
Week 12

Database Systems -

Introduction to Databases and Data Warehouses

CHAPTER 4 - Update Operations, Update Anomalies, and Normalization (Part 5)

NORMALIZATION

■ Other Normal Forms

- Normalizing to 3NF
 - Sufficient for most real-world problems
- Higher normal forms beyond 3NF
 - Based on other functional dependencies
 - Mostly for theoretical study
 - In Appendix B

NORMALIZATION

■ Normalization Exceptions

- Typically normalize database relations to 3NF to eliminate unnecessary data redundancy and avoid update anomalies
- But normalization to 3NF should be done judiciously and pragmatically
 - Possible to deliberately not normalizing certain relations to 3NF for some benefit



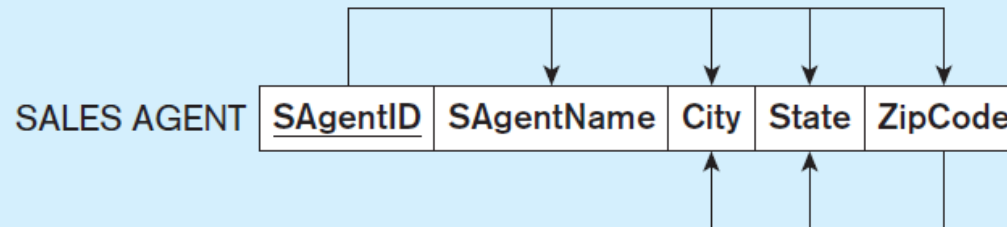
Normalization Exception Example: relation SALES AGENT

Relation SALES AGENT

SALES AGENT

<u>SAgentID</u>	SAgentName	City	State	ZipCode
SA1	Rose	Glen Ellyn	IL	60137
SA2	Sidney	Chicago	IL	60611
SA3	James	Chicago	IL	60610
SA4	Violet	Wheaton	IL	60187
SA5	Nicole	Kenosha	WI	53140
SA6	Justin	Milwaukee	WI	53201

Functional dependencies in the relation SALES AGENT (in 2NF)





Normalization Exception Example:

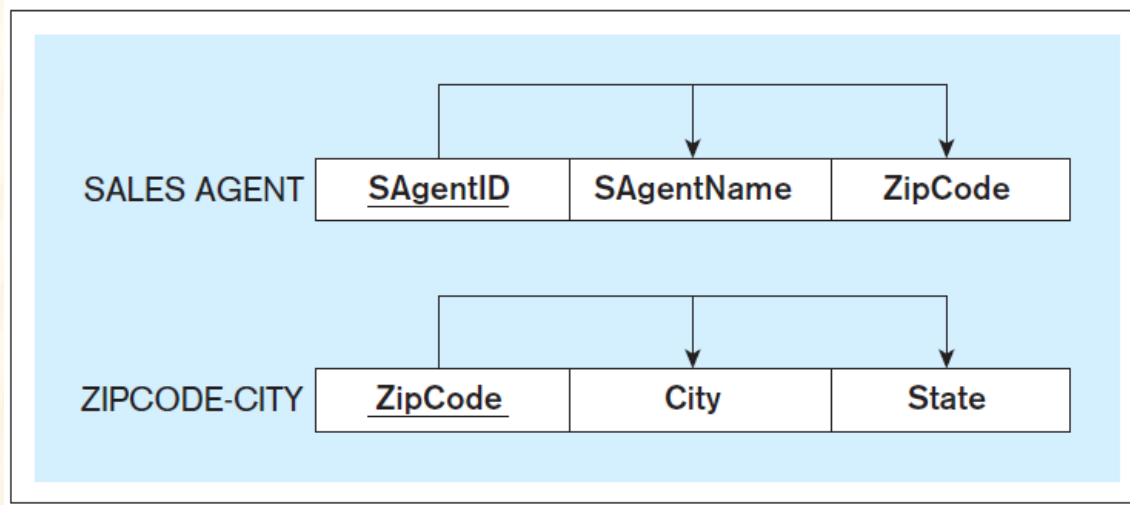
— Should relation SALES AGENT be further normalized?

