName, u.LastName, COUNT(*) AS MatchNum
FROM Hobby h JOIN User u on h.StudentId = u.UserId
WHERE h.StudentId <> '1039'
AND h.SportId IN (SELECT SportId
FROM Hobby
WHERE StudentId = '1039')
GROUP BY h.StudentId, u.FirstName, u.LastName
ORDER BY COUNT(h.SportId) DESC
LIMIT 15
```

(2) For each sport, find top 2 users who reserve the court related to this sport most frequently.

```
With newTable AS
(SELECT UserId, SportId,
row_number() OVER (PARTITION BY SportId ORDER BY cnt DESC) AS rk
FROM
(SELECT r.UserId, c.SportId, COUNT(r.ReservationId) AS cnt
FROM Reservation AS r NATURAL JOIN Court AS c
GROUP BY r.UserId, c.SportId) AS a)
SELECT s.SportName, u.FirstName, u.LastName
FROM newTable AS n NATURAL JOIN User AS u NATURAL JOIN Sport AS s
WHERE n.rk <= 2
ORDER BY s.SportName ASC
LIMIT 15;</pre>
```

5. Screenshots of Top 15 Rows

(1) Top 15 users who share most number of sport hobbies with user whose id is 1039.

	StudentId	FirstName	LastName	MatchNum
Þ	454	Dorothy	Betancourt	4
	801	Olivia	Campbell	2
	531	Diane	Campbell	2
	433	Danielle	Snider	2
	773	Gene	Nichols	2
	312	Julie	Davis	2
	1428	George	Kujawa	2
	1300	Terry	Franklin	2
	1027	Jan	Jackson	1
	1058	George	Gable	1
	1060	Anthony	Bethune	1
	997	Charles	Dean	1
	1089	Linda	Gay	1
	1093	Thuy	Long	1
	1113	Sharon	Johnson	1

(2) Top 2 users who reserve the court related to each sport most frequently.

	SportName	FirstName	LastName
⊳	American Football	Hector	Mares
	American Football	Cathy	King
	Badminton	Kimberly	Crabtree
	Badminton	Bertha	Black
	Basketball	Antonio	Jensen
	Basketball	Jerome	McCoy
	Boxing	Susan	Helmer
	Boxing	Michael	Dobson
	Cycling	Hillary	Johnson
	Cycling	Daniel	Price
	Figure Skating	Larry	Ashley
	Figure Skating	Joe	Ward
	Football	Melvin	Crosby
	Football	Ernesto	Bolden
	Squash	Larry	Williams

6. Trying at Least 3 Indices on Each Advanced Query

Query 1 - No index

Query 1 - Index On User.FirstName

```
| -> Limit: 15 row(s) (actual time=7.669..7.672 rows=15 loops=1)
    -> Sort: `count(h.SportId)` DESC, limit input to 15 row(s) per chunk (actual time=7.668..7.670 rows=15 loops=1)
    -> Stream results (cost=1005.83 rows=1005) (actual time=0.253..7.564 rows=233 loops=1)
    -> Group aggregate: count(h.SportId), count(e) (cost=1005.83 rows=1005) (actual time=0.247..7.356 rows=233 loops=1)
    -> Nested loop inner join (cost=905.33 rows=1005) (actual time=0.113..7.090 rows=243 loops=1)
    -> Nested loop inner join (cost=005.35.08 rows=1005) (actual time=0.113..7.090 rows=243 loops=1)
    -> Filter: (h.StudentId <> '1039') (cost=201.83 rows=1005) (actual time=0.061..0.983 rows=1005 loops=1)
    -> Index range scan on h using PRIMARY (cost=201.83 rows=1005) (actual time=0.061..0.983 rows=1005 loops=1)
    -> Single-row index lookup on hobby using PRIMARY (UserId=h.StudentId) (cost=0.25 rows=1) (actual time=0.004..0.007 rows=1 loops=243)
```

Query 1 - Index On User.LastName

