

## Computação em nuvem: gerenciamento de dados

IN1128/IF694 – Bancos de Dados Distribuídos e Móveis  
 Ana Carolina Salgado – acs@cin.ufpe.br  
 Bernadette Farias Lôscio – bfl@cin.ufpe.br



Cin.ufpe.br

## Computação em Nuvem

- Computação em nuvem (*Cloud Computing*) é uma tendência recente de tecnologia cujo objetivo é proporcionar serviços de Tecnologia da Informação (TI) sob demanda com pagamento baseado no uso.
- O usuário terceiriza tarefas complexas.
  - Armazenamento de dados, administração do sistema, implantação de aplicativos.
- A complexidade de gerenciar a infraestrutura do software/hardware se desloca das organizações para o provedor de cloud.
- Amazon, Google, Microsoft e Yahoo.



## Computação em Nuvem – Motivação

- Experiência com datacenters muito grande
  - Economia de escala sem precedentes
  - Transferência de risco
- Fatores de tecnologia
  - Internet de banda larga difundida
  - Maturidade de tecnologias de virtualização
- Fatores de negócios
  - Custo inicial mínimo
  - Modelo de pagamento baseado no uso pay-as-you-go



## Computação em Nuvem - Características principais

- **Sempre disponível**
- **Serviço automático sob demanda**
- **Elasticidade**
  - Capacidade de atender a necessidades de diferentes aplicações
  - Recursos alocados ou desalocados conforme sejam requisitados
- **Preço sob demanda**
  - pay-as-you-go



## Computação em Nuvem - Características principais

- **Surgiu da evolução/cominação de diferentes modelos computacionais propostos para suportar aplicações na Web:**
  - Arquiteturas orientadas a serviços (SOA)
    - Comunicação de alto nível entre aplicativos através de serviços web.
  - Cluster
    - Gerenciamento recursos de computação
  - Virtualização
    - Abstração da máquina física
  - Computação autônoma (automática)
    - Permite a auto-gestão das infraestruturas complexas.
  - Grid computing
    - Paradigma computacional de computação distribuída.



## Computação em Nuvem - Tipos de serviços oferecidos

- **Cloud computing pode oferecer:**
  - Infrastructure-as-a-Service (IaaS)
  - Platform-as-a-Service (PaaS)
  - Software-as-a-Service (SaaS)
- **Service Level Agreement (SLA)**
  - Rege as responsabilidades, garantias e compromissos dos serviços com o cliente.



### IaaS - Infrastructure-as-a-Service

- **Consiste em prover uma infraestrutura computacional (redes e recursos de armazenamento) como serviço**
  - O provedor de Cloud cuida da criação e manutenção da infraestrutura
- **Possibilidade de adicionar ou liberar recursos conforme necessário**
  - O usuário só é cobrado pelos recursos consumidos.
- **Virtualização de servidores**
- **Ex: Amazon Web Services**



### PaaS - Plataform-as-a-Service

- **Entrega de uma plataforma de computação com ferramentas de desenvolvimento e APIs como um serviço.**
  - Sem a necessidade de adquirir e gerenciar hardware e software.
- **Permite criar e implantar aplicativos personalizados diretamente na infraestrutura de Cloud, em máquinas virtuais, e integrá-las com os aplicativos fornecidos como SaaS.**
- **A Plataforma atende à demanda de desenvolver, compilar, debugar e testar uma aplicação, abstraindo hardware e sistema operacional.**
- **Ex: Salesforce, Google App Engine, Windows Azure**



### SaaS - Software-as-a-Service

- **Entrega de aplicativos de software como serviço.**
- **O aplicativo hospedado é totalmente operado em um servidor, não sendo necessário instalar o sistema no computador do cliente.**
- **Os usuários finais podem configurá-lo para atender as suas necessidades.**
- **Ex: CRM Salesforce, Google-Docs**



