



Presentation:

Dimensional Modelling 1

Erik Perjons

DSV, Stockholm University

The Concept of Cube and Towards Dimensional Modelling

OLAP tools

- OLAP tools supports so-called interactive and "ad-hoc querying", that is, systems that allow managers and analysts to ask questions that appear in their minds during interaction with the system - without the need to use SQL and other query languages.

OLAP tools

- The main idea of OLAP is that managers and analysts can interactively, quickly and easily choose facts that they want to study, such as sales figures or costs. They can also choose from what views they want to study these facts, that is, from what dimensions. Then they can add new or subtract existing dimensions and thus be able to study the facts, sales figures or cost, from other perspectives.

OLAP tools: User interface

Product group	Region	Quarter	Sales (in kSEK)
Group AB	Region ABC	Quarter 1- 2018	300
Group AB	Region DEF	Quarter 1- 2018	1 300

OLAP tools: User interface

TO DO: Drill-down (for example, by click on "Group AB") – in order to get detailed info about Group A och Group B

Product group	Region	Quarter	Sales (in kSEK)
Group AB	Region ABC	Quarter 1- 2018	300
Group AB	Region DEF	Quarter 1- 2018	1 300

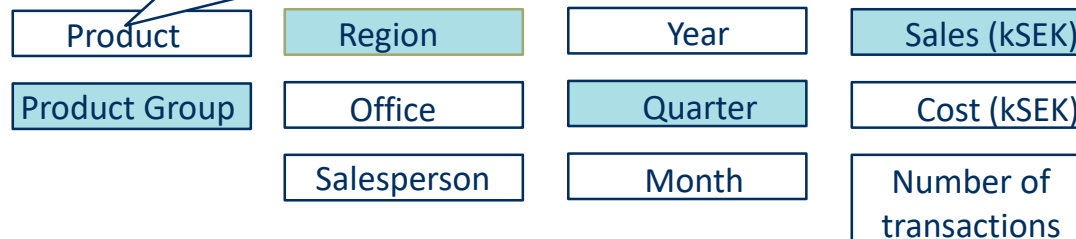
OLAP tools: User interface

DONE: Drill-down (for example, by click on "Group AB") – in order to get detailed info about Group A och Group B

Product Group	Region	Quarter	Sales (kSEK)
Group A	Region ABC	Quarter 1 - 2018	100
Group A	Region DEF	Quarter 1 - 2018	400
Group B	Region ABC	Quarter 1 - 2018	200
Group B	Region DEF	Quarter 1 - 2018	900

OLAP tools: User interface

TO DO: Add another column (e.g., add Product to Excel) – in order to get more detailed info

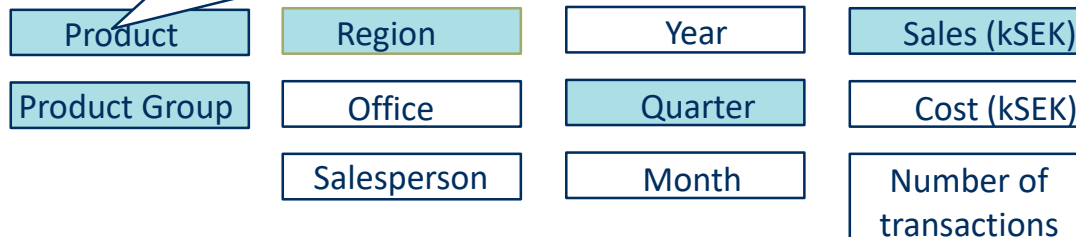


TO DO: Add another column (e.g., add Product to Excel) – in order to get more detailed info

Product Group	Region	Quarter	Sales (kSEK)
Group A	Region ABC	Quarter 1 - 2018	100
Group A	Region DEF	Quarter 1 - 2018	400
Group B	Region ABC	Quarter 1 - 2018	200
Group B	Region DEF	Quarter 1 - 2018	900

OLAP tools: User interface

DONE: Add another column (e.g., add Product to Excel) – in order to get more detailed info)

