تقلب در سازمان‌های بیمه و بانک باعث از دست رفتن سالیانه میلیاردها دلار میشود لذا صاحبان کسب وکار بایستی راه حل‌ها و یا سیستم‌های مناسبی را جهت کشف تقلب‌ها در نظر بگیرند. امروزه با توجه به گسترش اطلاعات و ارتباطات روابط بین سازمان‌ها، برای ذخیره‌سازی اطلاعات شبکه‌ای‌ نیاز به پایگاه‌داده‌های گرافی می‌باشد. مناسب‌ترین پایگاه‌داده‌ی گرافی پایگاه‌داده neo4j است که یکی از کاربردهای این پایگاه‌داده، کشف تقلب سازمان‌ها می‌باشد. همچنین می‌توان با بررسی دقیق روابط بین موجودیت‌ها الگوی مناسبی را جهت کشف تقلب در نظر گرفت تا آنچه مطابق با این الگو است از درون پایگاه‌داده‌گرافی بازیابی شود. به عنوان مثال در شکل الگوی سمت چپ را داریم و آنچه مطابق با این الگو است را از پایگاه داده استخراج میکنیم.

در این بخش به تشریح تعاریف ومفاهیم پرکاربردی که در این کار به کرات از آن‌ها استفاده خواهد شد می‌پردازیم. انواع پایگاه‌داده‌های رابطه‌ای و غیررابطه‌ای معرفی شده، معماری انواع پایگاه‌های غیررابطه‌ای بررسی می‌شود. یکی از انواع پایگاه‌داده‌های غیررابطه‌ای به نام پایگاه‌داده گرافی که در این کار تمرکز ما بر روی این نوع پایگاه داده خواهد بود، بررسی شده است.

* 1. مدل پایگاه داده

اصطلاح "مدل داده" به طور گستردهای در جامعه مدیریت اطلاعات مورد استفاده قرار گرفته است. مدل داده معانی مختلف را پوشش می‌دهد. به طور کلی، یک مدل داده (پایه) مجموعه‌ای از ابزارهای مفهومی است که برای مدل سازی موجودیت‌های جهان واقعی و روابط میان آنها استفاده میشود. مدل پایگاه‌داده از سه جز تشکیل شده است که عبارتند از مجموعه‌ای از انواع ساختارداده‌ها، مجموعه‌ای

از عملگرها جهت استنتاج قوانین و مجموعه‌ای از قوانین یکپارچه. با توجه به اهمیت فلسفی و عملی

مدلسازی مفهومی، مدل‌های پایگاه‌داده به ابزاری ضروری تبدیل شده‌اند. استفاده از انواع مدل پایگاه داده در موارد زیر کاربردی است:

* برای مشخص کردن نوع داده‌های مجاز
* پشتیبانی از روشی عمومی جهت طراحی پایگاه‌داده
* مقابله با تکامل پایگاه‌داده
* ایجاد خانواده زبان‌های سطح بالا برای پرسوجو و دستکاری داده‌ها
* تمرکز بر معماری سیستم‌های مدیریت پایگاه‌داده
* محرکی برای تحقیق در مورد خواص رفتاری ساختارهای اطلاعات متناظر
  1. انواع پایگاه داده

