

مدل بهبودیافته دولایه‌ای در طراحی سطح منطقی پایگاه داده تحلیلی

محیا ارومیه^۱ و نگین دانش پور^{۲*}

اطلاعات مقاله	چکیده
دریافت مقاله: ۱۳۹۴/۰۵/۲۸ پذیرش مقاله: ۱۳۹۶/۰۷/۲۵	پایگاه داده تحلیلی، مخزن داده‌ای متمرکز، جمع آوری شده از منابع اطلاعاتی مختلف و ناهمگن در یک محدوده وسیع زمانی و برای پشتیبانی از سیستم‌های تصمیم‌یار می‌باشد. پایگاه داده تحلیلی منبع داده‌ای است که در فرایند تصمیم‌گیری از طریق پردازش تحلیلی بر خط استفاده می‌شود. فرایند توسعه یک پایگاه داده تحلیلی با تحلیل پایگاه داده عملیاتی، شناسایی نیازهای تحلیلی و نهایتاً طراحی در سه سطح مفهومی، منطقی و فیزیکی انجام می‌شود. در این مقاله ابتدا مدل‌های طراحی موجود در سطوح مختلف پایگاه داده تحلیلی بررسی می‌شود. سپس مدل‌های سطح منطقی با توجه به خصوصیات مطرح در آن‌ها مقایسه و تحلیل شده، و نهایتاً مدلی بهبودیافته در این سطح، برای فرایند طراحی پایگاه داده تحلیلی، ارائه می‌شود که ترکیبی از دو مدل ستاره‌ای و دانه‌برفی، به صورت دولایه‌ای است. به منظور مقایسه مدل پیشنهادی با مدل‌های موجود، از معیار زمان پاسخگویی به پرس‌وجوها استفاده شده‌است. آزمایشات نشان می‌دهد که مدل پیشنهادی منجر به بهبود زمان پاسخ به پرس‌وجوها می‌شود. در واقع زمان پاسخ به پرس‌وجوها در مدل پیشنهادی (دولایه‌ای) نسبت به مدل ستاره‌ای بطور متوسط ۵۸/۵۳ درصد و نسبت به مدل دانه‌برفی بطور متوسط ۹۶/۶۱ درصد بهبود می‌یابد.
واژگان کلیدی: پایگاه داده تحلیلی در سطح مفهومی، پایگاه داده تحلیلی در سطح منطقی، مدل دانه‌برفی، مدل ستاره‌ای، طراحی پایگاه داده تحلیلی.	

۱- مقدمه

پایگاه داده تحلیلی^۳ منبع داده‌ای مبتنی بر موضوع^۴، مجتمع^۵، متغیر با زمان^۶ و غیر فرار^۷ است که معمولاً برای تصمیم‌گیری سازمانی استفاده می‌شود، پایگاه داده تحلیلی حجم زیادی از اطلاعات از منابع داده چندگانه را که برای پرس‌وجو و تحلیل استفاده می‌شود، ذخیره می‌کند [۱]. فرایند توسعه پایگاه داده تحلیلی با تحلیل پایگاه داده عملیاتی و شناسایی نیازهای تحلیلی شروع می‌شود و طراحی آن در سه سطح مفهومی، منطقی و فیزیکی انجام می‌شود. این مقاله بر روی سطح منطقی طراحی پایگاه داده

تحلیلی تمرکز دارد.

مدل‌های طراحی در سطح مفهومی به چهار گروه تقسیم می‌شوند [۲] که عبارتند از: ۱. مدل‌های رابطه موجودیت گسترش‌یافته، ۲. مدل‌های شی‌گرا، ۳. مدل‌های بر مبنای آنتولوژی، ۴. مدل‌های ادهاک^۸. این مدل‌ها را می‌توان بر اساس ۲۲ خصوصیت، ارزیابی و مقایسه نمود. این خصوصیات از مطالعات انجام شده بر روی مدل‌های مختلف استخراج شده‌اند [۳ و ۴]، و عبارتند از: استقلال در پیاده‌سازی^۹، جداسازی صریح ساختار و محتوا^{۱۰}، سلسله‌مراتب صریح^{۱۱}، رفتار متقارن برای ابعاد و معیارها^{۱۲}، سلسله‌مراتب

*پست الکترونیک نویسنده مسئول: ndaneshpour@sru.ac.ir

۱. دانشجو، دانشگاه آزاد اسلامی، واحد قزوین، دانشکده مهندسی کامپیوتر و

فناوری اطلاعات، قزوین، ایران

۲. استادیار، دانشکده مهندسی کامپیوتر، دانشگاه تربیت دبیر شهید رجایی،

تهران، ایران

۳ Data Warehouse

۴ Subject-oriented

۵ integrated

۶ time-variant

۷ non-volatile

۸ ad-hoc

۹ Implementation independent

۱۰ Explicit Separation of Structure and Contents

۱۱ Explicit hierarchies

۱۲ Symmetric treatment for dimensions and measures

کاملاً مدل دانه‌برفی (نرمال شده) و در لایه دوم کاملاً مدل ستاره‌ای (غیرنرمال) است. ایجاد جداول مدل ستاره‌ای براساس جداول لایه اول و نیاز مدیر پایگاه داده تحلیلی می‌باشد.

مدل پیشنهادی با توجه به مدل‌های مطرح شده قبلی از جمله خوشه‌ستاره‌ای و دانه‌ستاره‌ای که به صورت ترکیبی از مدل‌ها هستند، ایجاد می‌شود. مدل دانه‌ستاره‌ای به صورت تک لایه‌ای، و دارای ساختاری ترکیبی است؛ شامل ترکیبی از مدل ستاره‌ای و دانه‌برفی. این مدل ترکیبی از نرمال‌سازی (نگرش دانه‌برفی) و غیر نرمال‌سازی (نگرش ستاره‌ای) است. یعنی در مدل تعدادی از جداول بعد به صورت سلسله‌مراتبی (بر اساس نگرش دانه‌برفی) و تعدادی دیگر از جداول غیرنرمال (بر اساس نگرش ستاره‌ای) هستند و براساس این طراحی به پرس‌وجوهای مورد نیاز کاربر پاسخ می‌دهد. در مدل خوشه‌ستاره‌ای، در سطح نرمال‌سازی، جدولی که در موقعیت یک پدر و دو فرزند قرار دارد، یعنی در دو بعد مشترک است، تبدیل به یک زیر بعد می‌شود.

همچنین مدل پیشنهادی با توجه به هدف از مدل‌سازی بعدی که ایجاد ساختارهای پایگاه داده‌ای است که استفاده (نوشتن پرس‌وجوها و پاسخگویی سریع) از آن‌ها برای کاربران نهایی آسان باشد و این هدف با حداقل کردن شماری از جداول و روابط بین آن‌ها بدست می‌آید، ایجاد می‌شود. بنابراین طبق اهداف مدل‌سازی بعدی، در مدل پیشنهادی از مدل ستاره‌ای که دارای حداقل جداول و روابط است در لایه دوم استفاده می‌شود.

در واقع مدل پیشنهادی به صورت ترکیبی از مدل‌های ستاره‌ای و دانه‌برفی، با توجه به مزایای این دو مدل، در راستای رفع معایب آن‌ها پیشنهاد می‌شود. به این منظور مدل پیشنهادی، بر اساس ویژگی دولایه‌ای بودن اینگونه

چندتایی در ابعاد^{۱۳}، ویژگی‌های ابعاد^{۱۴}، پشتیبانی از تجمیع^{۱۵}، معیارهای پیچیده^{۱۶}، مدیریت سطوح مختلف دانه‌دانگی^{۱۷}، پشتیبانی از سلسله مراتب نامتوازن^{۱۸}، پشتیبانی از سلسله‌مراتب ضعیف^{۱۹}، پشتیبانی از رابطه چندبند^{۲۰}، سلسله‌مراتب تعمیم یافته/خصوصی‌شده^{۲۱}، مدیریت تغییرات در طول زمان^{۲۲}، مدیریت عدم قطعیت^{۲۳}، شمایایی با چندین معیار^{۲۴}، خلاصه‌سازی^{۲۵}، توابع تجمیع تعریف شده از طریق کاربر^{۲۶}، اشتراک‌گذاری ابعاد^{۲۷}، تجمیع با کاهش ابعاد^{۲۸}، تجمیع با کاهش معیارها^{۲۹}، تجمیع‌سازی از تجمیع‌های قبلی^{۳۰}.

مدل‌های طراحی مطرح در سطح منطقی عبارتند از: مدل تخت^{۳۱} [۵ و ۶]، مدل مسطح^{۳۲} [۵ و ۶]، مدل ستاره‌ای^{۳۳} [۵-۷]، مدل فلکی^{۳۴} [۵ و ۶]، مدل کهکشانی^{۳۵} [۵ و ۶]، مدل دانه‌برفی^{۳۶} [۵-۷]، مدل ستاره‌ای^{۳۷} [۵ و ۸] و مدل خوشه‌ستاره‌ای^{۳۸} [۵-۷]. این مدل‌ها را می‌توان بر اساس هفت خصوصیت ارزیابی و مقایسه نمود. این خصوصیات از مطالعات انجام شده بر روی مدل‌های مختلف استخراج شده‌اند [۵]، [۹ و ۱۰] و عبارتند از: ۱. کارایی^{۳۹}، ۲. قابلیت استفاده^{۴۰}، ۳. قابلیت استفاده-مجدد^{۴۱}، ۴. انعطاف‌پذیری^{۴۲}، ۵. افزونگی^{۴۳}، ۶. پیچیدگی^{۴۴} و ۷. اعتبار^{۴۵}.

طراحی در سطح فیزیکی نیز نمایش واقعی از جداول فیزیکی در پایگاه داده است که به پیاده‌سازی پایگاه داده تحلیلی مربوط می‌شود [۱۰].

در این مقاله مدل‌های در سطح منطقی با توجه به خصوصیات مطرح در آن‌ها دسته‌بندی و ارزیابی می‌شوند. همچنین مدلی بهبودیافته به صورت دولایه‌ای، برای فرایند طراحی پایگاه داده تحلیلی در سطح منطقی ارائه شده که ترکیبی از دو مدل ستاره‌ای و دانه‌برفی است. یعنی لایه اول

³⁰ Aggregation from aggregation

³¹ Flat schema

³² Terraced schema

³³ Star schema

³⁴ Constellation schema

³⁵ Galaxy schema

³⁶ Snowflake schema

³⁷ Star flake schema

³⁸ Star cluster schema

³⁹ Efficiency

⁴⁰ Usability

⁴¹ Reusability

⁴² Flexibility

⁴³ Redundancy

⁴⁴ complexity

⁴⁵ Reliability

¹³ Multiple hierarchies in dimension

¹⁴ Dimension/level attributes

¹⁵ Support for aggregation

¹⁶ Complex Measures

¹⁷ Handling different levels of granularity

¹⁸ Support for non-onto hierarchies

¹⁹ Support for non-strict hierarchies

²⁰ Support for many-to-many relationships

²¹ Generalisation/specialisation hierarchies

²² Handling change over time

²³ Handling uncertainty

²⁴ Multi-cube/fact schema

²⁵ Summarisability

²⁶ User defined Aggregation functions

²⁷ Drill-across

²⁸ Dimensionless aggregation

²⁹ Measureless aggregation

طراحی پایگاه داده تحلیلی به نام مدل سازی بعدی پیشنهاد می دهد. مدل سازی بعدی روشی غالب برای طراحی پایگاه داده تحلیلی و دیتا مارت ها^{۴۸} است.

