TECHNICAL IMPLEMENTATION PROPOSAL

Enterprise Infrastructure and Sovereign Private Cloud

Corporate Data Centre and Private Cloud Platform

End to end implementation covering facility engineering, electrical and mechanical design, network fabric, compute, storage, cloud control plane, platform services, security, disaster recovery, migration and steady state operations. Delivered in 12 weeks from hardware on site.

Prepared by	Tech4Biz Solutions Pvt Ltd
Prepared for	Chief Information Officer, Corporate IT
Document type	Technical Implementation Proposal and Solution Design
Delivery window	12 weeks to accepted production platform, measured from hardware on site
Precondition	Hardware procurement runs ahead of the 12 weeks and is not included in it
Version	2.0
Date	July 2026
Classification	Confidential

About This Document

This document sets out how Tech4Biz will design, build and operate a corporate data centre and private cloud platform. It is written as an implementation specification rather than a capability overview. Every section states what gets built, how it is configured, who owns it, and what acceptance looks like.

The 12 week delivery window is achievable because the reference architecture is already running in production. Cloudtopiaa is a live sovereign infrastructure as a service platform operated by Tech4Biz on CtrlS Tier IV. The VVCE private cloud is a live multi-tenant deployment serving thousands of concurrent sessions. The blueprint, the automation and the runbooks exist. This programme configures a proven platform for a corporate estate rather than designing one from a blank page.

One precondition governs the whole plan. The 12 week clock starts on the day hardware is received and asset tagged at both sites. Hardware procurement carries its own lead time, it sits ahead of the 12 weeks, and it is not counted inside them. Section 15 states the indicative procurement window and what compresses it.

CONTENTS

Document Structure

Section	Title
01	Engagement Context, Objectives and Baseline Assumptions
02	Proven Delivery: Cloudtopiaa and VVCE
03	Solution Architecture Overview
04	Data Centre Physical Engineering
05	Facility Commissioning and Acceptance
06	Network Fabric and Addressing Design
07	Compute Design and Bill of Materials
08	Storage Design and Data Services
09	Private Cloud Platform Design
10	Platform Services: Containers, Databases and Virtual Desktops
11	Security, Identity and Compliance Architecture
12	Backup, Disaster Recovery and Business Continuity
13	Observability, Automation and Operations
14	Workload Discovery and Migration Plan
15	12 Week Implementation Plan
16	Testing, Acceptance and Handover Criteria
17	Commercial Model and Five Year Total Cost of Ownership
18	Risk Register and Designed Failure Modes
19	Governance, Responsibility Matrix and Next Steps

How to read this proposal

Sections 01 to 03 establish scope, proof and architecture. Sections 04 to 13 are the technical build specification and can be handed directly to an implementation team. Sections 14 to 16 cover delivery, migration and acceptance. Sections 17 to 19 cover commercials, risk and governance.

SECTION 01

Engagement Context and Objectives

The organisation needs infrastructure it controls. Public cloud has delivered speed but has produced unpredictable egress and licence costs, data residency exposure under DPDP, and a dependency that is hard to price at renewal. A private cloud closes those gaps if it is built as a platform rather than as a rack of servers.

1.1 Business Objectives

Objective	Measure of success
Speed to platform	Production ready private cloud accepted within 12 weeks of hardware arriving on site
Cost predictability	Five year TCO below equivalent public cloud spend, no variable egress component
Data sovereignty	All regulated and customer data resident in India, under organisational key custody
Developer velocity	Self-service environment provisioning in under fifteen minutes, no ticket required
Resilience	99.982 percent platform availability, tested failover to a geographically separate site
Auditability	Single evidence trail across identity, change, access and data for ISO 27001 and DPDP
Future readiness	Power, cooling and fabric headroom for GPU workloads without facility rework

1.2 Baseline Assumptions

The design below is sized against the following baseline. These figures drive every quantity in this document. Tech4Biz confirms or revises them in week 1, and the bill of materials is reissued if they move materially.

Parameter	Baseline
Deployment model	Colocation at a certified Tier III or Tier IV facility, primary and DR
Contracted footprint	20 racks primary, 10 racks DR
Rack density	12 kW average, two racks provisioned at 40 kW for future accelerated compute
Design IT load	Approximately 300 kW at primary, 130 kW at DR
Workload estimate	Approximately 600 virtual machines, 12 line of business applications, 2 ERP instances
End user computing	300 virtual desktop seats
User population	1,200 to 2,000 employees across three locations
Availability target	99.982 percent for Tier 1 applications
Growth assumption	25 percent compute growth year on year for three years
Regulatory scope	DPDP Act, ISO 27001, sector regulator controls as applicable

1.3 In Scope and Out of Scope

In scope	Out of scope
Facility engineering specification, electrical and mechanical design, commissioning	Civil construction of a greenfield building shell

In scope	Out of scope
Rack, power, cooling, containment, structured cabling, DCIM	Campus LAN and branch office refresh
Network fabric, firewalls, load balancers, WAN edge	Application code remediation and refactoring
Compute, storage, virtualisation and cloud control plane	Legal interpretation of regulatory obligations
Platform services: Kubernetes, databases, virtual desktops, object storage	Third party SaaS contracts and their exit
Security architecture, IAM integration, key management	Data cleansing inside application databases
Backup, DR site build and orchestrated failover	Ongoing managed services beyond hypercare unless contracted
Workload discovery, migration waves, cutover, runbooks, training	End user device refresh

Design decision: colocate, do not build

Tech4Biz recommends colocation over owning the facility, and the 12 week timeline assumes it. Building concrete converts a technology decision into a twenty year real estate commitment, it locks the density envelope at the point of lowest knowledge, and it puts uptime engineering inside a team whose day job is applications. Section 04 still specifies the full electrical and mechanical design, because that specification is what Tech4Biz validates the colocation provider against, and it is the build specification if the organisation chooses to own the facility.

SECTION 02

Proven Delivery: Cloudtopiaa and VVCE

The 12 week build window is not an estimate. It is the delivery time of a platform Tech4Biz has already built twice, once as a commercial sovereign cloud and once as a multi-tenant private cloud serving a live user base. The window measures the build, not the wait for hardware.

2.1 Cloudtopiaa: Sovereign Infrastructure as a Service

Dimension	Detail
What it is	Tech4Biz owned and operated sovereign IaaS platform, in commercial production
Facility	CtrlS Tier IV data centre, India resident
Capability	Compute, block and object storage, virtual networking, Kubernetes, self-service portal and API
Sovereignty	Full data residency within India, organisation held encryption keys, no foreign control plane dependency
What it proves	Tech4Biz operates a cloud control plane in production rather than only integrating one
Reuse in this build	Control plane blueprint, infrastructure as code modules, tenancy and quota model, operational runbooks

2.2 VVCE Private Cloud Lab Ecosystem

Dimension	Detail
Client	Vidyavardhaka College of Engineering, Mysuru, Karnataka
Scope	22 specialised technology labs across 9 departments, delivered as a phased rollout
Scale	Live deployment supporting thousands of concurrent browser based sessions, architected to scale further without re-architecture
Stack	Software defined storage on Ceph, Kubernetes orchestration, GPU accelerated compute, custom self-service portal
Fabric	25 GbE top of rack data switching with a separate 1 GbE out of band management plane for BMC access
Design discipline	Full hardware procurement specification, vendor neutral bill of materials, network architecture and scaling path documented before purchase
What it proves	Multi-tenant private cloud design, capacity engineering and phased rollout delivered against a fixed budget
Reuse in this build	Storage cluster design, out of band plane pattern, portal and catalogue model, capacity forecasting method

