



عنوان مقاله: مقدمه‌ای بر SCD در Power BI

نویسنده مقاله: تیم فنی نیک‌آموز

تاریخ انتشار: ۱۶ بهمن ۱۴۰۱

منبع: <https://nikamooz.com/introduction-to-scd>

تغییر آهسته ابعاد (SCD) یک مفهوم انبار داده است که توسط Ralph Kimball ابداع شده است. مفهوم SCD با انتقال مجموعه خاصی از داده‌ها، از یک وضعیت به وضعیت دیگر سروکار دارد. تصور کنید ما یک سیستم منابع انسانی (HR) داریم. Stephen Jiang یک مدیر فروش است که ۱۰ نماینده فروش را در تیم خود مدیریت می‌کند. تصویر زیر نمونه داده‌ها را نشان می‌دهد:

Full Name	Title	Manager
Stephen Jiang	North American Sales Manager	Brian Welcker
David Campbell	Sales Representative	Stephen Jiang
Garrett Vargas	Sales Representative	Stephen Jiang
Jillian Carson	Sales Representative	Stephen Jiang
José Saraiva	Sales Representative	Stephen Jiang
Linda Mitchell	Sales Representative	Stephen Jiang
Michael Blythe	Sales Representative	Stephen Jiang
Pamela Ansman-Wolfe	Sales Representative	Stephen Jiang
Shu Ito	Sales Representative	Stephen Jiang
Tete Mensa-Annan	Sales Representative	Stephen Jiang
Tsvi Reiter	Sales Representative	Stephen Jiang

حالا در نظر بگیرید Stephen Jiang به سمت معاونت فروش ارتقا یافته است؛ بنابراین تعداد تیم او از ۱۰ به ۱۷ افزایش یافت. Stephen همان شخص است، اما نقش او اکنون تغییر کرده است؛ همان‌طور که در شکل زیر نشان داده شده است:

Full Name	Title	Manager
Stephen Jiang	Vice President of Sales	Patti Fernandez
Amy Alberts	European Sales Manager	Stephen Jiang
Roger Hamilton	North American Sales Manager	Stephen Jiang
Syed Abbas	Pacific Sales Manager	Stephen Jiang
David Campbell	Sales Representative	Roger Hamilton
Garrett Vargas	Sales Representative	Roger Hamilton
Jae Pak	Sales Representative	Amy Alberts
Jillian Carson	Sales Representative	Roger Hamilton
José Saraiva	Sales Representative	Roger Hamilton
Linda Mitchell	Sales Representative	Roger Hamilton
Lynn Tsoflias	Sales Representative	Syed Abbas
Michael Blythe	Sales Representative	Roger Hamilton
Pamela Ansman-Wolfe	Sales Representative	Roger Hamilton
Rachel Valdez	Sales Representative	Amy Alberts
Ranjit Varkey Chudukatil	Sales Representative	Amy Alberts
Shu Ito	Sales Representative	Roger Hamilton
Tete Mensa-Annan	Sales Representative	Roger Hamilton
Tsvi Reiter	Sales Representative	Roger Hamilton

مثال دیگر زمانی است که در سیستم فروش، آدرس مشتری تغییر می‌کند. باز هم مشتری همان است، اما آدرس او در حال حاضر تغییر کرده است. از نقطه نظر انبار داده، بسته به نیازهای کسب و کار، گزینه‌های مختلفی برای مدیریت این موارد در انبار داده‌ها داریم. راه‌حل‌ها ما را به سمت انواع مختلف SCD هدایت می‌کند. توجه به این نکته ضروری است که داده‌ها در سیستم‌های تراکنش (در مثال‌های ما، سیستم منابع انسانی یا سیستم فروش) تغییر می‌کنند. ما داده‌ها را از سیستم‌های تراکنش از طریق فرآیندهای استخراج، تبدیل و بارگذاری (ETL) منتقل و تبدیل می‌کنیم و آن‌ها را در یک انبار داده ذخیره می‌کنیم. انبار داده جایی است که مفهوم SCD معنا پیدا می‌کند. SCD در مورد چگونگی اعمال کردن تغییرات در سیستم‌های منبع داده‌ها در انبار داده به کار می‌رود. این نوع تغییرات در سیستم منبع، گاهی اوقات اتفاق می‌افتد و همیشگی نیست. به همین علت، اصطلاح تغییر آهسته برای این مفهوم به کار رفته است. بسیاری از انواع SCD در طول این سال‌ها پیاده‌سازی شده‌اند که پرداختن به همه آن‌ها از حد این مقاله خارج است؛ اما برای نمونه، سه نوع اول را به شرح زیر پوشش می‌دهیم.

