

FUNDAÇÃO GETÚLIO VARGAS
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO

Programa Institucional de Bolsas de Iniciação Científica (PIBIC)

Modelo de Dados Flexível para Análise Fundamentalista Moderna

Um estudo geral com dados no Brasil

Guilherme de Araujo Gonçalves

Rodrigo Marotti Togneri

São Paulo – SP

2021

Modelo de Dados Flexível para Análise Fundamentalista Moderna

Um estudo geral com dados no Brasil

Resumo

A análise fundamentalista clássica se apoia em métricas derivadas de contabilidade empresarial e matemática financeira. Via de regra, para que ativos sejam comparados quanto a estas métricas, os dados financeiros das empresas representadas são comumente padronizados e estruturados em tabelas.

Porém, sabe-se que empresas de setores econômicos distintos possuem marcantes diferenças de estrutura contábil, e que ainda entre empresas de um mesmo setor há particularidades importantes à análise. Além disso, atualmente, outros dados igualmente importantes estão disponíveis: rede de participações e conglomerados, influências macro e microeconômicas nacionais e estrangeiras, e dinâmica de mercado.

Propõe-se assim um modelo de dados que atenda às necessidades de análise fundamentalista moderna e técnicas atuais de ciência de dados, utilizando um único modelo de dados flexível para encaixar as três principais demonstrações financeiras: balanço patrimonial (BP), demonstração de resultados do exercício (DRE) e fluxo de caixa (FC). O modelo deve permitir o armazenamento em um único banco de dados, com uma rápida forma de consulta e sem perda ou agrupamento de informações.

O projeto final foi estruturado com uma estrutura de dados não relacional em MongoDB e possui as demonstrações financeiras de todas as empresas listadas em bolsa no Brasil da forma mais primitiva apresentada à Comissão de Valores Mobiliários (CVM). Isso permite ao analista observar e entender melhor o comportamento contábil e de negócios das empresas analisadas de uma forma rápida e objetiva, sem que haja perda de informações como contas não recorrentes ou ajustes contábeis, casos que podem ser despercebidos quando se utiliza portais de informações ou agrupamentos tabulares, sendo que uma estrutura flexível melhora a acuracidade dos modelos de análise fundamentalista.

Palavras-chaves

Banco de dados, Análise Fundamentalista, Valuation, NoSQL, MongoDB

1. Introdução

Percebe-se atualmente uma transição de estruturas nos sistemas para usuários e internet. Muitas aplicações e empresas estão se utilizando dos conceitos de *serverless*, ou seja, de aplicações sem servidores ou infraestrutura próprios, sendo que há uma gama de empresas oferecendo essas funções como serviços.

Com os dados sendo armazenados em nuvem por questões de segurança, backup e maiores facilidades no acesso, o custo computacional encontra-se em grande relevância. Apesar das aplicações *serverless* reduzirem *CAPEX* com servidores e estruturas iniciais, o *OPEX* agora torna-se mais relevante, uma vez que os modelos de serviço em nuvem são *pay-per-use*. Dessa forma, reduzir o custo computacional de processamento e de armazenamento torna-se uma condição primária para aumento do lucro operacional de empresas que dependem de grandes quantidades de dados.

A literatura de estruturas de banco de dados demonstra muito sobre formas de organizações de tabelas, que são objetos primários para estruturação de bancos de dados relacionais. Contudo, com o aumento da quantidade de dados e do surgimento de áreas mais especializadas como Big Data e Ciência de Dados outro debate importante surge não só sobre as formas de organização dos dados tabulares, mas também se a estrutura tabular é a melhor forma de armazenamento de dados. A contraparte do modelo relacional é o modelo não relacional, conhecidos popularmente como banco de dados SQL e banco de dados NoSQL, respectivamente. O primeiro preza por estrutura enquanto o segundo preza por flexibilidade.

Alguns autores realizam a comparação direta entre as estruturações para consultas específicas, mas as pesquisas não demonstram um claro discernimento de qual estrutura é melhor, sempre deixando claro que esse processo depende muito da aplicação final e de como os dados serão utilizados. O debate geralmente é focado apenas nas questões computacionais, que são de grande importância para a velocidade da aplicação, além de serem mitigadores primários de custos. Porém, há também espaço para uma análise mais qualitativa das estruturas de dados, sendo que uma estrutura permite o armazenamento de mais de um tipo de dados, enquanto a outra permite apenas o tipo de dados já predefinido.

