

VEKTOR
BY
INVESTPUPPY

VEKTOR RESEARCH SERIES · 2026

Production Infrastructure: The Vektor Technical Architecture

*AWS services, architectural decisions, and design principles —
documented as a record of choices made, not a list of technologies used.*

For technical due diligence teams, CTOs, and institutional partners

IP-WP-ARCH-260501-1.0

September 2026

ABSTRACT

Most vendor technical documentation is a list of services used. This paper is a record of decisions made — including the reasoning behind each one and what it would take to change them. This paper documents the Vektor production infrastructure through that lens: the architectural decisions made at the design stage, the reasoning behind each decision, and the properties those decisions produce — modularity, scalability, and the ability to integrate with existing institutional infrastructure. The platform is currently deployed in Singapore (ap-southeast-1) — selected as the first market: a globally-respected financial centre with a strong concentration of target clients. The architecture is region-agnostic by design: any AWS region can be configured as the deployment target without architectural change. Where capabilities are on the roadmap rather than live today, this paper says so directly.

KEY TAKEAWAYS

— Vektor runs on AWS using a microservices architecture. The current deployment is in ap-southeast-1 (Singapore). The architecture is region-agnostic — any AWS region can be configured as the deployment target. A client with UK or Hong Kong data residency requirements deploys to a different region.

— SageMaker serves two distinct roles: JupyterLab research environment for universe screening, and a managed XGBoost inference endpoint for signal validation. These are separate concerns, treated separately.

— SQS decouples order generation from execution. A connectivity issue with IBKR does not cascade into the application layer.

— EventBridge + Step Functions provide observable, retryable workflow orchestration — producing an implicit audit trail for every automated operation.

— Cognito with Azure EntraID integration enables SSO for institutional firms running Microsoft 365 — designed in from the first prototype.

— The two-account AWS strategy (shared CI/CD + workload) means deploying to a new region is a CI/CD pipeline configuration, not an architecture change.

— All infrastructure is defined in AWS CloudFormation. Every resource is version-controlled, reproducible, and auditable — no manual console configuration exists in the production path.

— ECS services run on Fargate Spot capacity, delivering approximately 70% reduction in compute costs versus on-demand pricing.

investpuppy.com · contact@investpuppy.com · IP-WP-ARCH-260501-1.0 · Copyright 2026 InvestPuppy

1

PLATFORM OVERVIEW

What Vektor is built on, and why the infrastructure choices reflect the same design principles as the platform itself.

Vektor is deployed on Amazon Web Services using a microservices architecture. The platform is currently running in ap-southeast-1 (Singapore), chosen as the first market for its stature as a respected global financial centre with deep concentration of target clients and for alignment with the MAS regulatory context of the founding client base.

Region-agnostic by design. Singapore (ap-southeast-1) is the current deployment region — a configuration choice, not a structural dependency. A client with UK data residency requirements deploys to eu-west-2; Hong Kong to ap-east-1; US to us-east-1 or us-west-2. AWS data residency guarantees follow the region. The two-account CI/CD architecture means deploying to a new region is a pipeline configuration, not an architecture change.

Three design principles were established before the first line of code was written. Modularity: every capability is an independently deployable service. Scalability by default: the infrastructure scales horizontally without architectural changes. Governance before growth: authentication, encryption, deployment pipeline separation, and operational visibility were designed in from the start.

Current deployment status. The architecture described in this document is designed for institutional production. The current deployment is in its pre-track-record phase, running in ap-southeast-1 with IBKR connected via paper trading. Section 9 describes which specific capabilities are live today and which are on the near-term roadmap.

2

ARCHITECTURE DIAGRAM

The complete Vektor AWS stack — nine layers, organised by function.

LAYER	COMPONENTS
USERS	Stock Analyst · Portfolio Manager · Operations
FRONTEND	React Admin Portal (all three roles) · AWS Amplify hosting + CI/CD
API	AWS API Gateway · Single entry point · Auth enforced · Rate limiting on roadmap · All ECS services exposed
COMPUTE	ECS Microservices (Python / Flask) · ECS Service Connect: sub-100ms inter-service calls · Fargate Spot: ~70% compute cost reduction · SageMaker: JupyterLab Research + XGBoost Endpoint
MESSAGING	SQS: Market order queue · EventBridge: Scheduled triggers · Step Functions: Workflow orchestration · SNS + Lambda router: Operational notifications
DATA	RDS PostgreSQL · Schema separation: market_data / portfolio_data / customer_data · S3: Model artefacts + Backup

LAYER	COMPONENTS
EXECUTION	EC2 t3.large: IBKR Gateway + SQS consumer · IBKR API (Simulation → Live) · EC2 t3.medium: Bastion host
AUTH & SECURITY	Cognito + Azure EntraID SSO · KMS encryption at rest · CloudWatch logging
INFRASTRUCTURE	AWS ap-southeast-1 · CloudFormation IaC · Two-account strategy (Shared CI/CD → Workload) · Dev/prod by stack naming convention

3

SERVICE INVENTORY

Every AWS service in the Vektor stack with its function and the decision that placed it there.

