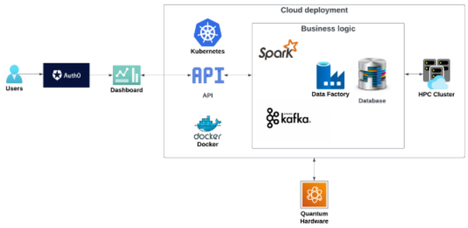
Technical Specification Double-side Page

1. **TECHNICAL SCOPE:** Summarize the solution developed during the EXPERIMENT phase: how have you finally addressed the challenge/Theme Challenges and tackled with its requirements and data. Include a diagram. Mention the different shared industrial and open data used.

QUANTUM-ASSISTED SYNTHETIC DATA FOR PROXY RECONSTRUCTION

The Proxy Reconstructor represents an innovative solution to a critical challenge encountered by financial institutions when calculating regulatory capital for assets that lack sufficient historical data. In the absence of this data, the inclusion of assets in portfolio investment decisions becomes unauthorised by regulators (ECB). To tackle this issue, proxies are utilized as substitutes or surrogates to estimate the behaviour of the target variable based on the observed behaviour of substitute variables. To do so, the solution generates traceable and explainable synthetic historical data by examining the correlations and the volatility ratio between the asset of interest (target) and other assets with extensive historical data. Those data series enable the computation of virtual risk metrics (VaR & SVaR) for regulatory capital, thereby allowing financial institutions to make legal and well-informed investments.



Quantum MADS addresses this challenge by integrating Quantum Computing, High-Performance Computing (HPC), and Machine Learning (ML). This groundbreaking approach confers a significant advantage, enabling financial institutions to encompass a vast universe of assets, enhancing the accuracy of data series reconstructions. In our case, this involves handling a dataset comprising more than 10,000 assets. Specifically, the DP (top-tier bank) has shared detailed P&L information for each asset dating back to 2007 or its IPO date. As a result, the assets under analysis, "targets" derive unique added value from three key perspectives: gaining insights into the behaviour of financial instruments; obtaining information on their associated risk metrics; and securing regulatory approval for investments in these assets.

2. **ALGORITHMS, TOOLS AND CONCLUSIONS:** Detail the algorithms and tools finally selected to accomplish the challenge/Theme Challenges. Summarize the main results that you have obtained during the EXPERIMENT phase: data, insights, conclusions and the main contributions to solve the challenge/Theme Challenges.

The solution involves two main goals: selecting relevant assets and applying pattern recognition techniques to extrapolate relationships. Our solution integrates High-Performance Computing (HPC), Artificial Intelligence (AI), and Quantum Computing techniques and can be divided into three primary processes.

1- Asset Selection: Initially, we identify the most suitable assets for target asset reconstruction by defining a similarity metric, considering volatility ratios and correlations, with HPC capabilities. This metric guides the Quadratic Programming Feature Selection (QPFS) algorithm, optimizing selection by maximizing relevance and minimizing redundancy. Working with a universe of around 10,000 diverse assets, our goal is to select a small subset for each target. To achieve this, we have developed a Quantum Unconstrained Binary Optimization (QUBO) formulation, optimized through quantum annealing based on the QPFS formulation. This quantum optimization process yields the optimal subset of size K as specified by the user.

2- Feature Engineering: After the asset subset is determined, we proceed to define features that characterize each day. These features enable the detection of patterns between the selected features and the target asset. We consider temporal characteristics such as lags, moving averages, and exponential moving averages. Lag values represent returns in previous days, while a moving average is the average of preceding days, and an exponential moving average is a weighted average that assigns greater weight to more recent days.

3- Model Training and Prediction: With established features, our goal is to utilize models for identifying patterns in new, unseen data. In this project, for traceability, we focus on linear regression models like Linear, Ridge, Lasso, and Linear Support Vector. These models share the purpose of being trained on features and target data within a defined time interval (training set). We adjust weights in a linear combination to minimize a specified loss function. These learned weights are then applied to predict the target asset's value in a new period (the test set). Most of these models have hyperparameters requiring selection, achieved through a grid search combined with cross-validation, where 80% of recent data is for training and 20% for validation. Additionally, users can specify the training interval by specifying dates.

3. **SCALABILITY AND FLEXIBILITY OF THE SOLUTION:** Explain how the solution copes with the challenge/Theme Challenges requirements and how can it be adapted to other similar problems. What work is still pending to create a real/stable product if any? What TRL level is it in? Is it a DVC?