Como pontuado pela literatura existente, a questão de qual alternativa é melhor depende da aplicação final e por isso, pretende-se comparar as duas estruturas de dados para um mesmo objetivo final. Assim, o intuito desse trabalho é demonstrar que um modelo de

dados flexível, com uma estrutura não relacional, permite realizar a análise e precificação de ativos com maior profundidade e com menor custo computacional frente ao modelo mais comum utilizado, sendo este o modelo relacional. Isso porque o modelo relacional pauta-se na universalização da estrutura, fazendo com que todas empresas e parâmetros a serem analisados encaixem em apenas uma estrutura fixa de linhas e colunas que serão utilizadas para os cálculos necessários. Assim, muitas vezes é necessário agrupar contas em uma mesma linha, perdendo particularidades do negócio e da estrutura contábil em questão.

Já no âmbito computacional, acredita-se que a estrutura de dados no modelo tabular acaba por se tornar mais demandante que a estrutura chave e valor de modelos não relacionais conforme a quantidade de dados aumenta, uma vez que no modelo tabular há sempre a verificação de chaves primárias, chaves secundárias, estrutura de relação entre linhas e colunas e dos tipos de dados inseridos. Isso dá maior segurança para o modelo tabular, mas deve causar maior custo computacional e reduz a flexibilidade do modelo de dados.

Este estudo então foi realizado de forma lógica a entender quais dados são necessários e como é feita a precificação de ativos utilizando a análise fundamentalista na seção de Teoria. Isso ajuda a compreender qual a utilização final do modelo de dados e quais dados devem ser obtidos para realizar a correta avaliação do valor justo de empresas. Além disso, a seção de Teoria também cobre aspectos importantes sobre as diferenças entre modelos de dados tabulares, em bancos de dados SQL, e modelos de dados voltados a objetos, ou NoSQL.

Já a seção de Métodos demonstra como foram obtidos os dados e como eles foram estruturados em um banco de dados não relacional, enquanto a seção de resultados demonstra a estrutura final do banco de dados juntamente com uma análise das suas vantagens e desvantagens em relação ao modelo tabular na utilização para a tarefa final pretendida que é a precificação de ativos.

A seção de Resultados demonstra a estrutura final obtida, bem como os principais achados do estudo, trazendo uma comparação entre os modelos de dados para a precificação de ativos. Por fim, a Conclusão demonstra possíveis utilizações e trabalhos futuros sobre o estudo.

2. Teoria

A busca por modelos que possam prever o movimento dos preços nos mercados globais remete há datas tão longínquas quanto as bolsas de valores. Com a evolução dos conhecimentos de matemática financeira, cálculo e estatística, os modelos tornaram-se cada vez mais sofisticados na tentativa de prever movimentos do mercado. O ponto principal dessa questão está na seleção de ativos, pois sabendo o movimento é possível saber qual a melhor hora de comprar ou de vender um determinado ativo, maximizando assim o retorno sobre as operações de compra e venda.

Contudo, mesmo com a evolução dos modelos, a previsão de preços continua uma tarefa desafiadora e complexa. A dinâmica dos mercados é ruidosa, caótica, não linear e não estacionária (FAMA, 1995). Essas características fazem com que procedimentos de base estatística tenham pouca acuracidade para prever eventos futuros em preços de ações. Além disso, há ainda um grande debate sobre a hipótese dos mercados eficientes (FAMA, 1995). Caso ela esteja correta, não há qualquer sentido em tentar prever valores pois os preços futuros seriam totalmente independentes dos passados, seguindo uma caminhada aleatória. Nesse caso, as anomalias nos preços seriam corrigidas quase que imediatamente devido à unanimidade das informações no mercado. Porém, até o momento, a comunidade científica não chegou a um consenso da eficiência ou da ineficiência dos mercados (ABU-MOSTAFA, ATIYA, 1996).

Ao longo do tempo foram consolidadas duas abordagens principais de analisar ações e realizar a seleção de ativos para a composição de um portfólio. A primeira é a análise baseada nos preços passados dos ativos e na premissa de que os comportamentos passados são válidos para prever movimentos futuros, consistindo na Análise Técnica. Já a outra abordagem consiste na Análise Fundamentalista que foca em calcular o valor intrínseco de cada ativo baseado em fatores fundamentais da contabilidade empresarial presente, bem como expectativas futuras de retorno da empresa analisada e de fatores macroeconômicos (HU, LIU, *et al.*, 2015, SHAH, ISAH, 2019).

