



Graph-Massivizer

Radu Prodan

University of Klagenfurt, Austria



This project has received funding from the European Union's Horizon Research and Innovation Actions under Grant Agreement № 101093202.

Graph-Massivizer

- Extreme and Sustainable Graph Processing for Urgent Societal Challenges in Europe
- HORIZON-CL4-2022-DATA-01-05
 - Extreme data mining, aggregation and analytics technologies and solutions (RIA)
- € 4,998,062.50
- 2023 – 2025



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

UNIVERSITY
OF TWENTE.



CINECA



metaphactory



Peracton



EVENTREGISTRY



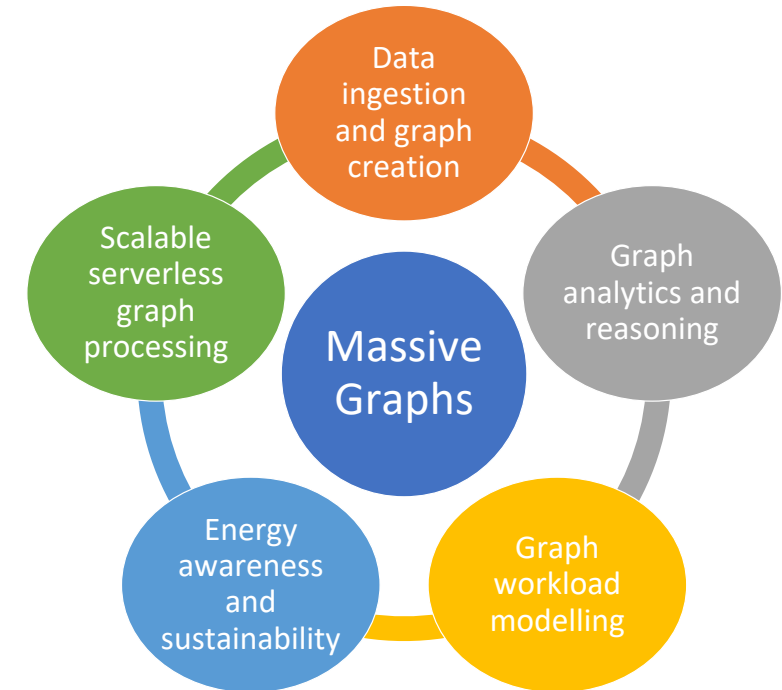
Goal

- High-performance and sustainable graph processing of extreme data
 - Holistic platform and integrated toolkit
 - Proper response for any need and any organisation by 2030
- Graphs
 - Universal mathematical abstractions
 - Capture, combine, model, analyse, and process knowledge about the real and digital worlds



