



Université Constantine 2
جامعة قسنطينة 2

حوسبة سحابية

-الدورة - 5

الفصل الخامس: تخزين البيانات في السحابة

دكتور منور ر.

كلية التقنيات الجديدة

rostom.mennour@univ-constantine2.dz

الطلاب المعنيين

الكلية/المعهد	قسم	مستوى	تخصص
التكنولوجيات الجديدة	إذا كان	سيد 1	سديا

القسم 1: أهمية التخزين

أهمية التخزين

كمية البيانات

• في عام 2008، قامت Google بمعالجة ما يقرب من 20 بيتابايت من البيانات يوميًا (لا تنشر Google معلومات عن كمية البيانات التي تعالجها اليوم).

• وفقًا لتقديرات شركة سيسكو، وصل الاستخدام الجماعي للإنترنت في جميع أنحاء العالم إلى زيتابايت واحد في عام 2016.

• وفقًا لشركة IDC لاستخبارات السوق، وصل حجم "مساحة البيانات العالمية" في عام 2018 إلى 18 زيتابايت. هذا هو إجمالي جميع البيانات التي تم إنشاؤها أو التقاطها أو نسخها.

(ومع ذلك، لا يتم تخزين جميع هذه البيانات والاحتفاظ بها).

• تم إنشاء الغالبية العظمى من البيانات العالمية في السنوات الأخيرة، ولا يُظهر هذا النمو المذهل للبيانات أي علامات على التباطؤ. في الواقع، تتوقع مؤسسة IDC أن البيانات العالمية سترتفع إلى 175 زيتابايت في عام 2025.

ماذا يعني

إذا كنت ستقوم بتخزين 175 زيتابايت على أقراص DVD، فستكون حزمة أقراص DVD الخاصة بك طويلة بما يكفي للتنقل الأرض 222 مرة.

إذا حاولت تنزيل 175 زيتابايت بمتوسط سرعة الاتصال بالإنترنت الحالي، فسيقوم التنزيل سيأخذك 1.8 مليار سنة. حتى لو ناشدت الجميع العالم للمساعدة في التنزيل، سيستغرق الأمر 81 يومًا أخرى.

أهمية التخزين

كمية البيانات

A DAY IN DATA

The exponential growth of data is undisputed, but the numbers behind this explosion - fuelled by internet of things and the use of connected devices - are hard to comprehend, particularly when looked at in the context of one day

500m
tweets are sent every day
Twitter



4PB
of data created by Facebook, including

350m photos
100m hours of video watch time
Facebook Research

294bn
billion emails are sent
Radicati Group

320bn

emails to be sent each day by 2021

306bn

emails to be sent each day by 2020

3.9bn
people use emails

4TB
of data produced by a connected car
Intel

ACCUMULATED DIGITAL UNIVERSE OF DATA

4.4ZB

PwC

2013

44ZB

2020

DEMYSIFYING DATA UNITS

From the more familiar 'bit' or 'megabyte', larger units of measurement are more frequently being used to explain the masses of data

Unit	Value	Size
b	bit	0 or 1
B	byte	1/8 of a byte
KB	kilobyte	1,000 bytes
MB	megabyte	1,000 ² bytes
GB	gigabyte	1,000 ³ bytes
TB	terabyte	1,000 ⁴ bytes
PB	petabyte	1,000 ⁵ bytes
EB	exabyte	1,000 ⁶ bytes
ZB	zettabyte	1,000 ⁷ bytes
YB	yottabyte	1,000 ⁸ bytes

*A lowercase "b" is used as an abbreviation for bits, while an uppercase "B" represents bytes.

65bn
messages sent over WhatsApp and two billion minutes of voice and video calls made
Facebook

Searches made a day **5bn**

Searches made a day from Google **3.5bn**



463EB

of data will be created every day by 2025
IDC

95m
photos and videos are shared on Instagram
Instagram Business

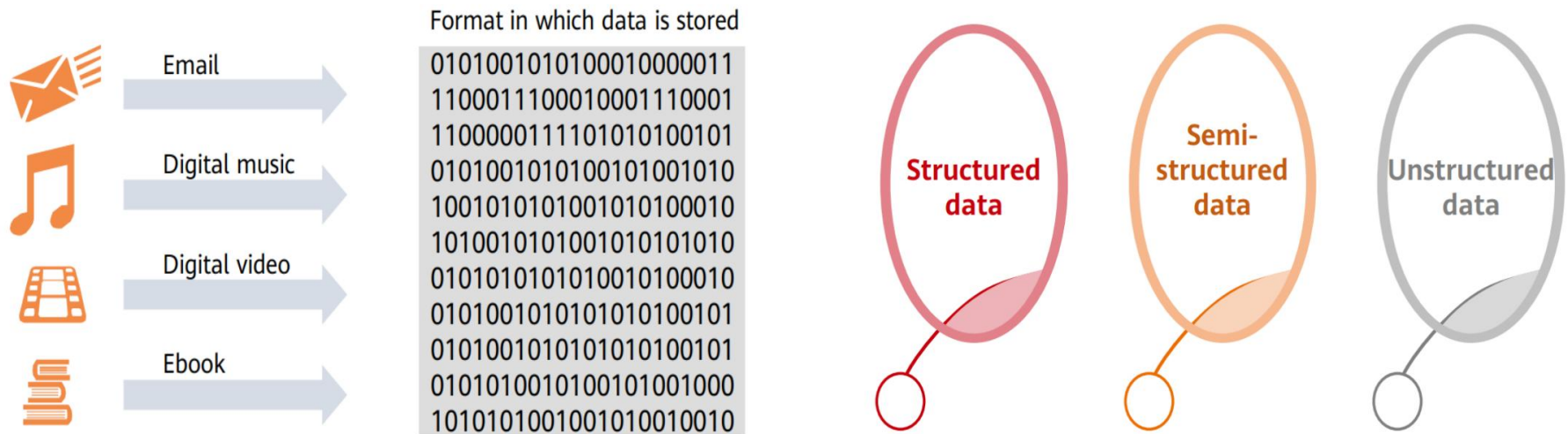
28PB
to be generated from wearable devices by 2020
Statista



