

إسم المادة: إدارة قواعد البيانات العملاقة

إسم المحاضر: م. خليل المحمد

الأكاديمية العربية الدولية – منصة أعد

# مقدمة في قواعد البيانات العملاقة - تعريف

قواعد البيانات العملاقة (Big Data Databases) هي أنظمة إدارة قواعد بيانات متخصصة تُستخدم لتخزين ومعالجة كميات هائلة من البيانات التي قد تكون معقدة ومتنوعة وسريعة التغير. تختلف قواعد البيانات العملاقة عن قواعد البيانات التقليدية (مثل قواعد البيانات العلائقية) من حيث الحجم، السرعة، والتنوع.

التميز عن قواعد البيانات التقليدية:

1. **التوسع الأفقي: (Horizontal Scalability)** قواعد البيانات العملاقة تُبنى غالباً لتكون قابلة للتوسع أفقيًا، مما يعني أنه يمكن إضافة مزيد من الخوادم لزيادة قدرة النظام، بدلاً من الاعتماد فقط على تحسين قدرات الخوادم الموجودة (التوسع العمودي).
2. **المعالجة الموزعة: (Distributed Processing)** تعتمد قواعد البيانات العملاقة على تقنيات المعالجة الموزعة مثل Hadoop و MapReduce، حيث يتم توزيع البيانات عبر عدة خوادم ومعالجتها بالتوازي، مما يعزز الكفاءة ويقلل من وقت الاستجابة.
3. **التكيف مع البيانات غير المنظمة:** بينما تعتمد قواعد البيانات التقليدية على بنى ثابتة مثل الجداول والعلاقات بين الجداول، تتميز قواعد البيانات العملاقة بالقدرة على التعامل مع بيانات غير منظمة أو شبه منظمة بكفاءة.

# أهمية قواعد البيانات العملاقة

## 1. تحليل البيانات الضخمة:

- تحقيق رؤى معمقة: إدارة قواعد البيانات العملاقة تُمكن الشركات والمؤسسات من تحليل كميات ضخمة من البيانات بسرعة وفعالية.
- الابتكار والتطوير: تحليل البيانات الضخمة يمكن أن يؤدي إلى تطوير منتجات وخدمات جديدة تلبي احتياجات العملاء بشكل أفضل، مما يعزز التنافسية في السوق.

## 2. التفاعل مع كميات هائلة من البيانات:

- إدارة البيانات المتدفقة: في عالم اليوم، تتدفق البيانات بسرعة غير مسبوقة من مصادر مختلفة.
- التكيف مع التنوع الكبير: تنوع البيانات، بما في ذلك البيانات غير المنظمة مثل النصوص والصور والفيديو، يتطلب أنظمة مرنة وقابلة للتوسع للتعامل معها بكفاءة.

## 3. تعزيز الكفاءة التشغيلية:

- تحسين العمليات: من خلال تحليل البيانات الكبيرة، يمكن للمؤسسات تحسين كفاءة العمليات التشغيلية.
- تقليل التكاليف: إدارة البيانات بفعالية تساهم في تقليل التكاليف من خلال تحسين استخدام الموارد وتجنب الإنفاق غير الضروري.

## أهمية قواعد البيانات العملاقة

### 4. الأمن السيبراني وإدارة المخاطر:

- الكشف عن التهديدات الأمنية: قواعد البيانات العملاقة تُستخدم في مراقبة وتحليل البيانات من أجل الكشف المبكر عن الأنشطة غير العادية أو التهديدات الأمنية.
- إدارة الأزمات: في حالة وقوع حوادث، يمكن استخدام البيانات الضخمة لتحليل الوضع بسرعة واتخاذ القرارات المناسبة للتعامل مع الأزمة.

### 5. الاستجابة للتحديات الاجتماعية والبيئية:

- التنبؤ بالأزمات: قواعد البيانات العملاقة تُستخدم في تحليل البيانات المتعلقة بالطقس والكوارث الطبيعية والصحية.
- تحقيق التنمية المستدامة: من خلال تحليل البيانات البيئية والاجتماعية، يمكن للمؤسسات تعزيز استدامة الموارد وتحقيق الأهداف البيئية.

### 6. تحقيق ميزة تنافسية:

- اكتساب رؤية استراتيجية: المؤسسات التي تدير وتستغل قواعد البيانات العملاقة بفعالية تكون قادرة على فهم السوق بشكل أعمق.
- التخصيص وتحسين تجربة العملاء: من خلال فهم سلوك العملاء بشكل أفضل، يمكن تخصيص المنتجات والخدمات لتلبية احتياجاتهم.

# تطور قواعد البيانات العملاقة

1. التطور التاريخي: البدايات مع قواعد البيانات التقليدية: (SQL) : ظهرت قواعد البيانات العلائقية في السبعينيات وأصبحت المعيار الذهبي لإدارة البيانات في العقود التالية. أنظمة مثل Oracle وMySQL وMicrosoft SQL Server هي أمثلة على قواعد بيانات SQL التي كانت تسيطر على المشهد.