Dentro da Análise Fundamentalista há ainda várias formas de realizar o *valuation* de empresas. Uma técnica comum é a comparação de múltiplos contábeis de mercado. Assumindo que duas empresas possuam a mesma taxa de crescimento, a que possuir um menor valor de preço dividido pelo lucro por ação é considerada melhor por estar mais barata

para um mesmo resultado. Assim, é comum realizar comparações entre companhias de um mesmo setor, pois teórica e historicamente a taxa de crescimento de empresas de um mesmo setor tende a ser próxima.(IMAM, BARKER, *et al.*, 2008)

Segundo (SHAH, ISAH, 2019) outra forma de realizar a precificação de ativos por meio de Análise Fundamentalista é o modelo de crescimento de Gordon (GORDON, 2013). Nesse modelo há a premissa de que os dividendos de uma companhia irão crescer a uma taxa constante eternamente, mas sempre menor que a taxa de desconto. Uma evolução desse modelo é o mais utilizado atualmente: o método do fluxo de caixa descontado. Olhando mais para a seleção de ativos e não tão especificamente para precificação, (DUTTA, BANDOPADHYAY, *et al.*, 2012) demonstrou a utilidade da análise fundamentalista por meio de múltiplos financeiros para classificar em um algoritmo de aprendizado de máquina as melhores e as piores ações para compra e expectativas de longo prazo, obtendo uma precisão de 74,6% (SHAH, ISAH, 2019).

Com o aumento significativo do número de empresas listadas nos mercados acionários houve também um aumento do volume de informações para análises de empresas. Além disso, a transformação digital promovida nos últimos anos gera cada vez mais fontes de dados que podem ser utilizados para pautar premissas de crescimento de países, setores e empresas. No fim, a consolidação dessas premissas pode alterar completamente o valor justo de uma ação ao utilizar métodos de Análise Fundamentalista, como o fluxo de caixa descontado, por exemplo. Dessa forma, obter meios de agrupar a maior parte dessas fontes de informações que sejam relevantes, bem como automatizar processos de coleta e disponibilização de dados torna-se imprescindível para o sucesso na análise de empresas e consequentemente na seleção de ativos.

Pensando nisso, as técnicas de Big Data estão sendo empregadas com maior frequência no setor financeiro. Além da análise direta de empresas, o uso de dados em finanças pode ajudar a modelar também outras questões que podem ser acopladas em análises de empresas, como probabilidade de falência (ALAKA, OYEDELE, *et al.*, 2019), índices de confiança (SAKAJI, KURAMOTO, *et al.*, 2019) e detecção de fraudes financeiras (TANG, KARIM, 2019). A maioria dos dados para esses tipos de análise de Big Data são não-estruturados e ruidosos, além de que com um aumento de volume tão significativo começam a aparecer problemas computacionais, principalmente em casos em que a velocidade na decisão é de extrema importância (ANDRIOSOPOULOS, DOUMPOS, *et al.*, 2019).

O formato de dados não tabular e o tempo de processamento se tornando cada vez mais relevantes indica que é necessário optar por arquiteturas de dados cada vez mais flexíveis e com maior agilidade. A principal arquitetura de banco de dados não estruturados utilizada atualmente é conhecida como NoSQL que trabalha com um sistema de documentos não relacionais e não tabular, contrariamente ao modelo mais conhecido de armazenamento de dados relacionais que foi popularizada pela linguagem SQL (HAN, HAIHONG, *et al.*, 2011). Esses tipos de banco de dados não relacionais existem desde 1960, mas começaram a ganhar mais popularidade com a crescente demanda por velocidade e flexibilidade advinda de empresas Web 2.0 que oferecem serviços via internet, como as redes sociais, por exemplo.

Segundo (HAN, HAIHONG, *et al.*, 2011), os modelos de dados NoSQL possuem algumas características que os tornam mais eficientes que modelos relacionais, principalmente para aplicações de alta escalabilidade, sendo os principais:

- **Velocidade de escrita e leitura:** um banco de dados relacional possui uma lógica complexa que é formulada conjuntamente com a inserção de dados. Com o aumento linear da quantidade de dados, a formulação e armazenamento das lógicas aumenta exponencialmente o custo computacional exigido. Esse problema não existe no caso de estruturas não relacionais porque não há lógica de estruturação.
- **Custo de armazenamento:** bancos de dados não estruturados possuem uma arquitetura menos complexa e não há armazenamento de lógica, fazendo com que os mesmos dados ocupem uma porção menor de espaço em disco do que os bancos de dados baseados em tabelas. Esse é um fator extremamente importante uma vez que a quantidade de informações está cada vez maior e o custo desse armazenamento torna-se cada vez mais relevante.
- **Concorrência e paralelismo:** a utilização de bancos de dados relacionais por múltiplos usuários ou programas ao mesmo tempo não é possível sem a interface de um gerenciador de banco de dados que garanta que a informação inserida por um usuário não seja mesclada ou substituída pela informação de outro usuário. Além de aumentar o custo computacional, também se dificulta o paralelismo das atividades, uma vez que as alterações de cada usuário devem ser executadas em ordem para garantir a estrutura do banco de dados. Em estruturas NoSQL há maior possibilidade de paralelização, uma vez que objetos diferentes podem ser inseridos de forma independente.