RACONTEUR

أهمية التخزين

المعطيات



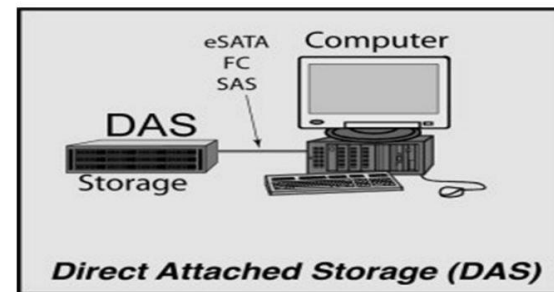
تُعرّف SNIA (رابطة صناعة شبكات التخزين) البيانات بأنها التمثيل الرقمي لأي شيء بأي شكل من الأشكال

القسم 2: تقنيات التخزين

تقنيات التخزين

التخزين المرفق المباشر

- التخزين المتصل المباشر (DAS) هو تخزين رقمي متصل مباشرة بالكمبيوتر الذي يصل إليه.
- يتكون نظام DAS النموذجي من جهاز تخزين بيانات متصل مباشرة بالكمبيوتر عبر محول الناقل المضيق (HBA).
- بين هاتين النقطتين، لا يوجد جهاز شبكة (مثل لوحة الوصل أو المحول أو جهاز التوجيه)، وهذه هي الميزة الرئيسية لـ DAS.
- البروتوكولات الرئيسية المستخدمة لاتصالات DAS هي eSATA و SATA و ATA و eMVN و USB 3.0 و USB و SAS و SCSI و EEEI. 1394.
- DAS هي علاقة 1:1 بين وحدة التخزين والمضيف.
- يعد DAS أيضًا أسرع بشكل عام من تخزين الشبكة، حيث تقتصر معظم الشبكات على 1 جيجابايت في الثانية، بينما يدعم USB 3.0 الشائع في محركات الأقراص الخارجية، ما يصل إلى 5 جيجابايت في الثانية.



تقنيات التخزين

شبكة التخزين المرفقة

- التخزين المتصل بالشبكة (NAS) هو خادم لتخزين بيانات الكمبيوتر على مستوى الملف (على عكس مستوى الكتلة) متصل بشبكة كمبيوتر يوفر الوصول إلى البيانات لمجموعة غير متجانسة من العملاء.
- وحدة NAS هي جهاز كمبيوتر متصل بشبكة توفر فقط خدمات تخزين البيانات المستندة إلى الملفات للأجهزة الأخرى الموجودة على الشبكة.
- ليس من الضروري وجود نظام تشغيل كامل على جهاز NAS، لذلك غالبًا ما يتم استخدام نظام تشغيل بسيط. مثال: FreeNAS أو NAS4Free.
- يستخدم NAS البروتوكولات المستندة إلى الملفات مثل NFS (الشائعة في أنظمة SMB (Server Letter Block) و UNIX) (المستخدمة مع أنظمة AFP و MS Windows) (المستخدمة مع أجهزة كمبيوتر Apple Macintosh). نادراً ما تقصر وحدات NAS العملاء على بروتوكول واحد.



تقنيات التخزين

شبكة منطقة التخزين

• شبكة منطقة التخزين (SAN) أو شبكة التخزين هي شبكة كمبيوتر توفر الوصول إلى تخزين البيانات الموحدة على مستوى الكتلة.

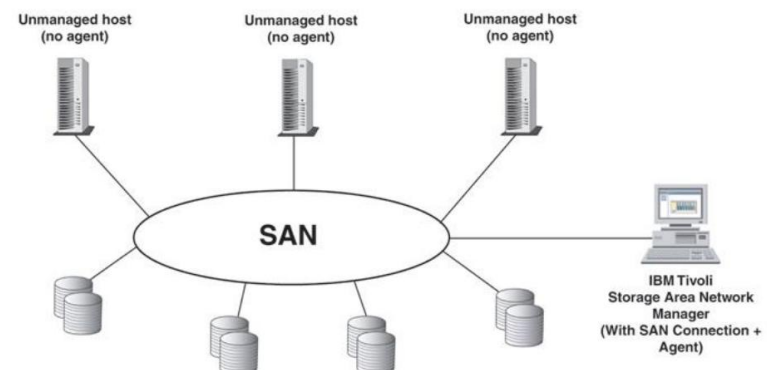
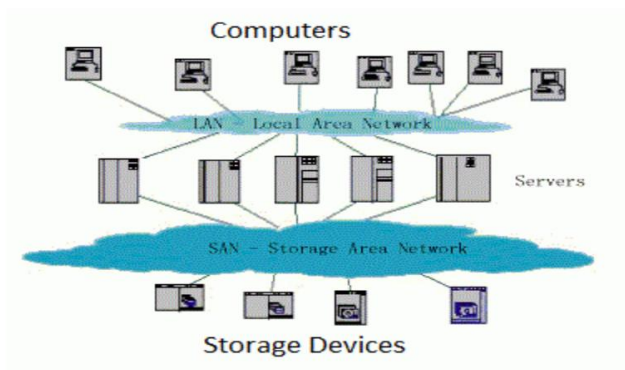
• يتم تمييز شبكة التخزين عن أنظمة التخزين الأخرى عن طريق الوصول على مستوى منخفض إلى الأقراص.

• لا تظهر صفائف التخزين كوحدات تخزين مشتركة على الشبكة. يمكن الوصول إليها مباشرة في وضع الحظر من خلال نظام ملفات الخادم.

• يرى كل خادم أن مساحة القرص الموجودة في حاوية SAN التي يمكنه الوصول إليها هي مساحة خاصة به

قرص صلب.

- تستخدم معظم شبكات التخزين بروتوكول SCSI للاتصال بين الخوادم ومحركات الأقراص. يتم استخدام طبقة التعيين للبروتوكولات الأخرى لتشكيل شبكة.



تقنيات التخزين

مقارنة

- يتمثل الاختلاف الرئيسي بين NAS وDAS في أن تخزين DAS لا يمكن الوصول إليه مباشرة إلا من المضيف الذي يتصل به. DAS لا يتضمن أي أجهزة شبكة أو بيئة تشغيل مرتبطة بها لتوفير وسيلة لمشاركة موارد التخزين بشكل مستقل.

• في حالة NAS، يكون مصدر التخزين متصلاً مباشرة بشبكة IP الخاصة به
مقابلة. يتضمن خادم NAS دعمًا لأنظمة ملفات الشبكة المتعددة (مثل NFS).

