



عنوان مقاله: مروری بر Data Warehouse و رویکرد آن در هوش تجاری  
تهیه کننده/گان: فاطمه احمدنژاد دیسفانی مدرک و رشته تحصیلی: کارشناسی ارشد مهندسی کامپیوتر (نرم افزار)  
رشته شغلی فناوری اطلاعات اداره کل/دفتر: مدیریت سامانه های عملیات شبکه و کسب و کار  
عنوان حوزه تحقیقاتی مورد نیاز شرکت: الزامات بهره گیری از پایگاه داده تحلیلی (Data Warehouse) به جهت توسعه  
دانشوردهای گزارش ساز مدیریتی شماره ردیف حوزه تحقیقاتی مورد نیاز شرکت: ۲۴

## عنوان: مروری بر Data Warehouse و رویکرد آن در هوش تجاری

چکیده: هدف اصلی Data Warehouse نگهداری حجم وسیعی از تاریخچه داده‌ها در قالبی است که تحلیل آن را برای حمایت از تصمیم گیری به موقع و بهتر در هر دو سطح عملیاتی و راهبردی تسهیل می‌کند. در انبارهای داده مبتنی بر ابر، کاربران تجاری می‌توانند از طریق منابع و مکان‌های پراکنده از نظر جغرافیایی مختلف به داده‌ها دسترسی پیدا کنند. تحلیل‌گران و تصمیم‌گیرندگان تجاری مشخصاً برای تحلیل و گزارش داده‌ها به انبارهای داده متکی هستند. دو عامل داده‌های زمانی و فضایی بشدت بر استراتژی‌های تصمیم‌گیری و بازاریابی اثر می‌گذارند و بسیاری از برنامه‌های کاربردی به مدلسازی و پردازش ویژه این نوع داده‌ها نیاز دارند، چون در پایگاه داده‌های چند بعدی متداول نمی‌توان به‌صورت بهینه آنها را پردازش کرد. صنعت مخابرات، که به‌تدریج با حجم عظیم داده‌ها تحت‌الشعاع قرار می‌گیرد، یکی از زمینه‌های کاربرد انبارداری داده‌های فضایی‌زمانی است.

کلید واژه: Data Warehouse، داده، هوش تجاری، پردازش تحلیلی

### مقدمه

هدف اصلی انبار داده (DW) نگهداری حجم وسیعی از تاریخچه داده‌ها در قالبی است که تحلیل آن را برای حمایت از تصمیم‌گیری به موقع و بهتر در هر دو سطح عملیاتی و راهبردی تسهیل می‌کند. با توسعه سریع اینترنت جامعه مدرن وارد عصر اطلاعات شده است. ما باید، با توجه به جنبه‌های مختلف روش‌های علمی، اطلاعات و داده‌های تجاری فراوانی را جمع‌آوری کنیم و از دیدگاه‌های مختلف آن‌ها را تحلیل کنیم؛ بدین جهت اثرگذاری این فعالیت‌ها از اهمیت بیشتری برخوردار شده است. سیستم انبار داده (DW)، به دلیل اثرگذاری بسیارش، در عملیات خود از این تکنیک‌ها استفاده می‌کند.

انبار داده نوعی راه‌حل است، روند پردازشی است که در آن داده‌های عملیاتی اصلی پردازش می‌شوند و به اطلاعات مفید تبدیل می‌شوند. کاربران می‌توانند از طریق تجزیه و تحلیل این اطلاعات تصمیمات استراتژیک بگیرند. بنابراین، در بسیاری از موقعیت‌ها به انبار داده «سیستم پشتیبانی تصمیم (DSS)» نیز می‌گویند [۱].

با توجه به ناهمگنی و نوسان عملکردهای کسب و کار و روابط میان آنها و همچنین اهداف مطابقت داده نشده بخش‌های مختلف یک سازمان، انبار داده (DW) بایستی نیازمندیهای موردنیاز برای حمایت کافی از فرایندهای تصمیم‌گیری، برنامه‌ریزی و هماهنگ‌سازی در سیستم کسب و کاری خاص را محقق نماید. از این رو باید انبار داده‌ها بوضوح به سمت نیازهای سیستم کسب و کار مورد نظر جهت دهی شود یعنی بتواند بصورت شفاف و بدون ابهام موقعیت سیستم را در نقطه یا دوره زمانی خاص بازتابی نماید. علاوه بر این انبار داده‌ها باید توالی از موقعیت‌های سیستم را در بر گیرد، یعنی بازه زمانی را تعیین کند تا از یک سو امکان تحلیل و مقایسه حالت‌های قبلی فراهم شود و از سوی دیگر پیش‌بینی و برنامه‌ریزی حالت‌های آتی سیستم را بتوان انجام داد. در نهایت برای در نظر گرفتن ناهمگنی منبع داده‌ها از لحاظ تکنولوژیکی و کسب و کار باید انبار داده بتواند داده‌ها را با یکدیگر ادغام کند که در این حالت ادغام باید به روشی انجام شود که داده‌های گردآوری شده تغییر معناداری نکنند [۲].

