



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

BANCO DE DADOS DE IMAGENS DE SENSORIAMENTO REMOTO

Alessandra Morais
Ivo Medeiros
Rogério Marinke
Wanessa da Silva

Trabalho de conclusão da
disciplina TÓPICOS EM
OBSERVAÇÃO DA TERRA:
CAP378

INPE
São José dos Campos
2015

Lista de Figuras

1.1	Imagem de radar de comprimento de onda de 23 cm exibindo alvos de textura lisa e alvos de textura rugosa. Fonte: Meneses (2012)	3
1.2	Exemplo da manifestação conjunta das Interações Macroscópicas e Microscópicas. Fonte: Meneses (2012)	4
1.3	Imagem de sensoriamento remoto e sua representação matricial multi-dimensional. Fonte SOUZA (2005)	6
1.4	Fluxo de dados genérico ilustrando as etapas de aquisição, correção de erros/artefatos e geração de informações úteis a partir de dados de imagem. Figura adaptada de Castelli & Bergman (2004)	7
1.5	Lago Paranoá de Brasília. Imagem adaptada de Meneses (2012)	8
1.6	Imagens com resolução radiométrica diferentes. Fonte: Meneses (2012)	9
1.7	Tipos de formato de armazenamento. Nota: figura retirada da Internet, fonte não determinada	9
2.1	Imagem e seus metadados (Fonte: Vinhas (2010))	14
2.2	Acesso a pedaços da imagem (Fonte: Vinhas (2010))	14
2.3	Suporte à reamostragem (Fonte: Vinhas (2010))	14
2.4	suporte a multi-resolução (Fonte: Vinhas (2010))	15
2.5	Arquitetura Dual (Fonte: adaptada de Queiroz & Lopes (2015))	16
2.6	Arquitetura Integrada baseada em SGBD relacional (Fonte: adaptada de Queiroz & Lopes (2015))	17
2.7	Exemplo de armazenamento utilizando BLOB (Fonte: adaptado de Vinhas (2010))	17
2.8	Exemplo de armazenamento do tipo espacial (raster) (Fonte: adaptado de Vinhas (2010))	18
2.9	Tamanho X Complexidade de banco NoSQL	21
2.10	Suporte espacial dos SGBD-R e NoSQL (Fonte: Queiroz et al. (2012))	24

Lista de Tabelas

2.1	Análise comparativa Relacional vs NoSQL	23
-----	---------------------------------------------------	----

Sumário

Introdução	1
1 Sensoriamento Remoto	2
1.1 Conceitos	3
1.2 Imagem	5
1.2.1 Aquisição, Transmissão, Recepção e Processamento	6
1.2.2 Resolução	7
1.2.3 Formatos de Armazenamento	8
1.3 Implicações: Armazenamento e Acesso	10
1.3.1 Metadados	10
1.3.2 Padrões de Acesso	10
1.3.3 Armazenamento Permanente	11
1.3.4 Volume, Variedade e Velocidade	11
2 Banco de Dados de Sensoriamento Remoto	12
2.1 Banco de Dados de Imagem	12
2.1.1 Arquitetura Dual	15
2.1.2 Arquiteturas integradas para gerência de dados	16
2.2 Tecnologias de Gerenciamento de Banco de Dados	18
2.2.1 Banco de Dados Relacional	19
2.2.2 Banco de Dados NoSQL	20
2.3 SGBD-R x NoSQL	22
Conclusions	25

Abstract

This work briefly presents concepts relating to remote sensing and the role that image database system plays into that domain. It also introduces relational database and NoSQL basis; beside the comparison between them. The objective is to show the current practice when storing images from satellite and the opportunities that could come into play deploying NoSQL technology.

Introdução

Imagens digitais têm um papel de destaque em vários setores da sociedade, elas estão presentes na agricultura e gestão florestal, no planejamento urbano, na medicina e outros. Segundo [Castelli & Bergman \(2004\)](#), o sensoriamento remoto foi uma das primeiras áreas de aplicação de imagens digitais. Definido no início da década de 60 por Evelyn L. Pruitt e colaboradores, o termo sensoriamento remoto remete a uma das mais bem sucedidas tecnologias de coleta automática de dados para o levantamento e monitoração dos recursos terrestres em escala global de acordo com [Meneses \(2012\)](#).

Dados provenientes da coleta automática apresentam diversas formas, no entanto, [Castelli & Bergman \(2004\)](#) destaca que as imagens digitais tendem a dominar os arquivos de dados, tanto em volume quanto em variedade. O custo considerável envolvido com a coleta e a questão temporal¹ fazem com que todos os dados considerados de valor sejam permanentemente preservados independente do volume e característica. O grande volume de imagens de valor produzidas, torna a aquisição, preparação, armazenamento, indexação, recuperação e distribuição dos dados uma tarefa desafiadora.

No que diz respeito ao armazenamento, a diversidade dos dados tanto em relação aos dados brutos quanto aos dados produzidos em diferentes níveis de processamento torna difícil a proposta de uma solução única que seja sempre aplicável e eficiente segundo [Castelli & Bergman \(2004\)](#). Na literatura, alguns autores classificam os dados de sensoriamento remoto como sendo parte do escopo do termo Big Data. Neste trabalho, são apresentados alguns dos desafios e soluções no armazenamento dos dados.

Este trabalho está organizado em duas partes. O Capítulo 1 resume os pontos destacados por [Castelli & Bergman \(2004\)](#) para a concepção de um banco de dados de imagens de sensoriamento remoto; neste Capítulo são apresentados os princípios físicos que compõem os fundamentos do sensoriamento remoto, o conceito de imagem, o sistema de coleta, os tipos de processamento efetuados sobre os dados e as várias maneiras em que os dados e as informações são acessados. No Capítulo 2 decorremos sobre Banco de Dados de Imagem, Banco de Dados Relacional, Banco de Dados NoSQL, e finalmente é realizada uma comparação entre os SGBG relacionais e NoSQL.

¹Dados do passado que são perdidos nunca poderão ser readquiridos

Capítulo 1

Sensoriamento Remoto

Os anos de 1960 ficaram conhecidos como sendo a década da corrida espacial. Os programas espaciais impulsionaram as pesquisas de equipamentos com a capacidade de “fotografar” à Terra sem o uso das tradicionais câmeras fotográficas. Esses equipamentos receberam o nome de sensores imageadores, sendo o produto gerado por eles denominado por imagem. Ao contrário das câmeras fotográficas que utilizam um mecanismo de tomada instantânea de área em quadro, as imagens são obtidas por meio de um processo de cobertura do terreno em forma de varredura. Segundo [Meneses \(2012\)](#), o termo Sensoriamento Remoto foi definido para designar o desenvolvimento desse novo conjunto de instrumentos cuja finalidade resume-se a obter imagens da superfície da Terra a distâncias remotas.

Sensores imageadores por satélite são capazes de obter imagens de toda a superfície do planeta num espaço de tempo relativamente curto e de modo periódico. Este fato propicia o monitoramento de mudanças que possam acontecer na superfície terrestre. As imagens coletadas são utilizadas por diversas áreas do conhecimento possuindo uma ampla gama de aplicações. São alguns exemplos de aplicações a caracterização e o estudo das variações de ozônio na atmosfera, a quantificação e a identificação das causas e efeitos da poluição, a previsão do tempo, o apoio a agricultura, o monitoramento de erupções vulcânicas, incêndios florestais e mudanças urbanas.

Os usuários (consumidores) de dados vão desde indivíduos e pequenas organizações com fins comerciais à grandes organizações e instituições ligadas a política pública e internacional. [Castelli & Bergman \(2004\)](#) destaca que é difícil prever a demanda e o padrão de uso desses dados, uma vez que, o sensoriamento remoto é uma área em que novas aplicações são frequentemente descobertas gerando novas comunidades de usuários. Além disso, o autor também alega que a preservação dos dados e informações relacionadas a eles, bem como, a disponibilidade dos mesmos de forma convenientemente acessível são tão importantes quanto a coleta dos dados em si. Na concepção de uma base de dados que visa o armazenamento e a futura disponibilidade dos mesmos [Castelli & Bergman \(2004\)](#) destaca a importância de compreender o sistema de coleta, os tipos de processamento efetuados sobre os dados e as várias maneiras em que os dados e as informações são acessados.

Este capítulo resume os pontos destacados pelo autor. A seção 1.1 aborda os princípios físicos que compõem os fundamentos do sensoriamento remoto. Na seção 1.2 é contextualizado o conceito de imagem, apresentando sua definição no contexto de sensoriamento remoto, o fluxo de atividades do processo de aquisição à disponibilização de imagens ao usuário final, as características quanto nível de detalhe que uma imagem comporta e possíveis formatos de armazenamento. Por fim, a seção 1.3 contextualiza as questões associadas à gestão das imagens digitais.

1.1 Conceitos

O Sensoriamento Remoto é uma ciência cuja base fundamenta-se nos princípios físicos da Radiação Eletromagnética (REM). A REM possui um comportamento dual, sendo ao mesmo tempo uma forma de onda e uma forma de energia. O princípio da dualidade da REM é a base da aquisição de dados (ie. imagens). Segundo [Meneses \(2012\)](#) ao se analisar uma imagem de sensoriamento remoto, a coexistência da radiação eletromagnética na forma de onda e energia deverá ser considerada para que se possa explicar tudo o que se observa nas imagens.