• محدودية قواعد البيانات التقليدية: إحدى المشاكل الرئيسية كانت أن قواعد البيانات العلائقية غير قادرة على التعامل بكفاءة مع البيانات التي لا تتبع هيكلًا معينًا (مثل النصوص، الصور، الفيديوها، وغيرها).

2. ظهور: NoSQL التحديات الجديدة والحلول البديلة: قواعد بيانات NoSQL لا تعتمد على الجداول أو العلاقات الثابتة، بل تستخدم نماذج بيانات مختلفة مثل الوثائق (Document-based)، والأعمدة العريضة (Wide-column stores)، والمفاتيح-القيم (Key-Value stores)، والرسوم البيانية (Graph databases).

3. الجيل الجديد مع: NewSQL : مواجهة التحديات المزدوجة: بالرغم من أن NoSQL قدمت حلولاً فعالة للتعامل مع البيانات الكبيرة وغير المنظمة، إلا أن بعض التطبيقات لا تزال تحتاج إلى ميزات مثل التوافق التام مع ACID (الذرة، التناسق، العزل، والاستمرارية) التي توفرها قواعد البيانات العلائقية.

# الأسس التقنية لإدارة قواعد البيانات العملاقة - البنية

## الأنظمة الموزعة:

### • كيفية توزيع البيانات:

- **Sharding التقسيم**: يتم تقسيم قاعدة البيانات إلى أجزاء أصغر، تُعرف باسم "shards"، بحيث يحتوي كل shard على جزء من البيانات الكلية.
- **Replication التكرار**: تُكرر البيانات عبر خوادم متعددة لضمان توافرها بشكل دائم. في حال فشل خادم معين، يمكن للخوادم الأخرى الاستمرار في تقديم البيانات دون انقطاع.

### • تحسين الأداء:

- **تقليل زمن الاستجابة**: عند توزيع البيانات عبر خوادم متعددة، يتمكن النظام من معالجة الاستعلامات بشكل أسرع، خاصة إذا كانت الخوادم موزعة جغرافيًا، مما يقلل من زمن الوصول.
- **التوسع الأفقي: (Horizontal Scaling)** يمكن إضافة المزيد من الخوادم بسهولة لتلبية الطلب المتزايد على الموارد دون الحاجة إلى ترقية الخوادم الموجودة.

### • التحديات:

- **التزامن: (Consistency)** ضمان أن تكون جميع نسخ البيانات متزامنة في الوقت الفعلي يمثل تحديًا، خاصة في الأنظمة التي تتطلب سرعة عالية في التحديثات.
- **إدارة الأخطاء**: التعامل مع الأخطاء مثل فشل أحد الخوادم يتطلب إدارة معقدة لضمان عدم فقدان البيانات أو توقف الخدمة.

# الأسس التقنية لإدارة قواعد البيانات العملاقة - البنية

## التخزين السحابي:

### أنواع خدمات التخزين السحابي

- 1. التخزين الكائني: (Object Storage) :** يخزن البيانات ككائنات مستقلة تحتوي على البيانات الفعلية وبيانات التعريف الخاصة بها.
  - الاستخدامات: مثالي لتخزين البيانات غير المنظمة مثل الصور، الفيديوهات، والملفات النصية.
  - أمثلة: Amazon S3، Google Cloud Storage، Microsoft Azure Blob Storage.
- 2. التخزين البلوكي: (Block Storage) :** يخزن البيانات في كتل ثابتة الحجم، كل كتلة يمكن الوصول إليها وإدارتها بشكل مستقل.
  - الاستخدامات: مناسب لقواعد البيانات التي تتطلب أداءً عاليًا مثل قواعد البيانات العلائقية والتطبيقات الحساسة للأداء.
  - أمثلة: Amazon EBS، Google Persistent Disk، Microsoft Azure Disk Storage.
- 3. التخزين الملفي: (File Storage) :** يوفر نظام ملفات يمكن الوصول إليه عبر الشبكة، مشابه للأنظمة التقليدية مثل NFS أو SMB.
  - أمثلة: Amazon EFS، Google Filestore، Microsoft Azure Files.

# الأسس التقنية لإدارة قواعد البيانات العملاقة - البنية

مزايا التخزين السحابي في إدارة قواعد البيانات العملاقة:

- التوسع الأفقي والرأسي:
  - التوسع الأفقي: إضافة المزيد من الخوادم لتحسين القدرة على التعامل مع الزيادة في حجم البيانات.
  - التوسع الرأسي: زيادة قدرات الخوادم الحالية مثل زيادة الذاكرة أو المعالجات لتحسين الأداء.
- الأداء العالي: توفر خدمات التخزين السحابي تقنيات متقدمة لتحسين سرعة الوصول إلى البيانات ومعالجتها.
- الأمان المتقدم: تشفير البيانات أثناء النقل والتخزين، بالإضافة إلى التحكم في الوصول وإدارة الهوية لضمان حماية البيانات الحساسة.
- إدارة البيانات المستمرة: تقديم أدوات لإدارة النسخ الاحتياطي والاستعادة، مما يضمن الحفاظ على البيانات وسلامتها في حالة حدوث أعطال أو فقدان بيانات.
- التكامل السلس مع الأدوات والتقنيات الأخرى: يمكن للتخزين السحابي التكامل بسهولة مع خدمات البيانات والتحليلات الأخرى مثل أدوات التحليل الكبيرة (Big Data Analytics) و منصات التعلم الآلي (Machine Learning Platforms).



