المحور الثاني: مستودع البيانات وإدارة قاعدة البيانات

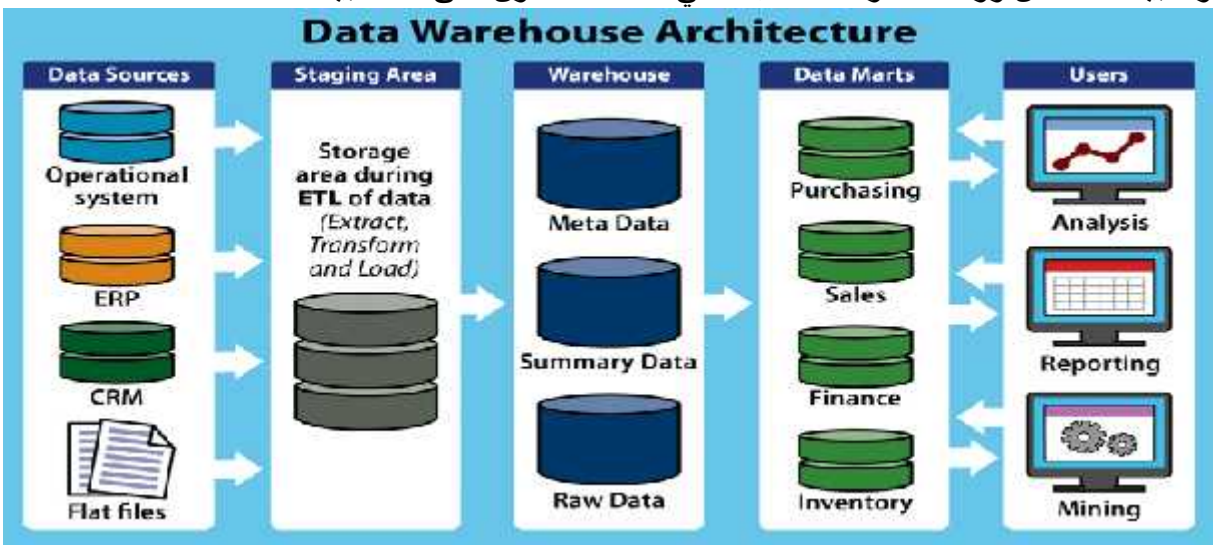
- **مستودع البيانات:** هو نوع من نظام إدارة البيانات تم تصميمه **لتمكين ودعم أنشطة ذكاء الأعمال** (BI)، وخاصةً التحليلات.

تهدف مستودعات البيانات إلى **إجراء الاستعلامات والتحليلات** وغالبًا ما تحتوي على أحجام كبيرة من البيانات القديمة.

عادةً ما يتم استخلاص البيانات داخل مستودع البيانات من مجموعة واسعة من المصادر، مثل ملفات **سجلّ التطبيقات وتطبيقات المعاملات**.

مستودع البيانات (DW) هو نظام تخزين رقمي يقوم **بتوصيل ومواءمة** كميات كبيرة من البيانات من العديد من المصادر المختلفة. والغرض منه هو **تغذية التحليل الذكي للأعمال** (BI) وإعداد التقارير والتحليلات ودعم المتطلبات التنظيمية - حتى تتمكن الشركات من تحويل بياناتها إلى رؤى شاملة واتخاذ قرارات ذكية معتمدة على البيانات.

تتدفق البيانات إلى مستودع بيانات من الأنظمة التشغيلية (**مثل** تخطيط موارد المؤسسة **وإدارة علاقات العملاء** (وقواعد البيانات والمصادر الخارجية مثل أنظمة الشركاء وأجهزة إنترنت الأشياء (IoT) وتطبيقات الطقس ووسائل التواصل الاجتماعي - عادة ما تكون على قاعدة بيانات منتظمة.



