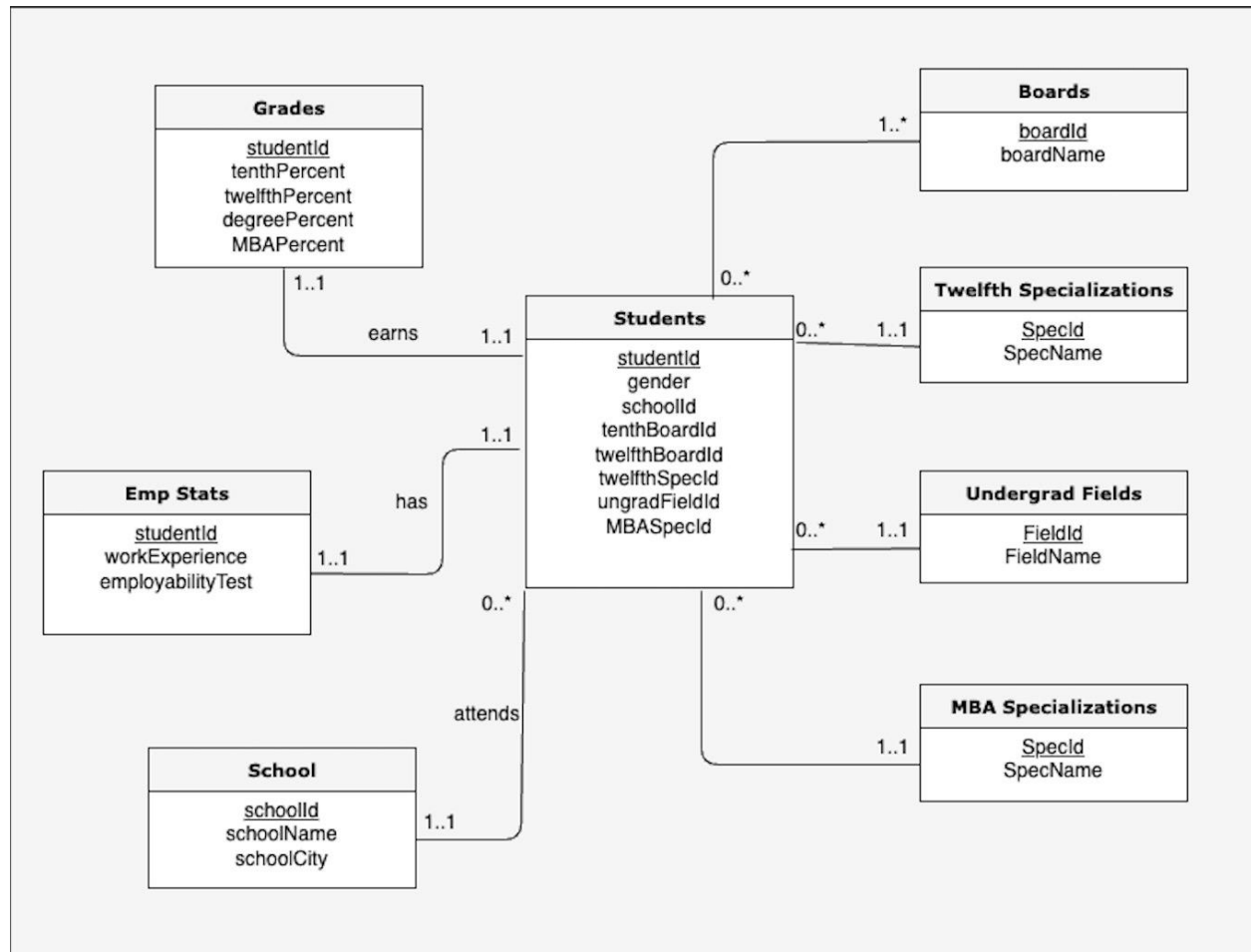## UML Diagram:



## Assumptions:
- We think each student should have one grade, and each grade should correspond to one student.
- We think that each student should have one set of employment statistics, and each set of employment statistics should correspond to one student.
- We think that each student should go to one school, and each school can have zero to many students.
- We think that each student should have one MBA specialization, and each MBA specialization should have zero to many students.
- We think that each student should have one undergrad field, and each undergrad field should have zero to many students.
- We think that each student should have one twelfth specialization, and each twelfth specialization should have zero to many students.
- We think that each student should have one to many boards, and each board should have zero to many students.

## Schema:
The schema we derived from the ER diagram is as follows:

**Students**(*studentId*:INT [PK], *gender*:VARCHAR(10), *schoolId*:INT [FK to School.schoolId], *tenthBoardId*:INT [FK to Boards.boardId], *twelfthBoardId*:INT [FK to Boards.boardId], *twelfthSpecId*:INT [FK to TwelfthSpecializations.SpecId], *ungradFieldId*:INT [FK to UndergradFields.fieldId, *MBASpecId*:INT [FK to MBASpecializations.SpecId])

**School**(*schoolId*:INT [PK], *schoolName*:VARCHAR(100), *schoolCity*:VARCHAR(100))

**EmpStats**(*studentId*:INT [PK], *workExperience*:INT, *employabilityTest*:INT)

**Grades**(*studentId*:INT [PK], *gender*:VARCHAR(10), *tenthPercent*:INT, *twelfthPercent*:INT *undergradPercent*:INT, *MBAPercent*:INT)

**Boards**(*boardId*:INT [PK], *boardName*:VARCHAR(100))

**TwelfthSpecializations**(*specId*:INT [PK], *specName*:VARCHAR(100))

**UndergradFields**(*fieldId*:INT [PK], *fieldName*:VARCHAR(100))

**MBASpecializations**(*specId*:INT [PK], *specName*:VARCHAR(100))


## Normalization:

The schema is in 3rd Normalization Form (3NF). 3NF states that relation R is in 3rd normal form if whenever there is a nontrivial dependency $A_1, A_2, ..., A_n$ -> B for R, then $\{A_1, A_2, ..., A_n \}$ is a super-key for R, or B is part of a key. This holds for our schema.

- All the attributes in the Students relation can be obtained from *studentId*, which is a super key. There are no other functional dependencies in this relation so it does not violate 3NF.
- All the attributes in the Grades relation can be obtained from *studentId*, which is a super key. There are no other functional dependencies in this relation so it does not violate 3NF.
- All the attributes in the EmpStats relation can be obtained from *studentId*, which is a super key. There are no other functional dependencies in this relation so it does not violate 3NF.
- All the attributes in the Boards relation can be obtained from *boardId*, which is a super key. There are no other functional dependencies in this relation so it does not violate 3NF.
- All the attributes in the TwelfthSpecializations relation can be obtained from *specId*, which is a super key. There are no other functional dependencies in this relation so it does not violate 3NF.
- All the attributes in the UndergradFields relation can be obtained from *specId*, which is a super key. There are no other functional dependencies in this relation so it does not violate 3NF.
- All the attributes in the MBASpecialization relation can be obtained from *specId*, which is a super key. There are no other functional dependencies in this relation so it does not violate 3NF.

We used 3NF normalization, rather than Boyce Codd Normal Form (BCNF), because it is lossless and dependency-preserving whereas BCNF is only lossless, not dependency-preserving.