Análise de Composição de Carteira de Fundos de Investimento

Andréia M. Luna1, Anderson S. Nascimento1

1Pontifícia Universidade Católica do Rio de Janeiro – Rio de Janeiro – RJ – Brasil

am.luna@hotmail.com, prof.anderson@ica.ele.puc-rio.br

**Abstract.** This paper describes the use of graph databases to analyse investment funds’ portfolios. The set of financial assets an investment fund owns is known as its portfolio, which is managed by his investment adviser. The choice of the graph-based data model to represent this problem domain was motivated by the need to produce analyses focused on the relationships between the entities: a fund invests in assets; a fund invests in shares of other funds; assets are issued by financial institutions or companies; funds’ portfolios are managed by individuals or legal entities.

**Resumo.** Este artigo descreve a utilização da abordagem baseada em banco de dados de grafos para analisar a composição de carteira de fundos de investimento. A composição da carteira de um fundo é o conjunto de ativos financeiros nos quais o patrimônio do fundo está aplicado, cuja seleção é de responsabilidade de seu(s) gestor(es). A escolha do modelo de dados baseado em grafos para representar este domínio de problema foi motivada pela necessidade de se produzir análises centradas nos relacionamentos entre as entidades: fundo aplica em ativos; fundo aplica em cotas de outros fundos; ativos são emitidos por instituições financeiras/empresas; fundos são geridos por pessoas físicas ou jurídicas.

# 1. Introdução

De acordo com dados da Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais – ANBIMA, a indústria de fundos de investimento encerrou o ano de 2020 com captação líquida positiva, apesar de todo o impacto da crise financeira causada pela pandemia do coronavírus [ANBIMA 2021].

Tal fato ilustra a relevância do mercado de fundos de investimento brasileiro, evidenciada também pelo surgimento de diversas plataformas online de análise de fundos, tais como: [Mais Retorno](https://maisretorno.com/), [Info Fundos](https://infofundos.com.br/), [Comparador de Fundos](https://www.comparadordefundos.com.br/) e [Yubb](https://yubb.com.br/). Essas plataformas disponibilizam informações tais como rentabilidade histórica, evolução do patrimônio, nível de risco e composição da carteira, entre outras.

Este artigo descreve a realização de uma prova de conceito com o objetivo de analisar a composição de carteira de fundos de investimento experimentando uma abordagem baseada em banco de dados de grafos (*graph database*). A composição da carteira de um fundo é o conjunto de ativos financeiros nos quais o patrimônio do fundo está aplicado. São exemplos de ativos financeiros: Letras Financeiras, debêntures, CDBs (Certificados de Depósito Bancário), ações, opções, contratos de Swap e cotas de outros fundos. A seleção dos ativos em que os recursos do fundo serão aplicados é de responsabilidade de seu(s) gestor(es). A escolha do modelo de dados baseado em grafos para representar este domínio de problema foi motivada pela necessidade de se produzir análises centradas nos relacionamentos entre as entidades: fundo aplica em ativos; fundo aplica em cotas de outros fundos; ativos são emitidos por instituições financeiras/empresas; fundos são geridos por pessoas físicas ou jurídicas.

# 2. Descrição do negócio

Os Fundos de Investimento são uma forma organizada de captar recursos de investidores, individuais e institucionais, e aplicá-los em carteiras de ativos disponíveis no mercado financeiro e de capitais.

O **patrimônio líquido** de um fundo é calculado pela soma de todos os ativos que constituem a sua carteira de investimento e o valor em caixa, menos as suas despesas e obrigações.

O patrimônio dos fundos é investido em ativos financeiros, de emissores públicos ou privados, emitidos no Brasil ou no exterior. Esses ativos compõem o que se chama de **carteira do fundo**.

O **gestor da carteira** é o profissional (pessoa física ou jurídica credenciada pela CVM como administrador de carteiras de valores mobiliários), responsável pelos investimentos realizados pelo fundo. É quem decide quais ativos financeiros irão compor a sua carteira. No entanto, suas decisões de compra e venda de ativos são regidas pelos objetivos e política de investimento determinados no regulamento do fundo [Comissão de Valores Mobiliários].

O funcionamento dos fundos de investimento está sujeito às normas publicadas pela Comissão de Valores Mobiliários (CVM). Há diversos tipos de fundos e a sua constituição, funcionamento e divulgação de informações seguem o estabelecido em diferentes normas, por exemplo, a [Instrução CVM 555](http://conteudo.cvm.gov.br/legislacao/instrucoes/inst555.html).

Os fundos regulados pela Instrução CVM 555 (e alguns outros tipos) devem enviar o demonstrativo mensal da composição da sua carteira por meio do [Sistema de Envio de Documentos da CVM](http://sistemas.cvm.gov.br/?cvmweb).

