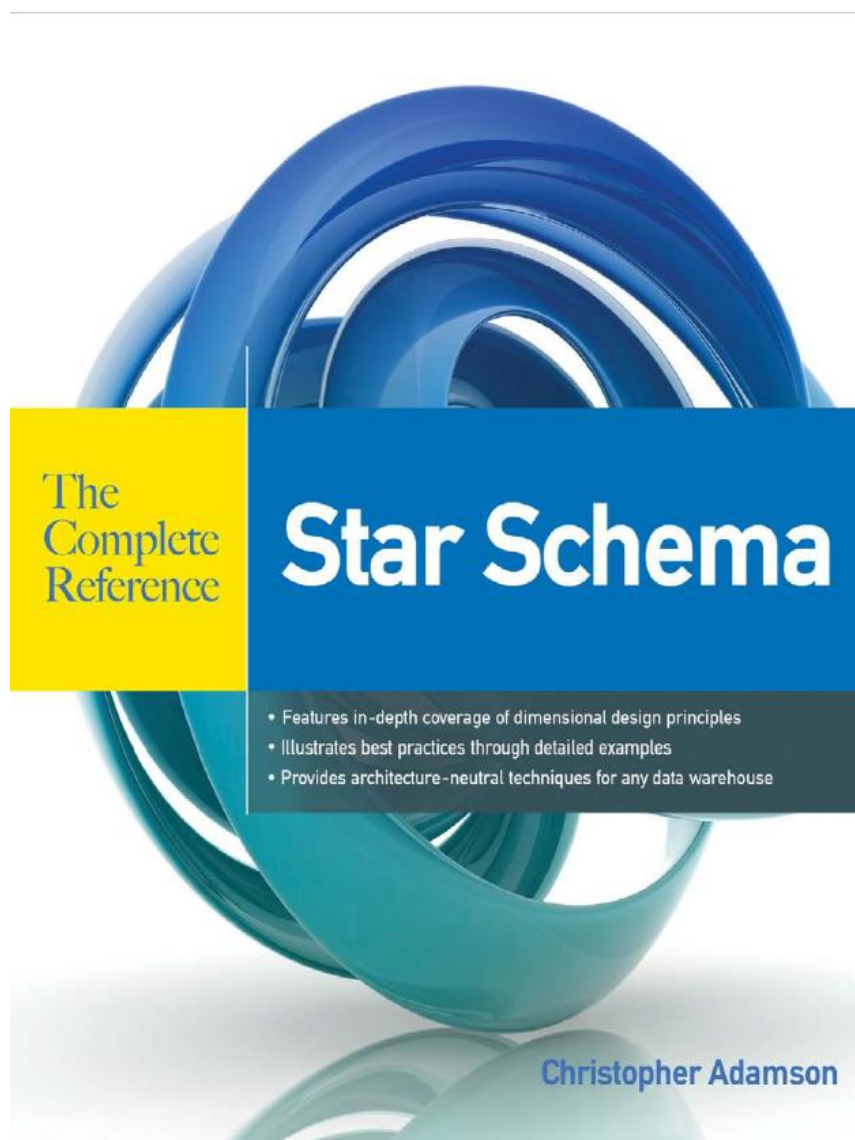


مدل سازی داده

طرح ستاره‌ای - بخش اول

Star Schema



این راهنما بخشی از مطالب مربوط به راهنمای Power BI و زبان DAX است که در کانال تلگرامی t.me/Power_of_BI منتشر می‌شود. برای دسترسی به سایر مطالب مربوط به زبان DAX، زبان M، پاورکوئری، اصول طراحی داشبوردها و چالش‌های دنیای واقعی به این کانال تلگرامی بپیوندید.

دانشگاه یزد - گروه مدیریت صنعتی

دکتر علیرضا ناصر صدرآبادی

زبان DAX برای بسیاری از کاربران Power BI چالش برانگیز است. در واقع، درصد زیادی از افرادی که با Power BI کار می‌کنند، در درک و استفاده از DAX مشکل دارند. اما اگر شما یک مدل داده‌ای خوب طراحی کرده باشید، حتی با دانش پایه از DAX و مفاهیم فیلترگذاری، می‌توانید کارهای زیادی انجام دهید. بنابراین، ساختن یک مدل داده‌ای قوی، برای تحلیل داده در Power BI یک ضرورت حیاتی است. برای ساختن یک مدل داده‌ای خوب، نیازی نیست سال‌ها تجربه داشته داشته باشید، کافی است اصول پایه مانند اطلاعات (Fact)، ابعاد تحلیل اطلاعات (Dimension) و طراحی مدل ستاره ای (Star Schema) را بدانید. راهنمای حاضر که به شکل خلاصه نویسی تهیه شده است به این اصول می‌پردازد.