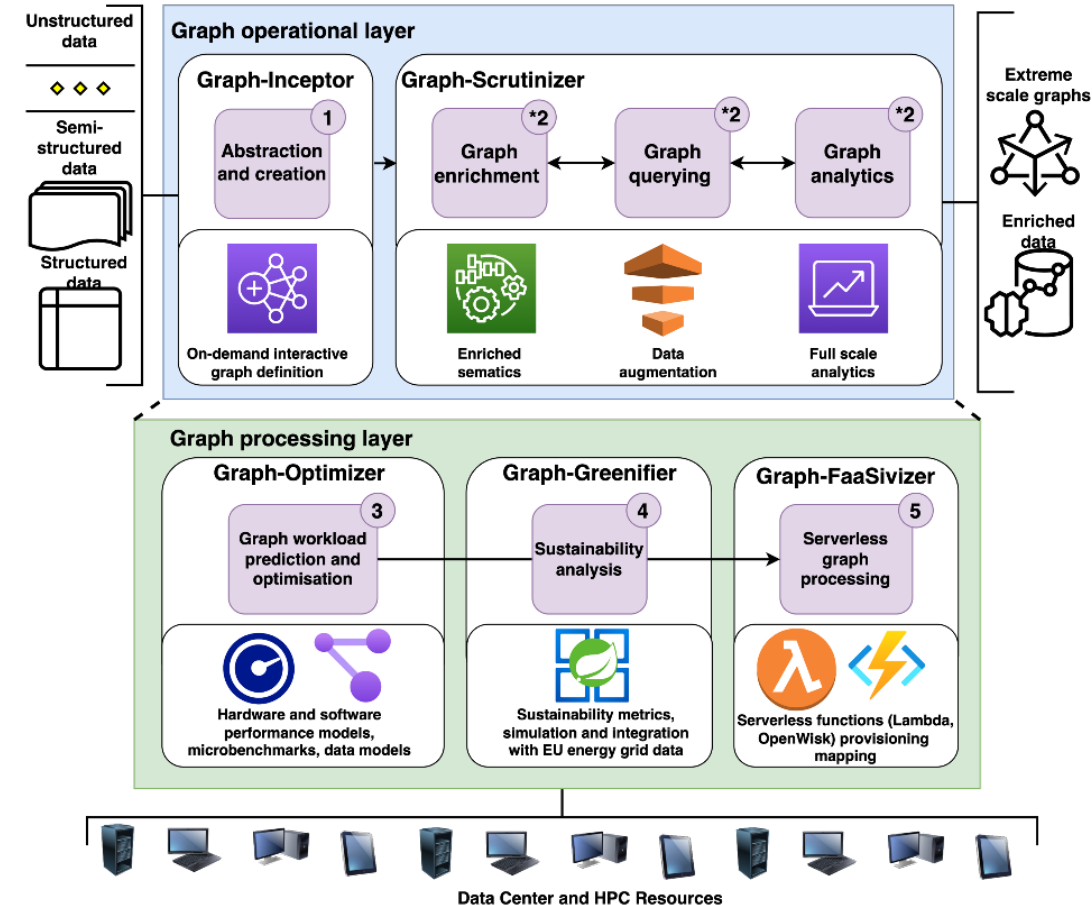
Massive Graph

- **Massive graph (MG)** representation of **extreme data**, integrating patterns and storing interlinked descriptions of objects, events, situations, concepts, and semantics
 - General graphs, knowledge graphs, property graphs
- **Extreme data challenge**, extending graph processing technology by orders of magnitude for several **“V”**-characteristics
 - **Volume** graph challenge by supporting up to billions of vertices and trillions of edges
 - **Velocity** graph challenge of dynamically changing topologies
 - **Viridescence** graph challenge for sustainable processing at scale



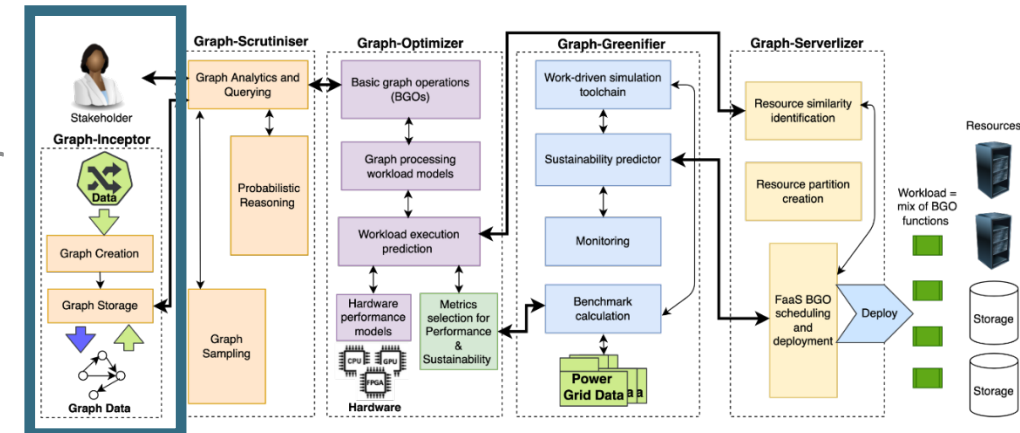
Graph-Massivizer Toolkit

- Graph operational layer
 - Graph-Inceptor: extreme data ingestion, MG creation and storage
 - Graph-Scrutinizer: MG analytic and reasoning
- Graph processing layer
 - Graph-Optimizer: workload modelling with performance and energy guarantees
 - Graph-Greenifier: sustainable and energy-aware MG processing
 - Graph-Choreographer: scalable serverless MG analytics over a codesigned computing continuum