O **Informe de Composição e Diversificação das Aplicações (CDA)** é o documento que descreve as aplicações que compõem as carteiras dos fundos de investimento (ativos financeiros ou cotas de outros fundos), entregue à CVM no [formato XML](https://cvmweb.cvm.gov.br/SWB/Sistemas/SCW/PadroesXML/PadraoXMLCDANetV4.aspx).

Assim, as carteiras individualizadas dos fundos podem ser consultadas no sítio da CVM, na página [Consulta a Carteiras de Fundos](https://cvmweb.cvm.gov.br/SWB/Sistemas/SCW/CPublica/FormBuscaPartic.aspx?TpConsulta=5). Além disso, a CVM disponibiliza essas carteiras para download em massa (bulk download) por meio de seu [Portal de Dados Abertos](http://dados.cvm.gov.br/dataset/fi-doc-cda).

# 3. Problema proposto

Existem diversas ferramentas online que permitem consultar as aplicações de um fundo, mas que se limitam a exibir as mesmas informações encontradas na [Consulta a Carteiras de Fundos](https://cvmweb.cvm.gov.br/SWB/Sistemas/SCW/CPublica/FormBuscaPartic.aspx?TpConsulta=5), no sítio da CVM.

Todas essas consultas apresentam o portfólio de investimentos que consta no documento **Composição e Diversificação das Aplicações (CDA)** do fundo (chamaremos esta versão da carteira de "**simples**"). No entanto, nenhuma das consultas avaliadas põe em perspectiva o relacionamento entre fundos.

Nos casos de aplicações em cotas de outros fundos, as consultas tradicionais à carteira não evidenciam, por exemplo, quais empresas e outras instituições são de fato financiadas com parte dos recursos do fundo. No entanto, é possível fazer tal inferência, bastando "navegar" do fundo investidor para os fundos que recebem o investimento.

Considerando que o CDA informa a relação fundo investidor/fundo investido, definimos a **versão "expandida" da carteira** como sendo todos os ativos e fundos em que ele aplica, direta ou indiretamente. Ou seja, a carteira expandida de um fundo inclui a carteira do próprio fundo e os ativos que estão nas carteiras de todos os fundos em que ele investe.

Sejam:

**F0**: o fundo em análise

**C0**: a carteira informada no documento CDA do fundo F0, denominada **carteira simples** (conjunto com as aplicações em ativos e em cotas de outros fundos)

**Catvi**: a carteira de ativos do fundo Fi (a carteira informada no documento CDA excluindo-se do conjunto as aplicações em cotas de outros fundos)

Definimos, no escopo deste trabalho, a **carteira expandida de nível 1** (**Ce01**) do fundo F0:

**Ce01** = C0 + Catvi, para cada Fi que aparece em C0.

A **carteira expandida de nível 2** (**Ce02**) do mesmo fundo F0 será definida como:

**Ce02** = C0 + Cei, para cada Fi que aparece em C0.

Por fim, definimos **a carteira expandida de nível n** (**Ce0n**) do mesmo fundo F0:

**Ce0n** = C0 + Cein-1, para cada Fi que aparece em C0, onde n é a profundidade máxima da árvore de investimentos em cotas de fundos cuja raiz é Fi.

Estamos interessados em construir análises e consultas que explorem a recursividade expressa no conceito de carteira expandida.

# 4. Objetivo e escopo da Prova de Conceito

O objetivo da prova de conceito é analisar a composição de carteira de fundos de investimento experimentando uma abordagem baseada em banco de dados de grafos.

Um **banco de dados de grafos** é um tipo de banco NoSQL baseado na teoria dos grafos, ramo da matemática discreta. As principais definições dessa teoria estabelecem a nomenclatura utilizada nos bancos de grafos:

* Grafo: conjunto de vértices e arestas;
* Vértices: são os nós ou entidades do grafo;
* Arestas: representam os relacionamentos entre as entidades (vértices) do grafo

Embora os grafos sejam uma abstração natural para um amplo leque de problemas de modelagem, a tecnologia de armazenamento e consulta de informação baseada em grafos tem um nicho de aplicação bem definido. Bancos de grafos são a escolha apropriada quando precisamos modelar dados altamente relacionados e os relacionamentos entre as entidades são tão (ou mais) importantes quanto as entidades em si.

[Bechberger e Perryman 2020] listam as consultas recursivas entre os tipos de problemas para os quais os bancos de grafos permitem construir soluções mais simples e eficientes, quando comparados com os bancos de dados relacionais.

