Understanding Normalization

Voluntary Additional Learning Material

Flat Files

All information in one table
In a flat file, all information is stored in one table. There are no structures for indexing or recognizing relationships between records. This way of data organization leads to a lot of **redundancy**.

Datasheet View and Design View
In **Datasheet View**, you can see that each row contains one record. Each column contains information of a specific data category, such as **Company**, **City**, or **Sales Region**.

Example of Datasheet View

	A	B	C	D		E	F	G	H		I	J	K		L	M	N	O
1	OrderId	SalesDate	ClientID	Company		Country	State	City	Street		SalesRegion	SalesRepID	SalesRep		ProductName	Category	Release	Revenue
2	AB1037	1/11/2018	94128	Per Conubia Nostra PC		Canada	AB	Medicine Hat	P.O. Box 954, 658 Diam Rd.		Can02	EM346	Garner, Heidi		Crosser	FLUX	2015	4521.00
3	AB1090	4/5/2018	94127	Nec Diam Associates		Canada	AB	Edmonton	P.O. Box 754, 9203 Vulputate St.		Can02	EM346	Garner, Heidi		Crosser	FLUX	2015	4568.00
4	AB1095	4/9/2018	94129	Per Institute		Canada	AB	Smoky Lake	P.O. Box 214, 6024 Donec Avenue		Can02	EM346	Garner, Heidi		Minimizer	FLUX	2015	5130.00
5	AB1113	4/27/2018	94122	Magna Foundation		Canada	AB	Raymond	P.O. Box 706, 9886 Urna. St.		Can02	EM346	Garner, Heidi		Crosser	FLUX	2015	4541.00
6	AB1176	7/16/2018	94128	Per Conubia Nostra PC		Canada	AB	Medicine Hat	P.O. Box 954, 658 Diam Rd.		Can02	EM346	Garner, Heidi		Encoder	DARO	2017	3307.00
7	AB1179	7/17/2018	94129	Per Institute		Canada	AB	Smoky Lake	P.O. Box 214, 6024 Donec Avenue		Can02	EM346	Garner, Heidi		Capacitor	FLUX	2018	3404.00
8	AB1192	8/6/2018	94129	Per Institute		Canada	AB	Smoky Lake	P.O. Box 214, 6024 Donec Avenue		Can02	EM346	Garner, Heidi		Optimizer	FLUX	2016	4134.00
9	AB1265	11/20/2018	94124	Neque Tellus Inc.		Canada	AB	Edmonton	Ap #722-9472 Enim, Av.		Can02	EM346	Garner, Heidi		Wobble	DARO	2018	3479.00
10	AB1267	11/26/2018	94122	Magna Foundation		Canada	AB	Raymond	P.O. Box 706, 9886 Urna. St.		Can02	EM346	Garner, Heidi		Encoder	DARO	2017	6008.00
11	AB924	1/11/2017	94127	Nec Diam Associates		Canada	AB	Edmonton	P.O. Box 754, 9203 Vulputate St.		Can02	EM346	Garner, Heidi		Rise	KENO	2017	6029.00
12	AB944	3/23/2017	94129	Per Institute		Canada	AB	Smoky Lake	P.O. Box 214, 6024 Donec Avenue		Can02	EM346	Garner, Heidi		Minimizer	FLUX	2015	6404.00
13	AB1285	12/21/2018	36342	A Odio PC		United States	AL	Birmingham	282-3573 Gravida. Road		USA05	EM352	Flores, Madison		Pulse	KENO	2018	4867.00
14	AB1110	4/27/2018	71739	Placerat Velit Quisque Consulting		United States	AR	Fayetteville	Ap #763-1379 Proin Road		USA02	EM349	White, Carly		Optimizer	FLUX	2016	3678.00
15	AB1189	8/1/2018	72558	Et Netus Associates		United States	AR	Fayetteville	Ap #764-9540 Neque. Rd.		USA02	EM349	White, Carly		Maximizer	FLUX	2016	3200.00
16	AB1135	5/30/2018	86467	Quis Massa Mauris Industries		United States	AZ	Phoenix	P.O. Box 192, 3759 Eu Rd.		USA02	EM349	White, Carly		Crosser	FLUX	2015	4566.00
17	AB1147	6/14/2018	86467	Quis Massa Mauris Industries		United States	AZ	Phoenix	P.O. Box 192, 3759 Eu Rd.		USA02	EM349	White, Carly		Crosser	FLUX	2015	6455.00
18	AB1191	8/3/2018	86467	Quis Massa Mauris Industries		United States	AZ	Phoenix	P.O. Box 192, 3759 Eu Rd.		USA02	EM349	White, Carly		Wobble	DARO	2018	4166.00
19	AB1281	12/13/2018	86004	Nulla Tempor Augue PC		United States	AZ	Tucson	392-3616 Sit Av.		USA02	EM349	White, Carly		Rise	KENO	2017	3481.00
20	AB932	2/8/2017	86467	Quis Massa Mauris Industries		United States	AZ	Phoenix	P.O. Box 192, 3759 Eu Rd.		USA02	EM349	White, Carly		Crosser	FLUX	2015	6449.00
21	AB1027	11/21/2017	94130	Aptent Taciti Sociosqu Foundation		Canada	BC	Burnaby	P.O. Box 450, 9499 Maecenas Street		Can02	EM346	Garner, Heidi		Warper	DARO	2016	6143.00