- **مستودعات البيانات، ومتاجر البيانات، ومخازن بيانات التشغيل و بحيرات البيانات:**

.يؤدي **متجر البيانات غرض وظائف** مستودع البيانات ولكن ضمن نطاق محدود للغاية - **عادةً ضمن قسم واحد أو خط أعمال واحد.** وهذا بدوره يجعل إنشاء متاجر البيانات أسهل من مستودعات البيانات. **ومع** ذلك، فإنها تميل إلى تقديم عدم تناسق لأنه قد يكون من الصعب إدارة البيانات والتحكم فيها بشكل موحّد عبر العديد من متاجر البيانات. ويُعد متجر البيانات مستودع بيانات يخدم احتياجات فريق أو وحدة أعمال معينة، مثل الشؤون المالية أو التسويق أو المبيعات. إنه أصغر حجمًا وأكثر تركيزًا، وقد يحتوي على ملخصات للبيانات التي تخدم مجتمع المستخدمين بشكل أفضل. قد يكون متجر البيانات جزءًا من مستودع البيانات أيضًا.

تدعم مخازن بيانات التشغيل العمليات اليومية فقط، لذلك فإن عرضها للبيانات القديمة محدود للغاية. وعلى الرغم من أنها تعمل جيدًا كمصادر للبيانات الحالية وغالبًا ما تُستخدم في مستودعات البيانات على هذا النحو، فإنها لا تدعم الاستعلامات الثرية القديمة.

تعمل بحيرات البيانات على تخزين عدد كبير من البيانات المتباينة وغير المرشحة لاستخدامها لاحقًا لغرض معين. يتم تحصيل البيانات من تطبيقات خط الأعمال، وتطبيقات الأجهزة المحمولة، والوسائط الاجتماعية، وأجهزة إنترنت الأشياء(IoT) ، وغيرها، كبيانات خام في بحيرة بيانات. يتم استخلاص بنية مجموعات البيانات المختلفة وتكاملها واختيارها وتنسيقها في وقت التحليلات من قِبل الشخص الذي يجري التحليلات. عندما تحتاج المؤسسات إلى تخزين منخفض التكلفة للبيانات غير المنظمة وغير المنسقة من

مصادر متعددة، والتي تعتزم استخدامها لغرض معين في المستقبل، فقد تكون بحيرة البيانات هي الاختيار الصحيح.

- **أنواع مستودعات البياناتDWH)**

عادةً ما تستخدم أنظمة المؤسسات ثلاثة أنواع رئيسية من مستودعات البيانات:(DWH)

1. **مستودع بيانات المؤسسة :(EDW)** كمستودع بيانات مركزي، يوفر EDW نهجًا شاملاً لتنظيم البيانات وتقديمها.
2. **مخزن البيانات التشغيلية (ODS)** : هو نوع من مخازن البيانات.
3. **مارت البيانات :**تم تصميم سوق البيانات لبيانات الإدارات، مثل المبيعات والتمويل وسلسلة التوريد.

- **مكونات مستودع بيانات:**

1. **قاعدة البيانات المركزية.**
2. **تكامل البيانات.**
3. **بيانات التعريف.**
4. **أدوات الوصول إلى مستودع البيانات :من أمثلة أدوات الوصول**: أدوات الاستعلام وإعداد التقارير، وأدوات تطوير التطبيقات، وأدوات تعدين البيانات، وأدوات.OLAP

- **مزايا تخزين البيانات بالمستودع:**

هناك العديد من الفوائد لمستودع البيانات. إليك بعض منها:

- **تحليلات أعمال أفضل؛**
- **استعلامات أسرع؛**
- **جودة البيانات المحسنة؛**
- **الرؤية التاريخية.**

- **بنية مستودع البيانات:**

يمكن القول أن مستودع البيانات يعتمد على بنية ثلاثية المستويات تتضمن:

- **الطبقة السفلية (طبقة التخزين :(**تتكون هذه الطبقة من وسائط التخزين، ومستودع التعريف، وسوق البيانات، وخادم قاعدة البيانات
- **الطبقة الوسطى (طبقة المعالجة :(**الطبقة الوسطى هي نظام المعالجة التحليلية عبر الإنترنت .(OLAP)يعالج الاستعلامات المعقدة ويقدم النتائج في شكل مناسب لتحليل البيانات وذكاء الأعمال.
- **المستوى العلوي (طبقة الخدمات :(**تمثل هذه الطبقة الواجهة الأمامية للمستخدم مع لوحة معلومات مرئية لتمكين التحليلات وإعداد التقارير.