• في حالة شبكة SAN، تظهر صفائف التخزين كوحدات تخزين مشتركة على الشبكة. يمكن الوصول إليها مباشرة في وضع الحظر من خلال نظام الملفات الخاص بـ

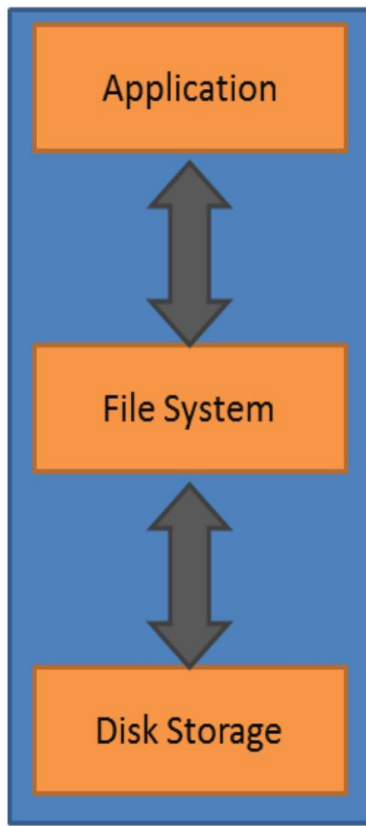
الخوادم.

• تشترك شبكة SAN مع DAS أكثر من NAS والفرق الرئيسي هو أن DAS هي علاقة 1:1 بين وحدة التخزين والمضيف بينما تكون شبكة SAN متعددة.

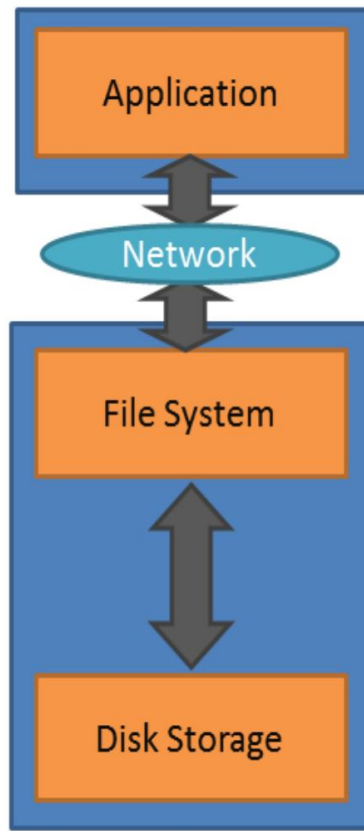
تقنيات التخزين

مقارنة

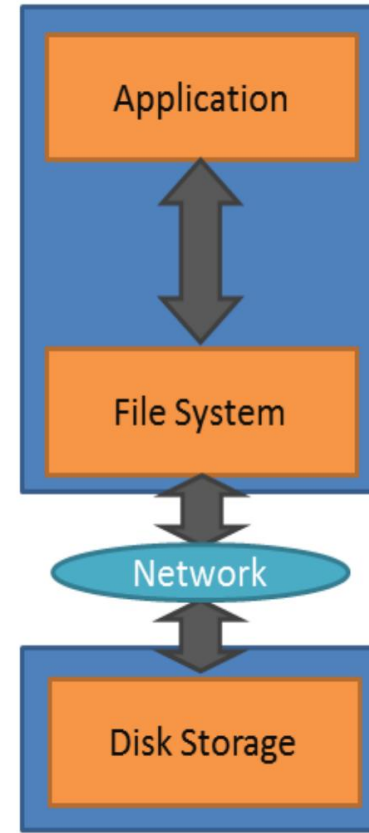
Direct Attached Storage



Network Attached Storage



Storage Area Network



القسم 3: أوضاع تخزين مختلفة

أوضاع تخزين مختلفة

كتلة التخزين

• ما هو تخزين الكتلة؟

• كتلة بيانات مجموعات التخزين في كتل يتم تخزينها كـ

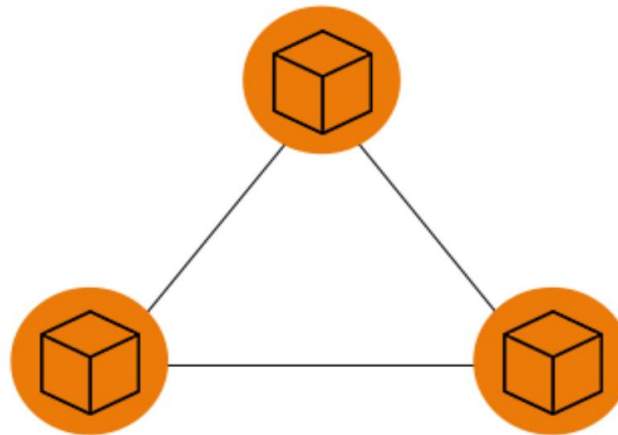
كعناصر منفصلة.

• يتم تعيين معرف فريد لكل كتلة من البيانات، مما يسمح لنظام التخزين بالاحتفاظ بالعناصر الصغيرة في الموقع الأكثر ملاءمة.

• غالبًا ما يتم تكوين وحدة تخزين الكتل لفصل البيانات عن بيئة المستخدم.

• عندما يطلب المستخدم الوصول إلى البيانات، يقوم برنامج التخزين الأساسي بإعادة تجميع كتل البيانات وتمثيلها لهم.

• عادةً ما يتم نشر نظام تخزين الكتل في بيئات شبكة منطقة التخزين (SAN) ويجب ربطه بخادم قيد التشغيل.



أوضاع تخزين مختلفة

كتلة التخزين

• مزايا تخزين الكتل

• الأداء: يعد تخزين الكتل سريعًا للغاية وغالبًا ما يتم استخدامه بواسطة التطبيقات التي تتطلب عمليات إدخال وإخراج سريعة في الثانية. (IOPS)

• المرونة: تتميز أحجام البيانات بالمرونة الشديدة، ويمكن توسيع الأحجام بسهولة دون توقف مع نمو أعمالك، كما يمكن إضافة أحجام كتل جديدة حسب الحاجة.

• السرعة: يمكن نقل وحدة تخزين الكتل من خادم إلى آخر بسهولة نسبية.

من خلال تغيير بسيط في مسار هدف SCSI، يمكن توجيه البيانات إلى خادم آخر.

• انخفاض تحميل FS: عندما يتمكن البرنامج من الكتابة مباشرة إلى كتل التخزين، فإنه يزيل الوقت الإضافي المفروض على أنشطة القراءة/الكتابة لنظام الملفات.

• التمهيد: يمكن لأنظمة التشغيل التمهيد مباشرة من كتلة التخزين المقدمة عبر شبكات التخزين. كل ما تحتاجه هو خادم فعلي أو افتراضي مزود بنظام BIOS قادر على التشغيل في وضع SAN.