### Computação em nuvem - Benefícios

- **Custo**
  - Cliente: a cobrança só é feita se os recursos forem consumidos.
  - Provedor: partilha os custos para vários clientes.
- **Facilidade de acesso e utilização:**
  - Acesso aos serviços a qualquer hora e em qualquer lugar com conexão à Internet.
  - A nuvem esconde a complexidade da infraestrutura de TI e abstrai a localização e distribuição dos dados.
- **Elasticidade:**
  - A capacidade de dimensionar os recursos de forma dinâmica para acomodar novas condições.
- **Melhor utilização de recursos computacionais**
- **Economia de tempo para criar a infraestrutura**
  - Auto-serviço (padronização + automação)
- **Diminuição do investimento em capital**



### Privacidade na nuvem

- **Esses serviços podem ser entregues de diversas formas**
  - Além dos tipos de serviços, existem diferentes tipos de nuvens quanto à privacidade.
  - **Nuvem pública**
    - A nuvem está disponível para qualquer pessoa na Internet
    - Mais utilizadas por pequenas empresas
  - **Nuvem privada**
    - Uso de tecnologias da Cloud para o gerenciamento de um centro de dados, como em uma rede privada, atrás de um firewall.
    - Pequena vantagem no custo
    - Utilização de redes internas.
    - Mais utilizadas por grandes empresas.



### Privacidade na nuvem

- **Nuvem privada virtual**
  - **Amazon e Google**
    - Têm proposto nuvens privadas virtuais com a promessa de um nível de segurança semelhante a uma nuvem privada, mas dentro de uma nuvem pública.
  - **Fornecer uma Rede Privada Virtual (VPN) com serviços de segurança para os clientes.**
  - **Usadas para desenvolver nuvens híbridas.**
    - Utiliza nuvens privadas com e nuvens públicas.



## Gerenciamento de dados na nuvem



## Gerenciamento de dados em nuvem

- SGBDs em nuvem estão começando a ser utilizados e têm o potencial de atrair clientes de diversos setores do mercado
  - pequenas empresas com o objetivo de reduzir o custo total, por meio da utilização de infraestrutura e sistemas de terceiros
  - grandes empresas que buscam soluções que para gerenciar milhares de máquinas e permitir o atendimento de um aumento inesperado de tráfego



## Gerenciamento de dados em nuvem

- A infraestrutura de SGBDs em nuvem possui várias vantagens para os usuários:
  - previsibilidade e custos mais baixos, proporcional à qualidade do serviço (QoS) e cargas de trabalho reais
  - complexidade técnica reduzida, graças a interfaces de acesso unificado e a delegação de tuning e administração de SGBDs e
  - a elasticidade e escalabilidade, proporcionando a percepção de recursos quase infinitos.
- Os SGBDs em nuvem estão sendo disponibilizados como serviços, que encapsulam a complexidade do gerenciamento por meio de formas de acesso simples e garantias de acordos de nível de serviço



## Gerenciamento de dados em nuvem

- Requisitos para SGBD como um serviço

Requisitos do Usuário	
<i>U1</i>	API simples com pouca configuração e administração (ex. sem tuning)
<i>U2</i>	Alto desempenho (ex. vazão, escalabilidade)
<i>U3</i>	Alta disponibilidade e confiança (ex. hot stand-by, backup)
<i>U4</i>	Acesso fácil à características avançadas (ex. snapshot, evolução de esquema, mineração de dados)
Requisitos do Provedor	
<i>P1</i>	Atender o SLA do usuário (ex. potencialmente sob carga de trabalho dinâmica)
<i>P2</i>	Limitar hardware e custo de energia (ex. multiplexação intensiva)
<i>P3</i>	Limitar custo de administração (ex. custo com pessoal)
Requisitos extra de Nuvem Pública	
<i>P1</i>	Esquema de preço: barato, previsível e proporcional ao uso (elasticidade)
<i>P2</i>	Garantias de segurança e privacidade
<i>P3</i>	Baixa latência (relevante para OLTP e aplicações Web)

[Sousa et al. 2010] Sousa, F. R. C., Moreira, L. O., Macêdo J. A. F., Machado, J. C., "Gerenciamento de dados em nuvem: conceitos, sistemas e desafios", minicurso - SBB0 2010.