SALES AGENT Example in 3NF

+ Removed transitive FD set:

ZipCode -> City, State

+ **Not needed if not many sales agents are in same zip code**





NORMALIZATION

▪ Denormalization

- **Reverse** the effect of normalization by **joining normalized relations into a relation that is not normalized**, in order to improve **query** performance
 - **Normalization spreads data** across more relations
 - Slow down data retrievals
- Can use denormalization to deal with **normalization vs. performance** issue
 - Example in next slide
 - For relational databases with **few updates but many retrievals**
 - **Keep a normalized master version** of the database for update
 - **Periodically create a physical denormalized copy** of the same database from the master copy for data retrievals
 - Used in data warehouses and materialized views

Denormalization Example

– Pressly Ad Agency Example

Popular data query faster in ad campaign mix table

- ✖ For each mode in each ad campaign, print the campaign's ID, name, mgr ID, mgr name and mode's ID, media, range, and budget percentage.

RETRIEVED DATA

<i>AdCampaignID</i>	<i>AdCampaignName</i>	<i>Campaign MgrID</i>	<i>Campaign MgrName</i>	<i>ModeID</i>	<i>Media</i>	<i>Range</i>	<i>BudgetPctg</i>
111	SummerFun13	CM100	Roberta	1	TV	Local	50%
111	SummerFun13	CM100	Roberta	2	TV	National	50%
222	SummerZing13	CM101	Sue	1	TV	Local	60%
222	SummerZing13	CM101	Sue	3	Radio	Local	30%
222	SummerZing13	CM101	Sue	5	Print	Local	10%
333	FallBall13	CM102	John	3	Radio	Local	80%
333	FallBall13	CM102	John	4	Radio	National	20%
444	AutmnStyle13	CM103	Nancy	6	Print	National	100%
555	AutmnColors13	CM100	Roberta	3	Radio	Local	100%

Denormalization Example

Normalized Master Copy of Pressly Ad Agency Database

AD CAMPAIGN

<u>AdCampaignID</u>	AdCampaignName	StartDate	Duration	CampaignMgrID
111	SummerFun13	6.6.2013	12 days	CM100
222	SummerZing13	6.8.2013	30 days	CM101
333	FallBall13	6.9.2013	12 days	CM102
444	AutmnStyle13	6.9.2013	5 days	CM103
555	AutmnColors13	6.9.2013	3 days	CM100

CAMPAIGN MANAGER

<u>CampaignMgrID</u>	CampaignMgrName
CM100	Roberta
CM101	Sue
CM102	John
CM103	Nancy

MODE

<u>ModelID</u>	Media	Range
1	TV	Local
2	TV	National
3	Radio	Local
4	Radio	National
5	Print	Local
6	Print	National

AD CAMPAIGN MIX

<u>AdCampaignID</u>	<u>ModelID</u>	BudgetPctg
111	1	50%
111	2	50%
222	1	60%
222	3	30%
222	5	10%
333	3	80%
333	4	20%
444	6	100%
555	3	100%

Denormalization Example

Physical DeNormalized Copy of Pressly Ad Agency Database

AD CAMPAIGN MIX

<u>AdCampaignID</u>	AdCampaignName	StartDate	Duration	Campaign MgrID	Campaign MgrName	<u>ModelID</u>	Media	Range	BudgetPctg
111	SummerFun13	6.6.2013	12 days	CM100	Roberta	1	TV	Local	50%
111	SummerFun13	6.6.2013	12 days	CM100	Roberta	2	TV	National	50%
222	SummerZing13	6.8.2013	30 days	CM101	Sue	1	TV	Local	60%
222	SummerZing13	6.8.2013	30 days	CM101	Sue	3	Radio	Local	30%
222	SummerZing13	6.8.2013	30 days	CM101	Sue	5	Print	Local	10%
333	FallBall13	6.9.2013	12 days	CM102	John	3	Radio	Local	80%
333	FallBall13	6.9.2013	12 days	CM102	John	4	Radio	National	20%
444	AutmnStyle13	6.9.2013	5 days	CM103	Nancy	6	Print	National	100%
555	AutmnColors13	6.9.2013	3 days	CM100	Roberta	3	Radio	Local	100%



NORMALIZATION

▪ Denormalization

- **Not** a default process to be **undertaken in all circumstances**
- Should be **used judiciously**, after analyzing its costs and benefits

NORMALIZATION

■ ER-Modeling versus Normalization

- ER modeling followed by mapping into a relational schema is one of the most common database design methods
- When faced with a non-normalized table, instead of identifying functional dependencies and going through normalization to 2NF and 3NF, a designer can
 - Analyze the table (interact with users etc) and
 - Extract database system requirements (textual description) and
 - Create an ER diagram based on it and m
 - Map the ER diagram into a relational schema



