

What is NoSQL?

Mestrado Integrado em Engenharia Informática

Unidade Curricular Completar

What is NoSQL?

António Abelha

Departamento de Informática

Escola de Engenharia



What is NoSQL?

NoSQL é um termo que se refere a um tipo específico de modelo de bases de dados ou sistema de gestão de bases de dados (DBMS).

O termo **NoSQL** é muito amplo, que não se refere a um modelo específico de bases de dados. Refere-se a toda uma panóplia de diferentes modelos que têm como principal característica não se encaixarem no modelo relacional.

Embora as bases de dados NoSQL existam desde os anos 60 do século passado, foi apenas no início deste século que a abordagem NoSQL começou a ganhar força, e toda uma nova geração de sistemas NoSQL começou a chegar ao mercado.

What is NoSQL?

Um problema chamado **Big Data**

Uma das principais razões pelas quais a abordagem NoSQL começou a ser adoptada ocorreu devido ao *big data*.

Empresas como Google e Amazon começaram a ter grandes dificuldades a lidar com volumes de dados crescentes devido à sua crescente popularidade.

Isso levou ao seguinte:

What is NoSQL?

A Google desenvolveu o *Bigtable*

Um sistema de armazenamento distribuído para gerir dados estruturados, projetada para funcionar em segurança para petabytes de dados e milhares de máquinas.

O *Bigtable* é usado em dezenas produtos e projetos, incluindo Pesquisa, Analytics, Mapas e Gmail.

O objetivo do *Bigtable* é fornecer ampla aplicabilidade, escalabilidade, alto desempenho e alta disponibilidade.

(ver documento sobre Bigtable no *elearning*).



What is NoSQL?

A Amazon desenvolveu e implementou o *Dynamo*

Um sistema de bases de dados do tipo Key-Values (valor-chave) com alta disponibilidade que alguns dos principais serviços da Amazon usam para proporcionar ao utilizadores uma experiência designada por "sempre ativa". Para poder atingir um tão alto grau de disponibilidade, o *Dynamo* sacrifica a consistência em cenários adversos e de falha.

(ver documento sobre *Dynamo* no elearning)

What is NoSQL?

Mas a Google e a Amazon não são as únicas empresas que lidam com o problema do *big data*.

Muitas outras desenvolveram soluções inovadoras por razões semelhantes.

A inovação neste campo continua ainda hoje.

Temos tipo nos últimos anos uma explosão de novas tecnologias desenvolvidas fora da estrutura relacional das DBMS

“NoSQL”.

What is NoSQL?

Definição de **NoSQL**

????????? Muitas se discute sobre este assunto ????????????????

"**Sem SQL**" (estes sistemas não utilizam **SQL** - usam uma linguagem de consulta alternativa).

"**Não apenas SQL**" (sistemas que usam **SQL** em conjunto com outras tecnologias / outras linguagens de consulta).

Muitos argumentam que a única coisa que todas as bases de dados NoSQL têm em comum é que não seguem o modelo relacional.

"**NoREL**" seria um nome mais adequado.

What is NoSQL?

Características de uma base de dados **NoSQL**

O termo NoSQL refere-se a um grupo específico de **DBMSs** que partilham certas características:

- Não Relacionais
- Código aberto
- Sem esquema
- Escaláveis horizontalmente
- Não seguem os princípios do **ACID**
- Sem uma linguagem de consulta *padrão*.

Nem todas as bases de dados **NoSQL** possuem estas características.

No entanto, a maioria dessas características é inerentemente inexistente nas bases de dados relacionais.

What is NoSQL?

Modelo Não Relacional

Modelos de bases de dados não baseados no modelo relacional proposto por E.F. Codd em 1970.

DBMSs não relacionais porque na maior parte das situações o modelo relacional não é o mais adequado para os requisitos.

What is NoSQL?

Modelo Não Relacional

A situação mais comum ocorre nos casos em que os dados sejam não estruturados ou semiestruturados.

O volume de dados (big data) pode exigir uma nova maneira de abordar o problema do armazenamento e da recuperação desses dados.

Muitas vezes as organizações podem precisar de escalar os seus sistemas para centenas ou milhares de computadores.

Podemos ter muitas outras razões.



What is NoSQL?

Código aberto

A maioria dos **DBMSs** *NoSQL* é disponibilizado em código aberto.

Embora também existam muitos DBMSs modelo relacionais em código aberto.

Quase todo o movimento NoSQL tende a orientar-se para projetos de código aberto, com muitas organizações contribuindo para os esforços de desenvolvimento de uma única solução.

Código aberto não é necessariamente um "requisito NoSQL", mas é uma "observação NoSQL".

What is NoSQL?

Sem esquema

A maioria dos sistemas de bases de dados NoSQL não possui um esquema fixo.

Nas DBMS modelo relacional temos sempre de considerar o esforço de modelação antes da introdução de quaisquer dados.

Numa base de dados **NoSQL** não necessitamos desse requisito.

Neste contexto os modelos noSQL tendem a ser mais flexíveis com os dados que pode aceitar, tornando todo o processo de desenvolvimento muito ágil.

What is NoSQL?

Escaláveis horizontalmente

A maioria das bases de dados NoSQL tem um comportamento excelente em ambientes cluster.

Aqui que os dados são particionados em vários modos para que cada todo possa executar uma tarefa específica e muitas vezes independentemente.

Cada todo pode executar sua tarefa sem a necessidade de partilhar memória ou espaço em disco com outros processos.

Processo conhecido por *shared nothing architecture* (SN).



What is NoSQL?