Consultas recursivas são aquelas executadas múltiplas vezes em um loop, repetidamente chamando a si mesmas até que seja atingida uma condição de parada [Bechberger e Perryman 2020]. Consultas recursivas em um banco de dados relacional utilizando a linguagem SQL sofrem de dois problemas: ineficiência e complexidade. Por outro lado, os bancos de grafos são otimizados para responder a esse tipo de consulta de maneira mais rápida e com uma sintaxe muito mais simples.

O problema proposto para esse trabalho possui ao menos duas características que sugerem uma abordagem baseada em grafos:

* A recursividade da carteira "expandida" de aplicações de um fundo;
* A hierarquia existente entre os fundos (fundos investidores e fundos investidos), implícita na recursividade da carteira "expandida"

Dessa forma, adotaremos um banco de dados de grafos para modelar e armazenar as informações do domínio do problema. Além de consultar as carteiras simples e expandidas dos fundos, estamos interessados em descobrir:

* Fundos investidores e fundos investidos (direta ou indiretamente);
* Se há ciclos no grafo, ou seja, se há referências circulares na hierarquia de investimento entre fundos (Exemplo: o fundo A investe no fundo B; o fundo B investe no fundo C; o fundo C investe no fundo A);
* Quais fundos possuem maior diversidade de ativos em suas carteiras (simples ou expandidas);
* Quais ativos possuem maior diversidade de fundos investidores (diretos ou indiretos);
* Quais emissores de ativos possuem maior diversidade de fundos investidores (direta ou indiretamente).

# 5. Modelo Conceitual de Dados

**Fundos de investimento** *aplicam* em **ativos** financeiros ou em cotas de outros fundos. Os ativos são *emitidos* por instituições financeiras ou empresas (**emissores**). Fundos são *geridos* por pessoas físicas ou jurídicas (**gestores**), que executam a política de investimento do fundo.

Diagrama

Descrição gerada automaticamente

Figura 1. Modelo Conceitual de Dados

# 6. Modelo Lógico de Dados (*Graph Schema*)

O modelo lógico (esquema) do banco de dados de grafos construído durante a prova de conceito é ele mesmo representado por um grafo. Aqui, os vértices correspondem aos tipos de vértice (entidades) e as arestas, aos tipos de aresta (relacionamentos).

O esquema também exibe as propriedades de vértices e arestas. A propriedade '#' nas arestas representa a cardinalidade do relacionamento e deve ser lida na direção indicada pela seta [Ramachandran 2015].

Diagrama

Descrição gerada automaticamente

Figura 2. Modelo Lógico de Dados

# 7. Coleta e preparação dos dados

A carteira dos fundos, a identificação de seus gestores e muitas outras informações relevantes são dados públicos que podem ser consultados, de forma centralizada, no [Portal de Dados Abertos da CVM](http://dados.cvm.gov.br/). Essa é a fonte principal das informações utilizadas para a análise das carteiras. As informações de emissores de ativos, porém, foram complementadas com dados obtidos no site da [B]3.

Utilizamos a linguagem Python e o ambiente do Google Colab para coletar e preparar os dados para carga no banco de dados de grafos. Para cada tipo de vértice e cada tipo de aresta definidos no esquema lógico do banco, foi construído um processo de geração de arquivo de dados no formato csv, para importação no banco de grafos.

Fontes de dados utilizadas:

* [Fundos de Investimento: Informação Cadastral](http://dados.cvm.gov.br/dataset/fi-cad): informações básicas sobre os fundos, por exemplo: CNPJ, nome, tipo, situação atual do registro na CVM, quem são seus gestores, etc...
* [Fundos de Investimento: Documentos: Composição e Diversificação das Aplicações](http://dados.cvm.gov.br/dataset/fi-doc-cda): o conteúdo do documento CDA, ou seja, a carteira mensal de aplicações do fundo, descrevendo o valor financeiro aplicado e o tipo do ativo
* [Cadastro ISIN de ativos](https://sistemaswebb3-listados.b3.com.br/isinPage/?language=pt-br): base de dados que atribui a cada ativo um código único de identificação conhecido como ISIN (*International Securities Identification Number*). Além do código ISIN do ativo, a base informa também a sua descrição e o seu emissor. Esses dados são complementares ao conteúdo do documento CDA, que não traz o emissor de todos os ativos.

# 8. Mecanismo de banco de grafos

Um banco de dados de grafos combina tecnologia de persistência e uma linguagem de consulta baseados em estruturas de vértices e arestas para entregar um sistema otimizado para o armazenamento e a recuperação rápida de dados altamente conectados [Bechberger e Perryman 2020].

