SCC.221
DATA Engineering

2025 - Week 1 – Module Information.

Uraz C Turker

What will you learn this week?

Lancaster University

- Understand fundamental concepts of databases

Learning outcomes

Lancaster University

- Definitions for
  - Data
  - Data Engineering
  - Database
  - Database Management System
  - Basics of Entity Relationship Diagrams

What is a Data?

Lancaster University

What is data?

What is a Data?

Lancaster University

Data is anything that can be represented using Binary notation.

We can then encode data in File:

0	1	0	0	1	1	1	0
0	1	0	0	1	1	1	1
0	0	1	1	1	0	1	0

NO:

Source of Data

Lancaster University

- Data is being created whenever we:
  - Eat,
  - Travel,
  - Chat,
  - Talk over a line on a phone, etc.

A source of data: Games

The aim of having data?

- is to retrieve information.
- To achieve this, we need to process this data.
- What is the fundamental step to process data?
- To establish a flexible, understandable, and a common representation for the data.

What is Data Engineering?

Lancaster University

- “Data engineering is the process of designing and building systems that allow people to collect, manage, and analyse data. Data engineers work to make raw data usable for data scientists and business analysts so that organizations can use it to improve their performance.”

What is Data Engineering?

Lancaster University

- "Data engineering is the process of designing and building systems that allow people to collect, manage, and analyse data. Data engineers work to make raw data usable for data scientists and business analysts so that organizations can use it to improve their performance."

Data engineers are responsible for:

Lancaster University

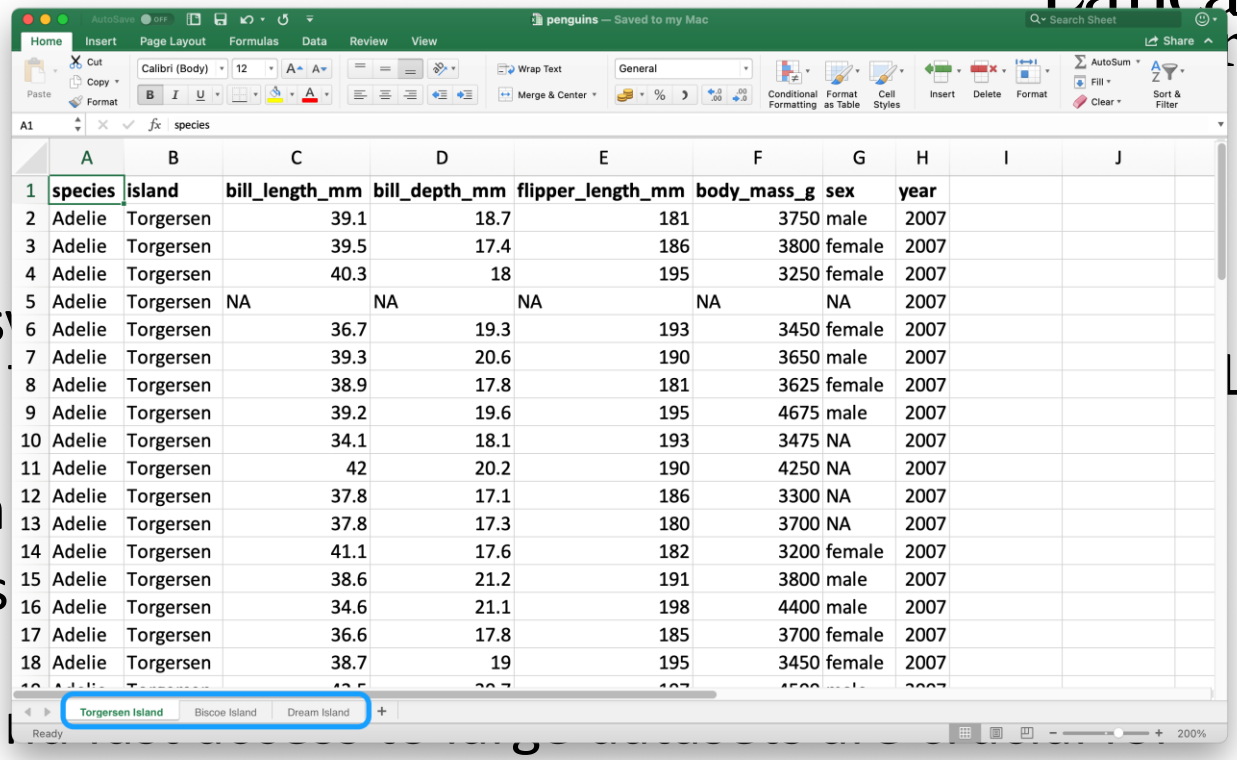
- **Data pipelines:** Creating data pipelines (flows) to manage and process large sets of data
- **Data integration:** Ensuring that data from different sources is integrated seamlessly
- **Data quality:** Ensuring that data is of high quality and that the data infrastructure is reliable and efficient
- **Data analysis:** Creating raw data analyses to provide predictive models and show trends
- **Data security:** Managing and storing data securely to protect it from loss or theft
- **Automation:** Creating ways to automate tasks within the data pipeline to improve efficiency
- Data engineers often work for larger organizations and may use tools and technologies such as Hadoop, MongoDB, and Kafka.

Data engineers are responsible for:

Lancaster University

- **Data pipelines:** Creating data pipelines (flows) to manage and process large sets of data
- **Data integration:** Ensuring that data from different sources is integrated seamlessly
- **Data quality:** Ensuring that data is of high quality and that the data infrastructure is reliable and efficient
- **Data analysis:** Creating raw data analyses to provide predictive models and show trends
- **Data security:** Managing and storing data securely to protect it from loss or theft
- **Automation:** Creating ways to automate tasks within the data pipeline to improve efficiency
- Data engineers often work for larger organizations and may use tools and technologies such as Hadoop, MongoDB, and Kafka.

Databases



	A	B	C	D	E	F	G	H	I	J
1	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	year		
2	Adelie	Torgersen	39.1	18.7	181	3750	male	2007		
3	Adelie	Torgersen	39.5	17.4	186	3800	female	2007		
4	Adelie	Torgersen	40.3	18	195	3250	female	2007		
5	Adelie	Torgersen	NA	NA	NA	NA	NA	2007		
6	Adelie	Torgersen	36.7	19.3	193	3450	female	2007		
7	Adelie	Torgersen	39.3	20.6	190	3650	male	2007		
8	Adelie	Torgersen	38.9	17.8	181	3625	female	2007		
9	Adelie	Torgersen	39.2	19.6	195	4675	male	2007		
10	Adelie	Torgersen	34.1	18.1	193	3475	NA	2007		
11	Adelie	Torgersen	42	20.2	190	4250	NA	2007		
12	Adelie	Torgersen	37.8	17.1	186	3300	NA	2007		
13	Adelie	Torgersen	37.8	17.3	180	3700	NA	2007		
14	Adelie	Torgersen	41.1	17.6	182	3200	female	2007		
15	Adelie	Torgersen	38.6	21.2	191	3800	male	2007		
16	Adelie	Torgersen	34.6	21.1	198	4400	male	2007		
17	Adelie	Torgersen	36.6	17.8	185	3700	female	2007		
18	Adelie	Torgersen	38.7	19	195	3450	female	2007		

- Databases are structured s
  data. They include various
  PostgreSQL) .
- Raw data is often stored in
- Data stored in databases is
  models, which define how
- Efficient query execution a
  databases.
- Secure storage and access control of sensitive data is handled within databases.
- Traditional and cloud-based databases are used to store data, but as the data scales, the infrastructure must also scale.

What we had and why did we invent DBMS?

Lancaster University

- Previously we had File-processing applications:
  - A set of files are processed to retrieve information.
  - Hard to implement,
  - Optimisation, performance, reliability, adoption, re-use were main issues.
- Standardisation of concepts.
  - Representation,
  - Coding,
  - Formatting, etc.
- Similar question “Why do we have an operating system?”

What is a Data, Database, Database Management System (DBMS)?

Lancaster University

Database: It represents a massive amount of persistent data according to a logical model

The logical model can be represented graphically by.

UML or Entity Relationship Diagrams.

The logical model allows us to reason about the logical representation of the data to gain the flexibility to establish an excellent data representation power to aid processing.

What is a Data, Database, Database Management System (DBMS)?

Lancaster University

DBMS: Provides efficient, reliable, convenient, safe, multi-user friendly storage and access to massive amounts of persistent data.

What is a Data, Database, Database Management System (DBMS)?

Lancaster University

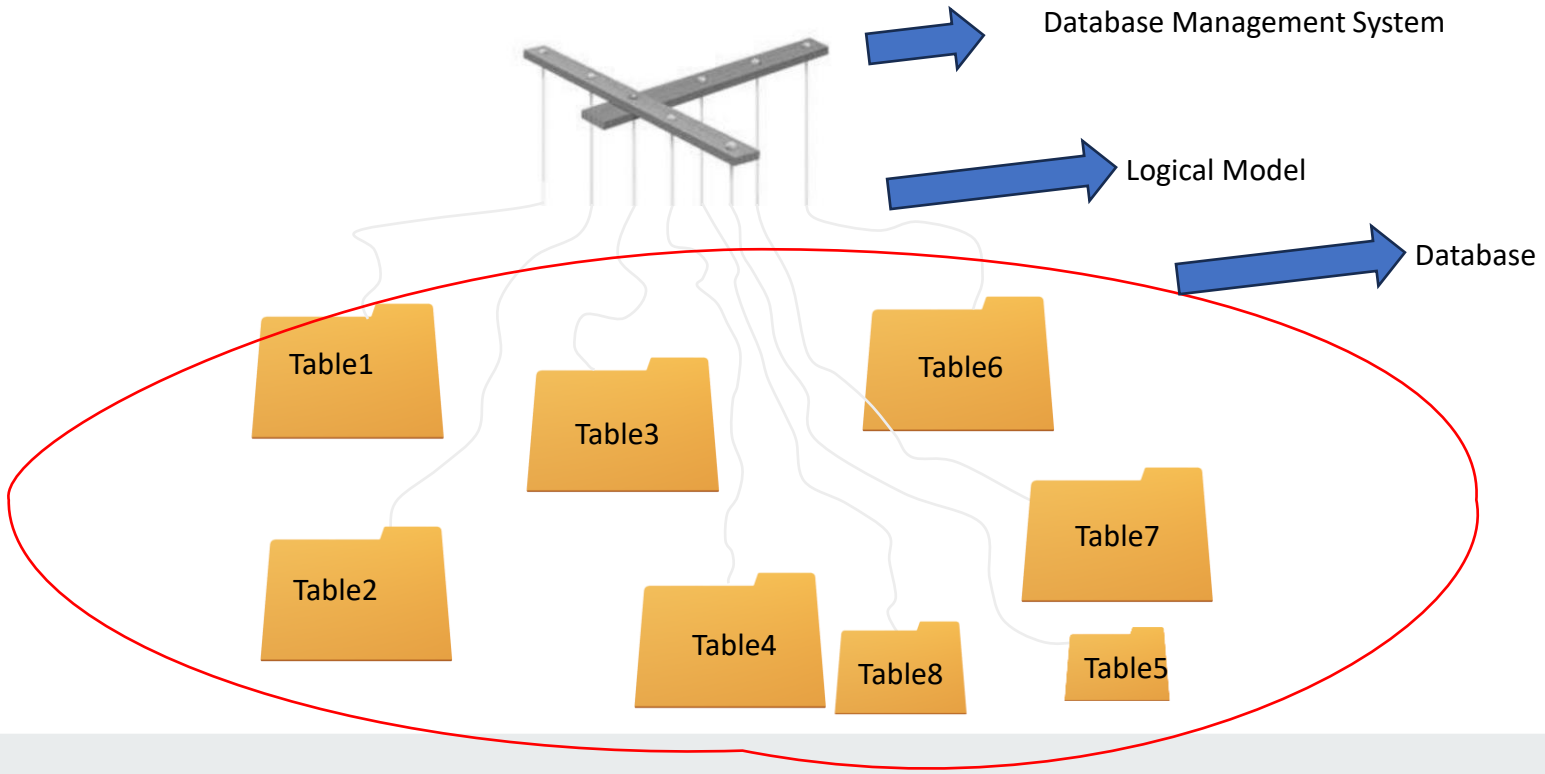
DBMS: Provides efficient, reliable, convenient, safe, multi-user friendly storage and access to massive amounts of persistent data.

This is why they are very popular.

Relation between DBMS, Database and the Logical Model.

