

پایگاه داده 2

مستند ETL آموزشگاه مجازی

استاد درس:

دکتر علیرضا بصیری

اعضای تیم:

محمدهادی حاجی‌حسینی، غزاله زهتاب، سجاد عباسی، آرش ماری‌اورپاد

پاییزو زمستان 1399

2	مقدمه
5	بعد ها
5	S_Dim_User: بعد
6	S_Dim_Course: بعد
9	S_Dim_CourseTopic: بعد
11	HR_Dim_Instructor: بعد
13	HR_Dim_Staff: بعد
15	HR_Dim_TicketCategory: بعد
16	فکت های مارت Course Education
16	C_Fact_CourseBuying: فکت تراکنشی
19	C_Fact_CourseBuying_Daily: فکت پرIODیک
21	C_Fact_CourseBuying_ACC: فکت تجمیعی
22	C_Fact_CourseDownloading: فکت تراکنشی
23	C_Fact_CourseDownloading_Daily: فکت پرIODیک
23	فکت های مارت Users Behavior
23	U_Fact_UserRating: فکت تراکنشی
24	U_Fact_PassedCourses: فکت تراکنشی
26	U_Fact_UserRate_Acc: فکت تجمیعی
28	U_Fact_Comments: فکت تراکنشی
29	U_Fact_CommetRating: فکت تراکنشی
30	U_Fact_InfluentialUsers_Acc: فکت
32	فکت های مارت Human Resources
32	HR_Fact_InstructorCourse_T: فکت تراکنشی
34	HR_Fact_InstructorCourse_ACC : فکت تجمیعی
34	HR_Fact_InstructorRate_Daily : فکت روزانه
35	HR_Fact_StaffPayment_T : فکت سالانه
37	HR_Fact_StaffPayment_Yearly : فکت سالانه
38	HR_Fact_Tickets_Daily : فکت روزانه

مقدمه

موضوع: آموزشگاه مجازی

پایگاه داده اصلی این انبار داده، پایگاه داده ای برای وبسایت فروش آنلاین آموزش است. به این صورت که کاربران وارد وبسایت میشوند و دوره ای (course) که میخواهند را تهیه میکنند. این وبسایت دارای افراد و role مختلفی از نوع کاربر (user)، مدرس (Instructor) و کارمند (Staff) هست. یوزر ها کورسی که میخواهند را تهیه میکنند. این کورس توسط مدرسی ارائه شده است. از طرفی کارکنان به سوالات و تیکت های مربوطه پاسخ می دهند یا وظایف دیگری مثل برنامه نویسی و ... دارند. این دیتابیس شامل ۲۳ تا جدول است.

انبار داده ای که برای این دیتابیس آماده شده است دارای ۷ بعد و ۳ مارت است که در کل دارای ۱۸ فکت است.

❖ شرح تک تک بعد ها و همینطور فکت ها در ادامه گزارش برای هر کدام ذکر شده است.

بعد ها:

1. S_Dim_User
2. S_Dim_Course
3. S_Dim_Date
4. C_Dim_CourseTopic
5. HR_Dim_Instructor
6. HR_Dim_Staff
7. HR_Dim_TicketCategory

مارت ها:

1. مارت Course_Education
برای فکت ها و گزارش هایی که کورس محور هستند که دارای فکت های زیر است:

- C_Fact_CourseBuying
- C_Fact_CourseBuying_Periodic
- C_Fact_Course_Buying_Acc
- C_Fact_CourseDownloading
- C_Fact_Course_Downloading_Periodic

2. مارت User Behavior

برای فکت ها و گزارش هایی که کاربر محور هستند. یعنی بررسی هایی که روی کاربران صورت میگیرد. دارای فکت های زیر است:

- U_Fact_UserRating
- U_Fact_PassedCourses
- U_Fact_UserRate_Acc
- U_Fact_Comments
- U_Fact_CommentRating
- U_Fact_InfluentialUsers_Acc

3. مارت Human Resources

برای فکت ها و گزارش هایی که با محوریت کارکنان هستند. یعنی مدرسان و کارمندان. این مارت دارای فکت های زیر است:

- HR_Fact_InstructorCourse_T
- HR_Fact_InstructorCourse_ACC
- HR_Fact_InstructorRate_Daily
- HR_Fact_StaffPayment_T
- HR_Fact_StaffPayment_Yearly
- HR_Fact_Tickets_Daily

بعد ها¹

بعد S_Dim_User:

دیاگرام:

[S_Dim_User]	
# user_id	Integer
o first_name	Variable characters (70)
o last_name	Variable characters (70)
o username	Variable characters (150)
o credit	Number
o phone_number	Variable characters (25)
o email_address	Variable characters (355)
o is_email_verified	Boolean
o description_verified	Variable characters (70)
o datetime_signed_up	Date & Time
o gender	Variable characters (10)
o date_of_birth	Date

شرح:

این بعد به دلیل کاربرد در مارت² های متفاوت یک بعد اشتراکی است.

این بعد اطلاعات مربوط به کاربران سایت را در انبار داده نگهداری می‌کند.

برای این که دفعه اول پر شود از پراسیجر³ First_Time_Fill_User_Dim استفاده می‌شود و در اپدیت های بعدی از پراسیجر Fill_User_Dim استفاده می‌شود.

فیلدها:

- user_id : کلید اصلی این بعد محسوب می‌شود، از نوع INT است . داده آن مستقیم از فیلدی با همین نام از جدول با نام User از پایگاه داده source منتقل می‌شود.
- first_name : نام هر کاربر در این فیلد ذخیره می‌شود، از نوع NVARCHAR(70) است. داده آن مستقیم از فیلدی با همین نام از جدول با نام User از پایگاه داده source منتقل می‌شود که در Staging area قرار گرفته است.
- last_name: نام خانوادگی هر کاربر در این فیلد ذخیره می‌شود، از نوع NVARCHAR(70) است. داده آن مستقیم از فیلدی با همین نام از جدول با نام User از پایگاه داده source منتقل می‌شود که در Staging area قرار گرفته است.

¹ Dimensions

² Mart

³ Procedure

- **username**: این فیلد **username** کاربر که در سایت ثبت نام کرده است را ذخیره می‌کند، از نوع **NVARCHAR(150)** است. داده آن مستقیماً از فیلدی با همین نام از جدول با نام **User** از پایگاه داده **source** منتقل می‌شود که در **Staging area** قرار گرفته است.
- **credit**: اعتباری است که سایت برای کاربران در نظر می‌گیرد و از فیدبک های انبار داده به سیستم محسوب می‌شود. از نوع **number** است و از فیلدی با همین اسم از جدول **User** پایگاه داده **source** منتقل می‌شود که در **Staging area** قرار گرفته است.
- **phone_number**: تلفن هر کاربر در این فیلد ذخیره می‌شود، از نوع **VARCHAR(25)** است. داده آن مستقیماً از فیلدی با همین نام از جدول با نام **User** از پایگاه داده **source** منتقل می‌شود که در **Staging area** قرار گرفته است.
- **email_address**: ایمیل هر کاربر در این فیلد ذخیره می‌شود، از نوع **NVARCHAR(355)** است. داده آن مستقیماً از فیلدی با همین نام از جدول با نام **User** از پایگاه داده **source** منتقل می‌شود که در **Staging area** قرار گرفته است.
- **is_email_verified**: این فیلد تایید شدن یا عدم تایید شدن ایمیل را توسط کاربر مشخص می‌کند از نوع **BIT** است. داده آن مستقیماً از فیلدی با همین نام از جدول با نام **User** از پایگاه داده **source** منتقل می‌شود که در **Staging area** قرار گرفته است.
- **description_verified**: این فیلد توضیح فیلد قبلی است. در واقع نشان می‌دهد که فیلد **is_email_verified** اگر یک باشد نمایانگر این است که ایمیل تایید شده و اگر صفر باشد تایید نشده است. و با عبارات "Email is verified" و "Email is not verified" در پراسیجر مربوطه پر می‌شود.
- **datetime_signed_up**: تاریخی که کاربر در سایت ثبت نام کرده است در این فیلد ذخیره می‌شود، از نوع **DATETIME** است. داده آن مستقیماً از فیلدی با همین نام از جدول با نام **User** از پایگاه داده **source** منتقل می‌شود که در **Staging area** قرار گرفته است.
- **gender**: جنسیت کاربر در این فیلد ذخیره می‌شود، از نوع **NVARCHAR(10)** است. داده آن مستقیماً از فیلدی با همین نام از جدول با نام **User** از پایگاه داده **source** منتقل می‌شود که در **Staging area** قرار گرفته است. این فیلد تنها می‌تواند دو مقدار **male** و **female** را بپذیرد.
- **date_of_birth**: تاریخ تولد وارد شده توسط کاربر در این فیلد ذخیره می‌شود، از نوع **DATE** است. داده آن مستقیماً از فیلدی با همین نام از جدول با نام **User** از پایگاه داده **source** منتقل می‌شود که در **Staging area** قرار گرفته است.

بعد S_Dim_Course:

دیگرام:

S_Dim_Course			
course_key	<pi>	Integer	<M>
course_id		Integer	
title		Variable characters (400)	
description		Variable characters (1024)	
price		Money	
course_type		Variable characters (255)	
course_level		Variable characters (255)	
pre_req		Variable characters (1024)	
language		Variable characters (1024)	
file_size		Integer	
how_to_download		Variable characters (1024)	
contact_way		Variable characters (1024)	
completion_percentage		Decimal	
length_in_calendar_time		Variable characters (1024)	
start_date		Date & Time	
datetime_content_update		Date & Time	
datetime_created		Date & Time	
price_starting_date		Date & Time	
price_end_date		Date & Time	
current_flag		Boolean	

شرح:

این بعد مسئول نگهداری اطلاعات مربوط به دوره‌ها (کورس‌ها) در انبار داده می‌باشد و فیلدهای آن بر اساس فیلدهای جدول OnlineCourse از سورس طراحی شده است. به دلیل استفاده‌ی چندین مارت از این بعد، یک بُعد اشتراکی به حساب می‌آید و به همین خاطر نام آن با حرف S (اول کلمه‌ی Shared) آغاز می‌شود. همچنین لازم به ذکر است که این بُعد SCD نوع دوم است (بر اساس فیلد price) و برای اولین بار با استفاده از پروسیجر S_First_Time_Fill_Course_DIm و برای دفعات بعدی با اجرای پروسیجر S_Fill_Course_Dim آپدیت و پر می‌شود.