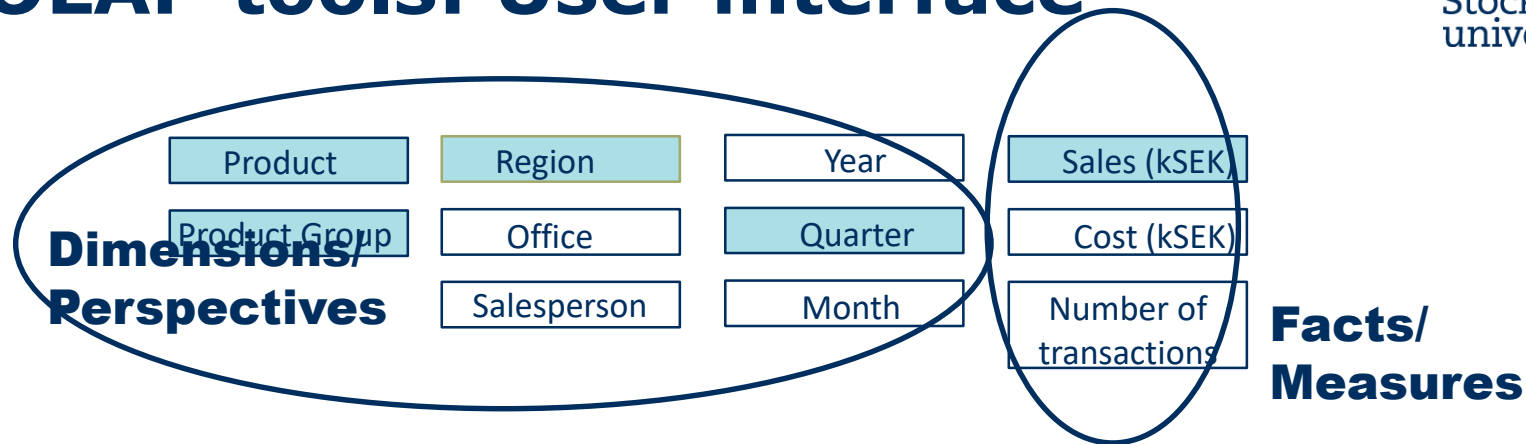


DONE: Add another column (e.g., add Product to Excel) – in order to get more detailed info)

Product	Product Group	Region	Quarter	Sales (kSEK)
Product 20011	Grupp A	Region ABC	Quarter 1- 2018	0
Product 20011	Grupp A	Region DEF	Quarter 1- 2018	10
Product 20012	Grupp A	Region ABC	Quarter 1- 2018	12
Product 20012	Grupp A	Region DEF	Quarter 1- 2018	45

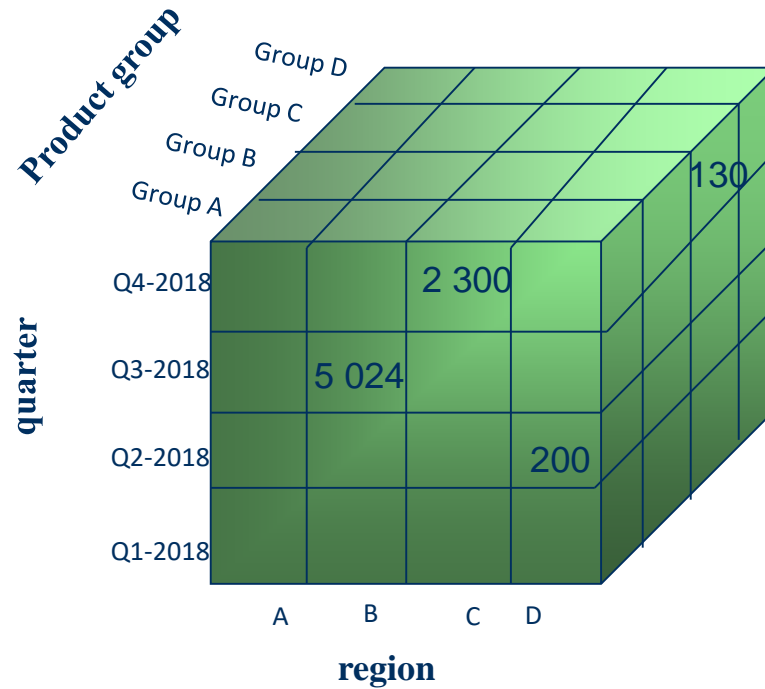
...