A maioria dos sistemas legado utilizados atualmente ainda se pauta na arquitetura de tabelas estruturadas e relacionais. Apesar do sucesso do modelo relacional, há ainda muito espaço para estudo de estruturas não relacionais. Os dados não tabulares não podem ser corretamente acoplados em tabelas, limitando ou até inviabilizando o uso dessas fontes de dados. Além disso, tabelas relacionais possuem um grande custo computacional acoplado, sendo que o tempo de processamento cresce exponencialmente com o volume de dados existente. Apesar do trabalho de (HAN, HAIHONG, *et al.*, 2011) apontar para uma possível superioridade computacional dos modelos NoSQL, trabalhos como (A. Flores, *et. al.*, 2018) e (Y. Li, S. Manoharan, 2013) demonstram ainda não haver um discernimento sobre qual modelo seria mais eficiente, indicando que mais estudos devem ser realizados e que eles devem ser mais direcionados para a aplicação final.

Dessa forma, o intuito desse trabalho é demonstrar que um modelo de dados flexível para análise e precificação de ativos permite análises mais profundas e com menor custo frente às principais técnicas e dados disponíveis. Esse modelo permite que as técnicas não suprimam particularidades de negócios das empresas e dos setores econômicos a partir da utilização de dados semi e não estruturados. Para exemplificação, o modelo deverá permitir a aplicação de técnicas de dados que contemplem particularidades de balanços patrimoniais quanto às formas que as empresas apresentam seus dados a partir de seus textos em notas explicativas. Outro exemplo é a manutenção de textos recentes no modelo de dados que tenham capacidade de impactar as precificações a curto e longo prazo, como notícias sobre empresas e setores. Todo esse processo pode ser feito apenas com o banco de dados não relacional, sem depender de uma aplicação específica para combinar todos os tipos de dados, uma vez que a característica intrínseca dessa estrutura é permitir a mescla de tipos de dados, havendo apenas a relação de chave e valor como estrutura.

3. Métodos

Para realização desta pesquisa, decidiu-se por realizar um estudo de caso empírico com dois bancos de dados de diferente estrutura, mas com o mesmo conteúdo: informações contábeis de empresas brasileiras. Esse é o mercado mais conhecido pelos autores, além de haver uma carência maior de acesso aos dados financeiros de forma organizada e estruturada em comparação com outros mercados, como o norte americano, por exemplo.

Assim, realiza-se a montagem um banco de dados não relacional com dados de demonstrações financeiras das empresas brasileiras listadas em bolsa de valores. Esse modelo de dados ainda não é muito explorado na literatura, principalmente para métodos de cálculos, uma vez que a maioria dos sistemas atuais sempre se utilizou de tabelas de dados para realização de operações aritméticas em bancos de dados.

A estruturação desse banco de dados não relacional permite avaliar suas vantagens e desvantagens na composição de uma ferramenta de precificação de ativos frente ao mesmo banco de dados em estruturas relacionais e indicar possibilidades de aplicações futuras, caso haja.

Para obtenção dos dados, foram verificadas ao menos cinco fontes de informações públicas de demonstrações financeiras, cotações, notas explicativas e qualquer outra informação que possa agregar valor à análise por meio de fluxo de caixa descontado.

As bases de dados gratuitas geralmente são limitadas em histórico e em informações que podem ser consultadas, além de possuírem um formato próprio dependendo de cada mantenedor do dado. Também não é garantido que esses mantenedores realizem a manutenção desses dados com compromisso real e contábil. Dessa forma, a melhor base pública encontrada foi a da própria Comissão de Valores Mobiliários (CVM) do Brasil. Nela é possível obter um histórico desde 2011, com as demonstrações padronizadas anuais e as demonstrações trimestrais.

Os dados financeiros então serão obtidos do portal de dados abertos da CVM. Nesse portal há todas as demonstrações financeiras enviadas obrigatoriamente ao órgão regulador, seja por se tratar de uma empresa listada em bolsa, ou por ser uma empresa financeira, ou de capital aberto de grande porte. O estudo será focado nos dados de empresas listadas visto que o objetivo é fornecer um banco de dados que possa ser utilizado para

precificação de ativos por meio de análise fundamentalista, sendo nesse caso as ações dessas empresas que são negociadas em bolsa de valores.