In **Design View**, the data itself is not visible. Only the **data fields** are visible. The data fields correspond to the column headings.

Field properties are assigned to each data field. The field properties define the allowed **type of content**, e.g. that the field "City" should always contain **text** and the field "Revenue" should always contain **numbers in a specific currency format**.

Data fields in Design View
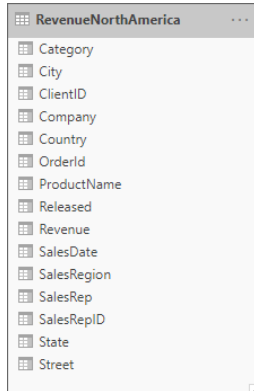
A common misconception

It is a common misconception that flat files are easy to handle just because it is not necessary to establish relationships between different tables.

In fact, flat files cause problems when data need to be changed, e.g. company name, product name, sales representatives, or a reorganization of sales regions.

Solving this task with "Find and Replace" is not a suitable workaround. Instead, take advantage of the many benefits of **normalization**. In this way, you can significantly reduce typos, avoid data duplication, and significantly increase processing speed.

Normalization

Master data and Transactions

Normalization requires several tables. In each table each **value** is only stored once.

After normalization, there are in most cases two types of tables (**master data** and **transactions**). **Master data** include **products**, **customers**, **sales regions**, and **sales representatives**.

In the **Revenue** table, the individual **transactions** are logged, i.e. which client ordered which product on which date. Note that the **Revenue** table does not store the clients' names, only the **Client ID**, and thus refers to a specific record in the **Clients** table.

The same structure applies to products.

You may ask yourself, what is the purpose of the **Order ID** in the table **Revenue**? The Order ID provides a unique ID for each entry in the **Revenue** table. In this way, another requirement for normalized tables is met, namely that each record has a unique identifier.

Data from the normalized table "Clients"

