Research Subteam: Presentation 2

Maxwell Bach, Okubay Gebrelibanos, Yili Hui, Arjun Patel, Manley Roberts

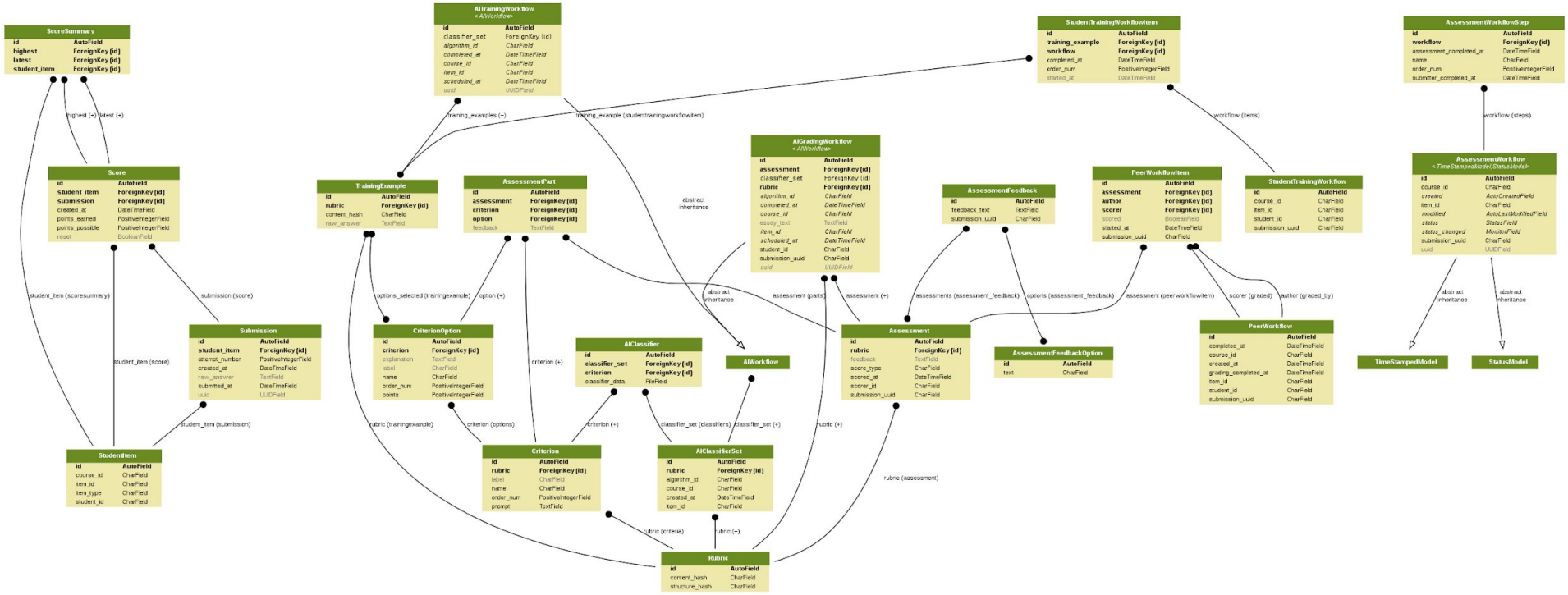
Review: Master Plan

- Aggregate/clean data
- Create a library of models
- Evaluate model success

Data: EdX on SQL

- Initial target attributes:
  - Correct Submissions Percent: Percentage of the total submissions that were correct.
  - Change in Weekly Average: Difference between current week average and previous week average quiz grade.
  - Average Raw Score: Average raw score on all assignments.
  - Pre-Submission Lead Time: Time between a quiz submission and deadline for all submissions; discretized buckets for $t \geq 7$ days, $3 \leq t < 7$, $1 \leq t < 3$, $0 \leq t < 1$, and late.
- Commonality!
  - All of these require understanding of the grading policy, due date, and type of each assignment (homework, quiz)