Lancaster University

Database: It represents a massive amount of persistent data according to a logical model



Databases in real life.. EVE Online

Lancaster University

https://docs.google.com/spreadsheets/d/1bVeIR2hB27e14mzHQfcwsJZP_Rg3u1BPmx6jDN_13fo/edit#gid=0

https://wiki.eveuniversity.org/Static_Data_Export

Databases in real life. Counter Strike.
https://www.csgodatabase.com/about/

Lancaster University

Representing the logic of the database

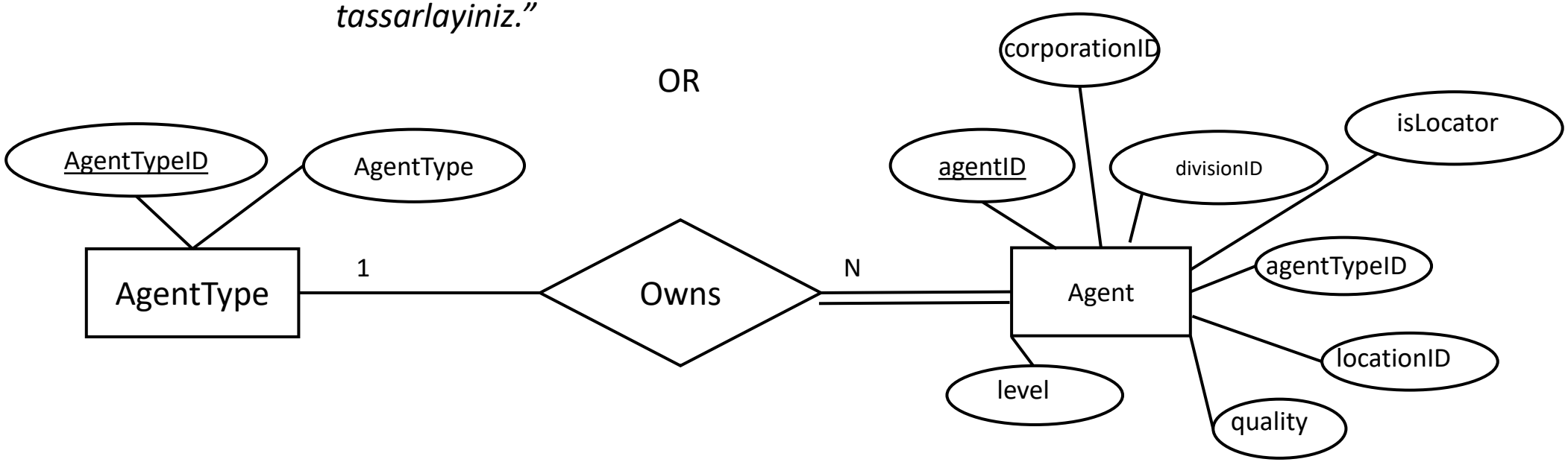
Lancaster University

- Designing a database requires documentation that will be used to i) reveal database requirements and ii) perform debugging/update and maintenance with ease.
- As in the case of many software engineering fields, the document is heavily on graphical representation (similar to UMLs).
- Among others, **Entity-Relationship diagrams** are used to represent the logical structure of the database.

Example

- In EVE Game there are:
- **Agents**
  - Where each agent has **agentID, divisionID, corporationID, locationID, level, quality, agentTypeID, isLocator**;
- **AgentTypes**
  - where each type has **agentTypeID, agentType**.

- *Query (Eng): "Please design a database where every Agent must have exactly one AgentType, but every AgentType can be associated with several Agents."*
- *Istem (Tr): "Lutfen, her **Agent'**in bir **AgentType**'inin oldugu ve her **AgentType**inin bir veya birden fazla **Agent** ile iliskisi oldugu bir veri tabani tassarlayiniz."*



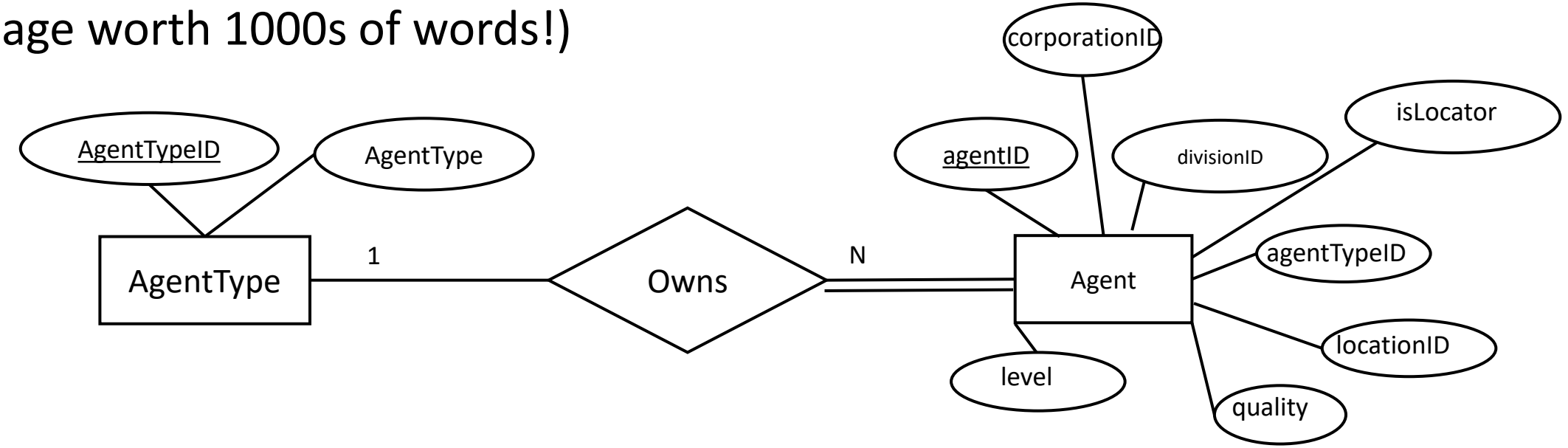
The logical model: Entity relationship model (Chens Representation).

Lancaster University

A graphical representation of the logical view of a Database.

Communication method (a single image worth 1000s of words!)

https://wiki.eveuniversity.org/Static_Data_Export
agtAgents.csv.bz2
agtAgentTypes.csv.bz2



Key concepts before proceeding.

Lancaster University

- Data Definition Language (DDL) is used to create tables.
- Data Manipulation Language (DML) is used to modify tables.

Key people before proceeding.

Lancaster University

- DBMS Implementer.
- DB Designer.
- DB Application Developer.
- DB Admin.

Lancaster University

The following slides are here to assure Uraz that everybody understands ERDs!

Please read Elmasri's (Fundamentals of database systems) book Chapter 3.

From you…

Lancaster University

Show chart data

What more support can be provided?

- Can you please add the slides as pdf?

Any other comments

- We can find power points only but it would be really helpful if we can get them as pdf format

From you…

Show chart data

What more support can be provided?

- Can you please add the slides as pdf?

Any other comments

- We can find power points only but it would be really helpful if we can get them as pdf format

Key concepts of ER diagrams: Entity

Lancaster University

- **Entity:** An object that has distinguishable **attributes**.
- Entity: An apple
  - Attributes:
    - Colour
    - Size
    - Origin
- Entity: AgentType
  Attributes:
  AgentTypeID
  AgentType

AgentTypeID	AgentType
1	NonAgent
2	BasicAgent

Entity set: A conceptual representation of entities w.r.t the same set of attributes is called an Entity Set.

Colour	Size	Origin
Red	Large	Spain
Green	Medium	Madagascar

Key concepts of ER: Entity

Lancaster University

- **Entity:** An object that has distinguishable **attributes**.
- Rectangle is used to represent an entity set.

Key concepts of ER: Entity

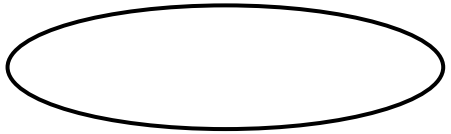
Lancaster University

- **Entity:** An object that has distinguishable **attributes**.
- Rectangle is used to represent an entity set.

Key concepts of ER: Entity

Lancaster University

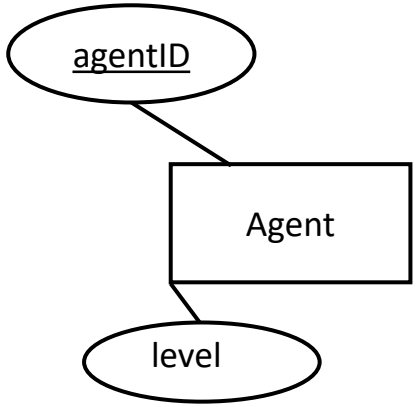
- **Entity:** An object that has distinguishable **attributes**.
- Rectangle is used to represent an entity set.
- Oval is used to represent an attribute.



Key concepts of ER: Entity

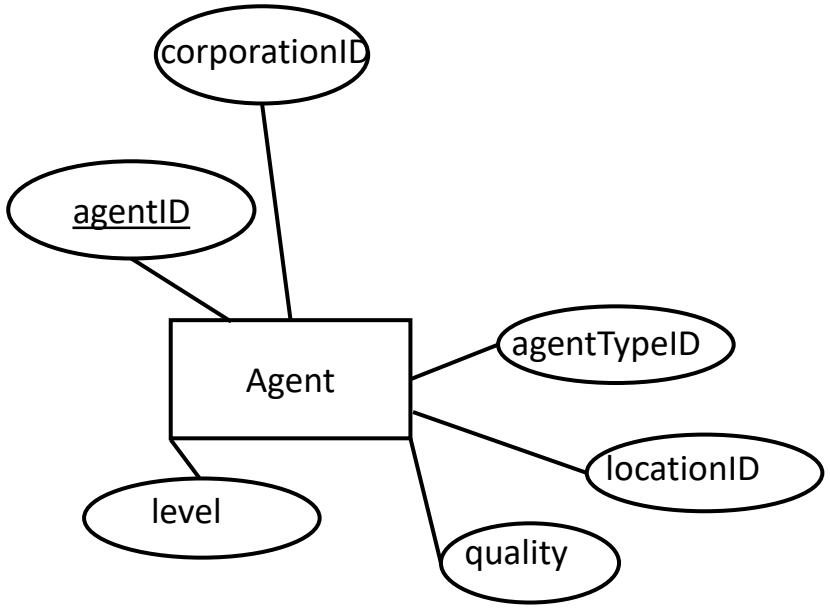
Lancaster University

- **Entity:** An object that has distinguishable **attributes**.
- Rectangle is used to represent an entity set.
- Oval is used to represent an attribute.



Key concepts of ER: Entity

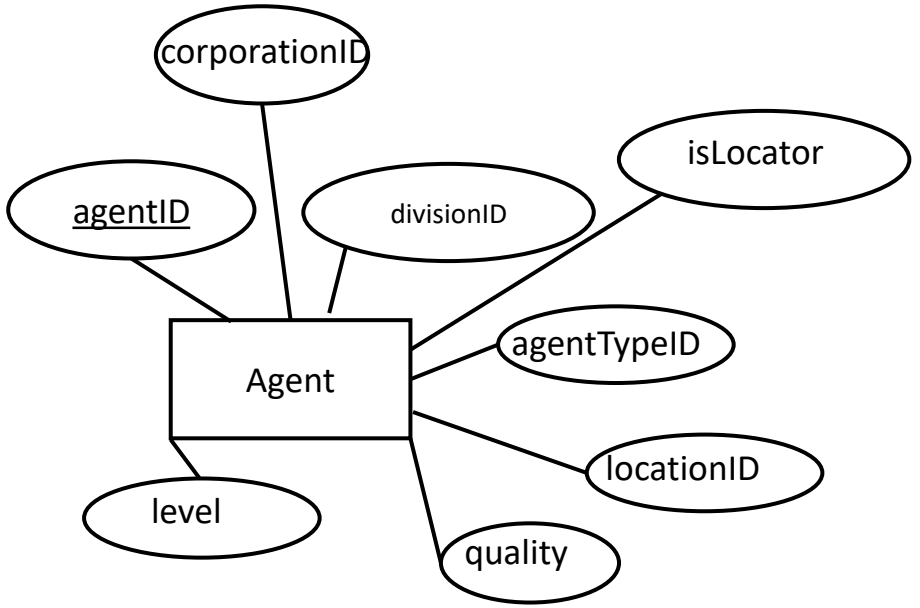
- **Entity:** An object that has distinguishable **attributes**.
- Rectangle is used to represent an entity set.
- Oval is used to represent an attribute.



Key concepts of ER: Entity

Lancaster University

- **Entity:** An object that has distinguishable **attributes**.
- Rectangle is used to represent an entity set.
- Oval is used to represent an attribute.