OLAP tools: User interface

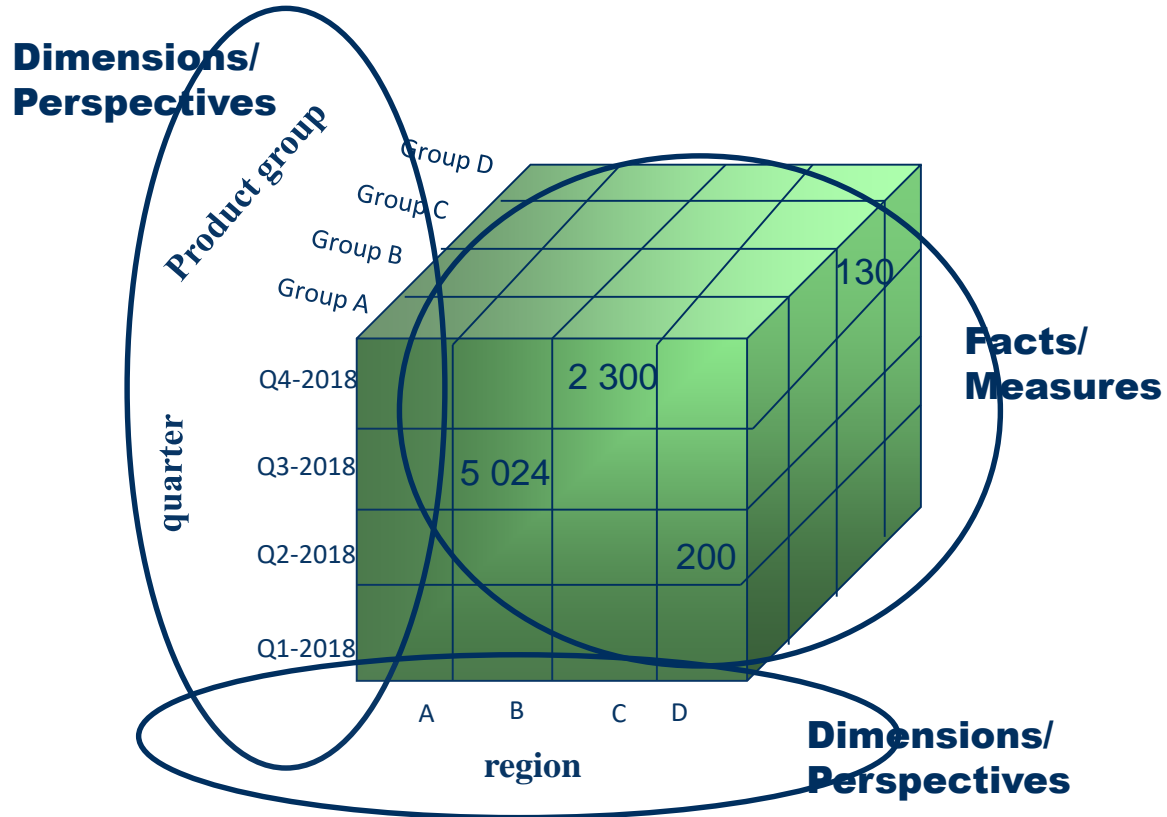


Product	Product Group	Region	Quarter	Sales (kSEK)
Product 20011	Grupp A	Region ABC	Quarter 1- 2018	0
Product 20011	Grupp A	Region DEF	Quarter 1- 2018	10
Product 20012	Grupp A	Region ABC	Quarter 1- 2018	12
Product 20012	Grupp A	Region DEF	Quarter 1- 2018	45
...				