SERVICE	CATEGORY	FUNCTION
ECS on Fargate	Compute	Runs all Python/Flask microservices. Fargate removes server management. Fargate Spot reduces compute cost by ~70%.
ECS Service Connect	Compute	Handles inter-service communication via local proxy. Sub-100ms service-to-service calls without additional infrastructure.
API Gateway	API	Single entry point for all frontend requests. Authentication and logging enforced. Rate limiting and WAF on roadmap.
RDS PostgreSQL	Data	Primary datastore. Schema-separated: market_data, portfolio_data, customer_data. Single instance with clear migration path to physical separation.
S3	Storage	Model artefact storage (XGBoost versions). Backup storage. Long-term audit data.
SageMaker	ML	Two roles: JupyterLab environment for research, managed XGBoost inference endpoint for signal validation.
SQS	Messaging	Market order queue between ECS and IBKR Gateway. Decouples order generation from execution. Pre-execution audit trail.
EventBridge	Orchestration	Scheduled triggers for daily data pipelines (8AM SGT price + FX collection).
Step Functions	Orchestration	Workflow orchestration for multi-step pipelines. Per-step logging creates an implicit audit trail.
SNS + Lambda	Notifications	Fan-out notification architecture. Services publish to SNS; Lambda router delivers to Slack and other channels.
Cognito	Auth	User authentication. Azure EntraID integration for Microsoft 365 SSO — designed in from first prototype.
AWS Amplify	Frontend	React admin portal hosting and CI/CD. Automatic build and deploy on code push.

SERVICE	CATEGORY	FUNCTION
CloudFormation	Infrastructure	All infrastructure defined as code. Version-controlled, reproducible, auditable. No manual console configuration in production path.
CloudWatch	Observability	Logging in place. Alerting and dashboards on near-term roadmap.
KMS	Security	Encryption at rest for RDS and S3.
EC2	Execution	t3.large for IBKR Gateway + SQS consumer. t3.medium for bastion host.

4

DATA FLOWS

How data moves through the platform from market feeds to funded positions.

FLOW 1 — DAILY MARKET DATA PIPELINE

EventBridge → Step Functions → ECS data service (equity prices: OHLCV for all configured instruments, 3-year daily history) → ECS data service (FX rates: all configured currency pairs at 8AM SGT) → RDS storage → SNS → Lambda → Slack notification on completion or failure.

FLOW 2 — RESEARCH TO FUNDED POSITIONS

Research phase: SageMaker JupyterLab — universe screening (500+ instruments → target portfolio), deterministic optimisation (MPT max-Sharpe selection, solved analytically — not sampled), technical indicator selection (grid search across six indicators × 3-year history), XGBoost validation (confidence score per signal).

Strategy record: ECS writes approved strategy to RDS — instruments, weights, indicators, Sharpe ratios — pending Portfolio Manager approval.

Strategy approval: PM reviews via Admin Portal → approval record written (PM identity, timestamp, strategy ID). Maker/checker approval workflow with RBAC enforcement planned for a subsequent release; current implementation allows PM to assign directly.

Client onboarding: PM creates client profile, assigns approved strategy, sets mandate capital.

Allocation: Allocation service fetches live prices, applies lot rounding (98.94% on SGX board lots; 99.94%+ on single-share-lot markets), calculates position set. PM confirms → audit record written.

Cash funding: Operations records wire transfer (currency, amount, reference). Cash account activated.

Order execution: ECS → SQS queue → EC2 → IBKR Gateway → IBKR API (currently simulation; live before first Proof Partner client onboards). SNS notification on completion.

5

DECISION LOG

Key architectural decisions with the alternative considered and the reason for the choice made.