Apesar de serem dados públicos, a obtenção do histórico das demonstrações financeiras é operacionalmente custosa e sujeita a erros devido aos arquivos estarem separados por ano e por tipo de demonstração financeira, sendo uma lista de arquivos para informações trimestrais do ano e outra lista para demonstração financeira padronizada de fim de exercício. Além disso, também há também outras duas separações dos arquivos, sendo dois de cada lista para cada demonstração financeira, um sendo sobre a Companhia Consolidada e outro sobre a Companhia Individual, totalizando 15 arquivos por ano de histórico, e 165 arquivos para o histórico desde 2011.

Dessa forma, a obtenção dos dados será feita de forma automática utilizando a linguagem de programação Python para download dos arquivos e organização em um único banco de dados, mitigando erros operacionais e deixando o processo mais veloz do que quando realizado manualmente. Após a obtenção dos dados, serão realizados testes para verificar qual a melhor forma de estruturação desses dados em um banco não relacional utilizando a arquitetura em nuvem oferecida pelo MongoDB.

Após a estruturação do banco de dados NoSQL, será feita uma comparação computacional direta entre este e um banco de dados SQL com as mesmas capacidades de processamento para algumas consultas específicas. Além disso, uma avaliação qualitativa será realizada entre os bancos de dados e suas funções, pretendendo deixar claro a utilidade ou não de um modelo não relacional na avaliação de empresas.

4. Resultados

4.1. Obtenção dos dados

Duas foram as principais fontes de dados, sendo o site da B3 a fonte de dados cadastrais para cada uma das empresas, enquanto o portal de dados abertos da CVM foi a fonte das demonstrações financeiras. As demonstrações utilizadas no trabalho foram o Balanço Patrimonial Ativo, Balanço Patrimonial Passivo, Demonstração de Resultado do Exercício, Fluxo de Caixa por Método Indireto e Fluxo de Caixa por Método Direto. Essas informações existem na forma como são entregues à CVM, sem qualquer tipo de tratamento adicional. Isso faz com que não haja padronização exata de códigos e linhas das demonstrações, uma vez que a CVM permite que os documentos sejam flexíveis para representar a natureza contábil da empresa com a maior fidelidade possível.

Um total de 165 arquivos da CVM foram unificados por meio de um script em Python para formar a base de dados a ser convertida em conjuntos de chave e valor para que haja uma forma de obter os dados de maneira estruturada ainda que armazenados sem estrutura no banco de dados NoSQL.

Um exemplo dos dados disponibilizados pela CVM está na Figura 1 com as informações financeiras de Balanço Patrimonial Ativo.

Figura 1 – Exemplo de arquivo disponibilizado pela CVM.

CNPJ_CIA	DT_REFER	VERSAO	DENOM_CIA	CD_CVM GRUPO_DFP	MOEDA	ESCALA_MOEDA	ORDEM_EXERC	DT_FIM_EXERC	CD_CONTA	DS_CONTA	VL_CONTA	ST_CONTA_FIXA
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010	1	Ativo Total	802819794.0	S
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010 1.01	1.01	Caixa e Equivalentes de Caixa	25219700.0	S
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010 1.01.01	1.01.01	Caixa e Depósitos Bancários	9816675.0	N
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010 1.01.02	1.01.02	Empréstimos a Instituições Finance	14889970.0	N
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010 1.01.03	1.01.03	Aplicações em Operações Compror	513055.0	N
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010 1.02	1.02	Aplicações Financeiras	143840090.0	S
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010 1.02.01	1.02.01	Aplicações Financeiras Avaliadas a	127232783.0	S
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010 1.02.01.01	1.02.01.01	Títulos para Negociação	52222158.0	S
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010 1.02.01.02	1.02.01.02	Títulos Disponíveis para Venda	75010625.0	S
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010 1.02.02	1.02.02	Aplicações Financeiras Avaliadas ac	16607307.0	S
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010 1.02.02.01	1.02.02.01	Títulos Mantidos até o Vencimento	16607307.0	S
00.000.000/0001-91	31/12/2010	3	BCO BRASIL S.A.	1023 DF Consolidado - Balanço Patrim	REAL	MIL	ÚLTIMO	31/12/2010 1.03	1.03	Empréstimos e Recebíveis	442022890.0	S

Fonte: elaborado pelo autor.

4.2. Modelo final de dados

Após o *download* e unificação das bases de dados, foi utilizado outro script em Python para organização dos dados no servidor do MongoDB. A ideia foi transformar cada

linha das tabelas da CVM em um conjunto de chaves e valores que marcasse claramente quais as informações a serem buscadas e acessadas.