A página [Best Graph Databases in 2021](https://www.g2.com/categories/graph-databases) apresenta uma lista de sistemas de gerenciamento de bancos de dados orientados a grafos, tanto comerciais quanto *open source*, com destaque para o Neo4j, Amazon Neptune, Azure Cosmos DB e DataStax. Nessa classificação, Neo4j e Amazon Neptune estão posicionados como líderes de mercado, entretanto, selecionamos o [Datastax Enterprise Graph](https://www.datastax.com/products/datastax-graph) para implementar a prova de conceito por causa das seguintes características:

* Baseado no [Apache Cassandra](https://cassandra.apache.org/), banco NoSQL *open source;*
* Suporte nativo ao [Apache TinkerPop](https://tinkerpop.apache.org/), mecanismo de grafo *open source*, e à linguagem de travessia de grafo [Gremlin](https://tinkerpop.apache.org/gremlin.html);
* Integrado ao DSE Search, mecanismo de indexação e busca de texto baseado nas plataformas *open source* [Apache Solr](https://solr.apache.org/) e [Apache Lucene](https://lucene.apache.org/);
* Inclui o Datastax Studio, ferramenta de desenvolvimento e teste de consultas nas linguagens CQL (Cassandra Query Language) e Gremlin, visualização e exploração de dados, em um ambiente interativo baseado em notebooks;
* Inclui o DataStax Bulk Loader, ferramenta de linha de comando para carga de dados, com suporte a arquivos no formato csv;
* Licenciado sem custo para utilização por até 90 dias em sistemas não produtivos[[1]](#footnote-1);
* Extensa documentação[[2]](#footnote-2) e material de treinamento[[3]](#footnote-3) *online*

# 9. Grafo de carteiras de fundos

O esquema do grafo carregado no banco DSE Graph está representado a seguir. A definição do esquema bem como todas as travessias e consultas ao grafo foram escritas na linguagem Gremlin, dentro do ambiente Datastax Studio. Os vértices e arestas foram criados a partir dos arquivos csv gerados na etapa de coleta e preparação, importados através da ferramenta DataStax Bulk Loader.

O grafo foi carregado com os dados obtidos em 31/03/2021 (posição da carteira dos fundos: novembro/2020).

Diagrama

Descrição gerada automaticamente

Figura 3. Esquema do Grafo de Carteiras de Fundos (DSE Graph)

## 9.1. Cardinalidade

O grafo analisado possui um total de 92.932 vértices e 548.095 arestas. Os principais números são: 20.564 fundos, 343.747 aplicações em ativos e 140.427 aplicações em cotas de fundos.

Gráfico, Gráfico de cascata

Descrição gerada automaticamente

Figura 4. Número de vértices por tipo

Gráfico, Gráfico de barras

Descrição gerada automaticamente

Figura 5. Número de arestas por tipo

## 9.2. Grau dos Vértices

A classificação dos vértices pelo número de arestas incidentes permite descobrir, por exemplo, quais fundos possuem maior diversidade de ativos em sua carteira (simples ou expandida).

* Fundos com maior diversidade de ativos na carteira simples:
  1. Caieiras FIM IE[[4]](#footnote-4): 1 0.175 aplicações em ativos;
  2. FALX FIM CP IE[[5]](#footnote-5): 8.852 aplicações em ativos;
  3. Gladius FIM CP IE[[6]](#footnote-6): 5.043 aplicações em ativos;
  4. FIM CP LS IE[[7]](#footnote-7): 4.598 aplicações em ativos;
  5. Legacy Capital Master FIM[[8]](#footnote-8): 3.228 aplicações em ativos

Os resultados acima contam apenas os ativos investidos de forma direta, ou seja, a busca no grafo não atravessou nenhuma aresta *aplicacao\_fundo*.

