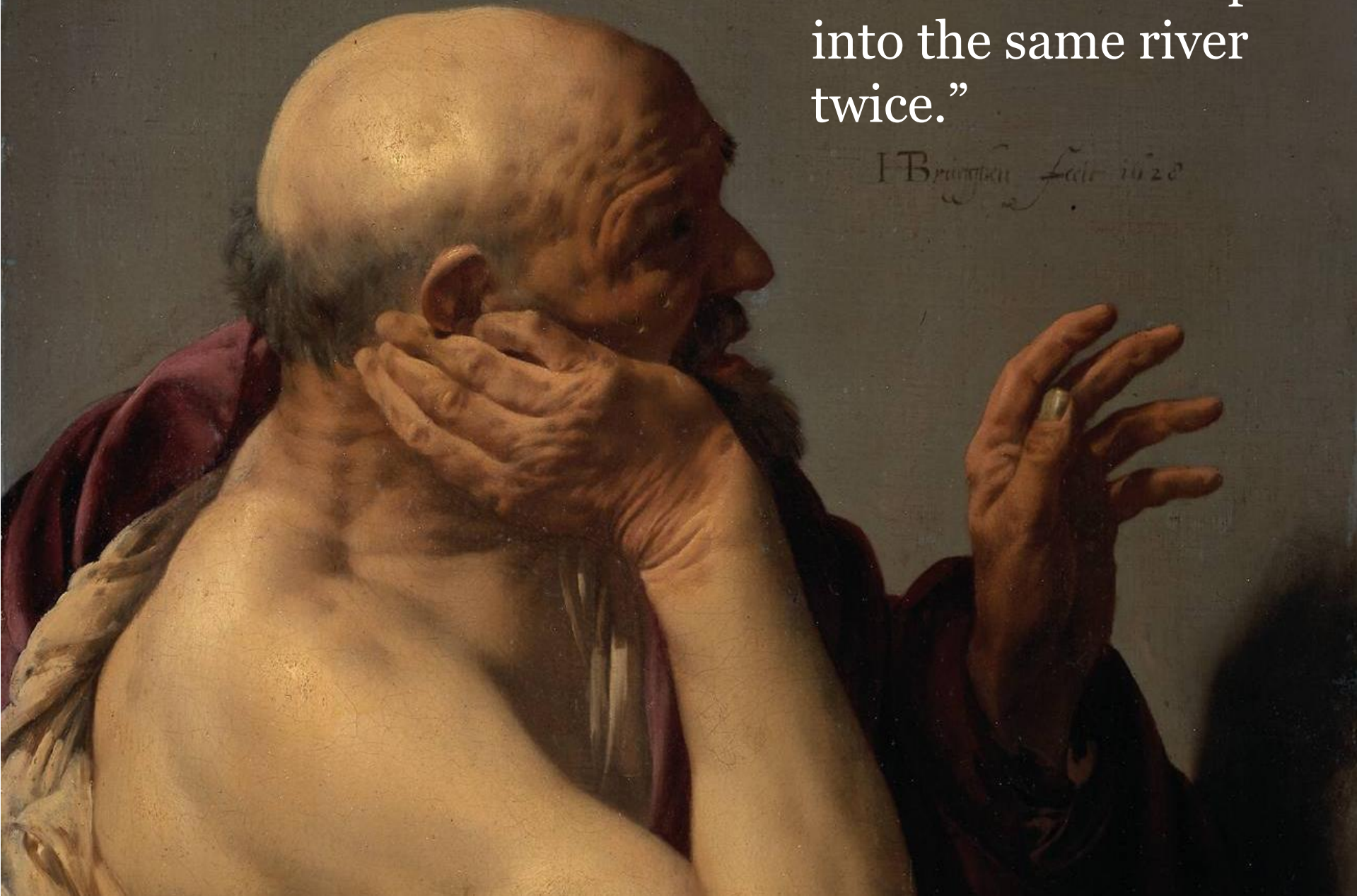




Lars Rönnbäck  
[ER09]

“You can never step  
into the same river  
twice.”