هوش تجاری فرایندی فناوری‌محور برای جمع‌آوری، یکپارچه‌سازی، تحلیل، و ارائه اطلاعات تجاری است. هوش تجاری شامل دامنه وسیعی از ابزارها، برنامه‌های کاربردی و روش‌هاست که به سازمان‌ها کمک می‌کنند داده‌ها را از منابع داخلی و خارجی جهت تحلیل و تصمیم‌گیری جمع‌آوری کنند [۳].

انبارهای داده را می‌توان به عنوان پایگاه‌داده‌هایی چندبعدی حاوی مقادیر عظیمی داده در نظر گرفت که در آنجا داده‌ها به عنوان مجموعه‌ای از جداول ابعاد و حقایق ساختار بندی شده‌اند. حقایق در واقع رویدادهایی هستند که در یک شرکت روی می‌دهند و

مستلزم فرآیند تصمیم‌گیری هستند. ابعاد ویژگی‌هایی هستند که حقیقتی بخصوص را توصیف می‌کنند. ابعاد داده‌های نسبتاً ایستای انبار داده هستند [۳].

حجم‌های عظیم داده به سرعت صنعت مخابرات را تحت‌الشعاع قرار می‌دهند. مشخصاً شرکت‌های مخابراتی برای نشان دادن تمامی داده‌های مربوط به تماس‌ها، قبض‌ها، و مشتریان به ده‌ها جدول رابطه‌ای نیاز دارند. برای هر رویداد کاربر بر روی شبکه مخابرات، یک رکورد جزئیات تماس تولید می‌شود که حاوی طیف گسترده‌ای از اطلاعات مربوط به هر تماس است؛ این اطلاعات شامل شماره تلفن هر دو طرف تماس، زمان شروع تماس، مدت زمان آن، و همچنین اطلاعات خاص‌تری از جمله شماره تلفنی که هزینه تماس را باید بپردازد، هزینه تماس، کل زمان استفاده در دوره قبض، کل هزینه پرداختی طی دوره قبض، نوع تماس (صوتی، پیامک)، شناسه تجهیزات مخابراتی که رکورد را می‌نویسد و بسیاری اطلاعات دیگر هستند. تمامی اطلاعات مختلفی که رکورد جزئیات تماس، ذخیره می‌کند را می‌توان به منظور کشف الگوهای مربوط به رفتار تماس و کاربرد اصلی سرویس استخراج کرد.

بدون شک تجلیل جزئیات تماس یکی از جذاب‌ترین حوزه‌ها برای کندوکاش در صنعت مخابرات است. داده‌های زمانی و فضایی ضروری‌ترین مولفه‌ها برای این کندوکاش هستند. شرکت‌های مخابراتی باید بدانند که مردم چه زمانی دوست دارند بیشتر صحبت کنند یا چه زمانی بیشتر تماس می‌گیرند؛ به عنوان مثال، به نظر می‌رسد که مردم معمولاً عصرها به مدت زمان طولانی‌تری صحبت می‌کنند اما در طول روز تماس‌های تلفنی بیشتری دارند. ویژگی‌های جغرافیایی فضایی رکورد جزئیات تماس برای تحلیل رفتار اجتماعی یا برنامه‌ریزی شهری منطقه‌ای جغرافیایی هم مهم است. شرکت‌ها با استفاده از این اطلاعات می‌توانند محصولاتی را که برای مشتریان‌شان جذاب‌تر هستند ارتقا ببخشند و توان رقابتی‌شان را بالا ببرند. برای این منظور می‌توان انبارهای داده را طراحی و تولید کرد و به کار گرفت [۳].