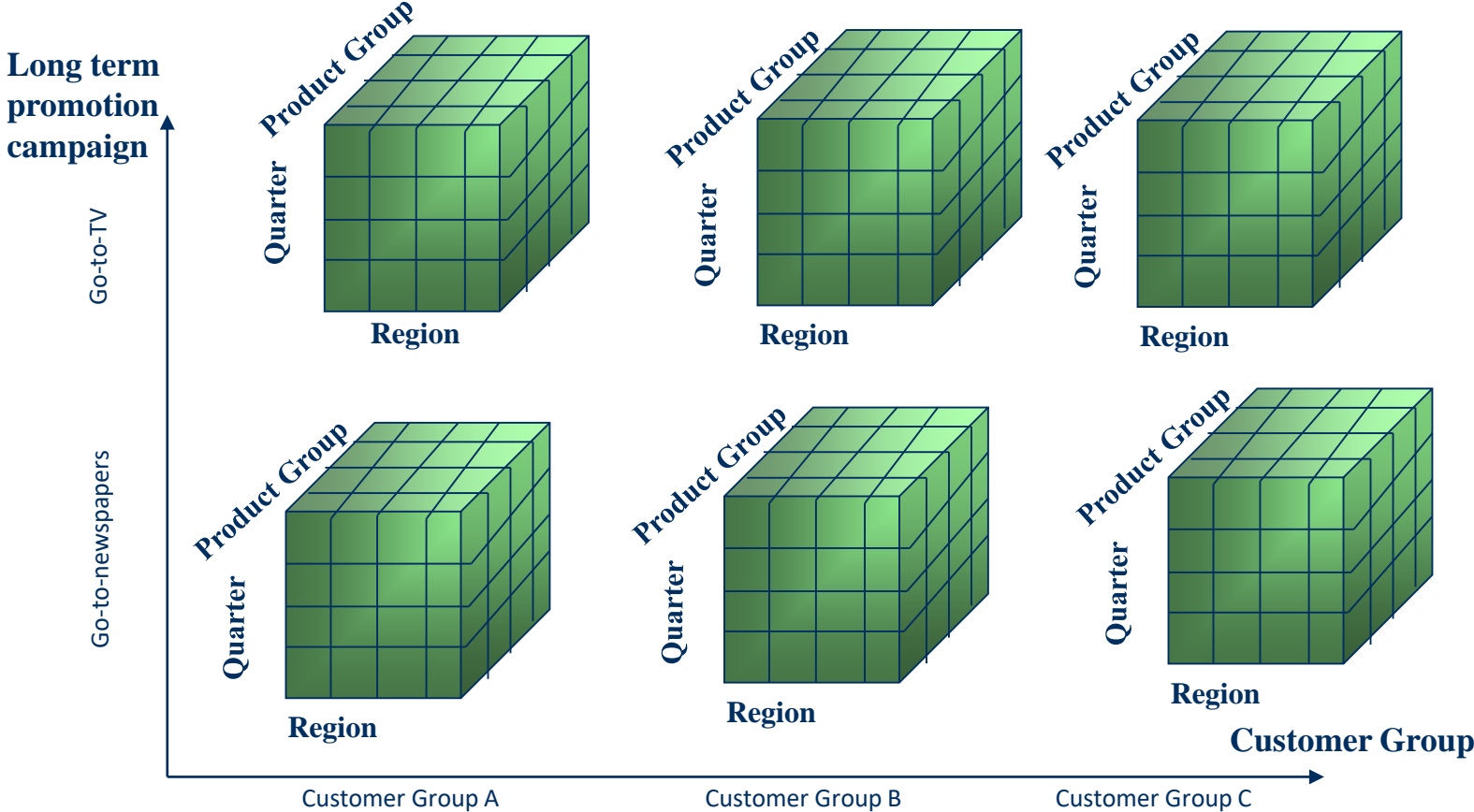
Cube – a multidimensional view on data



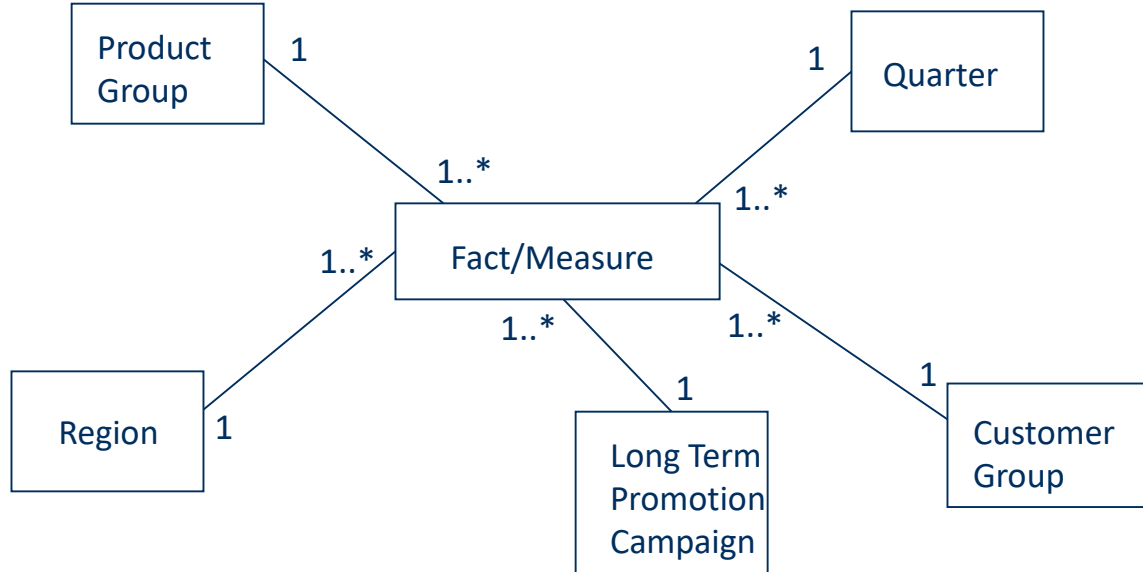
Cube – a multidimensional view on data