أوضاع تخزين مختلفة

كتلة التخزين

• عيوب تخزين الكتل

- مقفل على الخادم - يتم قفل حجم كتلة التخزين على الخادم
فردًا في وقت واحد ولا يمكن الوصول إليه في أي مكان آخر في البيئة من خادم مستهدف آخر. هناك بعض الحلول البرمجية لمكافحة هذا، ولكن كل ذلك يزيد من حمل الإدخال/الإخراج.
- البيانات التعريفية المحدودة: الكمية المحدودة من المعلومات المخزنة في البيانات التعريفية للتخزين الكتلي أقل بكثير من تلك الموجودة في التخزين على مستوى الملف أو الكائن. قد تعاني بعض التطبيقات التي تستخدم بيانات التعريف التفصيلية من هذا.
- التكلفة: في حالة السحابة، يمكن أن يكون تخزين الكتل خيارًا ميسور التكلفة
كثيراً؛ ومع ذلك، بالنسبة لمقدمي الخدمات المُدارة، تكون التكاليف مرتفعة نظرًا لأن شراء شبكات التخزين المرفقة (SAN) باهظ الثمن وغالبًا ما تتطلب خبراء تخزين متخصصين لإدارة تحسين مصفوفة البيانات.

أوضاع تخزين مختلفة

كتلة التخزين

• حظر حالات استخدام التخزين

• قواعد البيانات: تتطلب قواعد البيانات عمليات قراءة/كتابة سريعة ويتم تركيب وحدات تخزين الكتل مباشرة على نظام الملفات، وبالتالي فإن التأخير في كتابة البيانات واسترجاعها يكون في حده الأدنى.

• البيانات المؤقتة: مراحل مختلفة من دورة حياة تطوير البرمجيات

يمكن إنشاء كميات كبيرة من البيانات المؤقتة. يعد إنشاء وحدة تخزين كتلة للبيانات المؤقتة حالة استخدام جيدة عندما لا تكون مساحة التخزين المتوفرة مع المثل الخاص بك كافية أو عندما تحتاج إلى مساحة إضافية لفترة قصيرة من الوقت.

• النسخ الاحتياطي للبيانات: يُعد تخزين الكتل وسيلة نسخ احتياطي قابلة للتطوير ويمكن الوصول إليها بسرعة. قم بتخزين أي شيء قد تحتاج إلى نقله بسرعة إلى مثل آخر أو أي شيء قد تحتاجه على الإطلاق

لحظة.

• يمكنك التمهيد من صور القرص المثبتة على وحدة تخزين

في كتل.

أوضاع تخزين مختلفة

تخزين وضع الملف

• ما هو تخزين الملفات

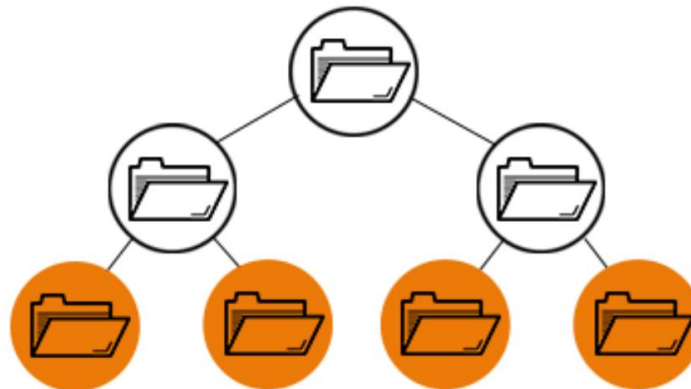
• يتم تخزين البيانات كقطعة واحدة من المعلومات بداخلها

من الملف.

• للوصول إلى البيانات، يجب أن يعرف الكمبيوتر المسار الذي يجب أن يسلكه للعثور عليه.

• يتم تنظيم البيانات المخزنة في الملفات واسترجاعها باستخدام بعض البيانات الوصفية التي تخبر الكمبيوتر بمكان وجود الملف.

- أنها تسمح لك بتخزين جميع أنواع البيانات تقريبًا.
- إنها مثالية لتخزين مجموعات معقدة من الملفات وسريعة التصفح.



أوضاع تخزين مختلفة

تخزين وضع الملف

•مزايا التخزين في وضع الملف

•البساطة: يعد تخزين الملفات الطريقة الأبسط والأكثر شيوعًا لتنظيم الملفات والمجلدات على محرك الأقراص الثابتة بجهاز الكمبيوتر أو جهاز NAS. ليست هناك حاجة لكتابة تطبيقات أو تعليمات برمجية للوصول إلى بياناتك .

•مشاركة الملفات: يعد تخزين الملفات مثالًا لمركزية الملفات ومشاركتها عبر شبكة محلية . (LAN) يمكن الوصول بسهولة إلى الملفات المخزنة على جهاز NAS بواسطة أي كمبيوتر على الشبكة يتمتع بالأذونات المناسبة.

•البروتوكولات الشائعة: يستخدم تخزين الملفات بروتوكولات على مستوى الملف تلك الشائعة مثل SMB (كتلة رسائل الخادم) أو CIFS (نظام ملفات الإنترنت المشترك) أو NFS (نظام ملفات الشبكة).

•حماية البيانات: تخزين الملفات على جهاز تخزين منفصل، متصل بالشبكة المحلية، يوفر حماية للبيانات في حالة فشل الشبكة.

الخادم.

•بأسعار معقولة: يتيح لك تخزين الملفات باستخدام جهاز NAS نقل الملفات من أجهزة الكمبيوتر الباهظة الثمن إلى جهاز تخزين متصل بشبكة LAN بتكلفة معقولة.

أوضاع تخزين مختلفة

تخزين وضع الملف

• حالات استخدام تخزين وضع الملف

• مشاركة الملفات المحلية: إذا كانت احتياجات تخزين البيانات الخاصة بك كذلك

متسقة وبسيطة بشكل عام، مثل تخزين الملفات ومشاركتها مع أعضاء الفريق في المكتب، خذ في الاعتبار بساطة التخزين على مستوى الملف.

• التعاون المركزي في الملفات: إذا قمت بتحميل الملفات وتخزينها ومشاركتها في مكتبة مركزية (توجد داخل مقر العمل أو خارج الموقع أو في السحابة)، فيمكنك التعاون بسهولة في الملفات مع مستخدمين داخليين وخارجيين أو مع ضيوف من الخارج جهاز الكمبيوتر الخاص بك/الشبكة.

• الأرشفة/التخزين: يمكنك أرشفة الملفات بطريقة فعالة من حيث التكلفة

أجهزة NAS في بيئة مركز بيانات صغيرة أو اشترك في خدمة تخزين الملفات المستندة إلى السحابة لتخزين بياناتك وأرشفتها.