2.3 What Transfers Directly to This Programme

Asset carried over	Effect on this build	Weeks saved
Infrastructure as code modules	Bare metal, fabric, storage and hypervisor build are executed rather than written	4 to 6
Control plane blueprint	Tenancy, quota, catalogue and metering are configured rather than designed	3 to 4

Asset carried over	Effect on this build	Weeks saved
Storage cluster reference design	Cluster topology, failure domains and tuning are known good	2 to 3
Operational runbooks	Failure scenarios, patching and capacity procedures already written and exercised	2 to 3
Vendor and procurement relationships	Pre-qualified SKU list and partner stock shorten hardware lead time	3 to 5
Acceptance test suite	Test cases, thresholds and evidence templates already exist	1 to 2

Why 12 weeks is credible and where it breaks

The compression comes from reuse, parallelism and colocation. It holds only if three conditions are met. Hardware is delivered and asset tagged before week 1, because the clock starts there and not at purchase order. The colocation contract is signed and the cage is ready by week 2. The design is frozen at the end of week 2 and is not reopened. If any of the three slips, the programme slips by the same amount, because there is no float left in a 12 week plan.

SECTION 03

Solution Architecture Overview

The platform is built as seven layers. Each layer has an independent failure domain, an independent lifecycle, and a defined interface to the layer above. This is what allows a storage refresh in year four without touching the cloud control plane.

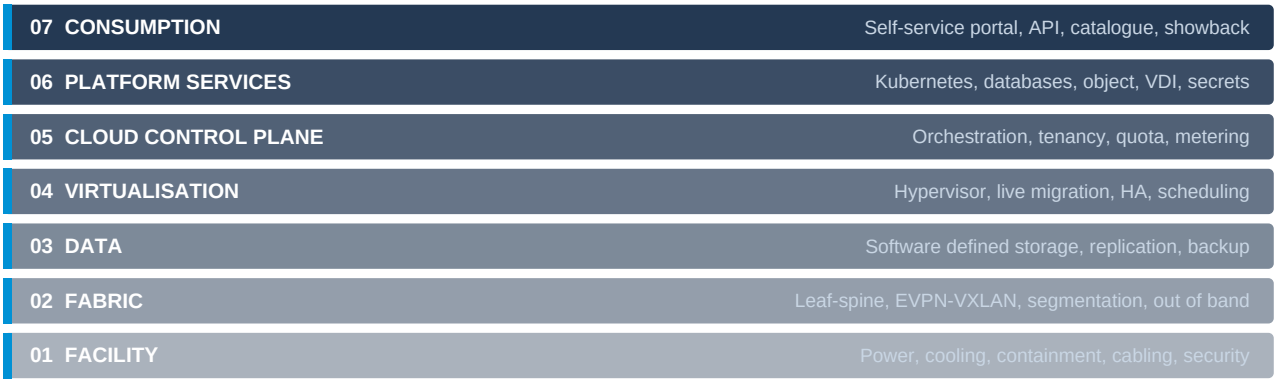

Figure 1. Seven layer platform architecture. Each layer is independently refreshable.

3.1 Architectural Principles

Principle	What it means in this build	Consequence if ignored
Two of everything, always	No single power path, switch, controller or fabric carries production alone	One component failure becomes a business outage
Software defined over appliance	Storage and network functions run on commodity hardware with open control planes	Forklift upgrades and vendor lock at renewal
Separate the planes	Production, management, storage and out of band traffic never share a fabric	A management flood takes production down with it
Automate the build	Infrastructure as code from bare metal upward, no hand configured production node	Configuration drift makes DR unprovable
Design for density now	Power, cooling and fabric headroom reserved for 40 kW racks	Year three AI programme requires a facility move
Prove resilience, do not claim it	Quarterly failover drills written into the operating model	DR is a line item that has never worked

3.2 Logical Topology

The primary site carries all production workloads. The DR site carries approximately 40 percent of primary capacity, sized to run Tier 1 and Tier 2 applications at reduced but acceptable performance. Replication runs asynchronously over a dedicated dark fibre or MPLS pair. The two sites sit in different seismic zones and on different grid feeds, separated by at least 250 km.

SECTION 04

Data Centre Physical Engineering

This section is the full facility specification. It is the document Tech4Biz uses to validate a colocation provider against contracted service levels, and it is the build specification if the organisation elects to own the facility.

4.1 Design Load and Topology

Parameter	Primary site	DR site
Design IT load	300 kW	130 kW
Total facility load with cooling and losses	Approximately 480 kW	Approximately 210 kW
Target design PUE	1.4 to 1.5 annualised	1.5 to 1.6 annualised
Electrical topology	2N UPS, N+1 generation	N+1 UPS, N+1 generation
Mechanical topology	N+1 precision cooling	N+1 precision cooling
Uptime classification	Tier III concurrently maintainable or better	Tier III
Concurrent maintainability	Any component serviceable without workload impact	Same

4.2 Electrical Design

Element	Specification
Utility supply	Two independent HT feeders from different substations, automatic changeover
Transformers	Two dry type transformers, each rated for full facility load, independently isolatable
UPS	Two independent strings, each rated 400 kVA, each capable of carrying 100 percent of IT load
Battery autonomy	Minimum 10 minutes at full IT load, VRLA or lithium ion with battery monitoring
Generation	Two diesel generators at 750 kVA each in N+1, automatic start within 15 seconds
Fuel	Bulk storage for 24 hours at full facility load, with a contracted priority refuelling agreement
Distribution	Independent A and B distribution boards, separate cable routes, no common failure point
Rack power	Dual feed to every rack, three phase intelligent PDUs, per outlet metering and switching
Breaker discipline	No rack loaded beyond 40 percent of breaker rating at steady state
Earthing and protection	Dedicated earth pit array, surge protection at each distribution stage, lightning protection

Dual feed with 40 percent steady state loading is deliberate. It means a full A side failure shifts the entire load to the B side without tripping a breaker, and it preserves the headroom required when a node group migrates during maintenance.

4.3 Mechanical and Cooling Design

Element	Specification
Cooling type	In-row precision cooling for the 12 kW rows, rear door heat exchanger or in-row for the 40 kW positions
Capacity	Approximately 85 TR at primary in N+1 configuration
Airflow	Approximately 120 CFM per kW, hot aisle containment with blanking panels and brush grommets
Inlet temperature	22 to 24 C at rack inlet, within ASHRAE A1 allowable envelope
Delta T	12 C design across the rack, monitored per aisle
Humidity	40 to 60 percent relative humidity, dew point controlled
Chilled water	Dual loop with isolation valves, N+1 pumps, thermal buffer for generator start transition
Liquid readiness	Facility water loop or rear door heat exchanger provisioned at the two high density positions
Leak detection	Zone based leak detection under every wet zone, alarmed to BMS and DCIM

4.4 Structured Cabling and Containment

- Overhead cable tray in three separate layers: fibre, copper data and power. Power never shares a tray with data.
- OS2 single mode fibre for spine to leaf and inter site, OM4 multimode intra rack, MPO trunk cabling for 100G.
- Category 6A shielded copper for the out of band management plane, physically separated and colour coded.
- Colour coding by plane: production, storage, management, out of band and DMZ each carry a distinct jacket colour.
- Every port labelled at both ends against the DCIM record, with patching documented before energisation.
- Structured patching at end of row, with documented spare capacity of 25 percent at every panel.

