

## مدلی سه لایه‌ای در طراحی سطح منطقی پایگاه داده تحلیلی

محیا ارومیه<sup>۱</sup>، دانشجوی کارشناسی ارشد؛ نگین دانش پور<sup>۲</sup>، استادیار

۱- دانشکده مهندسی کامپیوتر و فناوری اطلاعات - واحد قزوین - دانشگاه آزاد اسلامی - قزوین - ایران - mahya\_roumiye@yahoo.com

۲- دانشکده مهندسی کامپیوتر - دانشگاه تربیت دبیر شهید رجایی - تهران - ایران - ndaneshpour@srttu.edu

**چکیده:** پایگاه‌های داده عملیاتی و سیستم‌های پردازش تراکنشی برخط ساختارهای مناسبی برای نگهداری و جمع‌آوری اطلاعات هستند، اما چالش این ساختارها چگونگی استفاده و تفسیر این اطلاعات است. یک پایگاه داده تحلیلی، منبع داده‌ای است که داده‌ها را از سیستم‌ها و منابع اطلاعاتی دیگر مانند پایگاه داده‌های عملیاتی، سیستم کاربران و ... جمع‌آوری می‌کند و پس از همگن و یکپارچه‌سازی، به صورت چندبعدی دسته‌بندی و ذخیره می‌کند. این منبع داده برای عملیات گزارش‌گیری، تجزیه و تحلیل، کمک به تصمیم‌گیری و در معماری هوش تجاری مورد استفاده قرار می‌گیرد. فرآیند 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 سطح منطقی پایگاه داده تحلیلی بررسی می‌شود. سپس مدلی بهبودیافته در این سطح، برای فرآیند طراحی پایگاه داده تحلیلی، ارائه می‌شود که ترکیبی از سه مدل دانه‌برفی، خوشه‌ستاره‌ای و ستاره‌ای، به صورت سه لایه‌ای است. به منظور مقایسه مدل پیشنهادی با مدل‌های موجود، از معیار زمان پاسخ‌گویی به پرس‌وجوها استفاده شده است. آزمایش‌ها نشان می‌دهد که مدل پیشنهادی منجر به بهبود زمان پاسخ به پرس‌وجوها می‌شود.

**واژه‌های کلیدی:** پایگاه داده تحلیلی در سطح منطقی، مدل دانه‌برفی، مدل خوشه‌ستاره‌ای، مدل ستاره‌ای.

## Three-Layers Model Designing in the Logical Level Data Warehouse

M. Oroumiyeh<sup>1</sup>, MSc Student; N. Daneshpour<sup>2</sup>, Assistant Professor

1- Faculty of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran, Email: mahya\_roumiye@yahoo.com

2- Faculty of Computer Engineering, Shahid Rajaei Teacher Training University, Tehran, Iran, Email: ndaneshpour@srttu.edu

**Abstract:** The operational databases and online transaction processing systems are appropriate structures for maintenance and collection of data, but the challenge of structures is how to use and interpret this data. A data warehouse is a data source which collects data from other systems and data sources such as operational data bases, users systems, and etc; and stores and categorizes data in the form of multidimensional after homogenization and integration. This data source for reporting operation, analysis, to help make decisions and business intelligence. The process of developing a data warehouse is done through operational data base analyzing, analytical requirement identification and finally designing in conceptual, logical and physical levels. In this paper, the designing models at logical level of data warehouses are researched. Then an improved model is presented in this level that combines of three models (snow flake, star cluster and star) in the form of a three - layers. Queries response time measure is used to compare the proposed model with existing models. Experiments show that the proposed model improves the response time of queries.

**Keywords:** Data warehouse in logical level, snowflake schema, star cluster schema, star schema.

تاریخ ارسال مقاله: ۱۳۹۴/۰۹/۰۵

تاریخ اصلاح مقاله: ۱۳۹۵/۰۱/۱۴ و ۱۳۹۵/۰۳/۱۳

تاریخ پذیرش مقاله: ۱۳۹۵/۰۶/۰۸

نام نویسنده مسئول: نگین دانش پور

نشانی نویسنده مسئول: ایران - تهران - لویزان - دانشگاه تربیت دبیر شهید رجایی - دانشکده مهندسی کامپیوتر.

## ۱- مقدمه

پایگاه داده تحلیلی<sup>۱</sup> در دهه نود مطرح شد. یک پایگاه داده تحلیلی مجموعه‌ای مبتنی بر موضوع<sup>۲</sup>، مجتمع<sup>۳</sup>، متغیر با زمان<sup>۴</sup> و غیرفرار<sup>۵</sup> از داده‌ها است که معمولاً برای تصمیم‌گیری سازمانی استفاده می‌شود. پایگاه داده تحلیلی حجم زیادی از داده‌های منابع داده چندگانه را که برای پرس‌وجو و تحلیل استفاده می‌شود، ذخیره می‌کند [۱].

پایگاه داده تحلیلی از پایگاه‌های داده عملیاتی و یا سایر منابع داده توزیع‌شده سازمان‌ها و ارگان‌های متفاوت تهیه می‌شود و بستر مناسبی فراهم می‌آورد تا داده‌ها به‌منظور پاسخ‌گویی به پرس‌وجوهای تحلیلی به‌صورت بایگانی‌شده و سازمان‌یافته، ذخیره شوند. پایگاه داده تحلیلی شامل داده‌هایی است که برای انجام تصمیم‌گیری‌ها و تحلیل‌ها مناسب است. داده‌های موجود در پایگاه داده تحلیلی، غیر قابل تغییر، یعنی فقط خواندنی هستند و توسط کاربران قابل تغییر نیستند و هدف آن، بهبود کیفیت و سرعت تصمیم‌های کاربران دانش (مدیر، تحلیل‌گر) است.

فرآیند توسعه پایگاه داده تحلیلی با تحلیل پایگاه داده عملیاتی و شناسایی نیازهای تحلیلی شروع می‌شود و پس از آن مانند هر سیستم دیگری [۲-۴] طراحی آن صورت می‌گیرد. طراحی پایگاه داده تحلیلی در سه سطح مفهومی، منطقی و فیزیکی انجام می‌شود. این مقاله بر روی سطح منطقی طراحی پایگاه داده تحلیلی تمرکز دارد.

مدل‌های طراحی مطرح در سطح منطقی عبارت‌اند از: مدل تخت<sup>۶</sup> [۵، ۶]، مدل مسطح<sup>۷</sup> [۵، ۶]، مدل ستاره‌ای<sup>۸</sup> [۵-۷]، مدل فلکی<sup>۹</sup> [۵، ۶]، مدل کهکشانی<sup>۱۰</sup> [۵، ۶]، مدل دانه‌برفی<sup>۱۱</sup> [۵-۷]، مدل دانه‌ستاره‌ای<sup>۱۲</sup> [۵، ۸] و مدل خوشه‌ستاره‌ای<sup>۱۳</sup> [۵-۷]. این مدل‌ها را می‌توان بر اساس هفت خصوصیت ارزیابی و مقایسه نمود. این خصوصیت‌ها از مطالعات انجام‌شده بر روی مدل‌های مختلف استخراج شده‌اند [۵، ۹، ۱۰] و عبارت‌اند از: ۱. کارایی<sup>۱۴</sup>، ۲. قابلیت استفاده<sup>۱۵</sup>، ۳. قابلیت استفاده مجدد<sup>۱۶</sup>، ۴. انعطاف‌پذیری<sup>۱۷</sup>، ۵. افزونگی<sup>۱۸</sup>، ۶. پیچیدگی<sup>۱۹</sup> و ۷. اعتبار<sup>۲۰</sup>.

در این مقاله مدل‌های در سطح منطقی با توجه به خصوصیت‌های مطرح در آن‌ها بررسی می‌شوند. همچنین مدلی بهبودیافته به‌صورت سه‌لایه‌ای، برای فرآیند طراحی پایگاه داده تحلیلی در سطح منطقی ارائه می‌شود که ترکیبی از سه مدل دانه‌برفی، خوشه‌ستاره‌ای و ستاره‌ای است. آزمایش‌ها نشان داده که این مدل منجر به بهبود زمان پاسخ به پرس‌وجوها می‌شود.