انواع سیستم‌های اطلاعاتی

سامانه‌های اطلاعاتی به دو دسته کلی تقسیم می‌شوند. گروه اولی شامل سیستم‌هایی است که از اجرای فرآیندهای کسب‌وکار پشتیبانی می‌کنند و سامانه‌های عملیاتی نام دارند. گروه دوم مشتمل بر ابزارهایی است که برای تحلیل این فرآیندها به کار می‌روند و با عنوان سامانه‌های تحلیلی شناخته می‌شوند. اهداف متفاوت این دو دسته نشان می‌دهند که اصول طراحی آنها نیز با هم تفاوت دارند.

سامانه‌های عملیاتی جزئیات مربوط به رویدادهای مهم را به‌طور مستقیم ثبت می‌کنند. فعالیت‌هایی که توسط این سامانه‌ها ثبت می‌شوند، گاه با عنوان "تراکنش‌ها" شناخته می‌شوند و خود سامانه‌ها نیز با عنوان سامانه‌های پردازش تراکنش آنلاین (OLTP) نام‌گذاری می‌شوند. وقتی این سامانه‌ها در پایگاه داده رابطه‌ای پیاده‌سازی می‌شوند، طراحی بهینه‌ی آنها معمولاً در قالب سطح سوم قوانین نرمال سازی پذیرفته می‌شود. این طراحی، عملکرد بالا برای درج، به‌روزرسانی و حذف تراکنش‌های اتمیک را به صورت منظم و قابل پیش‌بینی پشتیبانی می‌کند. نرم‌افزارهای معمول حسابداری و مدیریت انبار و ... نمونه‌هایی رایج از این سامانه‌ها هستند.

سامانه تحلیلی فرآیند را مورد ارزیابی قرار می‌دهد و به سوالات کلیدی کسب و کار پاسخ می‌دهند و شاخص‌های کلیدی عملکرد "درصد تغییر میزان فروش فصل جاری نسبت به دوره مشابه سال قبل" را محاسبه می‌کنند. این شاخص‌ها به ارزیابی کلی فرآیند فروش مربوط می‌شوند و به سفارش‌های خاص و منفرد نمی‌پردازند. تعامل با سامانه‌های تحلیلی فقط از طریق پرس‌وجوها (کوئری‌ها) پی انجام می‌شود که اطلاعات مربوط به فرآیندهای کسب و کار را بازیابی می‌کنند و هیچ اطلاعاتی ایجاد یا اصلاح نمی‌شود. این پرس‌وجوها معمولاً حجم زیادی از تراکنش‌ها را در بر می‌گیرند، در حالی که سامانه‌های عملیاتی عمدتاً بر تراکنش‌های منفرد متمرکز هستند. پرسش‌های مشخصی که از سامانه تحلیلی پرسیده می‌شود کمتر قابل پیش‌بینی بوده و احتمال تغییر آنها در طول زمان بیشتر است. داده‌های تاریخی برای این سامانه‌ها بسیار اهمیت دارند، حتی پس از آنکه استفاده عملیاتی آنها به پایان برسد. سامانه‌های تحلیلی گاه با عنوان انبار داده (Data Mart) یا مخزن موضوعی داده (Data Mart) نیز شناخته می‌شوند. این راهنما به چگونگی طراحی مدل داده در سیستم‌های تحلیلی می‌پردازد. منبع اصلی مطالب با اندکی جرح و تعدیل کتاب Star Schema اثر قابل توجه Christopher Adamson است.