4.5 Fire, Safety and Physical Security

Control	Specification
Detection	VESDA aspirating smoke detection at rack row level, cross zoned with conventional detection
Suppression	Clean agent gaseous suppression, zoned, with pre-action interlock and manual abort
Compartmentation	Two hour fire rated separation between electrical, mechanical and white space
Access control	Mantrap entry, biometric plus card dual factor, cage level access list reviewed monthly
Surveillance	CCTV at cage, aisle and door level, minimum 90 day retention
Asset control	Every asset tagged and recorded in DCIM, quarterly physical reconciliation
Emergency power off	EPO with dual action and procedural control to prevent accidental activation

4.6 Building Management and DCIM

- BMS monitors electrical, mechanical, fire and access, and feeds alarms into the platform observability stack rather than sitting on a separate console nobody watches.
- DCIM holds the asset register, rack elevation, power chain mapping, port mapping and capacity model.
- Per outlet power metering feeds DCIM, producing per rack kW trending and live PUE reporting.
- Temperature and humidity sensors at three heights per cold aisle, front and rear of each row.

- The capacity model reports remaining power, cooling, rack unit and port capacity as four separate ceilings, because they exhaust at different times.

4.7 Rack Layout

Rack group	Qty	Density	Contents
Compute	8	12 kW	General purpose and memory optimised hypervisor nodes
Storage	4	10 kW	Storage nodes and NVMe capacity
Network	2	6 kW	Spine, leaf, border leaf, out of band, structured patching
Security and edge	2	6 kW	Firewalls, load balancers, WAN routers, DDoS appliance
Management	2	8 kW	Control plane, backup targets, jump hosts, DCIM and BMS
High density reserve	2	40 kW	Empty at go live, power, cooling and fabric provisioned

Future failure mode: the power envelope

The most common three year failure is not a hardware fault. It is discovering that the contracted power ceiling has been reached, that the facility cannot uprate the cage, and that the only route to expansion is a second site or a migration. Contract for a power ceiling of at least 1.8 times day one draw, and put a written uprate clause in the colocation agreement before signature.

SECTION 05

Facility Commissioning and Acceptance

A facility is not accepted because it was built. It is accepted because it was tested at full load, under failure, with evidence. Tech4Biz runs a five level commissioning process against the colocation provider or the build contractor.

Level	Stage	What is proven	Evidence
L1	Factory acceptance	Equipment meets specification before it ships	Witnessed FAT report per major asset
L2	Delivery and installation	Correct equipment installed correctly and safely	Installation checklist, photographic record
L3	Pre-functional	Each component energises and operates standalone	Start up report, settings record
L4	Functional	Each system operates to design under controlled conditions	Functional test script with measured results
L5	Integrated systems test	The whole facility behaves correctly under simulated full load and failure	IST report with load bank data and signed witness

5.1 Integrated Systems Test Scenarios

Scenario	Pass threshold
Utility feed A loss at full load bank	Zero interruption, UPS carries load, generator picks up within 15 seconds
Generator start failure on unit one	N+1 unit carries full facility load without interruption
UPS string A total loss	String B carries 100 percent of IT load with no output transient beyond tolerance
Loss of one cooling unit per row	Rack inlet temperature stays below 27 C for 30 minutes at full load
Chilled water pump failure	Standby pump engages, no thermal excursion beyond ASHRAE allowable
Full facility black start	Facility returns to full service within the contracted recovery time, correctly sequenced
Fire alarm and suppression interlock	Correct zoning and interlock behaviour, no false discharge, abort works
EPO functional test	Correct isolation, documented restoration procedure executed successfully

Load bank testing is not negotiable. A facility tested only with live IT load has never been tested at its design point, because the IT load at handover is a fraction of the design load. Tech4Biz requires load bank data at 100 percent design load before accepting the white space.

SECTION 06

Network Fabric and Addressing Design

The fabric is a two tier leaf-spine running EVPN-VXLAN over a BGP underlay. Three tier networks with spanning tree have no place in a build commissioned in 2026.

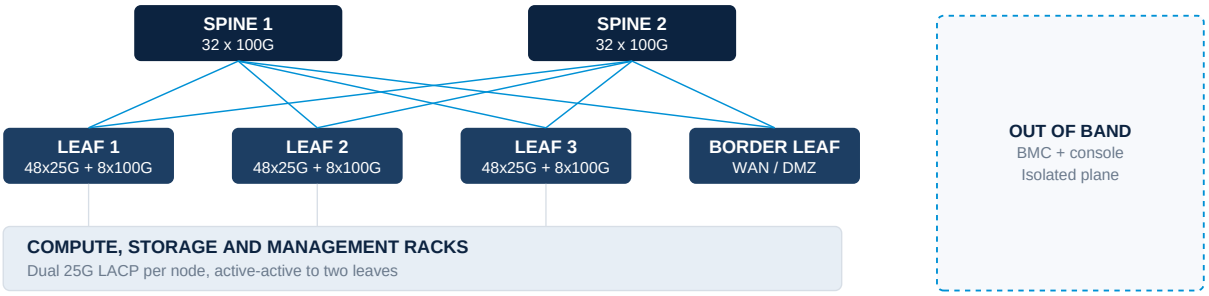

Figure 2. Leaf-spine fabric with border leaf pair and an isolated out of band management plane.

6.1 Fabric Specification

Element	Primary site	DR site	Notes
Spine	2 x 32 port 100G	2 x 32 port 100G	eBGP underlay, no host connections
Leaf	6 x 48 port 25G plus 8 x 100G uplink	4 x same	MLAG or EVPN multihoming pairs
Border leaf	2 x 48 port 25G	2 x same	WAN, DMZ, firewall and load balancer attachment
Out of band	2 x 48 port 1G	2 x same	Physically separate switches with a separate uplink
Host attachment	2 x 25G LACP per node	2 x 25G LACP	Active-active to two different leaves
Inter site	2 x 10G diverse path	2 x 10G diverse path	Dark fibre preferred, MPLS acceptable
Internet edge	2 x 1G burstable, diverse carriers	1 x 1G	Upstream DDoS scrubbing, BGP multihoming

6.2 Overlay and Segmentation

- **Underlay.** eBGP with a unique ASN per leaf and a shared spine ASN. Point to point /31 links. ECMP across all spine paths.
- **Overlay.** BGP EVPN with VXLAN encapsulation, Layer 2 and Layer 3 VNIs, anycast gateway on every leaf so workload mobility does not require re-addressing.
- **Tenancy.** One VRF per security zone: production, non-production, DMZ, management, backup, out of band.
- **Microsegmentation.** Distributed firewall enforced at the virtual NIC, policy expressed against workload tags rather than IP addresses.
- **East-west inspection.** Service insertion at the border leaf for zone crossing traffic that requires deep inspection.
- **Quality of service.** Storage replication and live migration traffic are policed so they cannot starve production east-west flows.

6.3 Addressing and Naming Plan

Plane	Allocation	Purpose
Underlay loopbacks	10.255.0.0/24	Router IDs and VTEP source addresses
Underlay point to point	10.255.1.0/24 as /31 blocks	Spine to leaf links
Production overlay	10.10.0.0/16	Tier 1 and Tier 2 application segments
Non-production overlay	10.20.0.0/16	Development, test, staging
Management	10.30.0.0/16	Hypervisor, control plane, monitoring
Storage	10.40.0.0/16	Replication and cluster traffic, jumbo frames enabled
Out of band	10.50.0.0/16	BMC, console servers, PDU, environmental, BMS
DMZ	172.16.0.0/20	Internet facing and partner facing services

The DR site mirrors this scheme with a distinct second octet, which keeps replication simple and makes failover routing deterministic. Naming follows a strict convention of site, role, sequence and environment, so that automation can infer placement from hostname alone.

