

Normalización de relaciones

| | | |
|-----------|---|----------|
| 1. | Normalización de relacións | 3 |
| 1.1 | Normalización e formas normais (FN)..... | 3 |
| 1.2 | Obxectivos da normalización | 4 |
| 1.3 | Dificultades no Deseño de BDR..... | 5 |
| 1.4 | Dependencias funcionais..... | 9 |
| | Definición formal de dependencia funcional (DF) | 9 |
| 1.5 | Proceso de normalización dunha relación..... | 10 |

1. Normalización de relacións

O obxectivo do desenvolvemento dun esquema de BD relacional é crear unha representación precisa e adecuada dos datos, as súas relación e as súas restricións. Esta representación realizarase mediante un conxunto de esquemas de relación que permitirán almacenar a información do “minimundo” cun mínimo de redundancias e facilitando a recuperación da información, evitando problemas de actualización, posibles incoherencias, etc.

1.1 Normalización e formas normais (FN)

En 1972 E.F. Codd definiu a Normalización como unha técnica para producir un conxunto de relacións con propiedades desexables, dados os requisitos de datos dunha organización.

Unha vez obtido o esquema de BD Relacional, formado por todos os esquemas de relación da BD, a partir da transformación do esquema conceptual (modelo entidade-interrelación estendido) necesitamos pautas ou teorías que axuden ao deseñador a identificar o agrupamento óptimo dos atributos para cada relación no esquema. Estas pautas poden ser:

- **Informais:**

- Diseñar esquemas de relación con significado fácil de entender e claras.
- Eficiencias de consulta e aproveitamento de espazo en disco.
- Redución de valores redundantes en tuplas para evitar anomalías de actualización, facilitando o mantemento e a evolución do deseño.
- Boa selección de claves.
- Redución de nulos nas tuplas, empregándoos en casos excepcionais (estes valores desperdician espazo de almacenamento e dificultan as operacións). Pode considerarse tamén un método formal.

- **Formais:** aplicando técnicas formais de Normalización. Este proceso identifica relacións baseándose na súa clave primaria (ou candidatas na FNBC) e das dependencias funcionais entre os seus atributos. O esquema relacional de BD pode ser normalizado ata unha forma normal específica, para evitar “propiedades non desexables”. As formas normais pódense clasificar en:

- 1ª, 2ª, 3ª Forma Normal (Codd, 1970/1971) e BNFC (R.F. Boyce e Codd, 1974) que constitúe unha definición máis forte da 3FN baseadas en dependencias funcionais. Eliminan case totalmente a posibilidade de anomalías de actualización. Estas formas normais resúmense requirindo que todos os atributos non clave sexan dependentes da "clave, clave completa, e nada máis excepto a clave".
- 4ª, e 5ª Forma Normal creadas por Fagin en 1977/1979, baseándose nos conceptos de dependencias multivaluadas e de reunión, respectivamente e 6ª Forma Normal (2002), que se basea no principio de que existen máis de dúas claves candidatas nunha táboa se terán que crear outras táboas con estes atributos claves. Estas últimas formas normais eliminan anomalías de actualización excepcionais.

Na seguinte táboa resúmense as distintas formas normais, indicando os seus creadores e a data en que se propuxeron:

| Forma normal | Definida por |
|--|---|
| Primera Forma Normal (1FN ou 1NF) | Existen dos versiones, la de E. F. Codd de 1970 y la de C. J. Date de 2003. |
| Segunda Forma Normal (2FN ou 2NF) | E. F. Codd en 1971. |
| Tercera Forma Normal (3FN ou 3NF) | E. F. Codd en 1971. |
| Forma Normal de Boyce-Codd (FNBC ou BCNF) | Raymond F. Boyce y E. F. Codd en 1974. |
| Cuarta Forma Normal (4FN ou 4NF) | Ronald Fagin en 1977. |
| Quinta Forma Normal (5FN ou 5NF) | Ronald Fagin en 1979. |
| Forma Normal de Dominio/Clave (FNDC ou DKNF) | Ronald Fagin en 1981. |
| Sexta Forma Normal (6FN ou 6NF) | C. J. Date, Hugh Darwen y Nikos Lorentzos en 2002. |