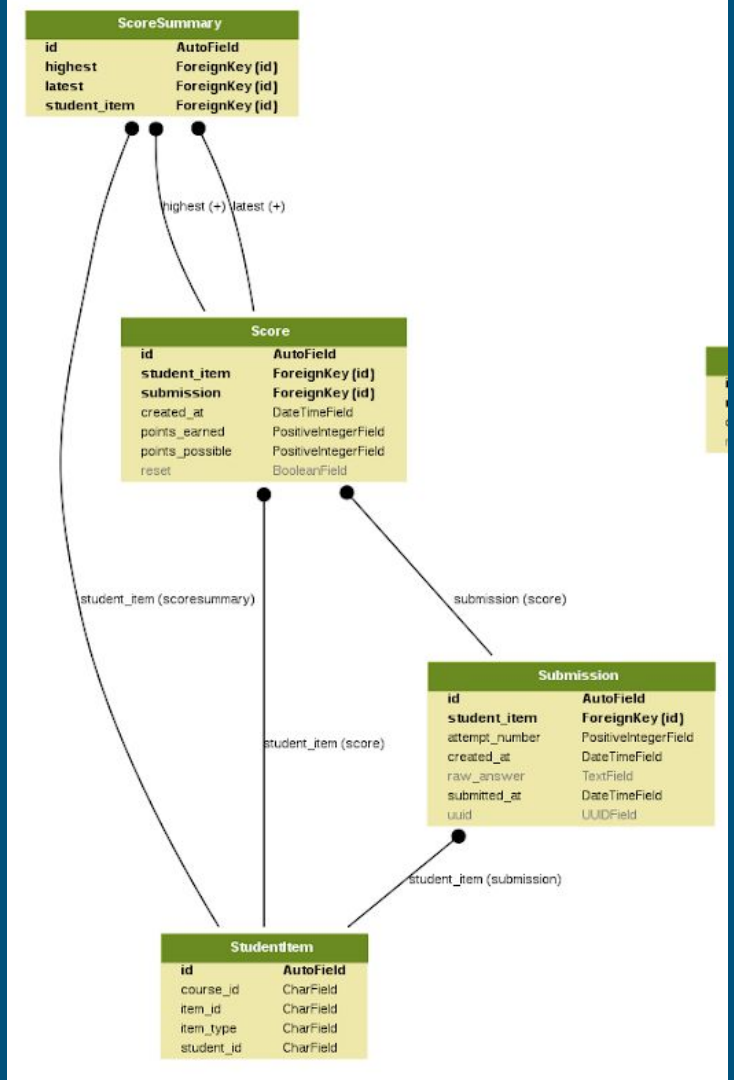
EdX on SQL: The Challenge



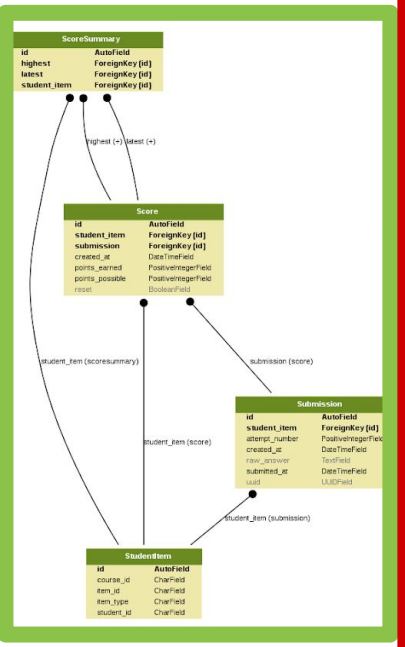
Populated Tables