در این مقاله در بخش ۲ مدل‌های مطرح در سطح منطقی با توجه به خصوصیت‌های مربوط در آن‌ها معرفی شده است. در بخش ۳ مدل پیشنهادی معرفی شده است. در بخش ۴ پیاده‌سازی و ارزیابی مدل پیشنهادی ارائه می‌شود؛ و در نهایت در بخش ۵ نتیجه‌گیری انجام شده است.

## ۲- بررسی کارهای مرتبط

در این بخش، مدل‌های ارائه‌شده در طراحی سطح منطقی پایگاه داده تحلیلی، بررسی و تحلیل می‌شوند.

**مدل تخت و مسطح [۵، ۶]:** ساده‌ترین مدل ممکن بدون از دست دادن اطلاعات می‌باشند. این مدل‌ها با تجمیع همه موجودیت‌های مدل داده شکل می‌گیرند و شماری از جدول‌ها را حداقل می‌کنند. دارای خصوصیت‌های کارایی، قابلیت استفاده و افزونگی در سطح بالا می‌باشند و نسبت به تمام مدل‌ها از افزونگی بالا و پیچیدگی پایین برخوردار هستند [۶].

**مدل ستاره‌ای [۷-۵]:** ساختار اصلی یک مدل چندبعدی است که بر اساس یک جدول مرکزی بزرگ (جدول حقیقت) و یک مجموعه از جدول‌های کوچک‌تر (ابعاد) سازمان‌دهی شده است. رکوردها در دو نوع از جدول‌ها ذخیره شده‌اند: یک جدول حقیقت و تعدادی جدول بعد. جدول حقیقت کلیدهایی از هر جدول بعد را نگهداری می‌کند. روابط بین هر جدول بعد و جدول حقیقت یک رابطه یک‌به‌چند است که شامل یک کلید منحصره‌فرد در جدول حقیقت است. جدول حقیقت به غیر از کلیدهای خارجی شامل معیارها نیز است. مدل ستاره‌ای به دلیل جدول‌های بعد غیرنرمال دارای افزونگی بالا است. بنابراین مدل ستاره‌ای خصوصیت‌های کارایی، قابلیت استفاده، انعطاف‌پذیری و افزونگی را در سطح بالایی دارا است و خصوصیت‌های قابلیت استفاده مجدد و پیچیدگی در پایین‌ترین سطح می‌باشند و به جز مدل‌های تخت و مسطح نسبت به مدل‌های دیگر پیچیدگی کم‌تری دارد ولی نسبت به مدل‌های دانه‌برفی، دانه‌ستاره‌ای و خوشه‌ستاره‌ای افزونگی بالاتری دارد [۵، ۶].

**مدل فلکی [۵، ۶]:** این مدل شامل یک مجموعه از مدل‌های ستاره‌ای است و جدول‌های حقیقت با هم ارتباط دارند. این ارتباط‌ها توانایی drill down بین سطوحی از جزئیات را فراهم می‌کند. مدل فلکی خصوصیت‌های کارایی، انعطاف‌پذیری و افزونگی را در سطح بالا و خصوصیت‌های قابلیت استفاده مجدد، پیچیدگی و اعتبار را به‌طور متوسط دارا است [۵].

**مدل کهکشانی [۵، ۶]:** این مدل یک مجموعه از مدل‌های ستاره‌ای است که دارای ابعاد مشترک می‌باشند. برعکس مدل فلکی، جدول‌های حقیقت در مدل کهکشانی نیاز به ارتباط مستقیم ندارند. مدل کهکشانی خصوصیت‌های کارایی، انعطاف‌پذیری و افزونگی را در سطح بالا و خصوصیت‌های قابلیت استفاده مجدد، پیچیدگی و اعتبار را به‌طور متوسط دارا است [۵].

**مدل دانه‌برفی [۷-۵]:** این مدل نتیجه‌ای از نرمال‌سازی یک یا چند بعد در مدل ستاره‌ای است. مدل دانه‌برفی دارای ساختار سلسله‌مراتبی از ابعاد است. مدل دانه‌برفی تکنیک کلی برای ذخیره‌سازی اطلاعات چند ارزشی در پایگاه داده تحلیلی است که نرمال‌سازی در جدول‌های بعد را به دنبال دارد. این مدل به دلیل جدول‌های نرمال دارای پیچیدگی بالا است. در مدل دانه‌برفی

همچنین ممکن است در سطح نرمال‌سازی، جدولی در موقعیت یک پدر و دو فرزند قرار بگیرد، یعنی در دو بعد مشترک باشد. در این شرایط مدل خوشه‌ستاره‌ای ایجاد می‌شود. به این منظور جدول مذکور تبدیل به یک جدول زیربهد می‌شود و با این عمل باعث کاهش پیچیدگی جدول‌ها در مدل دانه‌برفی می‌شود. این مدل، بهترین راه‌حل 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.

مدل پیشنهادی با توجه به مدل‌های مطرح‌شده قبلی از جمله خوشه‌ستاره‌ای و دانه‌ستاره‌ای که به صورت ترکیبی از مدل‌ها هستند، ایجاد شد.

مدل دانه‌ستاره‌ای دارای ساختاری ترکیبی است و در واقع شامل ترکیبی از مدل ستاره‌ای (غیرنرمال‌سازی) و دانه‌برفی (نرمال‌سازی) است. 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 زیربهد می‌شود.

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 جدول‌ها و روابط بین آن‌ها به دست می‌آید، ایجاد شد. بنابراین طبق اهداف مدل‌سازی بعدی، در مدل پیشنهادی از مدل ستاره‌ای که دارای حداقل جدول‌ها و روابط است در لایه سوم استفاده شد.

در واقع مدل پیشنهادی به صورت ترکیبی از مدل‌های ستاره‌ای، خوشه‌ستاره‌ای و دانه‌برفی، با توجه به مزایای این سه مدل و در راستای رفع معایب آن‌ها پیشنهاد شد. به این منظور مدل پیشنهادی، بر اساس ویژگی سه‌لایه‌ای بودن این‌گونه ایجاد شد که عیب مدل دانه‌برفی (لایه پایینی) که پیچیدگی بالای آن است و منجر به افزایش زمان پاسخ به پرس‌وجوها می‌شود را با افزودن دیدهای ذخیره‌شده مناسب برطرف می‌کند. در واقع برای ایجاد این دیدها، لایه میانی (مدل خوشه‌ستاره‌ای) ایجاد می‌شود. طبق تعریف مدل خوشه‌ستاره‌ای در سطح نرمال‌سازی، جدولی که در موقعیت یک پدر و دو فرزند قرار بگیرد، یعنی در دو بعد مشترک باشد تبدیل به یک جدول زیربهد می‌شود و با این عمل مدل خوشه‌ستاره‌ای باعث کاهش پیچیدگی جدول‌ها در مدل دانه‌برفی می‌شود. همچنین جدول زیربهد لایه خوشه‌ستاره‌ای، به صورت جدول غیرنرمال در لایه بالا (لایه ستاره‌ای) تبدیل می‌شود؛ بنابراین در لایه ستاره‌ای به دلیل عدم وجود جدول زیربهد، افزونگی نسبت به مدل خوشه‌ستاره‌ای افزایش می‌یابد.

در رابطه با عیب مدل ستاره‌ای افزونگی داده در بالاترین سطح است و افزونگی مشکلات زیر را می‌تواند ایجاد کند: هدر رفتن فضای ذخیره‌سازی، ایجاد تناقض در هنگام بهنگام‌سازی، حذف و درج، ولی از آن جایی که درج و تغییر اطلاعات در پایگاه داده تحلیلی توسط کاربران

خصوصیت‌های قابلیت استفاده مجدد، پیچیدگی و اعتبار در بالاترین سطح و خصوصیت افزونگی پایین‌ترین سطح می‌باشد. بنابراین نسبت به تمام مدل‌ها دارای بالاترین پیچیدگی و پایین‌ترین افزونگی است [۵، ۶].

**مدل دانه‌ستاره‌ای [۵، ۸]:** این مدل دارای ساختاری ترکیبی است و در واقع شامل ترکیبی از مدل ستاره‌ای و دانه‌برفی است. این مدل ترکیبی از نرمال‌سازی و غیرنرمال‌سازی است. مدل دانه‌ستاره‌ای به طور متوسط تمام خصوصیت‌ها را دارا است [۵].