در رابطه با طراحی کارهای زیادی انجام شده [۱۲- ۱۷] همچنین در [۱۸] طراحی انبار الگو^{۴۹} و مباحث کیفیتی مربوط به آن بررسی شده است. در واقع در این مقاله طراحی شمای دانه برقی با انبار الگو گسترش پیدا می کند. الگوها با سیستم مدیریت انبار الگو^{۵۰} مدیریت می شوند. انبار الگو موضوع محور است و از واژه محتوا^{۵۱} به عنوان یک جداکننده مجازی در میان الگوها استفاده می کند. در این مقاله چهار نوع محتوا به همراه مزایا و معایبشان بیان شده اند که عبارتند از: محتوای داده جهانی، محتوای داده دامنه-ای، محتوای سناریو و محتوای تکنیکها و نوعی از دانش. فرایند طراحی انبار الگو روالی است که با تحلیل و شناسایی نیازها شروع می شود و سپس طراحی مفهومی و منطقی صورت می گیرد. در نتیجه براساس روابطی ابتدا مدل مفهومی ایجاد شده است، و از شمای دانه برقی برای طراحی منطقی انبار الگو استفاده کرده است. جداول بعد در اینجا جداول نرمالی هستند که الگوها را ذخیره می کنند. هر بعد گروه خاصی از الگوها را نشان می دهد. در اینجا دو سطح از سلسله مراتب وجود دارد. سطح اول، نوع الگو براساس محتوای تکنیکی (تکنیکها شامل: clustering, classification, association) است. سطح دوم، نوع الگو بر اساس محتوای سناریو است. همچنین بعدهای دیگر می توانند شامل سلسله مراتب باشند و منجر به گسترش مدل شوند.

در [۱۹] یک روش اتوماتیک برای تولید الگوهای تحلیلی چندبعدی مفید از منابع داده مربوط^{۵۲} ارائه می شود. این الگوها به شمای ستاره ای چندبعدی از پایگاه داده تحلیلی رابطه ای شباهت دارند. پایه اساسی روش این مقاله یک چارچوب آماری است که مفهوم و نمونه داده^{۵۳} را می گیرد و از چارچوب آماری برای تولید ستاره های تحلیلی چند بعدی^{۵۴} استفاده می کند. در [۲۰] ارزیابی تجربی از متریک ها برای مدل های چندبعدی پایگاه داده تحلیلی در سطح مفهومی ارائه می شود. ویژگی های کیفی ادراکی^{۵۵} و

ایجاد می شود که عیب مدل دانه برقی (لایه پایینی) که پیچیدگی بالای آن است و منجر به افزایش زمان پاسخ به پرس و جوها می شود را با افزودن دیدهای ذخیره شده مناسب برطرف می کند. در واقع برای ایجاد این دیدها، لایه-ی بالایی (مدل ستاره ای) ایجاد می شود. در رابطه با عیب مدل ستاره ای افزونگی داده در بالاترین سطح است و افزونگی مشکلات زیر را می تواند ایجاد کند: هدر رفتن فضای ذخیره سازی، ایجاد تناقض در هنگام بهنگام سازی، حذف و درج ولی از آنجایی که درج و تغییر اطلاعات در پایگاه داده تحلیلی توسط کاربران عادی انجام نمی شود بلکه توسط مدیر سیستم و در واقع طراح و ایجادکننده آن صورت می گیرد، افزونگی اطلاعات منجر به تناقض نخواهد شد. از طرف دیگر این افزونگی ها در سطح ابعاد صورت می گیرد و از آنجایی که حجم جداول بعد در مقابل حجم جدول حقیقت ناچیز است، بنابراین این عیب تاثیر چندانی در درصد افزایش حجم کل داده ها ندارد.

در این مقاله در بخش ۲ مدل های مطرح در سطوح مفهومی و منطقی با توجه به خصوصیات مربوط در آن ها معرفی شده است. در بخش ۳ مدل پیشنهادی معرفی شده است. در بخش ۴ پیاده سازی و ارزیابی مدل پیشنهادی ارائه می شود؛ و در نهایت در بخش ۵ مدل های سطح منطقی، با توجه به وجود یا عدم وجود خصوصیات مذکور، در جداولی مقایسه شده اند. در پایان نتیجه گیری انجام شده است.

۲- بررسی کارهای مرتبط

طراحی پایگاه داده تحلیلی، یکی از مهم ترین مباحث در ایجاد آن، به منظور پاسخگویی مناسب به پرس و جوهای تحلیلی کاربران نهایی می باشد.

براساس کیمبال [۱۱] محیط پایگاه داده تحلیلی^{۴۶} متفاوت با محیط پایگاه داده عملیاتی^{۴۷} است و تکنیک های استفاده شده برای طراحی پایگاه داده عملیاتی، نامناسب برای طراحی پایگاه داده تحلیلی است. به همین دلیل کیمبال یک تکنیک جدید برای مدل سازی داده خصوصا برای

⁵¹ context

⁵² linked data

⁵³ instance data

⁵⁴ multidimensional analytical stars

⁵⁵ understandability

⁴⁶ OLAP

⁴⁷ OLTP

⁴⁸ Data Mart

⁴⁹ Pattern warehouse

⁵⁰ PWMS

متغیری و چند متغیری است، و روش های آموزش خودکار (درخت های تصمیم، طبقه بندی نیو بیز^{۵۷}) استفاده شده است و نهایتا عملکرد روش های آماری و آموزش خودکار با یکدیگر مقایسه شده اند. نتایج نشان می دهد که روش طبقه بندی نیو بیز پیش بینی بهتری از تحلیل رگرسیون منطقی و روش های درخت تصمیم دارد. اما از آنجایی که هدف از ارائه این مقاله ارائه یک مدل است، در زیر بخش های ۲-۱ و ۲-۲ فقط به بررسی مدل های ارائه شده پرداخته شده است.

همان طور که ذکر شد، پایگاه داده تحلیلی در سه سطح مفهومی، منطقی و فیزیکی طراحی می شود. با توجه به اینکه مرز مشخصی بین طراحی در سطح فیزیکی و پیاده سازی پایگاه داده تحلیلی وجود ندارد، در زیر بخش های بعدی، مدل های ارائه شده در سطوح مفهومی و منطقی شرح داده می شوند.

۲-۱- مدل های مفهومی

مدل ها در سطح مفهومی به چهار گروه تقسیم می شوند که عبارتند از:

۱. مدل های رابطه موجودیت گسترش یافته: که شامل مدل رابطه موجودیت چند بعدی^{۵۸} [۲۴]، مدل رابطه موجودیت چند بعدی گسترش یافته^{۵۹} [۲۵-۲۷]، مدل رابطه موجودیت ستاره ای^{۶۰} [۳]، مدل رابطه موجودیت ساختار یافته^{۶۱} [۲۸] و مدل داده چند بعدی مفهومی تصمیم یافته^{۶۲} [۴] است.

۲. مدل های شی گرا: که شامل یام دو^{۶۳} [۲۹ و ۳۰]، مدل پرو فایل زبان یک پارچه^{۶۴} [۳۱ و ۳۲]، مدل چند بعدی شی-گرای^{۶۵} [۳۳]، گلد^{۶۶} [۳۴] و مدل داده چند بعدی شی گرای مبتنی بر گراف^{۶۷} [۳۵] است.

۳. مدل های بر مبنای آنتولوژی: در این مدل ها از آنتولوژی برای کمک به طراحی استفاده می شود [۳۶ و ۳۷].

۴. مدل های اد ها ک^{۶۸}: شامل مدل داده مکعب تجمع چند بعدی^{۶۹} [۳۸] و مدل حقیقت بعدی^{۷۰} [۳۹ و ۴۰] است.

اثربخشی^{۵۶} با ترکیب های متنوع از متریک ها ارزیابی می شوند. از رگرسیون خطی چند گانه برای تحلیل و پیش بینی کیفیت مدل های چند بعدی استفاده می شود.

در [۲۱] متدولوژی برای طراحی یک شمای چند بعدی و کشف دانش مفید و مورد علاقه به فرم قوانین رابطه ای ارائه می شود. شمای چند بعدی با فیلتر کردن اطلاعات ابعاد و حقیقت از یک مجموعه اولیه از متغیر های حقیقت و ابعاد، طراحی شده است. این فیلترینگ روی استفاده از اینتر رویو PCA پایه ریزی شده است.

متدولوژی پیشنهاد شده در این مقاله سه هدف اصلی دارد: اولاً، برای طراحی و ساخت شمای چند بعدی به طور اتوماتیک و داده گرا تهیه شده است. دوماً، منجر به امکان تولید اطلاعات مکعب در سطوح مختلف از داده انتزاعی و شناخت تاثیر سطح داده انتزاعی روی اطلاعات می شود. سوماً، امکان کشف متغیر ها و پیش بینی قوانین رابطه ای را از ساختار مکعب چند بعدی در سطوح مختلف از داده انتزاعی می دهد.

در [۲۲] یک روش جدید ارزیابی برای پیدا کردن اتوماتیک تطبیق های پیچیده بین شمای از مخزن های داده پیشنهاد شده است. بر اساس ادعای نویسندگان، این اولین روشی است که توانایی کشف تطبیق های شمای پیچیده تنها با استفاده از نمونه های داده را دارد. تنها از داده ذخیره شده در مخزن ها برای این کار استفاده می شود. این مقاله بر استراتژی هایی تطبیقی که روی رکورد دوباره تکثیر شده (استراتژی مبتنی بر موجودیت) و تکنیک های اطلاعات باز یابی شده (استراتژی مبتنی بر ارزش) برای پیدا کردن تطبیق های شمای پیچیده در طول فرایند ارزیابی، پایه ریزی شده است. نتایج آزمایشات انجام شده نشان می دهد که این روش، توانایی پیدا کردن تطبیق های پیچیده با دقت بالا را دارد.

در [۲۳] تمرکز روی پیش بینی تاثیر متریک های ساختاری روی فهم شمای مفهومی است. به این منظور از روش های آماری (تحلیل رگرسیون منطقی) که شامل تحلیل یک

⁶⁴ UML profile

⁶⁵ Object Oriented Multi Dimensional model(OOMD)

⁶⁶ GOLD

⁶⁷ Graph Object Oriented Multidimensional Data Model (GOOMD)

⁶⁸ ad-hoc

⁶⁹ Multidimensional Aggregation Cube data model(MA

⁷⁰ Dimensional Fact Model(DFM)

⁵⁶ efficiency

⁵⁷ Naive Bayesian classifier

⁵⁸ Multidimensional Entity Relationship(MER)

⁵⁹ MultiDimER

⁶⁰ Star Entity Relationship(StarER)

⁶¹ Structured Entity Relationship Model(SERM)

⁶² Generalising Conceptual Multidimensional Data model(CGMD)

⁶³ Yet Another Multidimensional Model(YAM³)

۱۳. سلسله مراتب تعمیم یافته/خصوصی شده: بدان معنی که مدل، روابط تعمیم یافته/خصوصی شده را داشته باشد.

۱۴. مدیریت تغییرات در طول زمان: اگرچه تغییرات کسب و کار در مدل پایگاه داده تحلیلی منعکس می شود، با توجه به این خصوصیت، یک مدل باید این امکان را بدهد که داده در طول زمان مقایسه شود.

۱۵. مدیریت عدم قطعیت: این خصوصیت، احتمالات اضافی که وجود دارد را مدیریت می کند.