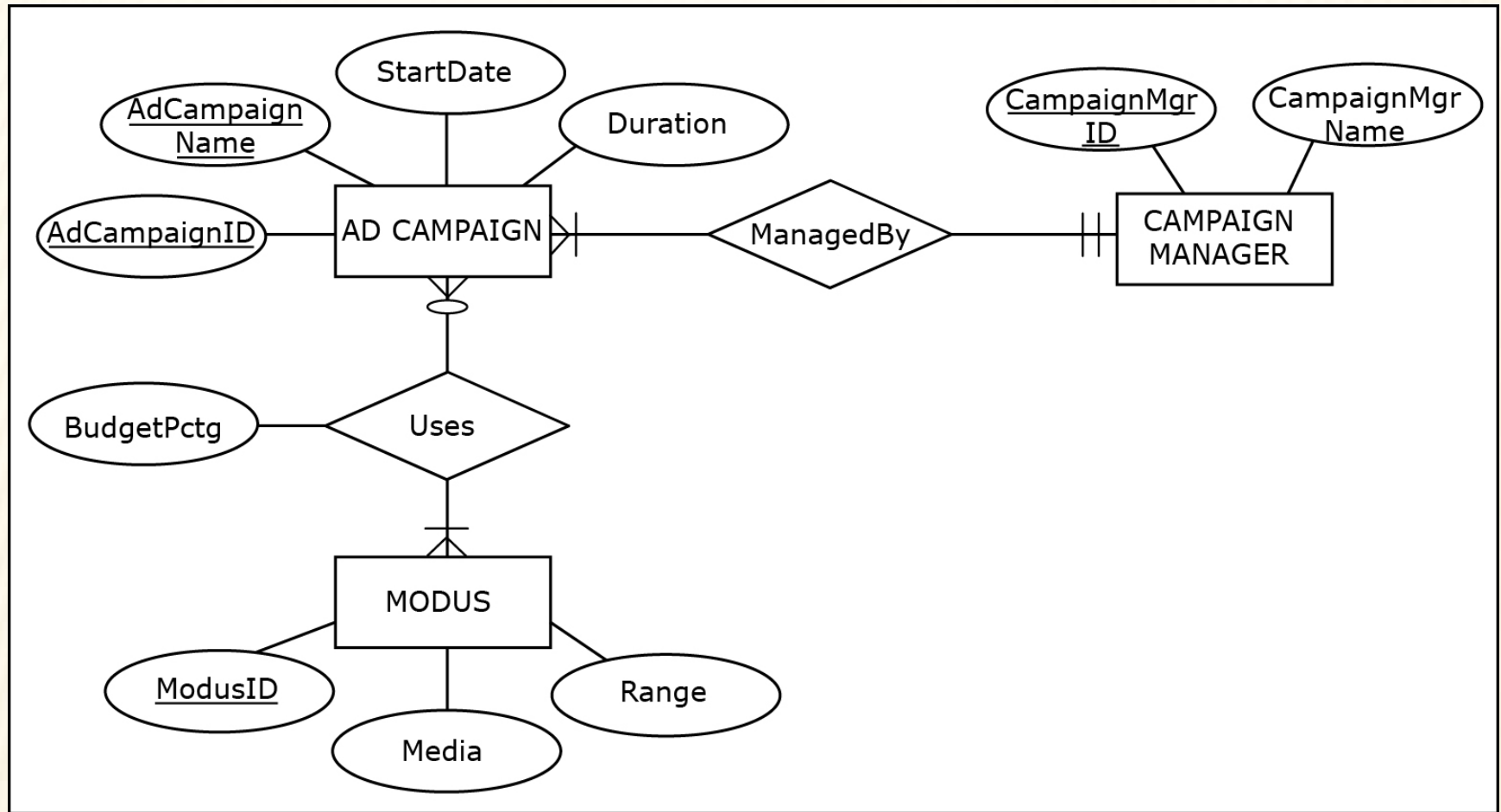
Pressly Ad Agency Database

relation AD CAMPAIGN MIX

AD CAMPAIGN MIX

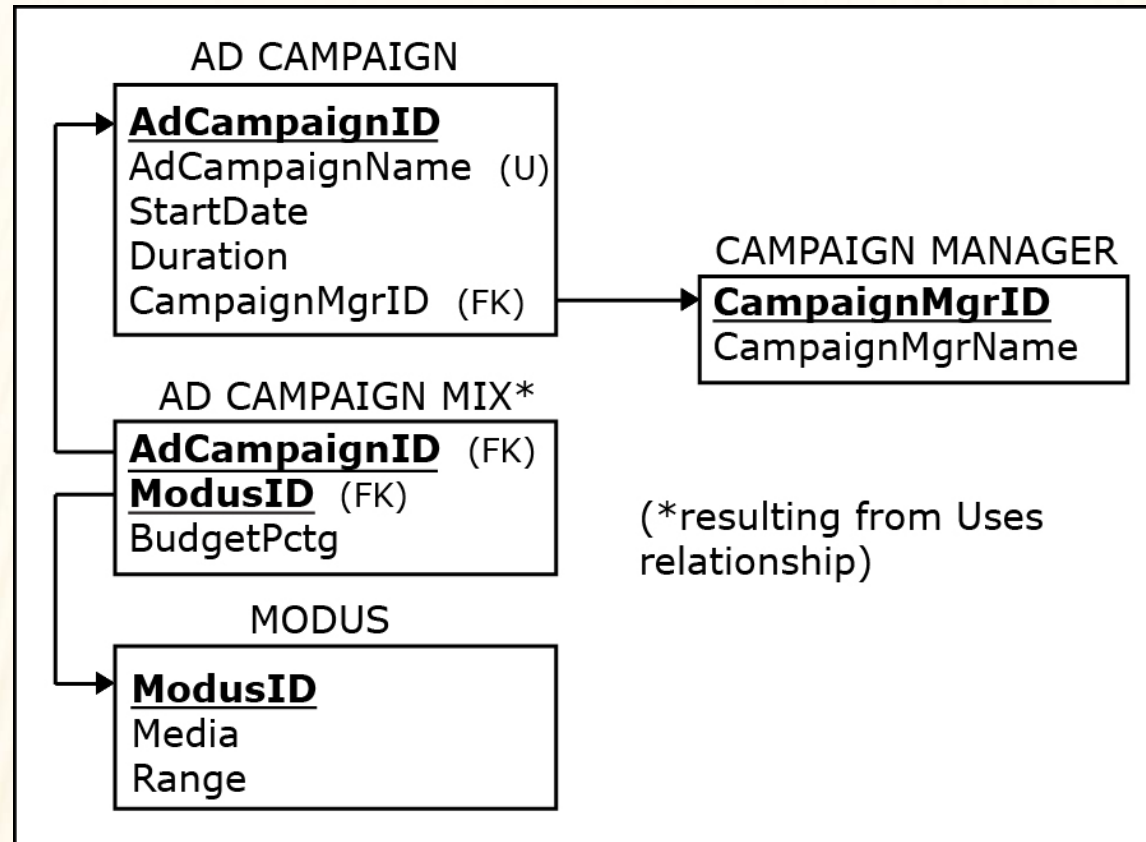
<u>AdCampaignID</u>	AdCampaignName	StartDate	Duration	Campaign MgrID	Campaign MgrName	<u>ModelID</u>	Media	Range	BudgetPctg
111	SummerFun13	6.6.2013	12 days	CM100	Roberta	1	TV	Local	50%
111	SummerFun13	6.6.2013	12 days	CM100	Roberta	2	TV	National	50%
222	SummerZing13	6.8.2013	30 days	CM101	Sue	1	TV	Local	60%
222	SummerZing13	6.8.2013	30 days	CM101	Sue	3	Radio	Local	30%
222	SummerZing13	6.8.2013	30 days	CM101	Sue	5	Print	Local	10%
333	FallBall13	6.9.2013	12 days	CM102	John	3	Radio	Local	80%
333	FallBall13	6.9.2013	12 days	CM102	John	4	Radio	National	20%
444	AutmnStyle13	6.9.2013	5 days	CM103	Nancy	6	Print	National	100%
555	AutmnColors13	6.9.2013	3 days	CM100	Roberta	3	Radio	Local	100%

ER-Modeling vs. Normalization example - ER diagram of the Pressly Ad Agency





ER-Modeling vs. Normalization example - mapped ER diagram, identical to the relational schema resulting from the normalization process (the relational schema of 3NF relations for the normalized Pressly Ad Agency example)