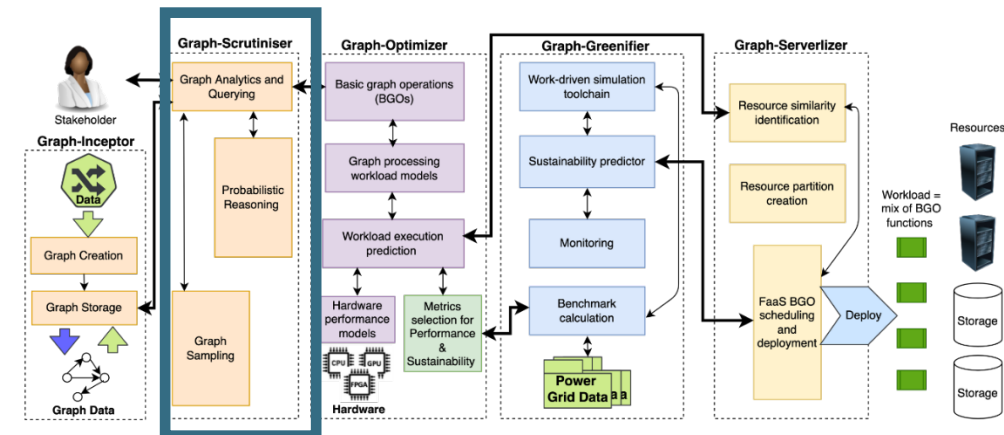
Graph-Inceptor

- Graph creation
 - ETL procedures to persist extreme data into MG of 10 billion vertices and 100 billion edges
 - Batch and stream data processing, compatible with Apache Spark
 - Streaming ingestion latency below 500 milliseconds for 95% of the data at 1,000s of new edges per second
- Graph storage
 - RDF graphs modelling semantic aspects
 - Property graphs modelled with edge properties
 - RDF-star graphs logically unifying the two views
 - Distributed for computation locality



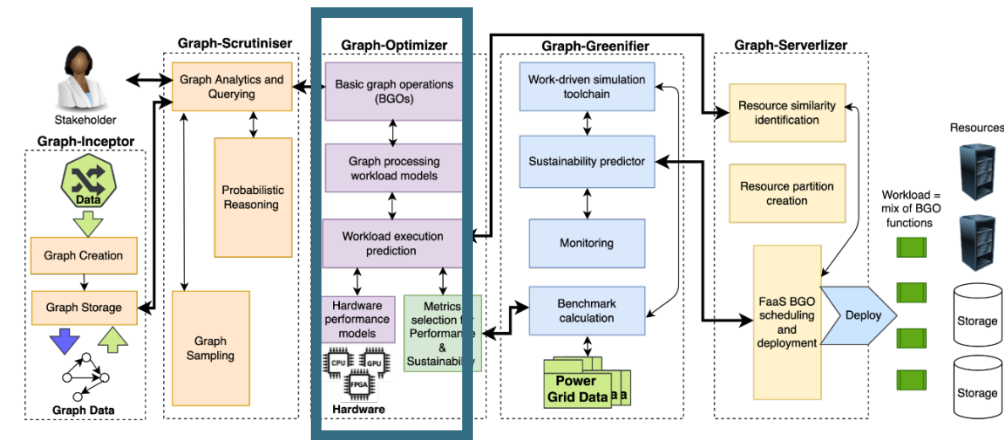
Graph-Scrutinizer

- Graphs analytics and querying
 - Sampling and analytics throughput of 500 giga edges per second (Apache GraphX)
 - Streaming-graph reasoning speed of 3 million triples per second (PyTorch Geometric)
 - Retain 90% performance for the last 10 million edges compared to batch algorithms
- Graph sampling
 - Optimised analysis and query, retaining critical information
- Graph enrichment with expanding data sets
 - Probabilistic reasoning
 - Graph discovery, low footprint graph generation, real-time, error-bounded queries



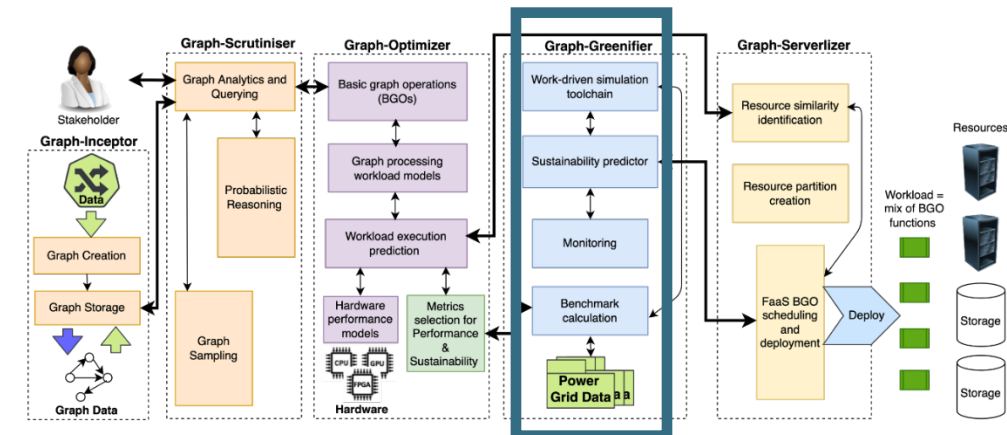
Graph-Optimizer

- Hardware operation (H-op) models relevant to BGO
 - Runtime performance and energy costs (CPU, GPU, FPGA)
 - Memory latency, bandwidth, BGO throughput, I/O bandwidth
- Graph processing and workload models
 - Model BGO as compositions of H-ops (Graph-BLAS, GBTL-CUDA)
 - Generate workload models using design space exploration
- Workload execution prediction
 - 80% accurate performance and energy consumption models for BGOs
 - Various heterogeneous multiprocessors, GPUs and FPGAs