فیلدها:

- **course_key** : کلید اصلی این بعد به حساب می‌آید و از نوع INT است. در واقع این فیلد یک Surrogate Key است و بُعد مذکور نیز همان طور که بیان شد یک SCD نوع ۲ است. مقادیر این فیلد به صورت اتوماتیک و افزایشی هنگام وارد شدن رکورد جدید به بُعد ساخته می‌شوند.
- **course_id** : از نوع INT است و مشخص کننده‌ی شناسه‌ی دوره‌ها (کورس‌ها) می‌باشد و محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area (پس از انتقال از سورس) به دست می‌آید.
- **title** : این فیلد از نوع NVARCHAR می‌باشد و مشخص کننده‌ی عنوان دوره است. محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **description** : حاوی توضیحاتی در رابطه با کورس و از نوع NVARCHAR می‌باشد. محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **price** : این فیلد معرف قیمت (هزینه) دوره می‌باشد و از نوع Money است. مقدار آن مستقیماً از جدول OnlineCourse از staging_area به دست می‌آید و با توجه به آپدیت شدن آن، بُعد مورد بحث SCD نوع دو است.

- **course_type**: این فیلد بیان کننده‌ی نوع کورس (دوره) می‌باشد و از نوع NVARCHAR است. در واقع مدرس دوره یا ادمین سایت می‌توانند با استفاده از این فیلد، توضیحاتی در مورد نوع دوره ارائه دهند. محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **course_level**: همانند course_type از نوع NVARCHAR است و حاوی توضیحاتی در مورد سطح دوره می‌باشد و محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **pre_req**: این فیلد برای ذکر پیش‌نیازهای دوره استفاده می‌شود و محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود. در ضمن از نوع NVARCHAR نیز می‌باشد.
- **language**: فیلدی از نوع NVARCHAR که مشخص کننده‌ی زبان دوره می‌باشد. (مثلا انگلیسی یا فارسی یا ...) و محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **files_size**: از نوع INT بوده و بیان‌کننده‌ی سائز (حجم) کلی فایل‌های دوره می‌باشد (مثلا با واحد MB) و محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **How_to_download**: فیلدی از نوع NVARCHAR و حاوی توضیحاتی در رابطه با انواع روش‌های بارگیری (دانلود) محتوا و فایل‌های مرتبط با دوره است. محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **contact_way**: فیلدی از نوع NVARCHAR و حاوی توضیحاتی در رابطه با انواع روش‌های برقراری ارتباط و تماس به مدرسان و مسئولین دوره است. محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **completion_percentage**: این فیلد از نوع Decimal بوده و مشخص کننده‌ی درصد تکمیل محتوای کورس می‌باشد. (ممکن است یک دوره در ابتدای قرارگیری در سایت به صورت تکمیل نشده باشد). محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **length_in_calendar_time**
- **start_date**: فیلدی از نوع DateTime و مشخص کننده‌ی تاریخ شروع دوره است. محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **datetime_content_update**: از نوع DateTime است و بیان کننده‌ی تاریخ آخرین آپدیت (به روزرسانی) محتوای دوره است. محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **datetime_created**: این فیلد حاوی تاریخ ایجاد و قرارگیری این دوره در سایت می‌باشد و از نوع DateTime است. محتوای آن از فیلدی با همین نام از جدول OnlineCourse از staging_area مقدار دهی می‌شود.
- **price_start_date**: این فیلد که از نوع DateTime است برای نگهداری تاریخ شروع تعیین یک قیمت برای یک کورس استفاده می‌شود. (در واقع به دلیل آنکه این بُعد یک SCD نوع دوم است، این فیلد استفاده می‌شود). محتوای آن در زمانی که یک کورس جدید تعریف می‌شود و یا قیمت یه کورس از پیش تعریف شده تغییر می‌یابد، مطابق زمان اجرای پروسیجر تنظیم می‌شود.
- **price_end_date**: همچون فیلد قبلی وجود این فیلد نیز به توجه به SCD نوع ۲ بودن بُعد مذکور بوده و از نوع DateTime است. این فیلد در واقع حاوی تاریخ پایان یک قیمت‌گذاری خاص برای یک دوره است. مقدار آن برای آخرین قیمت‌گذاری دوره‌ها NULL است. اما به محض قیمت‌گذاری جدید، مقدار آن به مقدار روز اجرای پروسیجر تغییر می‌یابد.
- **current_flag**: این فیلد از نوع INT بوده و تنها مقادیر 0 و 1 می‌گیرد. وجود آن به خاطر SCD نوع دو بودن بُعد مورد بحث و ۱ بودن آن به معنای آن است که رکورد مذکور، حاوی آخرین قیمت‌گذاری برای یک دوره است.

بعد S_Dim_CourseTopic:

دیاگرام:

C_Dim_CourseTopic		
<u>subtopic_id</u>	<u><pi></u> Integer	<u><M></u>
topic_id	Integer	
course_id	Integer	
topic_title	Variable characters (1024)	
topic_priority	Integer	
subtopic_title	Variable characters (1024)	
subtopic_priority	Integer	
course_title	Variable characters (1024)	

شرح:

از این بُعد برای نگهداری ارتباط میان course و topic و subtopic استفاده می‌شود. در واقع در طراحی سورس هر subtopic به وسیله‌ی کلید خارجی به یک topic متصل است و هر topic خود به وسیله‌ی کلید خارجی به یک course منتسب می‌شود. لازم به ذکر است که در این بعد میان course و topic و subtopic یک رابطه‌ی سلسله مراتبی وجود دارد. این بعد SCD نوع 1 با توجه به فیلد subtopic_priority می‌باشد.

فیلدها:

- subtopic_id: این فیلد کلید اصلی بعد مذکور به حساب می‌آید و از نوع INT است. محتوای آن شامل شناسه‌ی هر subtopic و با توجه به فیلدی با همین نام از جدول Subtopic از staging_area (پس از تکمیل شدن با توجه به سورس) پر می‌شود.
- topic_id: مشخص‌کننده‌ی شناسه‌ی هر تاپیک و دارای رابطه‌ی سلسله مراتبی با subtopic می‌باشد. نوع این فیلد INT بوده و محتوای آن با توجه به فیلدی با همین نام از جدول Subtopic از staging_area پر می‌شود.
- course_id: مشخص‌کننده‌ی شناسه‌ی هر دوره و از نوع INT می‌باشد. این فیلد به دو فیلد قبلی دارای رابطه‌ی سلسله مراتبی است و محتوای آن با توجه به فیلدی با همین نام از جدول Subtopic از staging_area پر می‌شود.
- topic_title: بیان‌کننده‌ی عنوان یک تاپیک و از نوع NVARCHAR می‌باشد. مقدار آن با توجه به نتیجه‌ی join جداول Subtopic و Topic (بر روی فیلد topic_id) از staging_area به دست می‌آید.
- topic_priority: فیلدی از نوع INT و مشخص‌کننده‌ی شماره‌ی تاپیک در یک دوره می‌باشد. مقدار آن با توجه به نتیجه‌ی join جداول Subtopic و Topic (بر روی فیلد topic_id) از staging_area و از فیلدی با نام مشابه به دست می‌آید.
- subtopic_title: این فیلد که از نوع NVARCHAR می‌باشد بیان‌کننده‌ی عنوان subtopic بوده و محتوای آن با توجه به فیلدی با همین نام از جدول Subtopic از staging_area پر می‌شود.
- subtopic_priority: فیلدی از نوع INT و مشخص‌کننده‌ی شماره‌ی subtopic در یک تاپیک می‌باشد. در ضمن محتوای آن با توجه به فیلدی با همین نام از جدول Subtopic از staging_area پر می‌شود.

- course_title : این فیلد از نوع NVARCHAR بوده و شامل توضیحاتی در مورد دوره می‌باشد. مقدار آن از نتیجه‌ی join جدول OnlineCourse با جدول حاصل از join جداول Topic و SubTopic و از فیلدی با همین نام به دست می‌آید.

بعد HR_Dim_Instructor:

دیاگرام:

HR_Dim_Instructor		
#	inst_id	Integer
o	first_name	Variable characters (70)
o	last_name	Variable characters (70)
o	email_address	Variable characters (355)
o	is_email_verified	Boolean
o	short_description	Variable characters (70)
o	description	Variable characters (500)
o	datetime_signed_up	Date & Time
o	original_phone_num	Variable characters (25)
o	effective_date	Date & Time
o	current_phone_num	Variable characters (25)

شرح:

این بعد حاوی اطلاعات مدرس ها می باشد و کاربرد اصلی آن در گزارش گیری های پیرامون وضعیت مدرس ها است. از همین رو این بعد در مارت منابع انسانی قرار گرفته است زیرا در طراحی انجام شده مدرس ها بخشی از منابع انسانی سازمان ما در نظر گرفته شده اند.

این بعد یک SCD نوع سوم است و احتمال تغییر شماره تماس مدرس در آن در نظر گرفته شده است؛ از همین رو شماره تماس قبلی مدرس نیز در آن ذخیره می شود. در نتیجه فیلد های مورد نیاز برای پیاده سازی SCD نوع سوم نیز برای این بعد در نظر گرفته شده اند که در ادامه به آن ها اشاره خواهد شد.

دو پراسیجر برای پر کردن این بعد در نظر گرفته شده است. از پراسیجر First_Time_Fill_HR_Dim_Instructor برای پر کردن این بعد برای اولین بار استفاده می شود. همچنین از پراسیجر Fill_HR_Dim_Instructor_SCD3 برای پر کردن این بعد در مراحل بعدی استفاده می شود.

فیلدها:

- inst_id: کلید این بعد است و از نوع integer در نظر گرفته شده است. این فیلد از روی inst_id داخل جدول Instructor سمت Staging_area مقداری می شود.
- First_name: نام مدرس در این فیلد قرار می گیرد و از نوع NVARCHAR(70) در نظر گرفته شده است. این فیلد از روی فیلدی با نام مشابه از داخل جدول Instructor سمت Staging_area مقداری می شود.
- Last_name: نام خانوادگی مدرس در این فیلد قرار می گیرد و از نوع NVARCHAR(70) در نظر گرفته شده است. این فیلد از روی فیلدی با نام مشابه از داخل جدول Instructor سمت Staging_area مقداری می شود.