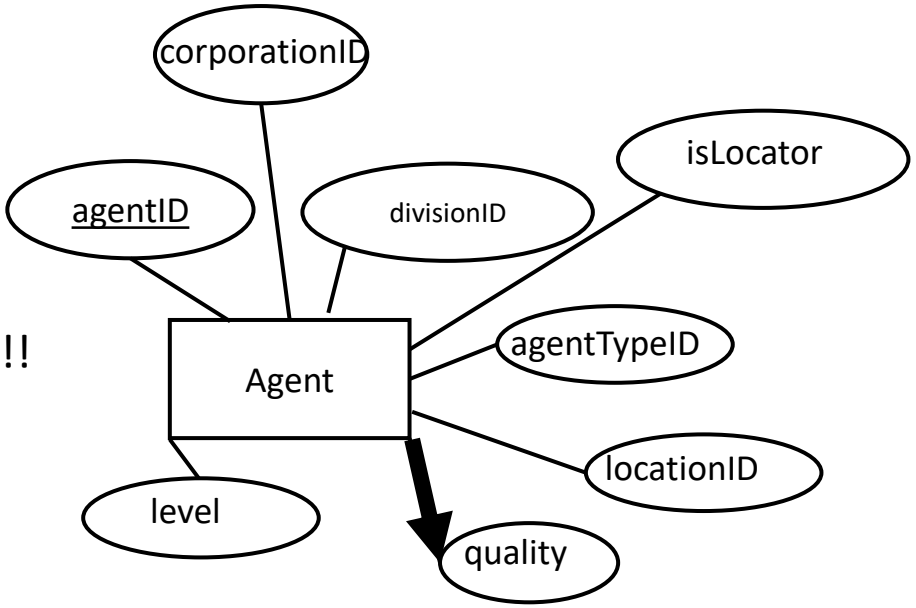


Key concepts of ER: Entity

Lancaster University

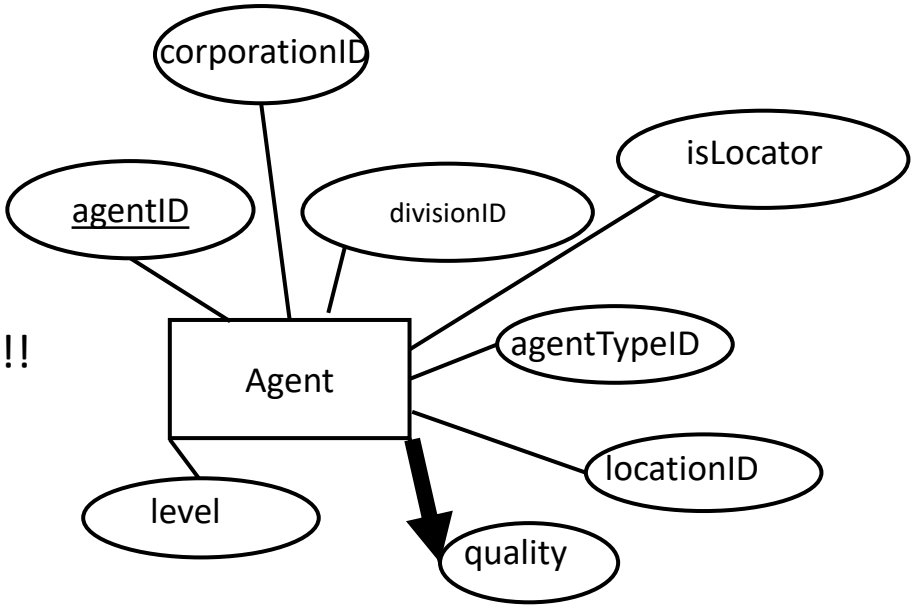
- **Entity:** An object that has distinguishable **attributes**.
- Rectangle is used to represent an entity set.
- Oval is used to represent an attribute.

WE NEVER USE ARROWS!!!!!



Key concepts of ER: Entity

- **Entity:** An object that has distinguishable **attributes**.
- Rectangle is used to represent an entity set.
- Oval is used to represent an attribute.



WE NEVER USE ARROWS!!!!!

Two semesters ago, one student failed all of the coursework!

Exercise…

Key concepts of ER: Entity

- **Entity:** An object that has distinguishable **attributes**.
- **Entity** *A Solar System* can have *RegionName, ConstellationName, System Name, X, Y, Z, Security, SecurityClass* as its **attributes**.

https://docs.google.com/spreadsheets/d/1bVeIR2hB27e14mzHQfcwsJZP_Rg3u1BPmx6jDN_13fo/edit#gid=0

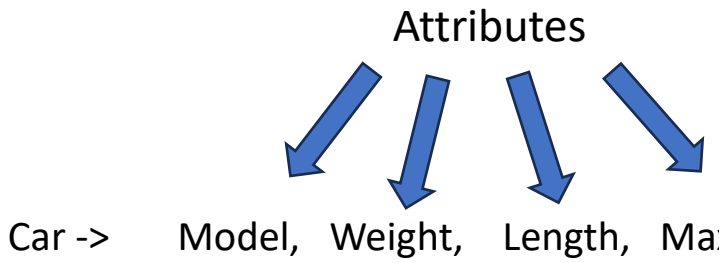
Eve Solar System as Entity Set

RegionName	ConstellationName	System Name	SystemID	X	Y	Z	Security	SecurityClass
Domain	Throne Worlds	Amarr	30002187	-20474870725000900	4023837993657140	-5762127890242100	1.0	A
Heimatar	Sveipar	Ammold	30002547	-8534866579833560	1874911505767550	4543015676849680	1.0	A
Lonetrek	Sela	Amsen	30001392	-145493013225457000	101594111567536000	137638171422790000	1.0	A
Sinq Laison	Nexus	Bourynes	30002715	-166183387531946000	44475501577819500	26973565770462400	1.0	A
Domain	Parud	Chaven	30003489	-180372429369955000	60008167808112100	-54794737517953000	1.0	A

Some other databases are given in XML format. https://archive.org/details/stackexchange

Key concepts of ER: Entity

Consider an Entity Set called CAR
ENTITY SET / Schema / Table



Car -> Model, Weight, Length, Max_Speed.

	Model	Weight	Length	Max_Speed
First Entity	BMW 3.21	1400	3.21	200
Second Entity	Toyota_Corolla	1300	3.16	200
Third Entity	Hyundai E.GLS	1400	3.16	210

Key concepts of ER: Entity

Lancaster University

Consider an Entity Set called CAR
ENTITY SET / Schema / Table

Attributes

Car -> Model, Weight, Length, Max_Speed.

	Model	Weight	Length	Max_Speed
First Entity	BMW 3.21	1400	3.21	200
Second Entity	Toyota_Corolla	1300	3.16	200
Third Entity	Hyundai E.GLS	1400	3.16	210
Fourth Entity	Hyundai E.GLS	1400	3.16	210

But we have redundancy here.. Third and Fourth Entities are the same!

Key concepts of ER: Entity

Lancaster University

Consider an Entity Set called CAR
ENTITY SET / Schema / Table

Attributes

Car -> Model, Weight, Length, Max_Speed.

	Model	Weight	Length	Max_Speed
First Entity	BMW 3.21	1400	3.21	200
Second Entity	Toyota_Corolla	1300	3.16	200
Third Entity	Hyundai E.GLS	1400	3.16	210

To avoid having duplicate entities, we need to be able to identify each row uniquely…
And, so select an attribute that distinguishes an entity within the Car entity set.

Key concepts of ER: Entity

Lancaster University

Consider an Entity Set called CAR
ENTITY SET / Schema / Table

Attributes

Car -> Model, Weight, Length, Max_Speed.

	Model	Weight	Length	Max_Speed
First Entity	BMW 3.21	1400	3.21	200
Second Entity	Toyota_Corolla	1300	3.16	200
Third Entity	Hyundai E.GLS	1400	3.16	210

To avoid having duplicate entities, we need to be able to identify each row uniquely…
And, so select an attribute that distinguishes an entity within the Car entity set.

Model uniquely identifies an entity (the whole row) within this entity set.

Key concepts of ER: Entity

Lancaster University

Consider an Entity Set called CAR
ENTITY SET / Schema / Table

Attributes

Car -> Model, Weight, Length, Max_Speed.

	Model	Weight	Length	Max_Speed
First Entity	BMW 3.21	1400	3.21	200
Second Entity	Toyota_Corolla	1300	3.16	200
Third Entity	Hyundai E.GLS	1400	3.16	210

To avoid having duplicate entities, we need to be able to identify each row uniquely…
And, so select an attribute that distinguishes an entity within the Car entity set.

Model uniquely identifies an entity (the whole row) within this entity set.

The attribute that uniquely identifies entities within the entity set is called a **'key attribute'**.

Key Attribute and Primary Key Attribute

Lancaster University

There can be more than one key attribute for a given Entity set.

Model	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1200	3.18	220
Hyundai E.GLS	1100	3.16	210

The designer has to select key attribute(s) as the **Primary Key** Attribute according to the context.

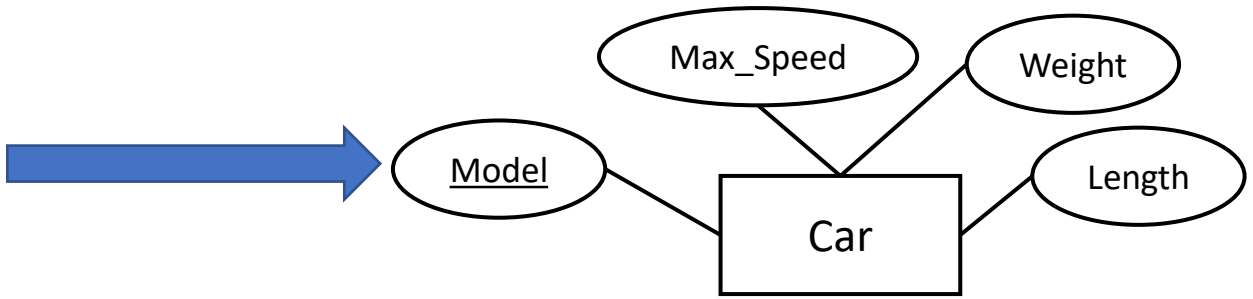
“Car Models should not repeat.”

Key concepts of ER: Entity

- Car -> Model, Weight, Length, Max_Speed.

Model	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210

The primary key attributes are represented with solid underlines (Model is selected).



Exercise…

Lancaster University

Key concepts of ER: Entity

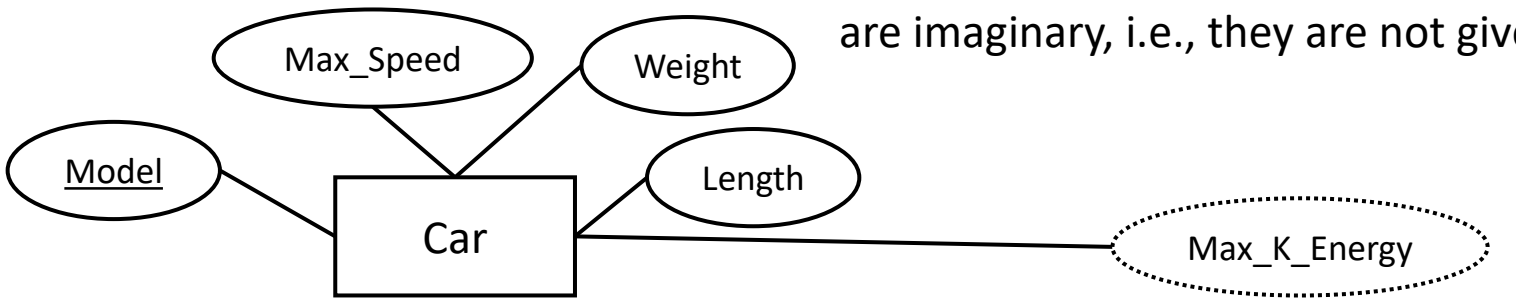
Lancaster University

- Car -> Model, Weight, Length, Max_Speed.

Model	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210

Some attributes are computed: Derived attributes.

Derived attributes are given using **dashed lines**. They are imaginary, i.e., they are not given in the table.