۱۶. شمایابی با چندین معیار: بدان معنی که مدل، توانایی پذیرش شمایابی با چندین معیار را داشته باشد.

۱۷. خلاصه سازی: بدان معنی که مدل، قابلیت خلاصه سازی را داشته باشد.

۱۸. توابع تجمیع تعریف شده از طریق کاربر: بدان معنی که مدل، پذیرش توابع تجمیع تعریف شده از طریق کاربر را داشته باشد.

۱۹. اشتراک گذاری ابعاد: بدان معنی که مدل، اشتراک گذاری ابعاد را امکان پذیر کند.

۲۰. تجمیع با کاهش ابعاد: بدان معنی که مدل، قابلیت تجمیع از طریق کاهش ابعاد را داشته باشد.

۲۱. تجمیع با کاهش معیارها: بدان معنی که مدل، قابلیت تجمیع از طریق کاهش معیارها را داشته باشد.

۲۲. تجمیع سازی از تجمیع های قبلی: بدان معنی که مدل، قابلیت تجمیع از طریق تجمیع های قبلی را داشته باشد.

هر یک از مدل های ارائه شده در این سطح، بعضی از خصوصیات فوق را دارا می باشند. به عنوان مثال، مدل رابطه موجودیت ستاره ای [۳] خصوصیات ۲، ۳، ۵، ۶، ۸ و ۱۲ را دارا است. مدل داده چند بعدی مفهومی تعمیم یافته [۴] به جز ۴، ۱۲ و ۱۴ بقیه خصوصیات را دارا است. مدل داده چند بعدی شی گرای مبتنی بر گراف [۳۵] خصوصیات ۱، ۲، ۴، ۷ و ۱۲ را دارا می باشد. مدل داده مکعب تجمیع چند بعدی [۳۸] بطور کامل خصوصیات ۲، ۳، ۵، ۶، ۸، ۹، ۱۰، ۱۱ و ۱۲ را دارا است. مدل حقیقت بعدی [۳۹ و ۴۰] به طور کامل خصوصیات ۱، ۲، ۳، ۵، ۶، ۸ را دارا است.

۲-۲- مدل های منطقی

بطور کلی، مدل ها در این سطح می توانند دارای خصوصیتی باشند که در زیر شرح داده شده اند [۳ و ۴]:

۱. استقلال در پیاده سازی: یک مدل باید در پیاده سازی استقلال داشته باشد. مدل های رابطه ای و مدل های بر پایه مدل ستاره ای این خصوصیت را بررسی نمی کنند. مدل های رابطه ای که برای دامنه ی خاص پیاده سازی شده اند این خصوصیت را ندارند.

۲. جداسازی صریح ساختار و محتوا: این خصوصیت مکعب چند بعدی، ابعاد و ارزش های هر سلول را جداگانه تعریف می کند.

۳. سلسله مراتب صریح: سلسله مراتب در ابعاد باید به طور صریح نشان داده شوند تا کاربران بتوانند به سطوح مختلف دسترسی یابند.

۴. رفتار متقارن برای ابعاد و معیارها: بدان معنی که مدل این امکان را بدهد که معیارها به عنوان ابعاد و برعکس ابعاد به عنوان معیارها رفتار کنند.

۵. سلسله مراتب چندتایی در ابعاد: گاهی برای یک بعد می توان دو مسیر جداگانه ایجاد کرد.

۶. ویژگی های ابعاد: بدان معنی که مدل، ویژگی ها را از سلسله مراتب تمیز دهد.

۷. پشتیبانی از تجمیع: بدان معنی که مدل، توانایی برای ایجاد تجمیع های معنی دار داشته باشد.

۸. معیارهای پیچیده: بدان معنی که مدل، معیارهای پیچیده و چندتایی برای حقیقت را دارا باشد.

۹. مدیریت سطوح مختلف دانه دانه گی: معیارها در سطوح مختلف تجمیع شده مدیریت می شوند.

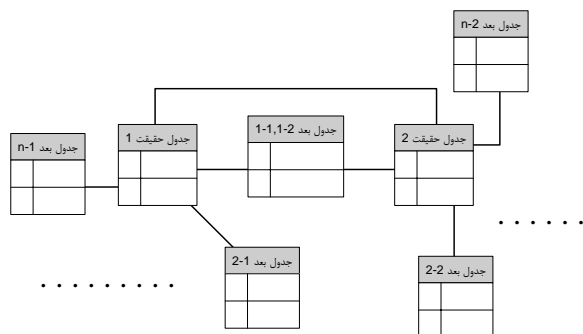
۱۰. پشتیبانی از سلسله مراتب نامتوازن: بدان معنی که مدل، سلسله مراتب با مسیرهایی در طول های مختلف داشته باشد.

۱۱. پشتیبانی از سلسله مراتب ضعیف: سلسله مراتبی هستند که حد^{۷۱} (حداقل و حداکثر) بین سطوح را نشان می دهند که حد پایین و حد بالای ارتباط داخل پراتز نوشته می شود.

۱۲. پشتیبانی از رابطه چند به چند: بدان معنی که مدل، رابطه چند به چند بین حقیقت ها و ابعاد و همچنین رابطه ای بین دو سطح سلسله مراتب را دارا می باشد.

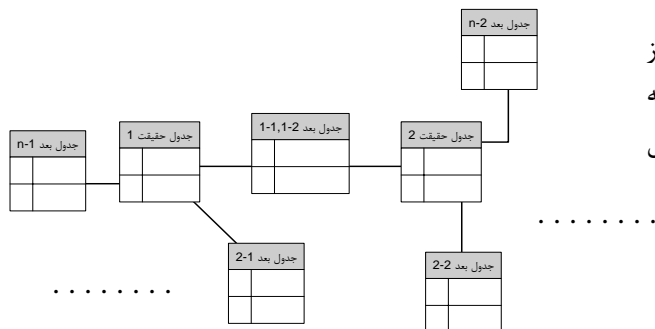
⁷¹cardinality

صورت مشترک بین جداول حقیقت باشد. همچنین بین دو جدول حقیقت ۱ و ۲ ارتباط وجود دارد.



شکل ۲: مدل فلکی

مدل کهکشانی [۵ و ۶]: این مدل یک مجموعه از مدل‌های ستاره‌ای است که دارای ابعاد مشترک می‌باشند. برعکس مدل فلکی، جداول حقیقت در مدل کهکشانی نیاز به ارتباط مستقیم ندارند (شکل ۳). شکل (۳) مانند شکل (۲) است با این تفاوت که ارتباطی بین جداول حقیقت وجود ندارد.



شکل ۳: مدل کهکشانی

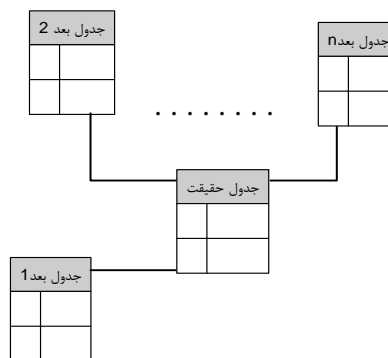
مدل دانه برفی [۵-۷]: این مدل نتیجه‌ای از تجزیه یک یا چند بعد در مدل ستاره‌ای است (شکل ۴). مدل دانه برفی دارای ساختار سلسله‌مراتبی از ابعاد است. مدل دانه برفی تکنیک کلی برای ذخیره‌سازی اطلاعات چند ارزی در پایگاه داده تحلیلی است که نرمال‌سازی در جداول بعد را به دنبال دارد و مدل به دلیل جداول نرمال دارای پیچیدگی بالا می‌باشد. طبق شکل (۴) مدل از یک جدول حقیقت و تعداد جداول بعد ۱ تا n ایجاد شده‌است که هر کدام از جداول بعد آن می‌تواند شامل جداول دیگر به صورت سلسله‌مراتبی باشد. برای جدول بعد ۱ جدول بعد ۱-۱ و برای جدول بعد ۲ جدول بعد ۱-۲ و جدول بعد ۲-۱ و ۱-۲ مشخص شده‌است.

شناخته شده‌ترین مدل‌ها در سطح منطقی عبارتند از: مدل تخت [۵ و ۶]، مدل مسطح [۵ و ۶]، مدل ستاره‌ای [۵-۷]، مدل فلکی [۵ و ۶]، مدل کهکشانی [۵ و ۶]، مدل دانه برفی [۵-۷]، مدل دانه ستاره‌ای [۵ و ۸] و مدل خوشه ستاره‌ای [۵-۷].

به دلیل تمرکز اصلی این مقاله بر روی سطح منطقی، مدل‌های این سطح همراه با شکل گرافیکی آن‌ها در ادامه ارائه می‌شود.

مدل ستاره‌ای [۵-۷]: ساختار اصلی یک مدل چندبعدی است که براساس یک جدول مرکزی بزرگ (جدول حقیقت) و یک مجموعه از جداول کوچک‌تر (ابعاد) سازمان‌دهی شده‌است (شکل ۱). رکوردها در دو نوع از جداول ذخیره شده‌اند: یک جدول حقیقت و تعدادی جداول بعد. جدول حقیقت کلیدهایی از هر جدول بعد را نگهداری می‌کند. روابط بین هر جدول بعد و جدول حقیقت یک رابطه یک به چند است که شامل یک کلید منحصر بفرد در جدول حقیقت است. جدول حقیقت به غیر از کلیدهای خارجی شامل معیارها نیز می‌باشد. مدل ستاره‌ای به دلیل جداول بعد دینرمال دارای افزونگی بالا می‌باشد.

طبق شکل (۱) مدل از یک جدول حقیقت که در مرکز قرار می‌گیرد و جداول بعدی که جدول حقیقت را احاطه می‌کنند و تعداد آن‌ها می‌تواند از ۱ تا n باشد تشکیل شده‌است.

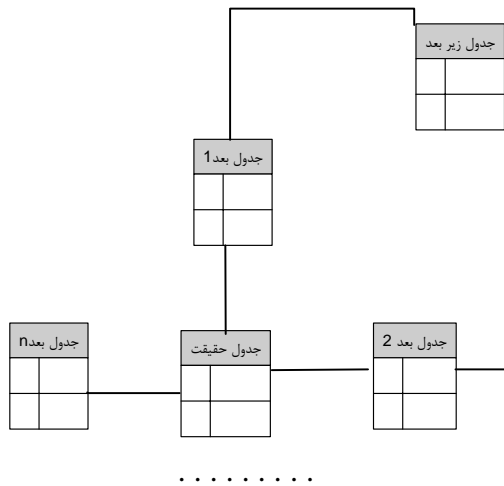


شکل ۱: مدل ستاره‌ای

مدل فلکی [۵ و ۶]: این مدل شامل یک مجموعه از مدل‌های ستاره‌ای است و جداول حقیقت با هم ارتباط دارند. این ارتباطات توانایی drill down بین سطوحی از جزئیات را فراهم می‌کند (شکل ۲).

طبق شکل (۲) دو جدول حقیقت با شماره‌های ۱ و ۲ مشخص شده‌است و تعداد جداول بعد هر جدول حقیقت می‌تواند از ۱ تا n باشد و یک یا چند جدول بعد می‌تواند به

می‌شود که در سطح نرمال‌سازی جدولی که در موقعیت یک پدر و دو فرزند قرار دارد، یعنی در دو بعد مشترک است، تبدیل به یک زیر بعد شود و با این عمل باعث کاهش پیچیدگی جداول می‌شود (شکل ۶). طبق شکل (۶) مدل از یک جدول حقیقت و تعداد جداول بعد از ۱ تا n تشکیل شده‌است و دو جدول بعد می‌توانند با جدول زیر بعدی مشترک در ارتباط باشند. به عنوان مثال جداول بعد ۱ و ۲ با جدول زیر بعدی بطور مشترک در ارتباطند.