ClientID	Company	Country	State	City	Street
16699	A Corporation	United States	HI	Honolulu	5574 Eu Street
74955	A Mi Foundation	United States	CO	Fort Collins	983-4649 Tincidunt Ave
36342	A Odio PC	United States	AL	Birmingham	282-3573 Gravida. Road
75878	Ac Incorporated	United States	KY	Covington	Ap #780-9193 Faucibus Rd.
94080	Ac Institute	United States	Delaware	Wilmington	836-2499 Aliquet Road
13922	Ac Orci PC	United States	GA	Atlanta	P.O. Box 596, 3765 Placerat, Rd.
94091	Accumsan Laoreet Institute	Canada	QC	Anjou	P.O. Box 910, 679 A St.
94135	Adipiscing Mauris Corporation	Canada	BC	Victoria	6602 Quisque Rd.
37624	Aenean Limited	United States	VA	Norfolk	Ap #553-5219 Egestas. Rd.
94100	Aenean Massa Integer Foundation	Canada	QC	Saint-Urbain	8698 Interdum Street
62202	Aliquam Enim Nec Associates	United States	CO	Hartford	688-1126 Tristique Street
86552	Aliquam Ultrices Consulting	United States	AZ	Mesa	Ap #184-8968 Vitae Street
37155	Aliquam Vulputate Ullamcorper Corporation	United States	IN	Indianapolis	Ap #293-3561 Eu St.
51444	Aliquet Odio Etiam Ltd	United States	IA	Des Moines	P.O. Box 158, 9397 Fames Rd.
88291	Amet Ltd	United States	MA	Augusta	974-1418 Feugiat Street
76555	Amet Ultricies Sem Consulting	United States	MN	Duluth	225-5750 Magna Avenue
76112	Ante Corporation	United States	IN	Indianapolis	428-9202 Quisque Street
94076	Ante Vivamus Non Associates	United States	WY	Casper	P.O. Box 595, 2425 Eu St.
91657	Aptent Taciti Inc.	United States	ID	Boise	Ap #828-6253 Ante. Avenue
94130	Aptent Taciti Sociosqu Foundation	Canada	BC	Burnaby	P.O. Box 450, 9499 Maecenas Street
94051	Arcu Consulting	United States	CA	Fresno	238-3538 Duis Road
33740	At Augue Institute	United States	NV	Paradise	Ap #760-966 Non Street

Data from the normalized table "Product"

ProductID	ProductName	Category	Released
25083	Crosser	FLUX	2015
25022	Minimizer	FLUX	2015
25109	Maximizer	FLUX	2016
25091	Optimizer	FLUX	2016
25023	Capacitor	FLUX	2018
25074	Warper	DARO	2016
25150	Encoder	DARO	2017
25175	Wobble	DARO	2018
25171	Pulse	KENO	2018
25110	Tilt	KENO	2016
25046	Risen	KENO	2017

Data from the normalized table "Revenue"

ClientID	ProductID	OrderId	Revenue	SalesDate
15280	25109	AB1289	4115	12/28/2018
59430	25074	AB1288	3164	12/26/2018
10527	25171	AB1287	5341	12/25/2018
94056	25023	AB1286	3883	12/24/2018
36342	25171	AB1285	4867	12/21/2018
94091	25091	AB1284	5013	12/20/2018
69979	25091	AB1283	4254	12/18/2018
94146	25083	AB1282	4023	12/14/2018
86004	25046	AB1281	3481	12/13/2018
10527	25150	AB1280	6887	12/12/2018

Note that every table has a unique identifier: **Product ID**, **Client ID** and **Order ID**. Now the preparations are done, and you can establish relationships between these tables. The relationships are established between the IDs, used in both tables. The table **Revenue** is using the **Client ID** to point to a specific record in the **Client** table.
The same works with **Product IDs**.

For the data used throughout this course, a normalized table structure looks like this:
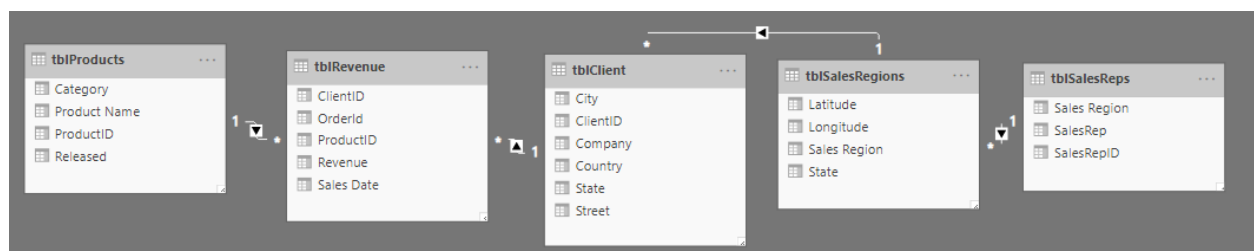
Querys

So-called **queries** let you create specific **Views** of data from the tables that are related to each other. You can now **ask questions to your database**.

Query example: Which client has bought which product?