*I Brington Fall 1928*





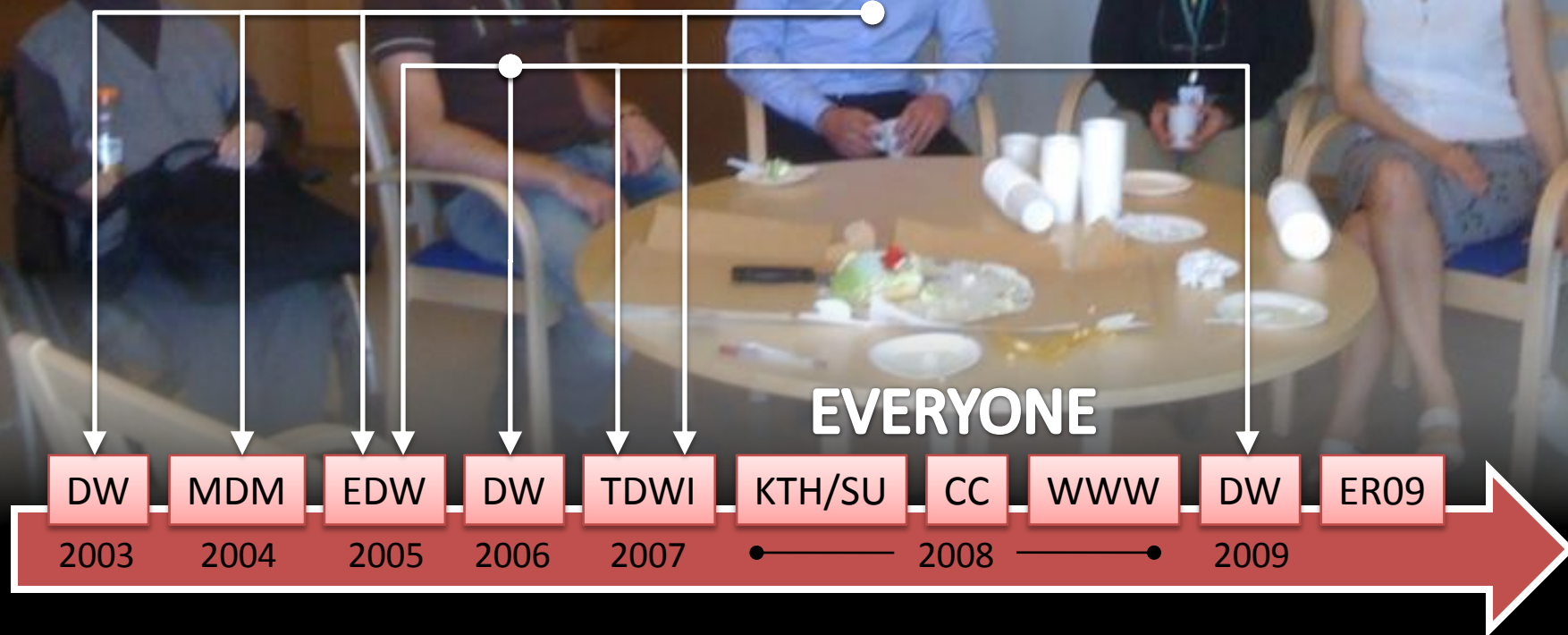
Paul  
Johannesson

Lars  
Rönnbäck

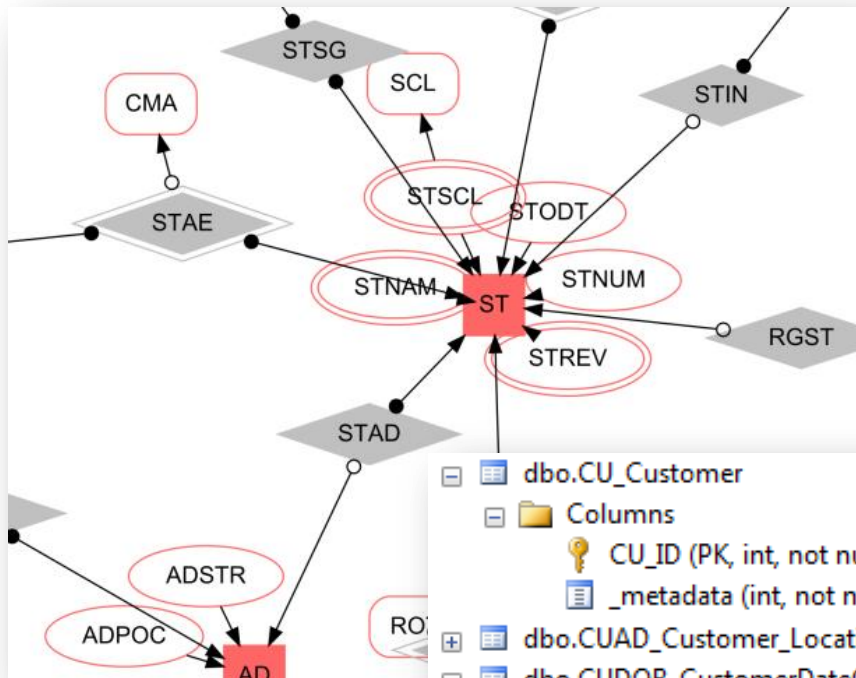
Olle  
Regardt

Maria  
Bergholtz

Petia  
Wohed



## Entity Relationship Modeling



## Sixth Normal Form

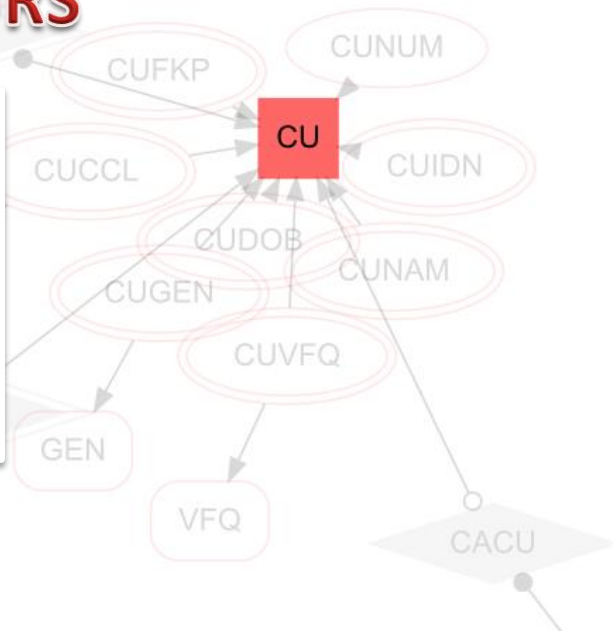
dbo.CU_Customer
Columns
CU_ID (PK, int, not null)
_metadata (int, not null)
dbo.CUAD_Customer_Location_Lives
dbo.CUDOB_CustomerDateOfBirth
Columns
CU_ID (PK, int, not null)
CUDOB_CustomerDateOfBirth (datetime, not null)
CUDOB_FromDate (PK, datetime, not null)
_metadata (int, not null)
dbo.CUNAM_CustomerName

```
select
    pCU.GEN_Gender,
    pCU.CUHAC_CustomerHairColor,
    COUNT(pCU.CUHAC_CustomerHairColor) as Customers
from
    pCU_Customer('1985-11-09') pCU
where
    pCU.CUDOB_CustomerDateOfBirth < '1980-01-01'
group by
    pCU.GEN_Gender,
    pCU.CUHAC_CustomerHairColor
```