A dualidade do comportamento no processo de aquisição e análise de imagens é formulada pelo modelo ondulatório (onda) e pelo modelo corpuscular (energia). O modelo ondulatório é responsável pela interação macroscópica entre um objeto e a REM incidente, ou seja, é a resposta da intensidade com que um objeto reflete a radiação eletromagnética em razão do tamanho da onda e a textura da superfície do objeto. Este modelo se aplica melhor para os grandes comprimentos de onda (ex. micro-ondas). A Figura 1.1 exemplifica o conceito por meio de uma imagem de radar, os alvos escuros representam superfícies especulares (água e solo gradeado) onde a REM foi refletida em direção oposta à direção de visada do sensor, de forma que nenhuma radiação voltou ao sensor; os demais níveis de cinza representam alvos com refletância difusa que retornaram ao sensor.



Figura 1.1: Imagem de radar de comprimento de onda de 23 cm exibindo alvos de textura lisa e alvos de textura rugosa. Fonte: [Meneses \(2012\)](#)

•

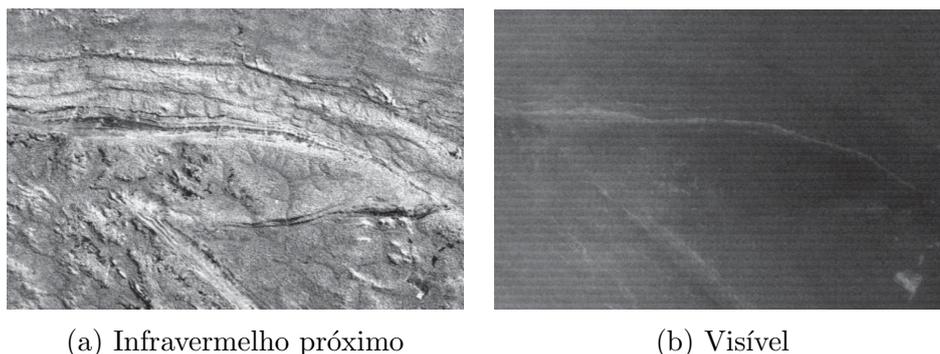


Figura 1.2: Exemplo da manifestação conjunta das Interações Macroscópicas e Microscópicas. Fonte: [Meneses \(2012\)](#)

perfície de um objeto interage com ela por meio de trocas de energia que ocorre através da absorção de parte da REM incidente. A parte da energia que não é absorvida é refletida para o sensor. São as quantidades de energia absorvidas que definem algumas informações quanto a composição do objeto observado. É válido destacar que a quantidade de energia transportada por uma onda é inversamente proporcional ao seu comprimento, de forma que comprimentos de ondas grandes não transportam energia suficiente para interagir ao nível microscópico.

O modelo corpuscular explica a interação microscópica entre um objeto e a REM incidente. Este modelo define que a REM ao atingir a superfície de um objeto interage com ela por meio de trocas de energia. A parte da energia que não é absorvida é refletida para o sensor. São as quantidades de energia absorvidas que definem algumas informações quanto a composição do objeto observado. É válido destacar que a quantidade de energia transportada por uma onda é inversamente proporcional ao seu comprimento, de forma que grandes comprimento de onda não transportam energia suficiente para interagir ao nível microscópico.

A Figura 1.2 exemplifica a manifestação conjunta dos modelos por meio de duas imagens tomadas de uma área com densa cobertura vegetal da região amazônica. As imagens foram captadas pelo mesmo sensor, sendo a Figura 1.2a da região do infravermelho próximo de $0,76\mu\text{m}$ a $0,9\mu\text{m}$ e a Figura 1.2b próximo da região do visível de $0,63\mu\text{m}$ a $0,69\mu\text{m}$. Na banda do infravermelho próximo as interações macroscópicas e as microscópicas se combinam favoravelmente para obter a imagem do relevo. É válido salientar que a radiação eletromagnética de cada comprimento de onda interage de modo distinto e com intensidades diferentes com os objetos.

A REM deve ser vista como um espectro contínuo, porém o espectro eletromagnético foi arbitrariamente dividido pelo homem em intervalos com base nos mecanismos físicos geradores e de detecção. Segundo [Meneses \(2012\)](#) conhecer os valores dos comprimentos de onda desses intervalos é essencial para os usuários do sensoriamento remoto. Devido a absorção da radiação eletromagnética pela atmosfera existem intervalos no espectro eletromagnético onde não é possível obter imagens a partir de sensores instalados em satélites. São intervalos que se pode obter imagens a partir de sensores:

- **Visível (0,45-0,76 μm):** região do espectro solar com a melhor janela atmosférica, sendo base utilizada para as atividades de sensoriamento remoto. Trata-se da região responsável pela interação com os minerais e que dá origem às suas cores e pigmentação da vegetação. Porém, esta sujeita a alto espalhamento da radiação pelos gases atmosféricos;
- **Infravermelho próximo (0,76-1,2 μm):** região do espectro com bastante transparência, mas com algumas bandas de absorções. Trata-se do intervalo onde ocorrem importantes interações da REM para a identificação da natureza de vários tipos de rochas;
- **Infravermelho de ondas curtas (1,2-3,0 μm):** região conhecida como espectral geológica devido aos vários minerais de alteração hidrotermal;
- **Infravermelho médio (3,0-5,0 μm):** região pouco utilizada com exceção de sensores meteorológicos e atmosféricos; é a região onde o Sol e a Terra não emitem quantidades suficientes de energia, somente alvos com elevadas temperaturas (vulcões e incêndios) podem ser detectados;
- **Infravermelho termal (5,00-1.0 μm):** conhecida como região termal devido a radiação emitida pelos objetos terrestres; é a faixa espectral para a detecção de quartzo de veios nas rochas;
- **Micro-ondas (3,00 μm -100cm):** região de uso de sensores ativos (radar); utilizam fontes artificiais para a geração da REM. Devido ao comprimento de onda, os radares têm a propriedade de operar em condições atmosféricas adversas (cobertura de nuvem e chuva) e podem operar tanto de dia como à noite.

1.2 Imagem

A forma como um sensor mede a radiação eletromagnética praticamente independe do comprimento de onda. Os sensores têm por finalidade captar a REM e transformar a energia transportada pela onda em um valor digital proporcional à intensidade desta energia (medição radiométrica). Segundo [Meneses \(2012\)](#), a radiometria inclui uma série de termos e leis para explicar o significado físico dos valores medidos pelo sensor. Está fora do escopo deste trabalho detalhar tais conceitos.

O produto final das medições de um sensor imageador é denominado por imagem, mais especificamente uma imagem digital. [Gonzalez \(2009\)](#) define imagem digital como sendo uma função bidimensional $f(x, y)$, onde x , y e o valor de amplitude $f(x, y)$ são quantidades discretas (ie. inteiros). No contexto de sensoriamento remoto é importante destacar a existência de imagens multi-dimensionais; essas imagens são produtos de sensores com a capacidade de coletar simultaneamente diferentes regiões do espectro eletromagnético (ie. bandas). Sendo assim, uma definição mais abrangente deve considerar que imagens digitais de sensoriamento remoto podem apresentar outras dimensões além das duas representadas no plano cartesiano das linhas (x) e colunas (y); neste sentido, imagens de sensoriamento remoto são

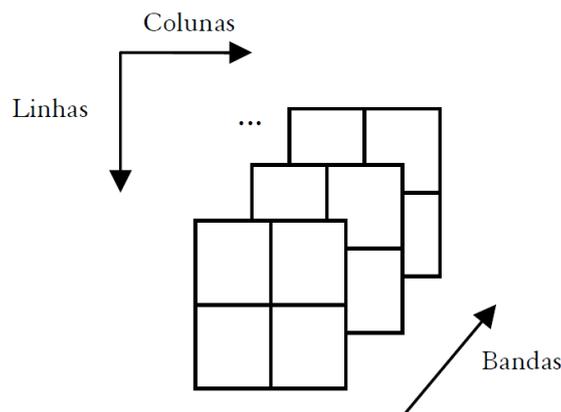


Figura 1.3: Imagem de sensoriamento remoto e sua representação matricial multi-dimensional. Fonte SOUZA (2005)

definidas como sendo uma função multi-dimensional. A Figura 1.3 exemplifica esse conceito.

As subseções seguintes descrevem as principais características das imagens no contexto do sensoriamento remoto.

1.2.1 Aquisição, Transmissão, Recepção e Processamento

A aquisição de uma imagem inicia no sensor imageador ao converter a energia transportada pela onda (sinal analógico) em um valor digital proporcional à intensidade desta energia. Os dados adquiridos a partir dos sensores nas plataformas de satélites necessitam ser eletronicamente transmitidos a Terra; segundo Levin (1999), os dados adquiridos podem ser transmitidos de três formas: transmissão direta para a estação receptora em solo se esta estiver na linha de visão do satélite, armazenamento temporário para aguardar uma oportunidade de transmissão direta ou uma série de transmissões entre satélites até a transmissão direta a Terra. Independentemente da forma transmitida o dado é recebido na estação de recepção em seu formato digital bruto.

Os dados recebidos estão sujeitos a erros e distorções introduzidas pelos sistemas de sensores devido aos erros instrumentais (ruídos espúrios), às interferências da atmosfera (erros radiométricos) e à geometria de imageamento (distorções geométricas). Esses dados (brutos) são referenciados como estando no nível zero de processamento. Meneses (2012) destaca que os erros ou distorções presentes se apresentam nas deformações de escala, incorreções nas posições espaciais dos pixels, na perda de contraste entre os alvos ou registros incorretos dos valores digitais dos pixels.