ADDITIONAL STREAMLINING OF DATABASE CONTENT

- **Designer-added entities (tables) and keys**
 - Even if a relation is in 3NF additional opportunities for streamlining database content may still exist
 - **Designer-added entities (tables) and designer-added keys** can be used for additional streamlining



Normalized Pressly Ad Agency Database

AD CAMPAIGN

<u>AdCampaignID</u>	AdCampaignName	StartDate	Duration	CampaignMgrID
111	SummerFun13	6.6.2013	12 days	CM100
222	SummerZing13	6.8.2013	30 days	CM101
333	FallBall13	6.9.2013	12 days	CM102
444	AutmnStyle13	6.9.2013	5 days	CM103
555	AutmnColors13	6.9.2013	3 days	CM100

CAMPAIGN MANAGER

<u>CampaignMgrID</u>	CampaignMgrName
CM100	Roberta
CM101	Sue
CM102	John
CM103	Nancy

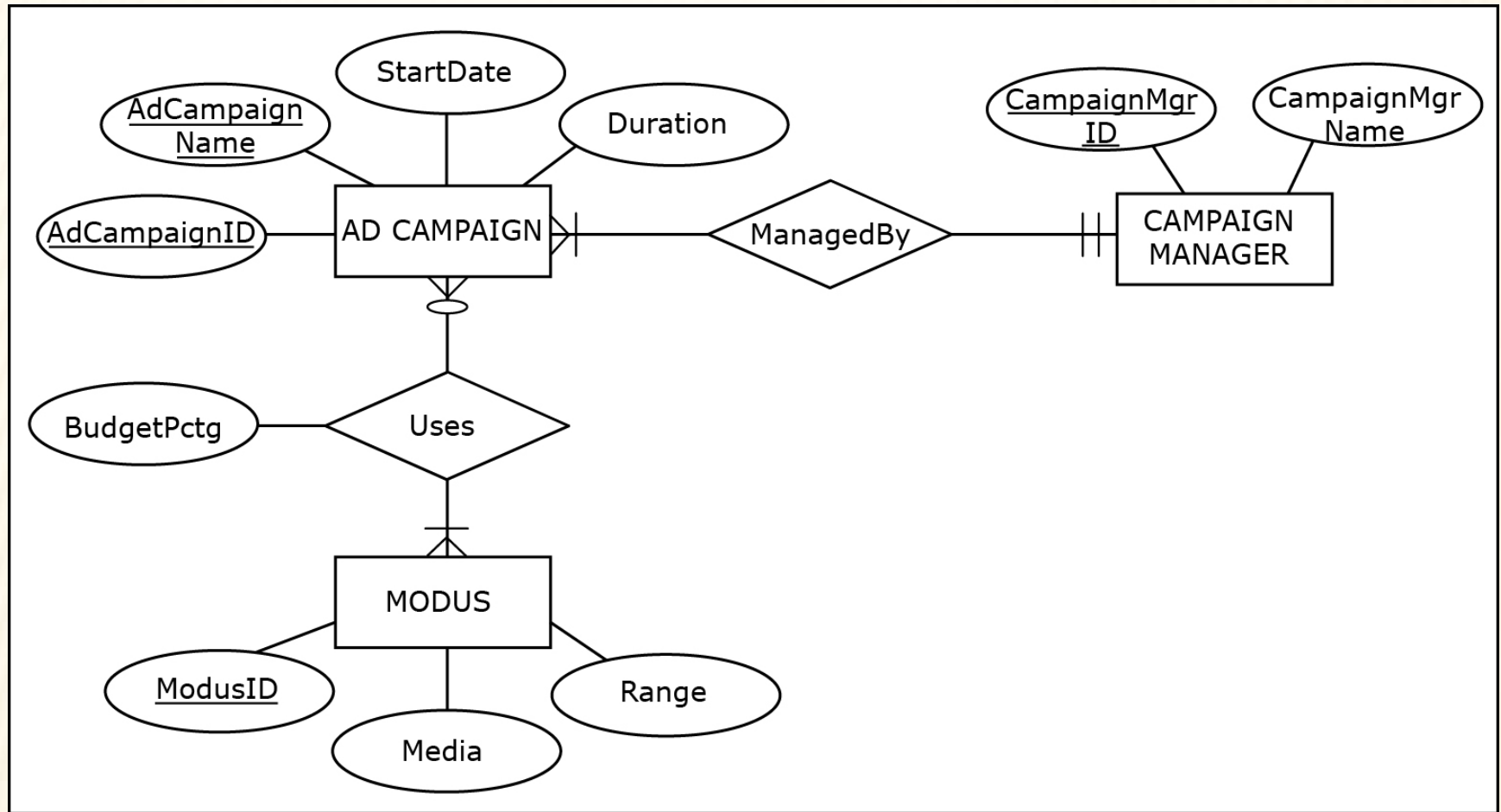
MODE

<u>ModelID</u>	Media	Range
1	TV	Local
2	TV	National
3	Radio	Local
4	Radio	National