Design decision: two physical fabrics

Production and out of band never share hardware in this design. The out of band plane carries BMC, console, PDU, environmental and BMS traffic on its own switches with its own uplink. When the production fabric is the thing that is broken, the out of band plane is how it gets fixed. This is the same pattern Tech4Biz deployed at VVCE, where a dedicated 1 GbE management plane sits alongside the 25 GbE data fabric.

SECTION 07

Compute Design and Bill of Materials

Compute is sized against measured peak with 30 percent headroom, plus enough spare capacity to absorb the loss of any two nodes without performance degradation.

7.1 Node Classes

Class	Primary	DR	Specification
General compute	12	6	2U dual socket, 2 x 32 core, 768 GB DDR5, 2 x 960 GB NVMe boot mirror, 2 x 25G, dedicated BMC
Memory optimised	3	2	2U dual socket, 2 x 32 core, 2 TB DDR5, for ERP and in-memory database tiers
Storage node	4	3	2U, 1 x 32 core, 256 GB, 12 x 15.36 TB NVMe, 2 x 25G storage, 2 x 25G front end
Management node	2	2	1U dual socket, 2 x 16 core, 256 GB, mirrored NVMe, hosts control plane and tooling
GPU reserve	0 at go live	0	Rack, power, cooling and fabric ports provisioned for 8 accelerated nodes

7.2 Capacity Position at Go Live

Resource	Raw	Usable after HA reserve	Day one utilisation
Physical cores	960	800	Approximately 55 percent
Memory	15.4 TB	12.8 TB	Approximately 60 percent
Flash capacity	737 TB raw	245 TB usable	Approximately 48 percent
Fabric bandwidth	2.4 Tbps bisection	Full	Under 15 percent

Day one utilisation is held near 55 percent deliberately. It absorbs the 25 percent annual growth assumption for two full years before any expansion purchase, and it leaves enough slack for a full node group evacuation during patching without touching workload performance.

7.3 Procurement Discipline

Procurement runs ahead of the 12 week build and is not counted inside it. Hardware lead time is a market condition rather than a delivery choice, and pretending otherwise is how programmes acquire a schedule nobody can hold. Section 15.1 states the indicative window. The discipline below is what shortens it.

- The bill of materials is drawn from a pre-qualified SKU list that Tech4Biz already procures against, which removes the evaluation cycle from the critical path.
- Alternate SKUs are pre-qualified for every long lead item, so a stock shortage produces a substitution rather than a delay.
- Partner stock is drawn on wherever the SKU allows it, which is the single largest available compression.
- Staged delivery is agreed with the distributor: network and storage first, compute second, because the fabric must be live before nodes are useful.
- The 12 week clock starts only when the full delivery is received and asset tagged at both sites, not on partial delivery.

7.4 Firmware and Lifecycle

- A single validated firmware baseline is defined per node class and enforced by automation. No node enters production off baseline.
- Firmware and hypervisor patching runs on a quarterly cadence, one fault domain at a time, with automatic workload evacuation.
- Hardware support runs four hour on site for production nodes and next business day for DR.
- Refresh is planned at year five, node class by node class, so that no single year carries the full capital event.

SECTION 08

Storage Design and Data Services

Storage is software defined and scales linearly. Traditional dual controller SAN is excluded because it caps growth at the controller and forces a forklift replacement at end of life. This is the same architectural position Tech4Biz took at VVCE and in Cloudtopiaa.

8.1 Storage Tiers

Tier	Media	Protection	Target performance	Consumers
Tier 0	NVMe all flash	Triple replica	Sub millisecond, 200k IOPS per volume	ERP database, transaction systems
Tier 1	NVMe all flash	Erasure coded 4+2	Under 2 ms, 60k IOPS per volume	Line of business applications, VDI
Tier 2	High capacity NVMe	Erasure coded 8+3	Under 10 ms	File shares, archives, image repositories
Backup	Object on capacity nodes	Erasure coded, immutable	Throughput oriented	Backup targets, long term retention

8.2 Cluster Topology and Failure Domains

- Failure domain is set at the rack level, so a full rack loss does not cause data unavailability.
- Cluster monitors are deployed in odd numbers across separate racks and separate power paths.
- Rebuild throughput is throttled against a policy so that recovery traffic cannot starve production input and output.
- The cluster network runs on its own VLAN with jumbo frames, separated from front end traffic at the fabric.
- Scale out to 32 nodes per cluster is supported without architectural change, roughly six times the day one footprint.

8.3 Data Services

- **Inline compression and deduplication** on Tier 1 and Tier 2. Tier 0 runs compression only, to protect latency.
- **Encryption at rest** with keys held in the organisation key management service, not in the storage platform. Rotation is scheduled and audited.
- **Snapshots** on an application consistent schedule, retained locally for 14 days.
- **Asynchronous replication** to the DR site with per volume recovery point policies.
- **Immutable object tier** with write once retention lock, which is the specific control that survives a ransomware event.
- **Quality of service** per volume, so a runaway test workload cannot starve production.

Future failure mode: the silent capacity cliff

Erasure coded clusters degrade sharply once they pass approximately 80 percent utilisation, because rebuild traffic competes with production for the same media. The operating model in Section 13 therefore sets a hard procurement trigger at 65 percent utilisation, not at 85 percent. Ordering storage after the cliff is reached is ordering it too late.

SECTION 09

Private Cloud Platform Design

This layer is what separates a private cloud from a virtualisation estate. Without tenancy, quota, catalogue, metering and an API, the platform is a set of servers that still needs a ticket.

9.1 Platform Stack Selection

Tech4Biz runs a formal five year licence cost evaluation across the candidates below during week 1 and 2. Post consolidation in the virtualisation market, licence repricing at renewal is now the largest single cost risk in this programme.

Option	Strengths	Risks	Best fit
Incumbent commercial hypervisor	Mature tooling, deep operational familiarity, broad ISV certification	Licence repricing at renewal, per core escalation, vendor concentration	Estates with heavy ISV certification dependency
Open source cloud platform	No licence escalation, full API surface, strong multi-tenancy and quota model	Requires genuine platform engineering capability in house or via partner	Organisations committed to a platform team, and the Cloudtopiaa lineage
Hyperconverged commercial stack	Fast to stand up, integrated storage and compute lifecycle	Node level licence coupling, storage cannot scale independently	Smaller estates that value speed over granularity
Lightweight open hypervisor with clustering	Low cost, simple operations, good fit for mid scale estates	Thinner enterprise support ecosystem, smaller talent pool	Cost sensitive estates under 500 virtual machines

9.2 Control Plane Components

Component	Function	Resilience
Orchestration	Placement, lifecycle, live migration, HA restart	Three node cluster, quorum based
Identity and tenancy	Organisations, projects, roles, quota, federation to corporate directory	Active-active, replicated
Self-service portal	Catalogue, request, approval, environment provisioning	Load balanced pair
API gateway	Programmatic access, rate limiting, audit of every call	Load balanced pair
Metering and showback	Per project consumption, cost allocation to business units	Primary with replica
Image registry	Golden images, signed, versioned and scanned	Replicated to DR
Secrets management	Credential and certificate custody, dynamic secrets	HA cluster with HSM seal

9.3 Tenancy and Quota Model

Level	Definition	Controls applied
Organisation	The enterprise boundary, one per legal entity	Global policy, key custody, audit scope

Level	Definition	Controls applied
Project	A business unit, product team or application group	Quota, budget, role assignment, network segment
Environment	Development, test, staging or production within a project	Approval workflow, backup class, change control
Workload	An individual instance, cluster or volume	Tags, microsegmentation policy, lifecycle owner

Quota is enforced at the project level across virtual CPU, memory, storage by tier, floating IP addresses, load balancers and Kubernetes clusters. Production quota changes require approval. Non-production quota is self-service within a ceiling, which is what makes the platform faster than the public cloud process it replaces.