## Gerenciamento de dados em nuvem - Características

- O gerenciamento de dados em nuvem pode ser organizado em duas classes de sistemas:
  - Classe 1: para apoiar aplicações com muitas atualizações
    - 1.1: o objetivo do sistema é apoiar uma única aplicação, com grandes quantidades de dados e
    - 1.2: o objetivo do sistema é apoiar um grande número de aplicações, cada uma com pequenas quantidades de dados
  - Classe 2: para análises dos dados e suporte a decisão



## Gerenciamento de dados na nuvem – Características

<i>Distribuição</i>	Poucos centros de dados
<i>Ambiente</i>	Recursos homogêneos em centros de dados
<i>Operações para acesso aos dados</i>	API simples, SQL ou variações
<i>Atualizações</i>	Suporte às atualizações concorrentes
<i>Transações</i>	ACID ou variações
<i>Replicação</i>	Garantia de QoS e transparência
<i>Granulosidade da Replicação</i>	Fina
<i>Controle Global</i>	Central ou Distribuído
<i>Alterações Dinâmicas</i>	Escalabilidade e suporte para cargas de trabalho inesperadas

[Sousa et al. 2010] Sousa, F. R. C., Moreira, L. O., Macêdo J. A. F., Machado, J. C., "Gerenciamento de dados em nuvem: conceitos, sistemas e desafios", minicurso - SBB0 2010.



### Gerenciamento de dados - Características

- Uma característica essencial no ambiente de nuvem é o gerenciamento autônomo
  - Hardware e software dentro de nuvens podem ser automaticamente reconfigurados, orquestrados e estas modificações são apresentadas ao usuário como uma imagem única



### Gerenciamento de dados na nuvem - Armazenamento e Processamento de Consultas

- Existem novos sistemas de arquivos, frameworks e propostas para o armazenamento e processamento de dados
  - Google File System (GFS) é um sistema de arquivos distribuídos proprietário desenvolvido pelo Google e projetado especialmente para fornecer acesso eficiente e confiável aos dados usando grandes clusters de servidores
  - Hadoop File System (HDFS) armazena grandes arquivos em várias servidores e obtém a confiabilidade por meio da replicação de dados. Similar ao GFS, os dados são armazenados em nós geograficamente distribuídos.



### Gerenciamento de dados na nuvem - Armazenamento e Processamento de Consultas

- Algumas propostas para o armazenamento e processamento utilizam a estrutura chave-valor em uma Distributed Hash Table (DHT)
- Outra abordagem para armazenar e processar dados em nuvem consiste em utilizar uma estrutura de colunas ou arrays multidimensionais
  - Os dados são organizados em tabelas e estas possuem diversas colunas.
  - Cada coluna armazena um valor, acessado por meio de uma chave
- Também existem as abordagens orientadas a documentos ou baseadas em grafos



### Gerenciamento de dados na nuvem - Transações/Consistência

- A utilização de transações distribuídas define o controle do processamento de dados em todo ambiente de computação em nuvem e tem a responsabilidade de garantir as propriedades ACID ou variações destas no ambiente
  - São usados protocolos de replicação de dados, terminação distribuída e sincronização de acesso devido à natureza compartilhada dos recursos
- Um ponto fundamental na construção de sistemas distribuídos e considerado por todos os sistemas em nuvem é o teorema Consistency, Availability, Partition Tolerance (CAP)



### Gerenciamento de dados na nuvem - Transações/Consistência

- Este teorema mostra que os sistemas distribuídos não podem assegurar as seguintes propriedades simultaneamente:
  - Consistência: todos os nós tem a mesma visão dos dados ao mesmo tempo.
  - Disponibilidade: falhas em nós não impedem os demais nós de continuar a operar.
  - Tolerância a partições: o sistema continua a operar mesmo com a perda arbitrária de mensagens.