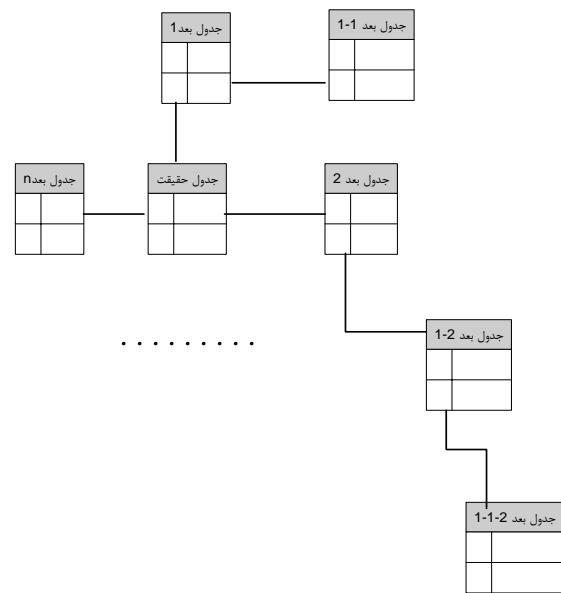


شکل ۶: مدل خوشه‌ستاره‌ای

بنابراین باتوجه به شرح مدل‌های منطقی ذکر شده، این مدل‌ها را می‌توان بر اساس هفت خصوصیت، ارزیابی و مقایسه نمود. این خصوصیات از مطالعات انجام شده بر روی مدل‌های مختلف استخراج شده‌اند [۵]، [۹ و ۱۰] و عبارتند از: ۱. کارایی، ۲. قابلیت استفاده، ۳. قابلیت استفاده مجدد، ۴. انعطاف‌پذیری، ۵. افزونگی، ۶. پیچیدگی و ۷. اعتبار.

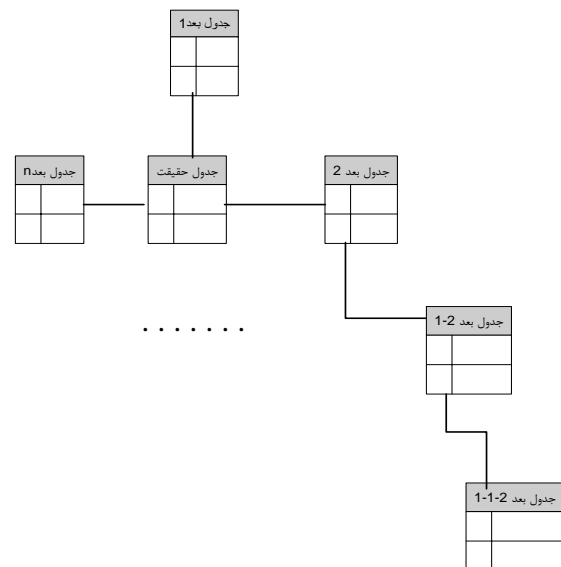
هر یک از مدل‌های ارائه شده در این سطح، بعضی از خصوصیات فوق را در سطح بالا، بعضی دیگر را در سطح متوسط و در نهایت بعضی‌ها را در سطح پایین دارا می‌باشند. مدل ستاره‌ای خصوصیات ۱، ۲، ۴ و ۵ را در سطح بالایی دارا است. مدل‌های فلکی و کهکشانی همانند مدل ستاره‌ای می‌باشند ولی پیچیدگی آن‌ها بیشتر می‌باشد. مدل دانه‌ستاره‌ای و همچنین خوشه ستاره‌ای به طور متوسط همه خصوصیات را دارا هستند. مدل دانه‌برفی خصوصیات ۳، ۶ و ۷ را در سطح بالایی دارا است.

بنابراین مدل‌های فلکی و کهکشانی علاوه بر افزونگی بالایی مدل ستاره‌ای، دارای پیچیدگی نیز می‌باشند. از طرفی مدل‌های دانه‌ستاره‌ای و خوشه‌ستاره‌ای برای تعادلی بین افزونگی مدل ستاره‌ای و پیچیدگی مدل دانه‌برفی ایجاد



شکل ۴: مدل دانه‌برفی

مدل دانه‌ستاره‌ای [۵ و ۸]: این مدل دارای ساختاری ترکیبی است و در واقع شامل ترکیبی از مدل ستاره‌ای و دانه‌برفی است. این مدل ترکیبی از نرمال‌سازی و غیر نرمال‌سازی است (شکل ۵). طبق شکل ۵ مدل از یک جدول حقیقت و تعداد جداول بعد ۱ تا n تشکیل شده است. طبق تعریف مدل، بعضی از جداول بعد به صورت سلسله‌مراتبی هستند. به طور مثال جدول بعد ۲ شامل جدول بعد ۱-۲ و جدول بعد ۱-۱-۲ می‌باشد.



شکل ۵: مدل دانه‌ستاره‌ای

مدل خوشه‌ستاره‌ای [۷-۵]: این مدل، بهترین راه حل ممکن برای تعادل بین افزونگی و پیچیدگی در دو مدل ستاره‌ای و دانه‌برفی است. در واقع این مدل همان مدل ستاره‌ای با تعریف دانه‌برفی است. این مدل زمانی ایجاد

می‌گیرد. سپس لایه دوم (بالا) که لایه مدل ستاره‌ای است، ایجاد می‌شود. بنابراین از مزایای این مدل که مهمترین آن‌ها سرعت پاسخ به پرس‌وجوها است بهره‌مند می‌شود. در واقع در مدل پیشنهادی (مدل دولایه‌ای)، با افزودن دیدهای ذخیره‌شده مناسب جهت پاسخ‌گویی به پرس‌وجوها در لایه ستاره‌ای، نرخ این افزایش زمان در لایه دانه‌برفی نیز به میزان قابل توجهی کاهش می‌یابد.

برای ساخت لایه بالایی یا همان ستاره‌ای باید از جداول لایه پایینی استفاده شود و جداول بعد باید به صورت غیر نرمال باشند. برای رسیدن از لایه دانه‌برفی به لایه ستاره‌ای از الگوی قطعه‌بندی عمودی ابعاد استفاده می‌شود [۴۲ و ۴۳]. به این منظور، ابتدا لازم است سطوحی از هر بعد که باید با هم ذخیره شوند مشخص شوند. به هر مجموعه از سطوح، قطعه گفته می‌شود. وقتی یک قطعه معتبر است که سطوحش به صورت سلسله‌مراتبی به هم مربوط باشند. بنابراین یک قطعه‌بندی زمانی کامل است که همه سطوح ابعاد آن، برای اینکه اطلاعات از بین نرود، حداقل به یک قطعه متعلق باشند. قطعه‌بندی عمودی ابعاد، زمان پاسخ به پرس‌وجوها را بهبود می‌دهد، اما (در لایه ستاره‌ای) به دلیل اینکه جداول بعد از موقعیت سلسله‌مراتبی (نرمال) به حالت دینرمال تبدیل می‌شوند، افزونگی افزایش می‌یابد [۴۲ و ۴۳].

بعد از قطعه‌بندی عمودی ابعاد، جداول لایه بالایی یا همان مدل ستاره‌ای ایجاد می‌شود. اما نکته مورد توجه در ارتباط با مدل پیشنهادی این است که وقتی از لایه دانه‌برفی به لایه ستاره‌ای می‌خواهد برسد، براساس نیاز مدیران پایگاه داده تحلیلی لایه ستاره‌ای ساخته می‌شود، بنابراین مدل قابلیت انعطاف‌پذیری بالایی دارد. به این صورت که جدول یا جداول بعد لایه مدل دانه‌برفی که به صورت نرمال است براساس نیاز مدیران در لایه ستاره‌ای تغییر پیدا خواهند کرد و جداول بعد دینرمال ایجاد می‌شود. در واقع صفات تمام جداول بعد در هر دو لایه وجود دارد، اما نحوه قرارگرفتن آن‌ها در جداول متفاوت است به عنوان مثال در لایه پایینی (مدل دانه‌برفی) تعداد جداول بعد آن نسبت به لایه بالایی (مدل ستاره‌ای) به دلیل نرمال‌سازی بیشتر است اما صفات آن‌ها یکسان است.

مثلا در مورد بعد خرده فروش که شامل سلسله مراتب بعد مکان جغرافیایی و بعد قلمرو فروش در لایه دانه‌برفی می‌باشد، بر اساس نیاز مدیر پایگاه داده تحلیلی بعد از

شده‌اند اما نسبت به مدل ستاره‌ای دارای پیچیدگی بیشتری می‌باشند. بنابراین با توجه به عیب مدل‌های موجود، پیچیدگی بالای مدل دانه‌برفی و افزونگی بالای مدل ستاره‌ای در بخش بعدی مدلی ارائه می‌شود که این عیب‌ها (افزونگی و پیچیدگی بالا) را تا حدی برطرف می‌کند و از مزایای مدل‌های موجود (کارایی، انعطاف‌پذیری و اعتبار) نیز استفاده می‌کند.

۳- معرفی مدل پیشنهادی

هدف از مدل‌سازی بعدی ایجاد ساختارهای پایگاه داده‌ای است که استفاده از آن‌ها برای کاربران نهایی آسان باشد. منظور از استفاده، نوشتن پرس‌وجوها و پاسخ‌گویی سریع به آن‌هاست. این هدف با حداقل کردن شماری از جداول و روابط بین آن‌ها بدست می‌آید (دینرمال سازی) [۶ و ۴۱]. بنابراین طبق اهداف مدل‌سازی بعدی، مدل ستاره‌ای دارای حداقل جداول و روابط می‌باشد و از آن برای ایجاد مدل پیشنهادی استفاده می‌شود.

در ادامه ابتدا مزایا و معایب مدل ستاره‌ای بررسی می‌شوند. مدل ستاره‌ای، ساده‌ترین ساختار است و شماری از جداول و روابط بین آن‌ها کاهش و در نتیجه شماری از اتصالات مورد نیاز در پرس و جوهای کاربر را کاهش می‌دهد. در واقع پیچیدگی پایینی دارد و به عملکرد پرس و جو سرعت می‌بخشد و از طرفی در رابطه با عیب آن می‌توان به افزونگی داده در بالاترین سطح اشاره کرد [۲ و ۵]. از طرفی مدل دانه‌برفی شکل نرمال شده مدل ستاره‌ای است که افزونگی داده را حداقل می‌کند. مدل دانه‌برفی ساختار سلسله‌مراتبی از هر بعد را به طور صریح نشان می‌دهد و درک و فهم آن آسان است؛ همچنین دارای قابلیت استفاده-مجدد می‌باشد. عیب این مدل پیچیدگی بالای آن است که منجر به افزایش زمان پاسخ به پرس‌وجوها می‌شود [۲ و ۵]. مدل پیشنهادی در سطح منطقی و براساس مدل‌های ستاره‌ای و دانه‌برفی طراحی می‌شود. پیشنهاد مطرح در این مقاله را می‌توان به صورت زیر بیان کرد:

مدل دولایه‌ای است. با توجه به مزایای مدل دانه‌برفی که شامل قابلیت استفاده مجدد و درک و فهم سلسله مراتب ابعاد می‌باشد، در لایه اول (پایین) لایه دانه‌برفی قرار می‌گیرد. جداول در این لایه کاملا نرمال و بدون هیچ افزونگی می‌باشند و تمام مراحل فرایند استخراج، تغییر شکل و بارگزاری (ای.تی.ال) در لایه دانه‌برفی انجام

و از سلسله مراتب مدل دانه‌برفی در مواقعی که نیاز به این جداول است استفاده می‌کند و این امر، اعتبار پاسخ به پرس‌وجوها را بالا می‌برد زیرا افزونگی حداقل شده‌است و از داده‌های تکراری جلوگیری می‌شود.