**مدل خوشه‌ستاره‌ای [۷-۵]:** این مدل، بهترین راه‌حل 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 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 زیربهد شود و با این عمل باعث کاهش پیچیدگی جدول‌ها شود. مدل خوشه‌ستاره‌ای به طور متوسط تمام خصوصیت‌ها را دارا است. همچنین پیچیدگی مدل خوشه‌ستاره‌ای از مدل دانه‌ستاره‌ای نیز بیش‌تر است [۵، ۶].

بنابراین با توجه به عیب مدل‌های موجود، پیچیدگی بالای مدل دانه‌برفی، در بخش بعدی مدلی ارائه می‌شود که پیچیدگی مدل دانه‌برفی را برطرف می‌کند و از مزایای مدل‌های موجود، کارایی (سرعت پاسخ‌گویی به پرس‌وجوها) مدل ستاره‌ای 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 آن‌ها است. این هدف با حداقل کردن شماری از جدول‌ها و روابط بین آن‌ها به دست می‌آید (غیرنرمال‌سازی) [۶، ۱۱].

بنابراین طبق اهداف مدل‌سازی بعدی، مدل ستاره‌ای دارای حداقل جدول‌ها و روابط است 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 جدول‌ها و روابط بین آن‌ها کاهش و در نتیجه شماری از اتصال‌های مورد نیاز در پرس‌وجوهای کاربر را کاهش می‌دهد. در واقع پیچیدگی پایینی دارد و به عملکرد پرس‌وجو سرعت می‌بخشد و از طرفی در رابطه با عیب آن می‌توان به افزونگی داده در بالاترین سطح اشاره کرد [۵، ۱۲].

مدل دانه‌برفی شکل نرمال‌شده 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 است. عیب این مدل پیچیدگی بالای آن است که منجر به افزایش زمان پاسخ به پرس‌وجوها می‌شود [۵، ۱۲].

وجود دارد، اما نحوه قرار گرفتن آن‌ها در جدول‌ها، متفاوت است. به‌عنوان مثال در لایه پایینی (مدل دانه‌برفی) تعداد جدول‌های بعد آن نسبت به دو لایه دیگر به دلیل ویژگی نرمال‌سازی بیش‌تر است اما صفات آن‌ها یکسان است.

جدول حقیقت لایه ستاره‌ای با توجه به تغییرات ایجادشده ساخته می‌شود. به این منظور از جدول حقیقت لایه خوشه‌ستاره‌ای یا همان جدول حقیقت لایه دانه‌برفی و جدول زیربعد ساخته‌شده در لایه خوشه‌ستاره‌ای که به‌صورت جدول بعد جدید برای لایه ستاره‌ای تبدیل شده، استفاده می‌شود. در واقع لایه میانی، جدول‌های بعد و جدول زیربعد آن، که از تجمیع جدول‌های بعد لایه پایینی به دست می‌آیند به‌شکل materialized view هستند و همچنین لایه بالایی (مدل ستاره‌ای) لایه materialized view از ابعاد و جدول حقیقت لایه میانی است.

به‌عنوان مثال، بعد خرده‌فروش و بعد مشتری هر دو شامل سلسله‌مراتب بعد مکان جغرافیایی و بعد قلمرو فروش در لایه دانه‌برفی هستند.

بنابراین برای تبدیل مدل دانه‌برفی به مدل خوشه‌ستاره‌ای، جدول بعد مکان جغرافیایی و بعد قلمرو فروش یکی شده و به‌صورت جدول زیربعد در می‌آید.

برای ساخت مدل ستاره‌ای بر اساس نیاز مدیر پایگاه داده تحلیلی، جدول زیربعد به‌عنوان جدول بعد جدید با کلید اصلی مکان جغرافیایی برای مشتری و خرده‌فروش در لایه ستاره‌ای به جدول حقیقت آن وصل می‌شود.

### ۳-۱- مزایای مدل پیشنهادی (مدل سه‌لایه‌ای)

در مدل پیشنهادی، فرآیند استخراج، تغییر شکل و بارگذاری سه فاز دارد. فاز اول در لایه پایینی به‌طور کامل انجام می‌گیرد. فاز دوم در لایه میانی صورت می‌گیرد و به دلیل اینکه نیاز به استخراج و تغییر شکل نمی‌باشد، بعد از تولید جدول زیربعد لایه میانی (که از جدول‌های لایه پایینی به‌صورت تجمیع‌شده به دست می‌آید) انجام می‌پذیرد. در نتیجه در فاز سوم در لایه بالایی مانند فاز دوم فقط بارگذاری صورت می‌گیرد. بنابراین سرعت این فرآیند در لایه بالایی افزایش پیدا می‌کند.

این مدل به دلیل وجود جدول‌های نرمال‌شده در لایه پایینی، امکان به‌روزرسانی و اعمال تغییرات در ساختار لایه ستاره‌ای 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 به‌روزرسانی مدل سه‌لایه‌ای است که می‌تواند متفاوت باشد و بسته به نیاز تغییر کند. به‌گونه‌ای 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،

عادی انجام نمی‌شود بلکه توسط مدیر سیستم و در واقع طراح و ایجادکننده آن صورت می‌گیرد، افزونگی اطلاعات منجر به تناقض نخواهد شد. از طرف دیگر این افزونگی‌ها در سطح ابعاد صورت می‌گیرد و به این دلیل که حجم جدول‌های بعد در مقابل حجم جدول حقیقت ناچیز است بنابراین این عیب تأثیر چندانی در درصد افزایش حجم کل داده‌ها ندارد.

مدل پیشنهادی در این مقاله، بر اساس مدل‌های ستاره‌ای، خوشه‌ستاره‌ای و دانه‌برفی طراحی می‌شود. پیشنهاد مطرح را می‌توان به‌صورت زیر بیان کرد:

مدل، سه‌لایه‌ای است. در مدل سه‌لایه‌ای، از مدل‌های دانه‌برفی، خوشه‌ستاره‌ای و ستاره‌ای استفاده می‌شود. به این صورت که در میان لایه دانه‌برفی و ستاره‌ای، مدل خوشه‌ستاره‌ای قرار می‌گیرد. به این منظور بر اساس مطالعات، مدل خوشه‌ستاره‌ای برای تعادل بین افزونگی و پیچیدگی جدول‌ها در دو مدل ستاره‌ای و دانه‌برفی ایجاد شده است و از داده‌های تکراری در جدول‌ها جلوگیری می‌کند.

بنابراین با توجه به مزایای مدل دانه‌برفی که شامل قابلیت استفاده مجدد و درک و فهم سلسله‌مراتب ابعاد است، در لایه اول (پایین) لایه دانه‌برفی قرار می‌گیرد. جدول‌ها در این لایه کاملاً نرمال و بدون هیچ افزونگی می‌باشند و تمام مراحل فرآیند استخراج، تغییر شکل و بارگذاری (ای.تی.ال) در لایه دانه‌برفی انجام می‌گیرد. سپس لایه دوم (میانی) که لایه مدل خوشه‌ستاره‌ای است و نهایتاً لایه سوم (بالا) که لایه مدل ستاره‌ای است، ایجاد می‌شود. بنابراین از مزایای این مدل که مهم‌ترین آن‌ها سرعت پاسخ به پرس‌وجوها است بهره‌مند می‌شود.

برای ساخت لایه میانی یا همان خوشه‌ستاره‌ای، از جدول‌های لایه پایینی (لایه دانه‌برفی) استفاده می‌شود و جدول‌های بعد باید به‌صورت غیرنرمال باشند. برای رسیدن از لایه دانه‌برفی به لایه خوشه‌ستاره‌ای از الگوی قطعه‌بندی عمودی ابعاد استفاده می‌شود [۱۳، ۱۴]. قطعه‌بندی عمودی ابعاد، زمان پاسخ به پرس‌وجوها را بهبود می‌دهد، اما در لایه خوشه‌ستاره‌ای به دلیل اینکه جدول‌های بعد از موقعیت سلسله‌مراتبی (نرمال) لایه دانه‌برفی به حالت غیرنرمال تبدیل می‌شوند، افزونگی افزایش می‌یابد [۱۳، ۱۴]. همچنین برای ساخت لایه بالایی یا همان ستاره‌ای باید از جدول‌های لایه میانی استفاده شود و جدول زیربعد به صورت غیرنرمال در لایه بالا (لایه ستاره‌ای) تبدیل می‌شود، بنابراین در لایه ستاره‌ای به دلیل عدم وجود جدول زیربعد، افزونگی نسبت به مدل خوشه‌ستاره‌ای افزایش می‌یابد.