## Temporal Databases

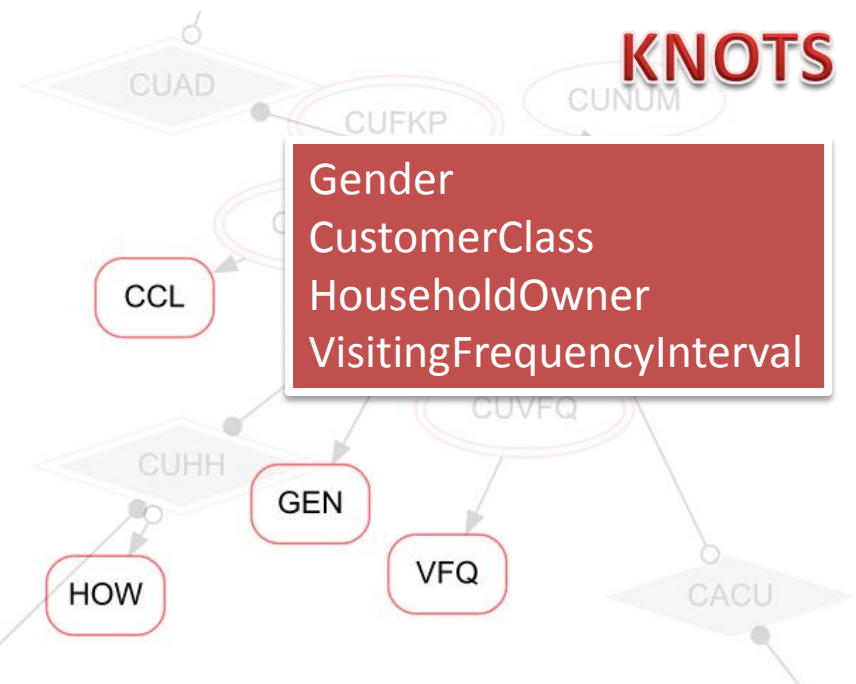
# ANCHORS

Customer  
Store  
Purchase  
Item  
PriceList  
Inventory



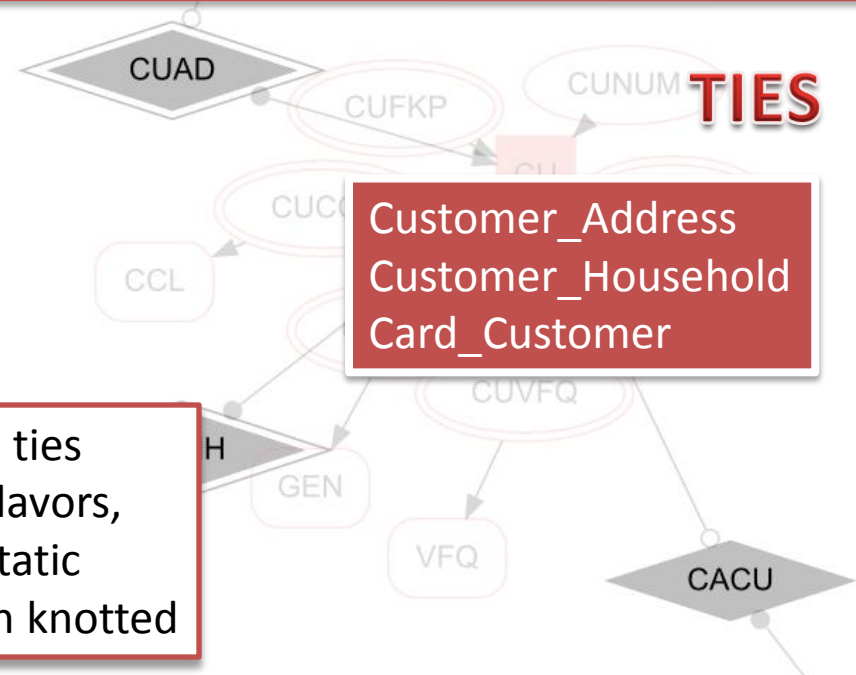
# KNOTS

Gender  
CustomerClass  
HouseholdOwner  
VisitingFrequencyInterval



# TIES

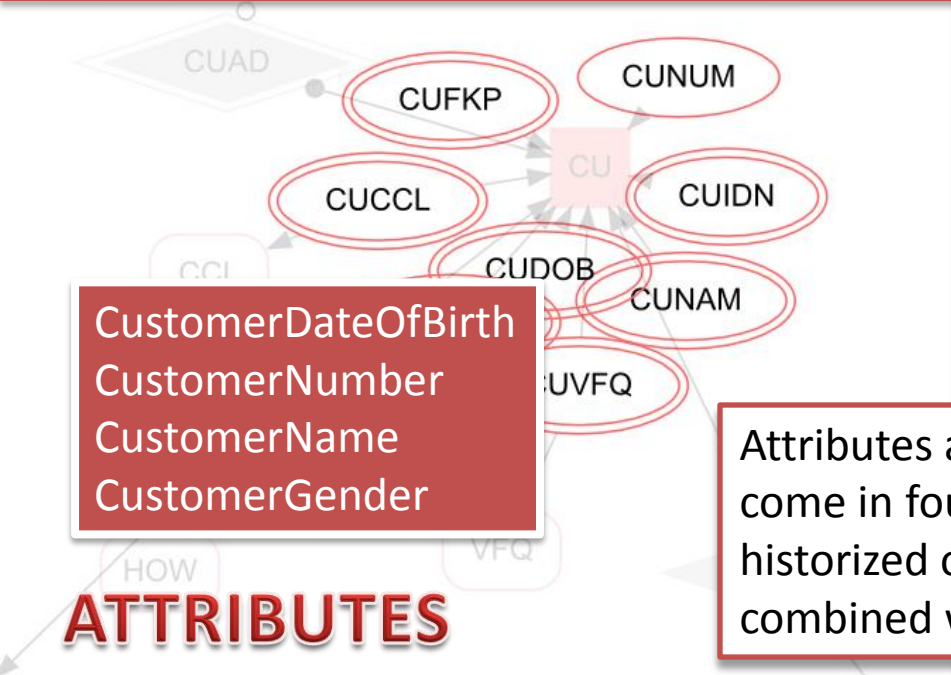
Customer\_Address  
Customer\_Household  
Card\_Customer