Quantum Mads offers an end-to-end SaaS solution with a versatile platform for customer request processing, dashboard insights, and CSV file downloads. The backend supports replicable processes for diverse financial use cases, emphasizing microservices in a scalable framework. DP's data seamlessly flows to the API, connecting to HPC resources for real-time data generation using quantum-based algorithms and ML. To ensure scalability and performance, microservices and auto-scaling via a load balancer are employed. The solution also provides robust data visualization for complex data analysis and pattern discovery.

Adhering to Data Value Chains principles, all process outputs follow standardized formats, ensuring secure data exchange among four stakeholders: the financial institution (DP), software provider, cloud & HPC provider, and quantum hardware provider. The product is now at TRL 6-7, having been tested in relevant environments. To complete the pilot phase, minor adjustments are needed, including refining visualizations, conducting extra testing for future stability, and completing process automation to eliminate human involvement

4. **DATA GOVERNANCE AND LEGAL COMPLIANCE:** Describe the security level of the solution, i.e. how authentication, authorization policies, encryption or other approaches are used to keep data secure. Explain how the solution is compliant with the current data legislations concerning security and privacy (e.g. GDPR). Describe in a convincing way how your solution realises a secure DVC, e.g. through usage of specific tools.

The proposed solution prioritizes data security through a robust set of measures that are very consolidated in the market. These measures mainly include: For user management and access to platform content, Auth0 is used, a CSA STAR certified service that operates under ISO 27001/27018, which ensure privacy protection in the cloud. With regard to data access to the business API, JWT and the OpenAPI standard are used. Therefore, JWT defines a compact and self-contained way for securely transmitting information between parties as a JSON object and OpenAPI restricts access to the source code. Besides, the hosting of business services and data storage is made of Microsoft Azure, certified by ISO/IEC 2700. In the case of third party's quantum machines that will be executing the algorithm, verification of compliance and security was done through assessments, and certifications included in the contract.

5. **QUALITY ASSURANCE AND RISK MANAGEMENT:** Describe the quality process followed for the final product. Technologically, which problems have you encountered and how you have solved them, and any processes followed that guarantee that the solution fulfills the challenge/Theme Challenges and data provider requirements.

The quality process ensures the product meets standards and performs effectively by optimizing data handling, improving efficiency, enhancing reliability, ensuring user-friendliness, and conducting rigorous testing. Weekly meetings with DP and rigorous software testing were integral to delivering a dependable, efficient solution meeting DP's needs. During development, we faced notable technological challenges. Initially, handling extensive data posed issues. Our asset selection strategy required computing correlations between all asset pairs. For context, with 10,000 assets, this entailed 10^8 correlations, presenting three primary constraints:

-Memory Size for the Correlation Matrix: The size of the correlation matrix corresponding to these computations was substantial and exceeded the capacity of standard laptops' memory. Consequently, we adopted the utilization of cloud-based High-Performance Computing (HPC) resources to accommodate this memory-intensive task.

-Memory Size for the Correlation Matrix: The size of the correlation matrix corresponding to these computations was substantial and exceeded the capacity of standard laptops' memory. Consequently, we adopted the utilization of cloud-based High-Performance Computing (HPC) resources to accommodate this memory-intensive task.

-Disk Storage Considerations: Using the standard CSV format for saving and loading this extensive matrix proved to be time-consuming. Thus, alternative data storage formats had to be explored to optimize efficiency.

-Computation Effort for Correlations: The computational challenge of calculating correlations required parallelization, even with multiple CPUs. To enhance efficiency, we segmented time series data into 90-day bins, effectively parallelizing correlation calculations between these bins, utilizing libraries like NumPy for inherent parallelization capabilities.

Moreover, our asset selection algorithm, based on Quadratic Programming Feature Selection (QPFS), requires solving a Quadratic Unconstrained Binary Optimization (QUBO) problem. However, the number of coefficients in this QUBO problem scales quadratically with asset count, creating a significant computational burden. In our scenario with 10,000 assets, defining the QUBO became excessively time-consuming. To address this, we implemented a solution by storing the common quadratic terms for all assets in the QUBO object. This stored QUBO object could then be loaded and combined with the individual linear terms for each asset selection, optimizing the process.

Annex 1. Means for accessing the MVP

Please, indicate in 1 page maximum the means for accessing the MVP for a potential customer (login information, website address, link to a demo video or whatever means are needed to check that the MVP exists and works).

In order to access to the MVP, the user only needs an internet browser for login in https://qmads-dashboard-6a0fd.web.app/strategy/algorithm-form with:

User: quantum-reach@quantum-mads.com
Pass: QMReach2023!!####

To facilitate understanding of the operation of the platform, a video showing the tour of it is attached below. It can be accessed from the following link (Google Drive Portfolio Folder):

https://drive.google.com/drive/folders/1OA6xH2MfaQ4rDRmb3brhbMy3WAZ9ppg9