نکته مورد توجه در ارتباط با مدل پیشنهادی این است که وقتی از لایه خوشه‌ستاره‌ای به لایه ستاره‌ای می‌خواهد برسد، لایه ستاره‌ای بر اساس نیاز مدیران پایگاه داده تحلیلی ساخته می‌شود، بنابراین مدل قابلیت انعطاف‌پذیری بالایی دارد. به این صورت که جدول زیربعد لایه خوشه‌ستاره‌ای که به‌صورت سلسله‌مراتب در دو بعد است بر اساس نیاز مدیران در لایه ستاره‌ای تغییر پیدا خواهد کرد و به جدول بعد غیرنرمال تبدیل می‌شود. در واقع صفات تمام جدول‌ها در هر سه لایه

طبق شکل ۱ در لایه اول (پایین) مدل دانه‌برفی قرار می‌گیرد که شامل جدول حقیقت و جدول‌های بعد از شماره ۱ تا  $n$  است. جدول بعد نرمال شده ۱-۲ و سلسله‌مراتب مربوط به آن (۱-۲) در جدول‌های بعد شماره ۱ و ۲ مشترک است. بنابراین با توجه به ویژگی‌های مدل خوشه‌ستاره‌ای، جدول بعد نرمال شده ۱-۲ و سلسله‌مراتب مربوط به آن (۱-۲)، در لایه میانی (لایه خوشه‌ستاره‌ای) به جدول زیر بعد تبدیل می‌شوند و این عمل باعث کاهش پیچیدگی جدول‌ها و در نتیجه کاهش زمان پاسخ به پرس‌وجوها نسبت به مدل دانه‌برفی می‌شود. همچنین با توجه به ویژگی مدل ستاره‌ای که جدول‌های بعد آن غیرنرمال است، لایه بالا (مدل ستاره‌ای) بر اساس جدول‌های بعد لایه میانی ساخته می‌شود. تمام جدول‌های بعد در لایه ستاره‌ای به صورت غیرنرمال هستند و جدول زیر بعد لایه میانی به عنوان جدول بعد جدید در لایه ستاره‌ای با جدول حقیقت آن رابطه یک‌به‌چند برقرار می‌کند و لایه ستاره‌ای ساخته می‌شود.

#### ۴- ارزیابی مدل پیشنهادی

زمان پاسخ به پرس‌وجوها همیشه مهم‌ترین معیار برای مقایسه مدل‌ها بوده است [۱۷-۱۵]. در این مقاله از این معیار برای ارزیابی مدل‌های دانه‌برفی، خوشه‌ستاره‌ای، ستاره‌ای و مدل پیشنهادی به نام مدل سه‌لایه‌ای استفاده می‌شود. در نتیجه مدل‌ها در محیط SQL Server بر روی پایگاه داده 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.

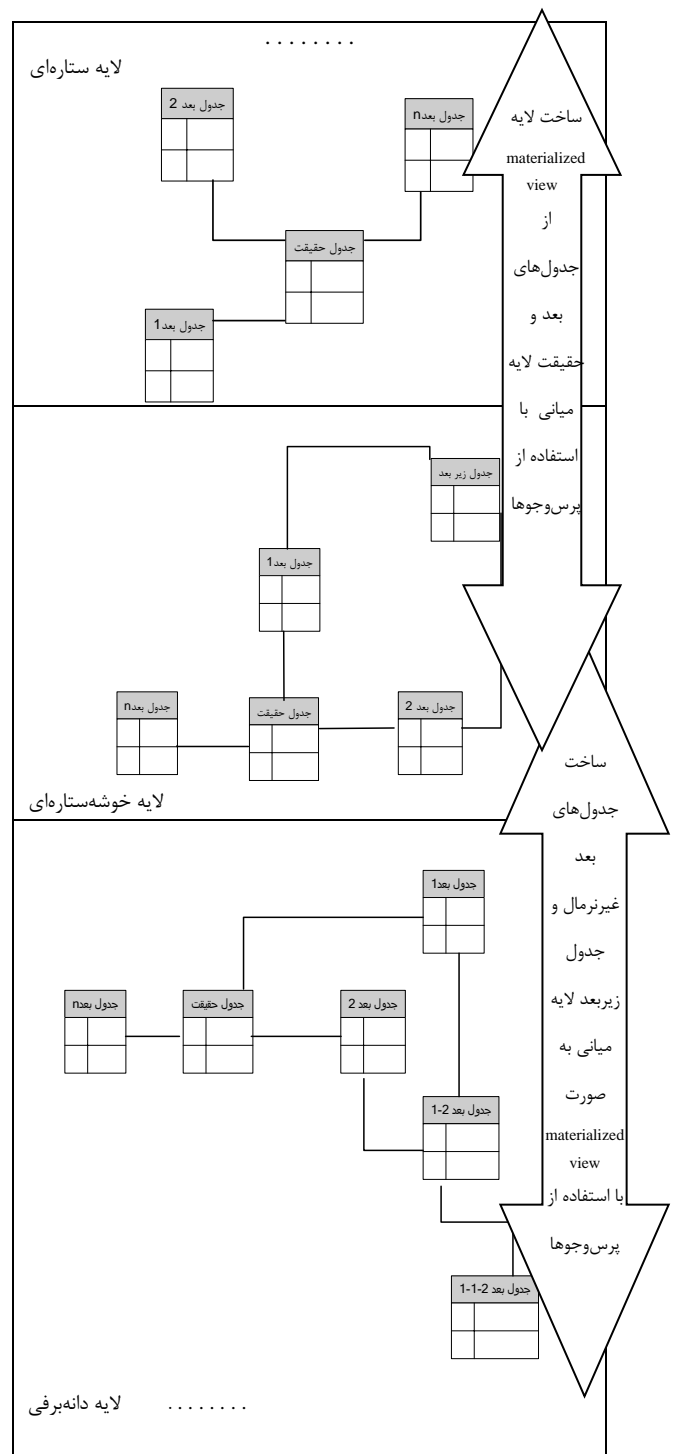
نوع پرس‌وجو معیار مناسب برای انتخاب مدل‌ها می‌باشد. تعداد پرس‌وجوها بر اساس گروه‌بندی با توجه به تعداد جدول‌های بعدی که با جدول حقیقت رابطه یک‌به‌چند دارند، به دست می‌آید.

بنابراین پرس‌وجوهایی بر اساس گروه‌بندی تولید می‌شوند، یعنی اگر  $k$  بعد وجود داشته باشد  $2^k$  پرس‌وجو بر اساس زیرمجموعه‌ای از ابعاد در گروه‌بندی ایجاد می‌شود. بنابراین پرس‌وجوهای یک تا چهار که در شکل‌های ۲ و ۳ به صورت  $q_1$  -  $q_4$  برای ارزیابی‌ها استفاده شده است مربوط به گروه‌بندی‌های مختلف هستند. همچنین پرس‌وجوهایی از نوع slice که یک ارزش از یک بعد و dice که محدوده‌ای از ارزش‌های یک یا چند بعد را انتخاب می‌کنند، به عنوان پرس‌وجوهای پنج و شش که در شکل‌های ۲ و ۳ به صورت  $q_5$  و  $q_6$  برای ارزیابی‌ها استفاده شده است، ایجاد می‌شوند. شکل ۲، این شش پرس‌وجو را برای انجام آزمایش‌ها بر روی پایگاه داده تحلیلی سه‌لایه‌ای فروش با توجه به مکان جغرافیایی مشتری به همراه مدل لایه سوم (لایه ستاره‌ای) به دلیل تغییرات نهایی که در لایه ستاره‌ای مدل سه‌لایه‌ای شکل گرفته است، نمایش می‌دهد. شکل ۳، این شش پرس‌وجو را برای انجام آزمایش‌ها بر روی پایگاه داده تحلیلی سه‌لایه‌ای فروش با توجه به مکان جغرافیایی خرده‌فروش نمایش می‌دهد. در این بخش فقط پرس‌وجوها

به‌روزرسانی آسنکرون یا غیرهمزمان برای سه لایه انجام می‌شود و انعطاف‌پذیری سیستم بالا می‌رود.