Táboa. Autores das Formas Normais

A normalización poderíase definir dunha maneira sinxela como o proceso de organizar e estruturar os datos de forma que se minimice a redundancia. Este proceso baséase nun conxunto de pautas, as regras de normalización, que dismínúen o risco de ter un deseño de base de datos defectuoso. Estas regras aplícanse ao modelo relacional da base de datos, normalmente obtido a partir do modelo entidade-interrelación.

As formas normais son incrementais, é dicir, se se cumpre a forma normal (n-ésima) cúmprese tamén a (n-1)-ésima. Unha relación pode atoparse en primeira forma normal e non en segunda forma normal, pero non ao contrario, é dicir os números altos de formas normais son mais restritivos.

O itinerario que se segue á hora de normalizar unha relación é o que se mostra no seguinte diagrama:

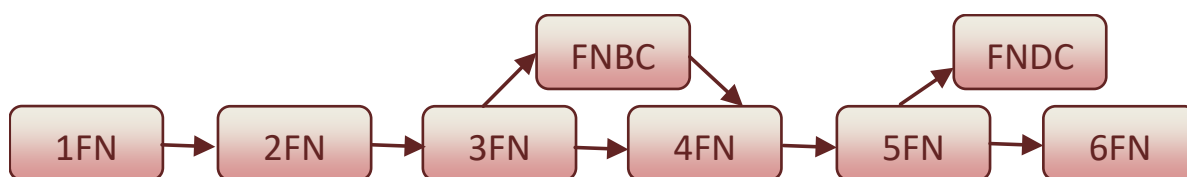


Figura. As formas normais

Existe un certo consenso en normalizar as táboas da base de datos ata a 3FN ou superior, de feito hai autores que opinan que as bases de datos en cuarta ou quinta forma teñen un redemento peor que as que se atopan na terceira o una FNBC. En cualquier caso debería ser obligatorio para cualquier diseñador chegar ata a forma normal de Boyce-Codd.

1.2 Obxectivos da normalización

Os principais obxectivos que persegue o proceso de normalización son:

- Evitar a redundancia de datos.

- Protexer e dar un mellor soporte á integridade dos datos.
- Eliminar as anomalías nos datos, tanto nas actualizacións como nas insercións e os borrados.
- Reducir na medida do posible o redeseño da base de datos cando esta se amplía.
- Facer máis entendible o modelo de datos a quenes vaian a utilizalo, posto que se modeliza mellor a realidade, o dominio do problema.
- Proporcionar independencia das linguaxes específicas para a consulta de datos.

Mellorar o rendemento da base de datos non está na lista de obxectivos, xa que a normalización céntrase exclusivamente en almacenar a información de maneira correcta.

1.3 Dificultades no Deseño de BDR

As bases de datos mal deseñadas teñen problemas de:

- Almacenamento redundante (varias copias da mesma información).
- Perdas non desexadas de información, ao modificar tuplas base, ou borrar tuplas poden levar a un estado non consistente.
- Imposibilidade de almacenar certa información.

Estas dificultades poden dar lugar a anomalías de actualización que aínda que non constitúen un error en si mesmas (si se fai ben a actualización non habería problemas) son un factor de erro, xerando inconsistencias semánticas e outros problemas que o modelo relacional podería non detectar, estas anomalías non detectadas evitaríanse coa normalización e permitirán un deseño máis robusto.