## ۱- ویژگی‌های انبار داده

انبار داده معمولاً به عنوان مجموعه‌ای از داده توصیف می‌شود که موضوع‌گرا، یکپارچه، نسبتاً پایدار، متغییر با زمان، زمانمند و غیرقابل‌نظرسنجی کردن است. انبار داده اساساً ویژگی‌های زیر را دارد:

۱. موضوع‌گرا: در مقایسه با ویژگی کاربرد محور پایگاه داده متداول، انبار داده بر تجزیه و تحلیل داده یا به عبارتی داده‌ای که مطابق با موضوع ذخیره می‌شود، تمرکز می‌کند؛ از این رو داده آن موضوع محور است. موضوع را می‌توان استاندارد برای طبقه‌بندی داده در سطوح بالاتر در نظر گرفت که هر موضوعی با تجزیه و تحلیل جامع حوزه مرتبط با آن مطابقت دارد.

۲. یکپارچگی: داده موجود در انبار داده از داده غیرمتمرکز استخراج می‌شود و سپس پردازش، جمع‌آوری و بازنگری روی آن اعمال می‌شود. برای یکپارچگی داده در انبار داده باید ناهماهنگی در داده اصلی از بین برود.

۳. زمانمندی: برای هدف تصمیم‌گیری، داده‌های انبار داده باید ویژگی زمان را داشته باشند. انبار داده مجموعه‌ای از داده‌های متغیر با زمان است. محدودیت زمانی ذخیره داده‌ها در انبار داده باید مقتضیات تصمیم‌گیری را برآورده سازد.

۴. پایداری نسبی: داده در پایگاه داده عملیاتی معمولاً باید مطابق با مقتضیات واقعی به‌روزرسانی شود. داده در انبار داده عموماً برای تجزیه و تحلیل استفاده می‌شود. عملکرد داده عموماً به معنی درخواست داده است و عملیات تغییر و پاکسازی به ندرت صورت می‌گیرد. همین که داده وارد انبار داده می‌شود، به مدتی طولانی حفظ می‌شود و تنها به بارگذاری و بازخوانی دوره‌ای نیاز دارد [۱].

## ۲- پرس و جوی در انبار داده

انبار داده با سیستم مدیریت اطلاعات تفاوت دارد، زیرا داده‌های مبتنی بر تاریخ را ذخیره می‌کند و عموماً هیچ تغییری در آن ایجاد نمی‌کند، بنابراین، از انبار داده اساساً به منظور پرس‌جو کردن استفاده می‌کنند. همچنین، بین پرس‌جوی سیستم مدیریت اطلاعات و پرس‌جوی انبار داده تفاوت عمده‌ای وجود دارد. هرچه استخراج‌کننده‌های داده به اعماق پایگاه‌های انبار داده بیشتر دسترسی پیدا کنند، نیاز به انواع داده پیچیده به پرس‌جوهای پیچیده‌تر تغییر پیدا می‌کند.

پرس‌جوی سیستم مدیریت داده معمولاً خیلی ساده است، معمولاً عملیات مرتبط‌کننده جداول وجود ندارد، میزان اطلاعات در هر دوره خیلی کم است، مسئله در این جا «دانستن کاری که باید انجام شود» است، مانند موجودی گرفتن از حساب. عملیات بر مبنای شاخص (Index) است، برای مثال شماره حساب می‌تواند به عنوان شاخص سیستم پس‌انداز استفاده شود. به دلیل این ویژگی‌ها، اندازه پایگاه داده تأثیر اندکی بر عملکرد سیستم مدیریت اطلاعات دارد.

پرس‌وجوی سیستم انبار داده معمولاً بسیار پیچیده است، معمولاً دو راه برای پرس‌وجو وجود دارد: یکی از راه‌ها، پرس‌وجو درباره انواع مختلف گزارش‌های کاری تولیدشده در پایگاه داده است. این نوع پرس‌وجو به خوبی از پیش برنامه‌ریزی شده است، راه دیگر پرس‌وجوی تصادفی و پویاست که در آن نتایج پرس‌وجو پیش‌بینی نمی‌شوند.

پرس‌وجوی انبار داده معمولاً این موارد را شامل می‌شود: ارتباط، گردآوری، طبقه‌بندی، مرتب کردن عملیات چندین جدول که معمولاً کل جدول را جست‌وجو می‌کنند. میزان اطلاعات برای هر پرس‌وجو زیاد است، برای پرس‌وجوی پویا بنابر نتایج قبلی جست‌وجوی بیشتری باید انجام شود، که به این فرایند استخراج داده می‌گویند [۱].