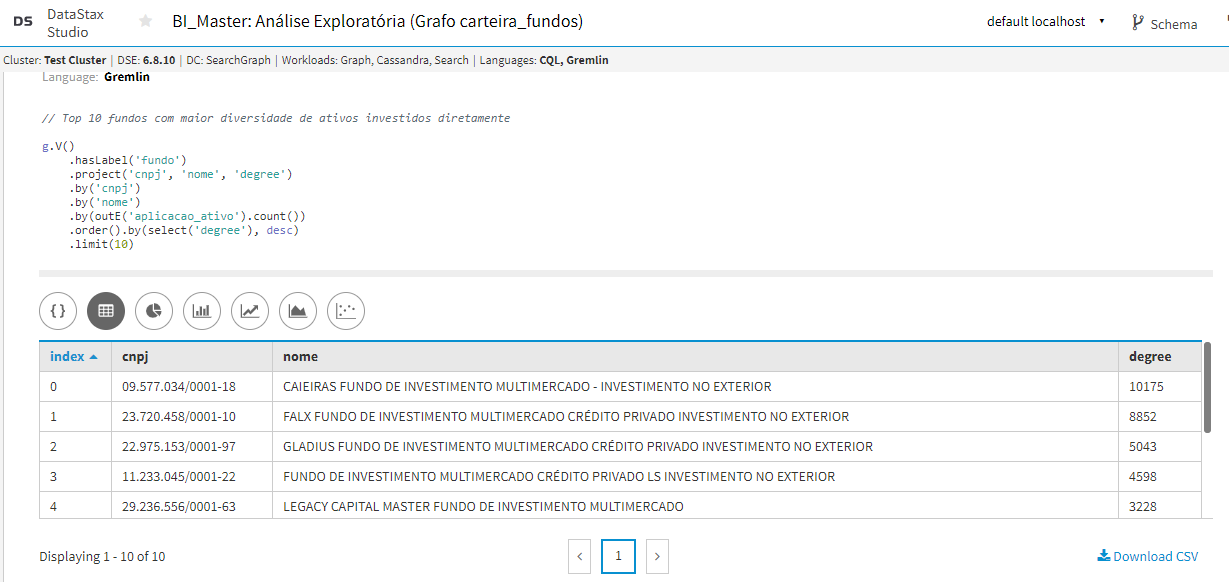


Figura 6. Fundos com maior diversidade de ativos (carteira simples)

* Fundos com maior diversidade de ativos na carteira expandida (nível n):
  1. Bradesco FIM CP IE Andrômeda[[9]](#footnote-9): 24.636 aplicações em ativos;
  2. Bradesco FIM CP IE CDI Mais[[10]](#footnote-10): 20.458 aplicações em ativos;
  3. Bradesco FIC FI MULT CP IE Riviera II[[11]](#footnote-11): 18.050 aplicações em ativos;
  4. Bradesco FIC FIM CP IE Darach[[12]](#footnote-12): 17.677 aplicações em ativos;
  5. Bradesco FIC FIM CP IE Gomez Kawall[[13]](#footnote-13): 17.561 aplicações em ativos;

Os resultados acima contam todos os ativos investidos pelo fundo de forma direta ou indireta. Para isso, a busca no grafo atravessou, a partir de cada fundo, todas as arestas *aplicacao\_fundo*, de forma recursiva, tantas vezes quanto possível.

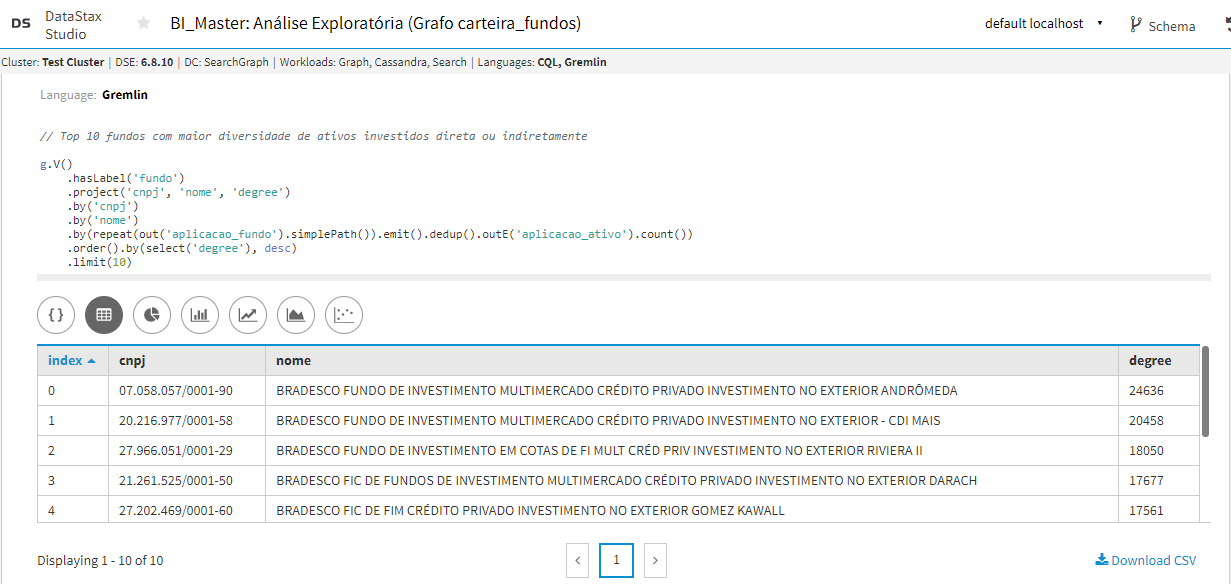


Figura 7. Fundos com maior diversidade de ativos (carteira expandida)

Consultas de natureza recursiva também foram utilizadas para determinar os fundos com maior diversidade de fundos investidos, os fundos com maior diversidade de fundos investidores e os ativos com maior diversidade de fundos investidores, para citar apenas alguns exemplos.