# الأسس التقنية لإدارة قواعد البيانات العملاقة - نظم إدارة قواعد البيانات DBMS

## • مميزات RDBMS:

- **تكامل البيانات:** دعم قوي للسلامة المرجعية، مما يضمن تكامل البيانات عبر الجداول.
- **التعامل مع المعاملات:** دعم ACID (Atomicity, Consistency, Isolation, Durability) لضمان أمان المعاملات.
- **قابلية التوسع الرأسي:** يمكن تحسين الأداء عن طريق زيادة موارد الخادم الواحد.

## • عيوب RDBMS:

- **القيود في التعامل مع البيانات غير المهيكلة:** قد تواجه صعوبة في التعامل مع البيانات الكبيرة أو غير المهيكلة.
- **قابلية التوسع الأفقي المحدودة:** تزداد التحديات عند محاولة التوسع الأفقي عبر خوادم متعددة.

## 1. نظم قواعد البيانات العلائقية (RDBMS): RDBMS (Relational Database Management System)

تتبع على نموذج البيانات العلائقي، حيث يتم تخزين البيانات في جداول تحتوي على صفوف وأعمدة. تم تصميم هذه النظم للتعامل مع البيانات المهيكلة، وتستخدم لغة الاستعلامات البنوية (SQL) للوصول إلى البيانات وإدارتها.

## • أمثلة على نظم RDBMS:

- **MySQL:** نظام قاعدة بيانات علائقية مفتوح المصدر، مشهور بمرونته وسهولة الاستخدام.
- **PostgreSQL:** نظام RDBMS مفتوح المصدر يدعم مجموعة واسعة من أنواع البيانات والوظائف المتقدمة.
- **Oracle Database:** نظام قاعدة بيانات قوي يدعم تطبيقات المؤسسات الكبيرة.
- **Microsoft SQL Server:** نظام RDBMS معروف بأدائه العالي وتكامله مع بيئة Microsoft.

## الأسس التقنية لإدارة قواعد البيانات العملاقة - نظم إدارة قواعد البيانات DBMS

- 2. نظم قواعد البيانات غير العلائقية (NoSQL): **NoSQL (Not Only SQL)** هي فئة من نظم إدارة قواعد البيانات التي تم تطويرها للتعامل مع البيانات غير المهيكلة، وشبه المهيكلة، والكميات الكبيرة من البيانات. هذه النظم لا تعتمد على الجداول الثابتة وتوفر مرونة أكبر في كيفية تخزين البيانات والوصول إليها.
- **أنواع: NoSQL**
  - **قواعد بيانات الوثائق:** مثل MongoDB ، حيث يتم تخزين البيانات في وثائق JSON أو BSON.
  - **قواعد بيانات الأعمدة العريضة:** مثل Cassandra ، حيث يتم تخزين البيانات في جداول ولكن يتم تنظيمها كأعمدة وليس صفوف.
  - **قواعد بيانات الرسوميات:** مثل Neo4j ، التي تُستخدم لتخزين البيانات في شكل عقد وحواف، مما يسهل عمليات الاستعلام المعقدة للعلاقات.
  - **قواعد بيانات المفتاح والقيمة:** مثل Redis ، حيث يتم تخزين البيانات كأزواج مفتاح وقيمة.
- **مميزات: NoSQL**
  - **قابلية التوسع الأفقي:** يمكن توزيع البيانات عبر خوادم متعددة مما يدعم التوسع الكبير.
  - **التعامل مع البيانات غير المهيكلة:** تدعم تخزين البيانات بأشكال مختلفة مثل الوثائق، الرسوميات، الأطر الزمنية، وغيرها.
  - **أداء عالي للقراءات والكتابات:** مثالية للتطبيقات التي تتطلب عمليات قراءة وكتابة سريعة.
- **عيوب: NoSQL**
  - **غياب المعاملات الصارمة:** قد لا تدعم بعض أنظمة NoSQL قواعد ACID بنفس الصرامة الموجودة في RDBMS.
  - **تعقيد الاستعلامات:** قد تكون الاستعلامات أكثر تعقيداً مقارنة بـ SQL في نظم RDBMS.

# استراتيجيات إدارة البيانات العملاقة - التخزين

1. **التخزين العمودي: (Vertical Partitioning)** : هو أسلوب تقسيم البيانات يعتمد على توزيع الأعمدة (أو الحقول) من الجداول على وحدات تخزين منفصلة.

- **مميزات التخزين العمودي:**

- تحسين أداء الاستعلامات المحددة: هذا النهج مفيد للغاية في الاستعلامات التي تحتاج إلى الوصول إلى عدد قليل من الأعمدة فقط.
- تخزين مضغوط: يمكن أن يؤدي التخزين العمودي إلى ضغط أفضل للبيانات نظرًا لتجانس البيانات المخزنة في كل جزء.
- إدارة سهلة للبيانات المتغيرة: يمكن تعديل الأعمدة أو إضافتها بدون التأثير على البيانات المخزنة في الأعمدة الأخرى.