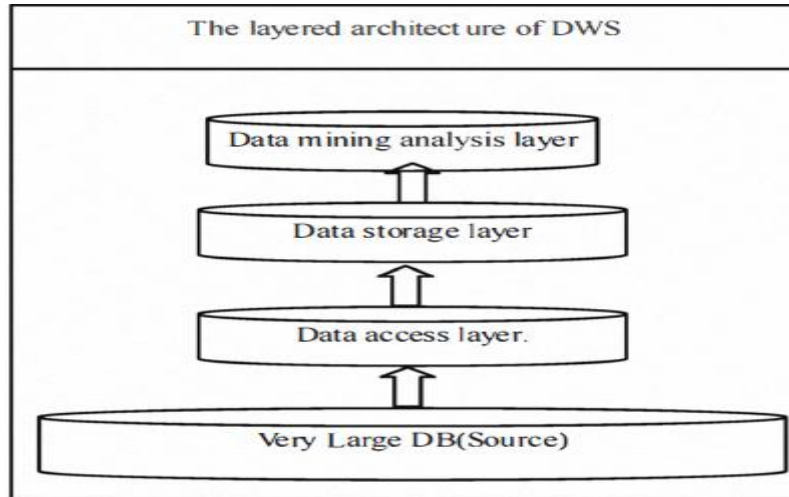
### ۳- مدیریت متادیتا در انبار داده

متادیتا معمولاً به صورت «داده درباره داده» توصیف می‌شود. در پایگاه داده، متادیتا توصیف چیزهای مختلف در پایگاه داده است؛ در پایگاه داده‌ی نسبی این نوع توصیف شرح جدول، ستون، پایگاه داده، دیدگاه و دیگر چیزها را شامل می‌شود. در کل، متادیتا نشان‌دهنده‌ی توصیف هر چیزی در انبار داده است. توصیف جامع متادیتا می‌تواند به تکمیل توصیف تمام چیزهای معنادار در پایگاه داده کمک کند. فهمیدن این توصیفات برای توسعه‌ی تمام جنبه‌های انبار داده، از توسعه‌ی روندهای استخراج داده گرفته تا انتقال داده در چندین انبار داده برای حفظ داده در انبار داده مطابق فرمت کلی، ضروری است.

در کل، متادیتا برای موارد و محتوای ذیل در انبار داده توصیف و توضیح ارائه می‌کند:

- توصیف سرور، پایگاه داده، جدول در انبار داده و نام‌ها و مکان‌هایشان
- توصیف قوانین تجزیه‌اتوماتیک و سلسله‌مراتب ابعادی کار، مانند محصول، بازار، فهرست حساب
- توصیف امنیت شخصی، کاری و سازمانی
- توصیف تغییر داده اصلی
- توصیف نحوه دسترسی
- توصیف نگاشت جدول‌های انبار داده
- توصیف استاندارد ترکیب
- توصیف هدر داده و اطلاعات مشخص شده به وسیله کاربر نهایی
- قوانین معمول محاسبات تعیین شده به وسیله کاربر نهایی
- توصیف جداول انبار داده و ویژگی‌هایشان
- توصیف فیزیکی جداول، ستون‌ها و ویژگی‌هایشان
- توصیف تاریخچه داده‌های استخراج شده
- توصیف مالک

معماری سیستم برای تمام سیستم‌های تصمیم‌گیری لازم است، از جمله انبار داده. شکل ۱ ساختار کلی سیستم انبار داده را نشان می‌دهد.



شکل ۱- معماری لایه ای سیستم انبار داده

لایه تجزیه و تحلیل استخراج داده: لایه استخراج داده شامل الگوریتم‌ها و ابزار استخراج داده مختلف است که می‌تواند مدل‌های وظیفه و فرم‌های سازمانی کاربردی و قابل تنظیمی ایجاد کند تا از استخراج داده در تصمیم‌گیری پشتیبانی کند. لایه دسترسی به داده: در لایه دسترسی به داده، داده مورد نیاز برای موضوعات تصمیم‌گیری از انواع مختلفی از پایگاه‌های کاری یا فایل‌های داده و دیگر منابع داده خارجی استخراج می‌شود، سپس پاکسازی، یکپارچه‌سازی و فرایندهای تبدیل ضروری انجام می‌شود. سپس این داده در انبار ذخیره می‌شود. لایه دسترسی به داده نقش مهمی در سیستم انبار داده ایفا می‌کند. لایه ذخیره‌سازی داده: لایه ذخیره‌سازی داده، داده‌های موضوعی مختلف را با ساختار سازمانی خاصی ذخیره می‌کند. انبار داده موضوعات مختلفی را شامل می‌شود [۱].