## 9.3. Detecção de Ciclos

Seja o subgrafo Gs(Vs, As), onde Vs é o conjunto dos vértices do tipo fundo e As é o conjunto das arestas do tipo *aplicacao\_fundo*.

A travessia no subgrafo Gs com o objetivo de verificar a existência de ciclos (caminhos fechados, onde os vértices de origem e destino coincidem) não retornou nenhum resultado.

Assim, nas carteiras analisadas, não foram encontradas referências circulares na hierarquia de investimento entre fundos.

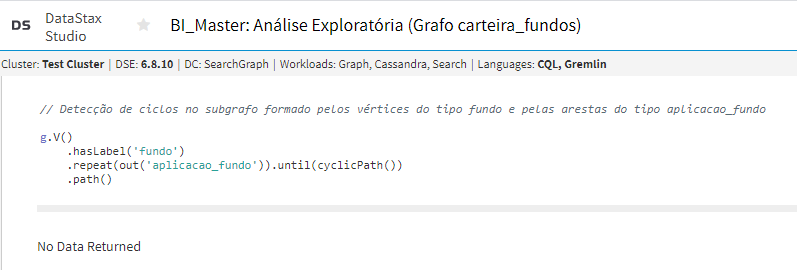


Figura 8. Busca de ciclos

## 9.4. Composição da carteira

Na seção 2 deste artigo, definimos subconjuntos e superconjuntos da carteira de investimentos dos fundos. A expressividade da linguagem Gremlin nos permite navegar através do grafo, especificando uma sequência de passos conhecida como travessia. Se o resultado obtido com essa travessia são as aplicações em ativos e em outros fundos, então temos a versão simples da carteira. Suprimir itens da carteira é uma questão de remover passos da travessia; incluir em um único resultado a carteira de todos os fundos relacionados por meio do investimento em cotas é uma questão de adicionar estruturas de repetição à sequência original de passos.

## 9.4.1. Carteira de ativos

Consulta à carteira de ativos de um fundo F0 (**Catv0**):

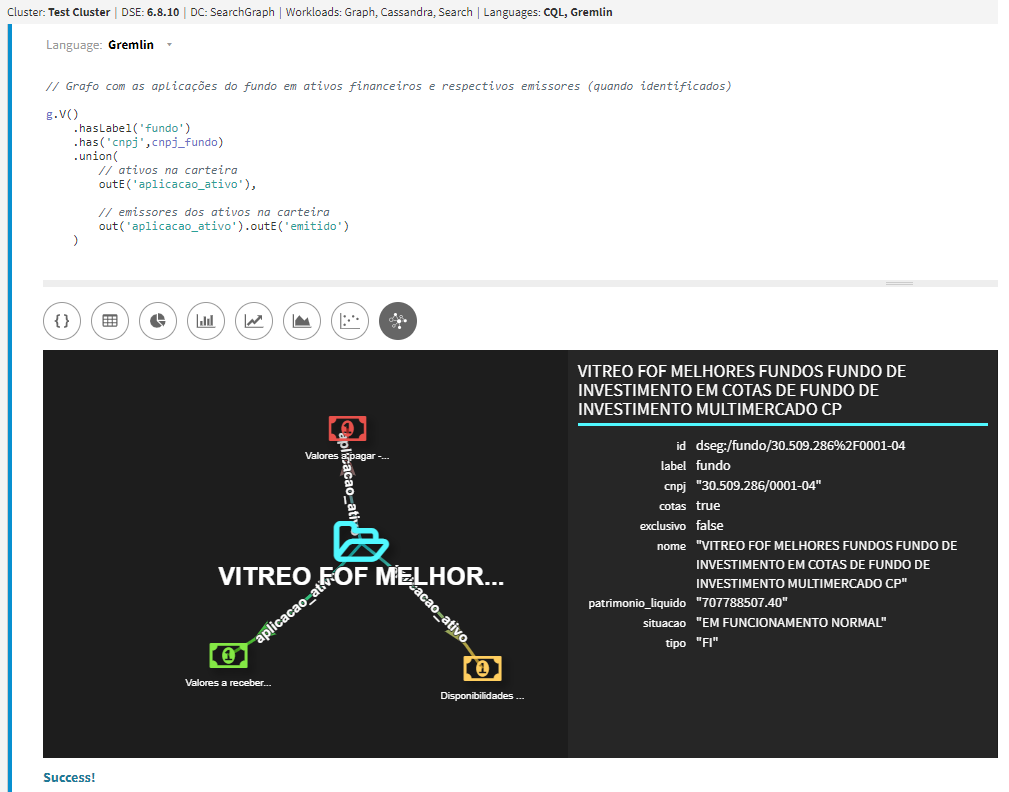


Figura 9. Carteira de ativos do fundo VITREO FOF Melhores Fundos FIC Multimercado CP (CNPJ: 30.509.286/0001-04)