- **عيوب التخزين العمودي:**

- أداء ضعيف في استعلامات التجميع: عند الحاجة إلى استعلامات تشمل أعمدة متعددة موزعة عبر أقسام مختلفة
- تعقيد في الإدارة: يتطلب التخزين العمودي إدارة معقدة للتأكد من توزيع الأعمدة بشكل مناسب وتجنب الأداء المنخفض.

- **تطبيقات التخزين العمودي:** يستخدم التخزين العمودي بشكل شائع في قواعد البيانات التحليلية وأنظمة تخزين البيانات التي تتطلب الاستعلامات التحليلية السريعة مثل قواعد البيانات. OLAP (Online Analytical Processing)

## استراتيجيات إدارة البيانات العملاقة - التخزين

2. التخزين الأفقي: (Horizontal Partitioning): هو أسلوب آخر لتقسيم البيانات، حيث يتم توزيع الصفوف من الجداول عبر وحدات تخزين مختلفة.

### • مميزات التخزين الأفقي:

- قابلية التوسع الأفقي: يُتيح التخزين الأفقي توزيع البيانات عبر خوادم متعددة، مما يمكن من توسيع النظام بزيادة عدد الخوادم لتلبية الزيادة في حجم البيانات.
- إدارة فعالة للأحجام الكبيرة من البيانات: يتم تقسيم البيانات إلى أجزاء أصغر يمكن إدارتها بسهولة أكبر.
- أداء محسن للقراءة والكتابة: يمكن تقليل الحمل على كل خادم عن طريق توزيع العمليات عبر أقسام متعددة.

### • عيوب التخزين الأفقي:

- تعقيد في استعلامات الانضمام: قد تكون الاستعلامات التي تحتاج إلى دمج البيانات من عدة أقسام أكثر تعقيدًا وتتطلب موارد إضافية.
- إدارة معقدة للتجزئة: تحديد الاستراتيجية المثلى للتجزئة يمكن أن يكون تحديًا، خاصة مع البيانات غير المتجانسة.
- تطبيقات التخزين الأفقي: يُستخدم التخزين الأفقي بشكل واسع في التطبيقات الكبيرة مثل الشبكات الاجتماعية.

## استراتيجيات إدارة البيانات العملاقة - التخزين

### 3. التخزين السحابي:

1. **قابلية التوسع:** إمكانية التوسع بسهولة: يتيح التخزين السحابي للشركات إمكانية توسيع سعة التخزين بسرعة وسهولة دون الحاجة إلى استثمارات كبيرة في البنية التحتية.
  - مرونة عالية: يمكن للمؤسسات توسيع موارد التخزين بشكل سريع لتلبية متطلبات الحمل الزائد أو المتغيرات الموسمية.
2. **الوصول عن بُعد:** الوصول من أي مكان: يمكن الوصول إلى البيانات المخزنة في السحابة من أي مكان وفي أي وقت، طالما كان هناك اتصال بالإنترنت.
  - تحسين الإنتاجية: إمكانية الوصول إلى البيانات بسرعة ودون تعقيد يمكن أن تزيد من كفاءة العمليات اليومية.
3. **الأمان والحماية:** أمان البيانات: تعتمد شركات التخزين السحابي على تقنيات متقدمة لحماية البيانات مثل التشفير وإدارة الهوية والوصول.
  - التحديثات الدائمة: توفر مزودي الخدمات السحابية تحديثات أمان دورية لضمان حماية البيانات من التهديدات الإلكترونية المتزايدة.
4. **التكلفة الفعالة:** تقليل التكاليف التشغيلية: يُعتبر التخزين السحابي خيارًا اقتصاديًا لأنه يلغي الحاجة إلى شراء وصيانة أجهزة تخزين محلية باهظة الثمن.
  - الدفع حسب الاستخدام: تدفع الشركات فقط مقابل ما تستخدمه من سعة تخزين، مما يجعل إدارة التكاليف أكثر مرونة. **التعافي من الكوارث: النسخ الاحتياطي واستعادة البيانات :** يوفر التخزين السحابي حلاً فعالاً للتعافي من الكوارث.

# استراتيجيات إدارة البيانات العملاقة - تقنيات تحسين الأداء

**1. التخزين المؤقت (Caching):** كيف يمكن أن يحسن أداء استرجاع البيانات: التخزين المؤقت هو تقنية تستخدم لتسريع الوصول إلى البيانات من خلال تخزين نسخ من البيانات التي يتم الوصول إليها بشكل متكرر في موقع أسرع للوصول، مثل الذاكرة العشوائية (RAM)

كيف يعمل التخزين المؤقت:

- التخزين المؤقت على مستوى التطبيقات: يتم استخدام الذاكرة المؤقتة (Cache) على مستوى التطبيق لتخزين البيانات التي يتم طلبها بشكل متكرر.
- التخزين المؤقت على مستوى قاعدة البيانات: بعض نظم إدارة قواعد البيانات توفر آليات تخزين مؤقت مدمجة.