انواع سیستم‌های اطلاعاتی

سیستم‌های عملیاتی

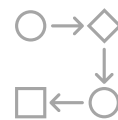


سیستم‌های تحلیلی



سیستم‌های عملیاتی

پشتیبانی از اجرای فرایندها



سرعت در انجام وظایف مشخص و تکراری



حذف داده‌های تکراری برای افزایش کارایی



سیستم‌های عملیاتی با عناوین زیر نیز شناخته می‌شوند:

On Line Transaction Processing (OLTP) System

Transaction System

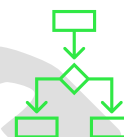
Source System

سیستم‌های تحلیلی

پشتیبانی از ارزیابی عملکرد فرایندها



سامانه پشتیبان تصمیم



وظایف متغیر در طول زمان



سیستم‌های تحلیلی با عناوین زیر نیز شناخته می‌شوند:

Data Warehouse System

Data Mart

اهداف متفاوت، طراحی متفاوت

سیستم عملیاتی: طراحی نرمال



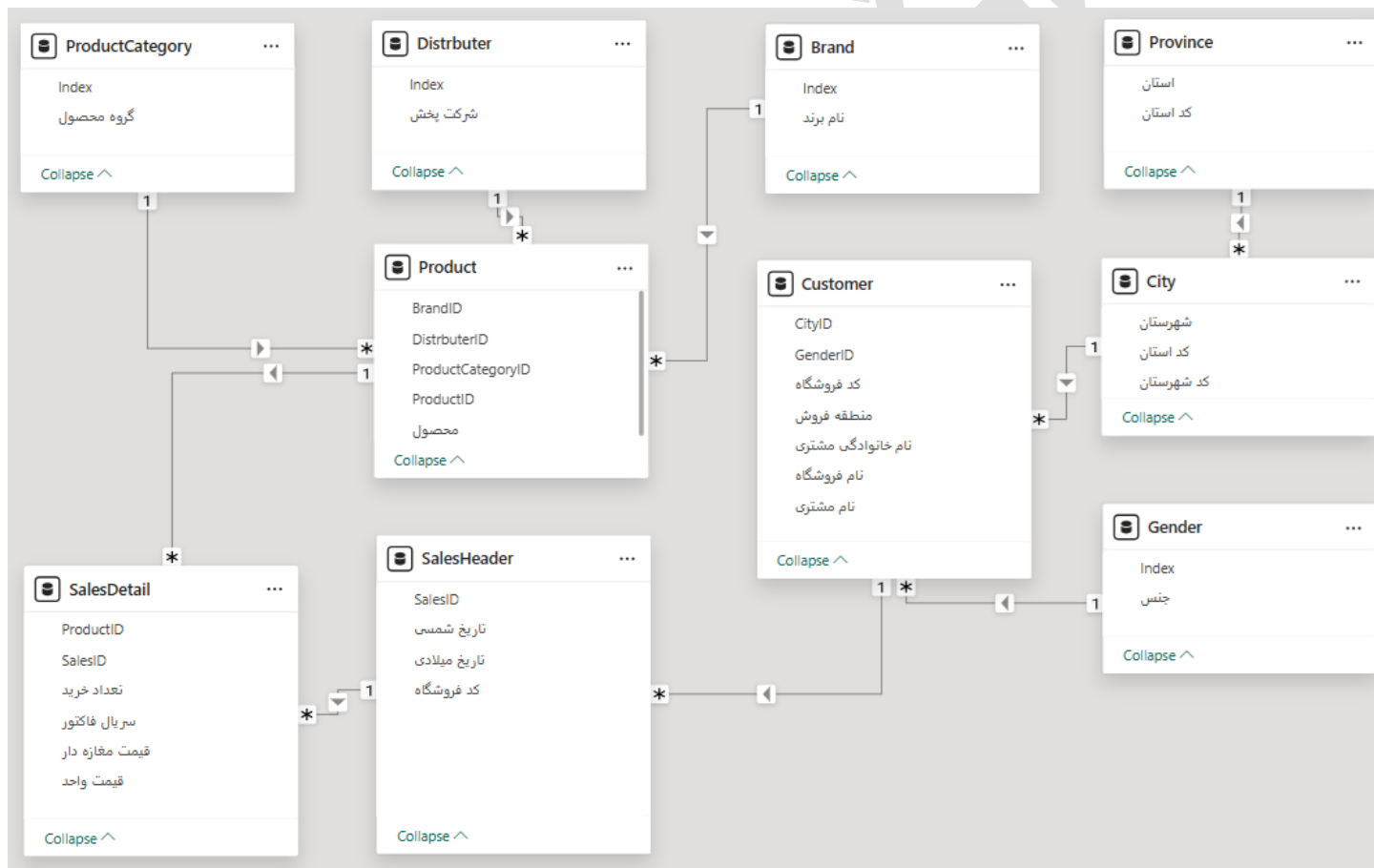
سیستم تحلیلی: طراحی مبتنی بر ابعاد



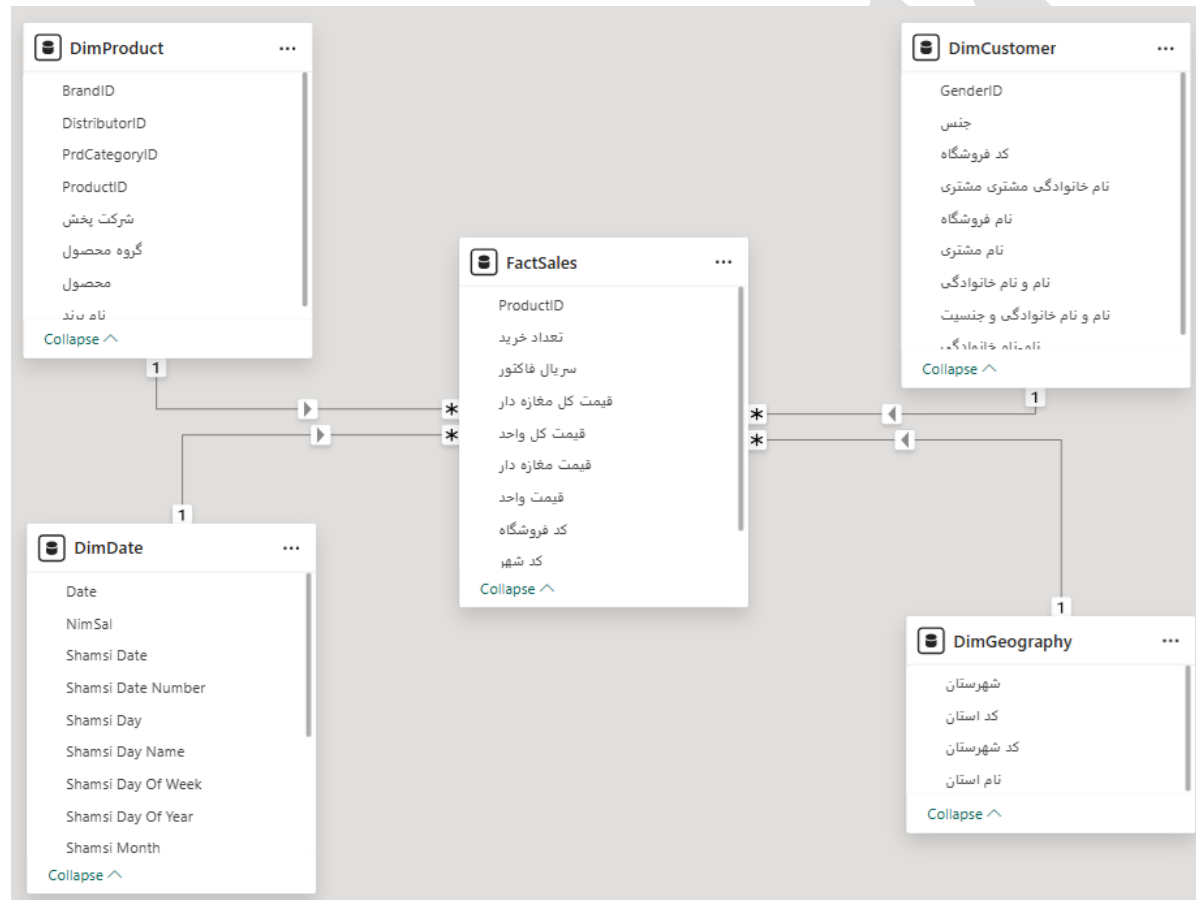
مثال: داده‌های فروش شکلات

قیمت فروش	قیمت مصرف کننده	سریال فاکتور	تعداد خرید	شرکت پخش	نام برند	گروه محصول	محصول	کد فروشگاه	منطقه فروش	شهر	استان	تاریخ شمسی
۳۹۷۰۵	۴۱۰۰۰	۱۰۱۵۱۲۶	۲۴	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-فندق- پیلوپک-۴۰گ	۴۰۰۱۵۴	۱	تهران	تهران	۱۴۰۳/۰۷/۲۵