- Email_address: آدرس پست الکترونیکی مدرس در این فیلد قرار می گیرد و از نوع NVARCHAR(355) در نظر گرفته شده است. این فیلد از روی فیلدی با نام مشابه از داخل جدول Instructor سمت Staging_area مقداردهی می شود.
- Is_email_verified: این فیلد وضعیت آدرس پست الکترونیکی وارد شده توسط کاربر را نگهداری می کند و از نوع bit است. اگر آدرس توسط کاربر تایید شده باشد این فیلد 1 و در غیر این صورت مقدار 0 دارد. این فیلد از روی فیلدی با نام مشابه از داخل جدول Instructor سمت Staging_area مقداردهی می شود.
- Short_description: به سبب اینکه فیلد قبلی از نوع 0 و 1 بود، این فیلد برای توضیح فیلد قبلی در این بعد قرار گرفته است و از نوع NVARCHAR(70) است. این فیلد از روی فیلدی با نام مشابه از داخل جدول Instructor سمت Staging_area مقداردهی می شود.
- Description: این فیلد حاوی کلمات کلیدی و نکاتی است که خود مدرس ترجیح می دهد درباره خود ذکر کند و از نوع NVARCHAR(500) است. این فیلد از روی فیلدی با نام مشابه از داخل جدول Instructor سمت Staging_area مقداردهی می شود.
- Datetime_signed_up: این زمان دقیق عضویت مدرس در مجموعه را شامل زمان و تاریخ نگهداری می کند و از نوع DATETIME است. این فیلد از روی فیلدی با نام مشابه از داخل جدول Instructor سمت Staging_area مقداردهی می شود.
- Original_phone_num: این فیلد برای نگه داشتن شماره تلفن قبلی مدرس است و از نوع NVARCHAR(25) است. در زمان ایجاد رکورد برای یک مدرس جدید این فیلد دارای مقدار NULL است. هرگاه شماره تلفن مدرس سمت source تغییر کند، پراسیجری که برای پر کردن این بعد طراحی شده است این مورد را چک می کند و مقدار قبلی شماره تماس مدرس را در این فیلد ذخیره می کند.
- Effective_time: این فیلد برای نگه داشتن زمانی است که شماره تماس مدرس تغییر کرده است؛ بنابراین تا زمانی که چنین تغییری رخ نداده باشد این فیلد مقدار NULL دارد. این فیلد از نوع DATETIME است. هرگاه شماره تلفن مدرس سمت source تغییر کند، پراسیجری که برای پر کردن این بعد طراحی شده است این مورد را چک می کند و زمان تغییر شماره تماس مدرس را در این فیلد ذخیره می کند.
- Current_phone_num: در این فیلد شماره تماس فعلی مدرس ذخیره می شود و از نوع NVARCHAR(25) است. هرگاه شماره تلفن مدرس سمت source تغییر کند، پراسیجری که برای پر کردن این بعد طراحی شده است این مورد را چک می کند و این فیلد را به شماره تماس فعلی مدرس آپدیت می کند.

بعد HR_Dim_Staff:

دیاگرام:

HR_Dim_Staff			
<u>staff_id</u>	<pi>	Integer	<M>
first_name		Variable characters (70)	
last_name		Variable characters (70)	
date_of_birth		Date	
email_address		Variable characters (355)	
is_email_verified		Boolean	
description_is_email_verified		Variable characters (200)	
phone_number		Variable characters (25)	
datetime_joined		Date & Time	
gender		Variable characters (10)	
short_description		Variable characters (500)	
username		Variable characters (150)	
job_title		Variable characters (400)	
job_description		Variable characters (1000)	

شرح:

این بعد اطلاعات کارکنان را نگه داری می‌کند. از آنجا که در طراحی، کارکنان جزو منابع انسانی سازمان در نظر گرفته شده اند این بعد در مارت منابع انسانی قرار گرفته است.

دو پراسیجر برای پر کردن این بعد در نظر گرفته شده است. پراسیجر HR_First_Time_Fill_Dim_Staff که برای پر کردن اولیه آن استفاده می‌شود و پراسیجر HR_Fill_Dim_Staff که برای پر کردن دفعات بعدی استفاده می‌شود. این بعد از روی join کردن دو جدول Staff و Job در Staging_area به دست می‌آید که در ادامه تک تک فیلد های آن توضیح داده می‌شود.

فیلد ها:

staff_id: کلید این بعد است. جنس آن Integer است. این فیلد از روی staff_id داخل جدول Staff در Staging_area پر میشود.

first_name: نام کارمند مورد نظر در این فیلد قرار می‌گیرد. جنس آن NVARCHAR(70) است. این فیلد از روی فیلدی با همین نام از جدول Staff در Staging_area پر میشود.

Last_name: نام خانوادگی کارمند مورد نظر در این فیلد قرار می‌گیرد. جنس آن NVARCHAR(70) است. این فیلد از روی فیلدی با همین نام از جدول Staff در Staging_area پر میشود.

date_of_birth: این فیلد برای نگهداری تاریخ تولد کارمند مورد نظر است. جنس این فیلد از نوع Date است. این فیلد از روی فیلدی با همین نام از جدول Staff در Staging_area پر میشود.

email_address: این فیلد برای نگهداری ایمیل کارمند مورد نظر است. جنس این فیلد از نوع NVARCHAR(355) است. این فیلد از روی فیلدی با همین نام از جدول Staff در Staging_area پر میشود.

is_email_verified: این فیلد وضعیت آدرس پست الکترونیکی وارد شده توسط کاربر را نگهداری می‌کند و از نوع bit است. اگر آدرس توسط کاربر تایید شده باشد این فیلد 1 و در غیر این صورت مقدار 0 دارد. این فیلد از روی فیلدی با نام مشابه از داخل جدول Staff سمت Staging_area مقداردهی می‌شود.

Description_is_email_verified: جنس این فیلد NVARCHAR(200) است. این فیلد برای توضیح فیلد بالا استفاده میشود. به این صورت که در حالت 1 بودن مقدار بیت فیلد قبلی یعنی فیلد is_email_verified و همچنین مقدار صفر بودن آن، در اینجا توضیح معنی آن نوشته میشود. در این حالت اگر مقدار فیلد قبلی 1 باید این مقدار برابر Email is Verified است. و اگر مقدار 0 داشته باشد آنگاه مقدار این فیلد برابر Email is not Verified میشود. این فیلد بر اساس مقدار فیلد is_email_verified در جدول Staff سمت Staging_area مقداردهی می‌شود. که با هر یک از شرط‌های ذکر شده در بالا یکی از دو مقدار Email is Verified و Email is not Verified را به خود می‌گیرد.

phone_number: این فیلد برای نگهداری شماره تماس کارمند مورد نظر استفاده میشود. جنس این فیلد از نوع NVARCHAR(25) است. این فیلد از روی فیلدی با همین نام از داخل جدول Staff سمت Staging_area مقداردهی می‌شود.

datetime_joined: این فیلد برای نگهداری تاریخ اضافه شدن کارمند به شرکت است. جنس آن DateTime است. این فیلد از روی فیلدی با نام مشابه از داخل جدول Staff سمت Staging_area مقداردهی می‌شود.

gender: این فیلد برای نگهداری جنسیت کارمند مورد نظر است. جنس آن NVARCHAR(10) است. این فیلد از روی فیلدی با نام مشابه از داخل جدول Staff سمت Staging_area مقداردهی می‌شود.

username: این فیلد username کارمند مورد نظر را ذخیره می‌کند. جنس آن از نوع NVARCHAR(150) است. داده آن مستقیم از فیلدی با همین نام از جدولی با نام Staff از پایگاه داده source منتقل می‌شود که در Staging area قرار گرفته است.

job_title: این فیلد برای نگهداری عنوان شغلی کارمند مورد نظر در اینجا ذخیره میشود. جنس این فیلد NVARCHAR(400) است. این فیلد از join دو جدول Staff و Job با اشتراک فیلد job_id صورت می‌گیرد که در Staging_area قرار دارد. از این join فیلدی با همین نام مقدار این فیلد را مشخص می‌کند.

job_description: این فیلد برای نگهداری توضیح شغل کارمند مورد نظر در اینجا ذخیره میشود. جنس این فیلد (NVARCHAR(1000) است. این فیلد از join دو جدول Staff و Job با اشتراک فیلد job_id صورت می‌گیرد که در Staging_area قرار دارد. از این join فیلدی با همین نام مقدار این فیلد را مشخص می‌کند.

بعد HR_Dim_TicketCategory:

دیاگرام:

HR_Dim_TicketCategory		
#	ticket_category_id	Integer
o	ticket_cat_title	Variable characters (250)
o	[description]	Variable characters (500)

شرح:

این بعد حاوی اطلاعات دسته بندی هایی است که برای انواع تیکت ها وجود دارد. از این بعد به منظور گزارش گیری های مربوط به عملکرد تیم پشتیبانی استفاده می شود.

دو پراسیجر برای پر کردن این بعد در نظر گرفته شده است. از پراسیجر HR_First_Time_Fill_Dim_TicketCategory برای پر کردن این بعد برای اولین بار استفاده می شود. همچنین از پراسیجر HR_Fill_Dim_TicketCategory برای پر کردن این بعد در مراحل بعدی استفاده می شود. اطلاعات این بعد به صورت مستقیم از جدولی در سمت Staging_area با نام مشابه این بعد تامین می شود.

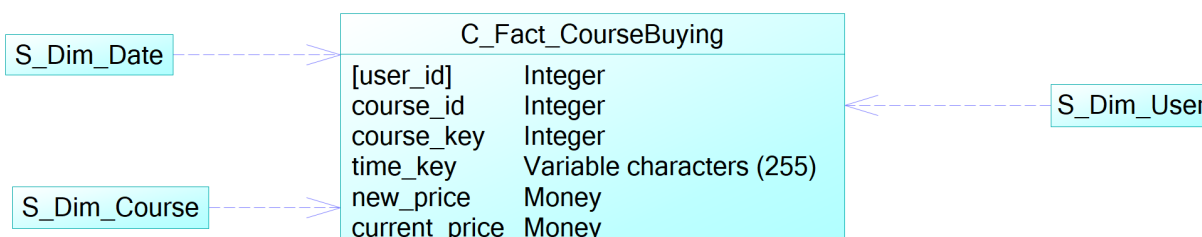
فیلدها:

- Ticket_category_id: این فیلد که کلید این بعد است حاوی شناسه هر دسته از تیکت هاست و از نوع integer است. از روی فیلدی با نام مشابه از جدول TicketCategory سمت Staging_area پر می شود.
- Ticket_cat_title: عنوان هر دسته از تیکت ها در این فیلد قرار می گیرد و از نوع (NVARCHAR(250) است. از روی فیلدی با نام title از جدول TicketCategory سمت Staging_area پر می شود.
- Description: این فیلد حاوی شرح جامع تری از هر دسته از تیکت هاست که برای گزارش گیری ها مفید است. از نوع (NVARCHAR(500) است و از روی فیلدی با نام مشابه از جدول TicketCategory سمت Staging_area پر می شود.

فکت های مارت Course Education

فکت تراکنشی C_Fact_CourseBuying:

دیاگرام:



شرح:

این فکت یک فکت تراکنشی است. هدف آن ذخیره این است که چه کسی (کدام user) کدام کورس (course) را در چه تاریخی خریده است. در کنار این موارد، قیمت اصلی دوره و همینطور قیمت نهایی (پس از اعمال تخفیف) نیز ذخیره می‌شود.

فیلدها:

- user_id :

جنس این فیلد، INTEGER است.