5	Print	Local
6	Print	National

AD CAMPAIGN MIX

<u>AdCampaignID</u>	<u>ModelID</u>	BudgetPctg
111	1	50%
111	2	50%
222	1	60%
222	3	30%
222	5	10%
333	3	80%
333	4	20%
444	6	100%
555	3	100%

ER-Modeling vs. Normalization example - ER diagram of the Pressly Ad Agency



Normalized Pressly Ad Agency Database

AD CAMPAIGN

<u>AdCampaignID</u>	AdCampaignName	StartDate	Duration	CampaignMgrID
111	SummerFun13	6.6.2013	12 days	CM100
222	SummerZing13	6.8.2013	30 days	CM101
333	FallBall13	6.9.2013	12 days	CM102
444	AutmnStyle13	6.9.2013	5 days	CM103
555	AutmnColors13	6.9.2013	3 days	CM100

CAMPAIGN MANAGER

<u>CampaignMgrID</u>	CampaignMgrName
CM100	Roberta
CM101	Sue
CM102	John
CM103	Nancy

MODE

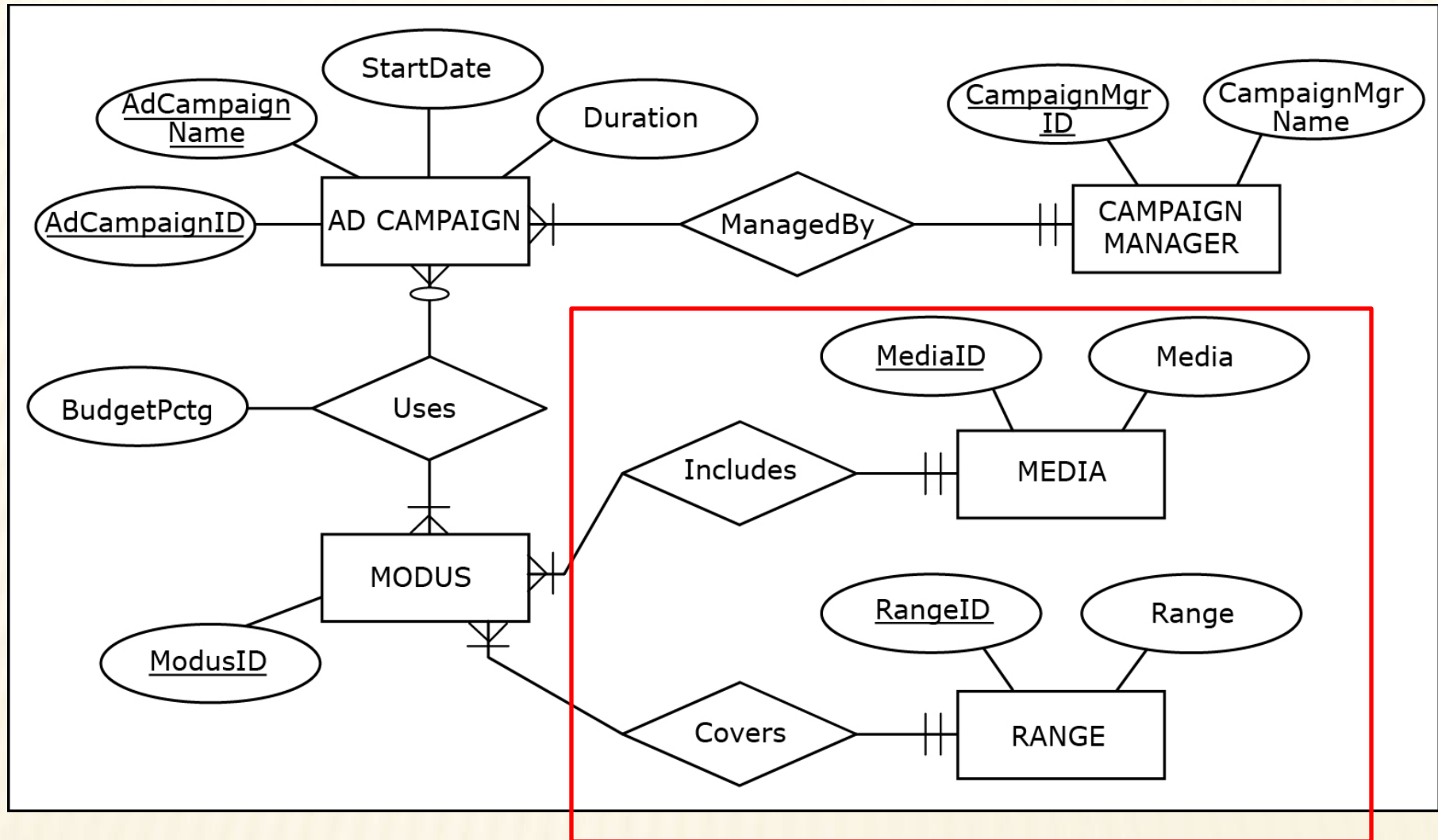
<u>ModelID</u>	Media	Range
1	TV	Local
2	TV	National
3	Radio	Local
4	Radio	National
5	Print	Local
6	Print	National