As correções nos dados são realizadas por meio de um conjunto de funções operacionais de processamento denominada por pré-processamento; essas técnicas não visam a transformação da imagem na forma de um produto final, mas destina-se a melhorar a qualidade dos dados. A aplicação de uma dada função para a correção de um determinado erro e/ou distorção depende dos objetivos do trabalho ou pesquisa

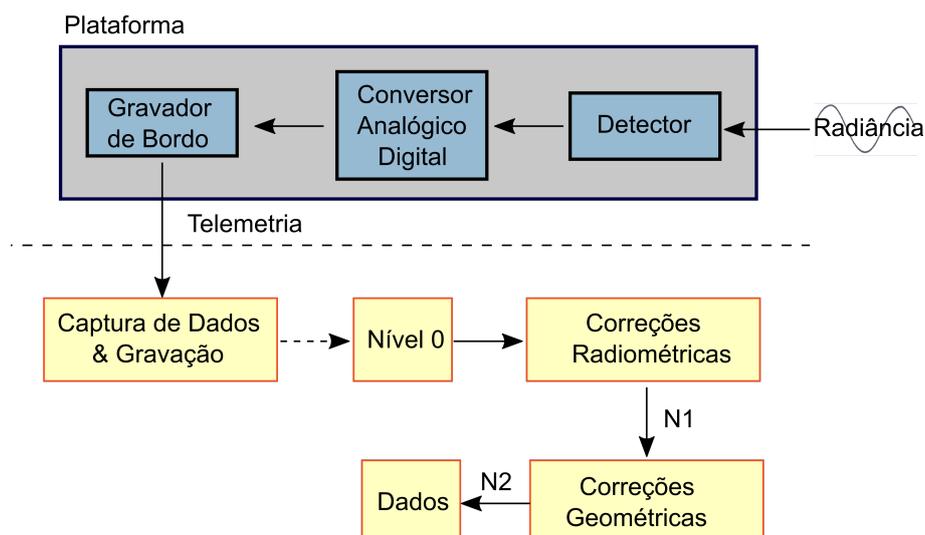


Figura 1.4: Fluxo de dados genérico ilustrando as etapas de aquisição, correção de erros/artefatos e geração de informações úteis a partir de dados de imagem. Figura adaptada de [Castelli & Bergman \(2004\)](#)

do consumidor final. A Figura 1.4 exemplifica um possível fluxo de atividades de pré-processamento, a cada manipulação empregada na imagem aumenta-se o nível de processamento.

[Meneses \(2012\)](#) destaca que com relação ao ruído se faz necessário corrigi-los antes da distribuição para eventuais usuários, para os demais casos o pré-processamento pode ser deixado a cargo do usuário.

1.2.2 Resolução

A resolução de uma imagem descreve o nível de detalhe que esta comporta. Em sensoriamento remoto a resolução é dada por quatro diferentes formas de medições que são relacionadas com as características do sensor imageador, a saber: a área do campo de visada do sensor (**espacial**), o comprimento de onda das bandas (**espectral**), os valores numéricos da medida de energia (**radiométrica**) e a data em que a imagem foi tomada (**temporal**).

A **Resolução Espacial** determina o tamanho do menor objeto que pode ser identificado numa imagem. Por definição, um objeto só será identificável se seu tamanho for no mínimo igual ou maior do que a célula de resolução. A Figura 1.5 exemplifica o conceito. Na prática é possível detectar objetos menores em algumas situações como destacado por [Levin \(1999\)](#). A célula de resolução trata-se da área no chão imageada pelo sensor, essa área é determinada através do cone angular de visibilidade do sensor (*Instantaneous Field of View* - IFOV) a partir de uma dada altitude a um determinado momento. [Levin \(1999\)](#) destaca a importância de distinguir o tamanho de um pixel da resolução espacial; se uma imagem obtida por um sensor é exibida na resolução máxima, então os tamanhos serão coincidentes, porém, é possível exibir uma imagem cuja resolução não coincida com a resolução espacial.

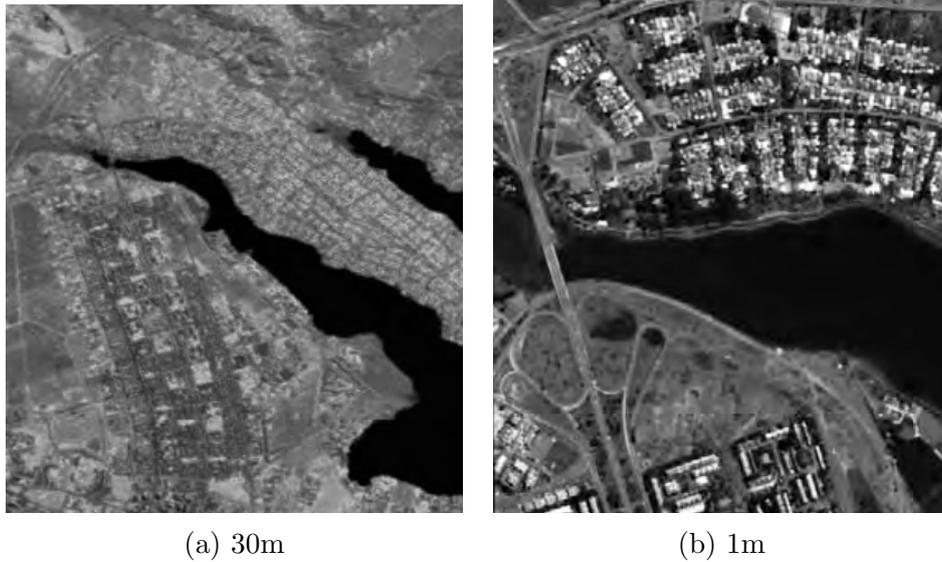


Figura 1.5: Lago Paranoá de Brasília. Imagem adaptada de [Meneses \(2012\)](#)

A sensibilidade de um sensor quanto a sua capacidade de detectar pequenas variações de energia refletida (ou emitida) por um objeto é dada por sua **Resolução Espectral**. Segundo [Meneses \(2012\)](#) o termo envolve pelo menos três parâmetros de medida, a saber: o número de bandas que o sensor possui, a largura em comprimento de onda das bandas e as posições que as bandas estão situadas no espectro eletromagnético. Como descrito na seção 1.1 cada objeto responde de forma diferente para uma dada região do espectro, sendo assim, quanto melhor a resolução espectral melhor é a capacidade do sensor de distinguir objetos. Comparativamente, um sensor apresenta uma melhor resolução espectral quando possui um maior número de bandas abrangendo diferentes regiões espectrais com larguras estreitas de comprimento de onda. Na prática, [Meneses \(2012\)](#) destaca que a resolução espectral aparenta melhorar a resolução espacial.

O número de níveis de energia que o sensor pode discriminar é definido pela **Resolução Radiométrica**; quanto maior for a capacidade de discriminação, maior será a resolução radiométrica. Essa medida é usualmente expressa em termos de números de dígitos binários (bits). A Figura 1.6 exhibe exemplos de imagens com resolução radiométricas diferentes. É válido destacar que essa medida influencia na qualidade visual da imagem.

Por fim, a **Resolução Temporal** refere-se a frequência que o sensor revisita uma área obtendo imagens periódicas ao longo de sua vida útil. Esta medida é fundamental para o acompanhamento ou detecção da evolução ou mudanças que possam ocorrer na Terra.

1.2.3 Formatos de Armazenamento

Uma imagem digital pode ser do tipo raster (também denominada por matricial ou bitmap) ou do tipo vetorial. [Levin \(1999\)](#) alega que em sensoriamento imagens

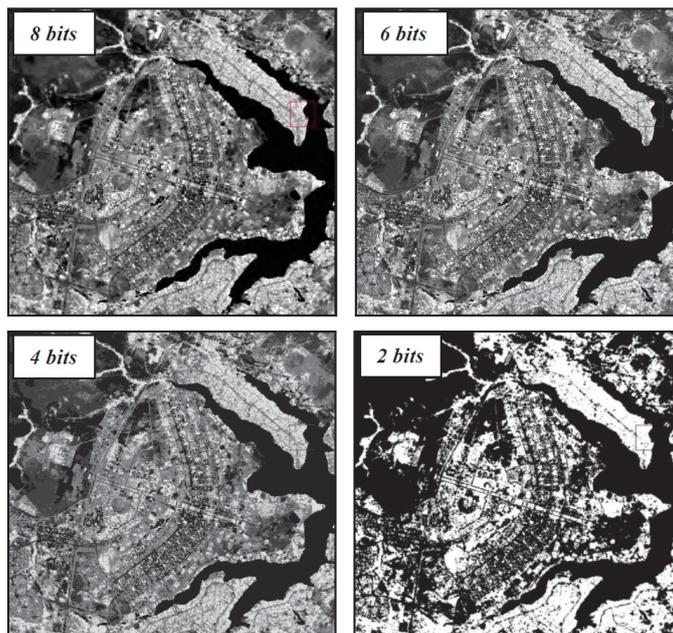


Figura 1.6: Imagens com resolução radiométrica diferentes. Fonte: [Meneses \(2012\)](#)

do tipo raster são predominantemente utilizadas; nessas imagens o primeiro pixel situa-se na posição do canto extremo superior esquerdo, as linhas são contadas de cima para baixo e as colunas da esquerda para a direita.

Conforme descrito na seção 1.1, imagens podem conter dados de diversas bandas. Essa característica implica que os dados podem ser organizados de diversas formas. [Levin \(1999\)](#) destaca que em geral, os dados são armazenados no disco em fluxo binário de bytes em três tipos básicos de formatos (formas); são eles bandas sequenciais (BSQ), bandas intervaladas por linha (BIL) ou bandas intervaladas por pixels (BIP).

