

Media service (SFU / MCU) — AI deployment prerequisites

Media SFU/MCU — AI deployment prerequisites (HVM vs Docker, BE isolation)

Table of contents

1. Executive answer — HVM vs Docker vs monolith BE
Recommendation (DR-aligned)
2. What you give the AI on day zero
 - 2.1 Access
 - 2.2 Machine identity
 - 2.3 PO-owned (snapshots, DNS, firewall)
 - 2.4 Integration secrets (existing BE — read-only API)
 - 2.5 Decisions PO should state (or defaults)
3. Recommended instance layout (up to 3 HVMs)
 - 3.1 Lab — single HVM (minimum handoff)
 - 3.2 Prod-shaped — three HVMs
4. Network, DNS, and firewall prerequisites
 - 4.1 Traffic paths
 - 4.2 Ports to plan (PO opens on FE/router → media HVM)
 - 4.3 DNS names (examples)
5. Integration with existing BE (isolation model)
6. SFU vs MCU stack choice
7. Packages and software (installed later by AI)
 - 7.1 Host (Alpine packages)
 - 7.2 Container images (typical compose)
 - 7.3 Repo deliverables (AI produces after deploy)
8. Sizing and bandwidth (200/200 Mbit context)
9. AI workflow after prerequisites are met
10. Out of scope for first media PoP
11. PO checklist (copy before handoff)
12. Changelog

Source: 20260612-MEDIASERVICE-AI-requirements.md — section links work in PDF viewers that support internal anchors.

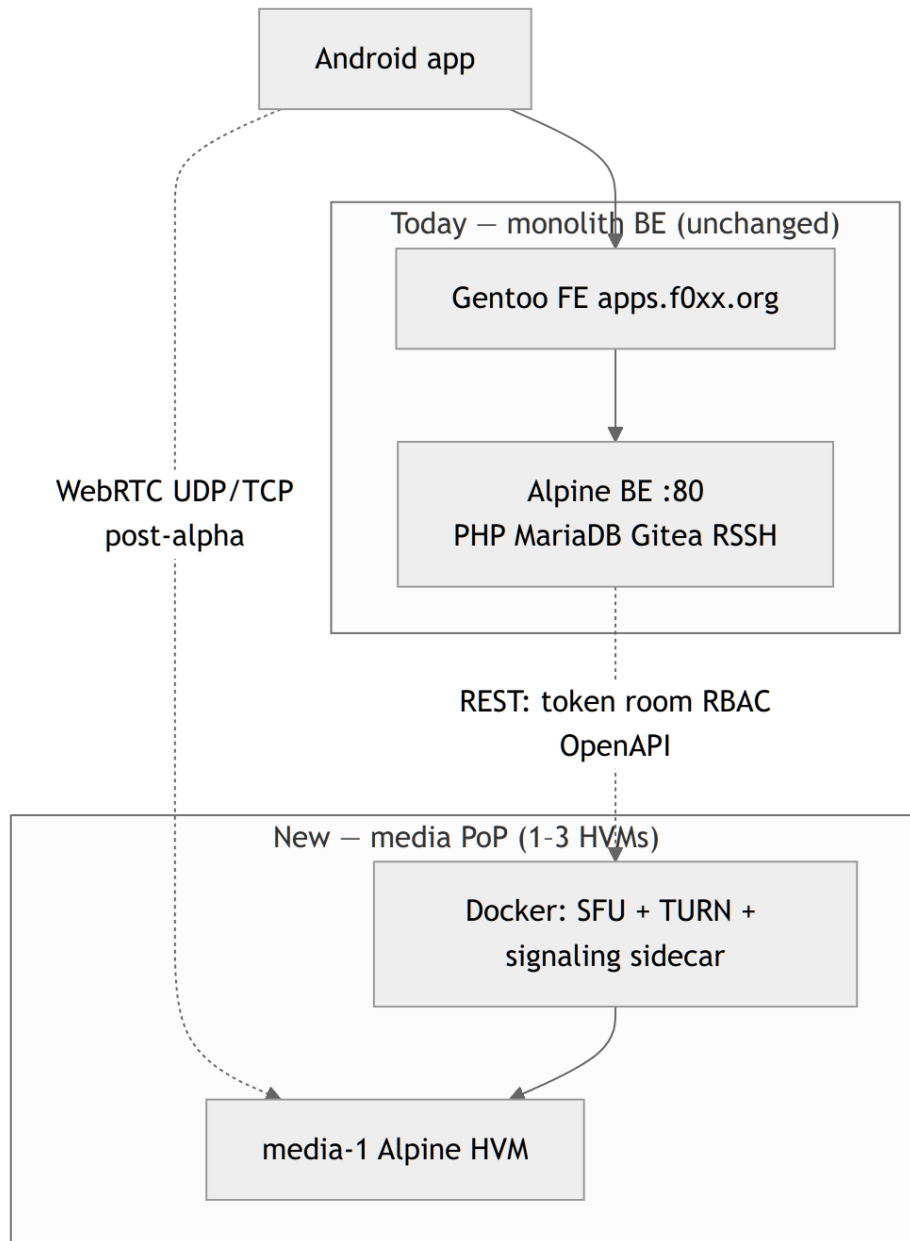
Document type: Requirements / handoff (pre-SPEC)
Date: 2026-06-12
Audience: PO / sysop → AI agent for full deploy + tune on fresh Alpine HVM(s)
Related: docs/specs/20100612_1_scaling.md, docs/DRs/20100612_1_scaling.md, examples/sfu/README.md, docs/20260608_BE_SERVICES_and_infra.md §SFU
PDF: tmp/20260612-MEDIASERVICE-AI-requirements.pdf — regenerate with .venv-pdf/bin/python docs/_pdf_build/md_to_pdf.py on the copy under tmp/
Code names (product): SPIDERWEB = SFU-alike, SPIDER = MCU-alike (draft glossary)