شکل گرافیکی مدل پیشنهادی:

نمایش گرافیکی مدل پیشنهادی در دو لایه است. در واقع لایه اول (پایین) مدل دانه‌برفی و لایه دوم (بالا) مدل ستاره-ای قرار دارد (شکل ۷).

همچنین نمایش گرافیکی مدل پیشنهادی در دو لایه دانه-برفی و ستاره‌ای در شکل (۸) بر روی پایگاه داده تحلیلی سیستم فروش ترسیم شده است. طبق شکل (۸)، جداول سلسله مراتب بعد خرده فروش لایه پایینی (لایه دانه‌برفی) طبق الگوی قطعه‌بندی عمودی ابعاد به جدول دینرمال براساس بعد قلمرو فروش در لایه بالا تبدیل می‌شود و همچنین جدول حقیقت لایه بالایی (لایه ستاره‌ای) براساس جدول حقیقت لایه پایینی (لایه دانه‌برفی)، بعد جدید قلمرو فروش و معیارهای جدید تعریف می‌شود. در واقع لایه بالا (لایه ستاره‌ای)، هم جدول بعد و هم جدول حقیقت آن به صورت materialized view پیاده‌سازی می‌شود.

۴- ارزیابی مدل پیشنهادی

یکی از مسائلی که به طراحی پایگاه داده تحلیلی مربوط می‌شود، انتخاب مناسب مدل می‌باشد. انتخاب مدل می‌تواند براساس نظر شخصی و نیازهای کسب‌وکار باشد. همچنین ابزار استفاده شده نیز در انتخاب مدل تاثیر دارند، مثل اوراکل و ام اس اس کیو ال^{۷۳} برای مدل ستاره‌ای [۷] و [۴۴] و دی بی دو^{۷۴} برای مدل دانه‌برفی [۴۴] مناسب می‌باشد و محیط هم یکی دیگر از معیارهای انتخاب مدل می‌باشد. این‌ها لازم هستند ولی کافی نیستند. معیارهای دیگری برای انتخاب مدل‌های پایگاه داده تحلیلی وجود دارد که می‌توان به نوع صفت، نوع جداول بعد، وجود شاخص و نوع پرس‌وجو اشاره نمود [۴۴].

زمان پاسخ به پرس‌وجوها همیشه مهمترین معیار برای مقایسه مدل‌ها بوده‌است [۴۴-۴۶]. در این مقاله از این معیار برای ارزیابی مدل‌های دانه‌برفی، ستاره‌ای و مدل پیشنهادی با نام مدل دولایه‌ای^{۷۵} استفاده می‌شود. در نتیجه مدل‌ها در محیط اس کیو ال بر روی پایگاه داده

قطعه‌بندی ابعاد، جدول بعد جدیدی با کلید اصلی قلمرو فروش در لایه ستاره‌ای ایجاد می‌شود. این روش را می‌توان براساس نیاز مدیران پایگاه داده تحلیلی فقط روی یک بعد یا تمام ابعاد پیاده کرد. همچنین باید جدول حقیقت لایه ستاره‌ای با توجه به تغییرات ایجاد شده ساخته شوند. به این منظور از جدول حقیقت لایه دانه برفی و یک بعد یا تمام ابعاد جدید استفاده می‌شود. در واقع لایه بالایی (مدل ستاره‌ای) لایه materialized view از ابعاد و جدول حقیقت لایه پایین می‌باشد.

مزایای مدل پیشنهادی (مدل دولایه‌ای):

در مدل پیشنهادی، فرایند استخراج، تغییر شکل و بارگزاری دو فاز دارد. فاز اول در لایه پایینی به طور کامل انجام می‌گیرد. فاز دوم در لایه بالایی صورت می‌گیرد و به دلیل اینکه نیاز به استخراج و تغییر شکل^{۷۲} نمی‌باشد، بعد از تولید جداول لایه بالایی (که از جداول لایه پایینی به صورت تجمیع‌شده (دینرمال) بدست می‌آیند) انجام می‌پذیرد. بنابراین سرعت این فرایند در لایه بالایی افزایش پیدا می‌کند.

این مدل به دلیل وجود جداول نرمال شده در لایه پایینی، امکان بروزرسانی و اعمال تغییرات در ساختار لایه ستاره-ای را بسیار ساده‌تر از روش‌های موجود، فراهم می‌کند. به این ترتیب، امکان افزودن بخشی از پرس‌وجوهای دلخواه به پایگاه داده تحلیلی میسر می‌شود.

مزیت دیگر، فرکانس بروزرسانی مدل دولایه‌ای است که می‌تواند متفاوت باشد و بسته به نیاز تغییر کند. به گونه‌ای که اگر تغییرات در منابع پایگاه داده به نحوی باشد که مدیر پایگاه داده تحلیلی تشخیص بدهد آن‌ها را در زمان‌های خاص به پایگاه داده تحلیلی اعمال کند، بروزرسانی آنسکرون یا غیرهمزمان برای دو لایه انجام می‌شود و انعطاف‌پذیری سیستم بالا می‌رود.

مزیت دیگر مدل این است که کاملاً نسبی است، یعنی زمان پرس‌وجو، متناسب با مدل ستاره‌ای است که از پایین‌ترین زمان پاسخگویی برخوردار است و از طرفی مزایای مدل دانه‌برفی را نیز دارد. این مدل از کارایی بالای مدل ستاره‌ای که عبارتند از پیچیدگی کمتر و سرعت پاسخگویی بالاتر؛ به دلیل جداول کمتر، و اتصالات کمتر، برخوردار می‌باشد؛

⁷⁴DB2

⁷⁵Two-Layer schema

⁷²transformation

⁷³MS SQL

تحلیلی که به پایگاه داده تحلیلی اعمال می‌شوند از نوع aggregate بوده و بسته به کاربرد مورد نظر می‌توانند بر روی ابعاد مختلف اعمال شوند و در مواردی نیز می‌توانند با شرط where همراه باشند. از آنجایی که احتمال ورود این پرس‌وجوها در شرایط مختلف متفاوت بوده و بستگی به اولین پرس‌وجوی اعمالی نیز دارد، در این مقاله ارزیابی‌ها برای دو نوع گروه‌بندی شده و گروه‌بندی نشده‌ای که با where همراه است بطور مجزا صورت گرفته است.

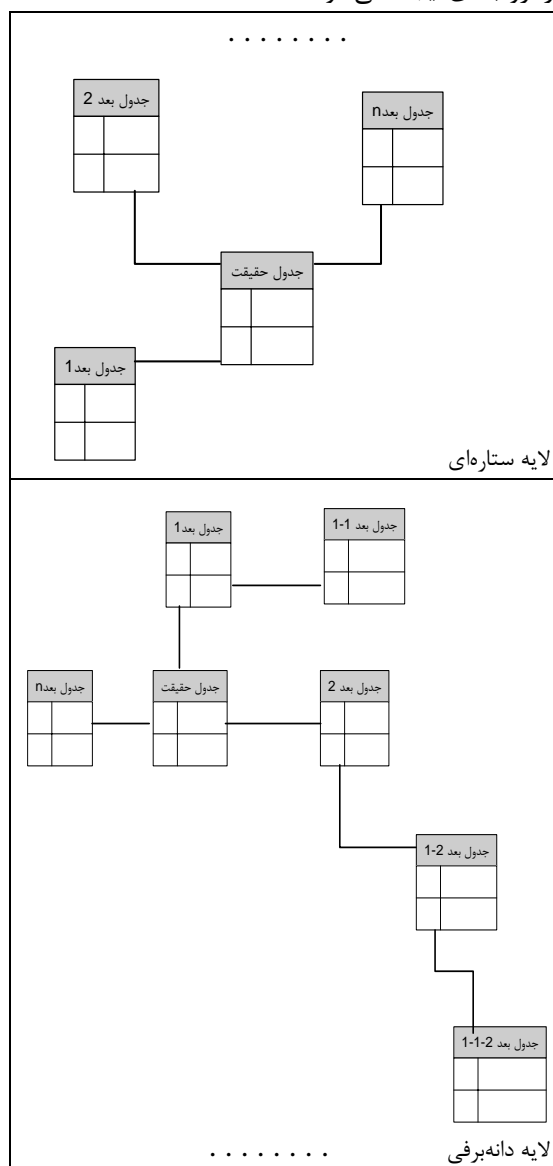
بنابراین پرس‌وجوهای یک تا چهار مربوط به گروه‌بندی‌های مختلف می‌باشند. همچنین پرس‌وجوهایی از نوع slice که یک ارزش از یک بعد و dice که رنجی از ارزش‌های یک یا چند بعد را انتخاب می‌کنند، به عنوان پرس‌وجوهای پنج و شش ایجاد می‌شوند. شکل (۹)، این شش پرس‌وجو را برای انجام آزمایشات بر روی پایگاه داده تحلیلی دولایه‌ای فروش به همراه مدل لایه دوم (لایه ستاره‌ای) نمایش می‌دهد. در این بخش فقط پرس‌وجوها در مدل دولایه‌ای ارائه شده‌اند. پرس‌وجوهای مطرح در مدل دولایه‌ای براساس سه بعد قلمرو فروش، محصول و زمان در گروه‌بندی‌های مختلف ایجاد شده‌اند.

به عنوان مثال لایه دوم (مدل ستاره‌ای) شامل سه بعد (محصول، زمان و قلمرو فروش) می‌باشد. بنابراین ۸ پرس‌وجو در گروه‌بندی‌های مختلف ایجاد می‌شود، به این صورت که یا فقط براساس یک بعد ((محصول)، (زمان) و (قلمرو فروش)) یا براساس دو بعد ((محصول، زمان)، (محصول، قلمرو فروش) و (زمان، قلمرو فروش)) و سه بعد ((قلمرو فروش، محصول، زمان)) است.

همچنین پرس‌وجوهایی از نوع slice که یک ارزش از یک بعد و dice که رنجی از ارزش‌های یک یا چند بعد را انتخاب می‌کنند، می‌توانند پرس‌وجوهای مختلف ایجاد کنند. بنابراین نحوه نوشتن پرس‌وجوها در این مدل تغییر داشته‌اند.

پرس‌وجوی یک بر اساس جدول قلمرو فروش و معیار جمع کل تخفیفات گروه‌بندی شده‌اند. پرس‌وجوی دو بر اساس دو بعد قلمرو فروش، زمان و معیار جمع کل فروش گروه‌بندی شده‌اند. پرس‌وجوی سه بر اساس دو بعد قلمرو فروش، محصول و معیار جمع کل فروش گروه‌بندی شده‌اند. پرس‌وجوی چهار بر اساس سه بعد قلمرو فروش، محصول، زمان و معیار جمع کل تخفیفات گروه‌بندی شده‌اند.