```
| -> Limit: 15 row(s) (actual time=7.773..7.776 rows=15 loops=1)
| -> Sort: `count(h.SportId)` DESC, limit input to 15 row(s) per chunk (actual time=7.772..7.774 rows=15 loops=1)
| -> Stream results (cost=1005.83 rows=1005) (actual time=0.329..7.646 rows=223 loops=1)
| -> Group aggregate: count(h.SportId), count(e) (cost=1005.83 rows=1005) (actual time=0.323..7.457 rows=233 loops=1)
| -> Nested loop inner join (cost=905.33 rows=1005) (actual time=0.186..7.200 rows=243 loops=1)
| -> Nested loop inner join (cost=505..58 rows=1005) (actual time=0.186..7.200 rows=243 loops=1)
| -> Filter: (h.StudentId <> '1039') (cost=201.83 rows=1005) (actual time=0.057..0.904 rows=1005 loops=1)
| -> Index range scan on h using PRIMARY (cost=201.83 rows=1005) (actual time=0.055..0.717 rows=1005 loops=1)
| -> Single-row index lookup on Hobby using PRIMARY (StudentId='1039', SportId-h.SportId) (cost=0.25 rows=1) (actual time=0.004..0.004 rows=0 loops=1005)
| -> Single-row index lookup on u using PRIMARY (UserId=h.StudentId) (cost=0.25 rows=1) (actual time=0.007..0.007 rows=1 loops=243)
```

Query 1 - Index On User.FirstName & User.LastName

```
| -> Limit: 15 row(s) (actual time=5.170..5.173 rows=15 loops=1)
    -> Sort: 'count(h.SportId)' DESC, limit input to 15 row(s) per chunk (actual time=5.169..5.171 rows=15 loops=1)
    -> Stream results (cost=1005.83 rows=1005) (actual time=0.168..5.081 rows=233 loops=1)
    -> Group aggregate: count(h.SportId), count(e) (cost=1006.83 rows=1005) (actual time=0.162..4.949 rows=233 loops=1)
    -> Nested loop inner join (cost=5065.33 rows=1005) (actual time=0.082..4.773 rows=243 loops=1)
    -> Nested loop inner join (cost=503.56 rows=1005) (actual time=0.082..4.773 rows=243 loops=1)
    -> Filter: (h.StudentId <> '1039') (cost=201.83 rows=1005) (actual time=0.049..0.650 rows=1005 loops=1)
    -> Index range scan on h using PRIMARY (cost=201.83 rows=1005) (actual time=0.047..0.469 rows=1005 loops=1)
    -> Single=row index lookup on hobby using PRIMARY (UserId=h.StudentId) (cost=0.25 rows=1) (actual time=0.005..0.005 rows=1 loops=243)
```

```
| -> Limit: 15 row(s) (cost=653.81 rows=15) (actual time=14.457..14.815 rows=15 loops=1)
    -> Nested loop inner join (cost=653.81 rows=463) (actual time=14.455..14.812 rows=15 loops=1)
    -> Nested loop inner join (cost=653.81 rows=463) (actual time=14.429..14.786 rows=15 loops=1)
    -> Sort: s.SportName (cost=6.58 rows=646) (actual time=6.172.a.017 rows=88 loops=1)
    -> Filter: (in.tk < 2) and (n.UserId is not null)) (cost=6.75,7.57 rows=10) (actual time=6.376..0.833 rows=0 loops=38)
    -> Index lookup on nu using cauto, key? (SportIda catual time=6.18.0.005 rows=20 loops=38)
    -> Materialize CTE newtable (cost=0.00.0.00 rows=0) (actual time=14.289..14.496 rows=1820 loops=1)
    -> Sort: a.SportId, a.cnt DESC (actual time=1.556..1.704 rows=1820 loops=1)
    -> Table scan on a (cost=227.61 rows=2001) (actual time=6.10..0.181 rows=1820 loops=1)
    -> Table scan on a (cost=227.61 rows=2001) (actual time=6.883..0.154 rows=1820 loops=1)
    -> Filter: (r.CourtId is not null) (cost=902.70 rows=2001) (actual time=0.697..1.520 rows=2000 loops=1)
    -> Filter: (r.CourtId is not null) (cost=902.70 rows=2001) (actual time=0.697..1.520 rows=2000 loops=1)
    -> Filter: (r.CourtId is not null) (cost=902.70 rows=2001) (actual time=0.697..1.520 rows=2000 loops=1)
    -> Filter: (r.CourtId is not null) (cost=902.70 rows=2001) (actual time=0.697..1.520 rows=2000 loops=1)
    -> Filter: (r.CourtId is not null) (cost=902.70 rows=2001) (actual time=0.697..1.520 rows=2000 loops=1)
    -> Filter: (r.CourtId is not null) (cost=902.70 rows=2001) (actual time=0.697..1.520 rows=2000 loops=1)
    -> Filter: (r.CourtId is not null) (cost=902.70 rows=2001) (actual time=0.697..1.520 rows=2000 loops=1)
    -> Filter: (r.CourtId is not null) (cost=902.70 rows=2001) (actual time=0.697..1.520 rows=2000 loops=1)
    -> Filter: (r.CourtId is not null) (cost=902.70 rows=2001) (actual time=0.697..1.520 rows=2000 loops=1)
    -> Filter: (r.CourtId is not null) (cost=902.70 rows=2001) (actual time=0.697..1.520 rows=2000 loops=1)
    -> Fi
```

Query 2 - Index on Reservation. UserId