ClientID	Company	Country	State	City	Street	ProductID	ProductName	Category
10527	Lorem Eu Inc.	United States	WA	Bellevue	Ap #826-9163 Purus, Avenue	25023	Capacitor	FLUX
10527	Lorem Eu Inc.	United States	WA	Bellevue	Ap #826-9163 Purus, Avenue	25046	Risen	KENO
10527	Lorem Eu Inc.	United States	WA	Bellevue	Ap #826-9163 Purus, Avenue	25074	Warper	DARO
10527	Lorem Eu Inc.	United States	WA	Bellevue	Ap #826-9163 Purus, Avenue	25083	Crosser	FLUX
10527	Lorem Eu Inc.	United States	WA	Bellevue	Ap #826-9163 Purus, Avenue	25091	Optimizer	FLUX
10527	Lorem Eu Inc.	United States	WA	Bellevue	Ap #826-9163 Purus, Avenue	25109	Maximizer	FLUX
10527	Lorem Eu Inc.	United States	WA	Bellevue	Ap #826-9163 Purus, Avenue	25110	Tilt	KENO
10527	Lorem Eu Inc.	United States	WA	Bellevue	Ap #826-9163 Purus, Avenue	25150	Encoder	DARO
10527	Lorem Eu Inc.	United States	WA	Bellevue	Ap #826-9163 Purus, Avenue	25171	Pulse	KENO
10527	Lorem Eu Inc.	United States	WA	Bellevue	Ap #826-9163 Purus, Avenue	25175	Wobble	DARO
11518	Eget Venenatis A Incorporated	United States	OK	Oklahoma City	P.O. Box 724, 3695 Maecenas Street	25046	Risen	KENO
11518	Eget Venenatis A Incorporated	United States	OK	Oklahoma City	P.O. Box 724, 3695 Maecenas Street	25083	Crosser	FLUX
12946	Montes Inc.	United States	OH	Akron	P.O. Box 256, 6118 Erat Ave	25022	Minimizer	FLUX
12946	Montes Inc.	United States	OH	Akron	P.O. Box 256, 6118 Erat Ave	25023	Capacitor	FLUX
12946	Montes Inc.	United States	OH	Akron	P.O. Box 256, 6118 Erat Ave	25091	Optimizer	FLUX
14039	Dolor Corp.	United States	MT	Missoula	P.O. Box 382, 9504 Nibh Avenue	25109	Maximizer	FLUX
14740	Vitae Purus Foundation	United States	OH	Cincinnati	Ap #781-7026 Ornare Rd.	25083	Crosser	FLUX
14740	Vitae Purus Foundation	United States	OH	Cincinnati	Ap #781-7026 Ornare Rd.	25091	Optimizer	FLUX
15280	Cursus Ltd	United States	KY	Lexington	Ap #815-6728 Libero. Rd.	25109	Maximizer	FLUX
15280	Cursus Ltd	United States	KY	Lexington	Ap #815-6728 Libero. Rd.	25171	Pulse	KENO
16770	Nulla PC	United States	IA	Des Moines	P.O. Box 934, 4267 Vitae St.	25091	Optimizer	FLUX
18105	Nec Metus Corp.	United States	WY	Casper	P.O. Box 973, 7729 Sit Av.	25091	Optimizer	FLUX
18850	Euismod In LLC	United States	MT	Missoula	P.O. Box 839, 9025 Faucibus Rd.	25022	Minimizer	FLUX
18850	Euismod In LLC	United States	MT	Missoula	P.O. Box 839, 9025 Faucibus Rd.	25091	Optimizer	FLUX
18850	Euismod In LLC	United States	MT	Missoula	P.O. Box 839, 9025 Faucibus Rd.	25109	Maximizer	FLUX
18850	Euismod In LLC	United States	MT	Missoula	P.O. Box 839, 9025 Faucibus Rd.	25175	Wobble	DARO
19366	Dui Company	United States	TN	Clarksville	266-1793 Neque Av.	25083	Crosser	FLUX

In Power BI, you can create various visualizations using the relationships between the tables. You can create lists like the above, and also charts, tables, cards, maps, sparklines and much more. And you can run calculations on fields containing values.

Links

If you want to learn more, this Wikipedia article is a good start. Try to understand the first, the second and third normal form (NF).

https://en.wikipedia.org/wiki/Database_normalization