این فیلد برای این است که بدانیم کدام یوزر این کورس را خریداری کرده است. این فیلد عملاً به بعد S_Dim_User متصل می‌شود و نشان می‌دهد که کدام یک از یوزر های داخل آن بعد، این کورس را خریداری کرده اند. این فیلد از جدول UserOnlineCourse پر می‌شود که در Staging Area قرار دارد و عیناً همان id ای است که در آن وجود دارد. زیرا بعد S_Dim_User یک بعد ساده است و نیازی به Surrogate Key نداشته است.

- course_id :

جنس این فیلد، Integer است.

این فیلد نشان می‌دهد که یوزر، کدام کورس را خریداری کرده است. این فیلد عملاً course_id را نگهداری می‌کند. این فیلد از خود جدول UserOnlineCourse پر می‌شود که در Staging Area قرار دارد که عیناً همان id در اینجا قرار می‌گیرد که نشان می‌دهد کدام کورس است که اینجا یوزر خریداری کرده است. این فیلد id کورس را نگه می‌دارد، در حالی که بعد S_Dim_Course به صورت SCD نوع دو ساخته شده؛ پس عملاً برای اتصال به آن، نیاز به فیلد دیگری داریم که فیلد course_key است که در ادامه بررسی می‌کنیم.

این فیلد از جدول UserOnlineCourse پر می‌شود که در Staging Area قرار دارد.

- **course_key : course_key**

جنس این فیلد INTEGER است.

این فیلد نشان می‌دهد که دقیقاً کدام کورس در بعد S_Dim_Course خرید شده است. با اینکه در فیلد قبلی، id کورس را نگه داری می‌کنیم اما چون بعد S_Dim_Course به صورت SCD نوع 2 است پس در آن Surrogate Key وجود دارد که با این کلید دقیقاً به همان رکورد دسترسی پیدا می‌کنیم. برای پیدا کردن این کلید برای کورس مربوطه، به دنبال کورسی در بعد S_Dim_Course در انبار داده می‌گردیم که همین course_id (که در فیلد قبلی آن را داریم) را برای کورس داشته باشد. اما چون بعد S_Dim_Course، از نوع SCD نوع 2 است و از هر course_id ممکن است چند تا داشته باشیم، پس course_key رکوردی را انتخاب می‌کنیم که current_flag آن مساوی با 1 باشد که این موضوع نشان می‌دهد که آخرین رکورد معتبر آن کورس در این بعد، این course_key است. و به این صورت course_key معتبر را برای course_id مورد نظر پیدا می‌کنیم.

- **time_key : این فیلد از نوع NVARCHAR(255) است.**

این فیلد مشخص می‌کند که این خرید که در این فکت ثبت شده است، دقیقاً در چه تاریخی بوده است. این فیلد عملاً به بعد S_Dim_Date اشاره می‌کند.

نحوه ساخت این فیلد به این صورت است که در ادامه گفته می‌شود. زمانی که می‌خواهیم کورس های ثبت نام شده توسط یوزر های مختلف را در این فکت اضافه کنیم، برای هر کدام تاریخی به نام datetime_user_enrolled وجود دارد که نشان می‌دهد هر در چه تاریخی یوزر در این کورس ثبت نام کرده است. حال ما معادل این تاریخ را در بعد تاریخ به دست می‌آوریم و کلید آن (TimeKey) را در اینجا نگهداری می‌کنیم. این کار را با استفاده از تابعی انجام می‌دهیم که خودمان نوشته ایم. این تابع DataWarehouse.dbo.S_Make_TimeKey نام دارد که از ما تاریخی را با جنس Date می‌گیرد و TimeKey معادل آن را از بعد S_Dim_Date بر می‌گرداند که در این فیلد ذخیره می‌کنیم.

- **Current_price : Current_price**


جنس این فیلد Money است.

این فیلد مشخص می‌کند که در لحظه که یوزر این کورس را خریداری کرده است، قیمت اصلی آن چه قدر بوده است. وقتی کسی کورسی را خریداری می‌کند، در آن لحظه آن کورس قیمتی را داراست. این قیمت در این فیلد ذخیره می‌شود. اما ممکن است یوزر با اعمال تخفیف، قیمت کمتری پرداخت کند که در این حالت فیلد بعدی رو ببینید. پس عملاً این فیلد، قیمت اصلی کورس است که هیچ تخفیفی روی آن قرار ندارد. این فیلد به صورت مستقیم از جدول Staging_area.dbo.UserOnlineCourse دریافت می‌شود.

- **new_price : new_price**

جنس این فیلد Money است.

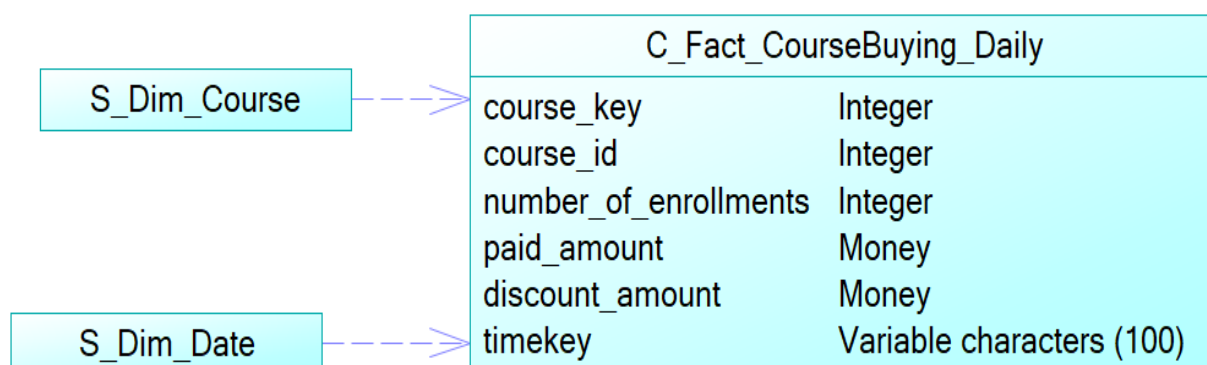
وقتی کسی کورسی را خریداری می‌کند، جدای از اینکه آن کورس چه قیمتی را داشته است (در فیلد بالا ذخیره شده است) ممکن است که با تخفیف، قیمت کمتری رو پرداخت کند. قیمت جدید در این فیلد ثبت می‌شود. این فیلد به این صورت پر می‌شود که با join دو جدول Staging_area.dbo.UserOnlineCourse و



Staging_area.dbo.UserOnlineCourse به دست می‌آید. چراکه در جدول Staging_area.dbo.OffPrice برای هر خرید گفته شده است که از کدام تخفیف استفاده شده است. تخفیفی که در جدول OffPrice در Staging_area ذخیره شده است. پس اگر این دو جدول را join بزنیم می‌توانیم ببینیم که قیمت آن بعد از اعمال تخفیف چه قدر می‌شود. البته دقت کنید که اگر محصولی بدون تخفیف خریداری شده باشد، آنگاه برای قسمت تخفیف آن (که کلید آن به جدول OffPrice است) در جدول UserOnlineCourse در Staging_area در آن فیلد مقدار NULL قرار می‌گیرد.

فکت پریودیک C_Fact_CourseBuying_Daily:

دیآگرام:



شرح:

در این جدول فکت که از نوع Daily Snapshot می باشد، برای هر کورس به صورت روزانه موارد زیر محاسبه و ذخیره می شوند:

- تعداد enrollment ها (ثبت نام ها) برای آن کورس
- هزینه ی پرداخت شده برای آن کورس
- میزان تخفیفی که کاربران در فرایند ثبت نام آن کورس از آن بهره مند شده اند

این فکت از روی فکت تراکنشی C_Fact_CourseBuying به دست می آید و در مارت Course Education قرار دارد.

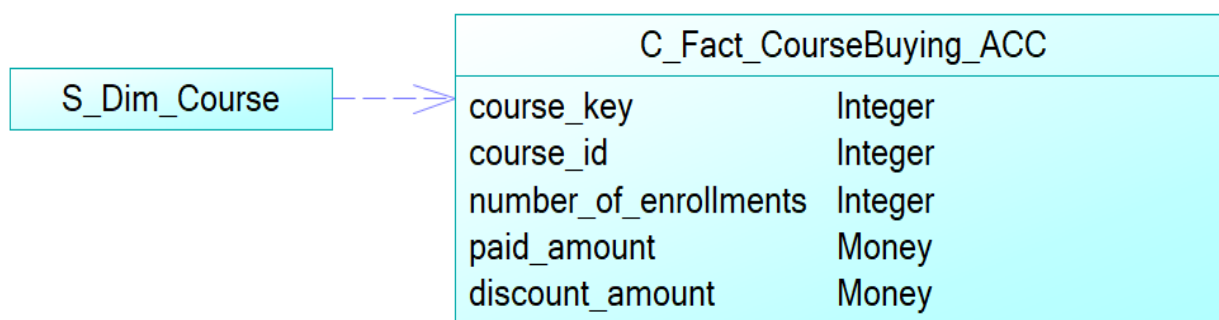
فیلدها:

- course_key: این فیلد از نوع INT می باشد و مقدار آن از فیلدی با همین نام از فکت تراکنشی C_Fact_CourseBuying به دست می آید که خود از بُعد S_Dim_Course پر شده است. در واقع این فیلد در بُعد C_Fact_CourseBuying یک surrogate key به حساب می آید.
- course_id: فیلدی از نوع INT و حاوی شناسه ی دوره ها که از محتوای آن از فیلدی با همین نام از فکت تراکنشی C_Fact_CourseBuying به دست می آید.
- number_of_enrollments: فیلدی از نوع INT و نشان دهنده ی تعداد ثبت نامی ها در یک کورس در یک روز خاص است. این فیلد در واقع یک measure است و از روی فکت تراکنشی C_Fact_CourseBuying محاسبه می شود.
- paid_amount: فیلدی از نوع Money و نشان دهنده ی جمع مبالغ پرداخت شده در یک روز برای یک کورس است. همانند فیلد قبلی این فیلد نیز یک measure است و با توجه به فکت تراکنشی C_Fact_CourseBuying محاسبه می شود.

- **discount_amount**: این فیلد نیز از نوع Money می باشد و جمع تخفیف هایی که کاربران برای خرید یک دوره در یک روز خاص از آن بهرمند شده اند را نشان می دهد. این فیلد نیز یک measure به حساب می آید و با توجه به محتوای C_Fact_CourseBuying محاسبه می شود.
- **time_key**: فیلدی از نوع NVARCHAR است که محتوای آن با توجه به دایمنشن S_Dim_Date مدیریت می شود. این فیلد تاریخ روزی را مشخص می کند که قرار است از یک کورس، گزارش های روزانه ای توضیح داده شده برای این فکت پرIODیک مشاهده شود.

فکت تجمیعی C_Fact_CourseBuying_ACC:

دیگرام:



شرح:

هر رکورد این جدول گزارشی است در این باب که برای یک کورس خاص:

- تا کنون چند نفر در این کورس ثبت‌نام کرده‌اند؟
- از نظر مالی این کورس چه میزان فروش داشته است؟
- ثبت‌نام کنندگان در این کورس تا کنون از چه میزان تخفیف بهره‌مند شده‌اند.