Supón o seguinte exemplo dunha táboa que contén a información sobre os pedidos de produtos que fan unha serie de clientes:

PEDIDO (*codPedido, codProducto, descripcion, prezoUnidade, unidades, total, data, idCliente, nomeCliente, cidade*)

| PEDIDO | | | | | | | | | |
|-----------|-------------|-------------|--------------|----------|-------|------------|-----------|-------------|------------|
| CodPedido | codProducto | descripcion | prezoUnidade | cantidad | total | data | idCliente | nomeCliente | cidade |
| 22546 | P111345 | silla | 180 | 12 | 2160 | 11/11/2014 | G948499 | Pirrel | Ourense |
| 23766 | P748765 | bolígrafo | 0,9 | 30 | 27 | 12/07/2015 | G948499 | Pirrel | Ourense |
| 17676 | P847877 | folios | 3 | 20 | 60 | 23/08/2014 | G948499 | Pirrel | Ourense |
| 34355 | P111345 | Silla | 180 | 10 | 1800 | 25/08/2013 | D482049 | Xen | Lugo |
| 77244 | P111345 | Silla | 180 | 15 | 2700 | 02/01/2015 | F985980 | Carpazo | Pontevedra |
| 86987 | P748765 | bolígrafo | 0,9 | 33 | 29,7 | 12/12/2012 | D482049 | Xen | Lugo |
| 55356 | P748765 | bolígrafo | 0,9 | 22 | 19,8 | 07/09/2014 | F985980 | Carpazo | Pontevedra |
| 56435 | P847877 | folios | 3 | 21 | 63 | 15/04/2015 | F985980 | Carpazo | Pontevedra |
| 12135 | P135455 | grapas | 0,75 | 23 | 17,25 | 02/01/2015 | J445532 | Lana | A Coruña |

Figura. Relación PEDIDO non normailizada

Na relación observase que a información dun cliente con máis dun pedido aparece repetida (*idCliente, nomeCliente* e *cidade* do mesmo). Isto pode conlevar certas anomalías de actualización:

- **Modificación do valor dun campo:** ¿qué ocorre se queremos modificar o id do cliente Carpazo.?

| PEDIDO | | | | | | | | | |
|-----------|-------------|-------------|--------------|-----------|-------|------------|-----------|-------------|------------|
| CodPedido | codProducto | descripcion | prezoUnidade | cantidade | total | data | idCliente | nomeCliente | cidade |
| 22546 | P111345 | silla | 180 | 12 | 2160 | 11/11/2014 | G948499 | Pirrel | Ourense |
| 23766 | P748765 | bolígrafo | 0,9 | 30 | 27 | 12/07/2015 | G948499 | Pirrel | Ourense |
| 17676 | P847877 | folios | 3 | 20 | 60 | 23/08/2014 | G948499 | Pirrel | Ourense |
| 34355 | P111345 | Silla | 180 | 10 | 1800 | 25/08/2013 | D482049 | Xen | Lugo |
| 77244 | P111345 | Silla | 180 | 15 | 2700 | 02/01/2015 | A124476 | Carpazo | Pontevedra |
| 86987 | P748765 | bolígrafo | 0,9 | 33 | 29,7 | 12/12/2012 | D482049 | Xen | Lugo |
| 55356 | P748765 | bolígrafo | 0,9 | 22 | 19,8 | 07/09/2014 | F985980 | Carpazo | Pontevedra |
| 56435 | P847877 | folios | 3 | 21 | 63 | 15/04/2015 | F985980 | Carpazo | Pontevedra |
| 12135 | P135455 | grapas | 0,75 | 23 | 17,25 | 02/01/2015 | J445532 | Lana | A Coruña |

Tuplas a modificar

Figura. Relación PEDIDO non normalizada que mostra anomalías de modificación

Ao modificar os valores dos atributos, debido a redundancia poderían xurdir problemas ao modificar e se non actualizamos os campos en todas as filas levaríamos a base de datos a un estado inconsistente.