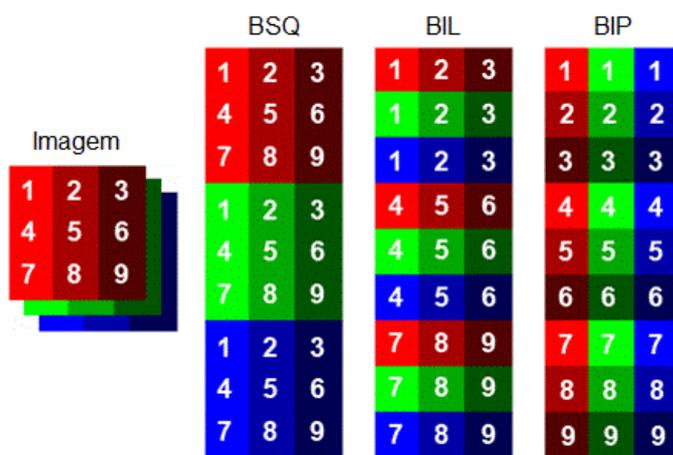


Figura 1.7: Tipos de formato de armazenamento. Nota: figura retirada da Internet, fonte não determinada

A Figura 1.7 exemplifica os três tipos de formato. No formato BSQ cada linha é seguida imediatamente pela próxima linha na mesma banda espectral, este formato

é indicado para acesso espacial. O formato BIL favorece o desempenho entre o processamento espacial e espectral, sendo o formato recomendado para a maioria das tarefas de processamento; as imagens armazenadas nesse formato tem a primeira linha do primeiro banda seguido pela primeira linha da segunda banda. Por fim, o formato BIP favorece o desempenho de processamento espectral, as imagens armazenadas nesse formato armazena o primeiro pixel de todas as bandas em ordem sequencial, seguida pelo segundo pixel para todas as bandas.

1.3 Implicações: Armazenamento e Acesso

As características das imagens de sensoriamento remoto e a dinâmica das atividades que a envolvem impõem certos requisitos sobre o armazenamento e acesso desses dados. As subseções a seguir apresentam os principais pontos destacados por [Castelli & Bergman \(2004\)](#) em relação a tais requisitos.

1.3.1 Metadados

A grosso modo uma imagem digital trata-se de um aglomerado de números. Essa característica impõe que informações adicionais sejam registradas e mantidas para o futuro gerenciamento das imagens que venham a compor um sistema de dados de sensoriamento remoto. Tais informações são ditas metadados, ou seja, são dados sobre outros dados.

Os metadados podem ser armazenados junto das imagens ou em base de dados separadas que contenham ponteiros que remetam a imagem. Essas informações consistem de vários atributos com diferentes granularidades; alguns desses atributos são destinados a descrever um grande conjunto de imagens, enquanto que outros destinam-se a descrever uma única imagem ou um subconjunto de imagens. [Levin \(1999\)](#) destaca que arquivos de imagens também podem conter metadados na forma de cabeçalho. São exemplos de atributos a origem dos dados (plataforma, sensor utilizado e outros), cobertura espacial (latitude e longitude dos cantos de um retângulo delimitador), cobertura temporal, contato (nome, e-mail e outras informações do provedor), tipo processamento utilizado, entre outros.

Existe na literatura padrões de metadados, esses padrões têm por objetivo prover uma terminologia comum e um conjunto de definições para documentar dados. Organizações como *Federal Geographic Data Committee* (FGDC) oferecem extensões de seus padrões para a documentação de dados de sensoriamento remoto (FGDC-STD-012-2002 [Authority \(2002\)](#)).

1.3.2 Padrões de Acesso

Imagens e metadados são acessados de diversas maneiras. Usuários podem ter interesse em acessar somente metadados com a finalidade de obter informações sobre

as características de uma coleção de imagens, por exemplo: a forma como as imagens são obtidas, os algoritmos utilizados para processamento entre outras questões. Além disso, usuários podem fazer uso dos metadados para localizar imagens que cubram uma dada área espacial.

Antes de acessar ou solicitar as imagens, usuários podem ter interesse em visualizar imagens com resoluções reduzidas a fim de verificar a qualidade dos dados, cobertura de nuvens e outras características. Usuários também podem ter interesse em poucos arquivos ou requerer coleções. Além disso, a aquisição dos dados pode ocorrer tanto na forma de download quanto por meio de mídias.

1.3.3 Armazenamento Permanente

Além do custo envolvido na aquisição de imagens por satélite, o registro contínuo e consistente dessas imagens é importante para muitas aplicações de sensoriamento remoto, como por exemplo, no estudo climático de longo prazo. Segundo [Castelli & Bergman \(2004\)](#), essa característica estabelece vários requisitos de armazenamento.

O autor destaca que os dados brutos e todos os dados auxiliares como metadados, documentação e softwares necessários para gerar níveis mais elevados de produtos devem ser armazenados permanentemente. A "sobrevivência" a longo prazo dos dados deve ser garantida e estes devem ser migrados com uma dada frequência para mídias mais atualizadas de forma a garantir que os dados continuem a ser legíveis.

1.3.4 Volume, Variedade e Velocidade

Sensoriamento remoto tem por característica a produção de grande volume de dados, segundo [Ma et al. \(2014\)](#) um único centro de dados de satélite tem potencial de gerar vários terabytes por dia. De acordo com o autor, a proliferação de dados vem lançando desafios significativos na gestão, processamento e interpretação dos mesmos. Na literatura é possível encontrar autores como [Guo et al. \(2014\)](#) que contextualizam as imagens de satélite no escopo do termo Big Data devido ao seu volume, variedade e velocidade (os "3 Vs" utilizados por alguns autores para definir o termo Big Data).

O aumento dos dados de sensoriamento remoto em volume e velocidade também traz um rápido crescimento nos metadados. [Ma et al. \(2014\)](#) cita que o programa de sistemas de dados de ciências da terra da NASA provavelmente ultrapassa 129 milhões de registros (metadados) e aumenta a uma taxa de 60 mil a cada dia. Devido a esse cenário existem trabalhos que consideram utilizar bancos de dados NoSQL ao invés dos tradicionais relacionais, são exemplos os trabalhos propostos por [Liu et al. \(2013\)](#), [Xiao & Liu \(2011\)](#) e [Chaokui & Wu \(2014\)](#).

Capítulo 2

Banco de Dados de Sensoriamento Remoto

Banco de dados de imagens de sensoriamento remoto são bancos que guardam imagens digitais adquiridas por sensores localizados em satélites, essas imagens correspondem a “fotografias” (matrizes numéricas multi-dimensionais) tomadas a grandes alturas a partir destes sensores. Tais sensores medem tanto a energia refletida da superfície terrestre e transmitida para a atmosfera, quanto a energia que é dispersa pela atmosfera.

Após a coleta da energia pelo sensor os dados coletados são transmitidos para estações terrestres, onde é feito o pre-processamento nas medições. Em seguida os dados são processados de acordo com o objetivo final de sua aplicação. O resultado de todo esse processo (e todos os passos intermediários que resultam em dados de valor) é então armazenado em uma ou mais bases de dados para posterior acesso, o qual irá prover as imagens e informações (metadados) a ela relacionada.

Queiroz (2012) destaca a predominância de dados científicos na forma de matrizes multi-dimensionais e a lacuna entre as tecnologias de banco de dados relacionais e a comunidade científica. Nesse capítulo, definimos o conceito de Banco de Dados de Imagem e apresentamos três diferentes arquiteturas na Seção 2.1. Na Seção 2.2 decorremos sobre as tecnologias de gerenciamento de Banco de Dados.

2.1 Banco de Dados de Imagem

Os bancos de dados de imagens são coleções de imagens, tipicamente associadas com as atividades de sensoriamento remoto, cujo papel principal é organização das imagens e seus metadados de maneira eficiente, o foco está no conteúdo das imagens. Essas imagens são informações massivas, que exigem do banco algumas características como:

- Armazenamento eficiente;

- Modelagem de dados que trate dos aspectos necessários para representar imagens no banco de dados;
- Suportar a manipulação das imagens no nível do banco de dados: recorte, reamostragem, etc; e
- Permitir o gerenciamento dos metadados da imagem;
- Capacidade de manipulação;
- Segurança;
- Backup e recuperação;
- Extensibilidade;e
- Flexibilidade.

Uma área cuja a aplicação também se ocupa com o conceito e manipulação de imagens de sensoriamento remoto são os chamados Banco de Dados Geográficos (BDG). Banco de Dados Geográficos (BDG) são aqueles que possuem dados com uma dimensão espacial, ou uma localização, diretamente ligada ao mundo geográfico real como as imagens de satélite de sensoriamento remoto [Vinhas \(2006\)](#). BDG são coleções de dados georreferenciados, manipulados por Sistemas de Informações geográficas (SIG). Os SIG são sistemas computacionais capazes de capturar, modelar, armazenar, recuperar, analisar e representar dados geográficos [Worboys & Duckham \(2004\)](#).

Ao longo dos anos, os SIGs foram implementados seguindo diferentes arquiteturas. Tais arquiteturas se diferenciam principalmente, na maneira e nos recursos utilizados para armazenar e recuperar dados espaciais. Essas arquiteturas evoluíram de forma a deixar, cada vez mais, a responsabilidade de gerenciamento dos dados para os Sistemas Gerenciadores de Banco de Dados (SGBDs) [Ferreira \(2003\)](#). O SGBD é um software ou coleção de programas que ajudarão no gerenciamento do Banco de Dados. Servindo para facilitar o processo de definição, construção e manipulação do Banco.

Atualmente a principal diferença entre os SIGs é a forma como os dados geográficos são gerenciados. De acordo com [Queiroz & Câmara \(2005\)](#) há basicamente três diferentes arquiteturas de SIG que utilizaram recursos de um SGBD: Dual, Integrada baseada em SGBDs relacionais e arquitetura Integrada com extensões espaciais.

Banco de dados geográficos armazenam tanto as imagens quanto informações a respeito destas (metadados). A Figura 2.1 apresenta um exemplo de uma imagem e seus metadados.