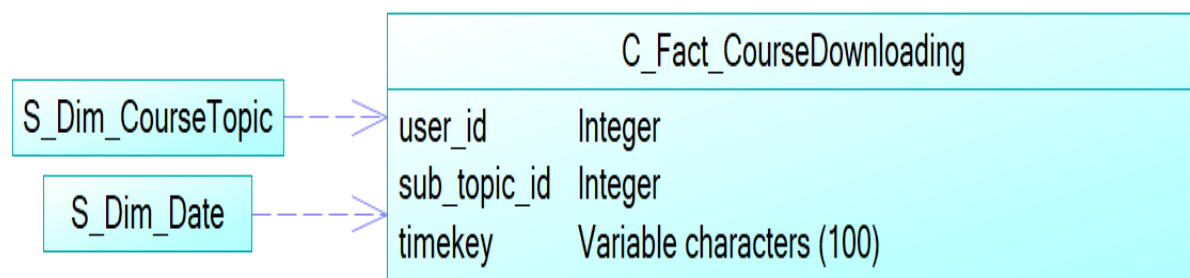
این فکت از نوع تجمیعی می‌باشد و اطلاعات آن از جدول فکت تراکنشی C_Fact_CourseBuying به دست می‌آید.

فیلدها:

- course_key: این فیلد از نوع INT می‌باشد و مقدار آن از فیلدی با همین نام از فکت تراکنشی C_Fact_CourseBuying به دست می‌آید که خود از بُعد S_Dim_Course پر شده است. در واقع این فیلد در بُعد C_Fact_CourseBuying یک surrogate key به حساب می‌آید.
- course_id: فیلدی از نوع INT و حاوی شناسه‌ی دوره‌ها که از محتوای آن از فیلدی با همین نام از فکت تراکنشی C_Fact_CourseBuying به دست می‌آید. این فیلد نشان‌دهنده‌ی شناسه‌ی یک دوره می‌باشد.
- number_of_enrollments: فیلدی از نوع INT و نشان‌دهنده‌ی تعداد کل ثبت‌نامی‌ها در یک کورس است. این فیلد در واقع یک measure است و از روی فکت تراکنشی C_Fact_CourseBuying محاسبه می‌شود.
- paid_amount: فیلدی از نوع Money و نشان‌دهنده‌ی جمع کل مبالغ پرداخت شده در برای یک کورس است. همانند فیلد قبلی این فیلد نیز یک measure است و با توجه به فکت تراکنشی C_Fact_CourseBuying محاسبه می‌شود.
- discount_amount: این فیلد نیز از نوع Money می‌باشد و جمع کل تخفیف‌هایی که کاربران برای خرید یک دوره از آن بهره‌مند شده‌اند را نشان می‌دهد. این فیلد نیز یک measure به حساب می‌آید و با توجه به محتوای C_Fact_CourseBuying محاسبه می‌شود.

فکت تراکنشی C_Fact_CourseDownloading:

دیاگرام:



شرح:

این فکت که یک فکت تراکنشی به حساب می‌آید جهت ثبت و بررسی درخواست‌های دانلود کاربران طراحی شده است.

در واقع هر رکورد این فکت نشان‌دهنده‌ی آن است که کاربری در تاریخی مشخص اقدام به دانلود محتوای یک subtopic مشخص از یک دوره کرده است.

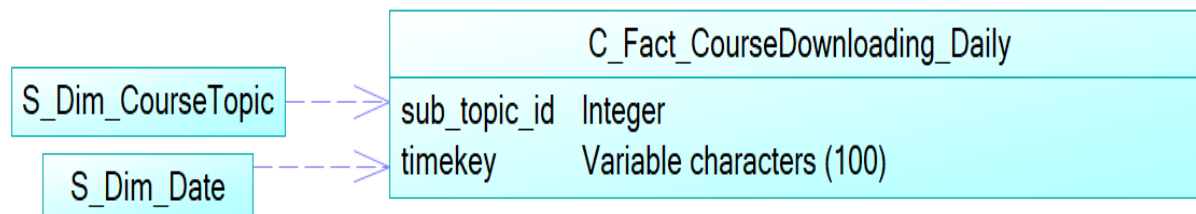
این فکت با توجه به جدول Downloaded از staging_area تکمیل می‌شود.

فیلدها:

- **user_id**: این فیلد از نوع INT می‌باشد و معرف شناسه‌ی کاربر بوده و از فیلدی با همین نام از جدول Downloaded از staging_area به دست می‌آید.
- **sub_topic_id**: این فیلد نیز از نوع INT می‌باشد و نشان‌دهنده‌ی شناسه‌ی یک subtopic است (که توسط کاربری دانلود شده است). محتوای این فیلد مستقیماً از فیلدی با همین نام و از جدول Downloaded از staging_area به دست می‌آید. مقدار ذخیره شده در این فیلد با مقادیر بعد S_Dim_CourseTopic نیز ارتباط دارد.
- **timekey**: این فیلد از نوع NVARCHAR بوده و نشان‌دهنده‌ی تاریخ اقدام یک کاربر به دانلود یک subtopic می‌باشد. این فیلد نیز از محتوای فیلدی با همین نام و از بُعد S_Dim_Date به دست می‌آید.

فکت پرودیک C_Fact_CourseDownloading_Daily:

دیاگرام:



شرح:

این فکت از نوع Daily Snapshot می‌باشد و با توجه به فکت تراکنشی C_Fact_CourseDownloading به دست می‌آید.

در این فکت می‌خواهیم گزارش کنیم که برای هر subtopic (و به تبع آن برای هر course) در هر روز، چه تعداد تلاش برای دانلود توسط کاربران صورت گرفته‌است.

این گزارش از جهت کنترل و بررسی منابع سرورها و طراحی زیرساخت حائز اهمیت می‌باشد.

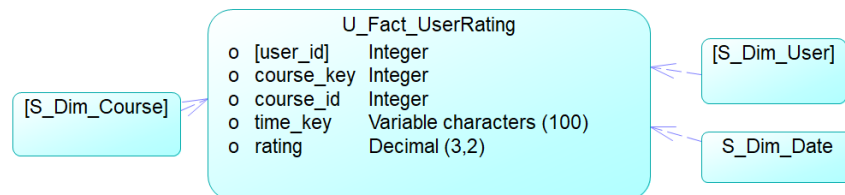
فیلدها:

- sub_topic_id: این فیلد نیز از نوع INT می‌باشد و نشان‌دهنده‌ی شناسه‌ی یک subtopic است (که توسط کاربری دانلود شده است). محتوای این فیلد مستقیماً از فیلدی با همین نام و از جدول Downloaded از staging_area به دست می‌آید. مقدار ذخیره شده در این فیلد با مقادیر بعد S_Dim_CourseTopic نیز ارتباط دارد.
- timekey: این فیلد از نوع NVARCHAR بوده و نشان‌دهنده‌ی تاریخ روزی است که می‌خواهیم بدانیم چه تعداد از کاربران اقدام به دانلود یک subtopic کرده‌اند. این فیلد نیز از محتوای فیلدی با همین نام و از بُعد S_Dim_Date به دست می‌آید.

فکت های مارت Users Behavior

فکت تراکنشی U_Fact_UserRating:

دیاگرام:



شرح:

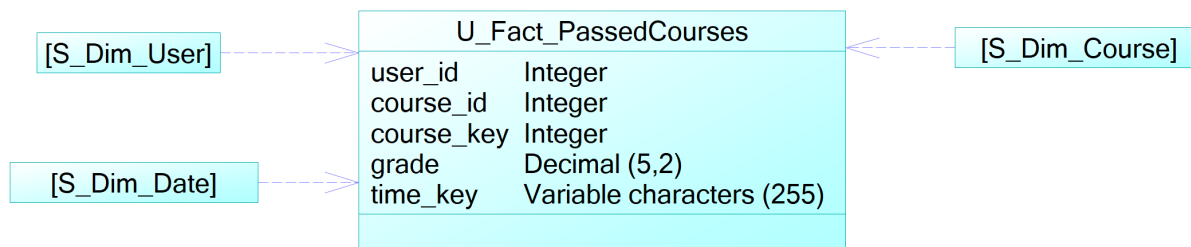
این فکت تراکنشی است. با این هدف طراحی شده است که در طول زمان rating که هر یوزر به هر course می‌دهد را نگهداری کند. لاگ تغییرات این فکت در U_UsersMart_log در انبار داده ثبت می‌شود.

فیلدها:

- user_id : از فیلدی با همین نام از جدول UserOnlineCourse که در staging_area است پر می‌شود.
- course_key : از فیلدی با همین نام از بعد S_Dim_Course وارد میشود از مقایسه ی course_id در این بعد و جدول UserOnlineCourse در staging_area حاصل می‌شود.
- course_id : از فیلدی با همین نام از جدول UserOnlineCourse که در staging_area است پر می‌شود.
- time_key: توسط function با اسم S_Make_TimeKey و ورودی datetime_of_rating از جدول UserOnlineCourse در staging_area پر می‌شود. خود S_Make_TimeKey در انبار داده است.
- rating: از فیلدی با همین نام از جدول UserOnlineCourse که در staging_area است پر می‌شود.

فکت تراکنشی U_Fact_PassedCourses:

دیاگرام:



شرح:

این فکت یک فکت تراکنشی است که هر رکورد آن نشان میدهد که یوزر مشخص شده چه کورسی را (هم با course_id و هم با course_key مربوطه) در چه تاریخی چه نمره ای برای آن آورده است. این فکت از جدول UserOnlineCourse در Staging_area ساخته می شود.

این فکت با استفاده از دو پراسیجر U_First_Time_Fill_PassedCourses_Fact_T و U_Full_PassedCourses_Fact_T پر میشوند. که اولی برای اجرا برای بار اول است و دومی برای اجرا برای دفعات بعدی.

فیلدها:

user_id: این فیلد برای نشان دادن یوزر مورد نظر است. جنس این فیلد Integer است. این فیلد از جدول UserOnlineCourse در Staging_area گرفته میشود که در آنجا نیز به همین نام است.

course_id: این فیلد برای مشخص کردن id کورسی هست که یوزر در آن نمره ای کسب کرده است. جنس این فیلد Integer است. این فیلد از جدول UserOnlineCourse در Staging_area گرفته میشود که در آنجا نیز به همین نام است.

course_key: این فیلد برای مشخص کردن کلید کورسی هست که یوزر در آن نمره ای کسب کرده است. جنس این فیلد Integer است. علت اینکه در این فکت هم course_id و هم course_key داریم این است که بعد S_Dim_Course به صورت SCD نوع 2 است. به این صورت که برای هر کورس در آن یک course_key به عنوان surrogate key وجود دارد. برای همین در این فکت جدای از course_id، یک course_key نیز ذخیره می کنیم که نشان دهد این رکورد در این فکت به کدام یک از رکورد های بعد course متصل است. برای به دست آوردن این course_key نیازه که به دنبال رکوردی در بعد S_Dim_Course بگردیم که course_id آن به همین صورت باشد و البته current_flag آن ۱ باشد.

grade: این فیلد برای نگهداری نمره یوزر است. جنس این فیلد Decimal(5,2) است. این فیلد نشان میدهد که یوزر مورد نظر در این درس چه نمره ای را به دست آورده است.