۳۹۷۰۵	۴۱۰۰۰	۱۰۱۵۱۲۶	۲۴	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-فندق- کارامل- پیلوپک-۴۰گ	۴۰۰۱۵۴	۱	تهران	تهران	۱۴۰۳/۰۷/۰۳
۱۹۸۵۲۵	۲۰۵۰۰۰	۱۰۱۵۴۹۳	۴	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-کاراملی-نوقا-مینیر-مینیر- لیوانی درب دار-۴۰گ	۴۰۰۱۵۷	۱	تهران	تهران	۱۴۰۳/۰۷/۱۲
۲۳۹۶۵	۲۴۵۰۰	۱۰۱۵۷۷۴	۴۸	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار- پیلوپک-۴۵گ	۴۰۰۱۵۷	۱	تهران	تهران	۱۴۰۳/۰۷/۰۴
۴۰۱۰۶	۴۱۰۰۰	۱۰۱۵۷۷۴	۲۴	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-فندق- پیلوپک-۴۰گ	۴۰۰۱۵۷	۱	تهران	تهران	۱۴۰۳/۰۷/۲۴
۱۱۷۳۷۰	۱۲۱۱۰۱	۱۰۱۱۶۴۴	۲۲	گسترش پخش سانا	مانی	نات بار	مانی-نات بار-کرنبری شکلاتی- پیلوپک-۳۵گ	۴۰۰۱۱۹	۱	تهران	تهران	۱۴۰۳/۰۷/۲۵
۱۵۹۴۴	۱۶۳۰۰	۱۰۱۱۲۵۴	۹۶	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-کاراملی-نوقا-نوقا- پیلوپک-۳۵گ	۴۰۰۱۱۹	۱	تهران	تهران	۱۴۰۳/۰۷/۲۴
۳۹۷۰	۴۱۰۰	۱۰۱۱۲۴۷	۱۰۰	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-بادام زمینی-مینیر- پیلوپک-۸گ	۴۰۰۱۱۲	۱۰	تهران	تهران	۱۴۰۳/۰۷/۱۲
۸۲۱۶	۸۴۰۰	۱۰۱۱۷۷۹	۱۸۰	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار- پیلوپک-۱۷گ	۴۰۰۱۱۲	۱۰	تهران	تهران	۱۴۰۳/۰۷/۱۹
۱۵۹۴۴	۱۶۳۰۰	۱۰۰۴۷۵۲	۴۸	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-کاراملی-نوقا-نوقا- پیلوپک-۳۵گ	۴۰۰۰۴۲	۱۰	تهران	تهران	۱۴۰۳/۰۷/۲۷