O modelo final de dados consiste em um objeto para cada empresa existente na bolsa de valores. Dentro desse objeto há as informações cadastrais obtidas do próprio site da B3. Além das informações cadastrais da bolsa brasileira, há também as informações financeiras obtidas da CVM, sendo essas as informações mais importantes para a realização do *valuation* das empresas.

Esquemas demonstrativo do documento que representa uma empresa e suas informações estão nas Figuras 2 e 3.

Figura 2 – Exemplo de um documento no banco de dados MongoDB.

```
_id: "07.526.557/0001-00"
Atividade Princ...: "Fabricação E Distribuição de Cervejas. Refrigerantes E Bebidas Não Car..."
Classificação S...: "Consumo não Cíclico / Bebidas / Cervejas e Refrigerantes"
Códigos de Nego...: "ABEV3-ABEV3-Códigos ISIN:-BRABEVACNOR1-Códigos CVM:-23264"
Nome de Pregão: "AMBEV S/A"
lastModified: 2021-04-06T10:55:46.531+00:00
> DFs: Object
> Cotações: Object
```

Fonte: elaborado pelo autor.

Figura 3 – Exemplo da estrutura das informações dentro de um documento.

```
_id: "07.526.557/0001-00"
Atividade Princ...: "Fabricação E Distribuição de Cervejas. Refrigerantes E Bebidas Não Car..."
Classificação S...: "Consumo não Cíclico / Bebidas / Cervejas e Refrigerantes"
Códigos de Nego...: "ABEV3-ABEV3-Códigos ISIN:-BRABEVACNOR1-Códigos CVM:-23264"
Nome de Pregão: "AMBEV S/A"
lastModified: 2021-04-06T10:55:46.531+00:00
< DFs: Object
  < 2013-12-31: Object
    < V2: Object
      < Consolidado: Object
      < Individual: Object
        < BPA: Object
          < 1: Object
            Valor: 43882834000
            Desc: "Ativo Total"
            Fixa: "S"
          < 1-01: Object
            Valor: 149230000
            Desc: "Ativo Circulante"
            Fixa: "S"
```

Fonte: elaborado pelo autor.

Pode-se notar que várias colunas repetidas no modelo tabular tornam-se um par de chave e valores no modelo NoSQL, sendo que isto tem a capacidade de reduzir espaço de armazenamento e de processamento, ao passo que gera uma semiestrutura para os dados no MongoDB.

4.3. O modelo NoSQL ocupa menor espaço que um modelo SQL

Em comparações diretas com o modelo de dados SQL, foi possível verificar um menor espaço de armazenamento. O armazenamento dos dados como obtidos da CVM em tabelas SQL utilizaram 5 tabelas, mais de 9 milhões de linhas e 2,19 GB de espaço.

Já para o modelo NoSQL, os mesmos dados puderam ser organizados em apenas uma coleção (equivalente à tabela do SQL), com um documento para cada empresa, totalizando 451 documentos com todas as informações de cadastro e financeiras inseridas em cada objeto e ocupando 0,51 GB de espaço.

Embora haja um a redução de quatro vezes no caso observado, constata-se na literatura que esse nem sempre é o caso. É provável que o que tenha levado a essa redução é o agrupamento de cada linha em um conjunto de várias chaves e valores, o que fornece certa estrutura para o modelo NoSQL e permite diminuir a quantidade de informações repetidas linha a linha no modelo SQL, como as datas de referência, CNPJs e Razões Sociais.

4.4. O modelo NoSQL possui maior velocidade de processamento

Quando analisada quantitativamente a velocidade de execução de consultas aos bancos de dados, o banco NoSQL novamente demonstrou maior performance. As consultas de todas as informações para uma determinada empresa demoraram em média de 2,13s para o banco de dados estruturado enquanto no MongoDB o processo demorou apenas 0,82s em média. Novamente, a grande quantidade de linhas e a necessidade de executar 5 consultas em tabelas diferentes para obter todos os dados financeiros é a mais provável causa da diferença entre o tempo de execução. Enquanto que para obter todos os dados de uma empresa o banco em SQL precisava consultar 5 tabelas com o número de linhas pertencente a cada empresa, que chegava a milhares dependendo do tamanho do histórico, a consulta no MongoDB era em apenas uma coleção e verificava apenas um CNPJ igual ao CNPJ pesquisado.

4.5. Operacionalmente, o modelo NoSQL é menos eficiente

Apesar de ser mais eficiente computacionalmente, o modelo de dados NoSQL demanda maior carga operacional de quem o estrutura. Por ser um banco de dados naturalmente sem estruturas, é necessário pensar na melhor forma de transformar as informações utilizadas tipicamente em maneiras tabulares para a organização em objetos. Dessa forma, o processo de arquitetura deve ser mais bem executado para um banco de dados NoSQL, já que ele não possui qualquer mecanismo que garanta estrutura ou tipos de dados.