فوائد التخزين المؤقت:

1. زيادة سرعة الوصول: يقلل التخزين المؤقت من زمن الوصول إلى البيانات، حيث يتم جلب البيانات من الذاكرة المؤقتة بدلاً من قراءة البيانات من القرص الصلب .
2. تقليل الضغط على قاعدة البيانات: بتقليل عدد الاستعلامات المرسلة إلى قاعدة البيانات، يساعد التخزين المؤقت في تخفيف الحمل على قاعدة البيانات وتحسين أدائها العام.
3. تحسين تجربة المستخدم: يوفر أوقات استجابة أسرع للتطبيقات والخدمات التي تعتمد على البيانات المخزنة مؤقتاً، مما يعزز تجربة المستخدم.

# استراتيجيات إدارة البيانات العملاقة - تقنيات تحسين الأداء

**2. التجزئة: (Sharding) تقسيم البيانات إلى أجزاء صغيرة لتحسين الأداء والتوسع :** التجزئة هي تقنية تستخدم لتوزيع البيانات عبر مجموعة من الخوادم أو قواعد البيانات، بحيث يتم تقسيم البيانات إلى أجزاء صغيرة تسمى "شاردات" (Shards). كل شارد يحتوي على جزء من البيانات ويخزن على خادم منفصل.

## كيف تعمل التجزئة:

- **تقسيم البيانات:** يتم تقسيم البيانات بناءً على معايير محددة مثل القيم المفهرسة أو نطاقات البيانات.
- **التوجيه إلى الشارد المناسب:** عندما يتم إرسال استعلام إلى قاعدة البيانات، يتم توجيهه إلى الشارد الذي يحتوي على البيانات المطلوبة.

## فوائد التجزئة:

1. **تحسين الأداء:** بتوزيع البيانات على خوادم متعددة، يمكن توزيع الأحمال وزيادة سرعة الوصول إلى البيانات، مما يحسن الأداء العام لنظام إدارة قواعد البيانات.
2. **زيادة التوسع:** يمكن إضافة المزيد من الشاردات (والخوادم) للتعامل مع كميات أكبر من البيانات وتوسيع النظام بسهولة.
3. **تقليل التأثيرات الجانبية:** إذا كان أحد الشاردات يعاني من مشاكل، فإن تأثيرها يكون محدودًا ولا يؤثر على باقي النظام.

# استراتيجيات إدارة البيانات العملاقة - النسخ الاحتياطي واستعادة البيانات

استراتيجيات النسخ الاحتياطي: كيفية ضمان عدم فقدان البيانات وتوفير خطط للتعافي

**مفهوم النسخ الاحتياطي:** النسخ الاحتياطي هو عملية إنشاء نسخ من البيانات لحمايتها من فقدان أو التلف الناتج عن الحوادث أو الأعطال. الهدف من النسخ الاحتياطي هو ضمان إمكانية استعادة البيانات في حالة حدوث مشكلة، مثل فقدان البيانات بسبب فشل النظام أو هجوم إلكتروني.

استراتيجيات النسخ الاحتياطي:

## 1. النسخ الاحتياطي الكامل: (Full Backup)

- **الوصف:** يتم نسخ جميع البيانات من النظام إلى موقع تخزين احتياطي. يُعتبر هذا النوع من النسخ الاحتياطي الأكثر شمولاً حيث يشمل كل الملفات والبيانات.
- **فوائد:** يوفر أقصى حماية حيث يتم تخزين جميع البيانات في كل عملية نسخ احتياطي.
- **تحديات:** يتطلب وقتاً طويلاً ومساحة تخزين كبيرة، مما قد يؤثر على أداء النظام أثناء تنفيذ النسخ الاحتياطي.



## استراتيجيات إدارة البيانات العملاقة - النسخ الاحتياطي واستعادة البيانات

**2. النسخ الاحتياطي التزايدى: (Incremental Backup):** يتم نسخ فقط البيانات التي تغيرت منذ آخر نسخ احتياطي. يتطلب هذا النوع مساحة تخزين أقل مقارنةً بالنسخ الاحتياطي الكامل.

- فوائد: أسرع في التنفيذ ويحتاج إلى أقل مساحة تخزين.
- تحديات: يتطلب عملية استعادة أكثر تعقيداً حيث يتعين دمج النسخ الاحتياطية التزايدية مع النسخ الاحتياطي الكامل.

**3. النسخ الاحتياطي التفاضلي: (Differential Backup):** يتم نسخ البيانات التي تغيرت منذ آخر نسخ احتياطي كامل. مثل النسخ الاحتياطي التزايدى، ولكنه يعتمد على النسخ الاحتياطي الكامل فقط.

- فوائد: أسرع من النسخ الاحتياطي الكامل وسهولة أكبر في الاستعادة مقارنةً بالنسخ الاحتياطي التزايدى.
- تحديات: يتطلب مساحة تخزين أكبر مقارنةً بالنسخ الاحتياطي التزايدى، حيث يتزايد حجم البيانات المخزنة بمرور الوقت.