CustomerDateOfBirth  
CustomerNumber  
CustomerName  
CustomerGender

# ATTRIBUTES

Attributes and ties  
come in four flavors,  
historized or static  
combined with knotted



# ANCHORS

**select** top 5 \* **from** CU\_Customer

CU_ID
1
2
3
4
5

# KNOTS

**select** top 5 \* **from** GEN\_Gender

GEN_ID	GEN_Gender
1	Male
2	Female

# ATTRIBUTES

**select** top 5 \* **from** CUDOB\_CustomerDateOfBirth

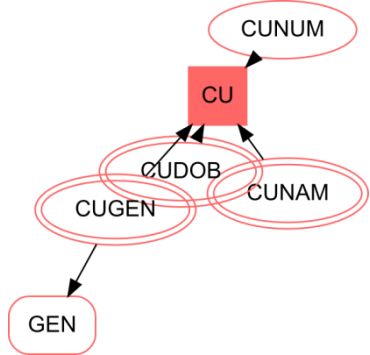
CU_ID	CUDOB_CustomerDateOfBirth
1	1905-03-02
2	1905-07-02
3	1908-09-14
4	1910-02-03
5	1912-04-01

# TIES

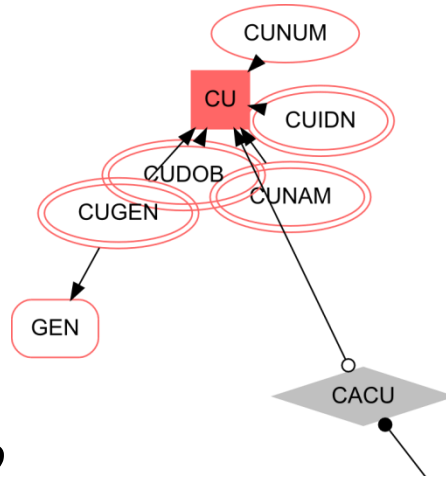
**select** top 5 \* **from** CUHH\_Customer\_Household

CU_ID	HH_ID	HOW_ID	CUHH_FromDate
1	1	1	2009-02-13
1	895	0	2009-09-21
2	2	1	2006-10-17
3	3	1	2002-08-20
4	4	1	1993-08-29