Os BDG além de poderem ser utilizados para visualização, distribuição ou ambos; oferecendo capacidade de consulta, navegação geográfica e acesso via Web Services. Também estão entre suas possíveis características:

- acesso a pedaços da imagem (Figura 2.2);
- suporte a reamostragem (Figura 2.3);

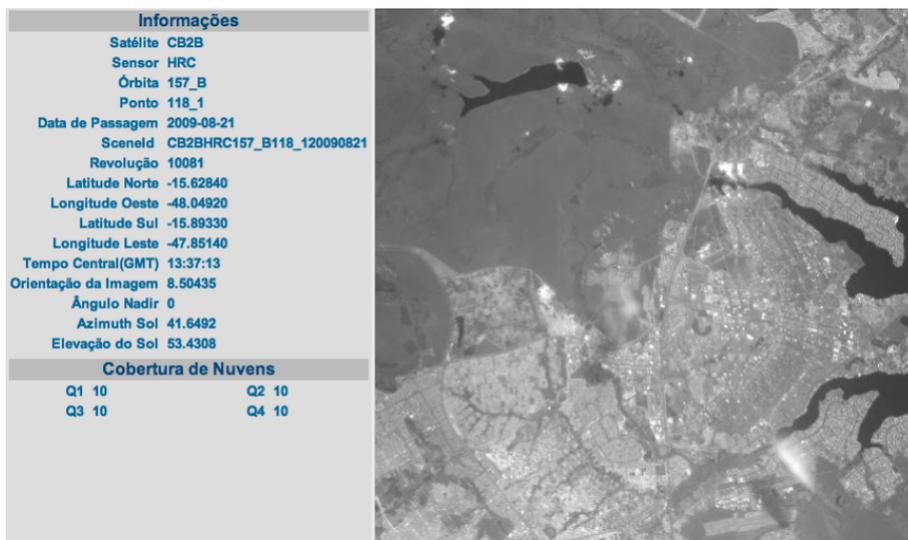


Figura 2.1: Imagem e seus metadados (Fonte: Vinhas (2010))

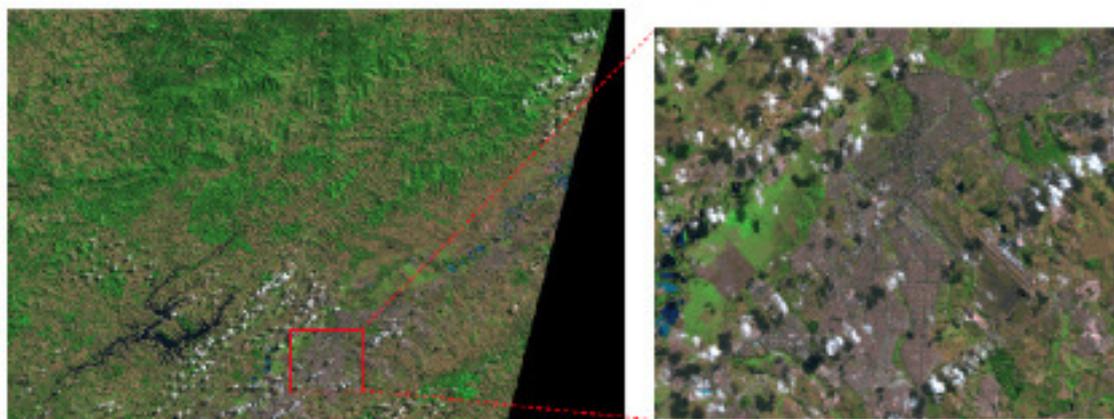


Figura 2.2: Acesso a pedaços da imagem (Fonte: Vinhas (2010))

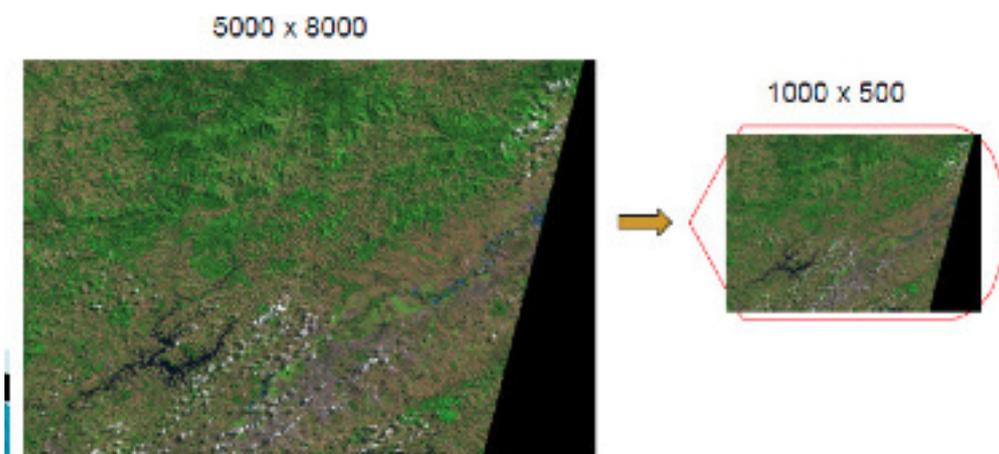


Figura 2.3: Suporte à reamostragem (Fonte: Vinhas (2010))

- suporte a multi-resolução (Figura 2.4);

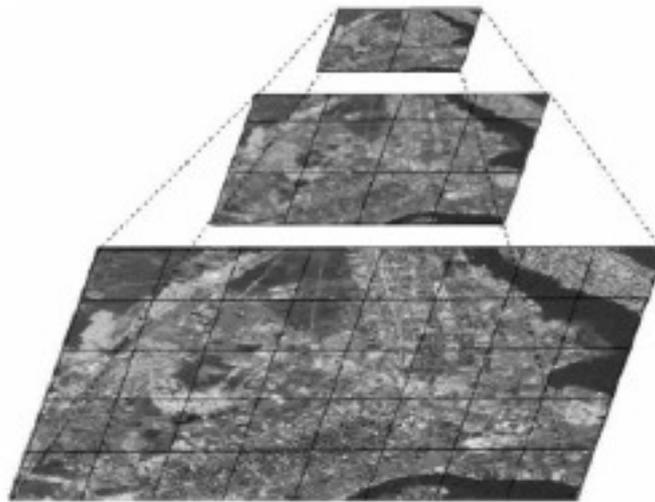


Figura 2.4: suporte a multi-resolução (Fonte: Vinhas (2010))

- indexação espacial;
- gerência dos Metadados.

2.1.1 Arquitetura Dual

Um SIG implementado com a estratégia dual utiliza um SGBD relacional para armazenar os atributos convencionais dos objetos geográficos (na forma de tabelas) e arquivos para guardar as representações geométricas destes objetos. No modelo relacional, os dados são organizados na forma de uma tabela onde as linhas correspondem aos dados e as colunas correspondem aos atributos Câmara et al. (1996).

A entrada dos atributos não-espaciais é feita por meio de um SGBD relacional e para cada entidade gráfica inserida no sistema é imposto um identificador único/ponteiro, por meio do qual é feita uma ligação lógica com seus respectivos atributos não-espaciais armazenados em tabelas de dados SGBD. A Figura 2.5 ilustra esse processo.

A principal vantagem desta estratégia é poder utilizar os SGBDs relacionais de mercado. No entanto, como as representações geométricas dos objectos espaciais estão fora do controle do SGBD, esta estrutura dificulta o equacionamento das questões de otimização de consultas, gerência de transações e controle de integridade e de concorrência. Segundo Schmitz (2010) estes problemas só podem ser resolvidos através de implementações sofisticadas das camadas superiores da arquitetura genérica, que operem coordenadamente o SGBD convencional.

Segundo Câmara & Queiroz (2001) as principais desvantagens desta arquitetura são:

- Dificuldade no controle e manipulação de dados espaciais;

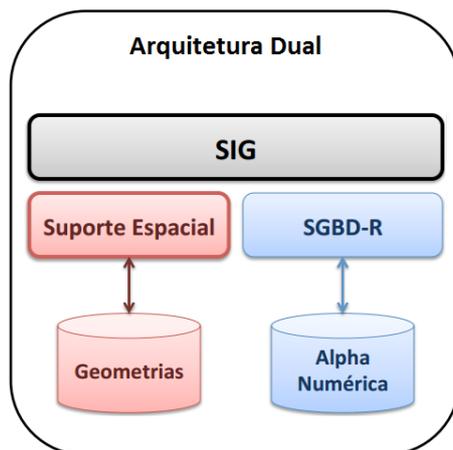


Figura 2.5: Arquitetura Dual (Fonte: adaptada de Queiroz & Lopes (2015))

- Dificuldade em manter a integridade entre a componente espacial e a alfanumérica;
- Consultas mais lentas devido ao processamento separado. A parte convencional da consulta é processada pelo SGBD separada da parte espacial, que é processada pelo aplicativo utilizando arquivos proprietários;
- Falta de interoperabilidade entre dados. Cada sistema produz seu próprio arquivo proprietário sem seguir o formato padrão, o que dificulta a integração dos dados.

Exemplos de sistemas comerciais baseados em estratégias Dual são o ARCVIEW-ESRI, o MGE-Integrgraph e o SPRING Câmara et al. (1996).

2.1.2 Arquiteturas integradas para gerência de dados

A arquitetura integrada permite que tanto o dado espacial quanto sua parte alfanumérica sejam armazenados em um SGBD. De acordo com Câmara & Queiroz (2001) sua principal vantagem é a utilização dos recursos de um SGBD para controle de integridade e concorrência, permitindo assim a manutenção da integridade entre a componente espacial e a alfanumérica. Há duas alternativas para a arquitetura integrada: a arquitetura integrada baseada em SGBD relacional e a arquitetura integrada com extensões espaciais.

Os softwares que utilizam a arquitetura integrada de SGBDs são o Oracle Spatial, PostGIS, IBM DB2 Spatial Extender, Informix Spatial Database.