**4. النسخ الاحتياطي في الموقع والنسخ الاحتياطي عن بعد: (On-site and Off-site Backup)**

- النسخ الاحتياطي في الموقع: يتم تخزين النسخ الاحتياطية في نفس الموقع الجغرافي الذي يحتوي على البيانات الأصلية.
- النسخ الاحتياطي عن بعد: يتم تخزين النسخ الاحتياطية في موقع جغرافي منفصل (مثل السحابة أو مركز بيانات بعيد) لحماية البيانات من الكوارث المحلية.

# إدارة الأمان والخصوصية في قواعد البيانات العملاقة \_ الأمان

1. التحكم بالوصول : التحكم بالوصول هو عملية ضمان أن الأشخاص أو الأنظمة المسموح لهم فقط يمكنهم الوصول إلى البيانات في قاعدة البيانات.

استراتيجيات التحكم بالوصول:

1. التحقق من الهوية: (Authentication) : يشير إلى عملية التحقق من هوية المستخدمين أو الأنظمة التي تحاول الوصول إلى قاعدة البيانات. يتم ذلك عبر أسماء المستخدمين وكلمات المرور أو أساليب التحقق الأخرى مثل التعرف على الوجه أو الرموز المرسلات عبر البريد الإلكتروني.

2. التحكم بالامتيازات: (Authorization) : بعد التحقق من الهوية، يتم تحديد مستوى الوصول المسموح به للمستخدم أو النظام.

○ أنواع التحكم بالامتيازات:

- التحكم بالوصول القائم على الدور: (RBAC) تعيين أدوار محددة للمستخدمين وتحديد ما يمكنهم فعله بناءً على أدوارهم.
- التحكم بالوصول القائم على الهوية: (ABAC) استخدام مجموعة من الخصائص (مثل المسمى الوظيفي، الموقع، الوقت) لتحديد الوصول.

3. سجلات الأنشطة: (Audit Logs) : تتبع الأنشطة التي تتم على قاعدة البيانات، بما في ذلك عمليات القراءة والكتابة والتعديلات.

# إدارة الأمان والخصوصية في قواعد البيانات العملاقة \_ الأمان

**مفهوم التشفير:** التشفير هو عملية تحويل البيانات إلى شكل غير قابل للقراءة إلا من قبل الأشخاص المصرح لهم باستخدام مفتاح تشفير. يهدف إلى حماية البيانات من الوصول غير المصرح به حتى إذا تم الحصول عليها.

## أنواع التشفير:

- 1. التشفير في الراحة: (Data-at-Rest Encryption) :** يشير إلى تشفير البيانات المخزنة في قاعدة البيانات أو على وسائط التخزين.
  - التقنيات الشائعة (AES (Advanced Encryption Standard : هو أحد أكثر تقنيات التشفير شيوعاً لتشفير البيانات المخزنة.
- 2. التشفير أثناء النقل: (Data-in-Transit Encryption) :** يشير إلى تشفير البيانات أثناء انتقالها عبر الشبكات.
  - التقنيات الشائعة (TLS (Transport Layer Security و SSL (Secure Sockets Layer هما تقنيتان تستخدمان لتشفير البيانات أثناء النقل.
- 3. تشفير الحقول: (Field-Level Encryption) :** يشير إلى تشفير بيانات محددة داخل قاعدة البيانات، بدلاً من تشفير قاعدة البيانات بالكامل.
  - فوائد: يوفر حماية مخصصة للبيانات الحساسة ويقلل من حجم البيانات المشفرة.

# إدارة الأمان والخصوصية في قواعد البيانات العملاقة \_ الخصوصية والامتثال

قوانين حماية البيانات: قوانين حماية البيانات تهدف إلى تنظيم كيفية جمع، استخدام، وتخزين المعلومات الشخصية.

1. اللائحة العامة لحماية البيانات: (GDPR) : اللائحة العامة لحماية البيانات هي قانون أوروبي ينظم حماية البيانات الشخصية للأفراد داخل الاتحاد الأوروبي.

## • المتطلبات الرئيسية:

- الشفافية: يتعين على المؤسسات إبلاغ الأفراد بوضوح عن كيفية استخدام بياناتهم. **الموافقة:** يجب الحصول على موافقة صريحة من الأفراد قبل جمع بياناتهم.
- حقوق الأفراد: تشمل حقوق الوصول إلى البيانات، تصحيح البيانات، طلب حذف البيانات (الحق في النسيان)، وحقوق نقل البيانات.
- إشعار بانتهاك البيانات: يجب إبلاغ السلطات المعنية والأفراد المتأثرين في حالة حدوث خرق للبيانات.

2. قانون حماية المعلومات الصحية القابلة للنقل والمساءلة: (HIPAA) : قانون HIPAA هو قانون أمريكي ينظم حماية المعلومات الصحية الشخصية للأفراد.

## • المتطلبات الرئيسية: حماية البيانات: يشمل تدابير للحفاظ على أمان المعلومات الصحية وحمايتها من الوصول غير المصرح به.

- الامتثال: يتطلب من المؤسسات الصحية أن تتبع معايير الأمان للبيانات واتباع إجراءات معينة للحفاظ على خصوصية المعلومات الصحية.