DECISION	ALTERNATIVE CONSIDERED	REASON
ECS Service Connect over standard DNS service discovery	Standard DNS	Sub-100ms inter-service calls without additional infrastructure. Consistent service-to-service communication with built-in load balancing.
Fargate Spot for 80–100% of ECS workloads	On-demand Fargate	~70% compute cost reduction. Acceptable for non-time-critical workloads. Fargate handles interruption recovery automatically.
SQS between ECS and IBKR Gateway	Direct API calls from ECS	Decouples order generation from execution. IBKR connectivity issues do not cascade into application layer. Pre-execution audit trail as SQS side effect.
EventBridge + Step Functions for pipeline orchestration	Lambda cron triggers	Complete workflow audit trail with per-step logging. Retryable at step level. Observable in CloudWatch.
SNS fan-out notification architecture	Direct Lambda notification calls	Services publish to SNS without knowledge of notification targets. Future multi-channel delivery requires no service changes.
SageMaker managed endpoint for XGBoost	Self-hosted inference	Model versioning and rollback independent of application deployment. Promotion requires explicit step, not automatic replacement.
Cognito + Azure EntraID SSO	Username/password auth only	Institutional firms run Microsoft 365. SSO removes credential management overhead. Designed in from first prototype — not retrofitted.
Two-account AWS strategy	Single account dev/prod separation	CodePipeline in shared account deploys cross-account. Deploying to new region is pipeline configuration. Workload accounts can be isolated per client.
CloudFormation for all infrastructure	Manual console configuration	Every resource version-controlled, reproducible, auditable. No configuration drift. Environment rebuild is deterministic.
Single RDS with schema separation	Separate databases per domain	Deliberate simplicity at prototype stage. Clear migration path to physical separation when compliance requirements demand it.
API Gateway as single entry point	Per-service API exposure	Auth and logging enforced at single point today. Rate limiting and WAF added at single point when ready.

DECISION	ALTERNATIVE CONSIDERED	REASON
ECS Service Connect over App Mesh	AWS App Mesh service mesh	Service Connect requires no additional infrastructure. App Mesh adds operational complexity not justified at current scale.
Fargate over EC2 for ECS tasks	EC2-backed ECS cluster	No server management. Auto-scaling without capacity planning. Higher per-unit cost offset by Fargate Spot savings.

6

SECURITY & DATA RESIDENCY

Encryption, authentication, network isolation, and data residency guarantees.

Encryption at rest: AWS KMS for RDS and S3. Encryption in transit: all communications over HTTPS/TLS.

Authentication: Cognito with Azure EntraID SSO. Infrastructure-level RBAC (AWS IAM) planned for a subsequent release. Current enforcement is at the application layer.

Data residency: Current deployment is ap-southeast-1. AWS data residency guarantees apply per region — data does not leave the configured region under normal operation. Clients with UK data residency requirements deploy to eu-west-2. Hong Kong to ap-east-1. Switzerland (nFADP/DSG) to eu-central-2 (Zurich). UAE (DFSA/FSRA) to me-central-1. Singapore (MAS TRM Guidelines) to ap-southeast-1. US to us-east-1 or us-west-2. Each region satisfies the data residency obligations of the applicable jurisdiction.

Network isolation: VPC with private subnets for compute and data layers. Bastion host for SSH access. API Gateway as the only public entry point.

7

SCALABILITY & REGIONAL DEPLOYMENT

How the architecture scales across users, mandates, and geographies.

Horizontal scaling: ECS services on Fargate scale independently. API Gateway handles traffic distribution. RDS scales vertically and, when required, via read replicas. No architectural changes are required to scale from 5 to 50 client mandates.

Regional deployment: the two-account CI/CD architecture makes adding a deployment region a pipeline configuration task. CodePipeline in the shared account builds Docker images, pushes to ECR, and deploys cross-account. For production client deployments, separate workload accounts per client region are the target architecture.

8

INTEGRATION ARCHITECTURE

How Vektor connects with existing DPM infrastructure.

REST API output: Vektor's allocation service exposes a REST API endpoint that returns a complete allocation table — instrument identifier, target quantity, target value, portfolio weight, and assigned indicator. A standard integration between this endpoint and an existing PMS typically takes two to four weeks for a developer familiar with the target system.

Market data substitution: Yahoo Finance is the current default data source. Bloomberg or any institutional market data feed can be substituted as the data source without architectural changes. This is a Proof Partners configuration option.

Execution venue: IBKR is the current execution venue via the IBKR Gateway and TWS API. Alternative execution venues using FIX protocol are a roadmap item. The SQS queue architecture means the execution layer is already decoupled from the rest of the application.

Current as of 4 August 2026.

9

WHAT IS NOT YET IN SCOPE

Honest current state — capabilities designed but not yet live, each with a defined trigger condition.

Current as of 4 August 2026.

