Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach

José Antonio Parejo and Antonio Ruiz-Cortés

SCORE Lab, I3US, Universidad de Sevilla, Seville, Spain
{japarejo,aruiz}@us.es

Abstract. This paper presents a hybrid quantum–classical services architecture for fleet management and vehicle routing in the NISQ era. We first introduce a reference architecture with 9 services that separates data acquisition, operational orchestration, and solving capabilities through classical and quantum routing services. We then formalize a baseline Multi-Depot Capacitated Vehicle Routing Problem (MD-CVRP) on a directed graph with explicit depot assignment, customer service uniqueness, flow conservation, capacity feasibility, and load propagation constraints. The paper also reviews state-of-the-practice approaches for solving MD-CVRP, from classical (exact and metaheuristic) methods to quantum-assisted QUBO/Ising workflows (QA and QAOA), and discusses their current practical limitations, including encoding growth, penalty calibration, feasibility repair, and quantum-hardware requirements. Based on this analysis, we outline a pragmatic research path on this problem: use strong classical baselines, apply quantum routines only to well-scoped subproblems, and report end-to-end performance with full accounting.

Keywords: Hybrid quantum–classical optimization · fleet management · MD-CVRP · service-oriented architecture

1 Introduction

Integrating quantum services into software compositions is a central challenge in Quantum Software Engineering [10]. Quantum computing may provide advantage for some problems once input size crosses specific thresholds [1]. However, we are still in the Noisy Intermediate-Scale Quantum (NISQ) era [11]: limited qubit counts, noise, and repeated shots often keep classical implementations competitive, especially for small or structurally simple instances.

This motivates hybrid applications in which classical and quantum services coexist. In such settings, runtime binding should consider not only non-functional properties (e.g., cost or latency) but also input characteristics. To the best of our knowledge, this input-dependent perspective is still under-represented in dynamic service binding proposals.

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

In this paper, we use the Hybrid Fleet Management Architecture, motivated by the ATHENA Project [12], as a case study. Route planning with multiple trucks, delivery points, and depots is a complex NP-complete variant of the Vehicle Routing Problem (VRP) named MD-CVRP. We make three contributions: (i) a reference architecture with 9 for hybrid quantum–classical services (ii) a MD-CVRP formalization with practical constraints for real distribution contexts (supermarkets, urban logistics, and last-mile delivery), and (iii) a review of current MD-CVRP solution alternatives (classical and quantum-assisted), with an analysis of their practical limitations. In this architecture, one service solves NP-hard VRP instances, motivating selective use of quantum resources as instance complexity increases.

2 The Fleet Management Services Architecture

The Fleet Management System (FMS) is designed as a cloud-native suite of 9 services. To ensure scalability and support hybrid execution, the architecture decouples logistics data management from the computationally intensive optimization logic. The system is built to support the following responsibilities:

1. **Master data management:** depots, customers, service locations, product catalogue and demand, vehicle profiles, driver profiles.
2. **Planning and optimization:** generation of feasible and cost-effective routes across depots and delivery points.
3. **Execution and monitoring:** dispatch management, vehicle telemetry, estimated time of arrival (ETA) monitoring, exception handling, and proof-of-delivery.
4. **Governance and observability:** authentication/authorization, auditing, compliance, and system telemetry.

2.1 Functional Layers and Service Inventory

We focus on domain-specific capabilities required to collect the data to formulate and solve MD-CVRP instances, omitting services such as Identity & Access control, customer data management, employee profile data, etc. Under this consideration we identify the following functional layers and services (see Fig. 1):

Data & Asset Layer: This foundational tier acts as the *System of Record* for the logistics domain. It manages the lifecycle of physical and digital assets, ensuring that the structural constraints of the problem—such as vehicle volumes, driver availability, and geographical distances—are consistently maintained and served to the upper layers.

- *Depot Service (D):* Manages depots and operational constraints (loading bays, opening hours, cut-off times, etc.).
- *Order & Demand Service (O:)* Transforms commercial orders into logistics demand units (capacity requirements and delivery points).

Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

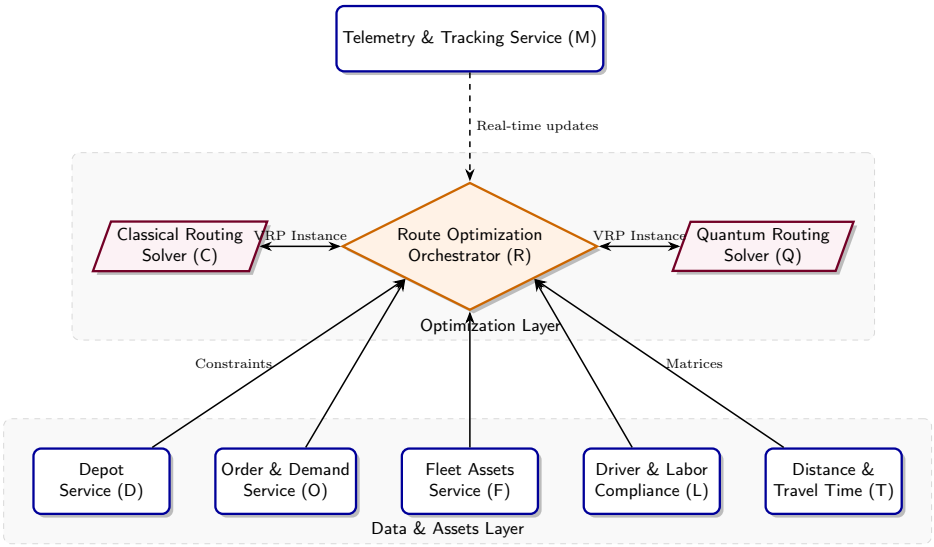

Fig. 1: Architecture of the Fleet Management System (FMS).

- *Fleet Assets Service (F):* Manages vehicle profiles and capabilities (capacity, compartments, refrigeration, costs), maintenance status, and availability windows.
- *Driver & Labor Compliance Service (L):* Provides driver availability and validates route-assignment feasibility according to shifts, working-time regulation constraints, minimum rest requirements, etc.
- *Distance Matrix & Travel Time Service (T):* Provides travel distances and dynamic travel times among depots and delivery points depending on the time and day according to traffic patterns.

Optimization & Governance Layer: This tier represents the *System of Intelligence* where business requirements are transformed into mathematical workloads. It is termed 'Governance' because the Orchestrator (R) performs policy-based dispatching, evaluating instance complexity and hardware reliability (via i-contracts [9]) to decide the most cost-effective execution path between classical and quantum solvers.

- *Route Optimization Orchestrator Service (R):* Coordinates the route planning jobs: collecting inputs (demand, depots, fleet, drivers, travel-time matrix), validating feasibility, selecting the appropriate quantum or classical routing solver, and invoking it.
- *Classical Routing Solver Service (C):* Solves the VRP using metaheuristics or MILP backends running over a classical computing platform.
- *Quantum Routing Solver Service (Q):* A quantum implementation (QA/QAOA) for well-scoped subproblems or experimental QUBO/Ising mapping.

Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

This work is licensed under a Creative Commons "Attribution-NonCommercial-NoDerivatives 4.0 International" license.

Execution Layer: This layer bridges the gap between the planned optimization and the physical operation. By ingesting real-time telemetry, it provides the feedback loop necessary for observability, allowing the system to monitor SLA compliance and trigger re-optimization if the operational reality deviates from the calculated plan.

– *Telemetry & Tracking Service (M):* Ingests IoT data (GPS, sensor data) and collects proof-of-delivery after each service.

2.2 Operational Workflow: The 24h Planning Scenario

To illustrate the interaction between services and the role of the orchestrator as a smart gateway, we define the execution flow of a standard 24-hour planning job through four sequential phases:

1. **Context Aggregation:** (R) initiates the process by fetching the "logistics world state": pending demands from (O), vehicle capacities (Q_k) from (F), and depot windows from (D), while filtering driver availability via (L).
2. **Spatial-Temporal Mapping:** (R) invokes (T) to generate the dynamic travel-time matrix (c_{uv}) for the target window, considering predicted traffic patterns.
3. **Hybrid Governance & Solving:** (R) formalizes the MD-CVRP instance and assesses its structural complexity. It chooses and invokes a suitable solving implementation such as (C) or (Q) for solving the problem instance, also possibly using other hybrid approaches, such as classical algorithms delegating sub-problems to quantum implementations as discussed in Section 4. This is governed by "i-contracts" [9] that weigh computational cost against the potential for higher-quality optima.
4. **Validation & Dispatch:** Once a solution is returned, (R) validates it against compliance rules (L) and persists the routes. During execution, (M) monitors real-time GPS data, allowing (R) to detect deviations and trigger re-optimization if necessary.