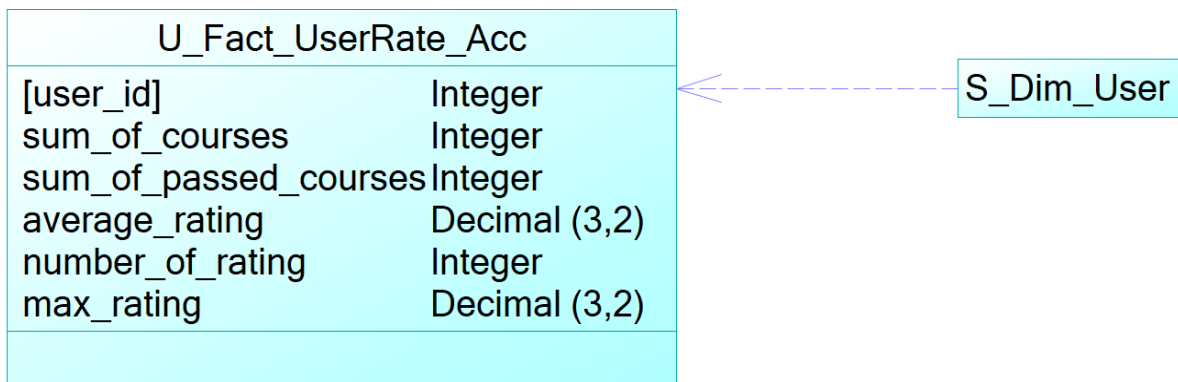
time_key:

این فیلد مشخص میکند که نمره ای که کاربر در این کورس به دست آورده است در چه تاریخی ثبت شده است. این فیلد از نوع NVARCHAR(255) است. این فیلد عملاً به بعد S_Dim_Date اشاره می کند. نحوه ساخت این فیلد به این صورت است که در ادامه گفته می شود. مقدار تاریخ اصلی آن از جدول UserOnlineCourse در Staging_area گرفته میشود که فیلدی به نام

datetime_of_grade برای هر رکورد در آن جا وجود دارد. حال ما معادل این تاریخ را در بعد تاریخ به دست می آوریم و کلید آن (TimeKey) را در اینجا نگهداری می کنیم. این کار را با استفاده از تابعی انجام می دهیم که خودمان نوشته ایم. این تابع DataWarehouse.dbo.S_Make_TimeKey نام دارد که از ما تاریخی را با جنس Date می گیرد و TimeKey معادل آن را از بعد S_Dim_Date بر می گرداند که در این فیلد ذخیره می کنیم. نتیجه این تبدیل را که کلیدی از بعد S_Dim_Date هست، در اینجا نگهداری می کنیم.

فکت تجمعی U_Fact_UserRate_Acc:

دیاگرام:



شرح:

این فکت یک فکت تجمعی است. این فکت نشان میدهد که هر یوزر از اول تا الان چه تعداد رای داده است و بالاترین رایی که داده چه قدر است. همچنین میتوان فهمید که میانگین رای هایی که تا کنون داده چه عددی است. تعداد کورس هایی که تا حالا شرکت کرده و تعداد کورس هایی که تا الان در آن ها نمره آورده است نیز در این فکت تجمعی نشان داده میشود. برای ساختن این فکت از جدول UserOnlineCourse در Staging_area و همینطور فکت تراکنشی U_Fact_UserRating که قبلا ساخته ایم در DataWarehouse استفاده می شود.

برای پر شدن این فکت از دو پروسیجر به نام های U_First_Time_Fill_UserRate_Fact_Acc و U_Fill_UserRate_Fact_Acc استفاده می شود. که اولی برای اجرای دفعه اول و بعدی برای اجرای دفعات بعدی است.

فیلدها:

user_id: این فیلد برای نشان دادن یوزر مورد نظر است. جنس این فیلد Integer است. این فیلد از فکت U_Fact_UserRating که یک فکت تراکنشی است که قبلا آن را ساخته ایم گرفته میشود که در DataWarehouse وجود دارد. این فیلد نشان میدهد که این اطلاعات در این رکورد مربوط به کدام یوزر است. این مقدار به بعد S_Dim_User اشاره میکند.

sum_of_courses: این فیلد برای نگهداری تعداد کورس هایی هست که کاربر از ابتدا تا الان در آن شرکت کرده است. جنس این فیلد Integer است. این تعداد با شمارش تعداد کورس هایی که کاربر ثبت نام کرده است از جدول UserOnlineCourse در Staging_area محاسبه می شود و در اینجا قرار میگیرد.

sum_of_passed_courses: این فیلد برای نگهداری تعداد کورس هایی هست که کاربر در آن ها نمره کسب کرده است. یعنی به طور مثال یوزری کورسی را تمام کرده و نمره ای برای اون ثبت شده است. تعداد این کورس ها در این قسمت نگهداری میشود. جنس این فیلد Integer است. این تعداد با شمارش تعداد کورس هایی که کاربر ثبت نام کرده است و در آن ها نمره ای دارد (چون اگر نمره ای نداشته باشد در آن جا برای فیلد نمره مقدار Null دارد) از جدول UserOnlineCourse در Staging_area محاسبه می شود و در اینجا قرار میگیرد.

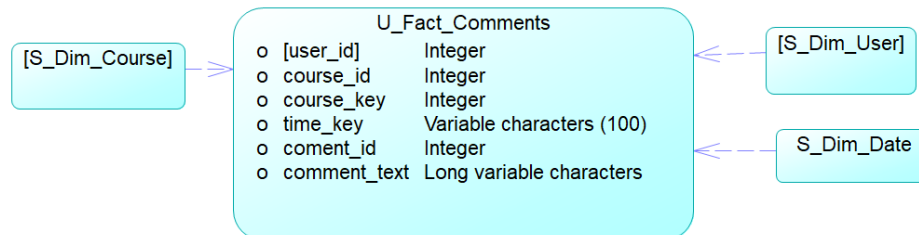
averge_rating: این فیلد برای نگهداری این است که به طور کلی میانگین امتیاز هایی که کاربر از ابتدا تا الان داده است چه مقدار بوده است. جنس این فیلد `Decimal(3,2)` است. این فیلد با محاسبه میانگین امتیاز هایی که در فکت تراکنشی ثبت شده است محاسبه میشود. فکت تراکنشی مربوطه به نام `U_Fact_UserRating` است که در `DataWarehouse` ساخته شده است.

number_of_rating: تعداد رای هایی هست که کاربر تا کنون داده است. یعنی عملا این تعداد مشخص میکند که میانگینی که در فیلد قبلی بررسی شده است برای چه تعداد رای محاسبه شده است. جنس این فیلد `Integer` است. برای محاسبه این کافی است که ببینیم کاربر تا کنون چه تعداد رای داده است که در فکت `U_Fact_UserRating` که در `DataWarehouse` است قبلا آنها به صورت جدا داشته ایم با جمع `incremental` میتوان این عدد را به دست آورد.

max_rating: بیشترین امتیازی است که کاربر تا کنون ثبت کرده است. جنس این فیلد `Decimal(3,2)` است. این عدد عملا از همان فکت تراکنشی امتیاز ها یعنی `U_Fact_UserRating` به دست می آید که قبلا ساخته ایم. در آن فکت برای هر روز بیشتری امتیاز را داریم. کفایت میان آن ها ماکزیم بگیریم که به صورت `incremental` میتوان آن را انجام داد.

فکت تراکنشی `U_Fact_Comments`:

دیاگرام:



شرح:

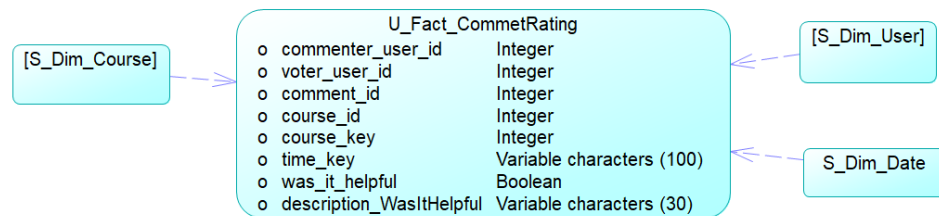
یک فکت تراکنشی برای نگهداری کامنت های کاربران برای course ها است. از جدولی به اسم Comment در staging_area پر می شود که در واقع از source آمده است. از این فکت برای تجمیعی بعدی استفاده می شود.

فیلدها:

- user_id : از فیلدی با همین نام از جدول Comment در staging_area وارد می شود.
- course_key : از فیلدی با همین نام از بعد S_Dim_Course وارد میشود از مقایسه ی course_id در این بعد و جدول Comment در staging_area.
- course_id : از فیلدی با همین نام از جدول Comment در staging_area وارد می شود.
- time_key : توسط function با اسم S_Make_TimeKey و ورودی datetime_created از جدول Comment در staging_area پر می شود. خود S_Make_TimeKey در انبار داده است.
- comment_id : از فیلدی با همین نام از جدول Comment در staging_area وارد می شود.
- comment_text : از فیلدی با همین نام از جدول Comment در staging_area وارد می شود.

فکت تراکنشی U_Fact_CommetRating:

دیاگرام:



شرح:

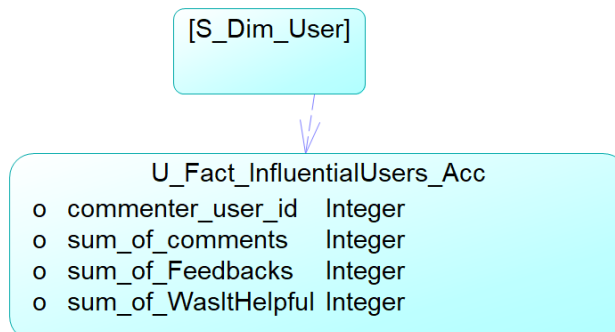
این فکت تراکنشی برای ثبت واکنش ها به هر کامنت قرار داده شده است. اطلاعات اصلی آن از inner join دو جدول source با اسم های Comment و CommentVote می‌آیند که روی comment_id جویین شده اند این جدول حاصل از جویین با اسم U_Fact_CommentRating_Temp ذخیره شده است.