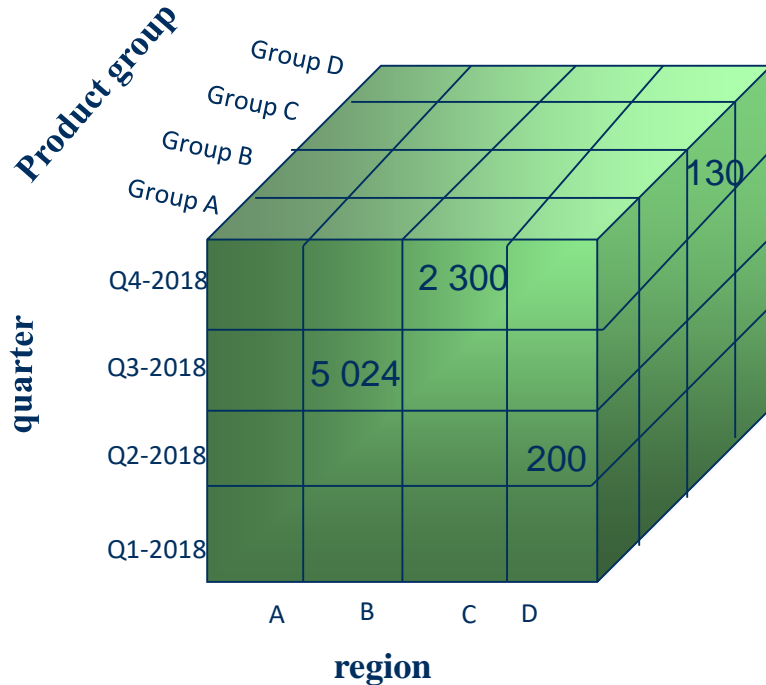
Cube – a multidimensional view on data



How to represent the “multidimensional cube” as a conceptual model?