3 The Multi-Depot Capacitated Vehicle Routing Problem (MD-CVRP)

The operational intelligence of the Fleet Management System (FMS) presented in the previous section relies on a formal optimization model. The *Route Optimization Orchestrator (R)* transforms the disparate data streams from the Data Layer into a *Multi-Depot Capacitated Vehicle Routing Problem (MD-CVRP)*. This model generalizes the classical VRP by allowing multiple starting locations and requiring the simultaneous optimization of customer-to-depot assignment and route sequencing.

In this architectural context, the MD-CVRP is not merely a theoretical abstraction, but the actual "job contract" that (R) dispatches to the solvers.

Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

The *Vehicle Routing Problem (VRP)* is a family of combinatorial optimization problems that generalizes the Traveling Salesman Problem (TSP) to realistic distribution settings. The VRP models the operational decisions required to distribute goods or services from one or several depots to a geographically dispersed set of customers using a fleet of vehicles.

In industrial contexts (such as retail distribution, urban logistics, or last-mile delivery), the VRP captures spatial, temporal, and operational constraints that are absent in the classical TSP. The VRP simultaneously combines assignment complexity (NP-hard), sequencing complexity (NP-hard), and scheduling feasibility (NP-hard in many variants). Consequently, large industrial instances are typically sub-optimally solved using hybrid metaheuristic and decomposition-based optimization methods rather than exact approaches.

The VRP involves a fleet $K = \{1, \ldots, m\}$. Each vehicle operates its own route, and collectively the fleet must service all customers.

For managers, this introduces route partitioning decisions, fleet sizing decisions, and load balancing considerations. In the MD-CVRP we focus on route partitioning decisions under the assumption of a fixed fleet size and trucks with limited capacity.

Thus, the problem is not only about sequencing nodes but also about assigning customers/deliveries to vehicles, taking into account that each vehicle k has maximum capacity Q_k. Each customer i has demand q_i. For route R_k assigned to vehicle k, feasibility requires that $\sum_{i \in R_k} q_i \leq Q_k$.

Capacity constraints transform the problem into a hybrid between routing and bin-packing: customers must be partitioned into feasible loads, and each partition must be optimally sequenced. Large distribution systems may involve several depots $D = \{d_1, d_2, \ldots, d_p\}$. Each vehicle is associated with one depot, and decisions include customer-to-depot assignment, depot workload balancing, and cross-docking considerations.

Additionally, in many real-world applications, each customer i must be served within a time interval $[a_i, b_i]$. If t_i denotes arrival time at customer i, then $a_i \leq t_i \leq b_i$. Time windows introduce scheduling feasibility constraints, waiting time modeling, and strong coupling between routing and temporal decisions.

Operational feasibility may require maximum route duration $T_k \leq T^{\max}$, where T_k denotes total route time of vehicle k. Route duration T_k typically includes, for each delivery: Travel time, Service time, and Waiting time. Such constraints reflect legal regulations (e.g., driving time limits) and operational policies. Additionally, per-driver working hour restrictions could also appear as constraints in the problem formalization.

In the formal model below, we focus on a baseline MD-CVRP with routing and capacity constraints; fixed vehicle costs, time windows, and route-duration constraints are omitted for the sake of simplicity and clarity.

Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

3.1 Problem Formalization

The problem is defined on a directed graph $G = (V, A)$. Below we detail the mathematical components and their provenance from the architecture described in Section 2:

- **Nodes (V):** The set of all locations $V = D \cup C$, where D is the set of depots provided by the *Depot Service (D)* and $C = \{1, \ldots, n\}$ is the set of customers/delivery points provided by the *Order Service (O)*.
- **Arcs and Costs (A, c_{uv}):** The set of directed arcs $A = \{(u, v) \in V \times V : u \neq v\}$. Each arc has a cost c_{uv} representing the travel time or distance, dynamically calculated by the *Distance Service (T)*.
- **Demands (q_i):** Each customer $i \in C$ has a deterministic demand $q_i \geq 0$, managed by the *Order Service (O)*.
- **Fleet (K, Q_k):** The set of vehicles $K = \{1, \ldots, m\}$ and their respective capacities Q_k are provided by the *Fleet Assets Service (F)*.
- **Depot Assignment (δ_{kd}):** A binary parameter provided by the *Depot Service (D)*, where $\delta_{kd} = 1$ if vehicle k is based at depot d.