۱۵۹۴۴	۱۶۳۰۰	۱۰۰۴۷۵۲	۲۴	صنایع غذایی شیرین عسل	نانی	نات بار	نانی-نات بار-سریال بار-غلات حجیم شده- پیلوپک-۲۵گ	۴۰۰۰۴۲	۱۰	تهران	تهران	۱۴۰۳/۰۷/۱۳
۴۰۱۰۶	۴۱۰۰۰	۱۰۰۴۷۵۲	۲۴	صنایع غذایی شیرین عسل	هیس	شکلات فیلد بار	هیس-شکلات فیلد بار-شیری فندق-رایس-کریسپی- پیلوپک-۴۲گ	۴۰۰۰۴۲	۱۰	تهران	تهران	۱۴۰۳/۰۷/۱۹
۱۳۷۵۰۰	۱۳۷۵۰۰	۱۰۰۴۱۵۲	۲۴	شایان کالا ایرانیان	استیکرز	شکلات فیلد بار	استیکرز-شکلات فیلد بار- پیلوپک-۵گ	۴۰۰۰۴۷	۱۰	تهران	تهران	۱۴۰۳/۰۷/۲۶
۱۵۷۹۳	۱۶۳۰۰	۱۰۰۳۸۴۱	۴۸	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-کاراملی-نوقا- پیلوپک-۳۵گ	۴۰۰۰۳۹	۱	تهران	تهران	۱۴۰۳/۰۷/۱۶
۱۵۷۸۵	۱۶۳۰۰	۱۰۱۹۷۳۵	۲۴	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-کاراملی-نوقا-نوقا- پیلوپک-۳۵گ	۴۰۰۱۹۲	۲	تهران	تهران	۱۴۰۳/۰۷/۱۴
۱۵۹۴۴	۱۶۳۰۰	۱۰۱۹۴۱۶	۴۸	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-کاراملی-نوقا-نوقا- پیلوپک-۳۵گ	۴۰۰۱۹۲	۲	تهران	تهران	۱۴۰۳/۰۷/۲۶
۲۰۲۹۵۵	۲۰۵۰۰۰	۱۰۱۴۸۸۸	۲	آیدین(داداش برادر)	آیدین	شکلات فیلد بار	آیدین-شکلات فیلد بار-بادام زمینی-مکث-کیفی-دو سر پرس کارتن-۲۰گ	۴۰۰۱۴۴	۲	تهران	تهران	۱۴۰۳/۰۷/۲۱
۲۰۲۹۵۵	۲۰۵۰۰۰	۱۰۱۴۸۸۸	۲	آیدین(داداش برادر)	آیدین	شکلات فیلد بار	آیدین-شکلات فیلد بار-قهوه-مکث-دو سر پرس-کیفی کارتن-۲۰گ	۴۰۰۱۴۴	۲	تهران	تهران	۱۴۰۳/۰۷/۰۹