Integrada baseada em SGBDs relacionais

A arquitetura integrada baseada em SGBD relacional utiliza-se de campos longos, chamados de BLOBs (Binary Large Object), para armazenar a componente espacial do dado. A Figura 2.6 apresenta essa arquitetura.

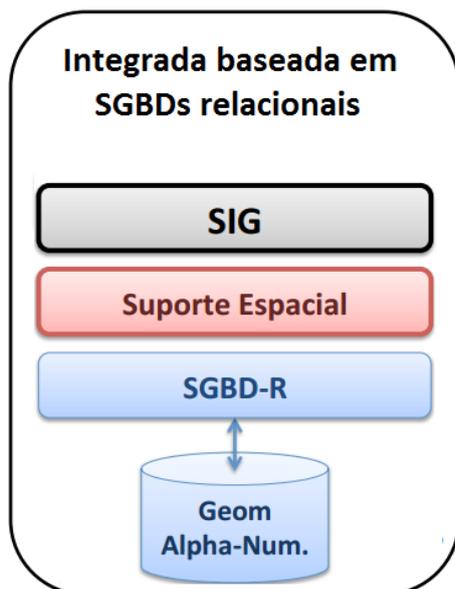


Figura 2.6: Arquitetura Integrada baseada em SGBD relacional (Fonte: adaptada de Queiroz & Lopes (2015))

A figura 2.7 exemplifica como é feito o armazenamento da componente espacial dos dados, utilizando BLOB (geometrias vetoriais).

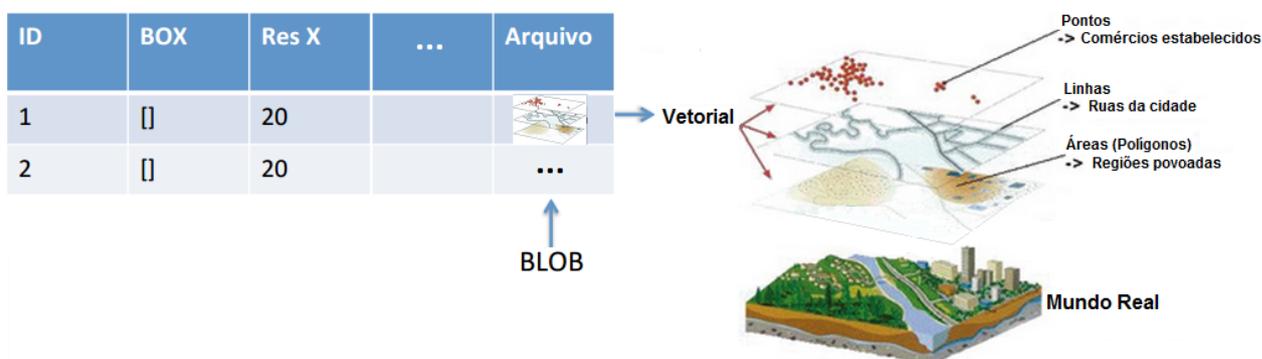


Figura 2.7: Exemplo de armazenamento utilizando BLOB (Fonte: adaptado de Vinhas (2010))

Suas principais desvantagens, segundo Queiroz & Câmara (2005), são:

- Não é capaz de capturar os dados espaciais: como o SGBD trata o campo longo como uma cadeia binária, não é possível conhecer a semântica do seu conteúdo;
- Métodos de acesso espacial e otimizador de consultas devem ser implementados pelo SIG: como o SGBD trata os dados espaciais como uma cadeia binária, não possui mecanismo satisfatórios para o seu tratamento;
- Limitações de linguagem SQL (Structured Query Language) para a manipulação dos dados espaciais: a SQL padrão oferece recursos ilimitados para o tratamento de campos longos.

Integrada com extensões espaciais

Essa arquitetura também conhecida como "Integrada baseada em extensões espaciais sobre os SGBDs objeto-relacionais". Os SGBDs Objeto-relacionais estendem os modelos relacionais, entre outras características, com um sistema de tipos de dados rico e extensível, oferecendo operadores que podem ser utilizados na linguagem de consulta. Possibilitam ainda a extensão dos mecanismos de indexação sobre os novos tipos. Essas características reduzem os problemas ocorridos na simulação de tipos de dados pelos SGBDs Relacionais, tornando os SGBDs Objeto-Relacionais uma solução atrativa para aplicações não-convencionais [Ferreira \(2003\)](#).

A figura 2.8 exemplifica como é feito o armazenamento da componente espacial do tipo raster (geometrias matriciais).



Figura 2.8: Exemplo de armazenamento do tipo espacial (raster) (Fonte: adaptado de [Vinhas \(2010\)](#))

Segundo [Ferreira \(2003\)](#) algumas das vantagens dessa arquitetura são:

- Permite definir tipos de dados espaciais, equipados com operadores específicos (operadores topológicos e métricos);
- Permite definir métodos de acesso específicos para dados espaciais.

2.2 Tecnologias de Gerenciamento de Banco de Dados

Segundo [Sadala & Fowler \(2012\)](#), os banco de dados relacionais têm sido a escolha padrão para o armazenamento de grandes volumes de dados, especialmente no mundo corporativo. O autor destaca que houve situações no passado em que outras tecnologias ameaçaram tomar uma fatia do mercado, mas nunca chegaram a tomar o campo de atuação dos banco de dados relacionais de fato.

Atualmente, existe o interesse pelos chamados NoSQL. [Queiroz \(2012\)](#) destaca que essa nova frente de tecnologias de armazenamento, gerenciamento e processamento de dados têm sido criadas por dois grupos distintos. O primeiro deles é composto pelos gigantes da Internet (ie. Google, Amazon e Facebook) e o segundo por comunidades científicas como a Astronomia e Sensoriamento Remoto. O autor também

alega que de acordo com [Gray et al. \(2005\)](#) sempre houve uma lacuna entre as tecnologias de banco de dados relacionais e a comunidade científica, devido as limitações na gerência e manuseio de dados multidimensionais. Na seção 2.1 já decorremos sobre alguns aspectos dos Banco de Dados Relacionais, nessa seção completaremos alguns conceitos (2.2.1) e apresentamos os Bancos de Dados NoSQL (2.2.2).

2.2.1 Banco de Dados Relacional

Introduzido por Codd em 1970 segundo um modelo formal baseado na teoria matemática das relações, o modelo relacional tornou-se um padrão para muitas aplicações. Esses bancos são caracterizados por uma composição de tabelas, as quais são um conjunto não ordenado de linhas. Cada linha contém os campos (atributos), estes identificados por um nome (nome do atributo). As relações são estabelecidas através de campos especiais denominados por chaves.

[Sadalage & Fowler \(2012\)](#) enfatiza que muitas coisas mudaram nas linguagens, arquiteturas e processos no campo da computação, mas poucas mudanças foram vistas quanto a forma de armazenamento e gerenciamento de dados desde de a proposta do modelo relacional. Além disso, [Queiroz \(2012\)](#) destaca que os Sistemas de Gerenciamento de Banco de Dados Relacionais (SGBD-R) têm sido empregados para o armazenamento e gerenciamento de dados nas mais diversas aplicações, incluindo as aplicações científicas. Segundo o autor, na última década esses sistemas passaram a lidar de maneira mais eficiente com dados científicos, graças a inclusão de tipos de dados, operadores e mecanismos de indexação.

No que diz respeito a dados geográficos, tanto o setor comercial quanto a comunidade de software livre propuseram extensões. Essas extensões implicam na disponibilidade de tipos, índices e funções espaciais. É valido destacar que o processo de manipulação dos dados é realizado por meio da linguagem Structured Query Language (SQL). Além disso, há também a existência de funções espaciais muito específicas disponíveis no SQL, como cálculo de histograma, área espacial funções para consulta aos valores dos pixels das imagens armazenadas etc, o que facilita a manipulação dos dados. São exemplos de extensões o PostGIS da PostGreSQL, o Oracle Spatial e o IBM DB2 Spatial Extender.

Segundo [Queiroz \(2012\)](#), uma das mais recentes funcionalidades integradas aos SGBD-R é o gerenciamento de dados na forma matricial (raster). A extensão PostGIS Raster, por exemplo, permite armazenar imagens georreferenciadas em um banco de dados. Esta extensão possibilita estratégias de armazenamento como a divisão da imagem em pequenos blocos (tiles) que são armazenadas em tabelas e o armazenamento de referências (links) para imagens gravadas em arquivos externos ao banco. Além disso, o autor destaca que é possível criar overviews (pirâmides) que aceleram a visualização por parte das aplicações.

Resumindo, o PostGIS Raster oferece um conjunto de funcionalidades que fornecem uma flexibilidade no modelo do banco, um repertório de funções e a simplicidade da API SQL [Obe \(2011\)](#). Essas funcionalidades contribuem para a evolução da

ferramenta e o uso por profissionais da área. Alguns aspectos necessitam de atenção, como a eficiência de processamento, versão ainda não estável e falta de front-ends amigáveis para a construção e visualização da base de dados.

Por fim, é válido destacar que ainda não há aplicações que usem extensivamente o PostGIS Raster, mas por se tratar de uma ferramenta livre há uma tendência que ela venha a se expandir e suas funcionalidades sejam aprimoradas. Muitas aplicações têm explorado o uso de banco de dados de imagens associados a aplicações Web/Desktop para consulta e manipulação da base de dados [Reis \(2013\)](#).

2.2.2 Banco de Dados NoSQL

A terminologia NoSQL foi usada pela primeira vez por [Strozzi \(2010\)](#) ao referir-se a uma solução de banco de dados que não oferecia suporte à linguagem de consulta SQL. Porém hoje esse termo é utilizado para designar os SGBDs não relacionais, que surgiram como uma alternativa aos sistemas de armazenamento tradicionais. Das suas principais características podem ser citadas o alto desempenho, a escalabilidade, a replicação e o suporte a dados estruturados [Agrawal et al. \(2008\)](#).