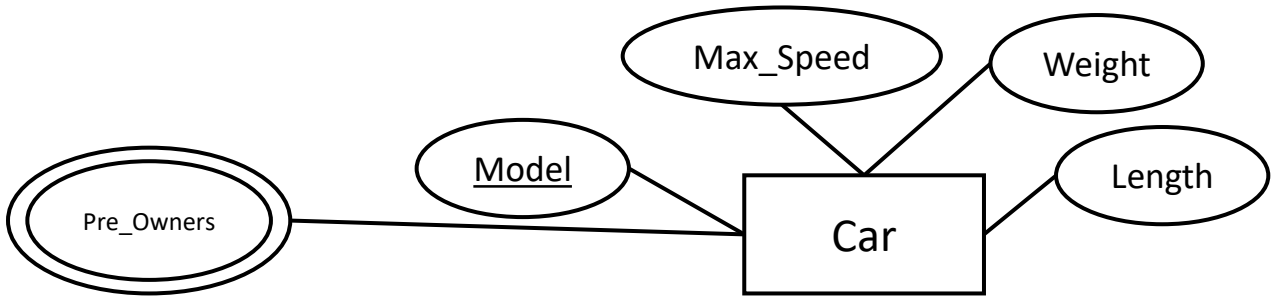
Key concepts of ER: Entity

- Car -> Model, Weight, Length, Max_Speed.

Model	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210

Some attributes have multi-values: Multi-valued attributes.

Multi-valued attributes are given **using double lines**.



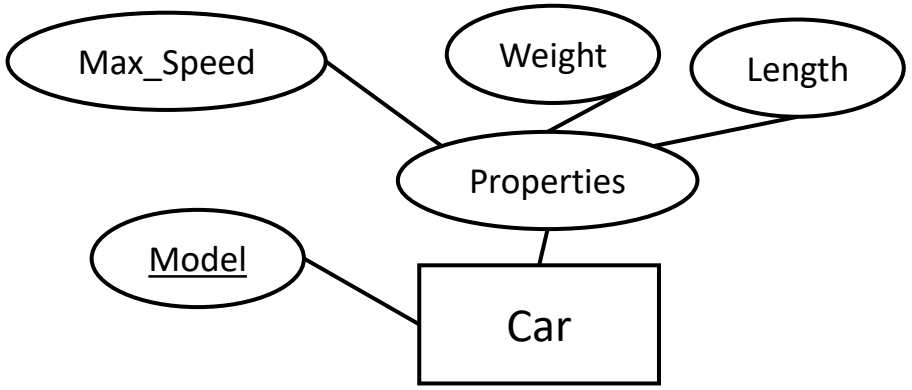
Key concepts of ER: Entity

Lancaster University

- Car -> Model, Weight, Length, Max_Speed.

Model	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210

We can create composite attributes using others.



Composite attributes are given as a **group of attributes**, and they are imaginary, i.e., a composite attribute is not given in the table.

Key concepts of ER: Entity

Lancaster University

- **Entity:** An object that has distinguishable attributes.
- What other entity sets are there for EVE Online?

https://wiki.eveuniversity.org/Static_Data_Export

Key concepts of ER: Entity

Lancaster University

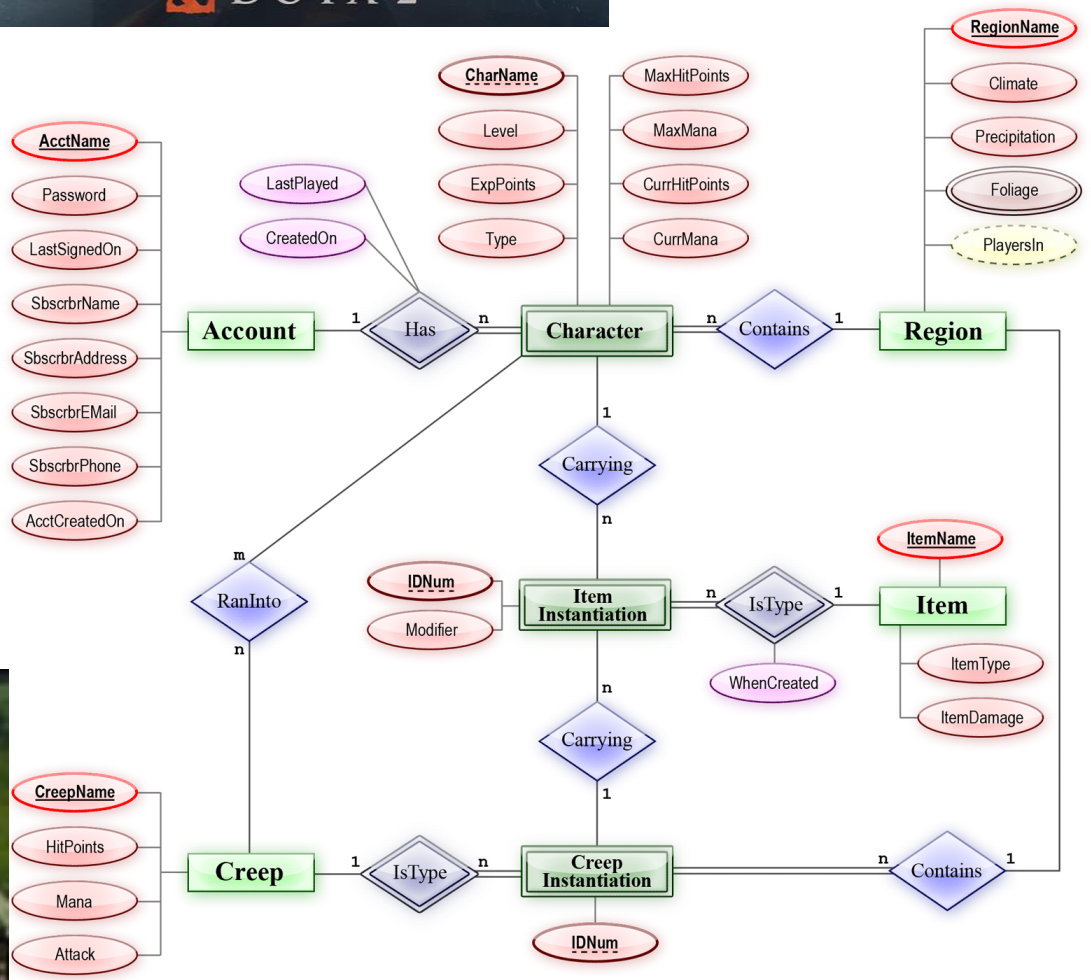
- **Entity:** An object that has distinguishable **attributes**.
- What entities and attributes can we have in multiplayer computer games?

Entity: Accounts

- Attributes:
  - AccountName
  - Password
  - ..

Entity: Character

- Attributes
  - CharacterName
  - Level
  - …



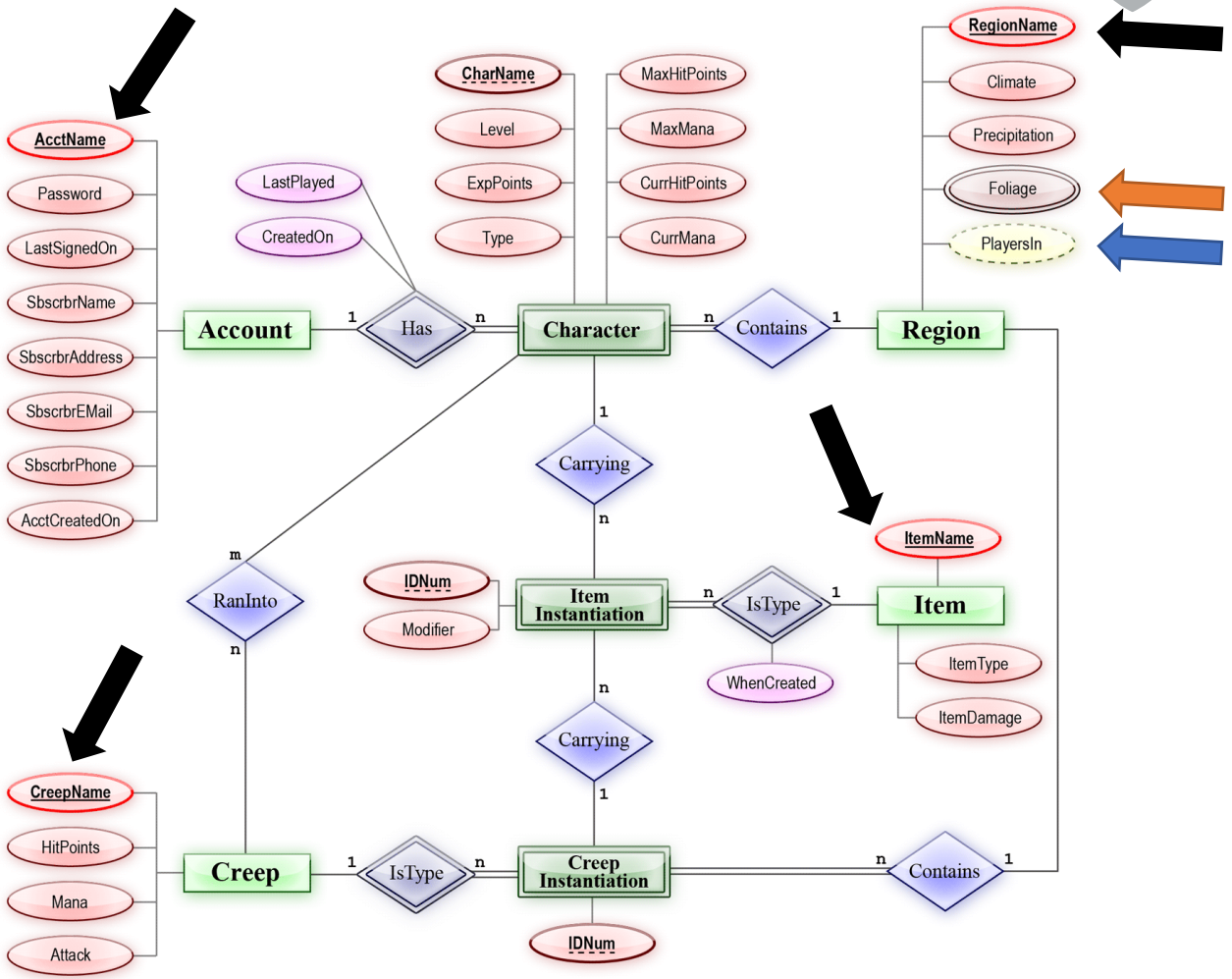
Exercise…

Lancaster University

Check yourself!

- **Which are multi-valued attributes?**
- Which are Derived Attributes?
- **Which are the primary key attributes?**

DOTA Database



Relations

Lancaster University

- In real life, an application usually interacts with several Entity Sets.
- Why?

Relations

Lancaster University

- In real life, an application usually interacts with several Entity Sets.
- Solar System

Relations

Lancaster University

- In real life, an application usually interacts with several Entity Sets.
- Solar System, Space Ship

Relations

Lancaster University

- In real life, an application usually interacts with several Entity Sets.
- Solar System, Space Ship, Commander. Etc..

Relations

Lancaster University

- In real life, an application usually interacts with several Entity Sets/Tables.
- Keeping everything in a single table will lead to poor performance as everything will be kept in a single file. Queries like "*Tell me the names of commanders that can pilot Space_Shuttle Falcon*" will take ages.
-

Relations

Lancaster University

- In real life, an application usually interacts with several Entity Sets/Tables.
- Keeping everything in a single table will lead to poor performance as everything will be kept in a single file. Queries like "*Tell me the names of commanders that can pilot Space_Shuttle Falcon*" will take ages.
- Therefore, a database has several tables instead of a single table.