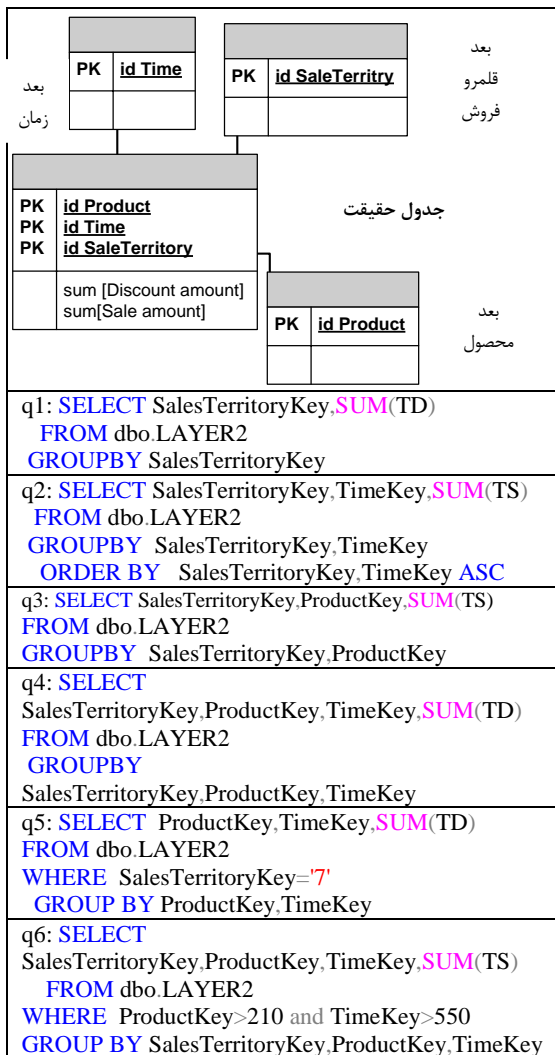
فروش پیاده‌سازی شده‌اند و معیار زمان پاسخگویی مربوط به شش نوع پرس‌وجوی مختلف که در ادامه توضیح داده می‌شوند، در نمودارها با یکدیگر مقایسه شده‌اند. نوع پرس‌وجو معیار مناسب برای انتخاب مدل‌ها می‌باشد. پرس‌وجوها براساس تعداد بعدی که جدول حقیقت را احاطه کرده‌اند بدست می‌آیند. بنابراین پرس‌وجوهایی براساس گروه‌بندی تولید می‌شوند، یعنی اگر k بعد وجود داشته باشد 2^k پرس‌وجو براساس زیرمجموعه‌ای از ابعاد در گروه‌بندی ایجاد می‌شود.



شکل ۷: مدل پیشنهادی، مدل دولایه‌ای

برای ارزیابی مدل پیشنهادی از معیار زمان پاسخ به پرس‌وجوها استفاده شده‌است و این ارزیابی برای انواع مختلف پرس‌وجوها صورت گرفته‌است. پارامترهای ارزیابی مطرح در مدل پیشنهادی، نوع پرس‌وجوها است. انواع پرس‌وجوهای

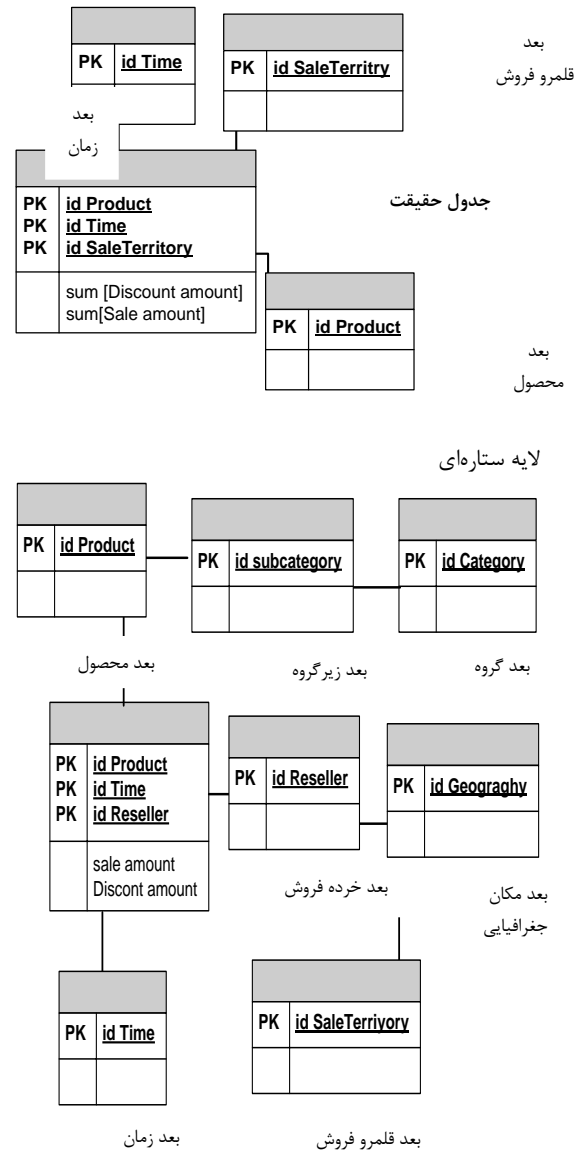
FROM dbo.LAYER2
GROUPBY SalesTerritoryKey



شکل ۹: شش نوع پرس و جو در مدل دولایه‌ای به همراه شکل گرافیکی لایه ستاره‌ای برای پایگاه داده تحلیلی سیستم فروش

این پرس و جو به لایه دوم وارد می‌شود و مجموع تخفیف‌ها را حساب می‌کند. این پرس و جو برای اینکه بر اساس کد قلمرو فروش گروه‌بندی شود از جدول قلمرو فروش لایه ستاره‌ای استفاده می‌کند. این جدول از الحاق سه جدول به نام بعد خرده فروش، بعد مکان جغرافیایی و بعد قلمرو فروش که به صورت سلسله‌مراتبی در لایه مدل دانه‌برفی قرار دارد، ایجاد شده است. لازم به ذکر است که جدول قلمرو فروش لایه ستاره‌ای به عنوان جدول بعد جدید با کلید اصلی قلمرو فروش در لایه ستاره‌ای به جدول حقیقت آن وصل شده است.

بنابراین برای نگهداری و مدیریت داده‌ها از SQL Server 2008 استفاده شده است. پیاده‌سازی الگوریتم نیز در نرم افزار C# نسخه ۲۰۱۲ انجام شده است.



لایه دانه‌برفی

شکل ۸: مدل دولایه‌ای برای پایگاه داده تحلیلی فروش

پرس و جوی پنج براساس یک ارزش از بعد قلمرو فروش که معادل "۷" و برابر کشور فرانسه است و براساس بعد محصول، زمان و معیار جمع کل تخفیفات گروه‌بندی شده است. پرس و جوی شش بر اساس رنجی از ارزش‌های بعد محصول که برابر با کد محصولات بیشتر از ۲۱۰ و همچنین برای بعد زمان که برابر کد زمان بیشتر از ۵۵۰ یعنی برابر سال‌های ۲۰۰۳ به بعد و براساس سه بعد (قلمرو فروش، محصول و زمان) گروه بندی شده است.

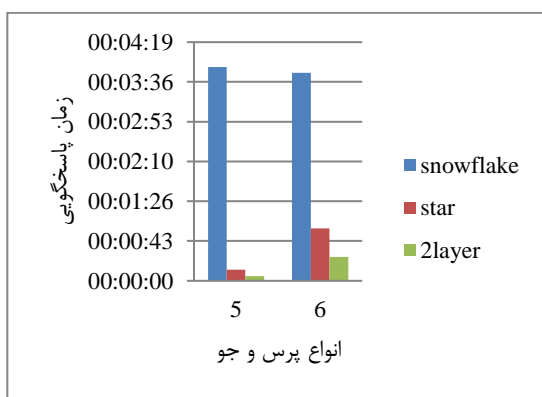
نقش لایه‌ها در هنگام پاسخگویی به پرس و جوها مطرح می‌شود و بستگی به پرس و جویی دارد که مدیر بر اساس نیاز خود بیان می‌کند. به عنوان مثال، پرس و جوی ۱ بررسی می‌شود:

q1: `SELECT SalesTerritoryKey,SUM(TD)`

user آن) اغلب در زمان‌های offline انجام و مدیریت می‌شود.

بنابراین با توجه به پرس‌وجوهای مذکور در ارتباط با بعد تغییر پیدا کرده، زمان پاسخ به پرس‌وجوها در هر سه مدل (دانه‌برفی، ستاره‌ای و دولایه‌ای) اندازه‌گیری می‌شود.

آزمایشات براساس دو نوع پرس‌وجوی slice و dice، در سه مدل دانه‌برفی، ستاره‌ای و دولایه‌ای انجام شده‌است و نتایج در شکل (۱۰) نشان داده شده‌است. طبق شکل (۱۰)، مدل دولایه‌ای سرعت پاسخگویی بالاتری در هر دو پرس‌وجو نسبت به دو مدل ستاره‌ای و دانه‌برفی دارد. همچنین در مدل دانه‌برفی بعد قلمرو فروش در سطوح سلسله‌مراتب آن قرارداد و به دلیل اتصالات بیشتر، پرس‌وجوی پنج زمان بیشتری نسبت به پرس‌وجوی شش می‌کند.



شکل ۱۰: نتایج آزمایشات براساس پرس‌وجوی slice, dice

آزمایشات بعدی براساس چهار نوع پرس‌وجو در گروه‌بندی‌های مختلف در سه مدل دانه‌برفی، ستاره‌ای و دولایه‌ای انجام شده‌است و نتایج در شکل ۱۱ نشان داده شده‌است. بنابراین مدل دولایه‌ای نسبت به مدل ستاره‌ای و دانه‌برفی زمان پاسخ به پرس‌وجوها را بهبود داده‌است. همچنین پرس‌وجوی چهار به دلیل اینکه در گروه‌بندی از تمام سه بعد قلمرو فروش، محصول و زمان استفاده می‌کند، زمان بیشتری برای پاسخگویی نسبت به پرس‌وجوهای دیگر صرف می‌کند.

با توجه به اینکه آزمایشات قبلی بر روی یک پرس‌وجوی خاص انجام گرفته‌است، در آزمایشات بعدی برای هر نوع پرس‌وجو، تعدادی پرس‌وجو نوشته شده و به مدل‌های ستاره‌ای، دانه‌برفی و دولایه‌ای اعمال شده‌است. بنابراین میانگین زمان پاسخگویی در هر سه مدل اندازه‌گیری می‌شود.

سخت‌افزاری که الگوریتم‌ها بر روی آن اجرا شده‌اند، دارای پردازنده 2.80GHz (Pentium® 4 CPU 2.80GHz) با حافظه اصلی 512 مگا بایت می‌باشد. تمامی الگوریتم‌ها تحت سیستم عامل Microsoft windows XP، Version 2002، service pack 3 اجرا شده‌اند.

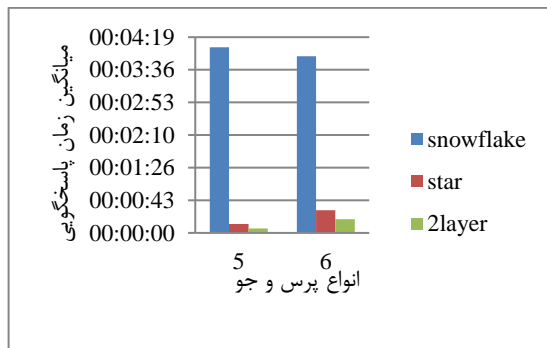
برای ایجاد رکورد با اندازه‌های بزرگ از جعبه ابزار Red Gate SQL Data Generator استفاده شده‌است.