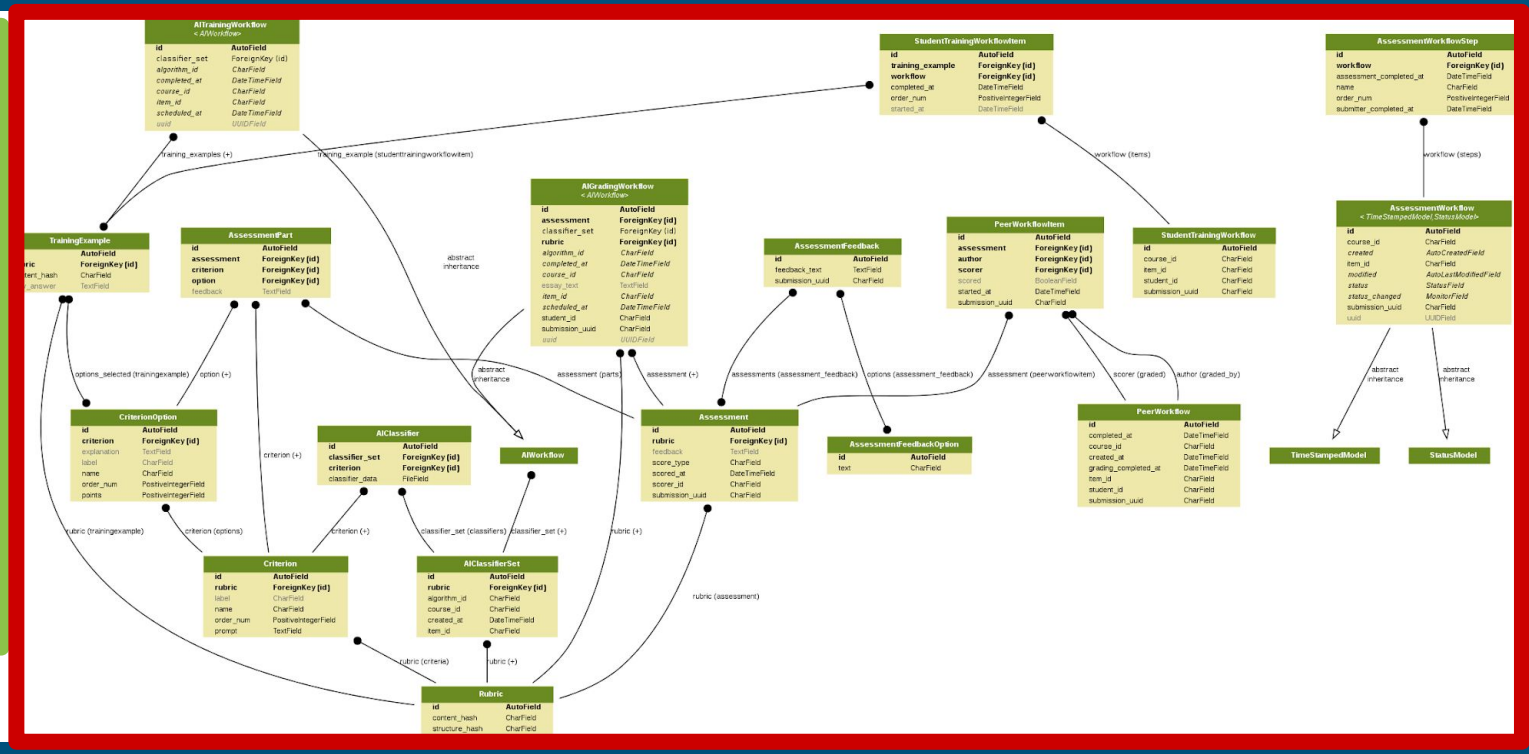
Joining Submission, StudentItem, Score, ScoreSummary gives us all information on a particular student response to an assignment