#### ۵- طرح های انبار داده

مدل داده مشکل اصلی انبار داده است. انبار داده دارای سه طرح جهت مدل‌سازی است: مدل ستاره‌ای، مدل برف ریزه و مدل ترکیبی [۱].

**مدل ستاره‌ای:** مدل ستاره‌ای الگوی مدل‌سازی‌ای است که به سمت بیرون منشعب می‌شود، یک موضوع در مرکز، چندین موضوع را به هم متصل می‌کند. موضوع مرکزی ستاره «جدول واقعیات» نام دارد. موضوع متصل شده «جدول ابعاد» نام دارد. پرس‌وجو از جدول واقعیات به منظور رسیدن به جدول نشانگر — که به جدول ابعاد اشاره دارد — انجام می‌شود. هنگام بازیابی چندین جدول ابعاد به طور همزمان، اطلاعات بسیاری را می‌توان به دست آورد. یک مدل ستاره منطقی از یک جدول واقعیات و چندین جدول ابعاد تشکیل می‌شود. طرح یک ستاره پیچیده شامل صدها جدول واقعیات و جدول ابعاد می‌شود. جدول واقعیات شامل معیارهای تجاری ابتدایی می‌شود و می‌تواند ده‌ها هزار خط داشته باشد.

**مدل برف ریزه:** مدل برف ریزه تعمیمی از مدل ستاره است، هر نقطه بیرون را به چندین نقطه متصل می‌کند، از جمله ایندکس، دامنه و دسته‌بندی دقیق در کنار محدوده. مدل برف ریزه نتیجه استانداردسازی بیشتر جدول ابعاد مدل ستاره‌ای است. جدول ابعاد مدل برف ریزه مبتنی بر تئوری عادی‌سازی است. در بعضی موارد، مدل برف ریزه بعد از استانداردسازی جداول داده شکل می‌گیرد تا سلسله‌مراتب جدول ابعاد را کاهش دهد. نکته مثبت این امر این است که فضای ذخیره‌سازی را تا حدودی کاهش می‌دهد و به دلیل استانداردسازی و دانه‌بندی کم، مدل برف ریزه انعطاف‌پذیری را افزایش می‌دهد و عملکرد پرس‌وجو را از طریق کمینه کردن ظرفیت ذخیره داده و ترکیب آن با جداول ابعاد کوچک‌تر بهبود می‌بخشد.

**مدل ترکیبی:** مدل ترکیبی از ترکیب مدل ستاره‌ای و مدل برف ریزه حاصل می‌شود، مدل ستاره‌ای شامل جداول واقعیات و جداول ابعاد استانداردسازی شده می‌باشد، درحالی‌که در مدل برف ریزه تمام جداول استانداردسازی می‌شوند. در مدل ترکیبی فقط بزرگ‌ترین جداول ابعاد استانداردسازی می‌شوند، این جداول عموماً داده کاملاً استانداردسازی شده را شامل می‌شوند [۱].

## ۶- مدل های داده در انبار داده

مدل های داده ابزاری هوشمند برای مشخص کردن ویژگی های آماری سیستم ها هستند، یعنی برای توضیح اشیا، ویژگی های آنها و روابط درونی شان. از آنجا که انبار داده به عنوان مدل سیستم کسب و کار واقعی تعریف می شود که مجموعه ای از همه حالات آن سیستم در طی دوره زمانی خاص را ارائه می کند، مهم است مدل داده هایی مشخص شود که نه تنها از مختصات سیستم پشتیبانی کند، بلکه تغییرات در سیستم کسب و کار یا منابع داده ها را مشخص نماید [۳].

در بالاترین سطح انتزاع، همه مدل های توصیف شده بر مبنای چند مفهوم بنیادین قرار دارند که در جدول ۱ نشان داده شده است.