3.2 Decision Variables

To formulate the optimization job, the Orchestrator defines three sets of variables:

- $x_{uvk} \in \{0, 1\}$: Equals 1 if vehicle k traverses arc (u, v).
- $y_{ik} \in \{0, 1\}$: Equals 1 if customer i is served by vehicle k.
- $u_{ik} \in [0, Q_k]$: Represents the cumulative load of vehicle k after serving customer i (used for subtour elimination).

3.3 Mathematical Model

The objective is to minimize the total routing cost (e.g., total time or distance):

$$\min \sum_{k \in K} \sum_{(u,v) \in A} c_{uv}\, x_{uvk}$$

Subject to the following operational constraints:

1. Coverage and Flow Conservation: Each customer must be served exactly once, and any vehicle entering a node must leave it:

$$\sum_{k \in K} y_{ik} = 1, \forall i \in C; \quad \sum_{(i,v) \in A} x_{ivk} = \sum_{(u,i) \in A} x_{uik} = y_{ik}, \forall i \in C, \forall k \in K$$

2. Depot Consistency: Vehicles must depart from and return to their assigned base depot as defined by (D):

$$\sum_{(d,v) \in A, v \in C} x_{dvk} = \sum_{(u,d) \in A, u \in C} x_{udk} = \delta_{kd}, \forall d \in D, \forall k \in K$$

Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

3. Capacity and Load Propagation: These constraints ensure that the vehicle capacity Q_k (from F) is never exceeded and prevent illegal subtours:

$$q_i y_{ik} \leq u_{ik} \leq Q_k y_{ik}, \forall i \in C, \forall k \in K$$

$$u_{jk} \geq u_{ik} + q_j - Q_k(1 - x_{ijk}), \forall k \in K, \forall i, j \in C, i \neq j$$

4. Domain Constraints: $x_{uvk}, y_{ik} \in \{0, 1\}$ and $x_{dd'k} = 0$ for $d \neq d'$, ensuring no inter-depot travel.

Remarks This baseline formulation can be extended to include fixed vehicle costs, time windows, heterogeneous fleets, or stochastic demand.

3.4 Relationship with the TSP

The TSP is a special case of the VRP obtained when $m = 1$, $Q_1 = \infty$, there are no time windows, we use one single depot, there are no fixed vehicle costs, and we have no duration limits.

Under these assumptions, the VRP reduces to:

$$\min \sum_{(u,v) \in A} c_{uv} x_{uv}$$

subject to standard subtour elimination constraints. Thus, TSP $\subset$ VRP.

In a retail distribution system, solving a VRP answers several questions: How many vehicles should be dispatched? Which customers does each vehicle serve? In what order? At what time does each service occur? Is the solution legally and operationally feasible? In contrast, the TSP answers only the optimal visiting order for a single route.

4 Approaches for solving the MD-CVRP

The *Route Optimization Orchestrator (R)* is responsible of choosing and invoking a specific problem solving implementation. In so doing, it must evaluate the problem instance against the advantages and drawbacks of the various solving paradigms available. This section analyzes the classical and quantum landscapes to justify the hybrid governance logic.

4.1 Classical baselines and state-of-the-practice

Any quantum(-assisted) proposal for MD-CVRP should be evaluated against strong classical baselines. For MD-CVRP, the main classical approaches fall into:

Exact and decomposition-based methods. State-of-the-art exact solvers for VRP families typically rely on *branch-and-cut(-and-price)* and *column generation* (set partitioning / route-based formulations), optionally with problem-specific cuts, stabilization, and advanced pricing. For multi-depot variants, exact methods remain feasible only for moderate instance sizes and require careful modeling of depot assignment and route feasibility.

Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

This work is licensed under a Creative Commons "Attribution-NonCommercial-NoDerivatives 4.0 International" license.

High-performance metaheuristics. For operational-scale instances, the dominant approach is hybrid metaheuristics such as: *Adaptive Large Neighborhood Search (ALNS)*, *Tabu Search*, and hybrid local search with ruin-and-recreate and intensification steps. These methods are robust, scalable, and provide high-quality solutions quickly; they are the appropriate reference point for "practical advantage" claims in logistics.

