Extreme and Sustainable Graph Processing for Green Finance Investment and Trading

Laurentiu Alexandru Vasiliu
laurentiu.vasiliu@peracton.com
Peracton Ltd.
Galway, Ireland

Dumitru Roman
Dumitru.Roman@sintef.no
SINTEF AS
Oslo, Norway

Radu Prodan
radu.prodan@aau.at
University of Klagenfurt
Klagenfurt, Austria

ABSTRACT

In this paper we present a case addressing the drawbacks of financial market data, its limited volumes, history, and sometimes the incomplete and erroneous datasets with variable quality, limited availability, and price barriers. The case aims to enable fast, semi-automated creation of realistic and affordable synthetic (extreme) financial datasets, unlimited in size and accessibility, ready to be commercialized. Peracton Ltd. intends to apply the resulting extreme financial data multiverse for testing and improving artificial intelligence (AI)-enhanced financial algorithms (e.g., using machine learning) focused on green investment and trading. Using synthetic data for testing financial algorithms removes critical biases, such as prior knowledge, overfitting, and indirect contamination due to real-world data scarcity, and ensures data completeness at an affordable cost. The availability of extreme (volumes) of synthetic data will consolidate further financial algorithms and provide a statistically relevant sample size for advanced back-testing.

CCS CONCEPTS

• **Software and its engineering → Software performance**; • **Social and professional topics → Computing and business**; • **Applied computing → Mathematics and statistics**; **Economics**.

KEYWORDS

Green Finance, Investment and Trading, Graph Massivizer, Graph processing, Finance Data Generation

ACM Reference Format:
Laurentiu Alexandru Vasiliu, Dumitru Roman, and Radu Prodan. 2023. Extreme and Sustainable Graph Processing for Green Finance Investment and Trading. In *Companion of the 2023 ACM/SPEC International Conference on Performance Engineering (ICPE '23 Companion), April 15–19, 2023, Coimbra, Portugal.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3578245.3585337

1 INTRODUCTION

The wide use, availability, accessible costs, interoperability, and analytical exploitation of financial data are essential for the EU data strategy. Knowledge graphs are crucial to innovation, competition, and prosperity and establish a strategic investment in technical processing and ecosystem enablers. Graphs are universal abstractions that capture, combine, model, analyze, and process knowledge about real and digital worlds into actionable insights through item representation and interconnectedness.

The improvement and optimization of green investments and trading face significant barriers. Historical securities' data, is not enough for in-depth, massive volume testing, de-risking financial algorithms, and training AI models due to erroneous, scattered, unavailable, proprietary, incomplete, un-audited or expensive data. The ever-growing complexity of AI-driven financial algorithms further amplifies the data scarcity for advanced training. The Graph-Massivizer project [1] aims, among others, to remove the limitations of current financial market data (limited volume, reduced accessibility, price barriers) by enabling fast, semi-automated creation of realistic and affordable synthetic extreme[1] financial datasets, unlimited in size and accessibility.

2 GRAPH MASSIVIZER

The Graph-Massivizer project researches and develops a high-performance, scalable, gender-neutral, secure, and sustainable platform based on the massive graph (MG) representation of extreme financial data in the form of general, knowledge, and property graphs. Graph-Massivizer supports the 'any-volume' graph challenge by processing up to billions of vertices and trillions of edges. It tackles the 'velocity graph' challenge of dynamically changing topologies and proposes a novel 'viridescence' graph challenge for sustainable processing at scale. Graph Massivizer's support for extreme financial data extends existing graph processing technological capabilities by orders of magnitude for at least one "v"-characteristic. The project delivers the Graph-Massivizer toolkit of five open-source software tools and FAIR graph datasets covering the sustainable lifecycle of processing extreme data as MG, displayed in Figure 1 and detailed below.

Data ingestion and graph creation translates extreme data from various static and event streams or follows heuristics to generate synthetic data, persist it, or publish it within a graph structure.

Graph enrichment, graph query, and graph analytics are three 'basic graph operations (BGO)' that analyze and expand extreme datasets using probabilistic reasoning and ML algorithms for graph pattern discovery, low memory footprint graph generation, and low latency error-bounded query response. The output of this phase is a new graph, a query, or an enriched structured dataset.

Graph workload modeling and optimization analyses and expresses the given graph processing workload into a workflow of

[1]By 'extreme data' we mean data volumes (processed or produced) from 1PB to 100EB and beyond.

ICPE '23 Companion, April 15–19, 2023, Coimbra, Portugal

ACM ISBN 979-8-4007-0072-9/23/04.
https://doi.org/10.1145/3578245.3585337

BGO. It further combines parametric BGO performance and energy models with hardware models to generate accurate performance and energy consumption predictions for the workload running on a given multi-node, heterogeneous infrastructure of CPUs, GPUs, and FPGAs. The predictions indicate the most promising combinations of BGO optimizations and infrastructure. Then, synthetic data will be created by instantiating the previously created graphs to generate unique financial extreme data sets.