جدول ۱: مفاهیم بنیادین مدل داده

شناسه	ویژگی	رابطه	شیء	
کلید اصلی	دامنه	کلید خارجی	رابطه	مدل رابطه ای
کلید اصلی/کسب و کار	ستلایت	لینک	هاب	مدل دیتا والت
کلید اصلی	ویژگی	همبستگی	آنکور/ گره (Knot)	مدل آنکور (قوسی)
کلید اصلی/کسب و کار	ویژگی	واقعیت (Fact)	بُعد	مدل بُعدی

در ادامه انواع مدل داده در انبار داده ارائه می شود:

۱. **مدل رابطه ای:** مدل رابطه ای مدل ریاضی رسمی است که برپایه مجموعه ای از محاسبات قرار دارد. توسعه پشتیبانی از سیستم های مدیریت پایگاه داده بلافاصله به پیدایش مدل رابطه ای (DBMS) منجر شد که سبب گردید تا این مدل ها به پرکاربردترین مدل ها برای توسعه سیستم های تراکنشی مبدل شوند. کاد مفاهیم اصلی را در ۱۹۶۹ ارائه نمود و سال بعد این تحقیق را بیشتر گسترش داد. مفهوم اصلی این مدل عبارت است از رابطه ای که به صورت زیرمجموعه ای از حاصلضرب دکارتی مجموعه های  $S_1, S_2, \dots, S_n$  است. هر یک از این مجموعه ها نشان دهنده دامنه ارتباط است. کاد همچنین مفهوم نرمال سازی را به عنوان رویه ای برای تضمین این مورد که مقادیر همه ویژگی های ارتباط همیشه بسیار کوچک هستند، ارائه کرد.

۲. **مدل Data Vault:** مدل دیتا والت (DV) مدل تجربی است که بیشتر از دو دهه است که از آن برای توسعه انبار داده ها استفاده می شود. این مدل را دن لینستت با هدف ایجاد قابلیت پیگیری کامل داده ها و همچنین قابلیت مقیاس بندی و انطباق پذیری بیشتر در مقایسه با مدل داده هایی مورد استفاده در آن زمان مطرح کرد. بیل اینمو که معماری انبار داده های نسل بعدی، DW.2.0، را پیشنهاد کرد مدل دیتا والت را انتخابی بهینه برای معماری DW 2.0 دانست.

هاب مفهوم بنیادین مدل DV است که نشان می دهد شی ای انتزاعی یا دنیای واقعی سیستم کسب و کار را می توان به شکل منحصر بفردی شناسایی کرد. این شناسه کلید کسب و کار واقعی است که به مرور زمان تثبیت می شود یا تمایل کمتری برای تغییر دارد. ساختار هابی که در معرض تغییر قرار دارد از طریق یک یا چند ستلایت مشخص می شود. ستلایت دامنه ای (احتمالاً پیچیده) است که یک یا چند ویژگی توصیفی هاب را نشان می دهد. هاب ها می توانند از طریق رویدادهای سیستم، روابط ساختاری یا سایر روابط، با یکدیگر ارتباط داشته باشند که از طریق مفهوم لینک نشان داده می شود.

۳. **مدل آنکور:** مدل آنکور ابتدا در سال ۲۰۰۴ به عنوان مدل (و روش) تجربی برای توسعه انبار داده معرفی شد و در سال ۲۰۱۰ بصورت رسمی درآمد. دلیل اصلی معرفی این مدل ارائه متادیتا DW قابل تعمیم و پشتیبانی از متدولوژی توسعه DW چابک بود. آنکور مفهوم اصلی است که مجموعه اشیا سیستم کسب و کار را نشان می دهد. مفهوم گره (Knot) با مفهوم آنکور شباهت دارد اما با این تفاوت مهم که این مفهوم مجموعه ای از اشیا را نشان می دهد که به مرور زمان تغییر نمی کنند. مفهوم پیوستگی (Tie) همراه با مفهوم نقش (Role) ارائه می شود که نگاهت بین دو مجموعه از اشیا را نشان می دهد. از آنجا که هر دو مدل آنکور و گره مجموعه ای از اشیا را نشان می دهند دو نوع نقش ارائه می شود که عبارتند از: نقش های آنکور و نقش های گره. از این رو پیوستگی، ارتباط بین دو یا چند آنکور (یا گره) را نشان می دهد که ترکیبی از حداقل دو نقش است.