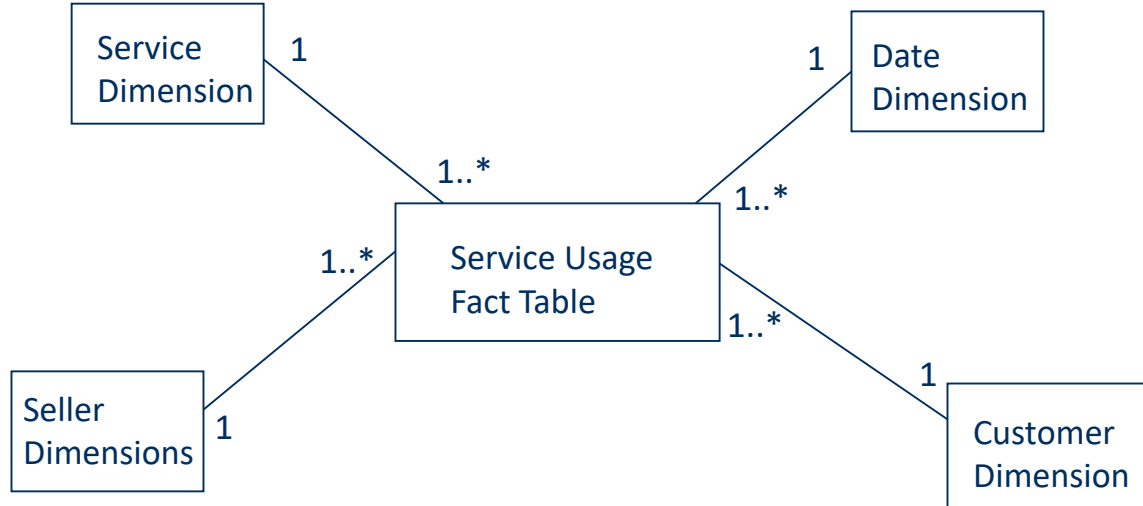
Cube – a multidimensional view on data



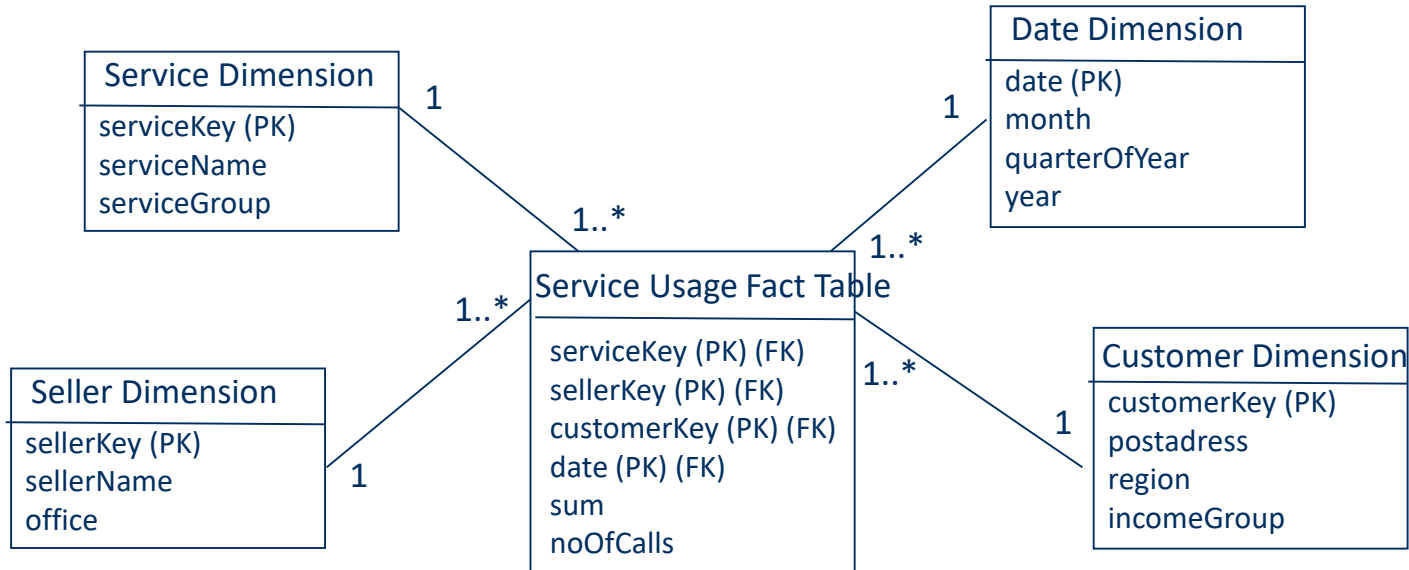
The cube is a concept that has influenced:

- user interface in OLAP tools
- database design, logical structure (star schema, dimensional modelling)
- physical storage ("Dimensional models implemented in multidimensional database environment (MOLAP) are referred to as OLAP cubes")

Dimensional modelling/Star schema



Dimensional modelling/Star schema



Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

1

1..*

1..*

1..*

1

1

1..*

Service Dimension

Key	Service	Service group
S1	Local call	Group A

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C213	S1	F13	991011	12:00	1

Query:
For how much
did customers in Sthlm
use service "Local call"
in october 1999?