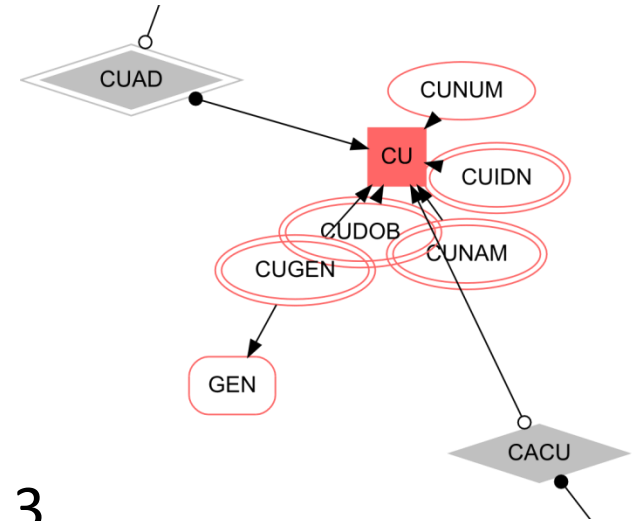
# Model Evolution



1

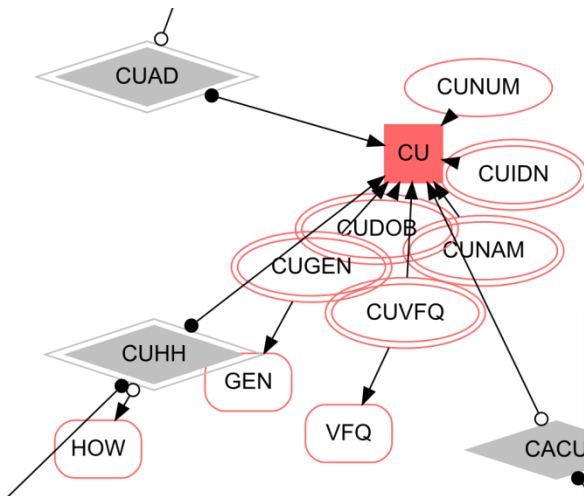


2



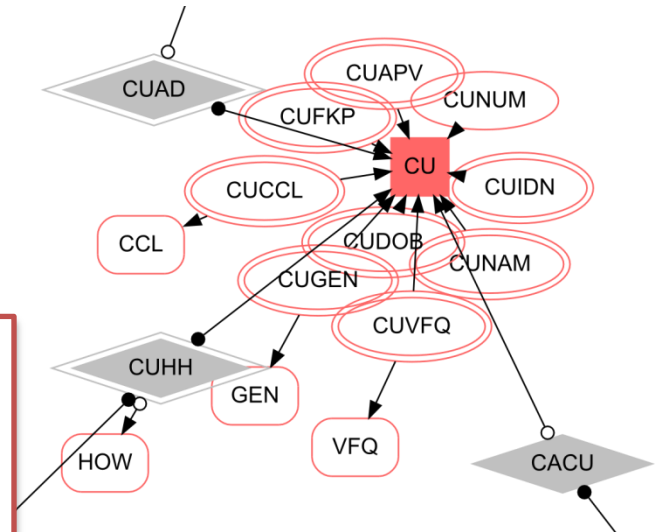
3

4



All previous versions of the schema are present and were never modified, allowing extensions to be made "online".

5





# Latest View

- [-] Views
  - [+] System Views
- [-] [Table Icon] dbo.ICU\_Customer
  - [-] Columns
    - [Table Icon] CU\_ID (int, not null)
    - [Table Icon] CUNUM\_CustomerNumber (int, null)
    - [Table Icon] CUNAM\_CustomerName (varchar(255), null)
    - [Table Icon] CUNAM\_FromDate (datetime, null)
    - [Table Icon] CUDOB\_CustomerDateOfBirth (datetime, null)
    - [Table Icon] CUDOB\_FromDate (datetime, null)
    - [Table Icon] CUFKP\_CustomerFirstKnownPurchase (datetime, null)
    - [Table Icon] CUFKP\_FromDate (datetime, null)
    - [Table Icon] CUIDN\_CustomerIdentification (varchar(255), null)
    - [Table Icon] CUIDN\_FromDate (datetime, null)
    - [Table Icon] CUAPV\_CustomerAmountPerVisit (money, null)
    - [Table Icon] CUAPV\_FromDate (datetime, null)
    - [Table Icon] CCL\_CustomerClassification (varchar(255), null)
    - [Table Icon] CUCCL\_FromDate (datetime, null)
    - [Table Icon] GEN\_Gender (varchar(255), null)
    - [Table Icon] CUGEN\_FromDate (datetime, null)
    - [Table Icon] VFQ\_VisitationFrequency (varchar(255), null)
    - [Table Icon] CUVFQ\_FromDate (datetime, null)

Joins all attributes and finds the attribute row with the latest FromDate if historized

# Point-in-Time Function

- [-] Programmability
  - [+] Stored Procedures
- [-] Functions
  - [-] Table-valued Functions
    - [-] [Table Icon] dbo.pCU\_Customer
      - [-] Parameters
        - [@ Icon] @timepoint (datetime, No default)

Joins all attributes and finds the attribute row with the latest FromDate earlier or on the given timepoint if historized

```
select
    pCU.GEN_Gender,
    pCU.CUHAC_CustomerHairColor,
    COUNT(pCU.CUHAC_CustomerHairColor) as Customers
from
    pCU_Customer('1985-11-09') pCU
where
    pCU.CUDOB_CustomerDateOfBirth < '1980-01-01'
group by
    pCU.GEN_Gender,
    pCU.CUHAC_CustomerHairColor
```