As consultas mais específicas também são mais simples de serem realizadas em um banco de dados estruturado. Neste, é possível especificar condições e agrupamentos em qualquer linha ou coluna do banco de dados. Já para o caso do MongoDB, é possível fazer agrupamentos, mas isso torna-se mais difícil quanto maior for a complexidade dos objetos. Uma simples cláusula de *SELECT* com *WHERE* num banco de dados SQL pode tornar-se várias cláusulas diferentes de *FIND* no MongoDB.

Quanto a flexibilidade, é possível pesquisar por um objeto por completo no MongoDB e tratar o restante das informações em outra linguagem que se julgue mais fácil, como Python, por exemplo. Esse processo também é possível em SQL, mas demanda maior conhecimento da integração entre as linguagens e perde parte da eficiência, já que os códigos em SQL costumam ser mais rápidos quando comparados com as mesmas manipulações em outras linguagens, como ocorre quando comparado com a manipulação de dados em Python.

Em resumo, o banco de dados não estruturado só resultará em consultas fáceis de serem realizadas caso ele seja desenhado para tal consulta. Nesse caso, uma consulta imprevista pode demandar um pedaço de código específico algumas vezes maior que o código para realizar a mesma consulta em um banco de dados SQL.

4.6. O modelo permite maior flexibilidade na precificação de ativos

Nesse tópico foi verificado maior ganho qualitativo para o banco de dados NoSQL em comparação com o banco de dados SQL. Enquanto os dados tabulares limitam-se sempre à mesma estrutura, nos dados não estruturados não há essa limitação.

Dessa forma, com um modelo mais enxuto, é possível armazenar mais informações de maneira mais eficiente computacionalmente e ainda assim não perder informações vitais para a avaliação de empresas. Um caso de exemplo é a possibilidade de verificar quais foram as mudanças de linhas contábeis de um ano para o outro e ainda assim permitir que operações sejam realizadas com cada linha de cada ano em separado, enquanto

num banco de dados SQL essas contas precisariam de unificação para realizar operações aritméticas.

Em resumo, a maior flexibilidade permite criar modelos mais específicos para cada empresa a ser analisada. Isso demanda um maior custo em tempos de programação do que formular um modelo geral, mas permite uma melhor análise. Formular o mesmo modelo flexível com bancos de dados estruturados é inviável caso sejam utilizadas informações como textos longos, notas explicativas ou notícias. Mesmo que essas informações não sejam utilizadas, formular um modelo flexível com dados tabulares também demanda mais tempo de programação do que com o banco de dados NoSQL pelos testes verificados no trabalho, pois haveria a necessidade de realizar um código condicional para cada empresa.

5. Conclusão

O intuito desse trabalho é demonstrar que um modelo de dados NoSQL permite realizar a análise e precificação de ativos com maior profundidade e com menor custo computacional frente ao modelo SQL.

Embora algumas métricas quantitativas podem ser observadas nesse estudo específico, salienta-se aqui que o banco de dados utilizado ainda foi pequeno para conclusões puramente computacionais, embora aponte-se uma superioridade do modelo NoSQL. Quando comparado com um banco de dados SQL em nuvem com as mesmas capacidades de processamento, as consultas NoSQL apresentaram maior velocidade e os dados demonstraram menor espaço de armazenamento.

Contudo, o grande ponto de destaque verificado com o estudo é que, qualitativamente, a estruturação de uma base de dados flexível auxilia a realização do *valuation* de empresas, pois permite compreender maiores aspectos sobre a estrutura contábil e de negócios da empresa a ser analisada. Verificou-se com a montagem do banco de dados não relacional que essa pode ser uma estrutura explorada mais a fundo para a avaliação de valor justo de *equities* de uma forma diferente da realizada atualmente. Quanto ao custo computacional, sugere-se realizar um estudo acoplando modelos de *valuation* mais complexos aos modelos de dados para tirar a conclusão correta.

Enquanto os modelos tradicionais pautam-se na universalização do modelo, agrupando linhas das demonstrações financeiras para realização dos cálculos com tabelas, o

modelo de dados não relacional permite a particularização dos modelos de *valuation*, sendo possível a montagem de um tipo de análise para cada setor, segmento ou até mesmo para cada empresa, fazendo com que não se desconsidere importantes linhas contábeis ou com que elas sejam agrupadas dentro de outra conta maior, processo necessário quando há universalização do modelo.