- **Inserción dunha tupla:**

- ¿Qué acontece si engadimos outro pedido, con código 90457, feito polo cliente Pirrel (con id de cliente G948499) de Ourense, de 45 unidades do produto P748786 Ficon prezo 0.90 €na data 01 de Xaneiro de 2015?.

| PEDIDO | | | | | | | | | |
|-----------|-------------|-------------|--------------|-----------|-------|------------|-----------|-------------|------------|
| CodPedido | codProducto | descripcion | prezoUnidade | cantidade | total | data | idCliente | nomeCliente | cidade |
| 22546 | P111345 | silla | 180 | 12 | 2160 | 11/11/2014 | G948499 | Pirrel | Ourense |
| 23766 | P748765 | bolígrafo | 0,9 | 30 | 27 | 12/07/2015 | G948499 | Pirrel | Ourense |
| 17676 | P847877 | folios | 3 | 20 | 60 | 23/08/2014 | G948499 | Pirrel | Ourense |
| 34355 | P111345 | Silla | 180 | 10 | 1800 | 25/08/2013 | D482049 | Xen | Lugo |
| 77244 | P111345 | Silla | 180 | 15 | 2700 | 02/01/2015 | A124476 | Carpazo | Pontevedra |
| 86987 | P748765 | bolígrafo | 0,9 | 33 | 29,7 | 12/12/2012 | D482049 | Xen | Lugo |
| 55356 | P748765 | bolígrafo | 0,9 | 22 | 19,8 | 07/09/2014 | A124476 | Carpazo | Pontevedra |
| 56435 | P847877 | folios | 3 | 21 | 63 | 15/04/2015 | A124476 | Carpazo | Pontevedra |
| 12135 | P135455 | grapas | 0,75 | 23 | 17,25 | 02/01/2015 | J445532 | Lana | A Coruña |
| 90457 | P748765 | bolígrafo | 0,9 | 45 | 40,5 | 01/01/2015 | G948499 | Pirrel | Ourense |

Información redundante

Figura. Relación PEDIDO non normalizada que mostra anomalías de inserción

Ao engadir un novo pedido dun cliente seguiríase a introducira información redundante.

- ¿Qué ocorre se o introducimos información dun cliente que non fixo pedidos?.

| PEDIDO | | | | | | | | | |
|-----------|-------------|-------------|--------------|-----------|-------|------------|-----------|-------------|------------|
| CodPedido | codProducto | descripcion | prezoUnidade | cantidade | total | data | idCliente | nomeCliente | cidade |
| 22546 | P111345 | silla | 180 | 12 | 2160 | 11/11/2014 | G948499 | Pirrel | Ourense |
| 23766 | P748765 | bolígrafo | 0,9 | 30 | 27 | 12/07/2015 | G948499 | Pirrel | Ourense |
| 17676 | P847877 | folios | 3 | 20 | 60 | 23/08/2014 | G948499 | Pirrel | Ourense |
| 34355 | P111345 | Silla | 180 | 10 | 1800 | 25/08/2013 | D482049 | Xen | Lugo |
| 77244 | P111345 | Silla | 180 | 15 | 2700 | 02/01/2015 | A124476 | Carpazo | Pontevedra |
| 86987 | P748765 | bolígrafo | 0,9 | 33 | 29,7 | 12/12/2012 | D482049 | Xen | Lugo |
| 55356 | P748765 | bolígrafo | 0,9 | 22 | 19,8 | 07/09/2014 | A124476 | Carpazo | Pontevedra |
| 56435 | P847877 | folios | 3 | 21 | 63 | 15/04/2015 | A124476 | Carpazo | Pontevedra |
| 12135 | P135455 | grapap | 0,75 | 23 | 17,25 | 02/01/2015 | J445532 | Lana | A Coruña |
| 90457 | P748765 | bolígrafo | 0,9 | 45 | 40,5 | 01/01/2015 | G948499 | Pirrel | Ourense |
| 44567 | L948566 | NULL | NULL | NULL | NULL | NULL | J554679 | keppo | Madrid |

Figura. Relación PEDIDO non normalizada que mostra anomalías de inserción.

Se queremos introducir os datos dun cliente que non fixo ningún pedido, teremos que xerar valores nulos.

O mesmo acontecería nas seguintes situacións: cambiar un cliente de cidade, modificar o prezo dun produto.