### SCD type zero (SCD 0)

با این نوع SCD، ما تمام تغییرات یک dimension را نادیده می‌گیریم. بنابراین، هنگامی که آدرس محل سکونت یک شخص در سیستم منبع تغییر می‌کند، مانند مثال سیستم منابع انسانی، ما این dimension را در انبار داده خود تغییر نمی‌دهیم. به عبارت دیگر، تغییرات درون منبع داده را نادیده می‌گیریم. SCD 0 به عنوان fixed dimensions نیز نامیده می‌شود.

### SCD type 1 (SCD 1)

با SCD type 1، داده‌های قدیمی را با جدید بازنویسی می‌کنیم و در واقع، داده‌های قدیمی نگه داشته نمی‌شوند. یک مثال عالی از این نوع، زمانی است که کسب و کار به آدرس قدیمی مشتری نیاز ندارد و فقط باید آدرس فعلی مشتری را حفظ کند.

### SCD type 2 (SCD 2)

با SCD type 2، تاریخچه تغییرات داده را در انبار داده نگه می‌داریم. یک مثال برای این نوع، زمانی است که کسب و کار نیاز به حفظ آدرس‌های قدیمی و فعلی مشتری دارد. در سناریوی SCD 2، ما باید تاریخچه را حفظ کنیم؛ بنابراین هر زمان که یک سیستم تراکنش تغییر کرد، یک ردیف جدید از داده‌ها را در انبار داده درج می‌کنیم.

درج یک ردیف جدید از داده‌ها باعث داده‌های تکراری در انبار داده می‌شود؛ در نتیجه نمی‌توانیم از ستون CustomerKey به عنوان کلید اصلی استفاده کنیم. به همین علت، باید مجموعه جدیدی از ستون‌ها را به صورت زیر معرفی کنیم:

- یک ستون کلید جدید اضافه می‌کنیم که منحصر به فرد بودن ردیف‌ها را در Customers dimension تضمین می‌کند. این ستون کلید، به سادگی یک ایندکس را ارائه می‌دهد که هر ردیف از داده‌های ذخیره شده در یک dimension انبار داده را نشان می‌دهد. کلید جدید به اصطلاح کلید جانشین (surrogate key) نامیده می‌شود. درحالی که surrogate key تضمین می‌کند که هر ردیف در dimension، منحصر به فرد است، ما

همچنان باید کلید اصلی سیستم منبع را نیز حفظ کنیم. طبق تعریف، کلیدهای اصلی سیستم منبع در مفاهیم انبار داده، کلیدهای تجاری (business keys) یا کلیدهای جایگزین (alternate keys) نامیده می‌شوند.

- ستون‌های Start Date و End Date نشان دهنده بازه زمانی است که طی آن بازه، یک ردیف از داده‌ها در وضعیت فعلی خود قرار دارد.
- ستون دیگری به نام status اضافه می‌کنیم که وضعیت هر ردیف داده را نشان می‌دهد.

SCD 2 رایج‌ترین نوع SCD است.

مثال قبلی خود را در مورد Stephen Jiang که از مدیر فروش به معاونت فروش ارتقا یافت، مجدداً مرور می‌کنیم. شکل زیر داده‌ها را قبل از اینکه Stephen ارتقا شغلی داشته باشد را نشان می‌دهد:

EmployeeKey	Full Name	Title	Manager	EmployeeBusinessKey	Start Date	End Date	Status
272	Stephen Jiang	North American Sales Manager	Stephen Jiang	502097814	4/08/2010		Current
273	Wanida Benshoof	Marketing Assistant	Debra Berger	323403273	7/08/2010		Current
274	Sharon Salavaria	Design Engineer	Roberto Tamburello	56920285	18/08/2010		Current
275	John Wood	Marketing Specialist	Debra Berger	222969461	7/09/2010		Current
276	Mary Dempsey	Marketing Assistant	Debra Berger	52541318	14/09/2010		Current

ستون EmployeeKey کلید جایگزین (Surrogate) برای dimension است و ستون EmployeeBusinessKey کلید تجاری (Business) که همان کلید اولیه customer در سیستم منبع است. ستون Start Date، تاریخ شروع کار Stephen Jiang به عنوان مدیر فروش آمریکای شمالی را نشان می‌دهد، ستون End Date خالی است و ستون Status، مقدار Current را نشان می‌دهد. حال نگاهی به داده‌ها پس از دریافت ارتقا توسط Stephen می‌اندازیم که در شکل زیر نشان داده شده است:

EmployeeKey	Full Name	Title	Manager	EmployeeBusinessKey	Start Date	End Date	Status
272	Stephen Jiang	North American Sales Manager	Brian Welcker	502097814	4/08/2010	12/10/2012	
273	Wanida Benshoof	Marketing Assistant	Debra Berger	323403273	7/08/2010		Current
274	Sharon Salavaria	Design Engineer	Roberto Tamburello	56920285	18/08/2010		Current
275	John Wood	Marketing Specialist	Debra Berger	222969461	7/09/2010		Current
276	Mary Dempsey	Marketing Assistant	Debra Berger	52541318	14/09/2010		Current
277	Stephen Jiang	Vice President of Sales	Patti Fernandez	502097814	13/10/2012		Current

همان‌طور که در شکل بالا نشان داده شده است، Stephen Jiang نقش جدید خود را به عنوان معاون فروش در تاریخ ۲۰۱۲/۱۰/۱۳ شروع کرده است و کار خود را به عنوان مدیر فروش آمریکای شمالی در تاریخ ۲۰۱۲/۱۰/۱۲ به پایان رسانده است.

بررسی می‌کنیم SCD 2 در مورد مدل‌سازی داده‌ها در Power BI به چه صورت است. سؤال اول این است: آیا می‌توانیم بدون داشتن انبار داده، SCD 2 را مستقیماً در Power BI Desktop پیاده‌سازی کنیم؟ برای پاسخ به این سؤال، باید به یاد داشته باشیم که هنگام ساخت یک مدل داده در Power BI، یک لایه معنایی (semantic layer) ایجاد می‌کنیم. در مورد اجزای مختلف یک BI solution، از جمله لایه معنایی، به مقالات دیگر مراجعه کنید. اما اینجا تأکید می‌کنم، لایه معنایی، طبق تعریف، نمایی از داده‌های منبع (معمولاً یک انبار داده) است که برای اهداف گزارش‌گیری و تحلیلی، بهینه‌سازی شده است. لایه معنایی جایگزین انبار داده یا نسخه دیگری از انبار داده نخواهد بود. بنابراین پاسخ منفی است، ما نمی‌توانیم SCD 2 را در Power BI پیاده‌سازی کنیم. بنابراین ما یا به یک انبار داده و یا به یک سیستم تراکنش دارای مکانیزمی برای پشتیبانی از حفظ سابقه داده‌ها نیاز داریم، مانند temporal mechanism.

temporal mechanism ویژگی‌ای است که برخی از سیستم‌های مدیریت پایگاه داده، رابطه‌ای مانند SQL Server برای ارائه اطلاعات در مورد داده‌های نگهداری شده در یک جدول ارائه می‌دهند که در هر زمان، فقط داده‌های فعلی را نگهداری می‌کنند.

پس از بارگذاری داده‌ها در مدل داده در Power BI Desktop، تمام داده‌های فعلی و سابقه داده‌ها را در جداول dimension داریم. بنابراین، هنگام کار با SCD ها باید دقت داشته باشیم. به عنوان مثال، شکل زیر فروش نمایندگی فروش را برای کارکنان نشان می‌دهد:

EmployeeKey	Full Name	Manager	Sales
290	Amy Alberts	Stephen Jiang	\$732,078.4446
289	David Campbell	Roger Hamilton	\$3,729,945.3501
284	Garrett Vargas	Roger Hamilton	\$3,609,447.2163
291	Jae Pak	Amy Alberts	\$8,503,338.6472
283	Jillian Carson	Roger Hamilton	\$10,065,803.5429
288	José Saraiva	Roger Hamilton	\$5,926,418.3574
282	Linda Mitchell	Roger Hamilton	\$10,367,007.4286
296	Lynn Tsoflias	Syed Abbas	\$1,421,810.9252
281	Michael Blythe	Roger Hamilton	\$9,293,903.0055
286	Pamela Ansman-Wolfe	Roger Hamilton	\$3,325,102.5952
295	Rachel Valdez	Amy Alberts	\$1,790,640.2311
292	Ranjit Varkey Chudukatil	Amy Alberts	\$4,509,888.933
287	Shu Ito	Roger Hamilton	\$6,427,005.5556
272	Stephen Jiang	Brian Welcker	\$1,092,123.8562
294	Syed Abbas	Stephen Jiang	\$172,524.4515
293	Tete Mensa-Annan	Roger Hamilton	\$2,312,545.6905
285	Tsvi Reiter	Roger Hamilton	\$7,171,012.7514
<b>Total</b>			<b>\$80,450,596.9823</b>

در نگاه اول، اعداد، صحیح به نظر می‌رسند اما ممکن است مفهوم اشتباهی را منتقل کنند. بستگی به این دارد که کسب و کار انتظار دارد در گزارش چه موردی را مشاهده کند. به شکل ۴ نگاه کنید که تغییرات Stephen را نشان می‌دهد. Stephen زمانی که مدیر فروش آمریکای شمالی بود، حجم فروش زیادی داشت (EmployeeKey 272) اما پس از ارتقا خود، نشان می‌دهد که Stephen دیگر فروش نداشته است (EmployeeKey 277). این مورد به این دلیل است که وقتی جدول بالا را ایجاد کردیم، SCD را در نظر نگرفتیم. برای وضوح بیشتر، باید ستون Status را به جدول اضافه کنیم.