Relations

Lancaster University

Plan S
Making full & immediate Open Access a reality

Template for 2022 annual reporting on Plan S approved Transformative Journals

Send to info@coalition-s.org no later than 1 June 2022

URL to page demonstrating how institutions purchasing a subscription to a Transformative Journal will pay only for remaining subscription	https://www.springernature.com/g
URL where data below is published:	https://www.springernature.com/g

MANDATORY COALITION S DATA FIELDS

ADDITIONAL DATA FIELDS PROVIDED BY SN

		2020 - Year 0 data				2021 - Year 1 data					2022 - Year 2 OA Target	Information on usage comparing subscription and OA content (2021 pubs)					Information on usage comparing subscription and OA content (2019 pubs)				Information on usage comparing subscription and OA content (2020 pubs)				
Name of journal	ISSN	Total number of research articles published in 2020	Total number of research articles published Open Access in 2020	Year 0 OA penetration rate (research articles)	Year 1 target [As agreed previously]	Total number of research articles published in 2021	Total number of research articles published Open Access in 2021	Year 1 OA penetration rate (Actual, research articles)	TJ target met Yes / No	If OA target not met, has a TJ exception been applied for? [Yes/No]	2022 Target (% of research content published OA)	Average number of citations for subscription research articles published in 2021	Average number of downloads for subscription research articles published	Average number of citations for OA research articles published in 2021	Average number of downloads for OA research articles published in 2021	Comments, observations	Average number of citations for subscription research	Average number of downloads for subscription	Average number of citations for OA research articles publish	Average number of downloads for OA research articles	Average number of citations for subscription research	Average number of downloads for subscription	Average number of citations for OA research articles publish	Average number of downloads for OA research articles	Year 1 OA penetration rate (Actual) (all article types
Advances in Neurodevelopmental Disorders	2366-7540	40	4	10%	15%	31	6	19%	Yes		24%	0	224	0	7800		0	265	0	2998	0	219	1	1864	18%
The Analysis of Verbal Behavior	2196-8926	11	0	0%	5%	19	1	5%	Yes		10%	0	91	0	2474		1	422	0	0	0	478	0	0	4%
Research in the Mathematical Sciences	2197-9847	32	7	22%	27%	69	16	23%	No	Yes - exceptio	32%	0	71	0	1930		3	175	3	1166	2	136	1	730	23%
Foundations of Chemistry	1572-8463	28	8	29%	34%	12	5	42%	Yes		48%	0	173	0	4148		1	306	0	4502	0	222	1	7911	24%
Continental Philosophy Review	1573-1103	29	5	17%	22%	33	14	42%	Yes		48%	0	267	0	6096		1	408	16	8279	1	346	1	3130	36%
International Journal for Philosophy of Religion	1572-8684	29	5	17%	22%	23	6	26%	Yes		31%	0	142	0	3232		1	319	3	3496	0	357	1	2571	19%
Archive for Mathematical Logic	1432-0665	45	9	20%	25%	55	12	22%	No	Yes - exceptio	30%	0	57	0	1152		1	92	1	662	0	116	0	1123	21%
Journal of Maritime Archaeology	1557-2293	16	1	6%	11%	18	5	28%	Yes		33%	0	140	1	2801		1	317	2	1778	1	174	4	2251	16%
Transportation Infrastructure Geotechnology	2196-7210	39	1	3%	8%	57	3	5%	No	Yes - exceptio	13%	0	104	1	2032		0	255	0	3481	1	172	0	1916	4%
Animal Cognition	1435-9456	96	27	28%	33%	114	35	31%	No	Yes - exceptio	38%	1	409	1	7421		6	649	6	2420	3	513	4	3612	30%
Multimedia Systems	1432-1882	74	1	1%	6%	136	3	2%	No	Yes - exceptio	11%	1	278	2	4975		3	357	2	2294	3	362	7	17882	2%
Studies in Comparative International Development	1936-6167	20	4	20%	25%	31	7	23%	No	Yes - exceptio	30%	1	749	1	13274		3	509	3	2799	3	516	3	4239	23%
Logica Universalis	1661-8300	24	8	33%	38%	24	4	17%	No	Yes - exceptio	44%	0	124	1	2109		1	215	3	2857	1	168	1	1653	17%
Foundations of Science	1572-8471	68	22	32%	37%	93	39	42%	Yes		48%	0	152	0	2557		2	269	2	5694	1	219	1	1370	40%
PalZ	1867-6812	21	10	48%	55%	54	30	56%	Yes		64%	1	152	1	2315		2	237	4	1186	2	207	2	2074	60%
Origins of Life and Evolution of Biospheres	1573-0875	10	4	40%	46%	10	2	20%	No	Yes - exceptio	53%	0	291	0	4280		2	711	6	1295	2	589	2	1953	29%
Eye	1476-5454	288	34	12%	17%	302	65	22%	Yes		27%	0	188	1	2757		5	585	7	2095	3	472	3	1622	20%
Monash Bioethics Review	1836-6716	15	6	40%	46%	16	2	13%	No	Yes - exceptio	53%	0	287	0	4029		0	333	2	8209	0	732	1	2219	17%
Journal of Logic, Language and Information	1572-9583	10	5	50%	58%	20	10	50%	No	Yes - exceptio	67%	0	132	0	1856		1	386	2	2173	1	238	0	1838	46%
Journal of Indian Philosophy	1573-0395	34	9	26%	31%	34	9	26%	No	Yes - exceptio	36%	0	141	0	1975		0	189	1	887	0	259	1	1921	25%
Journal of Fetal Medicine	2348-8859	24	3	13%	18%	19	1	5%	No	Yes - exceptio	23%	0	53	0	735		0	102	0	772	0	163	0	1640	2%
Health Services and Outcomes Research Methodology	1572-9400	24	3	13%	18%	32	5	16%	No	Yes - exceptio	23%	0	165	1	2220		2	303	1	1904	2	468	1	2335	15%
Linguistics and Philosophy	1573-0549	22	8	36%	42%	28	10	36%	No	Yes - exceptio	48%	0	183	0	2431		1	248	2	652	1	273	2	3674	36%

Relations

ISSN	Total number of research articles published in 2021	Total number of research articles published Open Access in 2021	Year 1 OA penetration rate (Actual, research articles)	TJ target met Yes / No	If OA target not met, has a TJ exception been applied for? [Yes/No]	2022 Target (% of research content published OA)
2366-7540	31	6	19%	Yes		24%
2196-8926	19	1	5%	Yes		10%
2197-9847	69	16	23%	No	Yes - exceptio	32%
1572-8463	12	5	42%	Yes		48%
1573-1103	33	14	42%	Yes		48%
1572-8684	23	6	26%	Yes		31%
1432-0665	55	12	22%	No	Yes - exceptio	30%
1557-2293	18	5	28%	Yes		33%
2196-7210	57	3	5%	No	Yes - exceptio	13%
1435-9456	114	35	31%	No	Yes - exceptio	38%
1432-1882	136	3	2%	No	Yes - exceptio	11%
1936-6167	31	7	23%	No	Yes - exceptio	30%
1661-8300	24	4	17%	No	Yes - exceptio	44%
1572-8471	93	39	42%	Yes		48%
1867-6812	54	30	56%	Yes		64%
1573-0875	10	2	20%	No	Yes - exceptio	53%
1476-5454	302	65	22%	Yes		27%
1836-6716	16	2	13%	No	Yes - exceptio	53%
1572-9583	20	10	50%	No	Yes - exceptio	67%
1573-0395	34	9	26%	No	Yes - exceptio	36%
2348-8859	19	1	5%	No	Yes - exceptio	23%
1572-9400	32	5	16%	No	Yes - exceptio	23%

ISSN	Average number of citations for subscription research articles published in 2021	Average number of downloads for subscription research articles published	Average number of citations for OA research articles published in 2021	Average number of downloads for OA research articles published in 2021
2366-7540	0	224	0	7800
2196-8926	0	91	0	2474
2197-9847	0	71	0	1930
1572-8463	0	173	0	4148
1573-1103	0	267	0	6096
1572-8684	0	142	0	3232
1432-0665	0	57	0	1152
1557-2293	0	140	1	2801
2196-7210	0	104	1	2032
1435-9456	1	409	1	7421
1432-1882	1	278	2	4975
1936-6167	1	749	1	13274
1661-8300	0	124	1	2109
1572-8471	0	152	0	2557
1867-6812	1	152	1	2315
1573-0875	0	291	0	4280
1476-5454	0	188	1	2757
1836-6716	0	287	0	4029
1572-9583	0	132	0	1856
1573-0395	0	141	0	1975
2348-8859	0	53	0	735
1572-9400	0	165	1	2220

ISSN	Average number of citations for subscription research	Average number of downloads for subscription	Average number of citations for OA research articles publish	Average number of downloads for OA research articles	Average number of citations for subscription research	Average number of downloads for subscription	Average number of citations for OA research articles publish	Average number of downloads for OA research articles	Year 1 OA penetration rate (Actual) (all article types
2366-7540	0	265	0	2998	0	219	1	1864	18%
2196-8926	1	422	0	0	0	478	0	0	4%
2197-9847	3	175	3	1166	2	136	1	730	23%
1572-8463	1	306	0	4502	0	222	1	7911	24%
1573-1103	1	408	16	8279	1	346	1	3130	36%
1572-8684	1	319	3	3496	0	357	1	2571	19%
1432-0665	1	92	1	662	0	116	0	1123	21%
1557-2293	1	317	2	1778	1	174	4	2251	16%
2196-7210	0	255	0	3481	1	172	0	1916	4%
1435-9456	6	649	6	2420	3	513	4	3612	30%
1432-1882	3	357	2	2294	3	362	7	17882	2%
1936-6167	3	509	3	2799	3	516	3	4239	23%
1661-8300	1	215	3	2857	1	168	1	1653	17%
1572-8471	2	269	2	5694	1	219	1	1370	40%
1867-6812	2	237	4	1186	2	207	2	2074	60%
1573-0875	2	711	6	1295	2	589	2	1953	29%
1476-5454	5	585	7	2095	3	472	3	1622	20%
1836-6716	0	333	2	8209	0	732	1	2219	17%
1572-9583	1	386	2	2173	1	238	0	1838	46%
1573-0395	0	189	1	887	0	259	1	1921	25%
2348-8859	0	102	0	772	0	163	0	1640	2%
1572-9400	2	303	1	1904	2	468	1	2335	15%

Relations

Lancaster University

- To answer queries like:
  - “Tell me a Commander that can pilot Space_Shuttle Falcon”.
- We need to **relate** entity sets logically.
- A **Relation** is a logical model for associating two or more **Entities**.
  - *Commander Turker can pilot Falcon.*
- A **Relationship set** is the set of all these relations.

Key concepts of ER: Relation

Lancaster University

- Relating two or more entity sets requires the designer to set **a mapping between the entities belonging the Entity Sets under Relation.**

Can more than one mechanic repair the same car?
Can a mechanic repair more than one car?

We will see…

Model Entity set

Brand	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210

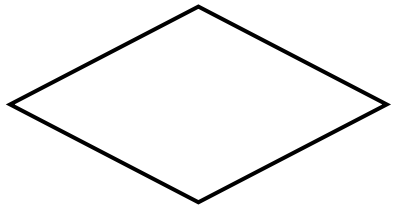
Mechanic Entity set

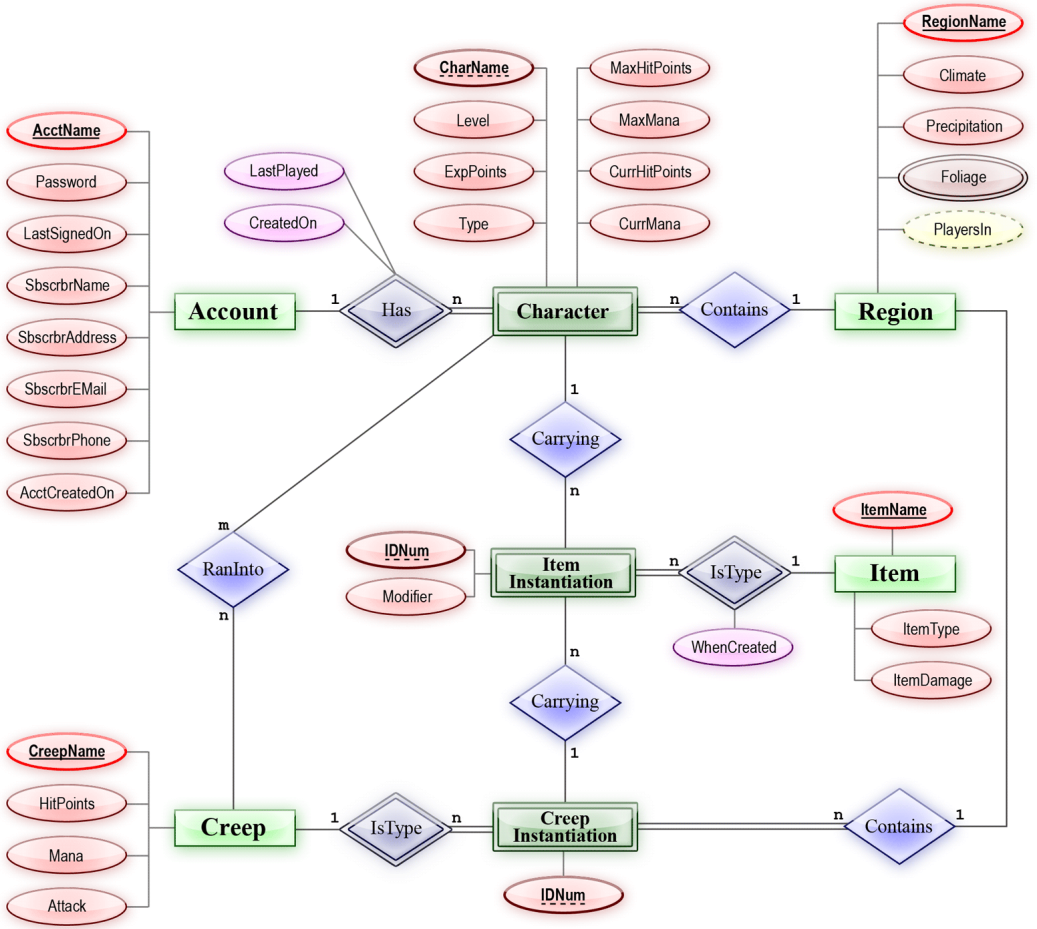
NIN	Name	Phone_Number
87542702	Tom	75315567, 75315264
68201937	Uraz	75335521, 75334567
23139827	Nick	75315544, 75315237