- **Eliminación dunha tupla:** ¿Qué ocorre se eliminamos o produto P748765 da relación PEDIDO?.

| PEDIDO | | | | | | | | | |
|-----------|-------------|-------------|--------------|-----------|-------|------------|-----------|-------------|------------|
| CodPedido | codProducto | descripcion | prezoUnidade | cantidade | total | data | idCliente | nomeCliente | cidade |
| 22546 | P111345 | silla | 180 | 12 | 2160 | 11/11/2014 | G948499 | Pirrel | Ourense |
| 23766 | P748765 | bolígrafo | 0,9 | 30 | 27 | 12/07/2015 | G948499 | Pirrel | Ourense |
| 17676 | P847877 | folios | 3 | 20 | 60 | 23/08/2014 | G948499 | Pirrel | Ourense |
| 34355 | P111345 | Silla | 180 | 10 | 1800 | 25/08/2013 | D482049 | Xen | Lugo |
| 77244 | P111345 | Silla | 180 | 15 | 2700 | 02/01/2015 | A124476 | Carpazo | Pontevedra |
| 86987 | P748765 | bolígrafo | 0,9 | 33 | 29,7 | 12/12/2012 | D482049 | Xen | Lugo |
| 55356 | P748765 | bolígrafo | 0,9 | 22 | 19,8 | 07/09/2014 | A124476 | Carpazo | Pontevedra |
| 56435 | P847877 | folios | 3 | 21 | 63 | 15/04/2015 | A124476 | Carpazo | Pontevedra |
| 12135 | P135455 | grapap | 0,75 | 23 | 17,25 | 02/01/2015 | J445532 | Lana | A Coruña |
| 90457 | P748765 | bolígrafo | 0,9 | 45 | 40,5 | 01/01/2015 | G948499 | Pirrel | Ourense |
| 44567 | L948566 | NULL | NULL | NULL | NULL | NULL | J554679 | keppo | Madrid |

Figura. Relación PEDIDO non normalizada que mostra anomalías de eliminación de tuplas

Ao eliminar a información do PEDIDO poderíamos perder todos os datos dos clientes con ese pedido. As inconsistencias afloran ao facer consultas: un cliente que xa non aparece, un produto con diferente prezo, etc.

- **Descomposición de reunión con perda (tuplas espúreas):**

Ao dividir unha relación de forma incorrecta pode producir resultados inconsistentes, por exemplo supoñamos a seguinte división en dúas relacións do exemplo:

| PEDIDOS | | | | |
|-----------|--------------|-------|------------|-------------|
| CodPedido | prezoUnidade | total | data | nomeCliente |
| 22546 | 180 | 2160 | 11/11/2014 | Pirrel |
| 23766 | 0,9 | 27 | 12/07/2015 | Pirrel |
| 17676 | 3 | 60 | 23/08/2014 | Pirrel |
| 34355 | 180 | 1800 | 25/08/2013 | Xen |
| 77244 | 180 | 2700 | 02/01/2015 | Carpazo |
| 86987 | 0,9 | 29,7 | 12/12/2012 | Xen |
| 55356 | 0,9 | 19,8 | 07/09/2014 | Carpazo |
| 56435 | 3 | 63 | 15/04/2015 | Carpazo |
| 12135 | 0,75 | 17,25 | 02/01/2015 | Lana |