- **تصميم مستودع بيانات:**

عندما تشرع مؤسسة في تصميم مستودع بيانات، يجب أن تبدأ بتحديد متطلباتها التجارية المحددة، والاتفاق على النطاق، وصياغة تصميم مفاهيمي. ويجب أن يعالج أي تصميم لمستودع البيانات ما يلي:

- محتوى بيانات محدد؛
- **العلاقات** داخل مجموعات البيانات وفيما بينها؛
- بيئة الأنظمة التي ستدعم مستودع البيانات؛
- أنواع تحويلات البيانات المطلوبة؛
- تكرار تحديث البيانات.

- **حالات استخدام مستودع البيانات:**

في البيع بالتجزئة: بالنسبة قطاع التجزئة، من الأمثلة الجيدة على ذلك سوق بيانات البيع بالتجزئة الذي يتضمن معلومات العملاء من سجلات النقد والقوائم البريدية والمواقع الإلكترونية وبطاقات التعليقات.

في التمويل: تحتاج البنوك وشركات التأمين والشركات التجارية وغيرها من الشركات ذات الصلة بالقطاع المالي إلى بيانات دقيقة في جميع الأوقات. يتيح التحقق الصحيح من صحة البيانات في قواعد البيانات والاتصال المناسب بالجداول الأخرى في قاعدة البيانات إمكانية تحقيق ذلك.

- **فوائد مستودع البيانات للمظلمات:**

فيما يلي فوائد رئيسية لتخزين البيانات لعملك:

1- **يوفر الوقت؛**
2- **يحسن جودة البيانات؛**
3- **يحسن ذكاء الأعمال؛**
4- **يؤدي إلى اتساق البيانات؛**
5- **.يعزز عائد الاستثمار(ROI) ؛**
6- **يخزن البيانات التاريخية؛**
7- **يزيد من أمن البيانات.**

مستودع البيانات وقاعدة البيانات

قاعدة البيانات هي مجرد تقنية تقليدية **لتخزين** البيانات (إلتقاط البيانات، استخراج البيانات...)
مستودع البيانات مخصص بشكل خاص لتحليل البيانات. إنه يحتفظ بكل شيء في مكان واحد من العديد من بنوك البيانات الخارجية.

الفرق الرئيسي بين قاعدة البيانات ومستودع البيانات:

- قاعدة البيانات عبارة عن مجموعة من البيانات ذات الصلة التي تمثل بعض عناصر العالم الحقيقي، في حين أن مستودع البيانات هو نظام معلومات يخزن البيانات التاريخية والتبديلية من مصادر فردية أو متعددة.
- تم تصميم قاعدة البيانات لتسجيل البيانات، في حين تم تصميم مستودع البيانات لتحليل البيانات.
- قاعدة البيانات عبارة عن مجموعة بيانات موجهة نحو التطبيق، في حين أن مستودع البيانات عبارة عن مجموعة بيانات موجهة نحو الموضوع.
- تستخدم قاعدة البيانات معالجة المعاملات عبر الإنترنت (OLTP) ، بينما يستخدم مستودع البيانات المعالجة التحليلية عبر الإنترنت.(OLAP)

النواحي	مستودع البيانات	قاعدة البيانات
الهدف	تخزين البيانات التاريخية في شكل منظم لتسهيل تحليل البيانات وإعداد التقارير.	التعامل مع كميات كبيرة من الاستعلامات لتخزين البيانات في الوقت الفعلي.
أسلوب المعالجة	OLAP	OLTP
مجل	يخزن البيانات من مصادر متعددة.	عادة ما تكون مصممة لوظيفة عمل محددة.
سهولة التحليل	أسهل لإجراء التحليلات.	التحليل معقد.

تطبيقات قواعد البيانات:

قطاع	الأستعمل
البنوك والمصارف	الاستخدام في القطاع المصرفي للحصول على معلومات العملاء والأنشطة المتعلقة بالحساب والمدفوعات والودائع والقروض وبطاقات الائتمان وما إلى ذلك.
الجامعات	لتخزين معلومات الطلاب وتسجيلات الدورات والكليات والنتائج.
اتصالات	يساعد على تخزين سجلات المكالمات والفواتير الشهرية وصيانة الرصيد وما إلى ذلك.
تمويل	يساعدك على تخزين المعلومات المتعلقة بالأسهم ومبيعات ومشتريات الأسهم والسندات.

المبيعات والإنتاج	تستخدم لتخزين العملاء والمنتجات والمبيعات.tails
تصنيع	يتم استخدامه لإدارة بيانات سلسلة التوريد وتتبع إنتاج العناصر وحالة المخزون.
ادارة موارد بشرية	تفاصيل حول رواتب الموظفين، والخصم، وتوليد شيكات الرواتب، وما إلى ذلك.