EmployeeKey	Full Name	Manager	Sales	Status
290	Amy Alberts	Stephen Jiang	\$732,078.4446	Current
289	David Campbell	Roger Hamilton	\$3,729,945.3501	Current
284	Garrett Vargas	Roger Hamilton	\$3,609,447.2163	Current
291	Jae Pak	Amy Alberts	\$8,503,338.6472	Current
283	Jillian Carson	Roger Hamilton	\$10,065,803.5429	Current
288	José Saraiva	Roger Hamilton	\$5,926,418.3574	Current
282	Linda Mitchell	Roger Hamilton	\$10,367,007.4286	Current
296	Lynn Tsoflias	Syed Abbas	\$1,421,810.9252	Current
281	Michael Blythe	Roger Hamilton	\$9,293,903.0055	Current
286	Pamela Ansman-Wolfe	Roger Hamilton	\$3,325,102.5952	Current
295	Rachel Valdez	Amy Alberts	\$1,790,640.2311	Current
292	Ranjit Varkey Chudukatil	Amy Alberts	\$4,509,888.933	Current
287	Shu Ito	Roger Hamilton	\$6,427,005.5556	Current
272	Stephen Jiang	Brian Welcker	\$1,092,123.8562	
294	Syed Abbas	Stephen Jiang	\$172,524.4515	Current
293	Tete Mensa-Annan	Roger Hamilton	\$2,312,545.6905	Current
285	Tsvi Reiter	Roger Hamilton	\$7,171,012.7514	Current
<b>Total</b>			<b>\$80,450,596.9823</b>	

اگر کسب و کار نیاز داشته باشد فقط ارزش‌های فروش را فقط برای کارکنانی نشان دهد که Status آن‌ها current است، در آن صورت، ما باید SCD را در معادله فاکتور بگیریم و مقادیر فروش Stephen را فیلتر کنیم. بسته به نیازهای کسب و کار، ممکن است لازم باشد ستون status را به عنوان فیلتر در بصری‌سازی‌ها اضافه کنیم؛ درحالی که در برخی موارد دیگر، ممکن است نیاز داشته باشیم معیارها را با افزودن ستون‌های Start Date و End Date و Status تغییر دهیم تا نتایج را فیلتر کنیم. وقتی از فیلترهای بصری برای کاهش فروش Stephen استفاده می‌کنیم، شکل زیر نتایج را نشان می‌دهد:

EmployeeKey	Full Name	Manager	Sales	Status
290	Amy Alberts	Stephen Jiang	\$732,078.4446	Current
289	David Campbell	Roger Hamilton	\$3,729,945.3501	Current
284	Garrett Vargas	Roger Hamilton	\$3,609,447.2163	Current
291	Jae Pak	Amy Alberts	\$8,503,338.6472	Current
283	Jillian Carson	Roger Hamilton	\$10,065,803.5429	Current
288	José Saraiva	Roger Hamilton	\$5,926,418.3574	Current
282	Linda Mitchell	Roger Hamilton	\$10,367,007.4286	Current
296	Lynn Tsoflias	Syed Abbas	\$1,421,810.9252	Current
281	Michael Blythe	Roger Hamilton	\$9,293,903.0055	Current
286	Pamela Ansman-Wolfe	Roger Hamilton	\$3,325,102.5952	Current
295	Rachel Valdez	Amy Alberts	\$1,790,640.2311	Current
292	Ranjit Varkey Chudukatil	Amy Alberts	\$4,509,888.933	Current
287	Shu Ito	Roger Hamilton	\$6,427,005.5556	Current
294	Syed Abbas	Stephen Jiang	\$172,524.4515	Current
293	Tete Mensa-Annan	Roger Hamilton	\$2,312,545.6905	Current
285	Tsvi Reiter	Roger Hamilton	\$7,171,012.7514	Current
<b>Total</b>			<b>\$79,358,473.1261</b>	

کار ب SCD ها همیشه به این سادگی نیست. گاهی اوقات، لازم است تغییراتی در مدل داده خود ایجاد کنیم. بنابراین، آیا همه توضیحات بالا به این معنی است که ما نمی‌توانیم هر نوع SCD را در Power BI پیاده‌سازی کنیم؟ پاسخ این است که بستگی به سناریو شما دارد. در برخی از سناریوها، می‌توانیم راه حلی شبیه به عملکرد 1 SCD را پیاده‌سازی کنیم، اما در اجرای عملکرد 2 SCD در Power BI شانس زیادی نداریم.