• النسخ الاحتياطي/التعافي من الكوارث: يمكنك تخزين النسخ الاحتياطية بسهولة

آمن على أجهزة تخزين منفصلة، متصلة بالشبكة المحلية. يمكنك أيضًا الاشتراك في خدمة تخزين الملفات المستندة إلى السحابة لنسخ ملفات البيانات الخاصة بك عبر مراكز بيانات متعددة متفرقة جغرافيًا والاستفادة من حماية البيانات الإضافية للمسافة والتكرار.

أوضاع تخزين مختلفة

تخزين وضع الكائن

• ما هو تخزين الكائنات؟

• تخزين الكائنات عبارة عن بنية لتخزين البيانات لمعالجة كميات كبيرة من البيانات غير المنظمة.

• الكائنات هي وحدات منفصلة من البيانات المخزنة في بيئة التخزين.

البيانات المسطحة هيكلية.

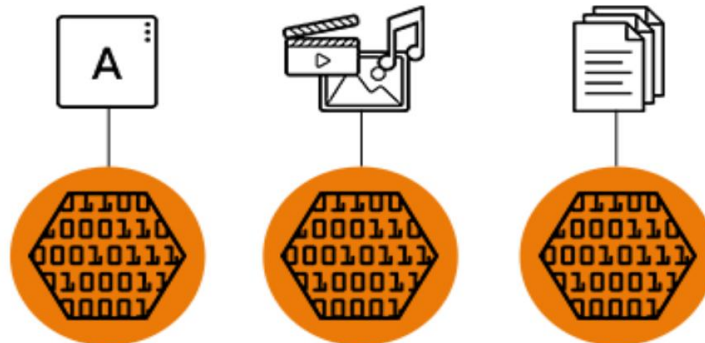
• كل كائن عبارة عن مستودع بسيط ومكتفي بذاته يتضمن البيانات وبيانات التعريف (المعلومات الوصفية المرتبطة بالكائن)

ورقم تعريف فريد (بدلاً من اسم الملف والمسار).

• يؤدي تخزين الكائنات إلى إزالة مشكلات التعقيد وقابلية التوسع من النظام

تنسيق الملف الهرمي مع المجلدات والدلائل.

• يتم الوصول إلى الكائنات (البيانات) الموجودة في نظام تخزين الكائنات من خلال واجهات برمجة التطبيقات (APIs).



أوضاع تخزين مختلفة

تخزين وضع الكائن

•مزايا التخزين في وضع الكائن

•تجسيم غير محدود: يتم تخزين الكائنات في بيئة البيانات مسطحة هيكلياً، داخل جهاز التخزين. يمكنك ببساطة إضافة المزيد من الأجهزة/الخوادم بالتوازي مع مجموعة تخزين الكائنات لمزيد من المعالجة ودعم الإنتاجية الأعلى التي تتطلبها الملفات الكبيرة مثل مقاطع الفيديو أو الصور.

•الحد من التعقيد: يؤدي تخزين الكائنات إلى إزالة التعقيد المتأصل في أنظمة الملفات الهرمي مع المجلدات والدلائل. يؤدي ذلك إلى تحسين الأداء، خاصة عند التعامل مع كميات كبيرة جداً من البيانات.

•التوفر: يمكنك تكوين أنظمة تخزين الكائنات بحيث تكون تكرار المحتوى. في حالة فشل قرص داخل المجموعة، يتوفر قرص مكرر ، مما يضمن استمرار النظام في العمل دون انقطاع أو تدهور الأداء.

•إدارة البيانات غير المنظمة: يتم اعتماد تخزين الكائنات بشكل متزايد في عصر الحوسبة السحابية وإدارة البيانات غير المنظمة والتي، وفقاً للمحللين، ستمثل في المستقبل القريب الغالبية العظمى من جميع البيانات حول العالم.

أوضاع تخزين مختلفة

تخزين وضع الكائن

• مزايا التخزين في وضع الكائن

• يعد تخزين الكائنات حلاً مثاليًا لتخزين وأرشفة ونسخ احتياطي وإدارة كميات كبيرة من البيانات الثابتة أو غير المنظمة.

• التطبيقات السحابية الأصلية: يسير تخزين الكائنات السحابية جنبًا إلى جنب تطوير التطبيقات السحابية. • إنشاء تطبيقات سحابية أصلية جديدة أو تحويل التطبيقات القديمة إلى

التطبيقات السحابية من الجيل التالي التي تستخدم تخزين الكائنات المستند إلى السحابة كتخزين دائم للبيانات.

• جمع وتخزين كميات كبيرة من إنترنت الأشياء وبيانات الهاتف المحمول غير المنظمة لتطبيقات الأجهزة الذكية الخاصة بك.

• تحديث مكونات التطبيق بسهولة وكفاءة.

• الذكاء الاصطناعي وتحليلات البيانات الضخمة:

• إنشاء مستودع بيانات مركزي، والاستفادة من تخزين العناصر وفعاليتها من حيث التكلفة

القابلة للتطوير.

• جمع وتخزين كميات غير محدودة من البيانات من أي نوع ومن أي مصدر. • الاستعلام عن هذه البيانات لإجراء تحليلات البيانات الضخمة والحصول على

رؤى قيمة للعملاء والعمليات والسوق.

القسم 4: نماذج الاتساق

نماذج الاتساق

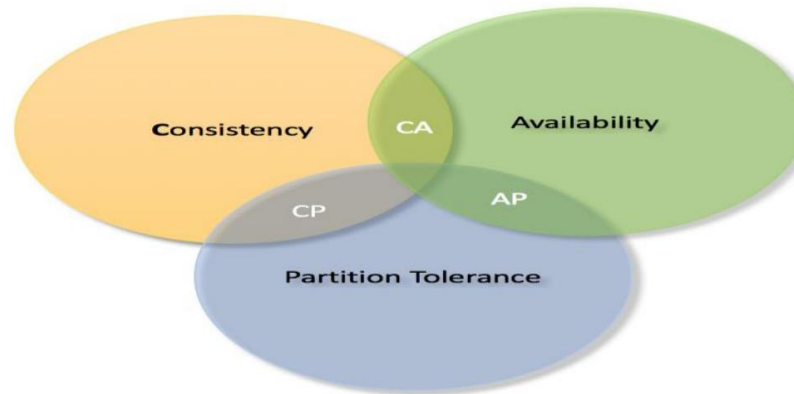
نظرية CAP

• تم تقديم نظرية CAP من قبل الدكتور بروير في عام 2000.