```
|-> Limit: 15 row(s) (cost=653.81 rows=15) (actual time=14.082..14.514 rows=15 loops=1)
-> Nested loop inner join (cost=653.81 rows=63) (actual time=14.082..14.511 rows=15 loops=1)
-> Nested loop inner join (cost=653.81 rows=643) (actual time=14.065..14.463 rows=15 loops=1)
-> Sort: s. SportName (cost=6.35 rows=64) (actual time=0.198..0.2 kors=38) loops=1)
-> Table scan on s (cost=5.35 rows=64) (actual time=0.897..0.124 rows=46 loops=1)
-> Filter: ((n.rk <= 2) and (n.Userid is not null)) (cost=0.75..7.57 rows=10) (actual time=0.806..0.375 rows=0 loops=38)
-> Index lookup on u using vato_key2* (SportId=s.SportId) (actual time=0.801..0.806 rows=26 loops=38)
-> Naterialize CTE newtable (cost=0.00..000 rows=0) (actual time=13.908..14.159 rows=1820 loops=1)
-> Window aggregate: row_number() OVER (PRANTITION BY a.SportId ORDER BY a.cnt desc) (actual time=10.675..11.743 rows=1820 loops=1)
-> Sort: a.SportId, a.cnt DESC (actual time=1.513..1.680 rows=1820 loops=1)
-> Table scan on a (cost=227.61 rows=2001) (actual time=0.001..0.157 rows=1820 loops=1)
-> Aggregate using temporary (actual time=0.001..0.155 rows=1820 loops=1)
-> Palse scan on a (cost=272.05 rows=002.70 rows=2002.70 rows=2001) (actual time=0.083..1.425 rows=2000 loops=1)
-> Filter: (r.CourtId is not null) (cost=202.35 rows=2001) (actual time=0.063..1.425 rows=2000 loops=1)
-> Table scan on r (cost=0.202..35 rows=2001) (actual time=0.061..1.218 rows=2000 loops=1)
-> Filter: (r.CourtId is not null) (cost=202.35 rows=2001) (actual time=0.061..1218 rows=2000 loops=1)
-> Single-row index lookup on c using PRIMARY (CourtId=r.CourtId) (cost=0.25 rows=1) (actual time=0.001..0.001 rows=1 loops=2000)
```

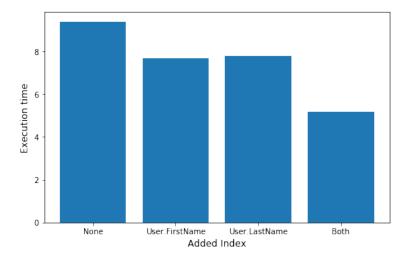
Query 2 - Index on Court.SportId

```
| -> Limit: 15 row(s) (cost=653.81 rows=15) (actual time=13.880..14.236 rows=15 loops=1)
| -> Nested loop inner join (cost=653.81 rows=643) (actual time=13.878..14.293 rows=15 loops=1)
| -> Nested loop inner join (cost=69.78 rows=64) (actual time=13.878..14.198 rows=15 loops=1)
| -> Sort: s.SportName (cost=5.35 rows=64) (actual time=2.878..14.198 rows=16 loops=1)
| -> Table scan on s (cost=5.35 rows=64) (actual time=6.891..0.115 rows=46 loops=1)
| -> Filter: ((n.rk <= 2) and (n.UserId is not null) (cost=0.75..7.57 rows=16) (actual time=0.361..0.368 rows=0 loops=38)
| -> Index lookup on u using quato_key2 (SportId=s.SportId) (actual time=0.805..0.805 rows=26 loops=38)
| -> Materialize CITe newtable (cost=0.80..0.80 rows=6) (actual time=1.698.rows=1828 loops=1)
| -> Validow aggregate: row,number() OVER (PARTITION BY a.SportId QRDER BY a.cnt desc) (actual time=10.836..11.788 rows=1820 loops=1)
| -> Table scan on a (cost=227.61 rows=2801) (actual time=0.818.01.151 rows=1820 loops=1)
| -> Table scan on a (cost=227.61 rows=2801) (actual time=0.812..11.093 rows=1828 loops=1)
| -> Table scan on a (temporary catual time=0.801..0.156 rows=1820 loops=1)
| -> Nested loop inner join (cost=20.276 rows=20.276) (actual time=0.856..826 rows=1820 loops=1)
| -> Nested loop inner join (cost=20.276 rows=20.210 (actual time=0.855..1.467 rows=2000 loops=1)
| -> Table scan on remporary (actual time=0.820..276 rows=2001) (actual time=0.855..1.246 rows=2000 loops=1)
| -> Filter: (r.CourtId is not null) (cost=20.235 rows=2001) (actual time=0.855..1.246 rows=2000 loops=1)
| -> Table scan on rows=2000 rows=2001 (actual time=0.855..1.246 rows=2000 loops=1)
| -> Table scan on rows=2000 rows=2001 (actual time=0.855..1.246 rows=2000 loops=1)
| -> Single-row index lookup on cusing PRIMARY (CourtId=r.CourtId) (cost=0.82.5 rows=1) (actual time=0.855.1.246 rows=2000 loops=1)
```

Query 2 - Index on Reservation. UserId and Court. SportId