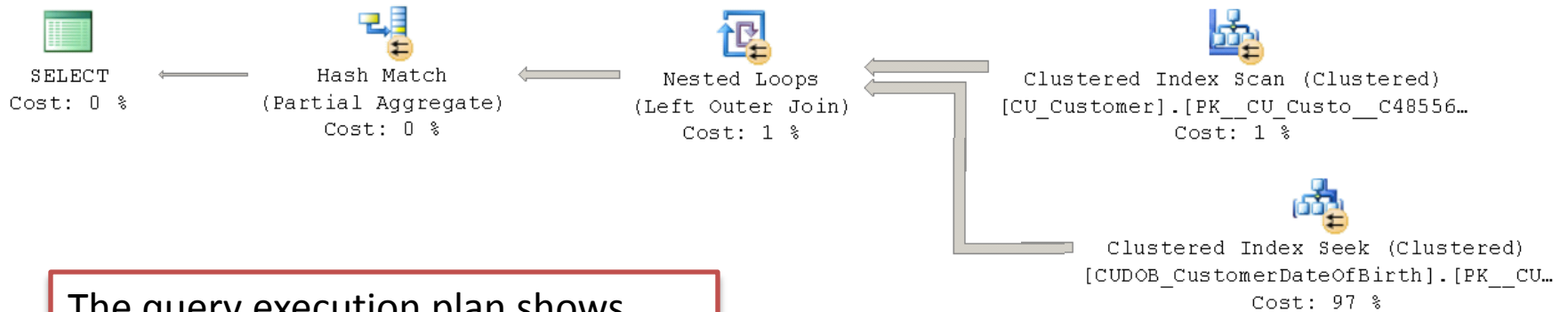


```

select
    YEAR(1CU.CUDOB_CustomerDateOfBirth) as YearOfBirth,
    COUNT(1CU.CUDOB_CustomerDateOfBirth) as Customers
from
    1CU_Customer 1CU
where
    YEAR(1CU.CUDOB_CustomerDateOfBirth) between 1950 and 1954
group by
    YEAR(1CU.CUDOB_CustomerDateOfBirth)
order by
    YEAR(1CU.CUDOB_CustomerDateOfBirth)

```

YearOfBirth	Customers
1950	20615
1951	19282
1952	20003
1953	19782
1954	19249



The query execution plan shows that only two tables are touched (the anchor and the selected attribute) despite of the fact that several others are joined into the view we are using.

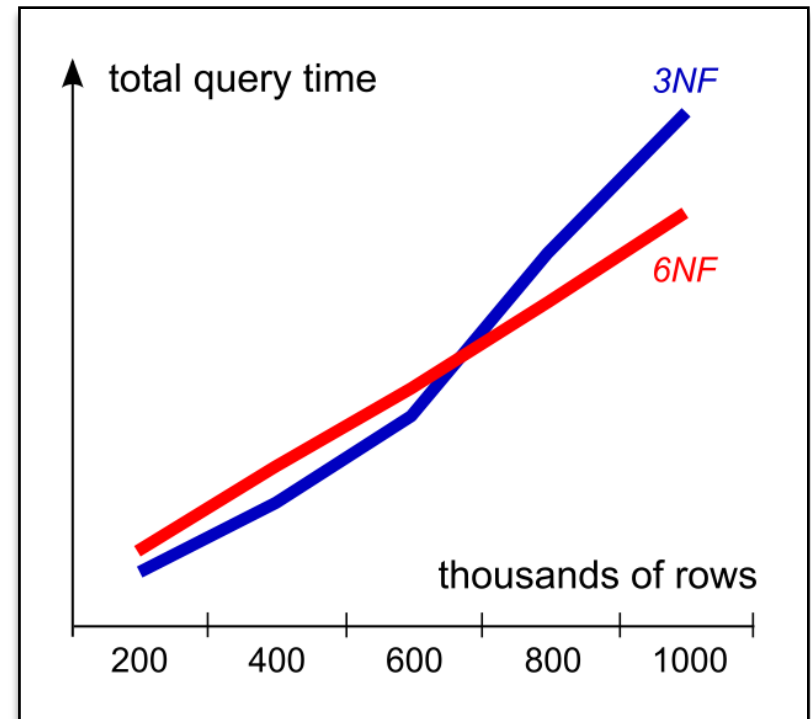
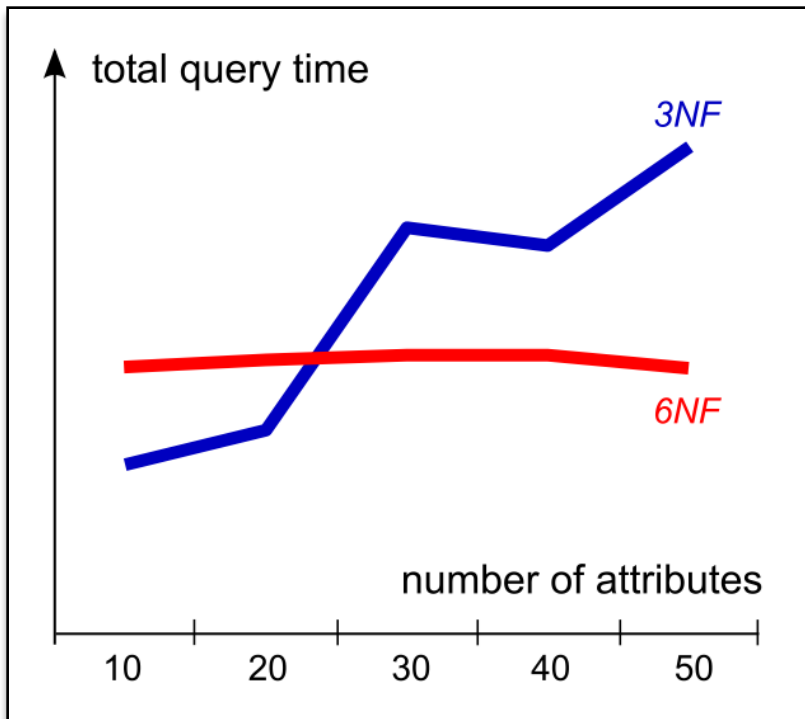
# Table Elimination