• تم إثبات النظرية لاحقًا بواسطة جيلبرت ولينش في عام 2002.

نظرية CAP

تنص هذه النظرية على أنه يمكن تحقيق اثنتين فقط من الخصائص الثلاث التالية على الأكثر في وقت واحد في نظام موزع: الاتساق والتوافر والتسامح مع التقسيم.



نماذج الاتساق

نظرية CAP

•الاتساق : خاصية الاتساق تضمن أن العملية أو المعاملة

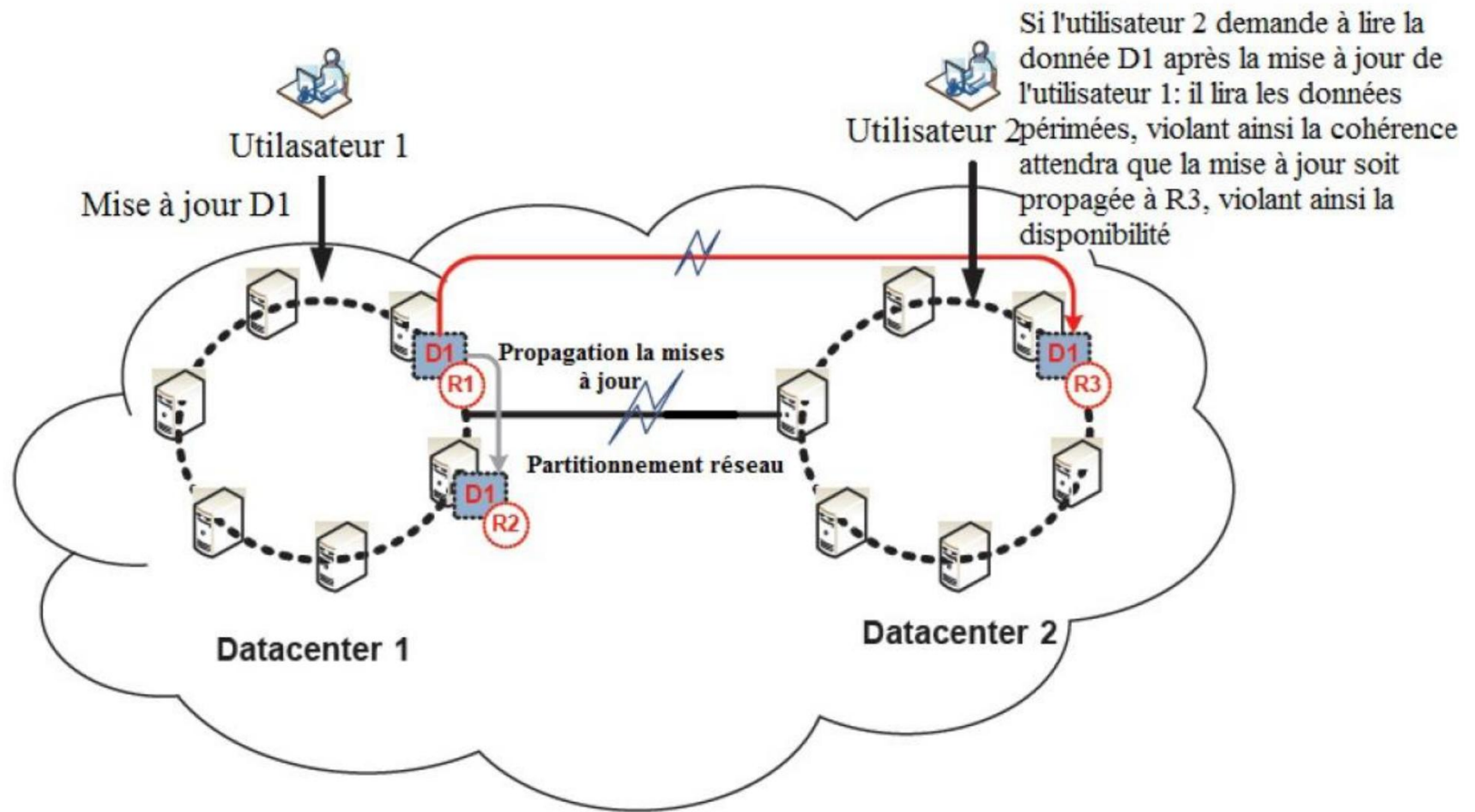
يتم تنفيذها ذريًا وتترك الأنظمة في حالة متماسكة أو تفشل بدلاً من ذلك. وهذا يعادل ضمان كل من ذرية العمليات واتساق خصائص البيانات (AC) لدلالات ACID (الذرية والاتساق والعزل والمتانة) في أنظمة إدارة قواعد البيانات العلائقية (RDBMS).

- التوفر : يتوفر نظام التخزين الموزع بشكل مستمر إذا كان كل طلب تتلقاه العقدة غير الفاشلة يجب أن يؤدي إلى استجابة. هذا هو حال اتفاقيات مستوى الخدمة (SLAs)(اتفاقيات مستوى الخدمة) التي يقدمها موفرو الحوسبة السحابية في الوقت الحاضر مما يضمن معدلات توفر أكبر من 99.99% من العمليات.

•التسامح مع التقسيم : في النظام المتسامح مع التقسيم، قد تفقد الشبكة الرسائل بين العقد من مكونات مختلفة (على سبيل المثال، مراكز البيانات). عند ظهور قسم الشبكة، يتم تعطيل اتصال الشبكة بين مكونين (الأرفف ومراكز البيانات وما إلى ذلك) ويتم فقدان كافة الرسائل. نظرًا لأنه يمكن توزيع النسخ المتماثلة عبر أقسام مختلفة في مثل هذه الحالة، فإن هذه الخاصية لها تأثير مباشر على الاتساق والتوافر.

نماذج الاتساق

نظرية CAP



نماذج الاتساق

نظرية CAP

• قدمت مضامين نظرية CAP مقايضات صعبة وأساسية للأنظمة الموزعة ومصممي الخدمات.

• الأنظمة المصممة ليتم نشرها على كيانات فردية، مثل من نظام RDBMS، يهدف إلى توفير كل من خصائص التوفر والاتساق، حيث أن التقسيم لا يمثل مشكلة.

• بالنسبة للأنظمة الموزعة التي تعتمد على الشبكات، مثل السحابة، يعد تحمل القسم أمرًا ضروريًا للغالبية العظمى منها.