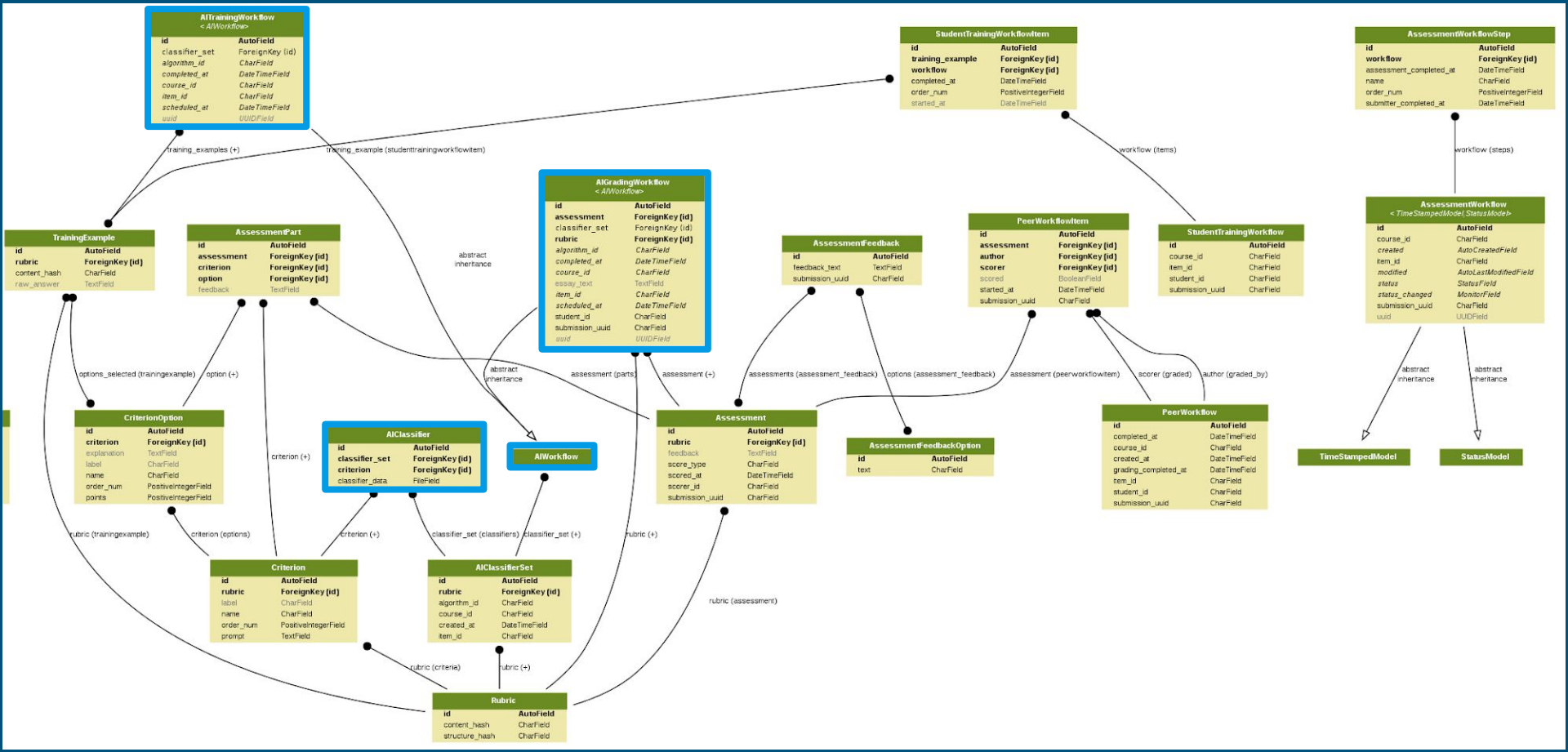
- What assignment is it?