The query optimizer will remove table  $T$  from the execution plan of a query if the following two conditions are fulfilled:

- no column from  $T$  is explicitly selected
- the number of rows in the returned data set is not affected by the join with  $T$

## Support

Microsoft SQL Server  
Oracle  
IBM DB2  
PostgreSQL  
MariaDB (fork of MySQL)  
Teradata (partial)



```

<knot mnemonic="ROI" name="Root" valueType="varchar(255)" idType="bit"/>
<knot mnemonic="ISZ" name="ItemSize" valueType="varchar(255)" idType="tinyint"/>
<knot mnemonic="ICO" name="ItemColor" valueType="varchar(255)" idType="tinyint"/>
<knot mnemonic="SGL" name="StoreGroupLevel" valueType="varchar(255)" idType="tinyint"/>
<knot mnemonic="IGL" name="ItemGroupLevel" valueType="varchar(255)" idType="tinyint"/>
<knot mnemonic="AEL" name="AreaLevel" valueType="varchar(255)" idType="tinyint"/>

```

```

<!-- ***** ANCHORS & ATTRIBUTES ***** -->
<anchor mnemonic="OF" name="Offer" idType="int">
  <attribute mnemonic="OFNUM" name="OfferNumber" valueType="int"/>
  <attribute mnemonic="OFVFR" name="OfferValidFrom" valueType="datetime"/>
  <attribute mnemonic="OFVTO" name="OfferValidTo" valueType="datetime"/>
  <attribute mnemonic="OFCON" name="OfferCondition" valueType="varchar(255)"/>
  <attribute mnemonic="OFEXP" name="OfferExplanation" valueType="varchar(255)"/>
  <attribute mnemonic="OFOTP" name="OfferType">
    <relation reference="OTP"/>
  </attribute>
  <attribute mnemonic="OFTGP" name="OfferTargetGroup">
    <relation reference="TGP"/>
  </attribute>
</anchor>
<anchor mnemonic="IN" name="Inventory" idType="int">
  <attribute mnemonic="INQUA" name="InventoryQuantity" valueType="int"/>
  <attribute mnemonic="INSTI" name="InventoryItemID" valueType="int"/>
</anchor>
<anchor mnemonic="RG" name="Register" idType="int">
  <attribute mnemonic="RGNUM" name="RegisterNumber" valueType="int"/>
  <attribute mnemonic="RGMAC" name="RegisterMAC" valueType="varchar(255)"/>
  <attribute mnemonic="RG RTP" name="RegisterRTP" valueType="varchar(255)"/>
  <relation reference="RTP"/>
</anchor>

```

The scripts for setting up the database, including all views and functions, can be automatically generated from a compact XML description.

### Pseudo loading code given "wide" source data:

- Check if there already is an associated surrogate key for each natural key
- For unknown individuals
  - Create and associate surrogate keys
  - Directly insert data into all relevant tables  
(most tables including the anchor)
- For known individuals
  - If this is a delta file, directly insert data into all relevant tables
  - If this is not a delta file, check if the value in the source differs from the latest value in the destination and insert if the data is new  
(few tables excluding the anchor)

Data loading templates can be made in which only the names of the tables and the join with the natural key have to be changed.

## **Ease of Modeling**

Simple concepts and notation  
Historization by design  
Iterative and incremental development  
Reduced translation logic

## **Simplified Maintenance**

Ease of temporal querying  
Absence of null values  
Reusability and automation  
Asynchronous arrival of data

## **High Performance**

High run-time performance  
Efficient storage  
Parallelized physical media access

[www.anchormodeling.com](http://www.anchormodeling.com)



**Affecto**