9.4 Service Catalogue

Catalogue item	What the user gets	Target provisioning time
Virtual machine	Sized instance from a hardened golden image, joined to directory, monitored and backed up	Under 5 minutes
Kubernetes cluster	Managed control plane, node pools, ingress, registry access, storage class	Under 20 minutes
Managed database	PostgreSQL or MySQL with HA, automated backup and point in time recovery	Under 15 minutes
Object storage bucket	Versioned bucket with lifecycle policy and access keys	Under 2 minutes
Load balancer	Layer 4 or Layer 7 with certificate binding and health checks	Under 5 minutes
Virtual desktop pool	Non-persistent or persistent pool with profile management	Under 30 minutes
Full environment	Network, compute, database and load balancer from a single blueprint	Under 15 minutes

9.5 Golden Image Factory

- Images are built by pipeline on a weekly cadence, never by hand, and never edited in place.
- Every image is hardened to a CIS benchmark, vulnerability scanned, signed and versioned before publication.
- Old image versions are deprecated on a schedule, so long lived instances cannot pin an unpatched base indefinitely.
- Instances are treated as replaceable. Patching happens by redeploying from a newer image wherever the workload allows it.

9.6 Metering, Showback and Chargeback

Unit rates are published for virtual CPU hour, memory gigabyte hour, storage gigabyte month by tier, and egress. Consumption is metered per project and reported monthly. Showback runs from day one. Chargeback is optional and can be enabled once the business is comfortable with the numbers. Without this, the platform has no demand signal and capacity planning becomes guesswork.

9.7 Infrastructure as Code

- Bare metal provisioning, fabric configuration, hypervisor build and platform deployment are declared in code and version controlled.
- No production node is configured by hand. The build pipeline is the only path to production.

- Environment definitions are reproducible, which is what makes the DR site provably identical to primary.
- Drift detection runs continuously and raises an alert on any out of band change.
- These modules are carried over from Cloudtopiaa rather than written for this programme, which is where the largest schedule saving comes from.

SECTION 10

Platform Services: Containers, Databases and Virtual Desktops

Platform services are built at day one rather than bolted on later. Retrofitting them onto a virtualisation estate two years after go live is consistently more expensive than including them in the original build.

10.1 Kubernetes as a First Class Tenant

- Clusters are provisioned through the same self-service catalogue, with policy guardrails applied automatically at creation.
- Node pools draw from the same compute capacity pool, so there is no stranded container hardware.
- Container registry with vulnerability scanning and image signing enforced at admission control.
- Persistent volumes served by the same software defined storage through a CSI driver.
- Ingress, certificate management, service mesh and observability are delivered as platform defaults rather than as a per team problem.
- Cluster lifecycle, upgrade and node replacement are automated, so a Kubernetes version upgrade is a platform task rather than an application team task.

10.2 Database as a Service

Capability	Implementation
Engines	PostgreSQL and MySQL delivered through operators, with room to add engines later
High availability	Primary with synchronous standby, automated failover and connection routing
Backup	Continuous archiving with point in time recovery to any second within the retention window
Storage	Tier 0 volumes with quality of service guarantees, isolated from noisy neighbours
Access	Credentials issued dynamically by the secrets manager, short lived, never embedded in images
Observability	Query performance, replication lag and connection saturation exposed by default

10.3 Virtual Desktop Infrastructure

Element	Design
Scale	300 seats at go live, sized to 450 without additional compute purchase
Pool model	Non-persistent pools for standard users, persistent for engineering and design roles
Profile management	Profile containers on Tier 1 storage, separated from the desktop image
Graphics	Reserved capacity in the high density racks for GPU backed pools when CAD or design workloads require it
Access	Brokered through the DMZ with multi-factor authentication and device posture check
Storage impact	Modelled at boot storm peak rather than steady state, which is where most VDI deployments are undersized

Design decision: build the platform layer at day one

Kubernetes, managed databases and object storage are included in the 12 week build even where there is no funded consumer today. The marginal cost during the build is small because the control plane, storage and fabric are already being deployed. The cost of adding them in year two is a separate project with its own design, procurement and change window.

SECTION 11

Security, Identity and Compliance Architecture

Security is designed into the fabric and the control plane rather than added as a perimeter. The working assumption is that the perimeter will be crossed and that the blast radius is the thing under engineering control.

11.1 Control Layers

Layer	Controls implemented
Physical	Cage access control, biometric plus card, mantrap, CCTV retention, asset tagging and audit
Network	Zone segmentation by VRF, next generation firewall pair in HA, IPS, DDoS scrubbing at edge, TLS inspection at zone crossings
Workload	Microsegmentation at the virtual NIC, default deny east-west, policy by workload tag, host based intrusion detection
Identity	Federation to corporate directory, mandatory MFA, role based access with least privilege, no shared accounts
Privileged access	Privileged access management with session recording, just in time elevation, credential vaulting and rotation
Data	Encryption in transit and at rest, organisation held keys in HSM backed KMS, classification driven handling, DLP at egress
Application	Web application firewall, API gateway policy, secrets injected at runtime and never stored in images
Detection	Central log aggregation, SIEM correlation, behavioural analytics on privileged activity, defined incident runbooks

11.2 Key Custody

Encryption keys are held by the organisation in a hardware security module backed key management service. Keys are never held inside the storage platform or the hypervisor. This single design choice is what makes the sovereignty claim real rather than rhetorical, because it means no vendor and no operator can decrypt organisational data unilaterally. It is the same custody model Cloudtopiaa operates under.

11.3 Compliance Mapping

Framework	How this design supports it
DPDP Act 2023	Data residency within India, purpose limitation enforced through segmentation, consent artefacts logged immutably, breach detection and notification workflow, defined retention and erasure
ISO 27001	Control implementation evidenced through change records, access reviews, asset register and continuous log retention
ISO 22301	Documented business continuity plan, tested recovery objectives, quarterly drill evidence
Sector regulator	RBI, IRDAI or SEBI control mappings applied where the organisation is regulated, including localisation and audit access
Internal audit	Single evidence trail across identity, change, access and data movement, exportable on demand

Design decision: no shared administrative credential

Every administrative action is attributable to a named human or a named service identity. Break glass accounts exist, are vaulted, are alarmed on use, and require dual authorisation. This is the control that most often fails an audit in the second year, because the platform was built correctly and then operated with a shared account for convenience.

SECTION 12

Backup, Disaster Recovery and Business Continuity

Recovery objectives are set per application tier. A single blanket recovery target across the estate is either unaffordable or untrue, and usually both.