پایگاه‌داده‌ها در دونوع رابطه‌ای و غیر رابطه‌ای [1] دسته بندی می‌شوند. امروزه با گسترش ارتباطات، ذخیره‌سازی و بازیابی داده‌ها واطلاعات در پایگاه‌داده‌های رابطه‌ای، بدلیل تاثیر عواملی چون حجم زیاد، تنوع، سرعت تغییر بالا، دقت اطلاعات و یک سری عوامل دیگر، امری هزینه بر است. مشکل پایگاه‌داده‌های رابطه‌ای این است که هنگام پیوست جداول در پرسوجوها، مجموعه بزرگی ازاطلاعات بدون فیلترکردن نمایش داده می‌شوند. درنتیجه خیلی کند می‌باشند و پیچیدگی زمانی بالایی دارند. از طرفی، داده‌ها با گذشت زمان تغییر میکنند و تغییرات داخلی و خارجی یک سیستم و زمینه‌ای که در آن کار می‌شود، می‌تواند تاثیر قابل‌توجهی را در سرعت داشته باشد. عامل دیگر که باعث عدم عملکرد بهینه و مناسب پایگاه‌داده‌های رابطه‌ای شده است، تنوع در ساختارداده‌ها است. تنوع، به درجه نظم یا بی‌نظمی، تراکم یا عدم تراکم و اتصال یا عدم اتصال داده‌ها مربوط می‌شود. با این وجود نکته این است که اطلاعات در طبیعت لزوما به صورت جدولی نیست. فرض کنید که در یک رابطه دوستانه در یک شبکه اجتماعی بخواهیم دوستان دوستان دوستان را بازیابی کنیم. چنین پرس وجویی در پایگاه داده‌ی رابطه‌ای حجم عظیمی کد ایجاد کرده(پیچیدگی محاسباتی وفضایی دارد) و انسجام را از بین میبرد. لذا زمان انتقال به یک پایگاه‌داده‌ی غیررابطه‌ای فرارسیده است. از انواع پایگاه‌داده‌های غیررابطه‌ای میتوان به پایگاه‌داده‌های سندمحور، کلید مقدار، ستونگرا وگرافی اشاره کرد که هریک معماری و ویژگی عملیاتی متنوعی دارند و متناسب با منطق کسب وکار مورد استفاده قرار میگیرند. درپایگاه داده‌های سندگرا اسناد به صورت جفت کلید\_مقدار ذخیره می‌شوند. این مقادیر همانند فرمت‌هایی از جمله JSON و XML می‌توانند به صورت لیست یا نقشه باشند. اسناد می‌توانند با شناسه، ذخیره و بازیابی شوند. به عبارت بهتر یک پایگاه‌داده سندگرا، میتواند مشابه پایگاه‌داده کلید مقدار عمل کند. در موارد کلی، پایگاه‌داده‌های سند محور به شاخص‌ها می‌پردازند تا دسترسی به اسناد را براساس هریک از ویژگی‌های آنها سرعت بخشند. در این صورت با نوشتن شاخص داده، میتوان به جزئیات داده‌ها دسترسی داشت. درپایگاه‌داده‌های کلید\_مقدار، مقادیر مبهم ذخیره شده، با کلید بازیابی میشوند. هرداده‌ای با درهمسازی شناسه (کلید)، ذخیره می‌شود. با فرض این که چند باکت و چند ماشین داشته باشیم، تابع درهمساز طوری ساخته میشود که توزیع یکنواخت در باکت در دسترس را فراهم کند. به طوری که هیچ ماشین واحد به یک نقطه کانونی تبدیل نشود. با توجه به کلید، می‌توان از این آدرس برای ذخیره مقدار در باکت مربوط استفاده کرد. پایگاه‌داده‌ی ستونگرا، پایگاه‌داده‌ای است که مدل داده‌ای آن بر پایه جدولی با جمعیت کم است. یک ردیف می‌تواند شامل هر تعداد ستون دلخواه باشد. ساده‌ترین واحد ذخیره‌سازی ستون است که شامل یک جفت نام و مقدار می‌باشد. ترکیب هر تعداد ستون را ابرستون گویند. هرابرستون یک نام مختص به خود را دارد. اگر چندین ستون کنار هم در یک ردیف ذخیره شوند، به آن خانواده ستون گویند. اگرچندین ابرستون دریک ردیف قرارگیرند، به آن خانواده ابر ستون گویند. پایگاه‌داده‌های گراف به عنوان معروفترین ابزار ذخیره سازی و پرسوجو در گراف‌ها هستند. پایگاه‌داده‌های گراف برای حل اطلاعات پیچیده و مشکلات تجاری حائز اهمیت هستند. گراف‌ها یک مدل طبیعی برای طیف وسیعی از حوزه‌ها هستند. به عنوان نمونه برای توصیف نحوه عبور کاربر از نقطه A به نقطه B پایگاه داده‌گرافی راه حل مناسبی است. [2] انتخاب یک پایگاه‌داده غیررابطه‌ای وجایگزینی احتمالی آن با یک پایگاه‌داده رابطه‌ای، به دلیل عدم همسویی مناسب با نیازهای موجود یکی از اتفاقات مهمی است که این روزها در بسیاری از سازمان‌ها در حال وقوع است.