$\Sigma=37:00$

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

**Denormalized dimensions –
that is, they are not in 3NF
What is 3NF?**

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Asa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

If possible, organize attributes in hierarchies

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

The PK in the Fact table consist of all the FKs – it is a composite/concatenated key

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Surrogate keys are used
in the dimensional tables
Why surrogate keys?

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

Understand the grain of the fact table – and how it will be impacted by dimensions added and deleted.

Grain is the level of detail at which data is represented

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

What happened if we replace the Date dimensions with a Month dimension instead (where one row represent a month and not a date)?

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

How can you create aggregated dimensions based on the Customer dimension?

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

90 percent of the total space in the database consist of fact tables' data.
Consequences?

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

90 percent of the total space in the database consist of fact tables' data. Therefore fact tables need to be sparse – avoid textual descriptions

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

Textual descriptions
belongs in the dimensional
tables

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Asa S	Stockholm	Stockholm	A

Use labels that end user
can understand – for
example, avoid
abbreviations and codes –
since the users will
actually see the data in
their analysis actions

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact Table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

You can access the fact table via the dimensions tables. The dimensions are the entry points to the data, that is, entry points for approaching the fact table

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

“The database engine can make strong assumptions - by first constraining the heavily indexed dimensional tables, and then attacking the fact table all at once with the Cartesian product of the keys in the dimensional tables”

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Asa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

The simplicity and symmetry of the star schema makes it easy to understand and navigate

Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Service Dimension

Key	Service	Service group
S1	Local call	Group A
S2	Intern. call	Group A
S3	SMS	Group B
S4	WAP	Group C

Date Dimension

Date/ Key	Month	Quarter	Year
991011	9910	4 - 99	99
991012	9910	4 - 99	99

Service Usage Fact table

				Sum	Number of calls
C210	S1	F11	991011	25:00	3
C210	S3	F11	991011	05:00	1
C212	S2	F13	991011	89:00	1
C213	S1	F13	991011	12:00	1
C214	S4	F13	991012	08:00	1

Star schemas “are gracefully extensible to accommodate change”

