

数据库范式

约瑟夫·哈利特

2023 年 2 月 1 日



这到底是怎么回事？

数据库理论！ ▶ 到目

前为止,我们已经讨论了如何涂鸦数据库设计... ▶ 我们已经讨论了如何在
SQL 中创建表

这次:我们如何
设计易于使用的表格？

让我们再次从我们的记录数据库开始.....

我们可以按如下方式存储数据:

艺术家

专辑

披头士

黄色潜水艇、白色相簿、橡胶灵魂

牛奶罐

让它变得甜蜜

德累斯顿娃娃 是弗吉尼亚,不是弗吉尼亚,德累斯顿娃娃

请不。

这是一个糟糕的主意

►是的,我们有一张大桌子,看起来更整洁►但实际上用它来做任何事情要困难得多

例如: ►每个艺

术家有多少张专辑? ►将1993年以后Prince的所有专辑改为by a Love Symbol ►有多少艺人拥有同名专辑?

范式

正常形式可以防止这种疯狂行为▶使用它们
需要纪律,并记住规则... ▶但在短期到中期内,为了您的理智,这是值得的

第一范式

每列应包含一个（且仅一个）值

每一行都描述了每个艺术家的多张专辑……

| 艺术家 | 专辑 |
|--------|---------------------|
| 披头士 | 黄色潜水艇、白色相簿、橡胶灵魂 |
| 牛奶罐 | 让它变得甜蜜 |
| 德累斯顿娃娃 | 是弗吉尼亚,不是弗吉尼亚,德累斯顿娃娃 |

第一范式

让我们解决这个问题...

艺术家

专辑

披头士

黄色潜水艇

披头士

白色相簿

披头士

橡胶灵魂

牛奶罐

让它变得甜蜜

德累斯顿娃娃是弗吉尼亚

德累斯顿娃娃没有弗吉尼亚

德累斯顿娃娃 德累斯顿娃娃

让我们向表中添加更多数据

| 艺术家 | 专辑 | 年份 总理 |
|---------------|-------------|--------------|
| 披头士 | 黄色潜水艇 | 1969 哈罗德·威尔逊 |
| 披头士 | 白色相簿 | 1968 哈罗德·威尔逊 |
| 披头士 | 橡胶灵魂 | 1965 哈罗德·威尔逊 |
| 牛奶罐 | 让它变得甜蜜 | 1999 托尼·布莱尔 |
| 德累斯顿娃娃是弗吉尼亚 | | 2006 托尼·布莱尔 |
| 德累斯顿娃娃没有弗吉尼亚 | | 2008 戈登·布朗 |
| 德累斯顿娃娃 德累斯顿娃娃 | 2003 托尼·布莱尔 | |

第二范式

每个非键属性完全依赖于键

在这种情况下,关键是艺术家、专辑

►如果你想像泰勒·斯威夫特一样重新发行你所有的专辑,那么今年也可以说是.....

首相依赖钥匙吗?

►不。将其放在不同的表中。

现在看起来像

| 艺术家 | 专辑 | 年 | 年份 总理 |
|---------------|--------------|------|--------------|
| 披头士 | 黄色潜水艇 | 1969 | 1969 哈罗德·威尔逊 |
| 披头士 | 白色相簿 | 1968 | 1968 哈罗德·威尔逊 |
| 披头士 | 橡胶灵魂 | 1965 | 1965 哈罗德·威尔逊 |
| 牛奶罐 | 甜蜜蜜 | 1999 | 1999 托尼·布莱尔 |
| 德累斯顿娃娃 | Yes Virginia | 2006 | 2006 托尼·布莱尔 |
| 德累斯顿娃娃没有弗吉尼亚 | | 2008 | 2008 戈登·布朗 |
| 德累斯顿娃娃 德累斯顿娃娃 | | 2003 | 2003 托尼·布莱尔 |

第三范式

每个非键属性必须提供有关键的事实,整个键,仅此而已
 钥匙;所以帮助我科德。

让我们在我们的总理表中添加一些额外的信息.....

| 年份 | 总理 | 生日 |
|------|---------|------------|
| 1969 | 哈罗德·威尔逊 | 1916-03-11 |
| 1968 | 哈罗德·威尔逊 | 1916-03-11 |
| 1965 | 哈罗德·威尔逊 | 1916-03-11 |
| 1999 | 托尼·布莱尔 | 1953年5月6日 |
| 2003 | 托尼·布莱尔 | 1953-05-06 |
| 2006 | 托尼·布莱尔 | 1953-05-06 |
| 2008 | 戈登·布朗 | 1951-02-20 |

我们的关键是（年份,总理）;生日取决于总理。

►所以每个非键都依赖于键.....

►所以2NF

但不是 3NF,因为生日并没有告诉你关于整个钥匙的事实.....只是总理。

那就分了吧!

年份 总理

1969 哈罗德·威尔逊

1968 哈罗德·威尔逊

1965 哈罗德·威尔逊

1999 托尼·布莱尔

2003 托尼·布莱尔

2006 托尼·布莱尔

2008 戈登·布朗

总理

生日

哈罗德·威尔逊 1916-03-11

托尼·布莱尔 1953-05-06

戈登·布朗 1951-02-20

为什么这样更好?

►现在,如果我们需要更改 PM 的生日(或有关该密钥的任何其他事实) ... ►...那么我们只需要在一个地方更改它。

其他正常形式...

Boyce-Codd 范式 3NF 的稍强形式.....

► 有时称为 3.5 范式

表的每个可能的候选键也在 3NF 中。 ► 将 3NF 表拆分为具有单个候选键的表以获得 3.5NF。

第四范式 如果表中的多个属性

依赖于同一个键， ► 那么这些属性也应该是依赖的 ► 否则将它们分成单独的表.....

第五范式 它是第四范式,您不能将其拆分为更多单独的表。

这一切都变得有点数学化了……

您可以查找每种范式的正式定义▶（而且您应该）

但只要你尽可能将事情分开,你通常会意外地达到至少 3NF。

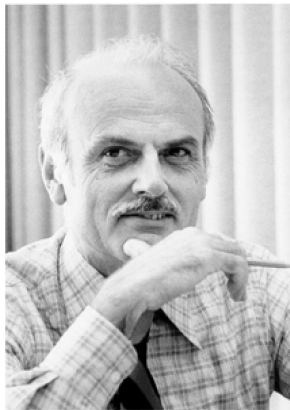
▶ …实际上来说,您可能会很好▶从长远来看,达到 5NF
确实会让事情变得更加灵活……▶但 3.5NF 数据库通常就足够好了。

最终设计是主观的（某种程度上）。▶ ……

但是灵活性的数学证明很好,对吧?

综上所述

每个非键属性都必须提供有关键的事实、整个键且仅提供键;所以帮助我科德。



特德·科德