• يقدم هذا من بين التنازلات الأخرى المستمدة من نظرية CAP، التماسك مقابل التماسك. التوفر كحل وسط كبير.

• وبذلك قدمت هذه النظرية العديد من التنازلات التي نسميها نماذج التماسك مثل: التماسك الضعيف، التماسك القوي، التماسك المتسلسل، التماسك السببي، إلخ.

القسم 5: قواعد البيانات الموزعة

ما هو نظام قاعدة البيانات الموزعة؟

قاعدة البيانات الموزعة:

نحن نعرّف قاعدة البيانات الموزعة على أنها مجموعة من قواعد البيانات المتعددة المرتبطة منطقيًا والموجودة في عقد النظام الموزع.

نظام قاعدة البيانات الموزعة:

يتم بعد ذلك تعريف نظام إدارة قواعد البيانات الموزعة (نظام إدارة قواعد البيانات الموزعة) على أنه نظام البرمجيات الذي يتيح إدارة قواعد البيانات الموزعة ويجعل التوزيع شفافاً للمستخدمين.

قواعد البيانات الموزعة

وعود نظام إدارة قواعد البيانات الموزعة

• الإدارة الشفافة للبيانات الموزعة والمكررة: تشير الشفافية إلى فصل دلالات المستوى الأعلى للنظام عن مشكلات التنفيذ ذات المستوى الأدنى. وبعبارة أخرى، فإن النظام الشفاف "يخفي" تفاصيل التنفيذ عن المستخدمين. تتمثل ميزة نظام إدارة قواعد البيانات الشفاف تمامًا في المستوى العالي من الدعم الذي يقدمه لتطوير التطبيقات المعقدة.

• الموثوقية من خلال المعاملات الموزعة: تهدف أنظمة إدارة قواعد البيانات الموزعة إلى تحسين الموثوقية لأنها تحتوي على مكونات مكررة وبالتالي القضاء على نقاط الفشل الفردية. إن فشل موقع واحد ، أو فشل رابط الاتصالات الذي يجعل الوصول إلى موقع واحد أو أكثر غير ممكن، لا يكفي لإسقاط النظام بأكمله. في حالة قاعدة البيانات الموزعة، هذا يعني أنه قد لا يمكن الوصول إلى بعض البيانات، ولكن مع العناية المناسبة، يمكن السماح للمستخدمين بالوصول إلى أجزاء أخرى من قاعدة البيانات الموزعة.

قواعد البيانات الموزعة

وعود نظام إدارة قواعد البيانات الموزعة

• تحسين الأداء: تعتمد الحجج المؤيدة لتحسين أداء أنظمة إدارة قواعد البيانات الموزعة بشكل عام على نقطتين. • أولاً، يقوم نظام إدارة قواعد البيانات الموزع بتجزئة قاعدة البيانات، مما يسمح بتخزين البيانات بالقرب من نقاط الاستخدام (وتسمى أيضاً منطقة البيانات). وهذا له ميزتان محتملتان:

- 1. نظراً لأن كل موقع يدير جزءاً فقط من قاعدة البيانات، فإن التعارضات الخاصة بخدمات وحدة المعالجة المركزية والإدخال/الإخراج ليست خطيرة كما هو الحال في قواعد البيانات المركزية.
- 2. تقلل المحلية من تأخيرات الوصول عن بعد التي عادة ما تكون متضمنة في شبكات المنطقة الواسعة. • النقطة الثانية هي أن التوازي المتأصل في الأنظمة الموزعة يمكن استغلاله في التوازي بين الاستعلامات وداخل الاستعلامات. يسمح توازي الاستعلام بالتنفيذ المتوازي لاستعلامات متعددة تم إنشاؤها بواسطة المعاملات المتزامنة، لزيادة إنتاجية المعاملات. يختلف تعريف التوازي داخل الاستعلام في نظام إدارة قواعد البيانات الموزع والمتوازي. • في الحالة الأولى، يتم تحقيق التوازي داخل الاستعلام عن طريق تقسيم استعلام واحد إلى عدد من الاستعلامات الفرعية، يتم تنفيذ كل منها في موقع مختلف، والوصول إلى جزء مختلف من قاعدة البيانات الموزعة.

• في نظم إدارة قواعد البيانات المتوازية، يتم تحقيق ذلك عن طريق التوازي بين المشغلين الداخليين والمشغلين الداخليين. يتم تحقيق التوازي بين المشغلين عن طريق تنفيذ عوامل استعلام مختلفة بالتوازي على معالجات مختلفة، بينما مع توازي المشغل الداخلي يتم تنفيذ نفس المشغل بواسطة العديد من المعالجات، كل منها يعمل على مجموعة فرعية من البيانات.

قواعد البيانات الموزعة

وعود نظام إدارة قواعد البيانات الموزعة

•قابلية التوسع: في البيئة الموزعة، يكون من الأسهل كثيرًا التوسع في زيادة أحجام قاعدة البيانات وأحمال العمل الأكبر. يمكن عادةً التعامل مع توسيع النظام عن طريق إضافة طاقة المعالجة والتخزين إلى الشبكة. من الواضح أنه قد لا يكون من الممكن تحقيق زيادة خطية في "الطاقة"، لأن هذا يعتمد أيضًا على حمل التوزيع. ومع ذلك، لا يزال من الممكن إدخال تحسينات كبيرة. هذا هو السبب في أن أنظمة إدارة قواعد البيانات الموزعة قد ولدت الكثير من الاهتمام بالبنى القابلة للتطوير في سياق الحوسبة العنقودية والحوسبة السحابية.

القسم 5: قواعد بيانات NoSQL

قواعد بيانات NoSQL

لماذا NoSQL

• لإدارة البيانات في السحابة، يمكننا دائمًا الاعتماد على نظام إدارة قواعد البيانات العلائقي. جميع أنظمة إدارة قواعد البيانات العلائقية لها نسخة موزعة، ويعمل معظمها في السحابة.

• على الرغم من أنهم كانوا قادرين على دمج الدعم لجميع أنواع البيانات (مثل كائنات الوسائط والمستندات) والوظائف الجديدة، فقد أدى ذلك إلى فقدان الأداء والبساطة والمرونة للتطبيقات ذات متطلبات الأداء المحددة والصارمة.

• بالنسبة لتطبيق معين، يوجد خادم قاعدة بيانات، والذي يوفر التسامح مع الخطأ وتوافر البيانات من خلال النسخ المتماثل. ومع زيادة عدد عملاء الويب، يصبح من السهل نسبيًا إضافة خوادم الويب. ومع ذلك، يصبح خادم قاعدة البيانات هو عنق الزجاجة وستتطلب إضافة خوادم قاعدة بيانات جديدة نسخ قاعدة البيانات بأكملها، الأمر الذي قد يستغرق وقتًا طويلًا.