#### ۳-۲- شکل گرافیکی مدل پیشنهادی

نمایش گرافیکی مدل پیشنهادی در سه لایه است. در واقع در لایه اول (پایین) مدل دانه‌برفی، در لایه دوم (میانی) مدل خوشه‌ستاره‌ای و در لایه سوم (بالا) مدل ستاره‌ای قرار دارد (شکل ۱).



شکل ۱: مدل سه‌لایه‌ای

<p>q1: <code>select</code> dbo.SUBDim.SalesTerritoryKey,<b>sum</b>(TD)  <code>from</code> dbo.LAYER3.dbo.SUBDim  <code>where</code> dbo.SUBDim.GeographyKey=dbo.LAYER3.[res-G-id]  <code>groupby</code> dbo.SUBDim.SalesTerritoryKey</p>
<p>q2: <code>select</code> dbo.SUBDim.SalesTerritoryKey,TimeKey,<b>sum</b>(TS)  <code>from</code> dbo.LAYER3.dbo.SUBDim  <code>where</code> dbo.LAYER3.[res-G-id]=dbo.SUBDim.GeographyKey  <code>groupby</code> dbo.SUBDim.SalesTerritoryKey,TimeKey  <code>orderby</code> dbo.SUBDim.SalesTerritoryKey,TimeKey <b>asc</b></p>
<p>q3: <code>select</code> dbo.SUBDim.SalesTerritoryKey,ProductKey,<b>sum</b>(TS)  <code>from</code> dbo.LAYER3.dbo.SUBDim  <code>where</code> dbo.SUBDim.GeographyKey=dbo.LAYER3.[res-G-id]  <code>groupby</code> dbo.SUBDim.SalesTerritoryKey,ProductKey  <code>orderby</code> dbo.SUBDim.SalesTerritoryKey,ProductKey</p>
<p>q4: <code>select</code>          dbo.SUBDim.SalesTerritoryKey,ProductKey,TimeKey,<b>sum</b>(TD)  <code>from</code> dbo.LAYER3.dbo.SUBDim  <code>where</code> dbo.SUBDim.GeographyKey=dbo.LAYER3.[res-G-id]  <code>groupby</code> dbo.SUBDim.SalesTerritoryKey,ProductKey,TimeKey  <code>orderby</code> dbo.SUBDim.SalesTerritoryKey,ProductKey,TimeKey</p>
<p>q5: <code>select</code> ProductKey,TimeKey,<b>sum</b>(TD)  <code>from</code> dbo.LAYER3.dbo.SUBDim  <code>where</code> dbo.LAYER3.[res-G-id]=dbo.SUBDim.GeographyKey and          dbo.SUBDim.SalesTerritoryKey=7  <code>groupby</code> ProductKey,TimeKey  <code>orderby</code> ProductKey,TimeKey <b>asc</b></p>
<p>q6: <code>select</code> dbo.SUBDim.SalesTerritoryKey,ProductKey,TimeKey,<b>sum</b>(TS)  <code>from</code> dbo.LAYER3.dbo.SUBDim  <code>where</code> dbo.SUBDim.GeographyKey=dbo.LAYER3.[res-G-id]and          ProductKey&gt;210 and TimeKey&gt;550  <code>groupby</code> dbo.SUBDim.SalesTerritoryKey,ProductKey,TimeKey  <code>orderby</code> dbo.SUBDim.SalesTerritoryKey,ProductKey,TimeKey <b>asc</b></p>

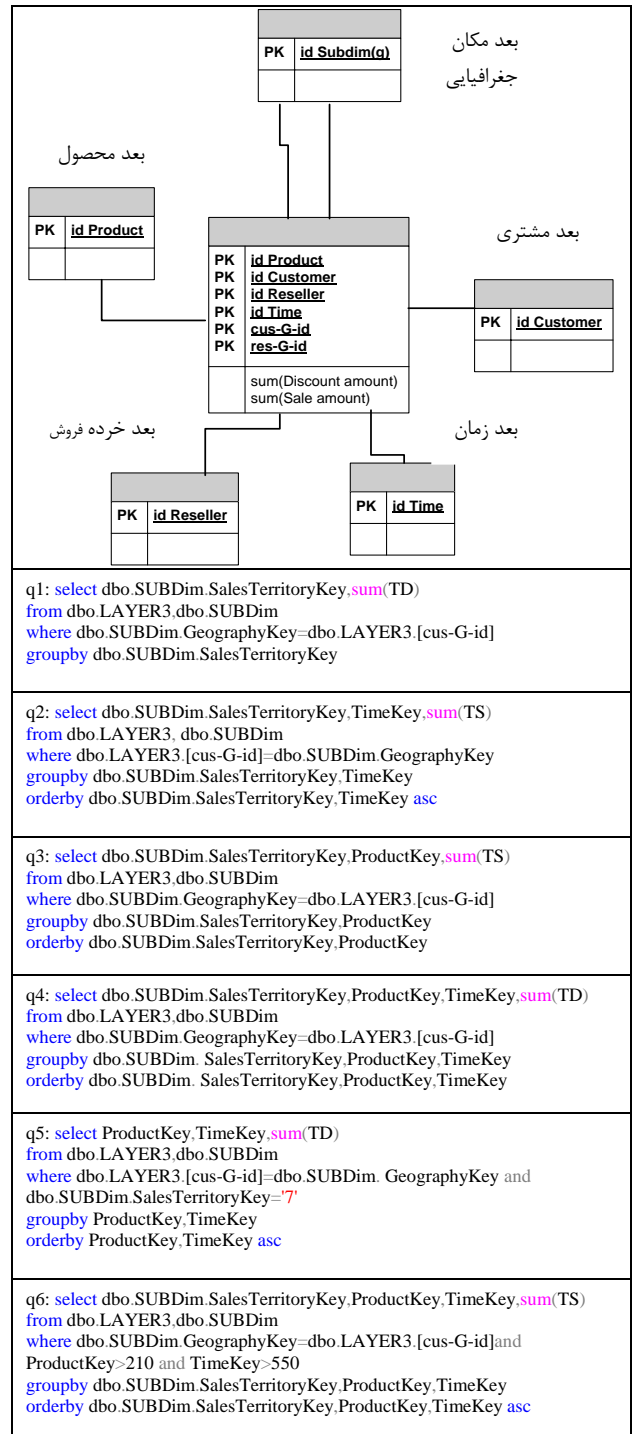
شکل ۳: شش نوع پرس‌وجو در مدل سه‌لایه‌ای با توجه به مکان جغرافیایی خرده‌فروش برای پایگاه داده تحلیلی سیستم فروش

به‌عنوان مثال در لایه سوم (مدل ستاره‌ای) شامل پنج بعد (محصول، مشتری، خرده‌فروش، زمان و مکان جغرافیایی) است. بنابراین ۳۲ پرس‌وجو در گروه‌بندی‌های مختلف ایجاد می‌شود، به این صورت که یا فقط بر اساس یک بعد (محصول، مشتری، خرده‌فروش، زمان و مکان جغرافیایی) یا بر اساس دو بعد ((محصول، مشتری)، (محصول، خرده‌فروش)، (محصول، زمان)، (محصول، مکان جغرافیایی)، (مشتری، خرده‌فروش)، (مشتری، زمان)، ...) یا سه بعد، یا چهار بعد و پنج بعد است.

همچنین پرس‌وجوهایی از نوع slice که یک ارزش از یک بعد و dice که دامنه‌ای از ارزش‌های یک یا چند بعد را انتخاب می‌کنند، می‌توانند پرس‌وجوهی مختلف ایجاد کنند.