Key concepts of ER: Relation

Lancaster University

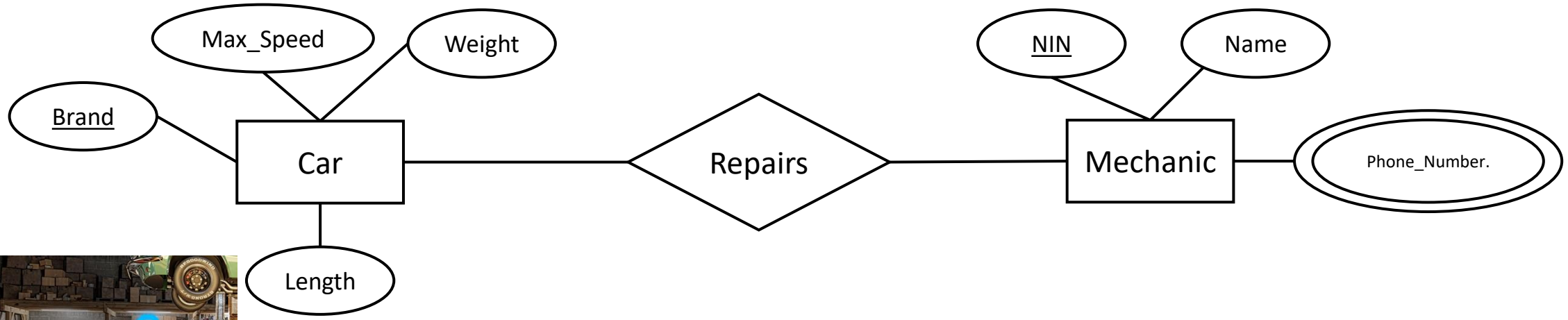
A diamond represents a **Relationship set**.





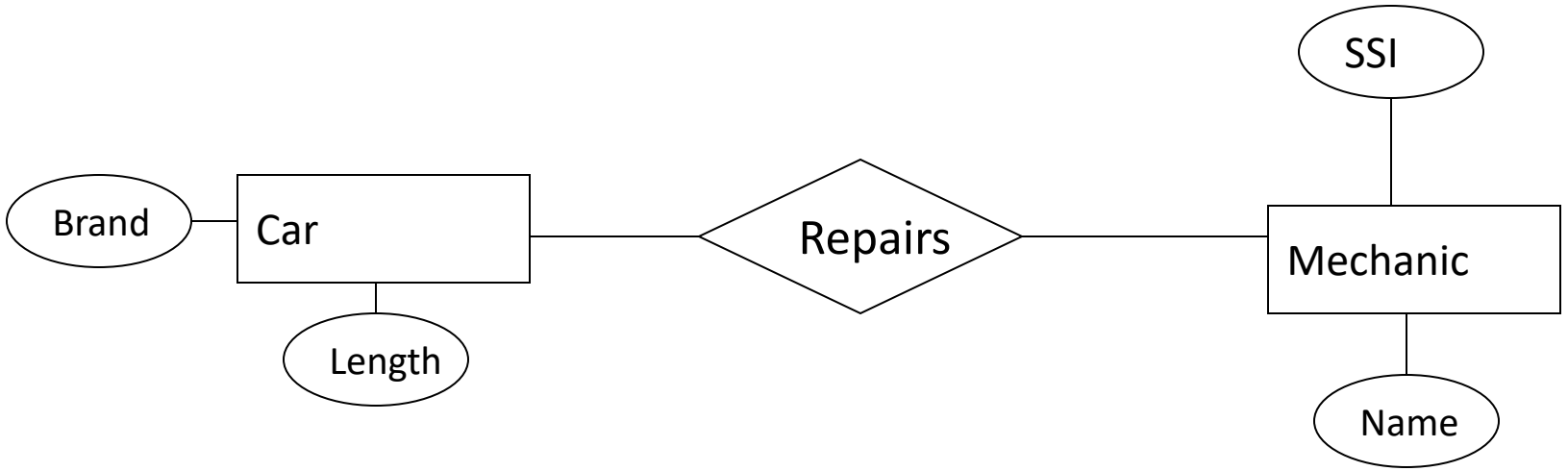
Key concepts of ER: Relation

A relationship set ‘Repairs’ relates two Entity Sets.



ER Model

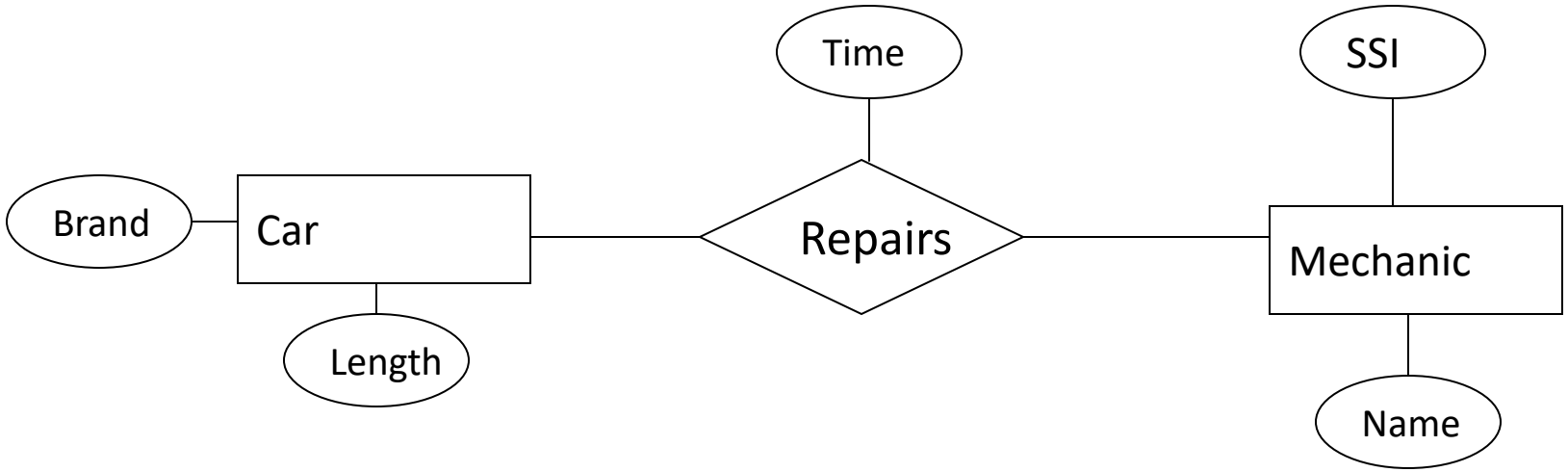
Lancaster University



- Relationship sets may also have attributes.

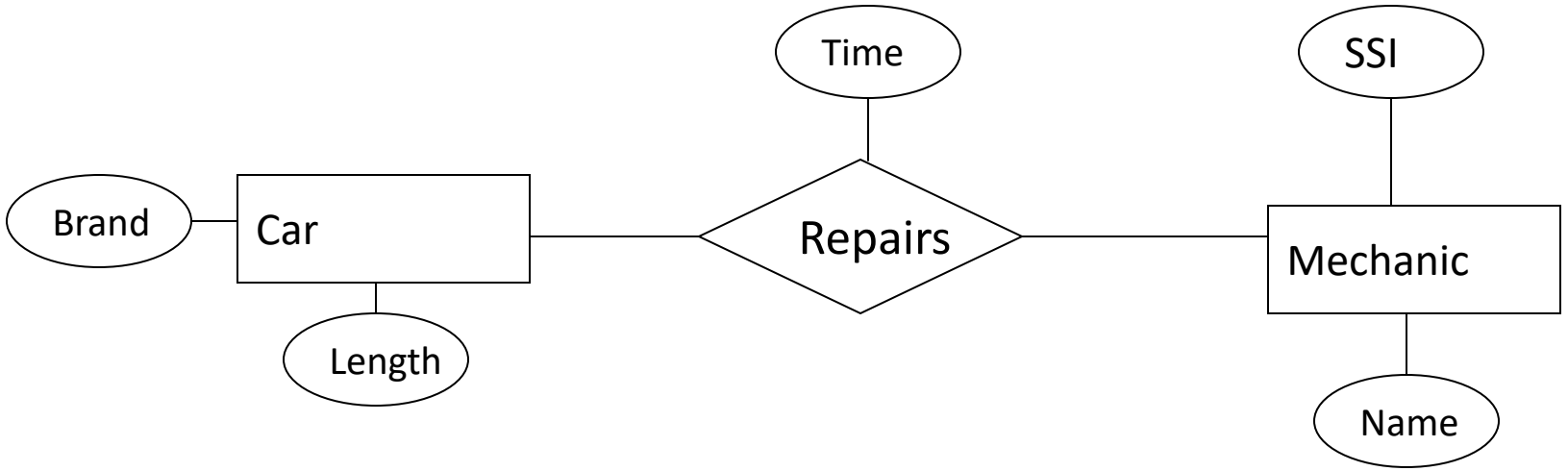
ER Model

Lancaster University



- Relationship sets may also have attributes.
- We will talk about the key in a relationship set later.

ER Model

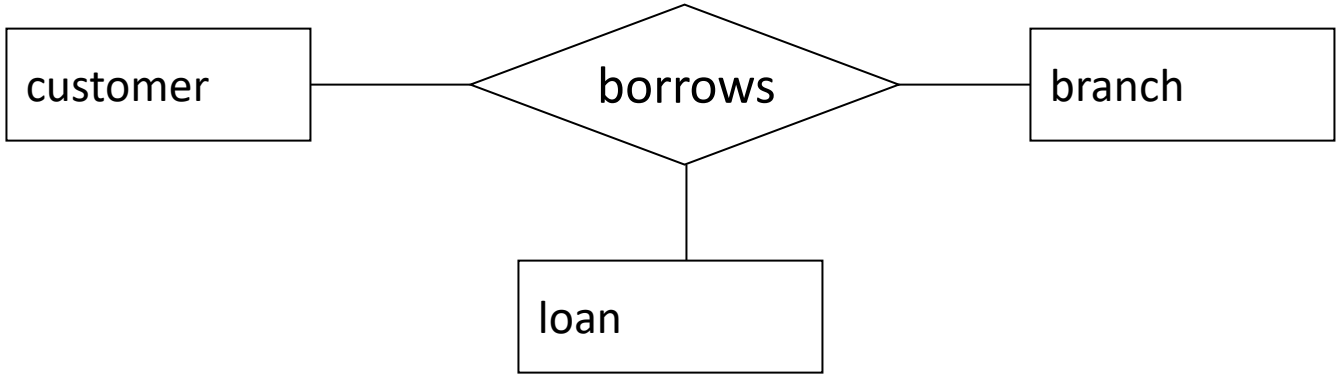


- **Degree of a relationship set** is the number of different entity sets that participate in a relationship
- **Binary relationship** sets involve two entity sets

ER Model

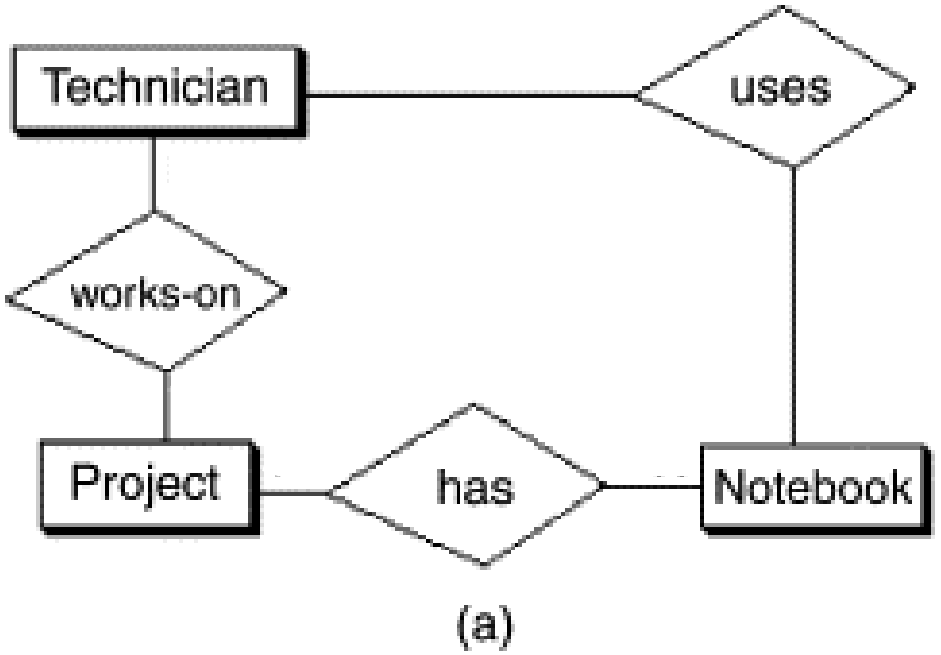
Lancaster University

- Ternary relationship sets involve three entity sets.