Escaláveis horizontalmente

As bases de dados relacionais também podem ser configurados para funcionar em ambiente *cluster*, a abordagem *RDBMS* tende a dificultar a sua configuração, quando comparada com as bases de dados *NoSQL*.

O desempenho também pode ficar comprometido quando escalamos uma base de dados relacional dessa maneira.

As bases de dados relacionais estão vocacionadas mais para a “expansão”, para adicionar mais recursos a uma única máquina de modo a obtermos uma máquina com melhor desempenho.

What is NoSQL?

Não seguem os princípios do ACID

A grande maioria das bases de dados NoSQL

Não é rigorosa no cumprimento as restrições de ACID (Atomicidade, Consistência, Isolamento, Durabilidade), em diferentes graus.

What is NoSQL?

ACID (**A**tomicidade, **C**onsistência, **I**solamento e **D**urabilidade)
(*Atomicity, Consistency, Isolation, Durability*)

é o conjunto das propriedades de uma transação numa bases de dados.

What is NoSQL?

ACID (**A**tomicidade, **C**onsistência, **I**solamento e **D**urabilidade)
(*Atomicity, Consistency, Isolation, Durability*)

Atomicidade:

Uma transação deve ter todas as suas operações executadas em caso de sucesso, em caso de falha, nenhum resultado de alguma operação refletido sobre a base de dados.

What is NoSQL?

ACID (**A**tomicidade, **C**onsistência, **I**solamento e **D**urabilidade)
(*Atomicity, Consistency, Isolation, Durability*)

Consistência:

A execução de uma transação deve levar a base de dados de um estado consistente a um outro estado consistente,

Uma transação deve respeitar as regras de integridade dos dados (como unicidade de chaves, restrições de integridade lógica, etc.).

What is NoSQL?

ACID (**A**tomicidade, **C**onsistência, **I**solamento e **D**urabilidade)
(*Atomicity, Consistency, Isolation, Durability*)

Isolamento:

O *isolamento* é um conjunto de técnicas que evitam que transações paralelas interfiram umas nas outras, como se todas essas transações fossem executadas sequencialmente (uma após a outra). Operações exteriores a uma dada transação jamais verão esta transação em estados intermediários.

Fornecer isolamento é o objetivo principal do controle da concorrência.

What is NoSQL?

ACID (**A**tomicidade, **C**onsistência, **I**solamento e **D**urabilidade)
(*Atomicity, Consistency, Isolation, Durability*)

Durabilidade:

O resultado da execução de uma transação em caso de sucesso (*commit*) devem persistir na base de dados mesmo em casos de falta de energia, bloqueamentos ou erros.
Devem garantir que os dados estarão disponíveis em definitivo.

What is NoSQL?

Não seguem os princípios do ACID

A porque a maioria das soluções NoSQL foi desenvolvida com o objetivo de fornecer alta disponibilidade e escalabilidade em ambiente *cluster*.

Ao fornecer alta disponibilidade em ambiente *cluster*, é necessário sacrificar a consistência ou a durabilidade (ou encontrar um equilíbrio entre os três).

Ao “facilitar” na consistência da informação, é possível obter uma maior disponibilidade.

What is NoSQL?

Não seguem os princípios do ACID

A abordagem NoSQL reconhece que existem alguns casos em que manter dados inconsistentes não é o “*fim do mundo*”.

Desde que tal seja assumido e essa inconsistência possa ser tratada de alguma forma, não é crítico apresentar dados que não sejam 100% consistentes.

No entanto, alguns sistemas NoSQL, como MarkLogic, Neo4j e OrientDB, suportam e respeitam as transações ACID.

What is NoSQL?

Sem uma linguagem de consulta *padrão*.

Ainda não temos uma linguagem de consulta padrão que seja suportada por todas as bases de dados NoSQL.

Embora já tenha havido tentativas de introduzir uma linguagem de consulta padrão para todas as bases de dados NoSQL, nenhuma foi adoptada.

Alguns DBMSs *NoSQL* têm sua própria linguagem de consulta, enquanto outros suportam várias linguagens, como **JSON**, **XQuery**, **SPARQL**, etc.

What is NoSQL?

As primeiras bases de dados NoSQL

Muitas bases de dados "não SQL" foram propostas ao longo dos anos. Antes mesmo do SQL ser proposto pela primeira vez em 1974. O primeiro RDBMS apenas foi lançado comercialmente em 1979.

Em seguida os três dos primeiros sistemas de gestão de bases de dados (NoSQL) apresentados:



What is NoSQL?

As primeiras bases de dados NoSQL

O ***MultiValue*** é uma base de dados NoSQL apresentada pela primeira vez em 1965 como parte do sistema operativo Pick.

O ***MUMPS*** (*Massachusetts General Hospital Utility Multi-Programming System*) desde 1966. *MUMPS* não tem esquema, e usa um motor de base de dados do tipo *key-values*, é um exemplo clássico de uma base de dados *NoSQL*.

O ***IBM IMS*** é a junção entre uma base de dados hierárquica e um sistema de gestão de informações desenvolvido em 1996.

What is NoSQL?

As primeiras bases de dados NoSQL

Na altura ninguém chamou a estas bases de dados “**NoSQL**”.
No caso das duas primeiras o **SQL** ainda não tinha sequer sido inventado.

Só depois de 2009 é que estes **DBMSs** mais antigos foram associados ao movimento **NoSQL**.

What is NoSQL?

Tipos de bases de dados *NoSQL*

A bases de dados NoSQL são actualmente classificadas em quatro tipos principais.

Algumas delas de dados são uma mistura entre diferentes tipos diferentes.
Em geral todas se encaixam nestas categorias.