Underused Tables






AITrainingWorkflow
< AIWorkflow>
id AutoField
classifier_set ForeignKey (id)
algorithm_id CharField
completed_at DateTimeField
course_id CharField
item_id CharField
scheduled_at DateTimeField
uuid UUIDField
StudentTrainingWorkflowItem
id AutoField
training_example ForeignKey (id)
workflow ForeignKey (id)
completed_at DateTimeField
order_num PositiveIntegerField
started_at DateTimeField
AssessmentWorkflowStep
id AutoField
workflow ForeignKey (id)
assessment_completed_at DateTimeField
name CharField
order_num PositiveIntegerField
submitter_completed_at DateTimeField
training_examples (+)
training_example (studenttrainingworkflowitem)
workflow (items)
workflow (steps)
AIGradingWorkflow
< AIWorkflow>
id AutoField
assessment ForeignKey (id)
classifier_set ForeignKey (id)
rubric ForeignKey (id)
algorithm_id CharField
completed_at DateTimeField
course_id CharField
essay_text TextField
item_id CharField
scheduled_at DateTimeField
student_id CharField
submission_uuid CharField
uuid UUIDField
AssessmentWorkflow
< TimeStampedModel,StatusModel>
id AutoField
course_id CharField
created AutoCreatedField
item_id CharField
modified AutoLastModifiedField
status StatusField
status_changed MonitorField
submission_uuid CharField
uuid UUIDField
TrainingExample
id AutoField
rubric ForeignKey (id)
content_hash CharField
raw_answer TextField
AssessmentPart
id AutoField
assessment ForeignKey (id)
criterion ForeignKey (id)
option ForeignKey (id)
feedback TextField
abstract inheritance
AssessmentFeedback
id AutoField
feedback_text TextField
submission_uuid CharField
PeerWorkflowItem
id AutoField
assessment ForeignKey (id)
author ForeignKey (id)
scorer ForeignKey (id)
scored BooleanField
started_at DateTimeField
submission_uuid CharField
StudentTrainingWorkflow
id AutoField
course_id CharField
item_id CharField
student_id CharField
submission_uuid CharField
options_selected (trainingexample)
option (+)
abstract inheritance
assessment (parts)
assessment (+)
assessments (assessment_feedback)
options (assessment_feedback)
assessment (peerworkflowitem)
scorer (graded)
author (graded_by)
abstract inheritance
abstract inheritance
CriterionOption
id AutoField
criterion ForeignKey (id)
explanation TextField
label CharField
name CharField
order_num PositiveIntegerField
points PositiveIntegerField
AIClassifier
id AutoField
classifier_set ForeignKey (id)
criterion ForeignKey (id)
classifier_data FileField
AIWorkflow
Assessment
id AutoField
rubric ForeignKey (id)
feedback TextField
score_type CharField
scored_at DateTimeField
scorer_id CharField
submission_uuid CharField
AssessmentFeedbackOption
id AutoField
text CharField
PeerWorkflow
id AutoField
completed_at DateTimeField
course_id CharField
created_at DateTimeField
grading_completed_at DateTimeField
item_id CharField
student_id CharField
submission_uuid CharField
TimeStampedModel
StatusModel
criterion (+)
rubric (trainingexample)
criterion (options)
criterion (+)
classifier_set (classifiers)
classifier_set (+)
rubric (+)
Criterion
id AutoField
rubric ForeignKey (id)
label CharField
name CharField
order_num PositiveIntegerField
prompt TextField
AIClassifierSet
id AutoField
rubric ForeignKey (id)
algorithm_id CharField
course_id CharField
created_at DateTimeField
item_id CharField
rubric (assessment)
rubric (criteria)
rubric (+)
Rubric
id AutoField
content_hash CharField
structure_hash CharField

Solution?

- NEW Grading Policy Data
  - Tells us how many of specific assignment types to expect
- NEW Additional Course Structure Data
  - What are our assignments? Problems in the flow of course sequence
  - Assignment context (which chapter, which content sequence, etc.)
    - Subsection of course?
  - Submission date
- Joining on this data may help make sense of submissions

What comes next

- MongoDB - Clickstream data
  - Just completed a crash course on dealing with this data (today)
- Continuing to make sense of SQL data (with new information)
- Piazza data - adding to SQL server and cleaning