## 9.4.2. Carteira simples

Consulta à carteira simples de um fundo F0 (**C0**):

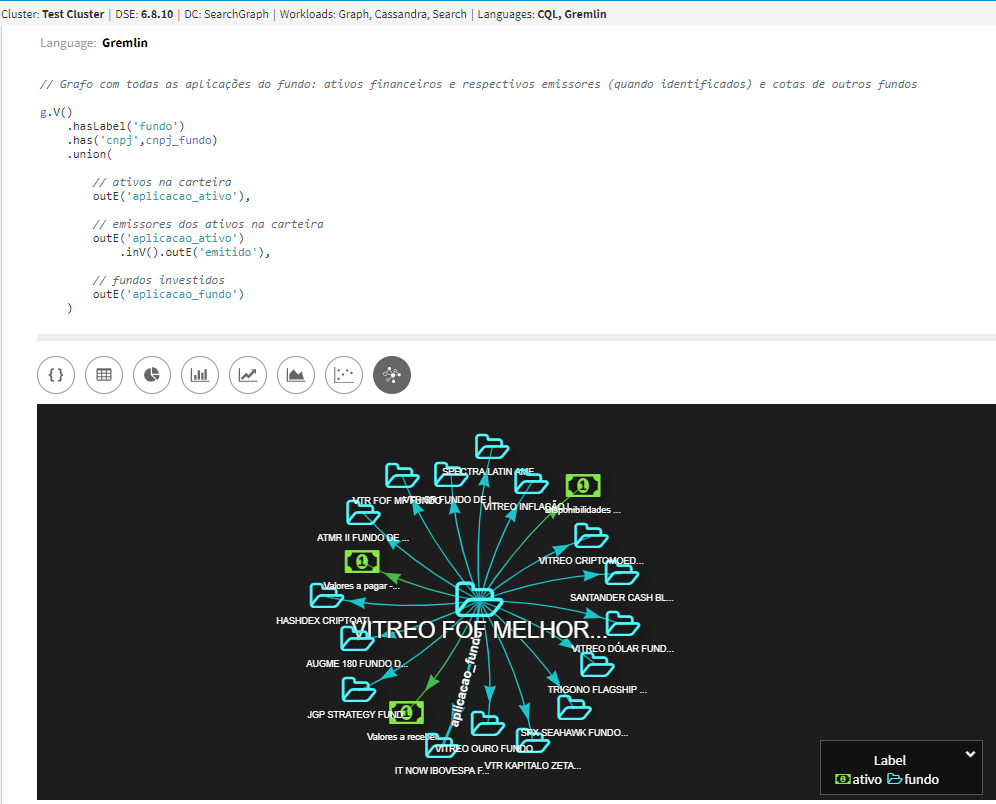


Figura 10. Carteira simples do fundo VITREO FOF Melhores Fundos FIC Multimercado CP (CNPJ: 30.509.286/0001-04)

## 9.4.3. Carteira expandida (nível 1)

Consulta à carteira expandida de nível 1 de um fundo F0 (**Ce01**):