| PEDIDO DE CLIENTE | | | | | | |
|-------------------|-------------|------------|-------------------------|-------------|-------------|----------|
| idCliente | nomeCliente | cidade | telefonos | codProducto | descripcion | cantidad |
| G948499 | Pirrel | Ourense | 988223344; 654327834 | P111345 | silla | 12 |
| G948499 | Pirrel | Ourense | 988123456; 654321648 | P748765 | bolígrafo | 30 |
| G948499 | Pirrel | Ourense | 988762311; 976534124 | P847877 | folios | 20 |
| D482049 | Xen | Lugo | 982098765; 653876098 | P111894 | Silla | 10 |
| F985980 | Carpazo | Pontevedra | 986341290; 645068234 | P111684 | Silla | 15 |
| D482049 | Xen | Lugo | 982009988; 618098345 | P748786 | bolígrafo | 33 |
| F985980 | Carpazo | Pontevedra | 986306187; 630982310 | P748866 | bolígrafo | 22 |
| F985980 | Carpazo | Pontevedra | 986976834; 639675143 | P847689 | folios | 21 |
| J445532 | Lana | A Coruña | 981675423; 623987407 | P135455 | grapas | 23 |

Figura. Descomposición errónea da relación PEDIDO

Para recuperar posteriormente a información completa dos pedidos dun cliente xuntaríamos ambas táboas polo nome do cliente (o que se denomina operación de reunión ou join), obtendose o seguinte resultado:

| PEDIDO DE CLIENTE | | | | | | | PEDIDOS | | | | |
|-------------------|-------------|--------|-------------------------|-------------|-------------|----------|-----------|--------------|-------|------------|-------------|
| idCliente | nomeCliente | cidade | telefonos | codProducto | descripcion | cantidad | CodPedido | prezoUnidade | total | data | nomeCliente |
| D482049 | Xen | Lugo | 982098765; 653876098 | P111894 | Silla | 10 | 34355 | 180 | 1800 | 25/08/2013 | Xen |
| D482049 | Xen | Lugo | 982098765; 653876098 | P111894 | Silla | 10 | 86987 | 0,9 | 29,7 | 12/12/2012 | Xen |
| D482049 | Xen | Lugo | 982009988; 618098345 | P748786 | bolígrafo | 33 | 86987 | 180 | 1800 | 25/08/2013 | Xen |
| D482049 | Xen | Lugo | 982009988; 618098345 | P748786 | bolígrafo | 33 | 86987 | 0,9 | 29,7 | 12/12/2012 | Xen |

Empregamos o atributo nome para localizar toda a información dos pedidos do cliente Xen

Tuplas
espúreas

Figura 2.9. Reunión da descomposición da relación PEDIDO que xenera tuplas espúreas

O mesturarse cada unha das tuplas de PEDIDO DE CLIENTE con PEDIDOS empregando o nome do cliente (neste exemplo “Xen”) aparecen máis tuplas asociadas a “Xen” pero con menos información. Estas tuplas que non existen na relación inicial e que son incorrectas, reciben o nome de “espúreas”.

Para resolver estes problemas debemos realizar un bo deseño, seguindo os pasos de transformación a partir dun modelo conceptual entidade-interrelación e comprobando se as relacións resultantes están normalizadas, e no caso de que non o estean descompondo o esquema de relación en varios esquemas de relación. Os esquemas resultantes cumprirán certas formas normais (preservándose a unión ou combinación de relacións). Deste xeito, minimizarase o grao de vulnerabilidade dunha táboa a inconsistencias e anomalías lóxicas.

1.4 Dependencias funcionais

O concepto de dependencia funcional é esencial para o deseño de esquemas de bases de datos sen redundancia.

Definimos a dependencia funcional (DF) da seguinte forma: Dado un esquema relacional $R(a_1, a_2, \dots, a_N)$ e sexan α e β dos conxuntos diferentes de atributos de dito esquema. O feito de que α determine funcionalmente a β , e que β dependa funcionalmente de α se expresa: $\alpha \rightarrow \beta$.

Pódese simplificar a definición dicindo que dados uns valores para os atributos de α , pódense determinar os correspondentes valores dos atributos β . Así, chámase α ao conxunto determinante e a β o conxunto dependente.

En xeral, unha dependencia funcional (df) $\alpha \rightarrow \beta$ é unha restrición entre dous conxuntos de atributos (descriptores) α e β e é unha propiedade inherente ao contido semántico dos datos, que se debe cumprir para calquera extensión dunha relación r co esquema de relación R .