12.1 Recovery Objectives by Tier

Tier	Examples	RPO	RTO	Mechanism
Tier 0	ERP transaction database, identity	Under 5 minutes	Under 30 minutes	Near synchronous replication, automated failover
Tier 1	Customer facing and core line of business	15 minutes	2 hours	Asynchronous replication, orchestrated runbook failover
Tier 2	Internal applications, reporting, collaboration	4 hours	8 hours	Asynchronous replication, scheduled recovery
Tier 3	Development, test, archives	24 hours	72 hours	Restore from immutable object tier

12.2 Backup Architecture

- Three copies of data on two media classes, with one copy immutable and logically separated from the production control plane.
- Backup catalogue and credentials are isolated from the production directory, so a directory compromise does not compromise recovery.
- Immutability is enforced by retention lock at the object tier, not by policy inside the backup application.
- Restore is tested monthly on a sampled basis and reported as a metric, not assumed.

12.3 Failover Orchestration

Failover is executed through an orchestrated runbook that sequences dependency order, remaps networking, validates application health and reports completion. The runbook is version controlled alongside the infrastructure code. Manual failover from written documentation is excluded from this design because it fails under time pressure.

> **Contractual requirement: quarterly drills**
> Tech4Biz recommends writing quarterly failover drills into the operating agreement with a named accountable owner and a reported pass or fail result. Disaster recovery that has never been exercised is a capital expense with no verified return. The overwhelming majority of failed recoveries are failures of untested process, not of technology.

SECTION 13

Observability, Automation and Operations

A well designed platform degrades within eighteen months if nobody funded the operating model. This is the part of the programme most often cut, and it is the part that determines whether the investment holds.

13.1 Observability Stack

Signal	Coverage	Retention
Metrics	Facility, BMS, hardware, fabric, hypervisor, storage, platform, application	13 months at reduced resolution
Logs	All infrastructure and platform components, security events, API calls	12 months hot, 7 years cold where regulated
Traces	Platform services and instrumented applications	30 days
Topology	Dependency mapping across workloads and services	Continuous
Capacity	Per resource pool trending with forecast to threshold	36 months

13.2 Operating Model

Function	Responsibility	Indicative FTE
Platform engineering	Control plane, automation, image factory, infrastructure as code pipelines	3
Infrastructure operations	Hardware, fabric, storage, capacity, patching, facility liaison	2
Security operations	Monitoring, incident response, access review, compliance evidence	2
Service management	Change, problem, request, vendor and SLA management	1
On call rotation	24 by 7 Tier 1 and Tier 2 escalation	Drawn from the above

13.3 Capacity Governance

- Compute procurement trigger at 70 percent sustained utilisation of usable capacity.
- Storage procurement trigger at 65 percent, ahead of erasure code rebuild degradation.
- Power procurement trigger at 60 percent of contracted rack ceiling.
- Port and rack unit capacity tracked separately, because they exhaust before power does in a low density row.
- Quarterly capacity review with a rolling twelve month forecast presented to the CIO.

Long term view: fund the operating model or do not build

Eight funded roles are the difference between a platform that improves for five years and one that decays into an unpatched estate nobody wants to touch. If the operating model cannot be funded internally, Tech4Biz recommends contracting it as a managed Scale Pod rather than leaving it unstaffed. An unstaffed private cloud is more expensive than public cloud within three years, because the failures are unbudgeted.

SECTION 14

Workload Discovery and Migration Plan

Migration is where private cloud programmes lose their schedule. The remedy is discovery that starts during procurement rather than at build kickoff, and wave based cutover with rehearsal.

14.1 Discovery

- Automated agentless discovery starts during the procurement window, well before build week 1, and runs continuously through the build. It captures peak rather than average.
- Dependency mapping by observed network flow, not by interview. Interview based dependency maps are consistently incomplete.
- Application tiering workshop with business owners to assign recovery objectives and migration priority.
- Licence and support impact assessment, since some ISV terms change materially on a new hypervisor.
- Output is a signed workload inventory that becomes the baseline for the bill of materials and the migration waves.

14.2 Migration Waves

Wave	Content	Method	In the 12 weeks
Wave 0	Platform validation workloads, synthetic and internal tooling	Rebuild on target	Yes, week 9
Wave 1	Development and test environments	Rebuild from code, cut over in place	Yes, week 10
Wave 2	Tier 2 internal applications	Block level replication with a short cutover window	Yes, weeks 11 to 12
Wave 3	Tier 1 line of business applications	Replication, rehearsal, weekend cutover with rollback	No, weeks 13 to 17
Wave 4	ERP and Tier 0 systems	Rehearsed twice, business sign off, extended freeze window	No, weeks 16 to 20
Wave 5	Virtual desktop estate	Pool rebuild, phased user migration by department	No, weeks 14 to 18

Waves 3 to 5 run after the 12 week platform acceptance, under hypercare. This is deliberate and it is the honest way to hold a 12 week commitment. Tier 0 and Tier 1 cutover is governed by business freeze windows and rehearsal outcomes rather than by the infrastructure schedule, and compressing it would trade platform risk for business risk.

14.3 Cutover Discipline

- Every wave has a written rollback that is tested before the cutover window opens, not drafted during it.
- Every Tier 1 and Tier 0 cutover is rehearsed end to end in a non-production copy before the live event.
- A go or no go checkpoint is held 48 hours before each window with named accountable signatories.
- Source environments are retained in a recoverable state for 30 days after cutover before decommissioning.

SECTION 15

12 Week Implementation Plan

Twelve weeks from build kickoff to an accepted production platform, including the disaster recovery site and a proven failover drill. The clock starts when hardware is on site. Procurement runs ahead of it and is not counted inside it.

15.1 Precondition: Hardware on Site

Hardware procurement carries a lead time set by the market, not by the delivery team. It sits ahead of the 12 weeks as a separate phase with its own schedule and its own risk. Quoting a single number that folds procurement into the build produces a date nobody can hold, and it moves the blame to the wrong place when a switch ships late.

Item	Indicative lead time	Compression available
Compute nodes	8 to 12 weeks	Partner stock, pre-qualified alternate SKUs
100G and 25G switching	10 to 16 weeks	Alternate vendor qualified in advance
High capacity NVMe media	8 to 14 weeks	Split order across two suppliers
Security appliances	6 to 10 weeks	Virtual form factor as an interim
Racks, PDU, cabling	4 to 8 weeks	Locally sourced
GPU accelerated nodes	16 to 30 weeks	Little. Order separately, do not gate the build on it

The realistic total is 10 to 16 weeks from purchase order to full delivery at both sites, assuming no GPU nodes in the day one order. Tech4Biz recommends releasing the purchase order against a frozen bill of materials and running discovery, facility contracting and design validation in parallel during that window, so that build week 1 starts with a signed design and a ready cage.

15.2 Programme Shape

Phase	Duration	What happens	Counted in the 12 weeks
A Procurement and preparation	10 to 16 weeks	Purchase order released, hardware manufactured and shipped, workload discovery runs, colocation contracted, design validated	No
B Build and acceptance	12 weeks	Everything in Section 15.3, ending in a signed acceptance and an operations team certified to run the platform	Yes
C Migration and hypercare	Weeks 13 to 24	Migration waves 3 to 5, ninety day embedded hypercare, second failover drill, legacy decommissioning	No

Phase A and Phase C are real work with real cost. They are excluded from the 12 week figure because the 12 weeks is a build commitment, and a build commitment should only cover what the delivery team controls.