بنابراین نحوه نوشتن پرس‌وجوها در این مدل تغییر داشته است. از این رو پرس‌وجوی یک، مجموع تخفیف‌ها را بر اساس جدول مکان جغرافیایی (یک بعد) گروه‌بندی می‌کند. پرس‌وجوی دو، مجموع فروش را بر اساس دو بعد مکان جغرافیایی و زمان گروه‌بندی می‌کند. پرس‌وجوی سه، مجموع فروش را بر اساس دو بعد مکان جغرافیایی و محصول گروه‌بندی می‌کند. پرس‌وجوی چهار، مجموع تخفیف‌ها را بر اساس سه بعد مکان جغرافیایی، محصول و زمان گروه‌بندی می‌کند. پرس‌وجوی پنج، مجموع تخفیف‌ها را بر اساس یک ارزش از بعد

در مدل سه‌لایه‌ای ارائه شده است و پرس‌وجوهی مربوط به مدل‌های ستاره‌ای، خوشه‌ستاره‌ای و دانه‌برفی ذکر نشده است. پرس‌وجوهی مطرح در مدل سه‌لایه‌ای بر اساس پنج بعد خرده‌فروش، محصول، مشتری، مکان جغرافیایی و زمان در گروه‌بندی‌های مختلف ایجاد شده است.



شکل ۴: شش نوع پرس‌وجو در مدل سه‌لایه‌ای با توجه به مکان جغرافیایی مشتری به همراه شکل گرافیکی لایه ستاره‌ای برای پایگاه داده تحلیلی سیستم فروش

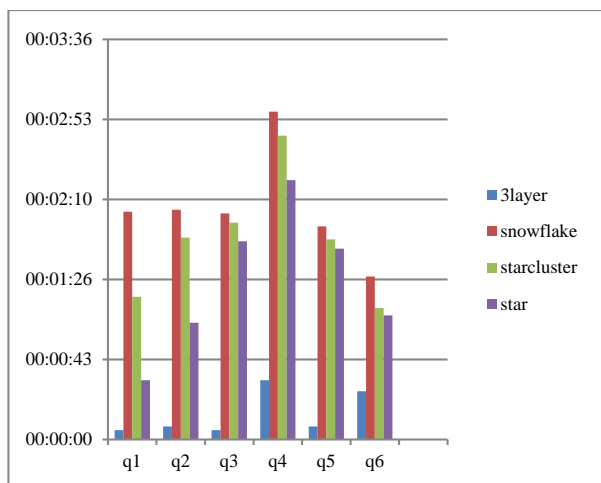
```
q5: select ProductKey,TimeKey,sum(TD)
from dbo.LAYER3.dbo.SUBDim
where dbo.LAYER3.[cus-G-id]=dbo.SUBDim. GeographyKey
and dbo.SUBDim.SalesTerritoryKey='7'
groupby ProductKey,TimeKey
orderby ProductKey,TimeKey asc
```

این پرس‌وجو به لایه سوم وارد می‌شود و مجموع تخفیف‌ها را بر اساس یک ارزش از بعد مکان جغرافیایی که قلمرو فروش آن معادل "۷" و برابر کشور فرانسه است، بر اساس بعد محصول و زمان گروه‌بندی می‌کند.

برای بررسی شرط از جدول زیر بعد لایه خوشه‌ستاره‌ای (نحوه ایجاد آن در مثال ۱ ذکر شد) که به‌عنوان جدول بعد جدید با کلید اصلی مکان جغرافیایی برای مشتری در لایه ستاره‌ای به جدول حقیقت آن وصل شده است، استفاده می‌کند.

بنابراین با توجه به پرس‌وجوهای مذکور در ارتباط با بعد جدید اضافه‌شده در لایه ستاره‌ای، زمان پاسخ به پرس‌وجوها در هر چهار مدل (دانه‌برفی، خوشه‌ستاره‌ای، ستاره‌ای و سه‌لایه‌ای) اندازه‌گیری می‌شود.

آزمایش‌ها بر اساس شش نوع پرس‌وجو، در چهار مدل دانه‌برفی، خوشه‌ستاره‌ای، ستاره‌ای و سه‌لایه‌ای با توجه به مکان جغرافیایی بعد مشتری انجام شده است و نتایج در شکل ۴ نشان داده شده است. طبق شکل ۴، مدل سه‌لایه‌ای سرعت پاسخ‌گویی بالاتری در هر پرس‌وجو نسبت به بقیه مدل‌ها دارد.



شکل ۴: نتایج آزمایش‌ها بر اساس شش نوع پرس‌وجو و با توجه به مکان جغرافیایی بعد مشتری

آزمایش‌های بعدی بر اساس شش نوع پرس‌وجو، در چهار مدل دانه‌برفی، خوشه‌ستاره‌ای، ستاره‌ای و سه‌لایه‌ای با توجه به مکان جغرافیایی بعد خرده‌فروش انجام شده است و نتایج در شکل ۵ نشان داده شده است. طبق شکل ۵، مدل سه‌لایه‌ای سرعت پاسخ‌گویی بالاتری در هر پرس‌وجو نسبت به بقیه مدل‌ها دارد.

مکان جغرافیایی که قلمرو فروش آن معادل "۷" و برابر کشور فرانسه است و بر اساس بعد محصول و زمان گروه‌بندی می‌کند. پرس‌وجوی شش، مجموع فروش را بر اساس محدوده‌ای از ارزش‌های بعد محصول که برابر با کد محصول‌های بیش‌تر از ۲۱۰ و همچنین برای بعد زمان که برابر کد زمان بیش‌تر از ۵۵۰ یعنی برابر سال‌های ۲۰۰۳ به بعد و بر اساس سه بعد (مکان جغرافیایی، محصول و زمان) گروه‌بندی می‌کند.

نقش لایه‌ها در هنگام پاسخ‌گویی به پرس‌وجوها مطرح می‌شود و بستگی به پرس‌وجویی دارد که مدیر بر اساس نیاز خود بیان می‌کند. به‌عنوان مثال، پرس‌وجوی ۱ بررسی می‌شود:

```
q1: select dbo.SUBDim.SalesTerritoryKey, sum(TD)
from dbo.LAYER3.dbo.SUBDim
where dbo.SUBDim.GeographyKey=dbo.LAYER3.[cus-G-id]
groupby dbo.SUBDim.SalesTerritoryKey
```

این پرس‌وجو به لایه سوم وارد می‌شود و مجموع تخفیف‌ها را حساب می‌کند. این پرس‌وجو برای اینکه بر اساس کد قلمرو فروش گروه‌بندی شود از جدول زیر بعد مکان جغرافیایی لایه خوشه‌ستاره‌ای استفاده می‌کند. این جدول از الحاق دو جدول به نام جدول بعد قلمرو فروش و جدول بعد مکان جغرافیایی که به‌صورت سلسله‌مراتبی در لایه مدل دانه‌برفی قرار دارد، ایجاد شده است. به این منظور الحاق دو جدول به‌صورت یک دید در می‌آیند و نهایتاً نتایج در جدول جدید زیر بعد مکان جغرافیایی وارد می‌شود.

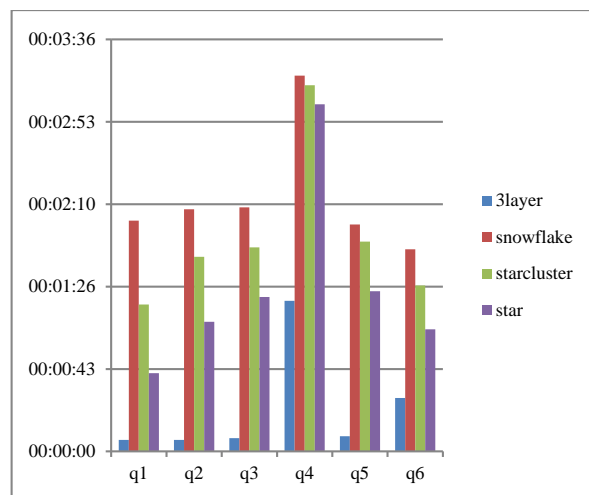
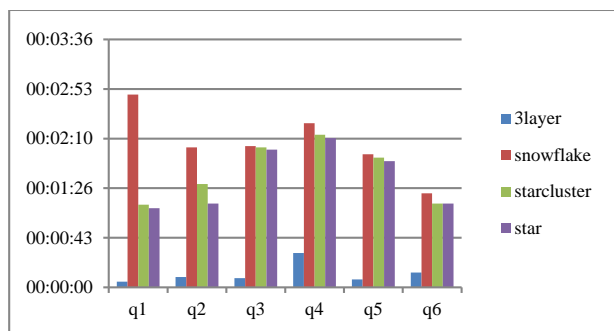
لازم به ذکر است که جدول زیر بعد لایه خوشه‌ستاره‌ای نیز به‌عنوان جدول بعد جدید با کلید اصلی مکان جغرافیایی برای مشتری و خرده‌فروش در لایه ستاره‌ای به جدول حقیقت آن وصل شده است. به این منظور جدول حقیقت لایه ستاره‌ای به‌صورت دید در می‌آید و نتایج در جدول جدید حقیقت لایه ستاره‌ای وارد می‌شود.

