


SCD (Slowly Changing Dimension)

☰ Tags	
📎 Image	



Медленно меняющимися измерения — измерения (колонки) значения которых со временем могут изменяться, но не часто. При этом изменяются неключевые атрибуты измерения.

Пример: «Клиент», т.к. клиент может изменить фамилию, номер паспорта и другие атрибуты, но происходит это очень редко.

Таким образом, такие изменения случаются хотя и редко, но являются непредсказуемыми, происходят под воздействием внешних факторов случайного характера и не укладываются в общую схему работы с данными. Это способно создать определенные проблемы, например, нарушить ссылочную целостность данных.

Основные типы SCD:

▼ SCD Type 0 (Стационарное измерение)



SCD 0 — Изменения в данных вообще не отслеживаются. Сохраняется только начальное состояние данных: Любые изменения в данных просто игнорируются, и они не обновляются.



Пример: Клиент с ID 1 имеет адрес "123 St." и даже если он переедет, эта информация не обновляется — старая запись сохраняется.

Используется: Используется, когда исторические данные **не должны меняться и всегда должны сохранять своё первоначальное значение**. Например, идентификатор клиента, который должен быть неизменным.

▼ SCD Type 1 (Перезапись)



SCD 1 — изменение данных приводит к **полной замене старого значения** на новое, без сохранения истории. Исторические данные не сохраняются, всегда представляется только актуальное состояние.



Пример: Если клиент изменил свой адрес, старый адрес будет заменён новым.



Плюсы: Простота реализации и управления данными.



Минусы: Потеря исторических данных.

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	IL

▼ SCD Type 2 (Сохранение истории)



SCD 2 — сохраняет полную историю изменений данных. Каждое изменение значения создаёт новую запись с новым значением, при этом старая запись остаётся в базе данных.

Обычно для каждой записи добавляются специальные столбцы: дата начала действия (`effective_date`), дата окончания действия (`end_date`) и статус актуальности (`current_flag`). Это позволяет хранить информацию обо всех изменениях и видеть, когда и как изменялись данные.



Пример: Если клиент изменил свой адрес, старая запись остаётся в базе данных с указанием срока действия, а новая запись создаётся с текущим адресом.



Плюсы:

- Полное сохранение истории изменений данных.
- Возможность проводить анализ на основе прошлых данных.



Минусы:

- Увеличение объёма данных, так как на каждое изменение создаётся новая запись.

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	Start_Date	End_Date
123	ABC	Acme Supply Co	CA	01-Jan-2000	21-Dec-2004
124	ABC	Acme Supply Co	IL	22-Dec-2004	

▼ SCD Type 3 (Атрибут предыдущего состояния)



SCD 3 — сохраняется только текущее и одно предыдущее значение атрибута. В базе данных создаются отдельные столбцы для хранения текущего значения и предыдущего значения. Таким образом, сохраняется ограниченная история изменений данных.



Пример: Если клиент изменил свой адрес, его старый адрес будет сохранён в столбце "старый адрес", а новый адрес — в столбце "новый адрес".



Плюсы: Простота в сравнении с SCD Type 2 и экономия места.



Минусы: Ограниченное хранение истории (только одно предыдущее значение).

Supplier_Key	Supplier_Code	Supplier_Name	Original_Supplier_State	Effective_Date	Current_Supplier_State
123	ABC	Acme Supply Co	CA	22-Dec-2004	IL

▼ SCD Type 4 (Историческая таблица)



SCD 4 — используется отдельная таблица для хранения всех исторических данных. Основная таблица содержит только текущие значения, а все предыдущие значения хранятся в отдельной исторической таблице.



Пример: Если клиент изменяет свой адрес, новая информация сохраняется в основной таблице, а старая информация перемещается в таблицу истории.



Плюсы: Облегчает управление историческими данными.

✗ Минусы: Усложняет структуру базы данных (две таблицы вместо одной).

SCD Type 4 current table

cust_id	customer_number	first_name	last_name
1	10001	Context	Wonderful


SCD Type 4 history table

cust_id	customer_number	first_name	last_name	effective_date
1	10001	Kontext	Wonderful	2022-01-01
1	10001	Context	Wonderful	2022-07-01

▼ SCD Type 6 (Гибрид Type 1, 2 и 3)



SCD 6 — является комбинацией типов **1, 2 и 3**. Он сохраняет и текущие значения, и исторические данные, как в Type 2, и включает атрибут предыдущего состояния, как в Type 3. Поддерживает все три вида данных: текущие, предыдущие и полную историю изменений.

 **Пример:** При изменении адреса клиента новая запись создаётся в соответствии с Type 2, текущий адрес сохраняется в таблице, а предыдущий — в отдельном столбце (Type 3).

✓ Плюсы: Полноценное решение для управления историческими данными и текущими состояниями.

✗ Минусы: Увеличение сложности и объёма данных.

ProductCode	↕ Cost	↕ HistoricalCost	↕ StartDate	↕ EndDate	↕ Current	↕
C123	125.50	125.50	20110228	99991231	Y	
F112	2425.00	2365.00	20120101	20120701	N	
G101	19.25	19.25	20110930	99991231	Y	
G101	19.25	21.25	20110501	20110930	N	
J245	550.50	450.50	20110430	20120701	N	

SCD (на английском)