- **عيوب قاعدة البيانات:**
- تكلفة الأجهزة والبرامج الخاصة بنظام قاعدة البيانات المطبق مرتفعة مما قد يزيد من ميزانية مؤسستك.
- غالبًا ما تكون العديد من أنظمة إدارة قواعد البيانات complex الأنظمة، لذلك يلزم تدريب المستخدمين على استخدام نظام إدارة قواعد البيانات.
- لا يمكن إجراء حسابات معقدة
- المشكلات المتعلقة بالتوافق مع الأنظمة الموجودة بالفعل
- قد يفقد أصحاب البيانات السيطرة على بياناتهم، مما يزيد من مشكلات الأمان والملكية والخصوصية.

- **عيوب مستودع البيانات:**
- تستغرق إضافة مصادر بيانات جديدة وقتًا، وترتبط بتكلفة عالية.
- في بعض الأحيان، قد لا يتم اكتشاف المشكلات المرتبطة بمستودع البيانات لسنوات عديدة.
- مستودعات البيانات هي أنظمة صيانة عالية. قد يستغرق استخراج البيانات وتحميلها وتنظيفها وقتًا طويلاً.
- قد يبدو مستودع البيانات بسيطًا، لكنه في الواقع معقد للغاية بالنسبة للمستخدمين العاديين. أنت بحاجة إلى توفير التدريب للمستخدمين النهائيين، الذين ينتهي بهم الأمر إلى عدم استخدام استخراج البيانات والمستودعات.
- على الرغم من أفضل الجهود المبذولة في إدارة المشاريع، فإن نطاق تخزين البيانات سيزداد دائمًا.

- **المقصود بـ OLAP وOLTP :**

المعالجة التحليلية عبر الإنترنت (OLAP) ومعالجة المعاملات عبر الإنترنت (OLTP) هما مفهومان متميزان في تخزين البيانات. كل منها يخدم غرضًا محددًا ويمتلك خصائص مختلفة.

OLAPهو نظام يدعم تحليل البيانات متعدد الأبعاد بسرعات عالية على أحجام ضخمة من البيانات. هذه البيانات بشكل عام من مستودع بيانات أو سوق بيانات أو أي مخزن بيانات آخر. يساعد OLAP في تحليل البيانات التاريخية وفهمها وهو مفيد لأداء هذه الوظائف: حسابات تحليلية معقدة، ذكاء الأعمال BI)، التنقيب عن البيانات، تحليل مالي، التنبؤ بالمبيعات.

من ناحية أخرى، يتم استخدام OLTP لمعالجة المعاملات ويتضمن عادةً استعلامات وتحديثات بسيطة على كمية كبيرة من البيانات في الوقت الفعلي من قبل عدد كبير من المستخدمين. تتم هذه المعاملات بشكل عام عبر الإنترنت.

وهو يدعم معالجة عالية التزامن ومُحسَّن لعمليات التحديث والحذف السريعة OLTP .هي العملية الرئيسية وراء معاملات أجهزة الصراف الآلي وعمليات الشراء في المتجر وحجوزات الفنادق.

OLAP	OLTP (التطبيقات بشكل بسيط)
تخزين ببانات التحليل	تخزين بيانات المعاملات
مفيد لتحليل الأعمال	مفيد لتشغيل الأعمال أي إدارتها
طبيعة البيانات تاريخية وملخصة	طبيعة البيانات هي حالية ومفصلة
يدعم القراءة فقط	يدعم CRUD أي بمعنى: خلق، وقراءة، وتحديث وحذف
البيانات متكاملة حسب مجال الموضوع	البيانات معزولة في التطبيقات

يقوم بتحليل البيانات المعقدة لاتخاذ القرارات الذكية	يُجري معاملات كبيرة من قبل عدة مستخدمين في **الوقت الفعلي**
مصممة للاستخدام من قبل مديري البيانات والعاملين في مجال المعرفة	مصممة للعاملين في الخطوط الأمامية مثل صرافى البنوك والصرافين **ومسؤولى مكتب الاستقبال**
يحتوي مصدر البيانات على مخططات متعددة **لدعم الاستعلامات المعقدة** من البيانات الحالية والتاريخية	يعتمد على أنظمة إدارة قواعد البيانات التقليدية لاستيعاب أحجام البيانات الضخمة في المعاملات **في الوقت الفعلي**