۲۳۴۸۶	۲۴۵۰۰	۱۰۱۷۷۷۰	۴۸	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار- پیلوپک-۴۵گ	۴۰۰۱۷۴	۲	تهران	تهران	۱۴۰۳/۰۷/۲۴
۱۵۶۲۵	۱۶۳۰۰	۱۰۱۷۷۷۰	۴۸	صنایع غذایی شیرین عسل	نانی	نات بار	نانی-نات بار-سریال بار-غلات حجیم شده- پیلوپک-۲۵گ	۴۰۰۱۷۴	۲	تهران	تهران	۱۴۰۳/۰۷/۱۹
۲۳۹۶۵	۲۴۵۰۰	۱۰۱۶۴۷۴	۴۸	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار- پیلوپک-۴۵گ	۴۰۰۱۶۱	۳	تهران	تهران	۱۴۰۳/۰۷/۱۳
۹۹۱۷۶	۱۰۱۲۰۰	۱۰۰۶۴۷۷	۶	شرکت دنیای شیرین شانجان	فریش	شکلات فیلد بار	فریش-شکلات فیلد بار-بادام-چهار تپه- پیلوپک-۳۰گ	۴۰۰۰۶۴	۴	تهران	تهران	۱۴۰۳/۰۷/۱۷
۹۹۱۷۶	۱۰۱۲۰۰	۱۰۰۶۴۷۷	۶	شرکت دنیای شیرین شانجان	فریش	شکلات فیلد بار	فریش-شکلات فیلد بار-بادام-هشت تپه- پیلوپک-۵گ	۴۰۰۰۶۴	۴	تهران	تهران	۱۴۰۳/۰۷/۱۰
۱۵۴۶۹	۱۶۳۰۰	۱۰۱۶۴۹۷	۲۴	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-کاراملی-نوقا-نوقا- پیلوپک-۳۵گ	۴۰۰۱۶۴	۴	تهران	تهران	۱۴۰۳/۰۷/۲۸
۳۰۹۳۹	۳۲۶۰۰	۱۰۱۶۴۹۷	۲۴	صنایع غذایی شیرین عسل	وینر	شکلات فیلد بار	وینر-شکلات فیلد بار-بادام زمینی- پیلوپک-۴۰گ	۴۰۰۱۶۴	۴	تهران	تهران	۱۴۰۳/۰۷/۱۷
۳۹۳۰۴	۴۱۰۰۰	۱۰۱۶۶۵۰	۲۴	صنایع غذایی شیرین عسل	نانی	شکلات فیلد بار	نانی-شکلات فیلد بار-فندق- کارامل- پیلوپک-۴۰گ	۴۰۰۱۶۴	۴	تهران	تهران	۱۴۰۳/۰۷/۰۸