Existem diferentes SGBDs não relacionais, apesar de possuírem certas características em comum, tais como serem livres de esquema, promoverem alta disponibilidade e maior escalabilidade, esses sistemas de bancos de dados NoSQL existentes possuem diversas singularidades. E eles são classificados de acordo com a forma como trabalham com os dados, sendo esta dividida em quatro categorias básicas: os sistemas baseados em armazenamento chave-valor, como é o caso do Amazon Dynamo; os sistemas orientados a documentos, entre os quais temos o CouchDB e o MongoDB; os sistemas orientados a coluna, que tem como exemplos o Cassandra e o BigTable; os sistemas baseados em grafos, como são os casos do Neo4j e do InfoGrid e os sistemas baseados em matrizes multidimensionais, como é o caso do SciDB e Rasdaman.

Na primeira categoria, existe uma coleção de chaves únicas e de valores, os quais são associados com as chaves.

No segundo grupo, não há qualquer tipo de estruturação pré-definida, como é o caso das tabelas do Modelo Relacional, sendo os documentos as unidades básicas de armazenamento. Um documento do CouchDB, por exemplo, é um objeto que possui um identificador único e que consiste de certos campos. Esse documento segue o formato JSON (JavaScript Object Notation), padrão que visa uma fácil legibilidade e independência de linguagem [ECMA \(1999\)](#).

Na terceira categoria, muda-se o paradigma de orientação a registros (ou tuplas) para orientação a atributos (ou colunas). O Cassandra possui colunas (tuplas que contém nome, valor e timestamp), famílias de colunas (um repositório para colunas, análogo a uma tabela do Modelo Relacional) e super-colunas (compostas por arrays de colunas) [Lakshman A \(2009\)](#).

Na quarta, os dados são armazenados em nós de um grafo cujas arestas representam o tipo de associação entre esses nós.

Na quinta e última categoria, um modelo de dados e uma linguagem de consulta baseados em matrizes multidimensionais é oferecido, dando origem ao que se convencionou chamar *arrays database*.

Quanto ao modo de armazenamento dos dados, têm-se os bancos que mantêm suas informações em memória realizando persistências ocasionais; aqueles que mantêm suas informações em disco, como são os casos do CouchDb e do MongoDB; e aqueles configuráveis, tais como BigTable e o Cassandra.

Nos bancos NoSQL, não há relacionamentos entre as tabelas como nos sistemas relacionais, e os dados não precisam ter um esquema fixo. Devido a estes e outros fatores anteriormente citados, os principais benefícios dessa abordagem são:

- maior disponibilidade;
- menor tempo de resposta para consultas;
- paralelismo de atualização de dados; e
- maior grau de concorrência.

Outra maneira ainda de avaliar os bancos de dados NoSQL é observando o *tradeoff* entre a capacidade de armazenamento e a complexidade da maneira como o banco de dados representa os dados internamente; o que é apresentado de maneira oportuna na Figura 2.9

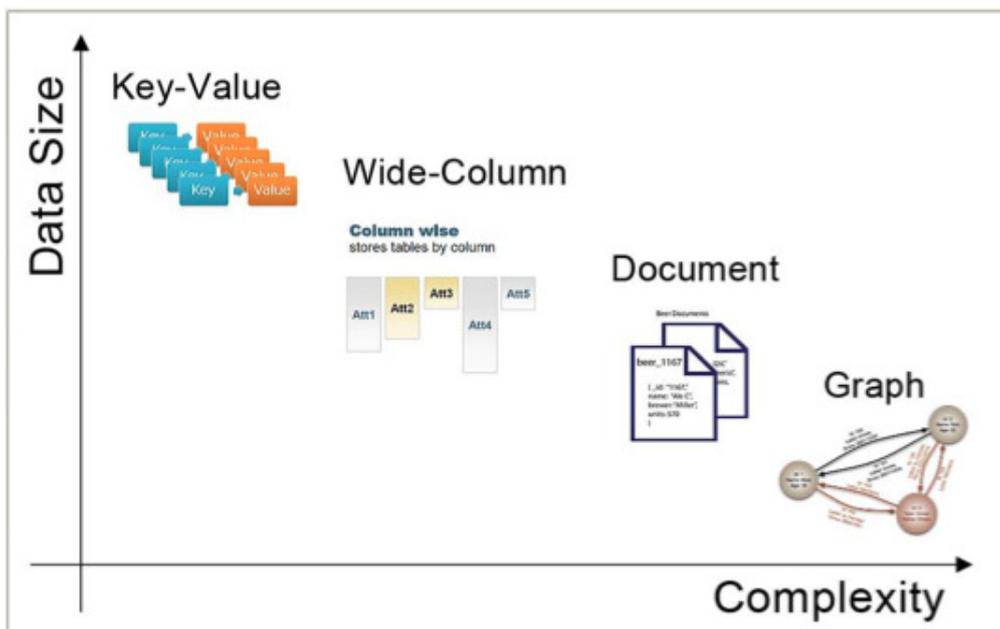


Figura 2.9: Tamanho X Complexidade de banco NoSQL

NoSQL e a Comunidade Científica

Dentre as tecnologias citadas a cima, [Queiroz \(2012\)](#) destaca que alguns trabalhos científicos estão começando a explorar o MongoDB para o armazenamento e proces-

samento de dados matriciais; este fato pode ser observado no trabalho proposto por Chaokui & Wu (2014). Na literatura, também é possível encontrar abordagens que exploram o HBase são exemplos desses trabalhos os artigos publicados por Xiao & Liu (2011) e Liu et al. (2013).

Queiroz (2012) também destaca que alguns grupos científicos optaram por construir seus próprios sistemas de armazenamento e processamento de dados. Segundo o autor, essa necessidade deu origem a um consorcio formado por diversos pesquisadores da área de banco de dados e representantes das comunidades científicas da Astronomia e Sensoriamento Remoto; desse consorcio surgiu o chamado SciDB.

Os bancos NoSQL SciDB e Rasdaman são os que tem maior notoriedade quando o assunto são dados científicos, em especial, quando fala-se de banco de imagens e da manipulação desses. A estrutura de matrizes multidimensional que o banco disponibiliza e as operações relacionadas a essa representação; já nasceram com o propósito e adequação necessária para prover o que um banco de dados de imagens demanda.

O SciDB, por exemplo, usa uma estratégia de armazenamento onde a matriz é quebrada em blocos que podem ser distribuídos em diversas máquinas para acelerar a realização de análises em matrizes muito grandes. Além disso, os blocos podem replicar partes das bordas para favorecer a computação de vizinhança. Estratégias de compactação com algoritmos que fazem um balanceamento entre capacidade de compressão e uso de CPU podem ser aplicadas a esses blocos. Os valores de cada célula são armazenados em arquivos separados, pois é comum, nas aplicações científicas, ter-se matrizes com células com grande número de valores e modelos, e processos que usam apenas alguns deles segundo Queiroz (2012).

2.3 SGBD-R x NoSQL

O Modelo Relacional tem sido amplamente utilizado em banco de dados geoespaciais durante as últimas décadas. Porém, com o crescimento cada vez mais intenso do volume de dados nesta área, certos fatores limitantes têm propiciado que modelos alternativos de banco de dados sejam utilizados em tais cenários. Motivados principalmente pela questão da escalabilidade do sistema, uma nova geração de bancos de dados – conhecidos como NoSQL (Not only SQL)– vem ganhando força e espaço.

Quando se analisa a possibilidade de se optar por uma estratégia NoSQL em detrimento de um SGBD tradicional, é preciso se levar em consideração algumas questões como os critérios de escalonamento, consistência de dados e disponibilidade.

A questão do escalonamento é essencial, e como os bancos NoSQL foram criados para este fim é neste ponto eles oferecem uma vantagem em relação aos SGBDs relacionais, que apresentam uma estruturação menos flexível e menos adaptada para cenários em que o escalonamento faz-se necessário. Bancos de dados relacionais escalam, mas quanto maior o tamanho, mais custoso se torna essa escalabilidade, seja pelo custo de novas máquinas, seja pelo aumento de especialistas nos bancos de

dados utilizados. Já os não relacionais (NoSQL), permitem uma escalabilidade mais barata e menos trabalhosa, pois não exigem máquinas extremamente poderosas e sua facilidade de manutenção permite que um número menor de profissionais seja necessário.

Já ao se substituir um SGBD relacional por uma solução NoSQL, a arquitetura perde em consistência, essa é uma desvantagem do emprego da abordagem NoSQL, mas que pode ser compensada com o ganho que se tem em flexibilidade, disponibilidade e performance.

Em relação ao grau de disponibilidade do sistema ele é otimizado em relação à abordagem relacional, pelo fato de que a queda de uma máquina não causa a interrupção de todo o sistema.

Um sumário da análise comparativa entre bancos relacionais e NoSQL é apresentado na Tabela 2.1.

Tabela 2.1: Análise comparativa Relacional vs NoSQL

	<i>RELACIONAL</i>	<i>NoSQL</i>
Armazenamento de Dados	Armazenado em um modelo relacional, com linhas e colunas. As linhas contêm todas as informações sobre uma entrada específica, e as colunas representam os atributos.	Engloba uma série de bases de dados, cada um com diferentes modelos de armazenamento de dados. Os principais são: document, graph, key-value e columnar.
Esquemas e flexibilidade	Cada registro está de acordo com um esquema fixo. Isto pode ser alterado, mas envolve a alteração de todo o banco de dados.	Os esquemas são dinâmicos. As informações podem ser adicionadas em tempo real, e cada linha (ou equivalente) não tem de conter dados para cada coluna.
Escalabilidade	A escalabilidade é vertical. Em essência, mais dados significam a necessidade de um servidor maior, o que pode ficar muito caro.	A escalabilidade é horizontal, ou seja, entre os servidores, tornando-o muito mais rentável do que a escala vertical.
ACID	A grande maioria dos bancos de dados relacionais são compatíveis com ACID (Atomicidade, Consistência, Isolamento, Durabilidade).	Variável entre as tecnologias, mas diversas soluções do tipo NoSQL sacrificam o ACID em razão do desempenho e da escalabilidade.