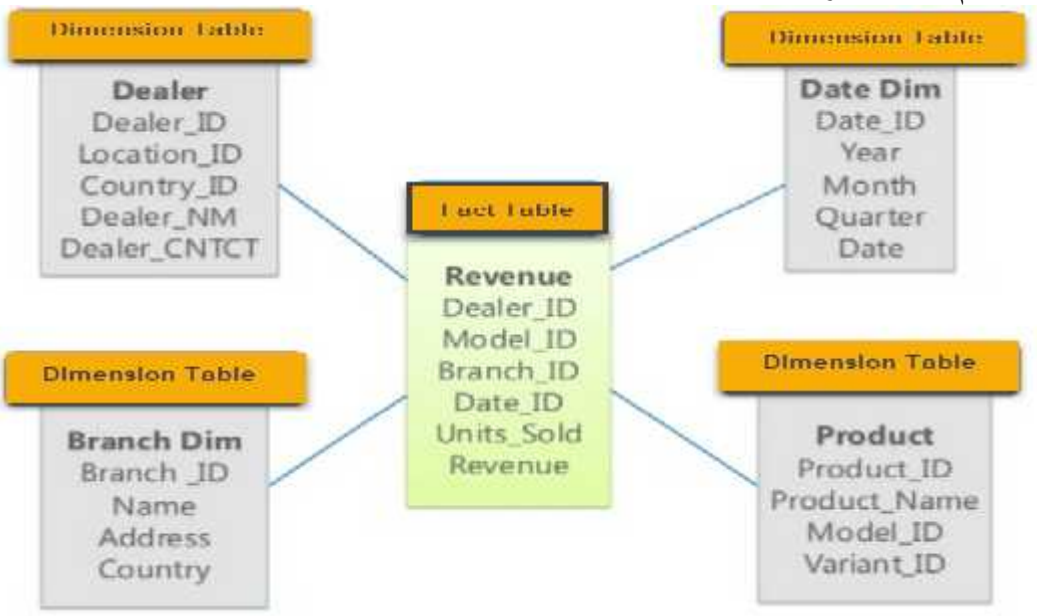
- **المخططات في مستودع البيانات:**

يحدد المخطط في مستودع البيانات طرقًا متعددة لتنظيم النظام باستخدام كيانات قاعدة البيانات مثل جدول الأبعاد وجداول البيانات الفعلية والاقتران المنطقي بها. هناك ثلاثة أنواع رئيسية من المخططات في مستودع البيانات.

1- مخطط مجموعة النجوم:

مخطط نجمة ينظم البيانات في جدول حقائق مركزي ومجموعة من جداول الأبعاد. يتم استخدامه لإلغاء تسوية البيانات وإضافة أعمدة زائدة عن الحاجة إلى جداول الأبعاد لاستعلامات البيانات بشكل **أسرع.**

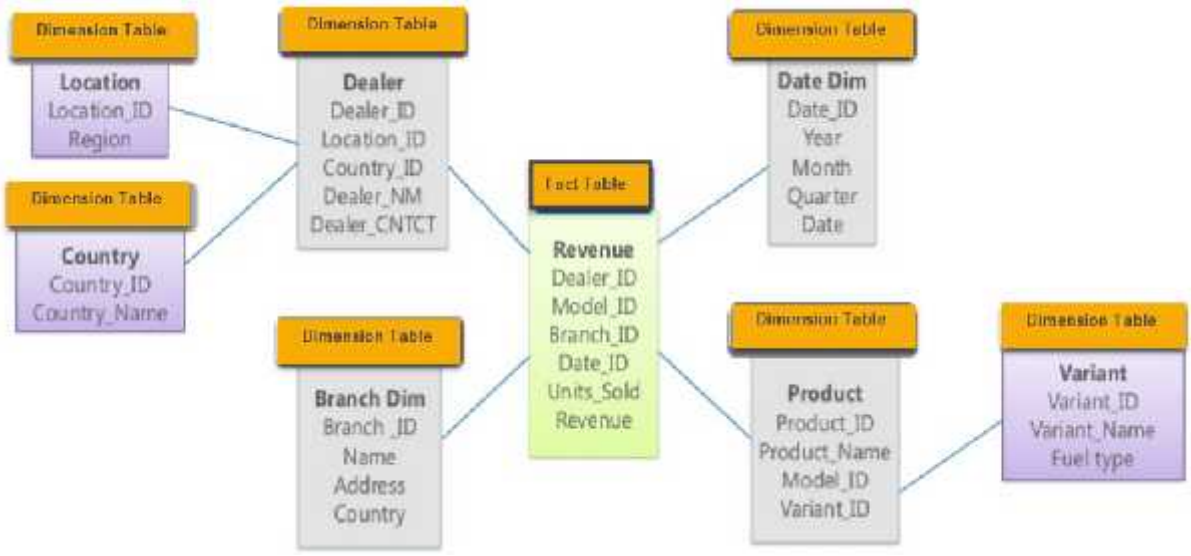
يحتوي جدول الحقائق على قياسات أو مقاييس البيانات، بينما توفر جداول الأبعاد سياق البيانات، **مثل** الوقت أو الموقع أو تفاصيل المنتج. يتكيف المخطط النجمي جيدًا ليلائم نماذج OLAP **للحصول على أداء** استعلام أفضل مقارنة بالبيانات العادية.