در نتیجه ابتدا باید شرایط اولیه طراحی پایگاه داده تحلیلی فراهم باشد که عبارتند از انتخاب فرایند تجاری که در اینجا سیستم فروش بررسی شده‌است، سپس انتخاب بعدها که زمان، خرده‌فروش و محصول انتخاب شده‌است و نهایتاً انتخاب معیار جدول حقیقت که مجموع تخفیفات یا مجموع فروش را بررسی می‌کند. برای شروع از جداول پایگاه داده Adventure work DW که از Microsoft SQL Server Product sample Database بارگذاری شده، استفاده می‌شود. این پایگاه داده بصورت چندبعدی می‌باشد و دانه آن عبارت است از: "چه محصولی در چه زمانی توسط کدام خرده‌فروش فروخته شده‌است." بر این اساس این پایگاه داده دارای یک جدول حقیقت و سه جدول بعد اصلی همراه با سلسله‌مراتب است.

حجم داده جداول بعد پایگاه داده تحلیلی بررسی می‌شود. تعداد رکوردهای جدول بعد خرده‌فروش ۷۰۱، جدول بعد محصول ۶۰۶ و جدول بعد زمان ۱۱۵۸ است. در طراحی پایگاه داده تحلیلی حجم جدول حقیقت معمولاً بالا است؛ بنابراین جدول حقیقتی با استفاده از بعدهای موجود خرده‌فروش، محصول و زمان ایجاد شده‌است و کلید سه بعد به عنوان کلید اصلی در جدول حقیقت تعریف و دو معیار مقدار فروش و مقدار تخفیف برای آن در نظر گرفته شده‌است. همچنین برای تولید داده با حجم بالا برای جدول حقیقت از نرم افزار Data Generator 2 red gate استفاده شده‌است. در نتیجه جدول حقیقت با ۵۰۰۰۰۰۰ رکود پر شده‌است.

بنابراین با توجه به اینکه پایگاه داده تحلیلی برای پاسخ به پرس‌وجوهای تحلیلی کاربران (select) طراحی و پیاده سازی می‌شود، در این مقاله نیز این نوع پرس‌وجوها که تنها پرس‌وجوهایی هستند که در زمان عملیاتی سیستم به آن اعمال می‌شوند، بررسی شده‌اند. عملیات دیگر شامل درج، حذف و بروزرسانی، از طریق مدیر پایگاه داده تحلیلی (نه

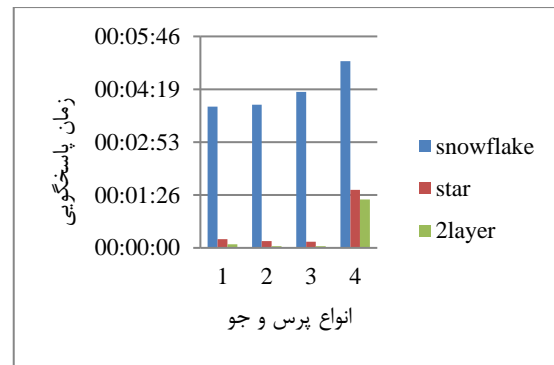
مقایسه نمود. این خصوصیات از مطالعات انجام شده بر روی مدل‌های مختلف استخراج شده‌اند [۵]، [۹ و ۱۰] و عبارتند از: ۱. کارایی، ۲. قابلیت استفاده، ۳. قابلیت استفاده مجدد، ۴. انعطاف پذیری، ۵. افزونگی، ۶. پیچیدگی و ۷. اعتبار.



شکل ۱۳: نتایج آزمایشات برای میانگین زمان پاسخگویی در پرس و جوی slice و dice

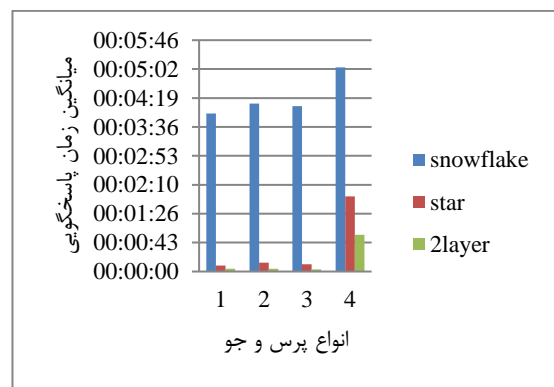
جدول ۱ دارای نه سطر است. در هر سطر مدل‌های منطقی مختلف براساس هفت خصوصیت ذکر شده، مقایسه و ارزیابی شده است. نتایج ارزیابی‌ها با واژه‌های "بالا"، "متوسط" و "پایین" مشخص شده است. منظور از "بالا" یعنی اینکه مدل این خصوصیت را در سطح بالایی دارا است و "متوسط" یعنی به طور متوسط مدل این خصوصیت را دارا است و در نهایت "پایین" یعنی مدل خصوصیت را در حد ضعیفی دارد.

همانطور که در جدول ۱ مشاهده می‌شود مدل تخت و مسطح، خصوصیات ۱، ۲ و ۵ را در سطح بالا دارا می‌باشند، و نسبت به تمام مدل‌ها از افزونگی بالا و پیچیدگی پایین برخوردار هستند. مدل ستاره‌ای خصوصیات ۱، ۲، ۴ و ۵ را در سطح بالا دارا است و خصوصیات ۳ و ۶ در پایین‌ترین سطح می‌باشند، و به جزء مدل‌های تخت و مسطح نسبت به مدل‌های دیگر پیچیدگی کمتری دارد. ولی نسبت به مدل‌های دانه‌برفی، دانه‌ستاره‌ای و خوشه‌ستاره‌ای افزونگی بالاتری دارد. مدل فلکی و کهکشانی خصوصیات ۱، ۴ و ۵ را در سطح بالا و خصوصیات ۳، ۶ و ۷ را به طور متوسط دارا می‌باشند. بنابراین نسبت به مدل دانه‌برفی افزونگی بالاتر و در مقایسه با مدل ستاره-ای پیچیدگی بیشتر دارند. همچنین برای پایگاه‌های داده تحلیلی بزرگ بکار می‌روند. مدل دانه‌برفی خصوصیات ۳، ۶ و ۷ را در بالاترین سطح و خصوصیت ۵ آن در پایین‌ترین سطح می‌باشد. بنابراین نسبت به تمام مدل‌ها دارای بالاترین پیچیدگی و پایین‌ترین افزونگی می‌باشد. مدل‌های



شکل ۱۱: نتایج آزمایشات براساس چهار نوع پرس و جوی گروه بندی شده

آزمایشات بعدی براساس میانگین زمان پاسخگویی چهار نوع پرس و جو در گروه‌بندی‌های مختلف در سه مدل دانه برفی، ستاره‌ای و دولایه‌ای انجام شده است. نتایج این آزمایشات در شکل (۱۲) نشان داده شده است. بنابراین مدل دولایه‌ای نسبت به مدل ستاره‌ای و دانه‌برفی زمان پاسخ به پرس و جوها را بهبود داده است.



شکل ۱۲: نتایج آزمایشات برای میانگین زمان پاسخگویی در چهار نوع پرس و جوی گروه بندی شده

همچنین نتایج آزمایشات دو نوع پرس و جوی slice و dice براساس میانگین زمان پاسخگویی در سه مدل در شکل (۱۳) نشان داده شده است که همواره مدل دولایه‌ای نسبت به دو مدل (ستاره‌ای و دانه‌برفی) زمان پاسخ به پرس و جوها را بهبود داده است.

۵- نتایج مقایسات

همانطور که در بخش ۲-۲ ذکر شد، مدل‌های طراحی در سطح منطقی عبارتند از: مدل تخت [۵ و ۶]، مدل مسطح [۵ و ۶]، مدل ستاره‌ای [۵-۷]، مدل فلکی [۵ و ۶]، مدل کهکشانی [۵ و ۶]، مدل دانه‌برفی [۵-۷]، مدل دانه ستاره-ای [۵ و ۸] و مدل خوشه‌ستاره‌ای [۵-۷]. این مدل‌ها را می‌توان بر اساس هفت خصوصیت ارزیابی و

۲، ۳، ۴ و ۷ را در سطح بالا و خصوصیات ۵ و ۶ را به طور متوسط دارا می‌باشد. مدل دولایه‌ای نسبت به مدل دانه‌برفی پیچیدگی و نسبت به مدل ستاره‌ای افزونگی کمتری دارد. همچنین نسبت به مدل دانه‌ستاره‌ای و خوشه‌ستاره‌ای انعطاف‌پذیری بالاتری دارد. بنابراین مدل پیشنهادی نسبت به مدل‌های موجود و با توجه به خصوصیات مطرح در آن‌ها کامل‌تر می‌باشد.

دانه‌ستاره‌ای و خوشه‌ستاره‌ای به طور متوسط تمام خصوصیات را دارا می‌باشند، از طرفی مدل‌های دانه‌ستاره‌ای و خوشه‌ستاره‌ای برای تعادلی بین افزونگی مدل ستاره‌ای و پیچیدگی مدل دانه‌برفی ایجاد شده‌اند اما نسبت به مدل ستاره‌ای دارای پیچیدگی بیشتری می‌باشند. همچنین پیچیدگی مدل خوشه‌ستاره‌ای از مدل دانه‌ستاره‌ای نیز بیشتر است و نسبت به مدل دانه‌برفی افزونگی بالاتری دارد. در نهایت مدل پیشنهادی (مدل دولایه‌ای)، خصوصیات ۱،

جدول ۱: مقایسه مدل‌ها در سطح منطقی

خصوصیات مدل	۱	۲	۳	۴	۵	۶	۷
تخت	بالا	بالا	پایین	متوسط	بالا	پایین	پایین
سطح	بالا	بالا	پایین	متوسط	بالا	پایین	پایین
ستاره‌ای	بالا	بالا	پایین	بالا	بالا	پایین	متوسط
فلکی	بالا	متوسط	پایین	بالا	بالا	متوسط	متوسط
کهکشانی	بالا	متوسط	پایین	بالا	بالا	متوسط	متوسط
دانه‌برفی	متوسط	متوسط	بالا	متوسط	پایین	بالا	بالا
دانه‌ستاره‌ای	متوسط	متوسط	متوسط	متوسط	متوسط	متوسط	متوسط
خوشه‌ستاره‌ای	متوسط	متوسط	متوسط	متوسط	متوسط	متوسط	متوسط
دولایه‌ای	بالا	بالا	بالا	بالا	متوسط	متوسط	بالا

۶- نتیجه‌گیری و کارهای آینده

فرایند توسعه یک پایگاه داده تحلیلی با تحلیل پایگاه داده عملیاتی، شناسایی نیازهای تحلیلی، و نهایتاً طراحی در سه سطح مفهومی، منطقی و فیزیکی انجام می‌شود. در این مقاله ابتدا مدل‌های طراحی موجود در سطوح مفهومی و منطقی پایگاه داده تحلیلی بررسی شد و مدلی بهبود یافته در سطح منطقی، برای فرایند طراحی پایگاه داده تحلیلی، ارائه شد که ترکیبی از دو مدل ستاره‌ای و دانه‌برفی، به صورت دولایه‌ای است. به منظور مقایسه مدل پیشنهادی با مدل‌های موجود، از معیار زمان پاسخگویی به پرس‌وجوها استفاده شد. آزمایشات نشان داد که مدل پیشنهادی منجر به بهبود زمان پاسخ به پرس‌وجوها می‌شود. سپس مدل‌های سطح منطقی با توجه به خصوصیات مطرح در آن‌ها مقایسه و تحلیل شد که مدل پیشنهادی (مدل دولایه‌ای) با توجه به عیب مدل‌های موجود، پیچیدگی بالای مدل دانه‌برفی و