Why these baselines matter for quantum approaches. Even if a quantum subroutine accelerates a subproblem (e.g., TSP within clusters), the global MD-CVRP objective depends critically on the customer-to-vehicle/depot partition. Classical methods already optimize this coupling effectively; therefore, quantum value must be demonstrated either as (i) improved solution quality at fixed time, (ii) reduced solving time at fixed quality, or (iii) improved scaling on carefully controlled substructures (given current q-bit limitations).

4.2 Quantum(-assisted) formulations and alternatives

Most quantum approaches for VRP map the routing problem to a binary optimization model: *QUBO* (Quadratic Unconstrained Binary Optimization) or an *Ising Hamiltonian*, then apply either *quantum annealing* (QA) or gate-model variational algorithms (primarily QAOA).

QA and QUBO encodings. A standard strategy is to encode routing decisions into binary variables and incorporate constraints as penalty terms:

$$\min_{x\in\{0,1\}^n} \; f(x) + \sum_{\ell} \lambda_\ell \, g_\ell(x), \tag{1}$$

where $f(x)$ represents the transportation cost and each $g_\ell(x)$ penalizes violations (capacity, degree, depot start/end, subtour elimination proxies, etc.). This yields a QUBO (or Ising) objective suitable for QA hardware.

Dedicated QA formulations have been proposed for multi-depot capacitated VRP variants, including dynamic settings, by constructing QUBO models that represent depot choices and capacity-feasible routing implicitly via penalties [7]. Such encodings are conceptually straightforward but often lead to large penalty weights, dense couplings, and embedding challenges on annealing hardware.

Because full VRP QUBOs scale quickly, several works advocate hybrid decomposition: cluster/partition the problem classically, then apply QA to smaller subproblems, and recombine solutions. A representative example is the family of hybrid QA algorithms that partition VRP instances (e.g., via clustering) and solve sub-QUBOs on QPU, with classical coordination and post-processing [4].

Gate-based variational methods (QAOA). Gate-based approaches frequently map VRP variants into Ising/QUBO Hamiltonians and apply QAOA as a hybrid quantum–classical heuristic. For instance, Azad et al. [2], use this approach to design a hybrid quantum–classical heuristic implemented in Qiskit. The authors report experiments on small instances (limited by qubit counts) and compare solution quality against classical solvers, illustrating both the promise of variational quantum heuristics and the limitations due to resource requirements

Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

This work is licensed under a Creative Commons "Attribution-NonCommercial-NoDerivatives 4.0 International" license.

as the problem size increases. Recent studies analyze feasibility and scaling in terms of qubit requirements and circuit depth, often reporting rapid growth in resource needs even for moderate instance sizes [6].

Hardware-aware analyses highlight that, beyond qubit counts, real-device connectivity and compilation overhead (SWAP insertion, increased depth, and noise accumulation) are the main limiting factors for VRP-like encodings [3].

A promising direction is to integrate quantum routines into classical decomposition frameworks. For example, combining a route-based master problem (column generation) with a quantum ansatz for the pricing/search component has been proposed for CVRP settings [7]. Although not inherently multi-depot, the methodology generalizes: depots affect the master problem and the feasibility of generated routes.

QA toolchains and implementations. Industrial and open-source ecosystems around QA provide reference implementations for VRP variants (including multi-depot/capacitated variants) using QUBO modeling and hybrid workflows (classical partitioning + QPU calls) [5]. These codebases are valuable for reproducibility, but they also illustrate the practical reliance on classical preprocessing and post-optimization.

Alternative non-quantum options. Given current hardware constraints, researchers also consider alternatives that retain "Ising/QUBO-like" modeling benefits without requiring large-scale fault-tolerant quantum computers. Digital annealers, specialized Ising/QUBO optimizers, and large-scale parallel metaheuristics can serve as strong baselines for QUBO-based approaches, often outperforming naive quantum encodings at current scales. Comparative studies emphasize that the QUBO modeling paradigm is not exclusive to quantum devices and can be exploited classically as well [8]. In classical VRP, the most principled "two-level" method is route-based decomposition: a master set-partitioning model and a pricing subproblem (often an elementary shortest path problem with resource constraints, ESPPRC). This approach provides a natural plug-in point for heuristic or quantum subroutines, but the pricing problem itself is highly nontrivial.