مثال فروش شکلات: طراحی نرمال



مثال فروش شکلات: طراحی مبتنی بر ابعاد



سوالات کلیدی کسب و کار

سیستم‌های تحلیلی برای پاسخ به سوالات کلیدی کسب و کار طراحی می‌شوند و به سوالاتی از جنس زیر پاسخ می‌دهند:

روند سفارش‌ها این ماه در مقایسه با ماه گذشته چگونه است؟

وضعیت سود در مقایسه با اهداف تعیین شده سه‌ماهه چگونه است؟

آیا یک کمپین تبلیغاتی خاص تأثیری بر فروش گذاشته است؟

بهترین مشتریان ما چه کسانی هستند؟

در همه ی این سوالات دو عامل مهم وجود داد:

* ارقام سنجیدنی مثل میزان فروش، سود و میزان اختلاف با هدف.

* عامل تحلیل که بستر تحلیل سنجیدنی‌ها است مثلاً میزان سود در بستر ماه تحلیل می‌شود.

بدون وجود بستر تحلیل، مقدار اندازه‌گیری شده هیچ مفهومی ندارد. فرض کنید در یک گزارش چنین بیان شود که سود برابر با ده میلیون تومان است. تا زمانی که مشخص نشود این سود مربوط به یک روز است یا یک ماه، مربوط به یک محصول است یا همه محصولات، در مقایسه با دوره قبل افزایش داشته یا کاهش و ... نمی‌توان آن را مثبت یا منفی ارزیابی کرد و بنابراین نمی‌تواند ابزار مناسبی برای کمک به تصمیم‌گیری باشد.

اطلاعات و ابعاد تحلیل اطلاعات

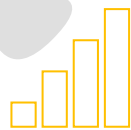
در طراحی مبتنی بر ابعاد "ارقام سنجیدنی" را اصطلاحاً اطلاعات (Fact) و "ارقام مربوط به بستر تحلیل" را اصطلاحاً بعد تحلیل اطلاعات (Dimension) می‌نامیم. مثلاً اگر وزن محصولات فروخته شده به تفکیک فصل‌های سال محاسبه شود، "وزن محصول" همان Fact و "فصل" معرف Dim است.

اطلاعات و ابعاد تحلیل اطلاعات

اقلام سنجیدنی "اطلاعات (Fact)" نام دارند .



بستر تحلیل اطلاعات "بُعد (Dimention)" نام دارد.



مثال‌های Dim و Fact برای مثال فروش شکلات

Dimention		Fact
محصول	اقلام ستون Fact می‌تواند به تفکیک هریک از اقسام ستون Dimention مورد تحلیل قرار بگیرد	تعداد فروش
گروه محصول		وزن فروش
برند		مبلغ فروش
مشتری		
گروه سن مشتری		
جنسیت مشتری		
تاریخ		
ماه		
فصل		
سال		
استان		
شهرستان		
منطقه		

ساماندهی ابعاد و اطلاعات در قالب جدول

Dim های مرتبط با هم که ماهیت وابسته دارند، در یک جدول قرار می‌گیرند.



Fact های سازگار با هم در یک جدول قرار می‌گیرند.



در جداول Fact سازگاری از دو بعد باید وجود داشته باشد؛

داده‌ها به صورت همزمان تولید شوند.

سطح ریزدانی (سطح جزئیات) داده‌ها یکسان باشد.

در طراحی مبتنی بر ابعاد تفاوت‌های مهمی نسبت به طراحی نرمال وجود دارد که مهم‌ترین مورد به قوانین نرمال‌سازی مربوط می‌شود. در سیستم‌های عملیاتی جداول پایگاه داده تا سطح سوم نرمال‌سازی از این قوانین پیروی می‌کنند تا با حداقل سازی فضای مصرفی حافظه به حداکثر سرعت و کارایی در ذخیره و بازیابی اطلاعات برسند. در طراحی مبتنی بر ابعاد اصلاً اقدامات ذخیره و بازیابی صورت نمی‌گیرد و اثربخشی باید در سرعت پاسخ‌گوئی کوئری‌های رخ بدهد که در راستای تولید گزارش‌های هوش تجاری مورد استفاده قرار می‌گیرند. این نیاز باعث می‌شود رعایت قوانین نرمال‌سازی در ساخت جدول‌های Fact و Dim جایگاهی نداشته باشد و به شکل‌های متفاوتی این قوانین نادیده گرفته شوند (بجز قانون وجود کلید منحصر به فرد).