2- مخطط ندفة الثلج:

مخطط ندفة الثلج هو شكل يرتبط جدول الحقائق بجداول أبعاد متعددة تمت تسويتها. وهذا يعني أنها مقسمة إلى جداول فرعية متعددة. بالمقارنة مع المخطط النجمي، تستفيد ندفة الثلج من تكرار البيانات المحدود، مما يساعد على التحسين تكامل البيانات **ولكن على حساب انخفاض أداء الاستعلام.**

أو يمكن القول أن مخطط ندفة الثلج في مستودع البيانات هو ترتيب منطقي للجداول في قاعدة بيانات **متعددة الأبعاد. مخطط ندفة** الثلج هو امتداد لمخطط النجوم، ويضيف أبعادًا إضافية. يتم تسوية جداول الأبعاد التي تقسم البيانات إلى جداول إضافية.



مثل عن مخطط ندفة الثلج snowflage schema

خصائص مخطط ندفة الثلج

- الميزة الرئيسية لمخطط ندفة الثلج هي أنه يستخدم مساحة قرص أصغر.
- تتم إضافة البعد إلى المخطط بشكل أسهل في التنفيذ
- بسبب انخفاض أداء الاستعلام الجداول المتعددة
- التحدي الأساسي الذي ستواجهه أثناء استخدام مخطط ندفة الثلج هو أنك تحتاج إلى إجراء المزيد من جهود الصيانة بسبب وجود المزيد من جداول البحث.

مساوئ مخطط ندفة الثلج

- العيب الكبير في مخطط ندفة الثلج هو زيادة الصيانة المطلوبة.
- الاستعلامات صعبة الفهم.
- يعني وجود عدد أكبر من الجداول المزيد من الصلات، وبالتالي وقت تنفيذ أطول للاستعلام.

الفرق بين مخطط النجوم ومخطط ندفة الثلج:

مخطط ندفة الثلج	مخطط النجوم
يتم تقسيم التسلسلات الهرمية إلى جداول مفصلة.	يتم تخزين التسلسلات الهرمية للأبعاد في جدول الأبعاد.
جدول حقائق واحد محاط بجدول الأبعاد والذي بدوره محاط بجدول الأبعاد	يحتوي على جدول حقائق محاط بجداول الأبعاد.
يتطلب مخطط ندفة الثلج العديد من الصلات لجلب البيانات.	في المخطط النجمي، تؤدي الصلة الفردية فقط إلى إنشاء العلاقة بين جدول الحقائق وأي جداول أبعاد.
مستوى عال من تكرار البيانات	تكرار البيانات على مستوى منخفض للغاية
تقسيم البيانات إلى جداول أبعاد مختلفة.	يحتوي جدول البعد الواحد على بيانات مجمعة.
يتم تمثيل مخطط Snowflake بجدول حقائق مركزي والذي من غير المحتمل أن يكون مرتبطًا بأبعاد متعددة.	يقدم استعلامات ذات أداء أعلى باستخدام Star Join Query Optimization. يمكن ربط الجداول بأبعاد متعددة.