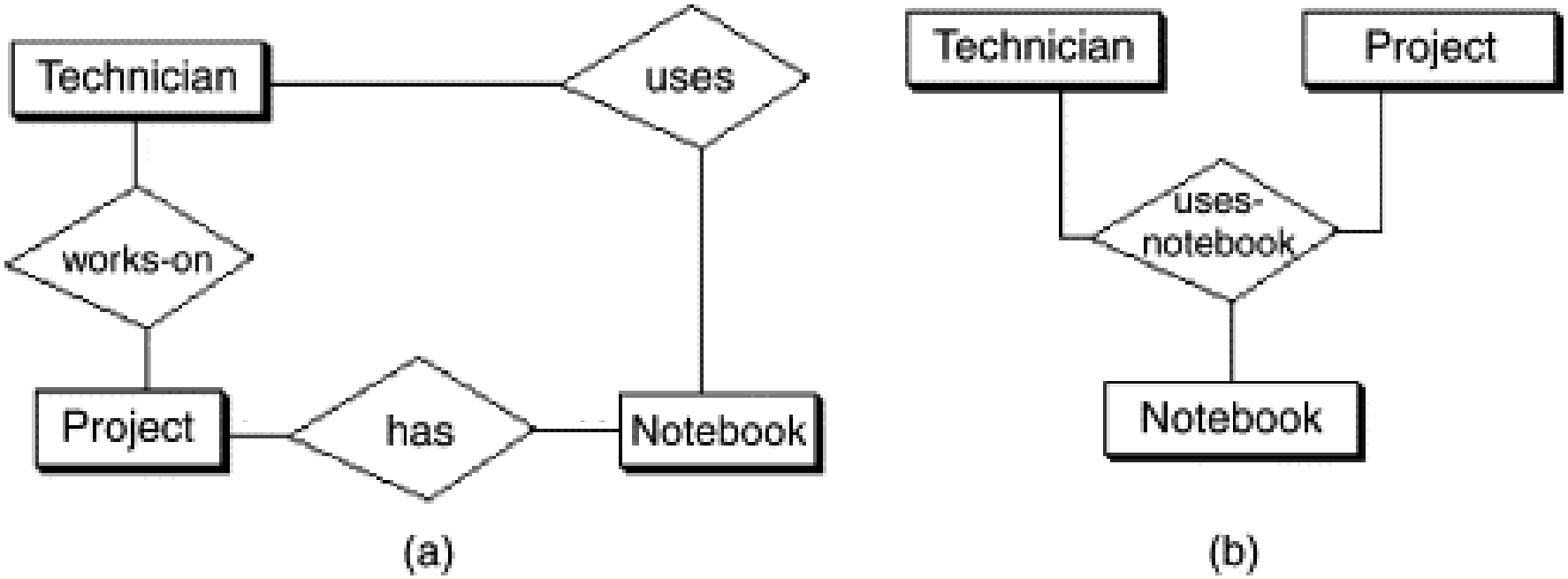
Ternary relationship sets.

Lancaster University



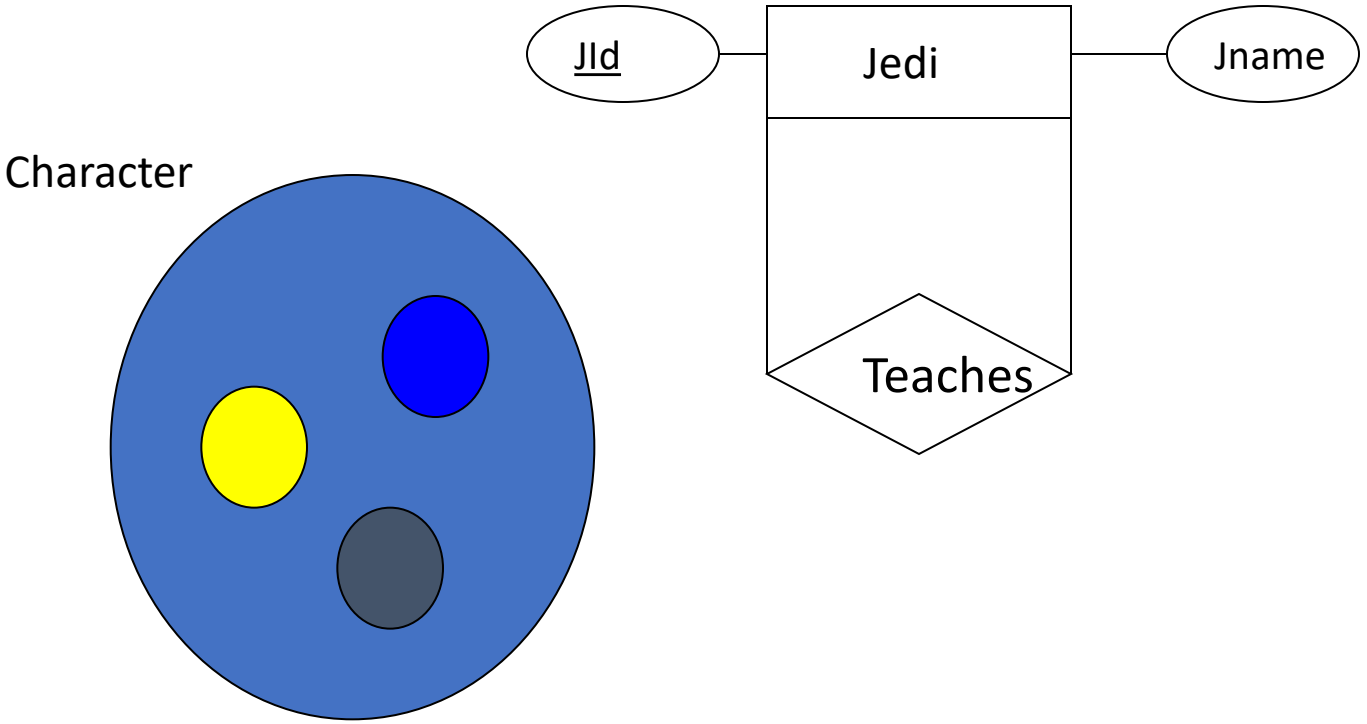
(a)

Ternary relationship sets.



ER Model

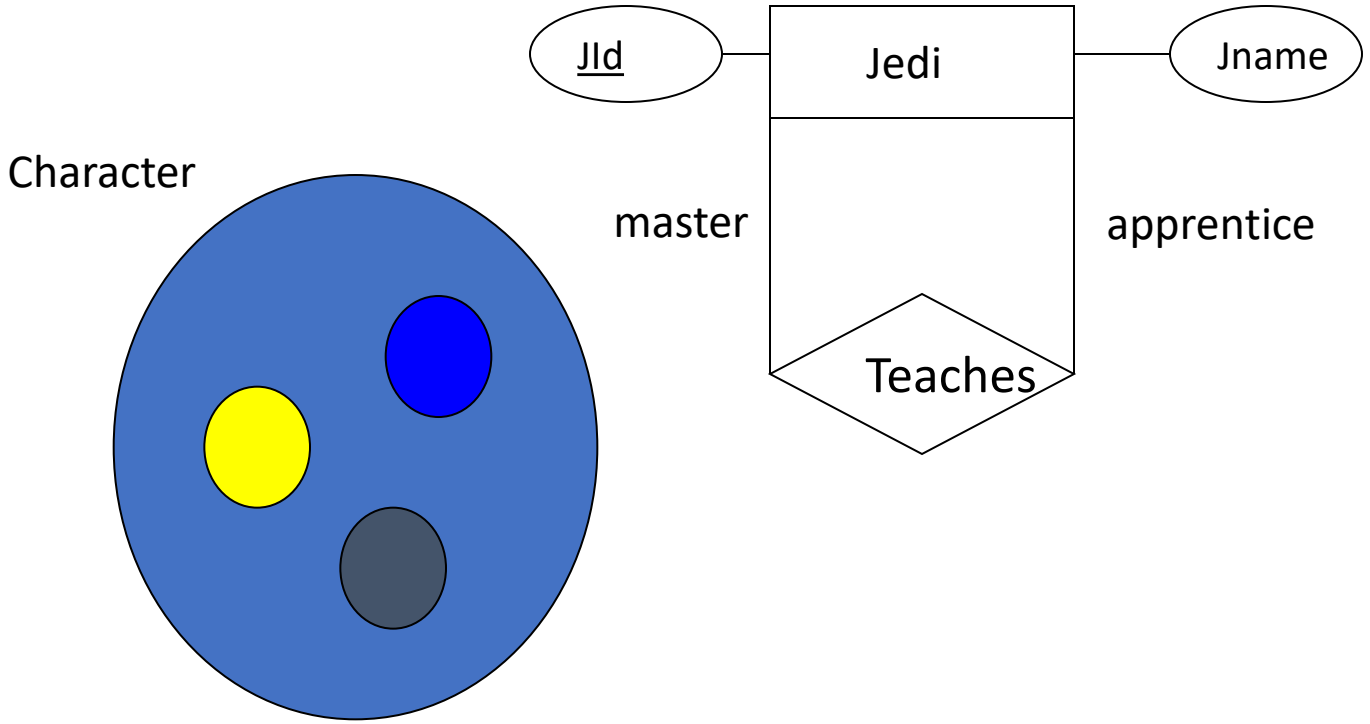
- We may have relationships among the entities that belong to the same entity set
- each entity has a role in such a relationship



ER Model

Lancaster University

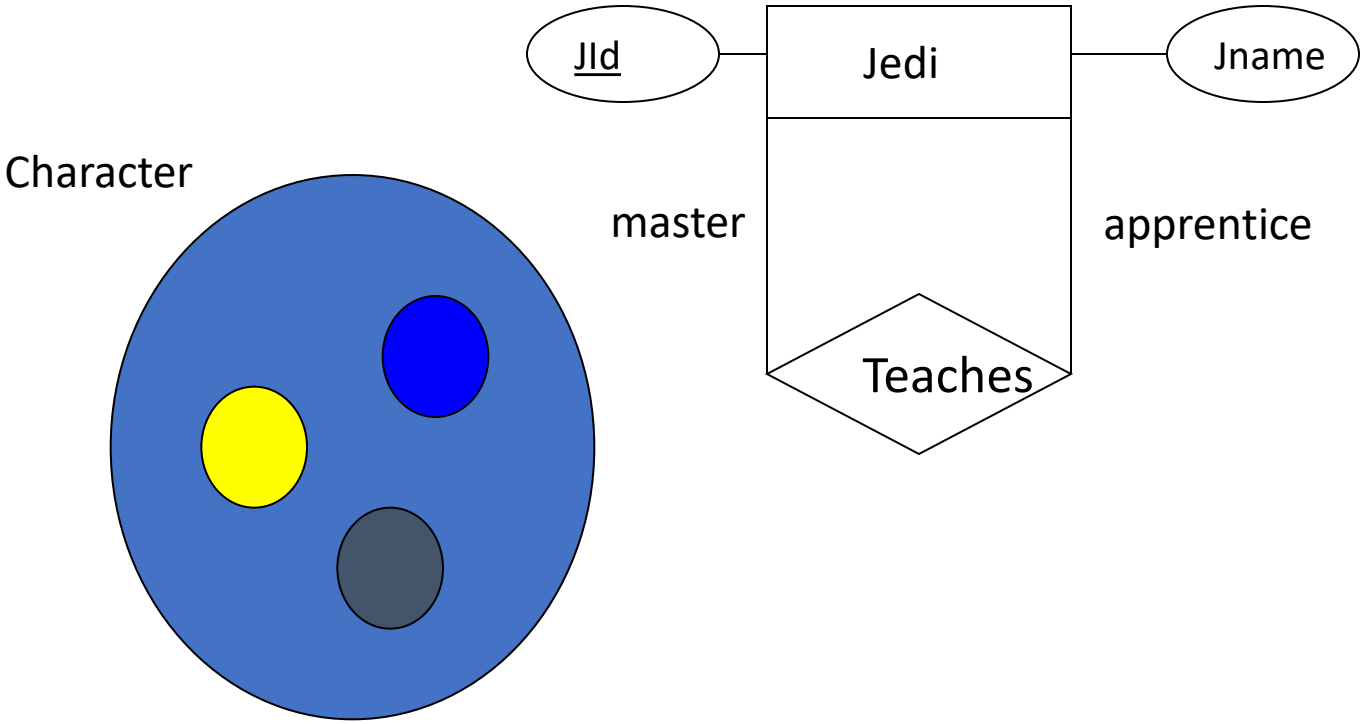
- We may have relationships among the entities that belong to the same entity set
- **each entity has a role in such a relationship.**