AD CAMPAIGN MIX

<u>AdCampaignID</u>	<u>ModelID</u>	BudgetPctg
111	1	50%
111	2	50%
222	1	60%
222	3	30%
222	5	10%
333	3	80%
333	4	20%
444	6	100%
555	3	100%

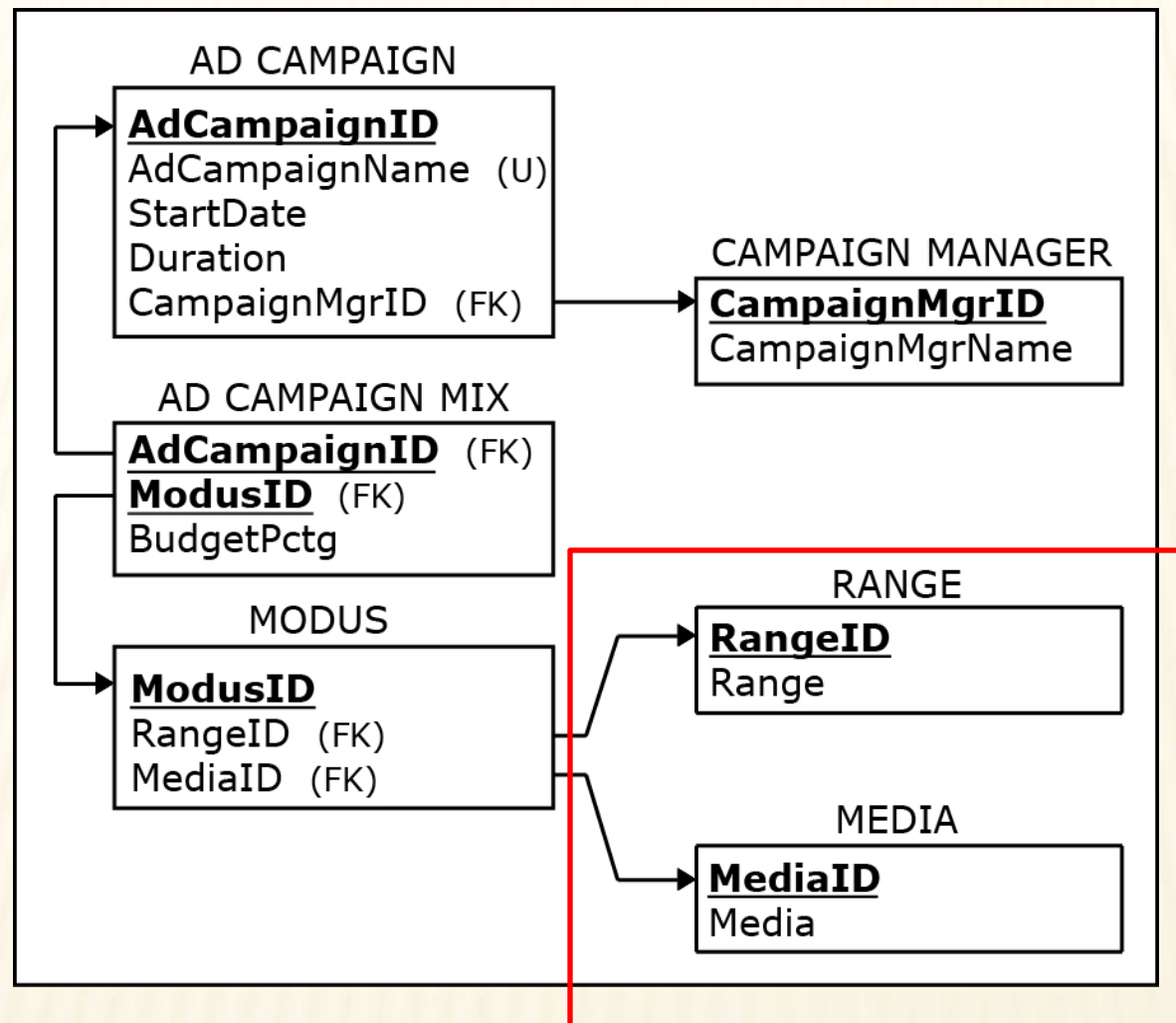


Designer-added entities (tables) and keys example - augmented ER diagram of the Pressly Ad Agency example





Designer-added entities (tables) and keys example - augmented ER diagram of the Pressly Ad Agency example mapped into a relational schema



Designer-added entities (tables) and keys example - mapped relations populated with data

AD CAMPAIGN

<u>AdCampaignID</u>	AdCampaignName	StartDate	Duration	CampaignMgrID
111	SummerFun20	6.6.2020	12 days	CM100
222	SummerZing20	6.8.2020	30 days	CM101
333	FallBall20	6.9.2020	12 days	CM102
444	AutmnStyle20	6.9.2020	5 days	CM103
555	AutmnColors20	6.9.2020	3 days	CM100

CAMPAIGN MANAGER

<u>CampaignMgrID</u>	CampaignMgrName
CM100	Roberta
CM101	Sue
CM102	John
CM103	Nancy

RANGE

<u>RangeID</u>	Range
L	Local
N	National

MEDIA

<u>MediaID</u>	Media
T	TV
R	Radio
P	Print

AD CAMPAIGN MIX

<u>AdCampaignID</u>	<u>ModusID</u>	BudgetPctg
111	1	50%
111	2	50%
222	1	60%
222	3	30%
222	5	10%
333	3	80%
333	4	20%
444	6	100%
555	3	100%