```
| -> Limit: 15 row(s) (cost=653.81 rows=15) (actual time=13.264..13.633 rows=15 loops=1)
-> Nested loop inner join (cost=653.81 rows=643) (actual time=13.263..13.631 rows=15 loops=1)
-> Nested loop inner join (cost=653.81 rows=643) (actual time=13.262..13.572 rows=15 loops=1)
-> Sort: s.SportName (cost=5.35 rows=46) (actual time=0.233..0.28 rows=38 loops=1)
-> Table scan on s (cost=5.35 rows=46) (actual time=0.822..0.186 rows=46 loops=1)
-> Filter: ((n.rk <= 2) and (n.UserId is not null)) (cost=0.75.7 rows=10) (actual time=0.842..0.186 rows=26 loops=38)
-> Index lookup on n using <auto_key2 (SportId=s.SportId) (actual time=0.001.0.005 rows=26 loops=38)
-> Naterialize CTE newtable (cost=0.00..00 rows=0) (actual time=0.001.0.005 rows=120 loops=1)
-> Sort: a.SportId, a.cnt DESC (actual time=1.537..1.681 rows=1820 loops=1)
-> Table scan on a (cost=227.61 rows=2001) (actual time=0.001.0.165 rows=1820 loops=1)
-> Aggregate using temporary (actual time=0.002..10.67 orows=1820 loops=1)
-> Naterialize (cost=0.00..0.00 rows=0) (actual time=0.002..0.167 rows=1820 loops=1)
-> Aggregate using temporary table (actual time=0.002..0.17 rows=2001) (actual time=0.002..1.71 rows=2001) (actual time=0.002..1.71 rows=2001) (actual time=0.002..1.71 rows=2000 loops=1)
-> Filter: (r.CourtId is not null) (cost=202.35 rows=2001) (actual time=0.062..1.71 rows=2000 loops=1)
-> Table scan on c (cost=202.35 rows=2001) (actual time=0.062..1.71 rows=2000 loops=1)
-> Filter: (r.CourtId is not null) (cost=202.35 rows=2001) (actual time=0.062..1.71 rows=2000 loops=1)
-> Single-row index lookup on u using PRIMARY (CourtId=r.CourtId) (cost=0.25 rows=1) (actual time=0.001..0.004 rows=1 loops=15)
```

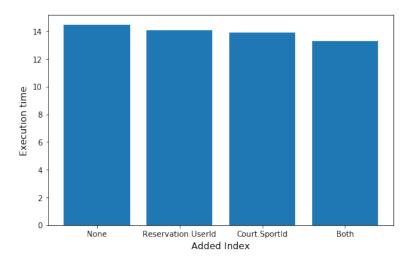
7. Indexing Analysis Report



Index candidates for this advanced query are set to be User.FirstName, User.LastName and both. We set these indices because our advanced query uses these two attributes to aggregate results and they might cost much time.

By observing execution time shown above, we find that adding more indices makes the whole query run more efficiently, and the improvement in time is significant. Execution time for using only User.FirstName or User.LastName are similar and it drops by 2 seconds from not using any index. Using both indices makes query execute most efficiently and its running time drops by 4 seconds from not using any index. Therefore, adding indices are good practices for this advanced query. We believe if we insert more data into schemas, adding indices of User.FirstName and User.LastName would improve execution time more significantly.

Query 2



In query 2, the Userld column of Reservation table and the Sportld column of Court table are used in the groupby statement, which means they need to be compared with each other, so we decide to use them two as the indexes. The bar chart above clearly shows that adding any index shortens the runtime than using the original primary key index. As a result, both of the indexes can be used to minimize the runtime and gain performance improvements.