A Figura 2.10 apresenta um quadro comparativo entre os recursos existentes nos SGBD-R e as várias classes de sistemas NoSQL. Nesta tabela, como referência para as funcionalidades espaciais atualmente encontradas em sistema relacionais, baseadas em software livre, são usados o PostgreSQL com sua extensão geográfica PostGIS.

	PostgreSQL/ PostGIS	MongoDB	CouchDB/ GeoCouch	Neo4J/ Spatial	Apache Cassandra	SciDB	LevelDB, Berkeley DB
Classe	Relacional	NoSQL Orientado a Documentos	NoSQL Orientado a Documentos	NoSQL Baseado em Grafos	NoSQL Orientado a Colunas	NoSQL Orientado a Matrizes	NoSQL Baseado em Chave-Valor
linguagem consulta	SQL	x	x	SQL, Gremlin	x	AQL, AFL	x
transações ACID	x	x	x	✓	x	x	✓
tipos geométricos	SFS	Pontos 2D	SFS	SFS	x	x	x
operadores espaciais	✓	x	x	x	x	x	x
indexação espacial	árvore-r	hash	árvore-r	árvore-r	x	x	x
projeções cartográficas	✓	x	x	x	x	x	x
armazenamento matricial	✓	x	x	x	x	✓	x
operadores para dados matriciais	✓	x	x	x	x	✓	x
representação topológica	✓	x	x	x	x	x	x
representação de redes	x	x	x	✓	x	x	x
integração com bibliotecas de acesso a dados geoespaciais	GDAL/OGR, GeoTools, TerraLib	x	GDAL/OGR, GeoTools	GeoTools	x	x	x
aplicativos geográficos	QGIS, uDIG, gvSIG, Kosmo, TerraView	QGIS (Experimental)	x	uDIG	x	x	x
serviços geoespaciais	MapServer, GeoServer, TerraOGC	x	x	x	x	x	x

Figura 2.10: Suporte espacial dos SGBD-R e NoSQL (Fonte: Queiroz et al. (2012))

Conclusão

Alguns dos grandes fornecedores de banco de dados relacionais como Oracle e PostgreSQL criaram extensões dos seus banco de dados para disponibilizar recursos relacionados a dados espaciais. Ferramentas como: PostGis/Raster, Oracle Spatial Geometry/GeoRaster possuem como características a disponibilidade de funções específicas para manipulação de dados espaciais, tipos de dados adequados e índices específicos para aprimorar a manipulação dos dados. Essas ferramentas têm em muito contribuído para a pesquisa e desenvolvimento da área, uma vez que muitas aplicações utilizam estas ferramentas para disponibilizar informações aos usuários de imagens espaciais.

Os banco de dados relacionais ainda são amplamente utilizados como os sistemas de armazenamento de imagens. Porém, ao longo desse trabalho, percebeu-se que há um bom número de iniciativas para aplicação de bancos NoSQL para compor o sistema de armazenamento de imagens; sejam elas de satélite, seja coletada por aviões, balões e/ou afins. Por isso, ainda há uma oportunidade muito grande de pesquisa e desenvolvimento na aplicação de banco de dados NoSQL; porém é importante que se resguarde a natureza de cada solução relacional e NoSQL; dado que em prol da performance, escalabilidade e as demais característica aclamadas pela comunidade NoSQL, abre-se mão por vezes de conceitos como atomicidade, consistência, isolamento e durabilidade; que garantem a integridade dos dados. Outro fator que ainda dificulta colocar os bancos NoSQL em operação é a falta de compatibilidade com sistemas legados que foram instituídos usando bancos de dados relacionais e exigiram um esforço de migração grande; devido a isso, hoje essa iniciativas têm sido localizadas em sistema pequeno e novos em caráter experimental.

Além dos fatores citados acima também é válido destacar a questão da quebra de paradigma no projeto e desenvolvimento de soluções NoSQL. A falta de familiaridade das equipes envolvidas e o conseqüente acréscimo de tempo no aprendizado de soluções NoSQL podem ser a justificativa para a escolha de soluções relacionais.

Referências Bibliográficas

- Agrawal, R., Ailamaki, A., Bernstein, A., Brewer, A., Carey, J., Chaudhuri, S., Doan, A., Florescu, D., Franklin, J. & Garcia, H. (2008), ‘The claremont report on database research’, *ACM SIGMOD Record* .
- Authority, T. V. (2002), Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata, PhD thesis, National Aeronautics and Space Administration.
- Castelli, V. & Bergman, L. D. (2004), *Image databases Search and retrieval of digital imagery*, John Wiley & Sons.
- Chaokui, L. & Wu, Y. (2014), The distributed storage strategy research of remote sensing image based on mongo db, *in* ‘Earth Observation and Remote Sensing Applications (EORSA), 2014 3rd International Workshop on’, IEEE, pp. 101–104.
- Câmara, G., Medeiros, C., Casanova, M., Hemerly, A. & Magalhães, G. (1996), *Anatomia de Sistemas de Informação Geográfica*, Escola de Computação SBC.
- Câmara, G. & Queiroz, G. (2001), ‘Arquitetura de Sistemas de Informação Geográfica. Introdução a ciência da geoinformação’.
URL: <http://www.dpi.inpe.br/gilberto/livro/>
- ECMA (1999), ‘Introducing JSON’.
URL: <http://json.org/>
- Ferreira, K. (2003), Interface para Operações Espaciais em Banco de Dados Geográficos, PhD thesis, Instituto Nacional de Pesquisas Espaciais.
- Gonzalez, R. C. (2009), *Digital image processing*, Pearson Education India.
- Gray, J., Liu, D. T., Nieto-Santisteban, M., Szalay, A., DeWitt, D. J. & Heber, G. (2005), ‘Scientific data management in the coming decade’, *ACM SIGMOD Record* **34**(4), 34–41.
- Guo, H., Wang, L., Chen, F. & Liang, D. (2014), ‘Scientific big data and digital earth’, *Chinese Science Bulletin* **59**(35), 5066–5073.
- Lakshman A, M. P. (2009), ‘Cassandra - a decentralized structured storage system’, *LADIS* .

- Levin, N. (1999), ‘Fundamentals of remote sensing’, *1st Hydrographic Data Management Course, IMO—International Maritime Academy, Trieste* p. 76.
- Liu, Y., Chen, B., He, W. & Fang, Y. (2013), Massive image data management using hbase and mapreduce, *in* ‘Geoinformatics (GEOINFORMATICS), 2013 21st International Conference on’, IEEE, pp. 1–5.
- Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A. & Jie, W. (2014), ‘Remote sensing big data computing: Challenges and opportunities’, *Future Generation Computer Systems* .
- Meneses, P. (2012), ‘Princípios do sensoriamento remoto’, *Introdução ao Processamento de Imagens de Sensoriamento Remoto. Brasília: UNB, CNPq* .
- Obe, R.; Hsu, L. (2011), ‘Postgis in action’, pp. 2–31.
- Queiroz, G. & Câmara, G. (2005), ‘Banco de dados geográficos’, *Curitiba: Mundo-Geo* .
- Queiroz, G. & Lopes, E. (2015), ‘Sistemas de Banco de Dados Geoespaciais Parte I: Tipos e Operadores Espaciais’.
URL: <http://wiki.dpi.inpe.br/lib/exe/fetch.php?media=sbde-vector-2015-07-16.pdf>
- Queiroz, G., Monteiro, A. & Câmara, G. (2012), ‘Geographic databases and nosql: Accomplishments and future directions’, *Revista Brasileira de Cartografia 2013* .
- Queiroz, G. R. (2012), CellDB: UMA ARQUITETURA PARA SUPORTE A PROBLEMAS DE MODELAGEM AMBIENTAL EM GRANDE ESCALA, PhD thesis, Instituto Nacional de Pesquisas Espaciais, Sao Jose dos Campos.
- Reis, V. L. F. K. (2013), ‘Explorando a extensão wkt raster do postgis para armazenamento e manipulação de imagens de sensoriamento remoto’.
- Sadalage, P. J. & Fowler, M. (2012), *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*, Pearson Education.
- Schmitz, L. (2010), *Geotecnologias Aplicadas ao SIG*, Apostila do curso especialização em geoprocessamento - Departamento de arquitetura Cieg.
- SOUZA, L. (2005), ‘Tratamento de dados matriciais na terralib’, *CASANOVA, MAC, G.; DAVIS JR., CA; VINHAS, L* pp. 441–476.
- Vinhas, L. (2006), Um Subsistema Extensível para o Armazenamento de GeoCampos em Bancos de Dados Geograficos, PhD thesis, Instituto Nacional de Pesquisas Espaciais.
- Vinhas, N. (2010), ‘Banco de Dados de Imagens. Palestra para a disciplina de Bancos de Dados Geograficos do Curso de Sensoriamento Remoto’.
URL: <http://www.dpi.inpe.br/cursos/ser303/Bancos%20de%20Dados%20de%20Imagens.pdf>
- Worboys, M. & Duckham, M. (2004), *GIS A Computing Pespective Second Edition*, CRC Press.

Xiao, Z. & Liu, Y. (2011), Remote sensing image database based on nosql database, *in* 'Geoinformatics, 2011 19th International Conference on', IEEE, pp. 1–5.