Nexford University

BAN6430: Data Mining

Module 1 Assignment: Developing ERD to Model Data

Professor: Rajan Thapaliya

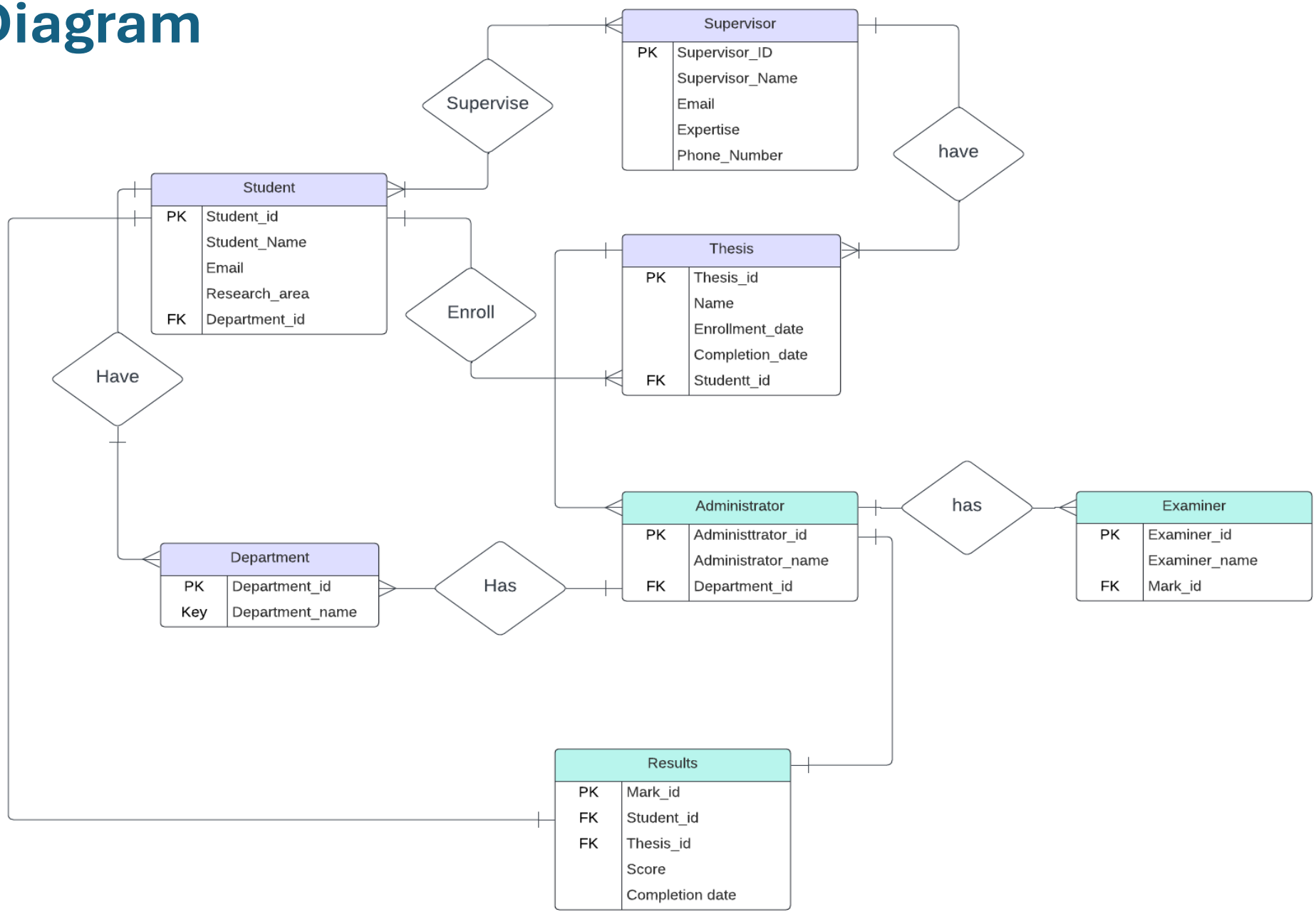
Presented by: Pharaoh Kipkirui Chirchir

Date: 04 Mar 2024

ERD Model Diagram

ERD Cardinality Key

- One
- One or Many
- Many



Data Mining Examples from ERD Model

Trends of Workload Allocation to Supervisors: The trends in supervisor allocation based on research areas and expertise can be analyzed by use of historical data.

Thesis Completion Analysis: Evaluate the time students complete their theses and identify factors affecting completion time, including the time it takes to mark the thesis.

Supervisor Workload: This can be achieved by analyzing the number of learners each supervisor is working for a certain period.

Departmental Performance: By analyzing learner's scores and completion rates, each department can be evaluated on how it is performing.

Research Area Popularity: Through analysis of research topic distribution, analysis can be done on the most common areas to learners.

Possible issues in ERD Model

- **Data Consistency and noise -** Inconsistency in data entry, especially when recording marks and completion dates, can lead to misleading data entry errors, hence the wrong decision.
- **Placement of Relationship Attributes -** Use of relationship attributes such as “date_marked” and mark in the Thesis relationship can lead to problems if the attributes have more complex relationships or dependencies.
- **Use of Entity issues versus attributes** - The "program" attribute in the "Student" entity is used to differentiate between undergraduate and postgraduate students. However, it could be modeled as a separate entity, especially if there is additional information specific to each program.
- **Binary versus nary relationship sets** - The relationships in the ERD are mainly between two entities. The “Administrator" relationship connects the "Thesis" entity with the “Examiner" entity. However, in a more complex scenario, a thesis might involve multiple examiners or administrators.
- **Use of entity sets versus relationship sets** - The entities "Supervisor" and "CoSupervisor" can be treated as separate entities, and each is directly related to the "Thesis" entity. This might suit the current scenario as it assumes that the relationship between the supervisors is fundamentally different, which might not be true in all the scenarios.