فیلدها:

- commenter_user_id : از همان جدول حاصل از جویین مطرح شده در شرح می‌آید مبدا آن جدول Comment در staging_area است.
- voter_user_id : از همان جدول حاصل از جویین مطرح شده در شرح می‌آید مبدا آن جدول CommentVote در staging_area است.
- course_id : از همان جدول حاصل از جویین مطرح شده در شرح می‌آید مبدا آن جدول Comment در staging_area است.
- course_key : از بعد course می‌آید و باتوجه به course_id مشخص می‌شود.
- time_key : توسط function با اسم S_Make_TimeKey و ورودی datetime_created از جدول CommentVote در staging_area پر می‌شود. خود S_Make_TimeKey در انبار داده است.
- comment_id : از همان جدول حاصل از جویین مطرح شده در شرح می‌آید مبدا آن جدول CommentVote در staging_area است.
- was_it_helpful : از همان جدول حاصل از جویین مطرح شده در شرح می‌آید مبدا آن جدول CommentVote در staging_area است.
- description_WasItHelpful : با توجه به فیلد was_it_helpful مقدار آن مشخص می‌شود. اگر was_it_helpful برابر با 1 باشد مقدار Helpful و اگر برابر با 0 باشد مقدار Not Helpful برای آن قرار داده می‌شود.

فکت U_Fact_InfluentialUsers_Acc:

دیاگرام:



شرح:

یک فکت تجمیعی است. این فکت با محوریت فکت U_Fact_CommentRating پر می‌شود. هدف آن پیدا کردن کاربرانی است که بیشترین کامنت تاثیرگذار را داشته اند. فیلد هایی که نگهداری می‌کند تعداد کل کامنت های یک کاربر، تعداد کل فیدبک هایی که دریافت کرده و تعداد دفعاتی که کامنت هایش مفید واقع شده اند است.

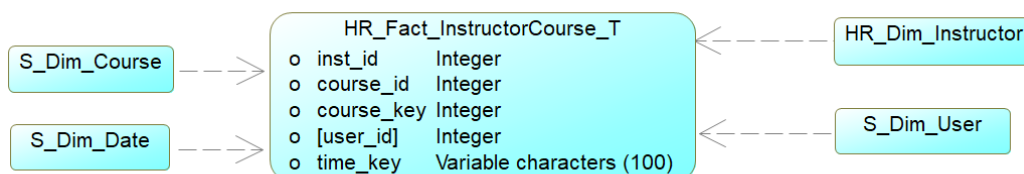
فیلدها:

- commenter_user_id: در هر روز اطلاعات مربوط به فکت U_Fact_CommentRating فراخوانی می‌شود و با خود فکت U_Fact_InfluentialUsers_Acc در حالت left join قرار می‌گیرد و این فیلد هم حاصل این جویین است. (ممکن است کاربری به تازگی کامنت گذاشته باشد که قبلا در فکت نبوده.)
- sum_of_comments: در هر روز اطلاعات مربوط به فکت U_Fact_CommentRating فراخوانی می‌شود و با خود فکت U_Fact_InfluentialUsers_Acc در حالت left join قرار می‌گیرد و این فیلد حاصل جمع خودش با تعداد کامنت های جدیدی است که وارد شده است.
- sum_of_Feedbacks: در هر روز اطلاعات مربوط به فکت U_Fact_CommentRating فراخوانی می‌شود و با خود فکت U_Fact_InfluentialUsers_Acc در حالت left join قرار می‌گیرد و این فیلد حاصل جمع خودش با تعداد فیدبک های جدیدی است که وارد شده است.
- sum_of_WasItHelpful: در هر روز اطلاعات مربوط به فکت U_Fact_CommentRating فراخوانی می‌شود و با خود فکت U_Fact_InfluentialUsers_Acc در حالت left join قرار می‌گیرد و این فیلد حاصل جمع خودش با تعداد فیدبک های مفید جدیدی است که وارد شده است.

فکت های مارت Human_Resources

فکت تراکنشی HR_Fact_InstructorCourse_T:

دیاگرام:



شرح:

یک فکت تراکنشی است که در آن به ازای ثبت نام هر کاربر در یک درس آنلاین، برای تمامی مدرس های مشارکت کننده در تدریس آن درس یک رکورد ثبت می شود. با کمک این فکت امکان تکمیل فکت های بعدی که با هدف گزارش گیری های پیرامون مدرس ها تشکیل شده اند فراهم می شود.

دو پراسیجر برای پر کردن این فکت در نظر گرفته شده است. HR_First_Time_Fill_Fact_InstructorCourse_T. همچنین از پراسیجر HR_Fill_Fact_InstructorCourse_T برای پراسیجری است که برای اولین بار این فکت را پر می کند. همچنین از پراسیجر HR_Fill_Fact_InstructorCourse_T برای پر کردن این فکت در مراحل بعدی استفاده می شود. این دو پراسیجر اطلاعات مورد نظر برای پر کردن این فکت را با join کردن دو جدول UserOnlineCourse و CourseInstructor، که سمت Staging_area قرار دارند، بر روی فیلد مشترک course_id به دست می آورند. حاصل این join در مراحل اجرای ETL در جدولی کمکی با نام HR_Fact_InstructorCourse_T_temp قرار می گیرد که در ادامه از نام فیلد های این جدول در پر شدن فیلد های جدول اصلی کمک گرفته شده است.

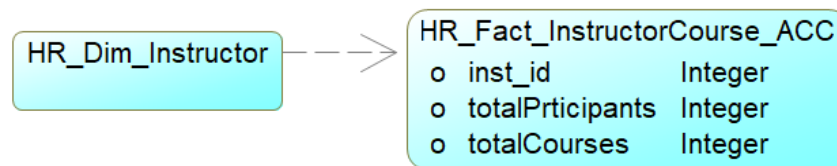
فیلدها:

- Inst_id: شناسه مدرس است؛ این شناسه که کلید بعد مدرس در سمت انبار داده نیز هست، برای هر مدرس یکتاست. این فیلد که از نوع Integer است از فیلدی هم نام از جدول HR_Fact_InstructorCourse_T_temp پر می شود.
- Course_id: شناسه درس است که برای هر درس یکتاست؛ این فیلد که از نوع integer است از روی فیلدی هم نام از جدول HR_Fact_InstructorCourse_T_temp پر می شود.
- Course_key: کلید درس است که از بعد S_Dim_Course دریافت می شود. به سبب اینکه بعد S_Dim_Course یک SCD نوع دوم است، ممکن است برای هر درس با شناسه یکتا چندین رکورد داشته باشد، بنابراین با کمک این کلید، رکورد های بعد نام برده یکتا می شوند و این فیلد است که کلید بعد S_Dim_Course در این فکت است.
- User_id: این فیلد که از نوع integer است شناسه کاربر را نگهداری می کند. این فیلد از روی فیلدی هم نام از جدول HR_Fact_InstructorCourse_T_temp پر می شود.

- Time_key: این فیلد کلید بعد زمان در این فکت است که از نوع NVARCHAR(100) است. این فیلد از روی فیلدی هم نام از جدول HR_Fact_InstructorCourse_T_temp پر می شود. (در خود جدول کمکی نام برده شده، این فیلد خود از روی فیلد datetime_user_enrolled که در جدول UserOnlineCourse قرار دارد پر می شود.)

فکت تجمیعی HR_Fact_InstructorCourse_ACC :

دیاگرام:



شرح:

یک فکت تجمیعی است که برای هر مدرس، آخرین وضعیت از نظر تعداد دانشجویان و تعداد درس هایی که او در آن ها به عنوان مدرس مشارکت داشته است را نگه می دارد. از نظر گزارش گیری های پیرامون وضعیت مدرس ها و میزان جذب دانشجو ها این فکت اهمیت قابل توجهی دارد.

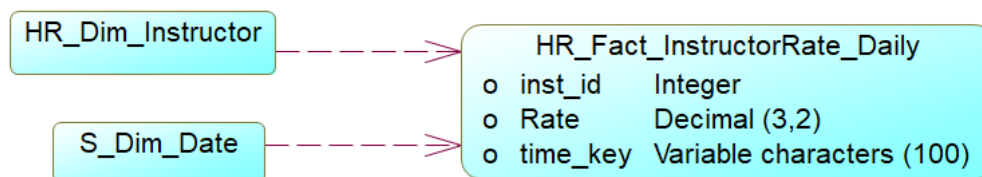
دو پراسیجر برای پر کردن این فکت در نظر گرفته است: HR_First_Time_Fill_Fact_InstructorCourse_ACC و HR_Fill_Fact_InstructorCourse_ACC که به ترتیب برای پر کردن این فکت برای اولین بار و برای سایر دفعات استفاده می شوند. این پراسیجرها اطلاعات مورد نیاز برای پر کردن این فکت را از فکت تراکنشی HR_Fact_InstructorCourse_T استخراج می کنند. نکته ی حائز اهمیت در پر کردن این فکت این است که علی رغم تجمیعی بود، به سبب اینکه measure خاصی که به صورت distributive و یا algebraic قابل محاسبه باشد ندارد. دو فیلد totalParticipants و totalCourses با شمارش تعداد رکورد های فکت تراکنشی HR_Fact_InstructorCourse_T به صورت مدیریت شده محاسبه می شوند.

فیلدها:

- **inst_id**: شناسه مدرس است؛ این شناسه که کلید بعد مدرس در سمت انبار داده نیز هست، برای هر مدرس یکتاست. این فیلد که از نوع Integer است از روی فیلدی هم نام در جدول HR_Fact_InstructorCourse_T پر می شود.
- **totalParticipants**: تعداد ثبت نام هایی که در همه درس های یک مدرس انجام شده است در این فیلد نگهداری می شود. این فیلد که از نوع integer است با شمارش رکورد های جدول HR_Fact_InstructorCourse_T به تفکیک اساتید پر می شود.
- **totalCourses**: تعداد درس هایی که یک مدرس در تدریس آن ها مشارکت داشته است در این فیلد نگهداری می شود. این فیلد که از نوع integer است با شمارش رکورد های جدول HR_Fact_InstructorCourse_T به تفکیک همزمان مدرس ها و درس ها پر می شود.

فکت روزانه HR_Fact_InstructorRate_Daily :

دیاگرام:



شرح:

این فکت از نوع پررودیک روزانه است و وضعیت امتیازی که هر مدرس کسب کرده است را به صورت روزانه ثبت می کند. با کمک این فکت امکان رصد کردن وضعیت عملکرد مدرس ها در طول زمان قابل بررسی است.