Sustainability analysis collects, studies, and archives performance and sustainability data from operational data centers and national energy suppliers on a large scale. This phase simulates multi-objective infrastructure sustainability profiles for operating graph analytics workloads, trading off performance and energy (e.g., consumption, CO2, methane, and greenhouse gas (GHG) emissions) metrics. Its ultimate purpose is to model the impact of specific graph analytics workloads on the environment for evidence-based decision-making.

Serverless BGO processing uses performance and sustainability models and data from the previous phases to deploy serverless graph analytics on the computing continuum. It relies on novel scheduling heuristics, infrastructure partitioning, and environment-aware processing for scalable orchestration of serverless graph analytics with an accountable performance and energy consumption tradeoff.

3 RELATED SOLUTIONS ON THE MARKET

Among the solution on the market we briefly discuss the following four alternatives.

1. synthetix.io is a decentralized finance (DeFi) platform that allows users to trade synthetic assets that track the value of real-world assets, such as currencies, commodities, and stocks. The platform uses a knowledge graph to create these synthetic assets by representing the relationship between various financial products and their underlying assets.

2. alethea.ai is a framework that uses knowledge graphs and AI to generate synthetic data for financial applications such as market volatility, credit risk and fraud detection.

3. intellegens.ai uses machine learning and knowledge graphs to create synthetic data for financial applications. The company's platform can generate synthetic data for various financial scenarios, such as portfolio optimization and risk analysis.

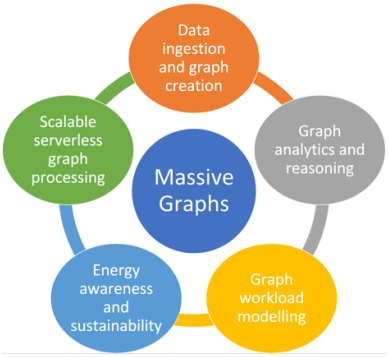

Figure 1: Graph Massivizer Software Tools

4. Datagenic.com uses a knowledge graph-based approach to create synthetic commodity price data for trading and risk management applications.

4 GREEN FINANCE USE CASE

Green finance targets various financial products and services that direct investments into green-oriented enterprises. It aims achieving economic growth and increased efficiency while reducing waste, pollution, and GHG emissions.

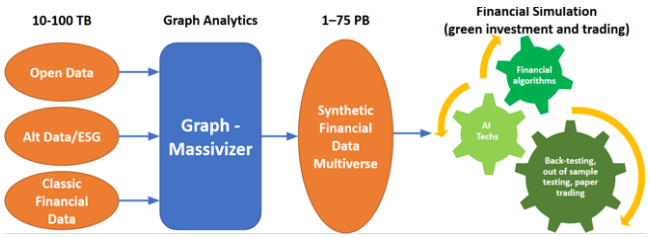

Figure 2: Green Finance Application

The Green Finance application, presented in Figure 2, aims to use the financial data multiverse for improved green AI-enhanced financial algorithms with reduced bias, risk, and higher performance. It aims to increase the investment return by a realistic 2-4% using mature and continuously-optimized financial algorithms. It further targets an increase in excess return (alpha) by 1-2% with a quick ratio higher 1.5, reflecting healthy investments with lower risk and higher returns. Graph-Massivizer extreme synthetic data will offer a technological advantage, supporting green investments, green algorithms, and further de-risking. By exposing and training the algorithms to 'never met before' data, the use of synthetic data will increase their resilience, performance and fitness and so further increase the trust in them and de-risk them further.

5 SUMMARY AND OUTLOOK

Graph-Massivizer will allow green financial investors to avail a massively increased financial data testing universe (synthetic multiverse) and get a competitive advantage on graph-based powerful analytics, with evidence of improved performance and sustainability. While other forms of analysis rely on assumptions about "what happened" or "what happens", correctly building and employing financial graphs to generate synthetic data for dedicated financial simulation environments will further consolidate financial models by exposing them to completely unknown situations.

ACKNOWLEDGMENTS

This work is partly funded through the projects Graph-Massivizer (HE 101093202), enRichMyData (HE 101070284), and DataCloud (H2020 101016835).

REFERENCES

[1] Radu Prodan, Dragi Kimovski, Andrea Bartolini, Michael Cochez, Alexandru Iosup, Evgeny Kharlamov, Jože Rožanec, Laurențiu Vasiliu, and Ana Lucia Vărbănescu. 2022. Towards Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe. In *2022 IEEE Cloud Summit*. 23–30. https://doi.org/10.1109/CloudSummit54781.2022.00010