MODUS

<u>ModusID</u>	<u>Media</u> ^{Id}	<u>Range</u> ^{Id}
1	T	L
2	T	N
3	R	L
4	R	N
5	P	L
6	P	N

Normalized Pressly Ad Agency Database

AD CAMPAIGN

<u>AdCampaignID</u>	AdCampaignName	StartDate	Duration	CampaignMgrID
111	SummerFun13	6.6.2013	12 days	CM100
222	SummerZing13	6.8.2013	30 days	CM101
333	FallBall13	6.9.2013	12 days	CM102
444	AutmnStyle13	6.9.2013	5 days	CM103
555	AutmnColors13	6.9.2013	3 days	CM100

CAMPAIGN MANAGER

<u>CampaignMgrID</u>	CampaignMgrName
CM100	Roberta
CM101	Sue
CM102	John
CM103	Nancy

MODE

<u>ModelID</u>	Media	Range
1	TV	Local
2	TV	National
3	Radio	Local
4	Radio	National
5	Print	Local
6	Print	National

AD CAMPAIGN MIX

<u>AdCampaignID</u>	<u>ModelID</u>	BudgetPctg
111	1	50%
111	2	50%
222	1	60%
222	3	30%
222	5	10%
333	3	80%
333	4	20%
444	6	100%
555	3	100%

ADDITIONAL STREAMLINING OF DATABASE CONTENT

- **Designer-added entities (tables) and keys**
 - Augmenting databases with designer added tables and keys is **not a default process** that is to be undertaken **in all** circumstances
 - Instead, augmenting databases with designer added tables and keys should be **done judiciously**, after analyzing **pros and cons** for each augmentation