ER Model

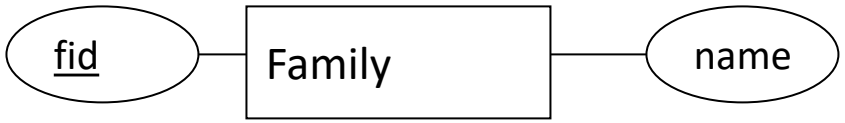
Lancaster University

- We may have relationships among the entities that belong to the same entity set
- each entity has a role in such a relationship.
- **What is the degree of the following relationship set (unary, binary, or ternary)?**



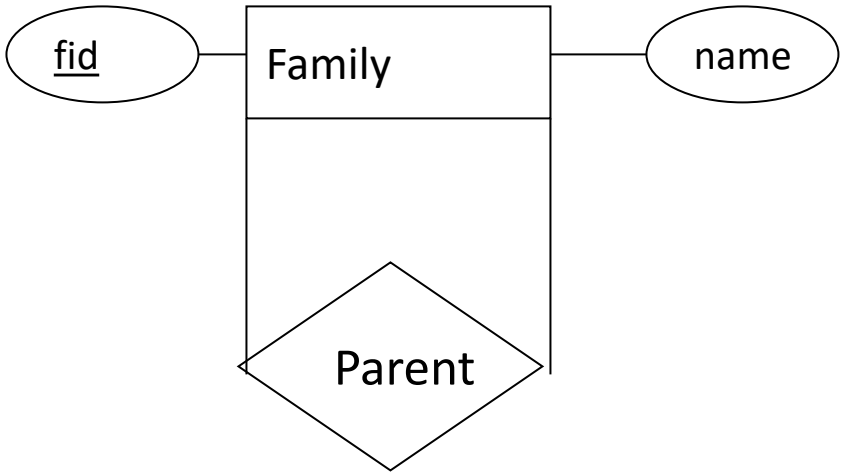
ER Model

Lancaster University



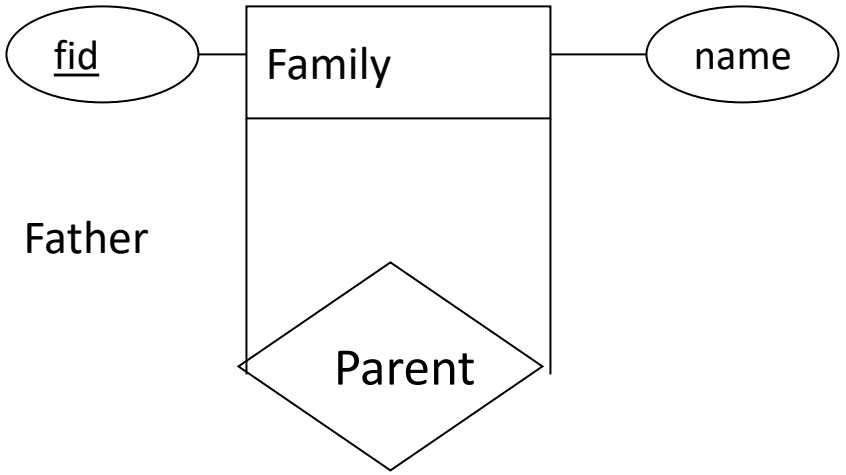
ER Model

Lancaster University



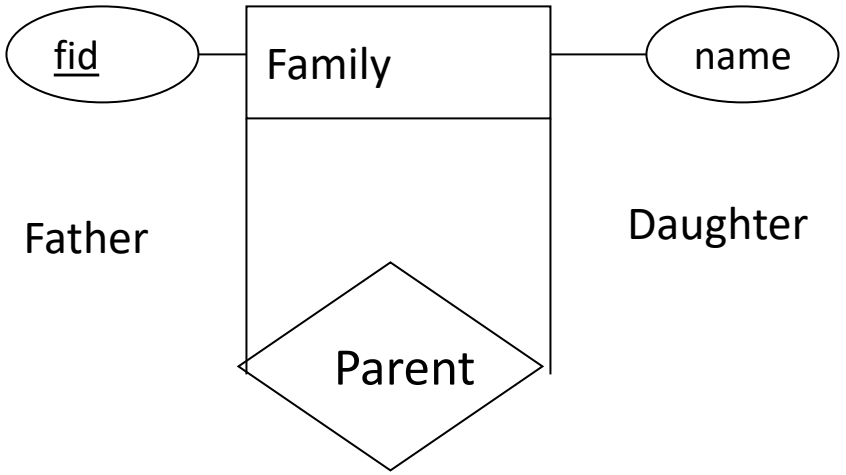
ER Model

Lancaster University



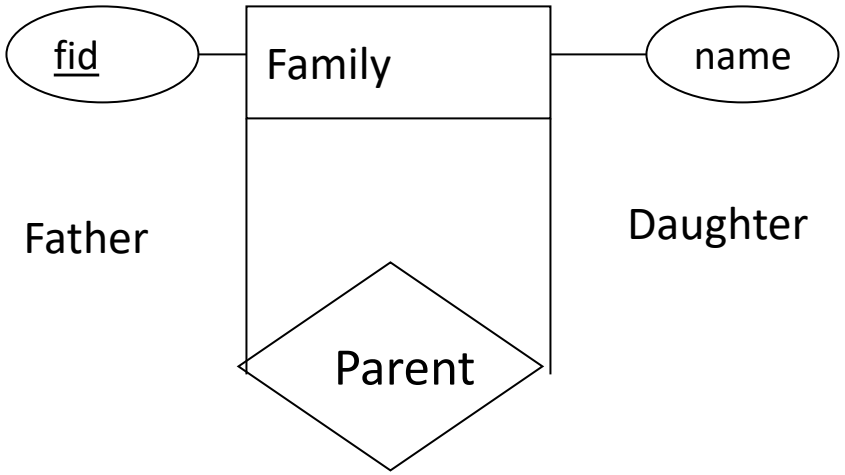
ER Model

Lancaster University



ER Model

Fid	Name	Name
12	Mr Bean	N/A
21	Thomas Shelby	Roby Shelby
55	Thanos	Gamora



Key concepts of ER: Cardinalities.

Lancaster University

VERY IMPORTANT CONTENT!!!!!

Model Entity set

Brand	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210

Mechanic Entity set

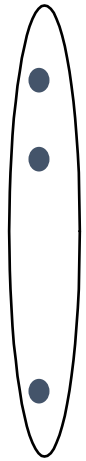
NIN	Name	Phone_Number
87542702	Tom	75315567, 75315264
68201937	Uraz	75335521, 75334567
23139827	Nick	75315544, 75315237

Given an entity (E) from one entity set, what is the relation of this entity with the entities in the other entity sets?
Can more than one mechanic repair BMW 3.21?
Can Tom repair more than one type of car?

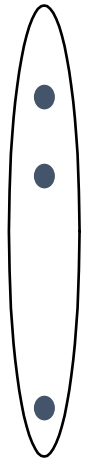
Key concepts of ER: Cardinalities.

Lancaster University

Brand	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210



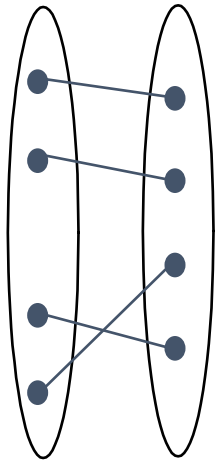
NIN	Name	Phone_Number
87542702	Tom	75315567, 75315264
68201937	Uraz	75335521, 75334567
23139827	Nick	75315544, 75315237



Mapping cardinalities

Lancaster University

- Recall: Relating two or more entity sets requires the designer to set **a mapping between the entities belonging the Entity Sets under Relation.**
- One-to-One relationship (ex: marriage relationship set between husbands and wives)



1-to-1

SSI	Name	Phone_Number
87542702	Tom	75315567, 75315264
68201937	Uraz	75335521, 75334567
23139827	Nick	75315544, 75315237

Brand	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210

- Relationship sets are created while creating tables in the relationship.
- This sometimes leads to new tables (as in this case) or leads to the addition of new attributes in the tables in the relationship (you will see)

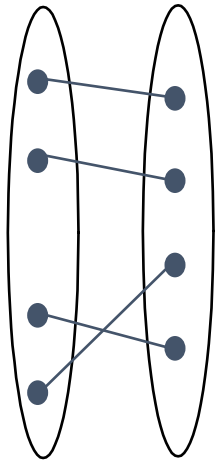
Mapping cardinalities

Lancaster University

- Recall: Relating two or more entity sets requires the designer to set **a mapping between the entities belonging the Entity Sets under Relation.**
- One-to-One relationship (ex: marriage relationship set between husbands and wives)

Name	Brand
Tom	Toyota..
Nick	Hyundai.
Uraz	BMW..

Relationship Set



1-to-1

SSI	Name	Phone_Number
87542702	Tom	75315567, 75315264
68201937	Uraz	75335521, 75334567
23139827	Nick	75315544, 75315237

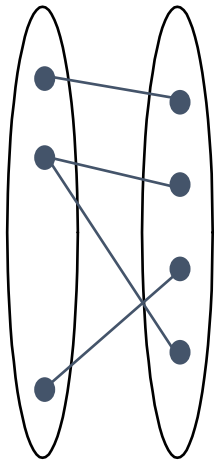
Brand	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210

Mapping cardinalities

- One-to-One (ex: marriage relationship set between husbands and wives)
- One-to-Many (ex?)

Name	Brand
Uraz	BMW
Uraz	Hyundai
Tom	Toyota

Relationship Set



1-to Many

NIN	Name	Phone_Number
87542702	Tom	75315567, 75315264
68201937	Uraz	75335521, 75334567
23139827	Nick	75315544, 75315237

Brand	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210

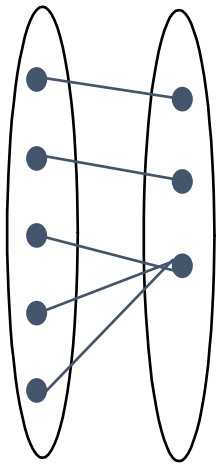
Mapping cardinalities

Lancaster University

- One-to-One (ex: marriage relationship set between husbands and wives)
- Many-to-One (ex?)

Name	Brand
Uraz	BMW
Nick	Toyota
Tom	BMW

Relationship Set



Many-to-1

NIN	Name	Phone_Number
87542702	Tom	75315567, 75315264
68201937	Uraz	75335521, 75334567
23139827	Nick	75315544, 75315237

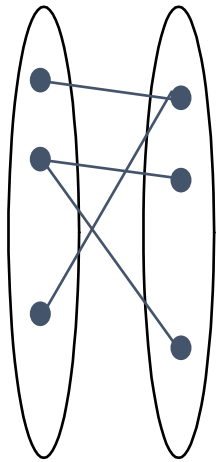
Brand	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210

Mapping cardinalities

- One-to-One (ex: marriage relationship set between husbands and wives)
- Many-to-Many (ex?)

Name	Brand
Uraz	BMW
Uraz	Hyundai
Nick	Toyota
Nick	BMW

Relationship Set



Many-to-Many

NIN	Name	Phone_Number
87542702	Tom	75315567, 75315264
68201937	Uraz	75335521, 75334567
23139827	Nick	75315544, 75315237

Brand	Weight	Length	Max_Speed
BMW 3.21	1400	3.21	200
Toyota_Corolla	1300	3.18	200
Hyundai E.GLS	1400	3.16	210