4.3 Practical limitations of hybrid MD-CVRP solving

Hybrid approaches are attractive because they attempt to use quantum resources only where they might help, while relying on classical methods for the rest. However, several recurring obstacles appear in the literature:

Resource scaling and encoding overhead. Many encodings require a number of binary variables that grows at least quadratically with the number of customers (e.g., edge-based or position-based encodings), which translates directly to qubit requirements in gate-based approaches [6]. Multi-depot features (multiple starts/ends, depot assignment) add further variables and constraints.

QUBO/Ising penalty terms must be large enough to enforce feasibility but not so large that they dominate the energy landscape, harm optimization, or exceed hardware parameter ranges. Choosing penalty weights is instance-dependent

Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

This work is licensed under a Creative Commons "Attribution-NonCommercial-NoDerivatives 4.0 International" license.

and often becomes a major tuning burden, especially when capacity and depot constraints interact.

Feasibility and subtour elimination. Classical VRP uses strong subtour elimination (cuts) or flow-based constraints. In QUBO form, subtour elimination is typically approximated or encoded indirectly, which can produce low-energy infeasible solutions requiring substantial classical repair. Hybrid QA approaches explicitly acknowledge and address this by partitioning the instance and/or applying classical repair and improvement steps [4].

Hardware realism: connectivity, embedding, and noise. On QA hardware, dense QUBOs require minor embedding onto sparse device graphs, increasing the number of physical qubits via chains and reducing effective precision. This is especially problematic for VRP encodings with many interacting constraints.

On gate-model devices, compiling QAOA for dense problem graphs can require many SWAP gates due to limited connectivity, raising depth and error rates. Hardware-grounded VRP studies emphasize this as a bottleneck [3].

End-to-end performance and benchmarking. In many hybrid pipelines, the dominant compute cost remains classical: clustering/partitioning, generating candidate routes, and repairing infeasible solutions. Moreover, when the quantum subproblem is reduced enough to be hardware-feasible, the remaining search space may already be small enough for strong classical heuristics to solve quickly, limiting the opportunity for observable advantage.

Claims of improvement can be misleading if: (i) classical baselines are weak (e.g., naive heuristics), (ii) quantum methods are compared on different stopping criteria, or (iii) overheads (embedding, compilation, repeated sampling, repair) are excluded. Open-source QA/VRP pipelines demonstrate that hybrid success depends heavily on classical engineering and careful accounting of overheads [5].

4.4 Pragmatic outlook for future MD-CVRP research

A pragmatic research path that aligns with current evidence is:

1. Use strong classical MD-CVRP solvers (exact for small instances; ALNS/OR-Tools-style metaheuristics for larger instances) as baselines.
2. Use quantum components only on well-scoped subproblems where the oracle/encoding is cheap (e.g., selection among a bounded pool of candidate tours, or local neighborhood exploration).
3. Report end-to-end results with full overheads and ablation studies (classical-only vs. hybrid, sensitivity to penalty weights, pool size, noise).

This is consistent with both QA-based partitioning strategies [4] and recent gate-based feasibility/scaling analyses [6,3].

5 Conclusions

This paper contributes three complementary outcomes for hybrid quantum–classical logistics software. First, we provide a formal baseline MD-CVRP model

Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.

that is suitable for service-oriented integration and explicit enough to support implementation and benchmarking. Second, we define a reference fleet-management services architecture where an orchestration layer can invoke either classical or quantum routing solvers under a common workflow.

Finally, our analysis of current QA/QAOA-based VRP approaches indicates that near-term value is more likely to appear in carefully bounded hybrid subproblems than in end-to-end replacement of strong classical MD-CVRP pipelines. Therefore, practical progress should prioritize rigorous comparisons against competitive classical baselines, complete overhead reporting, and ablation studies that isolate the real contribution of each quantum component.

Acknowledgments. This work was partially supported by grants PID2021-126227NB-C21 and PID2021-126227NB-C22 funded by MCIN / AEI / 10.13039 / 501100011033 / FEDER, UE "ERDF"; PID2024-155693NB-C44 funded by MICIU / AEI / 10.13039/ 501100011033 / FEDER, UE; and by "Smart Computer systems Research and Engineering (SCORE)", funded by the Consejería de Transformación Económica, Industria, Conocimiento y Universidades of Junta de Andalucía under the "Programa de Fortalecimiento de Institutos Universitarios de Investigación de la Universidades Andaluzas, Centros e Infraestructuras para la adquisición del sello 'Severo Ochoa' o 'María de Maeztu'".