قواعد بيانات NoSQL

ما هو نوسقل

• أحد الأنواع المهمة من تقنيات إدارة البيانات الجديدة هو NoSQL والذي يعني "ليس SQL فقط" على النقيض من منهج "حجم واحد يناسب الجميع" لنظام إدارة قواعد البيانات العلائقية.

• أنظمة NoSQL عبارة عن مخازن بيانات متخصصة تلبي متطلبات إدارة البيانات السحابية والويب.

• أدت تقنيات إدارة البيانات الجديدة هذه إلى تقديم مجموعة غنية من الخدمات التي يمكن استخدامها لإنشاء تطبيقات سحابية كثيفة البيانات يمكنها التوسع وإظهار الأداء العالي.

• ومع ذلك، فقد أدى ذلك أيضًا إلى تنوع واسع في واجهات تخزين البيانات وفقدان نموذج البرمجة المشترك. وبالتالي، فإن هذا يجعل من الصعب جدًا على المستخدم إنشاء تطبيقات تستخدم مخازن بيانات متعددة، على سبيل المثال، نظام الملفات الموزعة، ونظام إدارة قواعد البيانات الارتباطي، ونظام إدارة قواعد البيانات NoSQL.

• ظهر مصطلح NoSQL (وليس SQL فقط) لأول مرة في أواخر التسعينيات للإشارة إلى مخازن البيانات الجديدة المصممة لتلبية متطلبات إدارة بيانات الويب. بدلاً من قواعد البيانات العلائقية، فإنها تدعم نماذج بيانات ولغات مختلفة بخلاف SQL القياسية.

• تركز هذه الأنظمة بشكل عام على قابلية التوسع والتسامح مع الأخطاء

التوافر، وأحياناً على حساب الاتساق.

قواعد بيانات NoSQL

فئات مخزن بيانات NoSQL

كاريكاتير رئيسي - القيم

• في نموذج بيانات القيمة الرئيسية، يتم تمثيل كافة البيانات كأزواج قيمة المفتاح، حيث يقوم المفتاح بتعريف القيمة بشكل فريد.

• لا تحتوي مخازن القيمة الأساسية على مخططات، مما يوفر مرونة كبيرة

وقابلية التوسع.

• أنها توفر عادة واجهة بسيطة مثل وضع (مفتاح، قيمة)، قيمة = الحصول على (مفتاح)، حذف (مفتاح).

• هناك نموذج موسع لتخزين قيمة المفتاح قادر على تخزين السجلات، كقوائم لأزواج السمات والقيمة. تسمى السمة الأولى بالمفتاح الرئيسي أو المفتاح الأساسي، وتقوم بتعريف السجل بشكل فريد بين مجموعة من السجلات.

• على سبيل المثال: DynamoDB، Cassandra، Riak، Oracle NoSQL Data Base.

قواعد بيانات NoSQL

فئات مخزن بيانات NoSQL

كاريكاتير قائم على الوثائق

• مخازن المستندات هي مخازن متقدمة لقيمة المفتاح، حيث يتم تعيين المفاتيح لقيم نوع المستند، مثل JSON، أو YAML، أو XML.

• يتم عادةً تجميع المستندات في مجموعات، والتي تلعب دورًا مشابهًا للجداول العلائقية.

- تختلف المستندات عن الصفوف العلائقية. المستندات تصف نفسها بنفسها، وتخزن البيانات والبيانات الوصفية (مثل علامات XML، وأسماء الحقول في كائنات JSON) ويمكن أن تكون مختلفة عن بعضها البعض داخل المجموعة.

• بالإضافة إلى ذلك، تكون هياكل المستندات ذات تسلسل هرمي، وذلك باستخدام بنى متداخلة، على سبيل المثال، الكائنات والمصفوفات المتداخلة في JSON. إن نمذجة قاعدة البيانات باستخدام المستندات تتطلب مجموعات أقل من الجداول العلائقية (المسطحة)، كما أنها تتجنب عمليات الربط المكلفة.

• على سبيل المثال: MongoDB، AsterixDB، CouchDB، RavenDB.

قواعد بيانات NoSQL

فئات مخزن بيانات NoSQL

كاريكاتير على أساس أعمدة واسعة

- تجمع مخازن الأعمدة بين بعض الخصائص الجذابة لقواعد البيانات العلائقية (على سبيل المثال، تمثيل البيانات كجداول) مع مرونة مخازن القيمة الأساسية (على سبيل المثال، البيانات غير المخططة في الأعمدة).

- يتم تعريف كل صف في جدول أعمدة عريض بشكل فريد بواسطة مفتاح و لديه عدد من الأعمدة المسماة.

- على عكس الجدول العلائقي، حيث يمكن أن تحتوي الأعمدة على قيم ذرية فقط، يمكن أن يكون العمود كبيرًا ويحتوي على أزواج متعددة من القيمة الرئيسية.

- تعمل مخازن الأعمدة الواسعة على توسيع واجهة مخزن القيمة الرئيسية من خلال بنى تعريفية أكثر تتيح التحليل والمطابقة التامة واستعلامات النطاق في عائلات الأعمدة.

- أصل مخازن الأعمدة الواسعة هو Google Bigtable.

قواعد بيانات NoSQL

فئات مخزن بيانات NoSQL

كاريكاتير على أساس أعمدة واسعة

- تجمع مخازن الأعمدة بين بعض الخصائص الجذابة لقواعد البيانات العلائقية (على سبيل المثال، تمثيل البيانات كجداول) مع مرونة مخازن القيمة الأساسية (على سبيل المثال، البيانات غير المخططة في الأعمدة).

- يتم تعريف كل صف في جدول أعمدة عريض بشكل فريد بواسطة مفتاح و لديه عدد من الأعمدة المسماة.

- على عكس الجدول العلائقي، حيث يمكن أن تحتوي الأعمدة على قيم ذرية فقط، يمكن أن يكون العمود كبيرًا ويحتوي على أزواج متعددة من القيمة الرئيسية.

- تعمل مخازن الأعمدة الواسعة على توسيع واجهة مخزن القيمة الرئيسية من خلال بنى تعريفية أكثر تتيح التحليل والمطابقة التامة واستعلامات النطاق في عائلات الأعمدة.

- أصل مخازن الأعمدة الواسعة هو Google Bigtable.