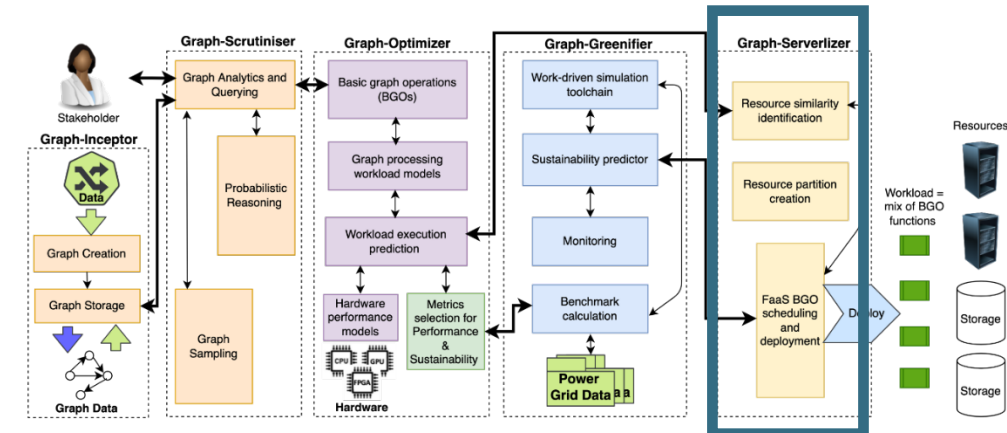
Graph-Greenifier

- Sustainability data collection
 - Monitoring infrastructure (hardware and software)
 - Power grid data interface (open market data)
- Work-driven graph processing simulation toolchain
 - Carbon footprint, CO₂ and methane emissions
 - Open DC and SimLess simulators
- Benchmarking, prediction and recommendations for sustainability-performance tradeoffs at scale
 - Sustainability labelling for graph analytics of 100 individual nodes and devices
 - Two-fold improvement in data centre energy efficiency and over 25% lower GHG emissions



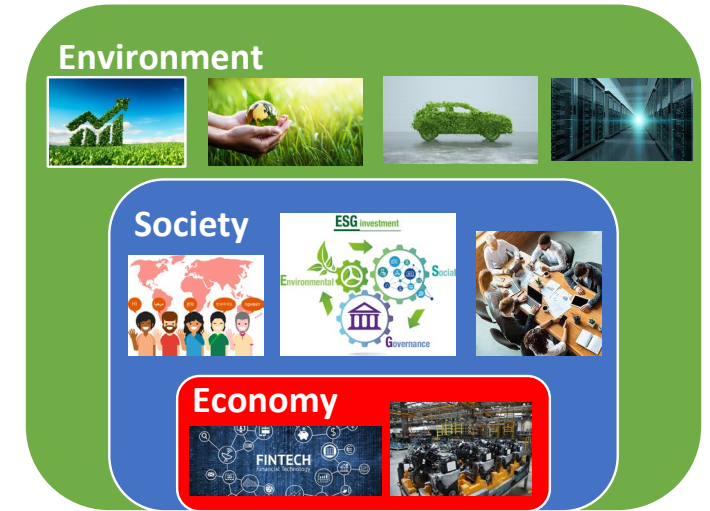
Graph-Choreographer

- Workload-infrastructure enactment
 - Multilayer infrastructure facilitation
 - Resource partitioning based on similarity relationships (HPC, cloud, edge)
- Sustainable BGO function operation and scheduling
 - OpenWhisk, Lambda, ...
 - 70% faster graph analytics over commercial AliGraph or other OSS solutions
 - 30% less energy for ETL storage than Amazon Redshift and OSS
- Cybersecure deployment
 - 40% faster runtime deployment compared to AWS CloudFormation
 - State-of-the-art security and privacy mechanisms



Use Cases

- UC-1: Green and sustainable finance
- UC-2: Global foresight for environment protection
- UC-3: Green AI for sustainable automotive industry
- UC-4: Data centre digital twin for sustainable exascale computing



	Volume	Velocity	Value	Veracity	Variety	Viscosity	Viridescence
UC-1	✓		✓	✓			✓
UC-2	✓				✓	✓	✓
UC-3		✓		✓	✓		✓
UC-4	✓	✓			✓		✓

LinkedIn: <https://www.linkedin.com/company/graph-massivizer-project/>

Twitter: @graphmassivizer

Thank you !



www.graph-massivizer.eu