AI Use Disclosure Artificial intelligence (AI)–based language tools were used solely to assist with language editing and improving clarity in the preparation of this manuscript. All generated content was carefully reviewed and revised by the authors, who assume full responsibility for the accuracy, integrity, and originality of the work. The AI tools did not contribute to the design, analysis, interpretation of the results, or scientific conclusions.

References

1. Aaronson, S., Chen, L.: Complexity-theoretic foundations of quantum supremacy experiments. In: Proceedings of the 32nd Computational Complexity Conference. CCC '17, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, DEU (2017)
2. Azad, U., Behera, B.K., Ahmed, E.A., Panigrahi, P.K., Farouk, A.: Solving vehicle routing problem using quantum approximate optimization algorithm. IEEE Transactions on Intelligent Transportation Systems **24**(7), 7564–7573 (2023). https://doi.org/10.1109/TITS.2022.3172241
3. Azfar, T., et al.: Quantum-assisted vehicle routing: Realizing qaoa-based approach on gate-based quantum computer (2025)
4. Borowski, M., Gora, P., Karnas, K., Błajda, M., Król, K., Matyjasek, A., Burczyk, D., Szewczyk, M., Kutwin, M.: New hybrid quantum annealing algorithms for solving vehicle routing problem. In: Krzhizhanovskaya, V.V., Závodszky, G., Lees, M.H., Dongarra, J.J., Sloot, P.M.A., Brissos, S., Teixeira, J. (eds.) Computational Science – ICCS 2020. pp. 546–561. Springer International Publishing, Cham (2020)

Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

This work is licensed under a Creative Commons "Attribution-NonCommercial-NoDerivatives 4.0 International" license.

5. D-Wave Systems, contributors: D-wave-vrp: Solve the vehicle routing problem (and variants) using quantum annealing. GitHub repository (2024), https://github.com/dwave-examples/D-Wave-VRP
6. Fitzek, D., Ghandriz, T., Laine, L., Granath, M., Kockum, A.F.: Applying quantum approximate optimization to the heterogeneous vehicle routing problem. Scientific Reports **14**(1), 25415 (2024)
7. Harikrishnakumar, R., Nannapaneni, S., Nguyen, N.H., Steck, J.E., Behrman, E.C.: A quantum annealing approach for dynamic multi-depot capacitated vehicle routing problem. arXiv preprint arXiv:2005.12478 (2020)
8. Karpas, I.: Vrp whitepaper: Qubo formulations and optimization platforms. Whitepaper (2024), https://lightsolver.com/wp-content/uploads/2024/02/VRP_Whitepaper.pdf
9. Molino-Peña, E., García, J.M., Ruiz-Cortés, A.: Integrating terms of service and service level agreements for automating cloud service management. In: Aiello, M., Deng, S., Murillo, J.M., Georgievski, I., Benatallah, B., Wang, Z. (eds.) Service-Oriented Computing. pp. 19–34. Springer Nature Singapore, Singapore (2026)
10. Murillo, J.M., Garcia-Alonso, J., Moguel, E., Barzen, J., Leymann, F., Ali, S., Yue, T., Arcaini, P., Pérez-Castillo, R., García-Rodríguez de Guzmán, I., Piattini, M., Ruiz-Cortés, A., Brogi, A., Zhao, J., Miranskyy, A., Wimmer, M.: Quantum software engineering: Roadmap and challenges ahead. ACM Trans. Softw. Eng. Methodol. **34**(5) (May 2025). https://doi.org/10.1145/3712002, https://doi.org/10.1145/3712002
11. Preskill, J.: Quantum computing in the nisq era and beyond. Quantum **2**, 79 (2018)
12. Ruiz-Cortés, A., Parejo Maestre, J.A.: An Initial Exploration of Pricing-driven Governance for Hybrid Quantum-Classical SaaS. Sistedes (2025), https://hdl.handle.net/11705/JISBD/2025/91

Parejo, J. A., Ruiz-Cortés, A.: Operationalizing Quantum Optimization for Multi-Depot Logistics: A Hybrid service-oriented Approach. In: Cetina, C. (ed.) Actas de las XXX Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2026). Sistedes (2026). https://hdl.handle.net/11705/JISBD/2026/124

This work is licensed under a Creative Commons “Attribution-NonCommercial-NoDerivatives 4.0 International” license.