بنابراین مجموع تخفیف‌ها از جدول حقیقت لایه سوم (مدل ستاره‌ای) در شرایطی که مکان جغرافیایی آن بر اساس بعد مشتری باشد، حساب می‌شود.

در واقع مدل پیشنهادی (مدل سه‌لایه‌ای)، با ایجاد دیدهای ذخیره‌شده مناسب در لایه میانی (خوشه‌ستاره‌ای) و نهایتاً لایه ستاره‌ای، جهت پاسخ‌گویی به پرس‌وجوها ایجاد شده است. پرس‌وجوهای مطرح‌شده 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 تا پرس‌وجوهای slice از نوع slice که یک ارزش از یک بعد و dice که دامنه‌ای از ارزش‌های یک یا چند بعد را انتخاب می‌کنند، ایجاد کنند. همچنین پرس‌وجوها بر اساس تعداد ابعاد مختلف به‌صورت گروه‌بندی ایجاد شده‌اند.

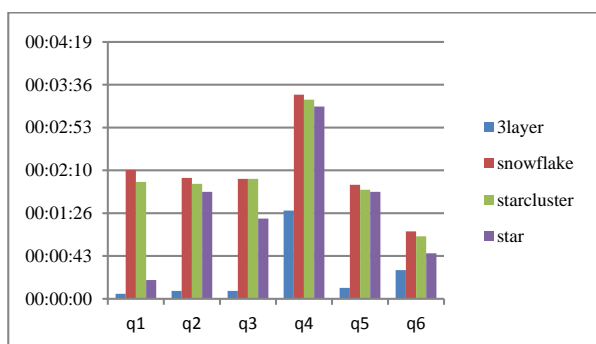
به‌طور مثال پرس‌وجوی ۵:

تأثیر دارد و در پرس‌وجوها، لایه پاسخ‌گو با توجه به جدول‌های بعدی که در پرس‌وجو مطرح شده‌اند، انتخاب می‌شوند. طبق شکل ۸، برای هر لایه شش نوع پرس‌وجو مطرح شده است و به صورت نمودار میله‌ای مقایسه شده‌اند. یعنی با افزایش تعداد جدول‌های الحاقی (ایجاد جدول‌های غیرنرمال) زمان پاسخ‌گویی به پرس‌وجوها بهبود می‌یابد.

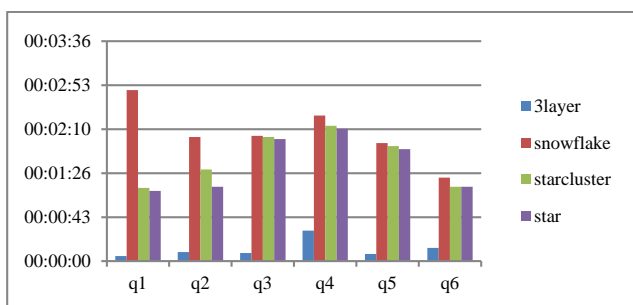


شکل ۵: نتایج آزمایش‌ها بر اساس شش نوع پرس‌وجو و با توجه به مکان جغرافیایی بعد خرده‌فروش

شکل ۶: نتایج آزمایش‌ها برای میانگین زمان پاسخ‌گویی بر اساس شش نوع پرس‌وجو و با توجه به مکان جغرافیایی بعد مشتری



شکل ۷: نتایج آزمایش‌ها برای میانگین زمان پاسخ‌گویی بر اساس شش نوع پرس‌وجو و با توجه به مکان جغرافیایی بعد خرده‌فروش



شکل ۸: نتایج آزمایش‌ها بر اساس تغییر تعداد جدول‌های الحاقی

در آزمایش بعدی تعداد رکوردهای جدول حقیقت تغییر داده شده‌اند و نتایج پرس‌وجوها بررسی شده است. انواع پرس‌وجوها، بر روی پایگاه داده تحلیلی فروش با ۳ سایز مختلف ۳۰۰۰۰۰، ۳۰۰۰۰۰۰ و ۳۰۰۰۰۰۰۰

با توجه به اینکه آزمایش‌های قبلی (شکل‌های ۴ و ۵) بر روی یک پرس‌وجوی خاص انجام گرفته بود، در آزمایش‌های بعدی (شکل‌های ۶ و ۷) برای هر نوع پرس‌وجو، تعدادی پرس‌وجو نوشته شد و به مدل‌های ستاره‌ای، خوشه‌ستاره‌ای، دانه‌برفی و سه‌لایه‌ای اعمال شد. میانگین زمان پاسخ‌گویی در مدل‌ها اندازه‌گیری شد که همچنان نتایج یکسان است. بنابراین آزمایش‌های بعدی بر اساس میانگین زمان پاسخ‌گویی شش نوع پرس‌وجو در مدل‌ها انجام شده است. نتایج این آزمایش‌ها در شکل‌های ۶ و ۷ نشان داده شده است. همان‌طور که مشاهده می‌شود، مدل سه‌لایه‌ای نسبت به مدل‌های دیگر زمان پاسخ به پرس‌وجوها را بهبود داده است.

آزمایش بعدی بر اساس تعداد جدول‌های الحاقی صورت می‌گیرد. به این صورت که در لایه اول رابطه بین جدول‌ها به صورت سلسله‌مراتبی است، که همان مدل دانه‌برفی را تشکیل می‌دهد و هیچ جدولی با هم پیوند نخورده است و آزمایش‌های انجام‌شده در مدل دانه‌برفی لایه اول سرعت پاسخ‌گویی پایینی را نشان می‌دهد و در لایه دوم مدل خوشه‌ستاره‌ای مطرح شد که یک مورد بین جدول قلمرو فروش و جدول مکان جغرافیایی الحاقی صورت گرفت و مورد دیگر، جدول محصول با سلسله‌مراتب مربوط به خودش پیوند خوردند و آزمایش‌ها نشان داد سرعت پاسخ‌گویی مدل خوشه‌ستاره‌ای از مدل دانه‌برفی لایه اول بهتر است. نهایتاً لایه سوم، مدل ستاره‌ای است، به این صورت است که جدول مکان جغرافیایی که در هر دو بعد مشتری و خرده‌فروش مشترک است به شکل جدول جدا به نام جدول بعد مکان جغرافیایی قرار گرفت و این عمل سرعت پاسخ‌گویی آن را نسبت به دو لایه دیگر بهبود داد.

بنابراین در شکل ۸ سرعت پاسخ‌گویی به پرس‌وجوها با تغییر تعداد جدول‌های الحاقی نشان داده شده است. همان‌طور که مشاهده می‌شود تعداد جدول‌های الحاقی در زمان پاسخ‌گویی به پرس‌وجوها



گاهی با توجه به وضعیت ممکن است داده‌ها از چند پایگاه داده نرمال شده استخراج شوند و در یک منبع غیرنرمال قرار گیرد. این روش برای مخزن داده پایگاه داده تحلیلی استاندارد خوبی است زیرا هدف اصلی از ایجاد پایگاه داده تحلیلی، افزایش سرعت پاسخ‌گویی به پرس‌وجوها است و از آنجایی که درج و تغییر اطلاعات در آن توسط کاربران عادی انجام نمی‌شود بلکه توسط مدیر سیستم و در واقع طراح و ایجادکننده آن صورت می‌گیرد، افزونگی اطلاعات منجر به تناقض نخواهد شد. از طرف دیگر این افزونگی‌ها در سطح ابعاد صورت می‌گیرد که حجم آن‌ها در مقابل حجم جدول حقیقت ناچیز است و بنابراین افزایش حجمی که ایجاد می‌کند نیز تأثیری در کل نخواهد داشت.