جدول Dimention

جدول عریض است یعنی تعداد ستونهای آن زیاد است.



تعداد سطرهای جدول کم است.



هر جدول Dim یک ستون شامل کلید منحصر به فرد دارد.



DimGepgraphy	DimProduct	DimCustomer	DimDate
Location_Key 🔑	Product_Key 🔑	Customer_Key 🔑	Date_Key 🔑
استان	محصول	مشتری	تاریخ
شهرستان	گروه محصول	گروه سن مشتری	ماه
منطقه	برند	جنسیت	فصل
	شرکت پخش		سال

با توجه به ماهیت طراحی تحلیلی، اطلاعات تکراری و ترکیبی دارد

نام و نام خانوادگی و جنسیت	نام و نام خانوادگی	نام-نام خانوادگی	جنس	GenderID	نام خانوادگی مشتری مشتری	نام مشتری	نام فروشگاه	کد فروشگاه
آقای سعید حسینی	سعید حسینی	سعید-حسینی	مرد	1	حسینی	سعید	فروشگاه حسینی	400154
آقای متین راد	متین راد	متین-راد	مرد	1	راد	متین	بازارچه	400157
آقای ایمان صادقیان	ایمان صادقیان	ایمان-صادقیان	مرد	1	صادقیان	ایمان	سوپر محله	400119
آقای حسین لطفی	حسین لطفی	حسین-لطفی	مرد	1	لطفی	حسین	فروشگاه اخراجی ها	400112
آقای رضا طراوت	رضا طراوت	رضا-طراوت	مرد	1	طراوت	رضا	سوپر 8	400042
آقای مهدی ناصری	مهدی ناصری	مهدی-ناصری	مرد	1	ناصری	مهدی	شکلات تو	400047
آقای علی نمکی	علی نمکی	علی-نمکی	مرد	1	نمکی	علی	اجیلی	400039
خانم حمیده نجاریان	حمیده نجاریان	حمیده-نجاریان	زن	2	نجاریان	حمیده	سوپر کیان	400192
آقای محمد صرافت	محمد صرافت	محمد-صرافت	مرد	1	صرافت	محمد	فروشگاه شایان	400144
آقای هیراد کرانه	هیراد کرانه	هیراد-کرانه	مرد	1	کرانه	هیراد	فروشگاه نوین	400174
خانم مریم مهبد	مریم مهبد	مریم-مهبد	زن	2	مهبد	مریم	بازارک	400161

جدول Fact

جدول عمیق است یعنی تعداد سطرهای آن زیاد است.



تعداد ستون‌های جدول کم است.



متشکل از ستون‌های سنجیدنی و کلید جدول‌های Dim است.



Fact
Date_Key 🔑
Customer_Key 🔑
Product_Key 🔑
Location_Key 🔑
تعداد فروش
وزن فروش
مبلغ فروش

جدول Fact همانند جدول Dim می‌تواند شامل اطلاعات تکراری و وابسته به هم باشد. مثلاً می‌تواند شامل قیمت واحد، مقدار فروش و حاصلضرب آن‌ها یعنی مبلغ قابل پرداخت نیز باشد.

* کلید جدول لزوماً کلید اصلی (Primary Key) نیست و می‌تواند کلید جایگزین (Surrogate Key) باشد که در طی فرایند پاکسازی داده ایجاد می‌شود اما در این مرحله آموزش از پرداختن به آن خودداری می‌شود تا آموزش در این گام شامل پیچیدگی نشود اما در مثال تکمیلی با این کلید و کارکرد آن در نگه‌داری اطلاعات تاریخی آشنا خواهید شد.

Star Schema

طرح ستاره‌ای

هنگامی که جداول Dim و Fact با ساختاری که گفته شد ایجاد شوند و رابطه بین آن‌ها بر اساس فیلد مشترک Data Model ایجاد شود ساختاری که شکل می‌گیرد را طرح ستاره‌ای یا طرح‌واره ستاره می‌نامیم. این نام‌گذاری به دلیل شباهت ظاهر مدل داده به ستاره است (زمانی که جدول Fact در وسط قرار بگیرد و جداول Dim در اطراف آن واقع شوند). در تصویر زیر طرح ستاره‌ای داده‌های مثال فروش شکلات را مشاهده می‌کنید.