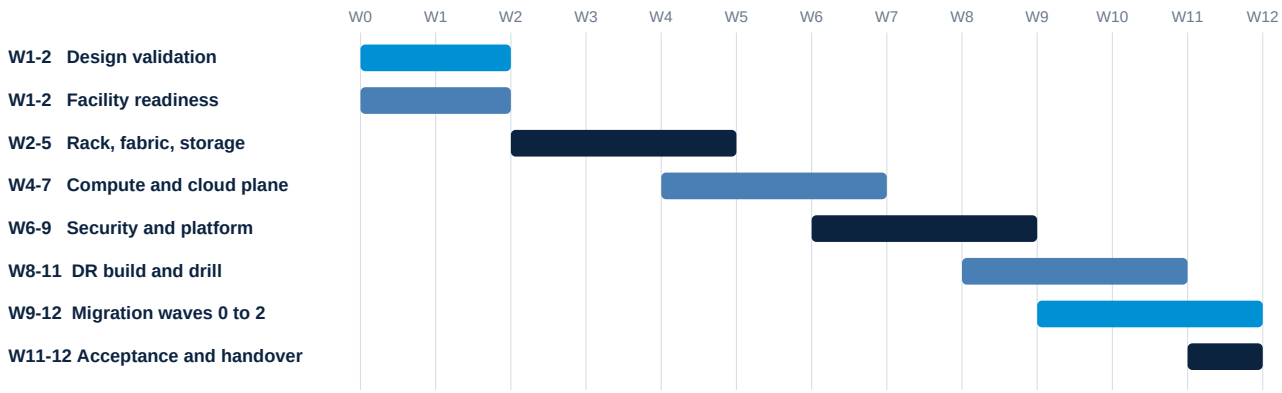

Figure 3. The 12 week build phase, starting from hardware on site. Workstreams overlap deliberately. No stream waits for the one above it to finish.

15.3 Week by Week Detail

Weeks	Workstream	Key activities	Exit criteria
1 to 2	Design validation and freeze	Discovery output reconciled against the delivered hardware, low level design confirmed, IP plan, platform stack decision closed	Signed low level design, change control in force
1 to 2	Facility readiness	Cage build completed, power and cooling validated against Section 04, commissioning evidence reviewed, carriers live	White space accepted, DCIM reporting
2 to 5	Rack, fabric and storage	Rack and stack, structured cabling, spine and leaf built from code, storage cluster deployed and tuned	Fabric and storage acceptance tests passed
4 to 7	Compute and cloud control plane	Hypervisor deployment, control plane, tenancy and quota model, self-service portal, API, metering and showback	Self-service provisioning demonstrated end to end
6 to 9	Security and platform services	Firewall policy, microsegmentation, IAM federation, KMS and secrets, SIEM integration, Kubernetes, database service, VDI	Security acceptance and penetration test passed
8 to 11	DR build and drill	DR site built from the same code, replication configured, orchestrated runbooks, first full failover drill	Documented failover within stated RTO and RPO
9 to 12	Migration waves 0 to 2	Platform validation workloads, development and test, Tier 2 internal applications	Waves 0 to 2 live on the platform
11 to 12	Acceptance and handover	Full acceptance test suite, runbook handover, operations training, hypercare start	Signed acceptance, operations team certified

15.4 What Sits Outside the 12 Weeks

Activity	Window	Why it sits outside
Hardware procurement and delivery	10 to 16 weeks before week 1	Lead time is a market condition, not a delivery choice. The build clock starts on receipt
Workload discovery	Runs during procurement	Needs elapsed time to capture peak, and it costs nothing to run early
Migration waves 3 to 5	Weeks 13 to 20	Tier 0 and Tier 1 cutover is governed by business freeze windows and rehearsal, not by infrastructure readiness

Activity	Window	Why it sits outside
Hypercare	Weeks 12 to 24	Ninety days of embedded Tech4Biz support after platform acceptance
Second failover drill	Week 20	Proves the platform recovers with production workloads on it, not only at build state
Legacy decommissioning	Weeks 20 to 24	Source environments are held recoverable for 30 days after each cutover

State the clock start in the contract

Tech4Biz recommends that the agreement defines week 1 as the first Monday after full hardware receipt and asset tagging at both sites, confirmed in writing by both parties. Without that definition, a shipping delay becomes a delivery dispute. With it, the 12 week commitment is something the delivery team can actually be held to.

Critical path inside the 12 weeks

Three dependencies carry the plan once the clock starts. The cage is ready by week 2. The design is frozen at the end of week 2 and is not reopened. A single named client decision owner can sign without a committee cycle. A 12 week plan has no float, so a one week slip in any of them is a one week slip in go live.

SECTION 16

Testing, Acceptance and Handover Criteria

Acceptance is defined before the build starts. Every test below has a written pass threshold, an evidence artefact and a named signatory. The test suite is carried over from prior Tech4Biz deployments rather than written during this programme.

Test	What is proven	Pass threshold
Power path failure	Loss of A feed does not interrupt any workload	Zero workload interruption across a full rack row
Cooling failure	Loss of one cooling unit holds inlet temperature	Inlet below 27 C for 30 minutes at full load
Fabric resilience	Loss of one spine and one leaf holds connectivity	No packet loss beyond 200 ms convergence
Node failure	HA restarts workloads on surviving capacity	All Tier 1 workloads restarted within 5 minutes
Storage node failure	Cluster survives node loss and rebuilds	No data loss, rebuild within SLA, latency within 20 percent
Storage performance	Tiers meet stated IOPS and latency	Tier 0 sub millisecond at 200k IOPS sustained
Live migration	Workloads move without interruption	Zero dropped sessions across 100 migrations
Self-service provisioning	Environment delivered without a ticket	Under 15 minutes from request to usable
Kubernetes cluster build	Cluster provisioned with guardrails applied	Under 20 minutes, policy enforced at admission
Backup and restore	Recovery of a full application stack	Restore within stated RTO from an immutable copy
Failover drill	Full site failover to DR	All Tier 0 and Tier 1 within stated RTO and RPO
Penetration test	External and internal security posture	No critical or high findings unremediated at handover
Segmentation validation	East-west policy enforcement	Zero unauthorised flows across zone boundaries

16.1 Handover Package

- As built low level design, updated to reflect the delivered state rather than the designed state.
- Full infrastructure as code repository with pipeline documentation and access handover.
- Operational runbooks for every routine and every failure scenario tested above.
- Facility commissioning evidence pack, including load bank data and integrated systems test results.
- Capacity model with the forecast and procurement triggers pre-populated.
- Training delivered to the named operations team, with competence assessed rather than assumed.
- Ninety day hypercare with Tech4Biz engineers embedded and a defined exit checklist.

SECTION 17

Commercial Model and Five Year Total Cost of Ownership

Figures below are indicative and sized against the Section 01 baseline. They are confirmed against live vendor quotations at design freeze in week 2. They are presented as ranges because component pricing and currency both move.