Não foi testado ainda quais modelos de cálculo e predição do valor justo de *equities* seriam melhores se aplicados na estrutura de dados NoSQL. Contudo, sua grande flexibilidade não apresenta restrições para qualquer método presente na literatura atual. O banco de dados pode ser utilizado tanto para técnicas de Big Data, quanto para técnicas automatizadas ou mesmo apenas para modelos de consulta em que um analista fará a precificação de maneira mais manual. É necessário apenas a organização de scripts que busquem os dados necessários e estruturarem para cada modelo, uma vez que o modelo de dados NoSQL não é estruturado. Esse é um ponto em que há uma desvantagem do modelo NoSQL, pois a operacionalização de um modelo é geralmente mais custosa em tempo de programação já que se deve primeiro criar um script que busque e organize os dados da forma que são necessários para a utilização.

Pelo observado, pode-se montar um esquema de *valuation* para cada empresa, se isso for necessário, bem como para os segmentos ou setores, ou mesmo um modelo geral para todas as empresas. Enquanto os modelos gerais serão mais eficientes em comparações de empresas, os modelos específicos servem para estimar o valor justo de uma maneira mais refinada, chegando possivelmente a um valor mais correto. Essa flexibilidade permite ao analista obter mais de uma gama de valores justos partindo de um mesmo modelo de dados e utilizando-se de mais de uma resposta para realizar suas indicações de investimentos, sem desconsiderar particularidades de cada empresa e negócio.

6. Referências

ABU-MOSTAFA, Y. S., ATIYA, A. F. "Introduction to financial forecasting", *Applied Intelligence*, v. 6, n. 3, p. 205–213, 1996. DOI: 10.1007/BF00126626. Disponível em: <https://doi.org/10.1007/BF00126626>.

A. FLORES, *et al.*, "Performance Evaluation of NoSQL and SQL Queries in Response Time for the E-government," 2018 *International Conference on eDemocracy & eGovernment (ICEDEG)*, 2018, pp. 257-262, doi: 10.1109/ICEDEG.2018.8372362.

ALAKA, H., OYEDELE, L., OWOLABI, H., *et al.* "A Big Data Analytics Approach for Construction Firms Failure Prediction Models", *IEEE Transactions on Engineering Management*, 2019. DOI: 10.1109/TEM.2018.2856376. .

ANDRIOSOPOULOS, D., DOUMPOS, M., PARDALOS, P. M., *et al.* Computational approaches and data analytics in financial services: A literature review. *Journal of the Operational Research Society*. [S.l: s.n.]. , 2019

DUTTA, A., BANDOPADHYAY, G., SENGUPTA, S. "Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression", *International Journal of Business and Information*, 2012. .

FAMA, E. F. "Random Walks in Stock Market Prices", *Financial Analysts Journal*, n. February, p. 75–80, 1995. .

GORDON, M. J. "DIVIDENDS , EARNINGS , AND STOCK PRICES", v. 41, n. 2, p. 99–105, 2013. .

HAN, J., HAIHONG, E., LE, G., *et al.* "Survey on NoSQL database", *Proceedings - 2011 6th International Conference on Pervasive Computing and Applications, ICPCA 2011*, p. 363–366, 2011. DOI: 10.1109/ICPCA.2011.6106531. .

HU, Y., LIU, K., ZHANG, X., *et al.* "Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review", *Applied Soft Computing Journal*, v. 36, p. 534–551, 2015. DOI: 10.1016/j.asoc.2015.07.008. Disponível em: <http://dx.doi.org/10.1016/j.asoc.2015.07.008>.

IMAM, S., BARKER, R., CLUBB, C. "The Use of Valuation Models by UK Investment

Analysts", *European Accounting Review*, p. 503–535, 2008. DOI: 10.1080/09638180802016650. .

SAKAJI, H., KURAMOTO, R., MATSUSHIMA, H., *et al.* "Financial Text Data Analytics Framework for Business Confidence Indices and Inter-Industry Relations". 2019.

SHAH, D., ISAH, H. "Stock Market Analysis : A Review and Taxonomy of Prediction Techniques", *International Journal of Financial Studies*, v. 7, 2019. .

TANG, J., KARIM, K. E. "Financial fraud detection and big data analytics – implications on auditors' use of fraud brainstorming session", *Managerial Auditing Journal*, 2019. DOI: 10.1108/MAJ-01-2018-1767. .

Y. LI, S. MANOHARAN, "A performance comparison of SQL and NoSQL databases," 2013 *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, 2013, pp. 15-19, doi: 10.1109/PACRIM.2013.6625441.