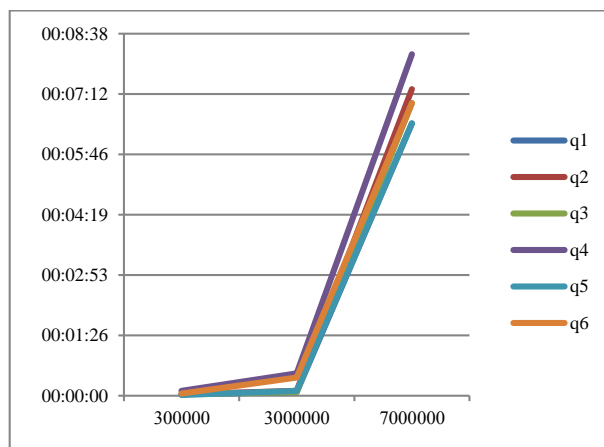
### ۵- نتیجه‌گیری و کارهای آینده

فرآیند 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 سطح منطقی پایگاه داده تحلیلی بررسی شد. سپس مدلی بهبودیافته در این سطح، برای فرآیند طراحی پایگاه داده تحلیلی، ارائه شد که ترکیبی از سه مدل دانه‌برفی، خوشه‌ستاره‌ای و ستاره‌ای، به صورت سه‌لایه‌ای است. به منظور مقایسه مدل پیشنهادی با مدل‌های موجود، از معیار زمان پاسخ‌گویی به پرس‌وجوها استفاده شده است. آزمایش‌ها نشان می‌دهد که مدل پیشنهادی منجر به بهبود زمان پاسخ به پرس‌وجوها می‌شود. همچنین مدل پیشنهادی (مدل سه‌لایه‌ای)، خصوصیت‌های کارایی، قابلیت استفاده، قابلیت استفاده مجدد، انعطاف‌پذیری و اعتبار را در سطح بالا دارا است. مدل سه‌لایه‌ای 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 خصوصیت‌های مطرح در آن‌ها کامل‌تر است.

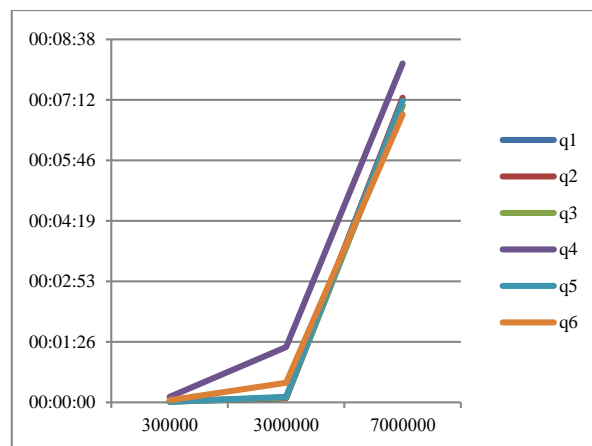
### مراجع

- [1] R. Jindal and Sh. Taneja, "Comparative study of data warehouse design approaches: a survey," *International Journal of Database Management Systems (IJDBMS)*, vol. 4, no. 1, 2012.
- [۲] فاطمه پیروزمند، نعمت‌آقا قهرمانی و محمدرضا عاروان، «طراحی کنترل‌کننده پیش‌بین مقاوم با استفاده از نامساوی‌های ماتریسی خطی برای سیستم کنترل وضعیت ماهواره»، *مجله مهندسی برق*، دوره ۴۴، دانشگاه تبریز، ۱۳۹۳.
- [۳] حسین مرادی فراهانی و جواد عسگری، «طراحی کنترل‌کننده عصبی-فازی نوع-۲»، *مجله مهندسی برق*، دوره ۴۳، دانشگاه تبریز، ۱۳۹۲.
- [۴] سعید ابادری، مجتبی برخوردار یزدی و عباس عرب دردردی، «طراحی کنترل‌کننده مقاوم SVC مبتنی بر WAMS با در نظر

۷۰۰۰۰۰ رکورد در جدول حقیقت db.LAYER3 اجرا شدند و نتایج آن برحسب زمان پاسخ‌گویی در شکل‌های ۹ و ۱۰ آورده شده است. همان‌طور که مشاهده می‌شود با افزایش حجم داده زمان پاسخ‌گویی به پرس‌وجوها افزایش می‌یابد. بدیهی است با افزودن دیدهای ذخیره‌شده مناسب جهت پاسخ‌گویی به پرس‌وجوها، نرخ این افزایش زمان نیز به میزان قابل توجهی کاهش می‌یابد.



شکل ۹: نتایج آزمایش بر اساس شش نوع پرس‌وجو و با توجه به مکان جغرافیایی بعد مشتری



شکل ۱۰: نتایج آزمایش بر اساس شش نوع پرس‌وجو و با توجه به مکان جغرافیایی بعد خرده‌فروش

رویکرد غیرنرمال مشکلاتی دارد، که از مهم‌ترین آن‌ها افزونگی است. افزونگی مشکلات زیر را می‌تواند ایجاد کند: هدر رفتن فضای ذخیره‌سازی، ایجاد تناقض در هنگام بهنگام‌سازی، حذف و درج. نرمال‌سازی تکنیک مهمی برای طراحی پایگاه داده‌های کارآمد است اما در ضمنی که افزونگی داده را کاهش می‌دهد سبب افزایش زمان اجرای پرس‌وجوها نیز می‌شود.

- [11] 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.
- [12] N. Arfaoui and J. Akaichi, "Data warehouse: conceptual and logical schema – survey," *International Journal of Enterprise Computing and Business Systems*, vol. 2, pp. 1-31, 2012.
- [13] V. Peralta and R. Ruggia, "Using design guidelines to improve data warehouse logical design," *Proc. of the International Workshop on Design and Management of Data Warehouses*, 2003.
- [14] V. Peralta, *Data Warehouse Logical Design from Multidimensional Conceptual Schemas*, Instituto de Computación, Universidad de la República, Julio HerrerayReisig 565, Montevideo, CP 11300, Uruguay, 2003.
- [15] M. H. Peyravi, "A schema selection framework for data warehouse design," *International Journal of Machine Learning and Computing*, vol. 2, no. 3, 2012.
- [16] M. Shekofteh, "Building quick and accurate structure for data warehouse," *International Journal of Advanced Computer Science*, vol. 2, no. 7, pp. 274-276, 2012.
- [17] 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*, pp. 198-204, 2012.
- گرفتن نامعینی تأخیر زمانی سیگنال‌های راه دور»، *مجله مهندسی برق*، دوره ۴۵، دانشگاه تبریز، ۱۳۹۴.
- [5] B. P. Başaran, *A Comparison of Data Warehouse Design Models*, M.Sc. Thesis, Computer Engineering, Atilim University, Turkey, 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," *In Proceedings of International Workshop on Design and Management of Data Warehouses (DMDW'2000)*, pp. 5-1, 5-12, 2000.
- [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*, pp. 1-12, 2008.
- [9] D. Mishra, A. Yazici, B. P. Başaran, "A case study of data models in data warehousing design," *Applications of Digital Information and Web Technologies*, pp. 314-319, 2008.
- [10] S. Nugawela, *Data Warehousing Model for Integrating Fragmented Electronic Health Records from Disparate and Heterogeneous Clinical Data Stores*, M.Sc. Thesis, School of Electrical Engineering and Computer Science Faculty of Science and Engineering Queensland University of Technology, 2013.

## زیرنویس‌ها

- <sup>1</sup> Data Warehouse
- <sup>2</sup> Subject-oriented
- <sup>3</sup> integrated
- <sup>4</sup> time-variant
- <sup>5</sup> non-volatile
- <sup>6</sup> Flat schema
- <sup>7</sup> Terraced schema
- <sup>8</sup> Star schema
- <sup>9</sup> Constellation schema
- <sup>10</sup> Galaxy schema
- <sup>11</sup> Snowflake schema
- <sup>12</sup> Star flake schema
- <sup>13</sup> Star cluster schema
- <sup>14</sup> Efficiency
- <sup>15</sup> Usability
- <sup>16</sup> Reusability
- <sup>17</sup> Flexibility
- <sup>18</sup> Redundancy
- <sup>19</sup> complexity
- <sup>20</sup> Reliability