A inherencia ao contido semántico significa que unha dependencia funcional non pode ser demostrada, pero si afirmada por observación da realidade.

A súa utilidade é a de describir mellor un esquema de relación, mediante a especificación de restricións (sobre atributos) que deben cumprirse sempre.

Definición formal de dependencia funcional (DF)

Sexa R un esquema de relación, no que o seu conxunto de atributos é A , e sexan os descriptores α e β subconxuntos de A , dise que

$\alpha \rightarrow \beta$ se " $t_1, t_2 \in r(R)$ e $t_1[\alpha] = t_2[\alpha]$ entón $t_1[\beta] = t_2[\beta]$ ".

β depende funcionalmente de α si dadas dous tuplas t_1 e t_2 (da extensión dunha relación $r(R)$) que teñan os mesmos valores nos atributos que forman α , tamén teñen os mesmos valores para os atributos de β (**Tamén se di que α determina funcionalmente a β**).

Dito doutro xeito que a un valor dado de α correspóndelle un único valor de β e isto sucede independentemente do instante considerado, se coñecemos os valores do compoñente α dunha tupla se coñecerán os valores del compoñente β desa mesma tupla (os valores do compoñente α dunha tupla determinan funcionalmente dos valores do compoñente β).

D.F. é una aserción ou afirmación sobre todos os valores posibles e non sobre os valores actuais, polo tanto, caracteriza a intensión e non a extensión de un esquema de relación.

Débese clarexar que $\alpha \rightarrow \beta$ **non** significa que “dado α poida deducirse o valor de β ”, ademais que $\alpha \rightarrow \beta$, dous tuplas son iguais en α son iguais en β non indica nada acerca de si $\beta \rightarrow \alpha$, pode ocorrer que sendo diferentes en α sexan iguais en β .

Un esquema relacional utilizando dependencias funcionais podería definirse como $r(R, f)$, con R conxunto de atributo e f conxunto de dependencias funcionais. Por exemplo:

$Actor = (\{idActor, nome, apleido1, apelido2\}, f)$.

Un atributo ou conxunto de atributos X dunha relación utilizando dependencias funcionais cumpre que $X \rightarrow A_1, A_2, \dots, A_n$.

Non existe ningún atributo ou conxunto de atributos Y , contido en X , tal que $Y \rightarrow A_1, A_2, \dots, A_n$, é dicir, a **clave é o conxunto mínimo de atributos que determina a tódolos demais**.

Por exemplo o nome dunha persoa depende funcionalmente do DNI; é dicir para un DNI concreto só hai un nome posible. Na relación do exemplo, o $idCliente$ non ten dependencia funcional co $codPedido$, xa que para un mesmo $idCliente$ pode haber máis dun pedido posible, eero o $nomeCliente$ sí que depende do $idCliente$.

1.5 Proceso de normalización dunha relación

O proceso de normalización somete a un esquema de relación a unha serie de probas para "certificar" se pertence ou non a unha certa forma normal.

A normalización dos datos define o proceso no que os esquemas de relación que non cumpren as condicións se descompoñen, repartindo os seus atributos entre esquemas de relación máis pequenos que cumpren as condicións establecidas; deste xeito se garantizas que non ocorran anomalías de actualización.

A **descomposición** expresada máis formalmente **consiste na substitución da relación $R(A_1, A_2, \dots, A_n)$ por unha serie de relacións R_1, R_2, \dots, R_n obtidas mediante proxeccións de R e tales que a relación resultante das reunións de R_1, R_2, \dots, R_n teña o mesmo esquema que R .**

As propiedades que debe verificar unha boa descomposición son:

- **A propiedade de reunión sen perda**, é dicir, que a extensión das relacións normalizadas nos permitan obter todos os valores de R .
- **A propiedade de conservación das dependencias**, que asegura que todas as dependencias funcionais estean representadas nalgunha das relacións individuais resultantes.

Non todas as descomposición son válidas ou boas porque nalgunhas prodúcense perdas de información.