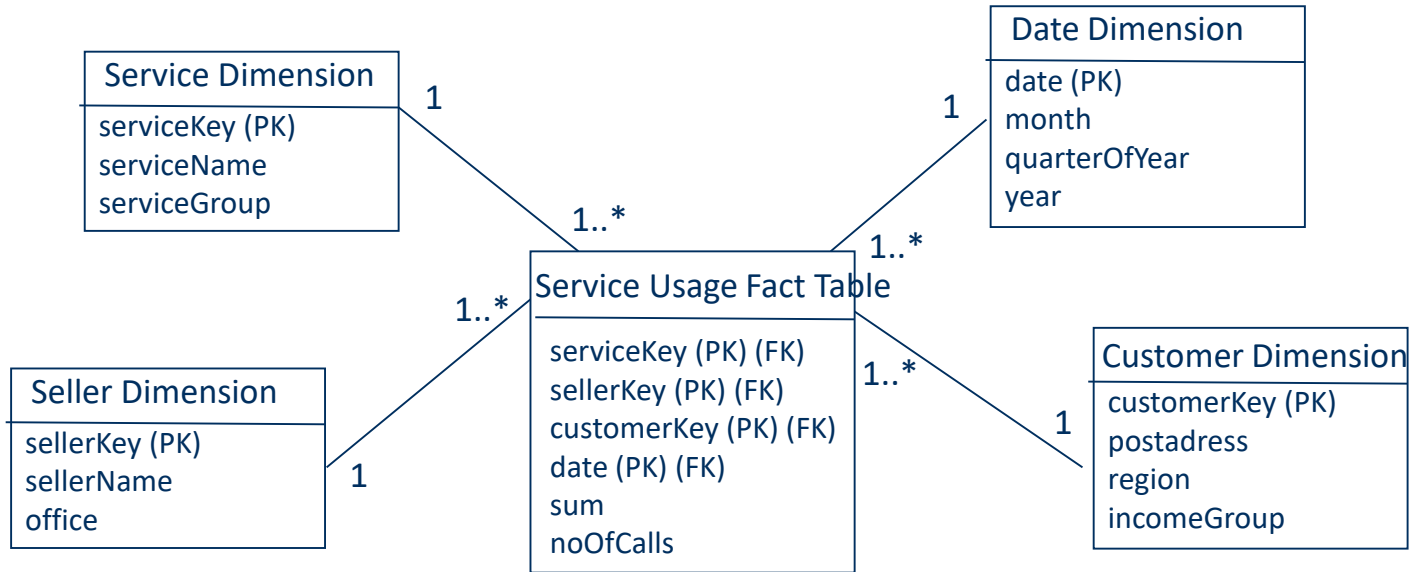
Seller Dimension

Key	Seller	Office
F11	Anders C	Sundsvall
F12	Lisa B	Sundsvall
F13	Janis B	Kista

Customer Dimension

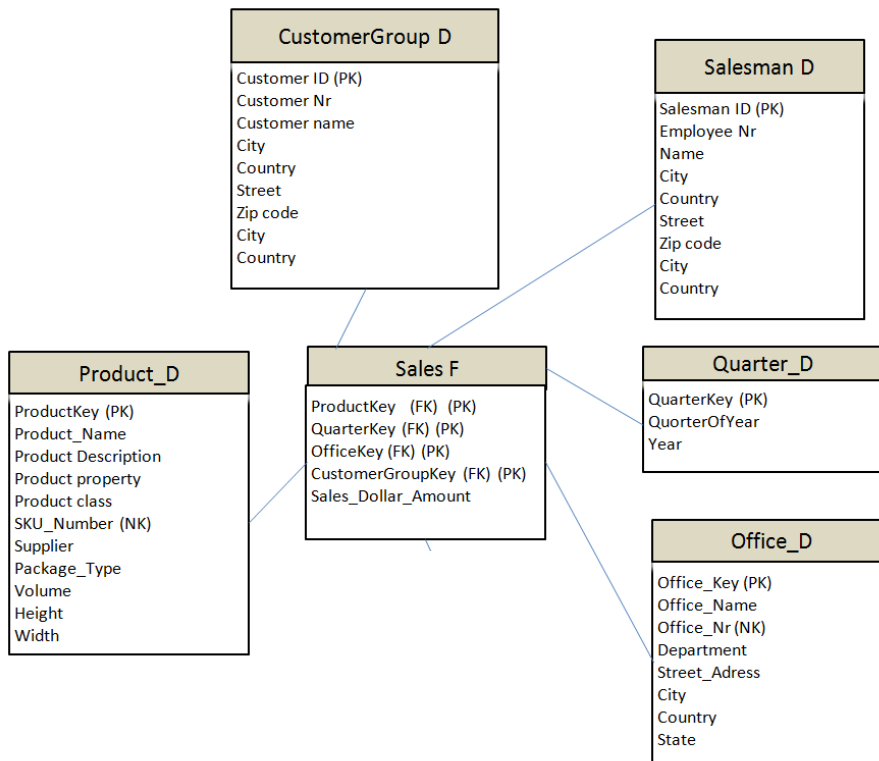
Key	Customer	Address	Region	Income group
C210	Anna N	Stockholm	Stockholm	B
C211	Lars S	Malmö	Skåne	B
C212	Erik P	Rättvik	Dalarna	C
C213	Danny B	Stockholm	Stockholm	A
C214	Åsa S	Stockholm	Stockholm	A

Dimensional modelling/Star schema

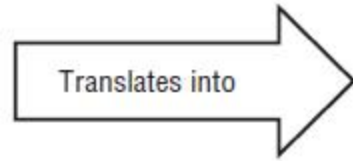


Towards Dimensional Modelling

- A single fact table
 - Facts, measurements
KPI's
- Tables for each dimension
- The dimension tables are denormalized
- Kimball approach



A business event/transaction



Retail Sales Facts
Date Key (FK)
Product Key (FK)
Store Key (FK)
Promotion Key (FK)
Customer Key (FK)
Clerk Key (FK)
Transaction #
Sales Dollars
Sales Units