CAPABILITY	CURRENT STATUS & TRIGGER
Live trade execution	IBKR is connected to a paper trading account. Live execution will be activated before the first Proof Partner client onboards. This is a defined trigger condition, not an open-ended roadmap item.
Infrastructure-level RBAC	Role enforcement is currently at the application layer. Infrastructure-level RBAC via AWS IAM is planned for the next prototype iteration.
Maker/checker approval workflow	Current implementation allows the PM to assign strategies directly. Maker/checker with RBAC enforcement is planned for the next prototype iteration.
API Gateway rate limiting and WAF	Authentication and logging are enforced today. Rate limiting and WAF protection are on the near-term roadmap.

CAPABILITY	CURRENT STATUS & TRIGGER
Production observability	CloudWatch logging is in place. Alerting, dashboards, and automated drift detection are on the near-term roadmap.
Production environment promotion	The current deployment is a dev environment. Promotion to a production environment will occur before the first Proof Partner client onboards.
External PMS integrations	The REST API output is available. Certified integrations with specific PMS platforms are a roadmap item based on Proof Partner requirements.

10

CONCLUSION

Architecture as a statement of values.

The choices documented in this paper are not a list of technologies adopted because they were available. They are a record of decisions made against explicit criteria: modularity so that any component can be replaced without disrupting the others; scalability without architectural rework; governance designed in from the start rather than added in response to requirements.

The platform is in its pre-track-record phase. The architecture is designed for production. The gap between the current state and full production deployment is implementation and integration work — not architectural change. The Decision Log captures the reasoning behind each major choice so that future development decisions build on the same principles rather than cutting across them.

For AI governance and ML model management, see WP-08. For compliance architecture, see WP-06. For risk disclosure, see WP-09. All documents at investpuppy.com.

investpuppy.com · contact@investpuppy.com · IP-WP-ARCH-260501-1.0 · Copyright 2026 InvestPuppy

The Full Document Suite

Organised by access tier — not a catalogue, a map.
Full reading-order guidance in the Due Diligence Navigation Guide.

UNGATED — 29 DOCUMENTS

VEKTOR COMMERCIAL

- Vektor At a Glance
- Vektor Brand Story
- Vektor Platform Brochure
- Due Diligence Navigation Guide
- What Vektor Is Not
- Why Not the Incumbents?
- Why Vektor
- Reference Output — SGX Equity Mandate

VEKTOR RESEARCH PAPERS

- WP-09: Risk and Limitation Disclosure
- WP-10: Evaluating Performance

OUR BETTER WAY

- BW-01: Technical Debt
- BW-02: Vektor House Implementation
- BW-03: The Configurability Gap
- BW-04: AI as Instrument

UNVARNISHED

- UNV-01: Why Implementations Fail
- UNV-02: Nobody Owns It
- UNV-03: Complexity as Cover
- UNV-04: Plan Last
- UNV-05: The Change Request Trap
- UNV-06: Invisible Stakeholders
- UNV-07: The Hammer, the Tool, and the Methodology
- UNV-08: Life After Live
- UNV-09: The U in UAT
- UNV-10: Nine Women
- UNV-11: The AI Revolution — Automating Bad Practice at Scale
- UNV-12: Model Bank — the perfect solution for banks that don't exist
- UNV-13: The parameter that broke the camel's back

WHITE PAPERS

- WP-12: Rebalancing: Drift Thresholds and Disclosure
- WP-13: Portfolio Valuation Methodology

investpuppy.com/documents

REGISTRATION REQUIRED — 13 DOCUMENTS

VEKTOR COMMERCIAL

- Vektor FAQ
- One Platform, Three Delivery Models
- Reference Output — SGX · NASDAQ · LSE · 21 May 2026

VEKTOR RESEARCH PAPERS

- WP-Index · Research Series Index
- WP-01: Quantitative Portfolio Construction
- WP-02: Signal Optimisation
- WP-03: Capital Allocation
- WP-04: Indicator Selection
- WP-05: Multi-Currency
- WP-06: Audit and Compliance

● WP-07: Technical Architecture (you are here)

- WP-08: AI and ML Philosophy

WHITE PAPERS

- WP-11: Signal Robustness Architecture — Why walk-forward validation was considered and discarded

How to register: investpuppy.com/research

NDA REQUIRED — 8 DOCUMENTS

VEKTOR COMMERCIAL

- First 90 Days
- Mutual Non-Disclosure Agreement
- Proof Partners Programme
- Workflow Integration Guide
- Vektor Operational Walkthrough — Client to Portfolio Valuation
- Proof Partners Agreement

INVESTPUPPY

- Alloy Partners Programme

OTHER

- Vektor OpenWealth Integration — Commercial Overview

Available once a Proof Partners conversation begins. Start at investpuppy.com/proof-partners