۴. مدل **Dimensial (بعدی)**: مدل بعدی احتمالاً پرکاربردترین مدل داده DW است. بزرگترین مزیت مدل بعدی این است که بازیابی و تحلیل داده‌ها را تسهیل می‌کند تا کاربران محیط کسب و کار بتوانند به آسانی در DW تحقیق کنند. دو مفهوم اصلی مدل بعدی عبارتند از ابعاد و واقعیت‌ها. واقعیت‌ها نوعاً مجموعه رخدادهایی را نشان می‌دهند که بین سیستم کسب و کار رخ داده‌اند درحالی‌که ابعاد اطلاعات توصیفی در خصوص واقعیت‌ها را ارائه می‌کند. بعد می‌تواند نرمال سازی نشده باشد یعنی وقتی کل سلسله مراتب بعد به شکلی زاید در جدولی واحد ذخیره می‌شود (طرح ستاره) می‌توان از جدول‌های مربوط به سلسله مراتب یک بعد در ابعاد دیگر استفاده کرد [۱].

جدول ۲ مقایسه ای از مدل های داده ارائه شده را نشان می دهد:

جدول ۲: مقایسه مدل های داده

بعدی	آنکور	Data vault	رابطه ای	
بله	بله	بله	خیر	مفاهیم درونی
خیر	تا حدودی	تا حدودی	خیر	مقاومت در برابر تغییر
بله	تا حدودی	تا حدودی	خیر	جنبه زمانی
خیر	تا حدودی	بله	خیر	کامل بودن
خیر	بله	بله	خیر	قابلیت پیگیری

## ۷- پردازش تحلیلی آنلاین

پردازش تحلیلی آنلاین (OLAP) یکی از مولفه‌های هوش تجاری است. پردازش تحلیلی آنلاین نمایی چندبعدی از داده‌ها را جهت تحلیل برای کاربر ایجاد می‌کند. پردازش تحلیلی آنلاین به چند روش انجام می‌شود که شامل MOLAP، ROLAP، و HOLAP است. در روش MOLAP (پردازش تحلیلی آنلاین چندبعدی) داده‌های مورد استفاده برای تحلیل در پایگاه داده‌های چندبعدی ویژه ذخیره می‌شوند. روش ROLAP مستقیماً با پایگاه داده‌های رابطه‌ای کار می‌کند. روش HOLAP نوع دوگانه OLAP است که روش‌های MOLAP و ROLAP را بایکدیگر ترکیب می‌کند و به طراح اجازه می‌دهد بخشی از داده را که به روش MOLAP ذخیره می‌شود و بخشی را که به روش ROLAP ذخیره می‌شود انتخاب کند.

داده‌های هوش تجاری معمولاً در انبارهای داده ذخیره می‌شوند. هدف از انبارداری داده ساخت مخزنی عظیم از داده‌های یکپارچه است که برای اهداف تحلیلی بهینه‌سازی شده‌اند. امروزه، چالش کلان داده باعث تغییر انبارهای داده سنتی به انبارهای داده مبتنی بر ابر شده است که از منابع ذخیره‌سازی نامحدود و دسترسی ایمن مبتنی بر اینترنت از هر جایی از دنیا بهره می‌برند [۳].

## ۸- پژوهش های مرتبط با طراحی انباره داده

پیشرفت‌های هوش تجاری بر صنعت مخابرات تاثیرگذار می باشد. تاکنون راه کارهای مختلفی برای طراحی انبار داده‌های صنعت مخابرات پیشنهاد شده است، که اکثر آنها بر مبنای شرکت‌ها و کشورهای بخصوصی بوده‌اند. نمونه‌ای از این راه کار روشی است که برای طراحی انبار داده و کاربرد آن در بزرگترین شرکت ارائه‌دهنده خدمات مخابراتی ایتالیا، یعنی سیستم اطلاعاتی Telecom Italia، بر مبنای سطح رایانشی استفاده شده است. برای بکارگیری این روش از ابزارهای نرم‌افزاری آزمایشی استفاده شده است. در این طرح، طراحی انبارهای داده مناسب برای تحلیل داده‌های مشتری و تماس‌های تلفنی هم در نظر گرفته شده است. مثال دیگر پژوهشی است که در آن از داده‌های یکی از بزرگترین شرکت‌های مخابراتی آمریکا، یعنی شرکت Pacific Bell، برای ساخت انبار داده‌ای مختص تصمیم‌گیری و هوش مصنوعی برای افرادی که با تماس‌های تلفنی، خطوط تلفنی، و کلاهبرداری سروکار دارند، و همچنین دسترسی به اطلاعات مختصر و دقیق، استفاده شده است.