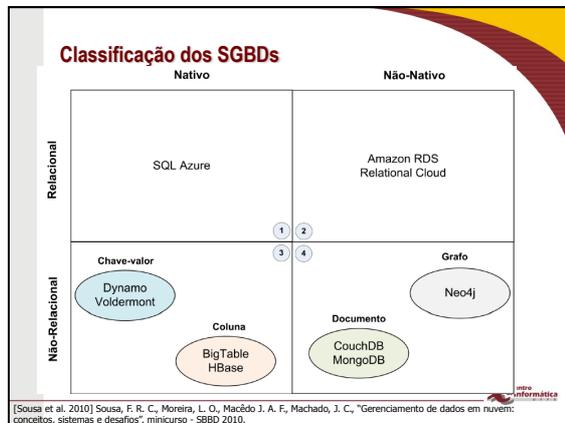
### Gerenciamento de dados na nuvem - Transações/Consistência

- Um sistema distribuído pode suportar apenas duas dessas três propriedades ao mesmo tempo.
- O teorema CAP tornou-se um modelo popular para compreender aspectos de sistemas distribuídos.
- Estas propriedades não devem ser interpretadas no sentido de que o sistema é disponível ou consistente, e sim quando ocorre uma falha de rede, é necessário escolher qual propriedade torna-se mais importante para o sistema.



### Gerenciamento de dados na nuvem - Transações/Consistência

- Algumas abordagens para o gerenciamento de dados em nuvem têm utilizado diferentes formas de consistência
- Uma alternativa é utilizar a abordagem Basically Available, Soft state, Eventually consistent (BASE)
  - Basicamente disponível: o sistema parece estar em funcionamento todo o tempo;
  - Em estado leve: o sistema não precisa estar sempre consistente;
  - Eventualmente consistente: o sistema torna-se consistente em um determinado momento.

### Classificação dos SGBDs

- No primeiro quadrante estão os SGBDs relacionais desenvolvidos nativamente para nuvem.
  - Consideram características da computação em nuvem juntamente com aspectos do modelo relacional.
  - Ex: Microsoft SQL Azure.
- No segundo quadrante estão SGBDs relacionais que não foram concebidos para nuvem, mas que podem ser executados em uma infraestrutura baseada em nuvem.
  - Ex: Amazon Relational Database Service (Amazon RDS) e o Relational Cloud



### Classificação dos SGBDs

- No terceiro quadrante estão os sistemas considerados nativos para nuvem que não utilizam o modelo relacional, tais como os sistemas que utilizam o modelo chave-valor e baseado em coluna.
  - Ex: Amazon Dynamo e o Voldemor (modelo chave-valor) . BigTable, HBase e Cassandra são exemplos de sistemas baseados em coluna
- No quarto quadrante estão os sistemas não-nativos que não utilizam o modelo relacional, tais como grafo, documento ou XML.
  - Ex: Neo4j é um exemplo dos sistemas baseados em grafo e CouchDB e MongoDB são exemplos de sistemas orientados a documento.



### Gerenciamento de dados – Quadro comparativo dos SGBDs

Característica	S3 SimpleDB	BigTable GAE	Cassandra	CouchDB	SQL Azure	Relational Cloud
Modelo	Chave-valor	Coluna	Coluna	Documento	Relacional	Relacional
Armazenamento	Hash consistente	Índices	Índices	Árvore B+	Tabela	Tabela
Linguagem de Consulta	API simples	API simples	API simples	API simples	SQL	SQL
Transações	Não	Não	Sim simplificada	Não	Sim	Sim
Consistência	Eventual	Forte	Eventual	Eventual	Forte	Forte
Escalabilidade	Alta	Alta	Alta	Média	Média	Baixa
Disponibilidade	Alta	Alta	Alta	Alta	Alta	Média

[Sousa et al. 2010] Sousa, F. R. C., Moreira, L. O., Macêdo J. A. F., Machado, J. C., "Gerenciamento de dados em nuvem: conceitos, sistemas e desafios", minicurso - SBBD 2010.