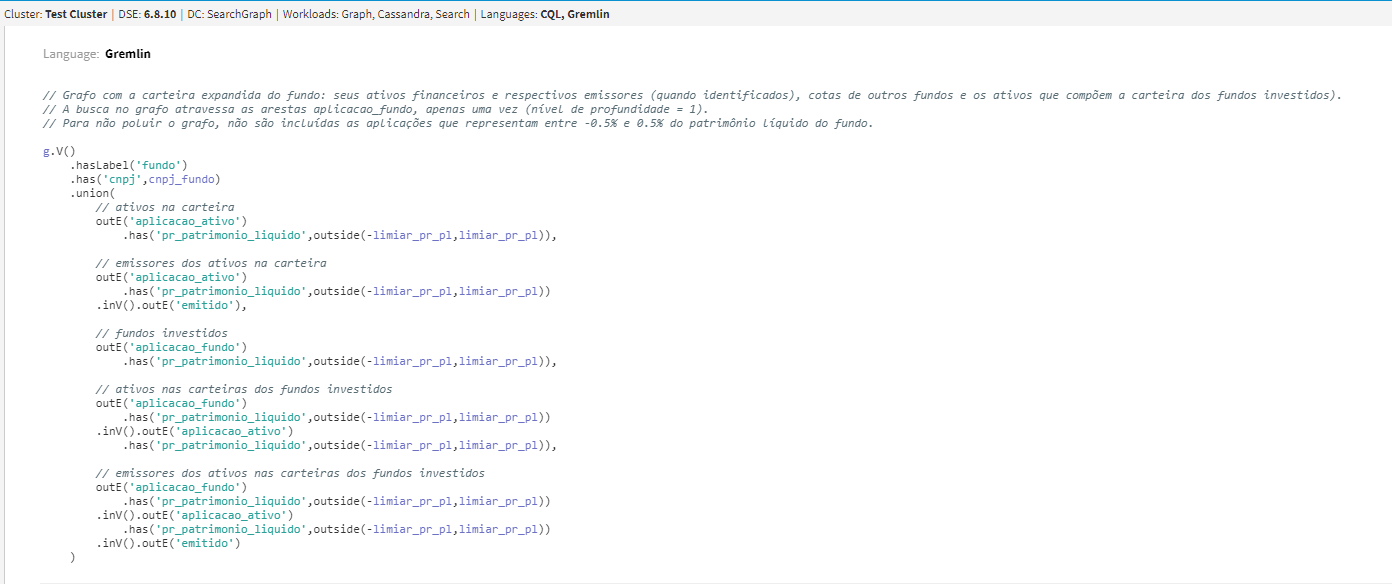


Figura 11. Consulta à carteira expandida nível 1

## 9.4.4. Carteira expandida (nível 2)

Consulta à carteira expandida de nível 2 de um fundo F0 (**Ce02**):



Figura 12. Consulta à carteira expandida de nível 2

# 10. Conclusão

Todas as etapas da prova de conceito estão documentadas no [GitHub](https://github.com/amlua/bi-master), incluindo:

* O código Python utilizado para coletar e preparar os dados para carga no banco de dados de grafos;
* Os arquivos de dados fonte obtidos para o processo de preparação;
* Os procedimentos de instalação, configuração e carga do servidor de banco de dados DSE Graph;
* O código Gremlin construído para explorar e consultar o grafo de carteira de fundos, bem como os resultados obtidos.

As consultas ao grafo exemplificadas neste artigo fornecem respostas às perguntas levantadas na seção 4, que são casos de uso típicos das soluções orientadas a grafos. Além disso, tipos de consultas comumente construídas com a linguagem SQL sobre dados em bancos relacionais também foram experimentados nesta prova de conceito, utilizando a linguagem Gremlin e a travessia de grafos.

Não obstante, avaliamos que o modelo de dados baseado em grafos se mostrou de fato adequado ao problema central proposto: examinar a composição das carteiras de aplicações nos casos dos fundos investidores em cotas de outros fundos. O desenho do esquema lógico reflete como as relações no domínio do problema são descritas em linguagem natural, o que é importante para simplificar o trabalho de especificar como “navegar” sobre essas relações. O mecanismo de grafo e a linguagem de travessia permitem descrever e executar recursivamente percursos complexos na estrutura do grafo em busca de informações, de forma sucinta e eficiente.

# Referências

ANBIMA (2021) “Fundos encerram ano com captação líquida positiva de R$ 156,4 bilhões”, https://www.anbima.com.br/pt\_br/noticias/fundos-encerram-ano-com-captacao-liquida-positiva-de-r-156-4-bilhoes-8A2AB2B67692226E0176F83667DE2D01-00.htm

Comissão de Valores Mobiliários. “Fundos de Investimento”, https://www.investidor.gov.br/menu/Menu\_Investidor/fundos\_investimentos/introducao

Bechberger, D. e Perryman, J. (2020) “Graph Databases in Action”, Manning Publications; 1ª edição.

Ramachandran, S. (2015) “Graph Database Theory: Comparing Graph and Relational Data Models”, https://www.lambdazen.com/assets/pdf/GraphDatabaseTheory.pdf

1. <https://www.datastax.com/legal/datastax-enterprise-terms> [↑](#footnote-ref-1)
2. <https://docs.datastax.com/> [↑](#footnote-ref-2)
3. <https://academy.datastax.com/> [↑](#footnote-ref-3)
4. CNPJ: 09.577.034/0001-18 [↑](#footnote-ref-4)
5. CNPJ: 23.720.458/0001-10 [↑](#footnote-ref-5)
6. CNPJ: 22.975.153/0001-97 [↑](#footnote-ref-6)
7. CNPJ: 11.233.045/0001-22 [↑](#footnote-ref-7)
8. CNPJ: 29.236.556/0001-63 [↑](#footnote-ref-8)
9. CNPJ: 07.058.057/0001-90 [↑](#footnote-ref-9)
10. CNPJ: 20.216.977/0001-58 [↑](#footnote-ref-10)
11. CNPJ: 27.966.051/0001-29 [↑](#footnote-ref-11)
12. CNPJ: 21.261.525/0001-50 [↑](#footnote-ref-12)
13. CNPJ: 27.202.469/0001-60 [↑](#footnote-ref-13)