## إدارة الأمان والخصوصية في قواعد البيانات العملاقة \_ الخصوصية والامتثال

**مفهوم سياسات الخصوصية:** سياسات الخصوصية هي وثائق توضح كيفية جمع، استخدام، وتخزين المعلومات الشخصية للأفراد. تهدف إلى إبلاغ الأفراد بكيفية التعامل مع بياناتهم وضمان أنها تُعامل بطرق تتوافق مع القوانين واللوائح ذات الصلة.

**عناصر أساسية لسياسات الخصوصية:**

### 1. وصف جمع البيانات:

- **المعلومات التي يتم جمعها:** يجب توضيح أنواع المعلومات التي يتم جمعها، مثل المعلومات الشخصية، بيانات الاتصال، والمعلومات المالية.
- **طرق الجمع:** شرح كيف يتم جمع البيانات، مثل التسجيل عبر الإنترنت، استبيانات، أو جمع البيانات تلقائيًا.

### 2. استخدام البيانات:

- **الأغراض:** تحديد الأغراض التي تُستخدم البيانات من أجلها، مثل تحسين الخدمات، إرسال إشعارات، أو إجراء تحليلات.
- **مشاركة البيانات:** شرح ما إذا كانت البيانات تُشارك مع أطراف ثالثة، مثل شركاء الأعمال أو مقدمي الخدمات، وكيفية حماية البيانات أثناء مشاركتها.

## إدارة الأمان والخصوصية في قواعد البيانات العملاقة \_ الخصوصية والامتثال

### 3. حقوق الأفراد:

- الوصول والتصحيح: توضيح كيفية تمكين الأفراد من الوصول إلى بياناتهم وتصحيحها إذا كانت غير دقيقة.
- الانسحاب: توفير معلومات حول كيفية الانسحاب من تلقي الرسائل أو الخدمات غير المرغوب فيها.

### 4. إجراءات الأمان:

- حماية البيانات: وصف التدابير المتخذة لحماية البيانات من الوصول غير المصرح به أو التلاعب.
- إجراءات التشفير: شرح كيفية استخدام التشفير لحماية البيانات المخزنة والمنقولة.

### 5. الإشعارات والتحديثات:

- الإشعارات: توضيح كيفية إخطار الأفراد بأي تغييرات في سياسة الخصوصية.
- التحديثات: ضمان تحديث سياسة الخصوصية بانتظام لتكون متوافقة مع القوانين والتكنولوجيا الحديثة.

# أدوات وتقنيات في إدارة قواعد البيانات العملاقة - الإدارة

أدوات إدارة قواعد البيانات العملاقة:

**1. Apache Hadoop:** هو إطار عمل مفتوح المصدر لإدارة وتحليل البيانات الكبيرة. يعتمد Hadoop على نموذج البرمجة الموزعة الذي يقسم البيانات عبر مجموعة من الخوادم، مما يتيح معالجة كميات ضخمة من البيانات بكفاءة.

المكونات الرئيسية:

- **HDFS (Hadoop Distributed File System):** نظام ملفات موزع يتيح تخزين البيانات عبر العديد من الخوادم بطريقة موثوقة وقابلة للتوسع.
- **MapReduce:** نموذج برمجة لمعالجة البيانات، حيث يتم تقسيم المهمة إلى مراحل (Map) و (Reduce) لتوزيع وتحليل البيانات بشكل متوازي.

**2. Apache Spark:** هو إطار عمل مفتوح المصدر لمعالجة البيانات الكبيرة بشكل أسرع من Hadoop، مع دعم للمعالجة في الذاكرة (In-Memory Processing)، مما يحسن الأداء بشكل كبير.

الميزات الرئيسية:

- **معالجة في الذاكرة:** يقوم بتخزين البيانات في الذاكرة بدلاً من قراءتها وكتابتها من/إلى القرص الصلب بشكل متكرر، مما يسرع عمليات التحليل.
- **دعم متعدد الأنماط:** يدعم Spark أنماط معالجة مختلفة مثل التحليل التفاعلي، التحليل المعتمد على الجداول، والتعلم الآلي.

# أدوات وتقنيات في إدارة قواعد البيانات العملاقة - الإدارة

## 3. NoSQL Databases:

- الوصف: قواعد البيانات غير العلائقية (NoSQL) مصممة للتعامل مع البيانات غير المنظمة أو شبه المنظمة وتوفير أداء عالي في البيئات الموزعة.
- أنواع قواعد البيانات: NoSQL
  - قواعد البيانات الوثائقية: مثل MongoDB و CouchDB، تخزن البيانات كوثائق مثل (JSON) مما يتيح مرونة في تخزين البيانات.
  - قواعد البيانات العمودية: مثل Cassandra و HBase، تستخدم تخزين البيانات بشكل عمودي مما يساعد في تحسين الأداء للقراءات والكتابات على نطاق واسع.
  - قواعد البيانات المفتاحية-القيمة: مثل Redis و Riak، تستخدم لتخزين البيانات كمفتاح وقيمة، وهي فعالة في معالجة البيانات السريعة.