17.1 Indicative Capital Expenditure

Component	Primary site	DR site	Notes
Compute nodes	INR 2.4 to 2.9 Cr	INR 1.2 to 1.5 Cr	17 nodes primary, 11 DR
Storage nodes and media	INR 1.3 to 1.6 Cr	INR 0.9 to 1.1 Cr	NVMe pricing is the largest variable
Network fabric and optics	INR 1.3 to 1.6 Cr	INR 0.8 to 1.0 Cr	Spine, leaf, border, out of band, cabling
Security appliances	INR 0.5 to 0.7 Cr	INR 0.3 to 0.4 Cr	Firewall and load balancer pairs
Racks, PDU, DCIM, KVM	INR 0.2 to 0.3 Cr	INR 0.1 to 0.2 Cr	Includes environmental instrumentation
Software licences, year one	INR 1.0 to 1.4 Cr	INR 0.4 to 0.6 Cr	Highly dependent on the Section 9.1 selection
Professional services	INR 1.2 to 1.6 Cr	Included	Design, build, commissioning, migration, hypercare
Indicative total	**INR 7.9 to 10.1 Cr**	**INR 3.7 to 4.8 Cr**	**Programme total INR 11.6 to 14.9 Cr**

17.2 Indicative Annual Operating Expenditure

Item	Annual	Basis
Colocation, primary	INR 2.4 to 2.9 Cr	20 racks at 12 kW average, power inclusive
Colocation, DR	INR 1.1 to 1.4 Cr	10 racks at 10 kW average
Connectivity	INR 0.5 to 0.7 Cr	Internet, inter site, carrier diversity
Hardware support and AMC	INR 0.8 to 1.0 Cr	Approximately 10 percent of hardware capital
Software subscription renewal	INR 0.9 to 1.3 Cr	Section 9.1 selection dependent
Operations team	INR 1.4 to 1.8 Cr	Eight FTE, internal or contracted Scale Pod
Indicative total	**INR 7.1 to 9.1 Cr**	**Per year at steady state**

17.3 Five Year Comparison

Scenario	Year 1	Years 2 to 5	Five year total
Private cloud, this design	INR 19 to 24 Cr	INR 28 to 36 Cr	INR 47 to 60 Cr
Equivalent public cloud	INR 13 to 17 Cr	INR 58 to 76 Cr	INR 71 to 93 Cr
Indicative delta	**Public cloud lower**	**Private cloud lower**	**Saving of 30 to 38 percent**

The crossover point falls between month 26 and month 32. Public cloud remains cheaper through year one because it carries no capital event. The private cloud advantage compounds from year two onward, and it widens

further if data egress volumes or licence entitlements grow, which are the two costs that escalate least predictably in a public cloud contract.

17.4 Payment Milestones

Milestone	Trigger	Share of services value
M0 Bill of materials freeze	Purchase order released, start of the procurement phase	20 percent
M1 Design freeze	End of build week 2, signed low level design and cage accepted	15 percent
M2 Infrastructure acceptance	End of build week 5, fabric and storage acceptance passed	15 percent
M3 Platform and DR acceptance	End of build week 11, security tests and first failover drill passed	30 percent
M4 Handover complete	End of build week 12, acceptance signed and operations team certified	20 percent

SECTION 18

Risk Register and Designed Failure Modes

The risks below are the ones that actually end these programmes. Each has a designed mitigation rather than an acknowledgement.

Risk	Impact	Likelihood	Mitigation designed into this build
Hardware lead time overruns	Build start date moves, 12 week window shifts with it	High	Procurement excluded from the 12 weeks by design, pre-qualified SKU list, partner stock, alternate SKUs, split orders
Partial delivery treated as start	Build begins without a full set, stalls mid stream	Medium	Contract defines week 1 as full receipt and asset tagging at both sites, confirmed in writing
Design reopened after week 2	Cascading delay across every parallel workstream	High	Single named client decision owner, formal change control with cost and schedule impact stated
Power envelope undersized	Expansion blocked in year three, forced site move	High	Contract for 1.8 times day one draw with a written uprate clause
Hypervisor licence repricing	Operating cost step change at renewal	High	Five year licence modelling in week 1, open platform kept as a live alternative
Discovery understates workload	Compute or storage undersized at cutover	Medium	Discovery from day 1 capturing peak, 30 percent headroom, expansion without redesign
DR never exercised after handover	Recovery fails at the moment of need	High	Quarterly drills contracted, orchestrated runbook, reported pass or fail
Operations model unfunded	Platform decays within 18 months	High	Eight FTE model costed in Section 17, managed Scale Pod alternative offered
Storage vendor lock	Forklift replacement at end of life	Medium	Software defined storage on commodity hardware, no proprietary controller
Skills gap at handover	Operations team cannot run the platform	Medium	Competence assessed rather than assumed, 90 day hypercare, runbooks exercised by the client team
Ransomware reaching backups	Recovery impossible, extortion leverage	Medium	Immutable object tier with retention lock, backup credentials isolated from the production directory
Application incompatibility	Migration wave blocked	Medium	ISV certification checked in week 1, rehearsal before every Tier 1 cutover

Risk	Impact	Likelihood	Mitigation designed into this build
Facility accepted without load bank	Design load never proven, failure surfaces in year two	Medium	Level 5 integrated systems test with load bank data required before white space acceptance

The three failures worth planning against most

Across corporate private cloud programmes, three failures recur. The power ceiling arrives earlier than anyone modelled. The licence renewal costs more than the original purchase. And nobody funded the team, so a well engineered platform is unpatched within two years. All three are commercial failures rather than technical ones, and all three are cheapest to fix before the design is frozen in week 2.

SECTION 19

Governance, Responsibility Matrix and Next Steps

19.1 Responsibility Matrix

Workstream	Tech4Biz	Client	Third party
Low level design	Accountable	Approve	Consulted
Facility contract	Consulted	Accountable	Colocation provider delivers
Facility commissioning witness	Accountable	Informed	Provider or contractor executes
Procurement	Responsible	Accountable for purchase orders	OEM and distributor deliver
Build and integration	Accountable	Informed	OEM support consulted
Security policy definition	Responsible	Accountable	Auditor consulted
Application migration	Responsible	Accountable for business sign off	ISV consulted
Acceptance testing	Responsible	Accountable for sign off	Independent tester for penetration test
Steady state operations	Consulted or contracted	Accountable	OEM support

19.2 Governance Cadence

- Daily standup during weeks 1 to 3 and weeks 9 to 12, the two periods with no float.
- Weekly delivery review with the joint programme team, risks and blockers only.
- Fortnightly steering committee with the CIO, covering schedule, spend and decisions required.
- Formal design authority sign off at each acceptance gate, with named signatories.

19.3 What Tech4Biz Needs to Proceed

Input required	Why it matters	Needed by
Named decision owner with signing authority	A 12 week plan cannot absorb a committee cycle on design or acceptance	Kickoff
Purchase order release against a frozen bill of materials	Starts the procurement clock. The 12 week build clock starts on delivery, not on this	As early as possible
Confirmed rack count and kW per rack	Drives facility contract, bill of materials and every cost figure in Section 17	Before purchase order
Decision on accelerated compute in scope	Determines density reservation and cooling readiness. GPU lead times can reach 30 weeks, so order separately	Before purchase order
Build versus colocate decision	Colocation is assumed. Building the facility changes the programme from 12 weeks to roughly 40	Kickoff
Access for discovery tooling	Produces the workload inventory the bill of materials is sized against. Runs during procurement	Before purchase order

Input required	Why it matters	Needed by
Application tier assignment	Sets recovery objectives and migration wave order	Week 2
Regulatory scope confirmation	Determines control set, residency and audit obligations	Week 1

19.4 Immediate Next Step

Tech4Biz proposes a two week paid discovery and design sprint. It produces the confirmed workload inventory, the facility shortlist with commercial terms, the frozen low level design and a bill of materials priced against live vendor quotations. The output is a document the organisation can take to its board with real numbers rather than ranges, and it is what the purchase order is released against. Procurement then runs for 10 to 16 weeks while discovery and facility contracting continue in parallel. The 12 week build clock starts the Monday after hardware lands.

Prepared by	Tech4Biz Solutions Pvt Ltd
Registered office	Bengaluru, Karnataka, India
Web	tech4biz.in
Reference deployments	Cloudtopiaa sovereign IaaS, VVCE Private Cloud Lab Ecosystem
Validity	This proposal is valid for 45 days from the date of issue
Classification	Confidential. Not for distribution outside the recipient organisation