تحلیل فضایی زمانی رد گوشی های همراه برای شناسایی مبدأ افرادی که در رویدادهای بخصوص شرکت می کنند در پژوهشی ارائه شده است. نتایج نشان می دهند که رابطه محکمی میان رویدادها و افراد شرکت کننده وجود دارد، چون افرادی که در مجاورت رویدادی زندگی می کنند ترجیحاً جذب آن می شوند. برای انجام این مطالعه، از مسیرها، یعنی توالی هایی از موقعیت های مکانی ای که کاربر از آنها بازدید کرده است به ترتیب وقوع، که با استفاده از اندازه گیری های مداوم موقعیت مکانی فرد به دست آمده اند، استفاده شده است. این مطالعه به فرآیند تصمیم گیری درباره مدیریت رویدادها و همچنین کاهش بار ترافیکی که پیامدهای جدی ای را برای مدیریت شهر در پی دارد کمک می کند.

در پژوهش دیگری طرحی از انبار داده ای که داده های مخابرات را از منابع مختلف مثل رکورد جزئیات تماس و داده های مدیریت رابطه مشتری، با یکدیگر ترکیب می کند، ارائه می شود و سپس معماری یکپارچه ای را برای ترکیب داده های به دست آمده از پایگاه داده های عملیاتی اپراتورهای مخابراتی مختلف با انبار داده مرکزی برای پردازش OLAP و تحلیل تجاری پیشنهاد می شود.

در مطالعه دیگر انبار داده ای برای شرکتی مخابراتی در اندونزی بر مبنای چهار طرح، یعنی استفاده از تلفن، استفاده از اینترنت، صورتحساب، و پرداخت ساخته می شود. برای پروفایل بندی مشتریان مخابرات بر مبنای نوع استفاده مشتری از خدمات، صورتحساب مشتری و پرداخت مشتری از دسته بندی مشتری با کمک الگوریتم خوشه بندی میانگین های K استفاده می شود [۳].

## ۹- نتیجه گیری

شرکت ها منابع مالی عظیمی را به هوش تجاری اختصاص می دهند تا تصمیمات تجاری را بهتر درک کنند و انبارهای داده برای این منظور از یک ابزار مدیریت ضروری تشکیل شده اند. انبارهای داده، داده های عملیاتی را به قالبی قابل درک تر و کاربر پسندتر برای تحلیل و تصمیم گیری تبدیل می کنند. تحلیل داده ها با استفاده از انبار داده در صورت استفاده از عملیات های جدیدی از جمله انباشتن (-roll up) و استخراج کردن (drill-down) برای بازیابی داده ها در چندین سطح به میزان قابل توجهی بهبود می یابد. بعدها می توان از تکنیک های استخراج داده برای یافتن الگوهای پنهان و پیش بینی رفتارهای مشتری و روندهای آتی استفاده کرد و به کسب و کارها کمک کرد تصمیمات فعالانه و دانش محور بگیرند.

کارآیی انبار داده برای شرکت مخابرات چندجنبه دارد: با تأمین منبع اصلی اطلاعات برای ساخت گزارش ها می توان تحلیل داده ها را ساده تر کرد و کاربران می توانند با تحلیل سریع تر داده ها و تولید ساده تر گزارش ها جزئیات داده ها را استخراج کنند. علاوه بر این، می توانند اطلاعات را به گونه ای که قبلاً هرگز ممکن نبود به قسمت های مختلف تقسیم کنند و کوئری های چندبعدی را بر روی داده های فضایی و زمانی اجرا کنند. نتایج این تحلیل می تواند منجر به تصمیمات بهتر شود و محصولات را برای مشتریان جذاب تر کند. در این مقاله سعی گردید به مرور ویژگی های انبار داده، معماری، طرح ها و مدل های داده در انبار داده پرداخته شود. پردازش تحلیلی آنالین (OLAP) به عنوان یکی از مولفه های هوش تجاری ارائه شد و خلاصه ای از پژوهش های مرتبط با انبار داده مطرح گردید.

## مراجع

- [1]: R.Hou "Research and analysis of data warehouse technologies", IEEE International Conference on Computer Science and Network Technology, 2011
- [2]: I.Bojicic, Z.Marjancvic, N.Turajlic, M.Petovic, M.Vuckovic, V.Jovanovic,"A Comparative Analysis of Data Warehouse Data Models", 6<sup>th</sup> International Conference on Computer Communications and Control(ICCCC), 2016
- [3]: G.Garani, A.Chernov, M.Butakova "A Data Warehouse Approach for Business Intelligence", IEEE 28<sup>th</sup> International Conference on Enabling Technologies:Unfrastructure for Collaborative Enterprises(WETICE), 2019