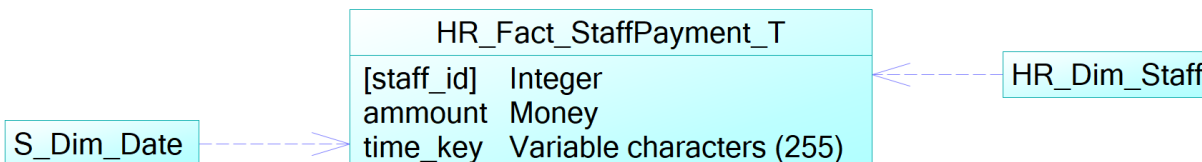
دو پراسیجر برای پر کردن این فکت در نظر گرفته است: HR_First_Time_Fill_Fact_InstructorRate_Daily و HR_Fill_Fact_InstructorRate_Daily که به ترتیب برای پر کردن این فکت برای اولین بار و برای سایر دفعات استفاده می شوند. این پراسیجرها اطلاعات مورد نیاز برای پر کردن این فکت را از جدول Instructor در سمت Staging_area استخراج می کنند.

فیلدها:

- **Inst_id**: شناسه هر مدرس است که برای هر مدرس یکتاست و کلید بعد HR_Dim_Instructor نیز هست. این فیلد از نوع integer است و از روی فیلد با نام مشابه از جدول Instructor در سمت Staging_area پر می شود.
- **Rate**: امتیاز مدرس در این فیلد ذخیره می شود. از نوع decimal است و تا دو رقم اعشار اخذ می کند. از روی فیلد با نام rating از جدول Instructor در سمت Staging_area پر می شود.
- **Time_key**: کلید زمان از بعد تاریخ است. زمان کنونی پراسیجر پر کننده این جدول با کمک تابع S_Make_TimeKey به صورت کلید تبدیل شده و ذخیره می شود.

فکت سالانه HR_Fact_StaffPayment_T :

دیاگرام:



شرح:

این فکت برای نگهداری آمار پرداخت حقوق ها و مبالغ به کارمندان است. به این صورت که هر ردیف آن این موضوع را نشان میدهد: چه کارمندی در چه زمانی چه مقداری را دریافت کرده است. که همانطور که در دیاگرام مشخص است برای staff_id به بعد HR_Dim_Staff ارتباط داریم و از time_key به بعد S_Dim_Date ارتباط داریم.

این فکت با استفاده از دو پراسیجر HR_First_Time_Fill_StaffPayment_Fact_T برای پر کردن اولین بار و همینطور HR_Full_StaffPayment_Fact_T برای دفعات بعدی پرد میشود.

فیلدها:

staff_id: این فیلد نشان میدهد که این پرداخت مربوط به کدام کارمند است. جنس آن Integer است. این فیلد به رکوردی در بعد Staff اشاره می‌کند. مقدار آن از جدول Payment در Staging_area گرفته میشود که دقیقاً فیلدی به همین نام برای هر رکورد در آن جا وجود دارد.

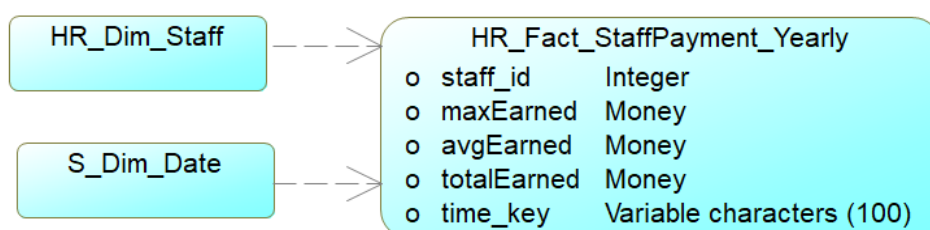
ammount: مقدار پولی است که برای پرداختی به کارمند مورد نظر داده شده است. جنس این فیلد از نوع Money است. مقدار آن از جدول Payment در Staging_area گرفته میشود که دقیقاً فیلدی به همین نام برای هر رکورد در آن جا وجود دارد.

time_key:

این فیلد مشخص میکند که این پرداختی که به کارمند مورد نظر انجام شده است در چه تاریخی انجام شده. این فیلد از نوع NVARCHAR(255) است. این فیلد عملاً به بعد S_Dim_Date اشاره میکند. نحوه ساخت این فیلد به این صورت است که در ادامه گفته میشود. مقدار تاریخی آن از جدول Payment در Staging_area گرفته میشود که فیلدی به نام date_of_payment برای هر رکورد در آن جا وجود دارد. حال ما معادل این تاریخ را در بعد تاریخ به دست می‌آوریم و کلید آن (TimeKey) را در اینجا نگهداری می‌کنیم. این کار را با استفاده از تابعی انجام می‌دهیم که خودمان نوشته ایم. این تابع DataWarehouse.dbo.S_Make_TimeKey نام دارد که از ما تاریخی را با جنس Date می‌گیرد و TimeKey معادل آن را از بعد S_Dim_Date بر می‌گرداند که در این فیلد ذخیره می‌کنیم. نتیجه این تبدیل را که کلیدی از بعد S_Dim_Date هست، در اینجا نگهداری می‌کنیم.

فکت سالانه : HR_Fact_StaffPayment_Yearly

دیاگرام:



شرح:

این فکت که از نوع پرودیگ سالانه است شامل اطلاعات پرداختی با نگاه سالیانه است. هدف این فکت دریافت گزارش های مربوط به پرداخت ها به کارکنان است. از این گزارشات برای تغییر و بهبود سیستم دستمزد و پاداش ها می توان استفاده کرد.

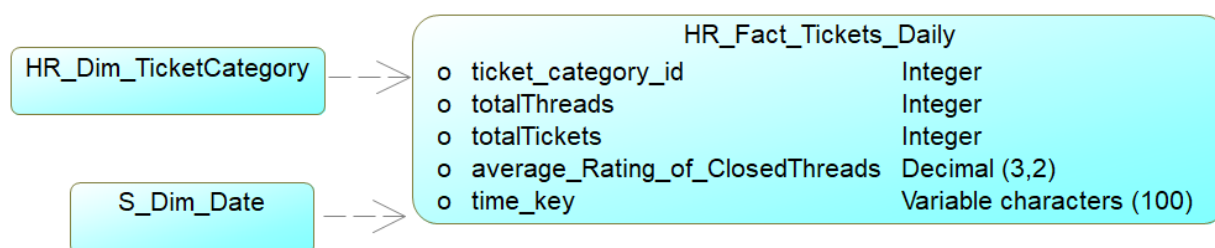
دو پراسیجر برای پر کردن این فکت در نظر گرفته است: HR_First_Time_Fill_Fact_StaffPayment_Yearly و HR_Fill_Fact_StaffPayment_Yearly که به ترتیب برای پر کردن این فکت برای اولین بار و برای سایر دفعات استفاده می شوند. این پراسیجرها اطلاعات مورد نیاز برای پر کردن این فکت را از فکت HR_Fact_StaffPayment_T استخراج می کنند.

فیلدها:

- **Staff_id**: این فیلد شناسه ی کارکنان را نگه می دارد که کلید بعد HR_Dim_Staff است. این فیلد که از نوع integer است از فیلدی با نام مشابه از فکت HR_Fact_StaffPayment_T پر می شود.
- **maxEarned**: برای کارمند، بیشترین دریافتی در طول 12 ماه سال (هرماه یک پرداختی) در این فیلد ذخیره می شود. این فیلد که از نوع Money است با Max گیری روی فیلد amount روی داده های یک سال گذشته از فکت HR_Fact_StaffPayment_T به تفکیک کارکنان به دست می آید.
- **avgEarned**: برا کارمندان میانگین دریافتی در طول 12 ماه یک سال در این فیلد ذخیره می شود. این فیلد که از نوع Money است با میانگین گیری روی فیلد amount روی داده های یک سال گذشته از فکت HR_Fact_StaffPayment_T به تفکیک کارکنان به دست می آید.
- **totalEarned**: برا کارمندان مجموع دریافتی در طول 12 ماه یک سال در این فیلد ذخیره می شود. این فیلد که از نوع Money است با تجمیع روی فیلد amount روی داده های یک سال گذشته از فکت HR_Fact_StaffPayment_T به تفکیک کارکنان به دست می آید.
- **Time_key**: کلید زمان از بعد تاریخ است. زمان کنونی پراسیجر پر کننده این جدول با کمک تابع S_Make_TimeKey به صورت کلید تبدیل شده و ذخیره می شود.

فکت روزانه HR_Fact_Tickets_Daily :

دیاگرام:



شرح:

این جدول به صورت روزانه مجموع تعداد تردهای ایجاد شده، مجموع تعداد بلیط های ایجاد شده، و میانگین رای داده شده به تردهای پایان یافته را بر اساس دسته بندی موجود برای بلیط ها ثبت می کند. هدف آن دریافت گزارش به منظور تحلیل بازخورد کاربران سیستم و بررسی روند پاسخگویی به دغدغه های کاربران از طرف تیم پشتیبانی است.

دو پراسیجر برای پر کردن این فکت در نظر گرفته است: HR_First_Time_Fill_Fact_Tickets_Daily و HR_Fill_Fact_Tickets_Daily که به ترتیب برای پر کردن این فکت برای اولین بار و برای سایر دفعات استفاده می شوند. این پراسیجرها اطلاعات مورد نیاز برای پر کردن این فکت را با join کردن دو جدول TicketThread و Ticket در سمت Staging_area روی فیلد مشترکشان یعنی ticket_thread_id به دست می آورند.

فیلدها:

- **ticket_category_id**: این فیلد کلید بعد HR_Dim_TicketCategory است که برای هر دسته از تیکت ها یکتاست. از نوع integer است و از روی فیلدی با نام مشابه از جدول TicketThread در سمت Staging_area به دست می آید (پس از عمل join ذکر شده در بالا)
- **totalThread**: در این فیلد تعداد کل ترد های هر دسته از تیکت ها در یک روز ثبت می شود. این فیلد از نوع integer است. پس از join دو جدول TicketThread و Ticket تیکت های هر روز استخراج می شوند. با تفکیک تیکت های استخراج شده بر اساس دسته و آیدی ترد یک جدول تولید می شود که با شمارش رکورد های آن به تفکیک ticket_category_id برای هر دسته تعداد ترد های آن دسته محاسبه می شود.
- **Totaltickets**: در این فیلد تعداد کل تیکت های هر دسته از تیکت ها در یک روز ثبت می شود. این فیلد از نوع integer است. پس از join دو جدول TicketThread و Ticket تیکت های هر روز استخراج می شوند. با تفکیک تیکت های استخراج شده بر اساس صرفا دسته یک جدول تولید می شود که با شمارش رکورد های آن به تفکیک ticket_category_id برای هر دسته تعداد تیکت های آن دسته محاسبه می شود.
- **average_Rating_of_ClosedThreads**: در این فیلد تاریخ بسته شدن یک ترد بررسی میشود. اگر برای روز در حال محاسبه بود، امتیاز ترد های از آن نوع را میانگین گرفته و در این فیلد قرار میدهد.

- Time_key: کلید زمان از بعد تاریخ است. زمان کنونی پراسیجر پر کننده این جدول با کمک تابع S_Make_TimeKey به صورت کلید تبدیل شده و ذخیره می شود.