# أدوات وتقنيات في إدارة قواعد البيانات العملاقة - التحليل

## أدوات تحليل الأداء:

### 1. أدوات مراقبة الأداء:

- **Nagios:** أداة مراقبة مفتوحة المصدر تساعد في تتبع الأداء وإدارة التنبيهات لتطبيقات وقواعد البيانات.
- **Prometheus:** أداة مراقبة تجمع البيانات من قواعد البيانات والخوادم وتوفر مؤشرات الأداء بشكل مرّن وقابل للتوسع.
- **Grafana:** أداة لتصور البيانات المجمعة من أدوات المراقبة، توفر لوحات بيانات تفاعلية لمراقبة الأداء وتحليل البيانات.

### 2. تحليل الأداء:

- **تحليل السجلات:** استخدام أدوات مثل ELK Stack (Elasticsearch, Logstash, Kibana) لتحليل سجلات الأداء واكتشاف المشكلات في الوقت الفعلي.
- **تحليل المقاييس:** تحليل مؤشرات الأداء الرئيسية مثل أوقات الاستجابة، استخدام الموارد، وسرعة المعالجة لتحديد المشكلات وتحسين الأداء.

# التحديات المستقبلية في إدارة قواعد البيانات العملاقة

## 1. مشاكل تأخير الاستجابة:

الشرح: مع تزايد حجم البيانات، قد تواجه أنظمة إدارة قواعد البيانات صعوبة في الحفاظ على سرعة الاستجابة المطلوبة. التأخير في الاستجابة يمكن أن يؤثر بشكل كبير على تجربة المستخدم وفعالية العمليات التجارية.

### الأسباب:

■ زيادة حجم البيانات: التعامل مع كميات ضخمة من البيانات يمكن أن يؤدي إلى أوقات استعلام أطول، خاصة إذا كانت البيانات موزعة عبر عدة خوادم.

■ التكامل والتكرار: عمليات التكامل بين البيانات المتكررة أو الموزعة قد تزيد من تعقيد الاستعلامات وتؤدي إلى تأخير إضافي.

### الحلول الممكنة:

■ استخدام تقنيات التخزين المؤقت: (Caching) تحسين أوقات الاستجابة من خلال تخزين البيانات المتكررة في الذاكرة.

■ تحسين استراتيجيات التجزئة: (Sharding) تقسيم البيانات بشكل أكثر كفاءة لتقليل الحمل على أي خادم واحد.

■ التحليل الموزع: استخدام تقنيات تحليل البيانات الموزعة مثل Apache Spark لتحسين سرعة المعالجة.

# التحديات المستقبلية في إدارة قواعد البيانات العملاقة

## 2. التوسع: (Scalability)

- الشرح: التوسع هو القدرة على زيادة الموارد (مثل الخوادم أو التخزين) بطريقة سلسلة وفعالة لتلبية الطلب المتزايد على البيانات.
- التحديات:
  - إدارة الموارد: زيادة الموارد قد يتطلب تعديلات في بنية النظام لضمان التنسيق الفعال بين الخوادم.
  - التكلفة: التوسع يمكن أن يكون مكلفاً، حيث يتطلب إضافة المزيد من الخوادم أو تحسين البنية التحتية.
- الحلول الممكنة:
  - التوسع الأفقي: (Horizontal Scaling) إضافة المزيد من الخوادم بدلاً من زيادة قدرة الخادم الواحد.
  - التوسع العمودي: (Vertical Scaling) تحسين قدرات الخوادم الحالية من خلال زيادة الموارد مثل الذاكرة والتخزين.

## دور الذكاء الاصطناعي والتعلم الآلي في تحسين إدارة قواعد البيانات

### 1. تحسين إدارة قواعد البيانات:

- **التنبؤ بالطلب:** الذكاء الاصطناعي والتعلم الآلي يمكن أن يساعد في التنبؤ بالأنماط المستقبلية لاستخدام البيانات، مما يمكن أن يحسن التخزين والتوسع.
  - **إدارة الأداء:** تقنيات الذكاء الاصطناعي يمكن أن تحلل باستمرار الأداء الفعلي لقواعد البيانات وتحديد المشكلات المحتملة مثل اختناقات الأداء أو استخدام الموارد بشكل غير فعال.
  - **التعرف على الأنماط الشاذة:** يمكن للتعلم الآلي أن يكتشف الأنماط غير المعتادة في البيانات التي قد تشير إلى مشكلات مثل الفساد في البيانات أو الهجمات الأمنية.
  - **تحسين جودة البيانات:** تقنيات الذكاء الاصطناعي يمكن أن تساعد في تنظيف البيانات من الأخطاء والبيانات غير المتناسقة بشكل تلقائي.
2. **تحليل البيانات الذكي:** أدوات الذكاء الاصطناعي والتعلم الآلي يمكن أن تدير وتحلل البيانات بشكل أكثر فعالية، مما يوفر رؤى قيمة دون الحاجة لتحليل يدوي طويل.

## نهاية المحاضرة

آمل أن تكونوا قد حققتم الفائدة