افزونگی بالای مدل ستاره‌ای و همچنین مزایای مدل‌های موجود (کارایی، قابلیت استفاده، قابلیت استفاده مجدد، انعطاف پذیری و اعتبار) توانست نسبت به مدل‌های موجود مدلی کامل‌تر باشد. یکی از کارهای پیشنهادی برای آینده، مدل سه‌لایه‌ای است. در مدل سه‌لایه‌ای، از مدل خوشه‌ستاره‌ای در مدل دولایه‌ای می‌توان استفاده نمود. به این صورت که در میان لایه دانه‌برفی و ستاره‌ای قرار می‌گیرد. به این منظور براساس مطالعات، مدل خوشه‌ستاره‌ای برای تعادل بین افزونگی و پیچیدگی جداول در دو مدل ستاره‌ای و دانه‌برفی ایجاد شده‌است و از داده‌های تکراری در جداول جلوگیری می‌کند. بنابراین در سطح نرمال سازی جداولی که در موقعیت یک پدر و دو فرزند می‌باشند، یعنی در دو بعد مشترکند (روبهم افتادگی) را تبدیل به یک زیر بعد می‌کند و با این عمل باعث کاهش پیچیدگی جداول می‌شود.

مراجع

- [1] R. Jindal, and SH. Taneja, "Comparative study of Data Warehouse Design Approches: A Survey", International Journal of Database Management Systems (IJDMS), Vol.4, No.1, 2012.
- [2] N. Arfaoui, and J. Akaichi, "Data warehouse: conceptual and logical schema – survey", International Journal of Enterprise Computing and Business Systems, Vol.2, 2012, pp. 1-31.
- [3] N. Tryfona, F. Busborg, and J.G. B. Christiansen, "starER: A Conceptual Model for Data Warehouse Design", Department of Computer Science, Aalborg University, Fredrik Bajersvej 7E, DK-9220, Aalborg Øst, Denmark, 1999, pp. 1-6.
- [4] A.S. Kamble, "A conceptual model for multidimensional data", Proceedings of the fifth on Asia-Pacific conference on conceptual modeling, Wollongong, NSW, Australia, 2008.
- [5] B. P. Başaran, "A comparison of data warehouse design models", 2005.
- [6] D. L. Moody, and M. A. R. Kortink, "From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design", Proceedings of International Workshop on Design and Management of Data Warehouses (DMDW'2000), 2000, Sweden, pp. 5-1, 5-12.
- [7] B. Seyed-Abbassi, "Teaching Effective Methodologies to Design a Data Warehouse", 2001.
- [8] D. L. Moody, and M. A. R. Kortink, "From ER Models to Dimensional Models Part II: Advanced Design Issues", Journal of Business Intelligence, 2008, PP. 1-12.
- [9] D. Mishra, A. Yazici, and B. P. Başaran, "A Case study of Data Models in Data Warehousing Design", 2008, pp. 314-319.
- [10] S. Nugawela, "Data warehousing model for integrating fragmented electronic health records from disparate and heterogeneous clinical data stores", School of Electrical Engineering and Computer Science Faculty of Science and Engineering Queensland University of Technology, 2013.
- [11] R. Kimball, "Data Warehouse Toolkit", John Wiley & Sons, Inc, New York, 1996.
- [12] A. Gosain, "Literature Review of Data model Quality metrics of Data Warehouse", International Conference on Intelligent Computing, Communication & Convergence (ICCC), Vol. 48, 2015, pp. 236 – 243.
- [13] J. Completo, R. S. Cruz, L. Coheur, and M. Delgado, "Design and Implementation of a Data Warehouse for Benchmarking in Clinical Rehabilitation", International Conference on Health and Social Care Information Systems and Technologies, Vol. 5, 2012, pp. 885 – 894.
- [14] M. Arora, and A. Gosain "Schema Evolution for Data Warehouse: A Survey", International Journal of Computer Applications, Vol. 22, No.6, 2011.
- [15] P. Jovanovica, O. Romeroa, A. Simitsisb, A. Abellóa, and D. Mayorova, "A requirement-driven approach to the design and evolution of data warehouses", Information Systems, Vol.44, 2014, pp. 94–119.
- [16] J. George, B. Vijay Kumar, and V. Santhosh Kumar, "Data Warehouse Design Considerations for a Healthcare Business Intelligence System", Proceedings of the World Congress on Engineering (WCE), Vol. 1, 2015.
- [۱۷] ز. زاهدین لباف، ن. دانشپور، "یافتن خودکار حقایق و ابعاد با روش اکتشافی در پایگاه داده تحلیلی". بیست و یکمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، تهران، ۱۸ تا ۲۰ اسفند، ۱۳۹۴.
- [18] V. Tiwari, and R. Singh Thakur, " Contextual snowflake modelling for pattern warehouse logical design", 2015.

- [19] V. Nebot, and R. Belanga, "Statistically-driven generation of multidimensional analytical schemas from linked data", Knowledge-Based Systems, Vol. 0, 2016, pp.1–15.
- [20] A. Gosain, and S. Mann, "Empirical validation of metrics for object oriented multidimensional model for data warehouse", The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2013.
- [21] M. Usman, R. Pears, and A. C. M. Fong, "Discovering diverse association rules from multidimensional schema", Expert Systems with Applications, Vol.40, 2013, pp. 5975-5996.
- [22] M. G. de Carvalho, A. H. F. Laender, M.A. Goncalves, and A. S. da Silva, "An evolutionary approach to complex schema matching", Information Systems, Vol. 38, 2013, pp. 302-316.
- [23] M. Kumar, A. Gosain, and Y. Singh, "Empirical validation of structural metrics for predicting understandability of conceptual schemas for data warehouse", The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2013.
- [24] C. Sapia, M. H. Blaschka, G. Fling, and B. Dinter, "Extending the E/R Model for the Multidimensional Paradigm", Proceeding of the International Workshop on Data Warehousing and Data Mining, Singapore, 1998, pp. 105-116.
- [25] E. Malinowski, and E. Zimányi, "Representing spatiality in a conceptual multidimensional model", Proceedings of ACM international work shop on Geographic information systems, ACM Press, New York, 2004, pp. 12-22.
- [26] E. Malinowski, and E. Zimányi, "Spatial Hierarchies and Topological Relationships in the Spatial MultiDimERmodel", Proceedings of the 22nd British National Conference on Databases, BNCOD22, number 3567 in Lecture Notes in Computer Science, Springer-Verlag, Sunderland, UK, 2005, pp. 17-28.
- [27] E. Malinowski, and E. Zimányi, "Hierarchies in a Multidimensional Model: From Conceptual Modeling to Logical Representation", Data & Knowledge Engineering, Vol.59, No.2, 2006, pp. 348-377.
- [28] M. Boehnlein, and A. U. Ende, "Deriving Initial Data Warehouse Structures from the Conceptual Data Models of the Underlying Operational Information Systems", Proceedings of the ACM Second International Workshop on Data Warehousing and OLAP (DOLAP'1999, Kansas City, 6. November), 1999, pp. 15-21.
- [29] A. Abellò, J. Samos, and F. Saltor, "YAM2 (Yet another multidimensional model): An extension of UML", International Database Engineering & Applications Symposium (IDEAS 2002), Edmonton, Canada, IEEE Computer Society, 2002, pp. 172-18.
- [30] A. Abellò, J. Samos, and F. Saltor, "YAM2: a multidimensional conceptual model extending UML", Information Systems (IS), Elsevier, 2006, pp. 541–567.
- [31] S. L. Mora, and J. Trujillo, "A comprehensive method for data warehouse design", Proceedings of the 5th International Workshop on Design and Management of Data Warehouses, Germany, Vol. 1, 2003, pp. 1-14.
- [32] S. L. Mora, J. Trujillo, and I. Y. Song, "A UML profile for multidimensional modeling in data warehouses", Data & Knowledge Engineering, 2006, pp. 725-769.
- [33] J. Trujillo, and M. Palomar, "An Object Oriented Approach to Multidimensional Database Conceptual Modeling (OOMD)", Conference on Information and Knowledge Management, ACM Press, New York, 1998, pp. 16-21.
- [34] J. Trujillo, "The GOLD Model: An Object Oriented Multidimensional Data Model For Multidimensional Databases", Proceedings of the 9th ECOOP International Workshop For PhD Students In Objects Oriented Systems, Lisboa, Portugal, 1999.

- [35] A. Sarkar, and S. CHoudhury , "Conceptual Level Design of Object Oriented Data Warehouse: Graph Semantic Based Model", 2009.
- [36] D. N. Xuan, L. Bellatreche, and G. Pierra, "A Versioning Management Model for Ontology-Based Data Warehouses", Proceedings of the 8th International Conference on Data Warehousing and Knowledge Discovery DaWaK 2006, Krakow, Poland, 2006, pp. 195-206.
- [37] D. Skoutas, and A. Simitsis, "Ontology-Based Conceptual Design of ETL Processes for Both Structured and Semi-Structured Data". International Journal of Semantic Web and Information Systems, Vol. 3, No. 4, 2007, pp. 1-24.
- [38] N. Tsois, and T. S. Karayannidis, "MAC: Conceptual data modeling forOLAP", Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW-2001), 2001, pp. 5-1, 5-13.
- [39] M. Golfarelli, D. Maio, and S. Rizzi, "The Dimensional Fact Model: a Conceptual Model for Data Warehouses", International Journal of Cooperative Information Systems, 1998, pp. 215-247.
- [40] M. Golfarelli, "From User Requirements to Conceptual Design in Data Warehouse Design – a Survey", 2010.
- [41] A. Gutiérrez, and A. Marotta, " An Overview of Data Warehouse Design Approaches and Techniques", Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay, 2000.
- [42] V. Peralta, and R. Ruggia, "Using Design Guidelines to Improve DataWarehouse Logical Design", Proc of the International Workshop on Design and Management of Data Warehouses, Berlin, 2003.
- [43] V. Peralta, "Data Warehouse Logical Design from Multidimensional Conceptual Schemas", Instituto de Computación, Universidad de la República, Julio Herrera y Reisig 565, Montevideo, CP 11300, Uruguay, 2003.
- [44] M. H. Peyravi, "A Schema Selection Framework for Data Warehouse Design", International Journal of Machine Learning and Computing, Vol. 2, No. 3, 2012.
- [45] M. Shekofteh, "Building Quick and Accurate Structure for Data Warehouse", International Journal of Advanced Computer Science, This work was supported by Department Of computer science, Sarvestan Branch, Islamic Azad University, sarvestan, Iran, Vol. 2, No. 7, 2012, pp. 274-276.
- [46] R. K. Shrivastava, M. Mahrishi, and P. Parnami, "An Optimized algorithm to select the appropriate Schema in Data Warehouses". International Journal of Engineering Sciences & Research Technology, 2012, pp. 198-204. ISSN: 2277-9655.