1. Executive answer — HVM vs Docker vs monolith BE

Recommendation (DR-aligned)

Approach	Verdict	Why
SFU/MCU on current androidcast BE monolith	No	PHP/MariaDB/RSSH/Gitea share CPU/RAM/NIC; pm.max_children=5 today; blast radius; INFRA.md keeps :8089 as other Janus stack
Fresh dedicated Alpine HVM(s)	Yes (primary)	Isolated media plane ; snapshots; UDP/WebRTC without touching hub/crashes
Docker on dedicated HVM	Yes (preferred delivery)	Reproducible Janus/mediasoup/coturn/ffmpeg; easy rollback; same host still “empty” at handoff
Docker on existing BE	No	Does not isolate bandwidth or failure domain

Practical model: PO provides **1–3 empty Alpine HVMs** with SSH + DNS. AI installs **Docker (+ compose)** on them and runs **SPIDERWEB (SFU)** or **SPIDER (MCU)** stacks in containers. **Control plane** (signaling auth, room lifecycle, RBAC) talks to existing crashes BE via **HTTPS/OpenAPI-style APIs** — no co-location of MariaDB or PHP-FPM with media.



Not too heavy: one HVM + Docker Compose is enough for **lab/stress**; three HVMs match **prod-shaped** split (SFU / TURN / transcode). Start with 1; split when metrics or PO asks.

2. What you give the AI on day zero

Minimum handoff — **no media packages required yet** on the empty machine.

2.1 Access

Item	Required	Example
SSH target	Yes	ssh media@media-lab.example.org or intra 10.7.x.x
Sudo	Yes	passwordless sudo for apk, rc-service, docker
SSH key	Yes	laptop → HVM; optional: HVM → git read for deploy scripts
Jump host	If no public SSH	ProxyJump via f0xx-monstro / alpine-be (document in ~/.ssh/config)

2.2 Machine identity

Item	Required	Notes
Hostname	Yes	e.g. media-1.intra.raptor.org
OS	Yes	Alpine 3.x x86_64 (match prod BE)

vCPU / RAM	Yes	See §8
Root disk	≥ 40 GiB SSD	containers + logs
Intra IP	Yes	10.7.0.0/8 address stable across reboots
Public reach	For external test	DNS name OR FE proxy only (see §4)

2.3 PO-owned (snapshots, DNS, firewall)

Item	Owner	AI needs
VM snapshots	PO	"Snapshot before AI session" + rollback name
External DNS	PO	A/AAAA or CNAME for WebRTC/signaling test
Firewall / DNAT	PO	UDP/TCP port list from §4 opened on router/FE
TLS certs	PO or AI via DNS-01	Let's Encrypt on FE or on media host

2.4 Integration secrets (existing BE — read-only API)

Item	Required	Purpose
Crashes base URL	Yes	https://apps.f0xx.org/app/androidcast_project/crashes
Service account or API token	Yes	Mint rooms / validate session (new sfu API or extend OpenAPI)
RBAC model	Reference	Same pattern as remote access / short links
Gitea / deploy	Optional	Pull android_cast branch feature/sfu-relay or deploy tarball

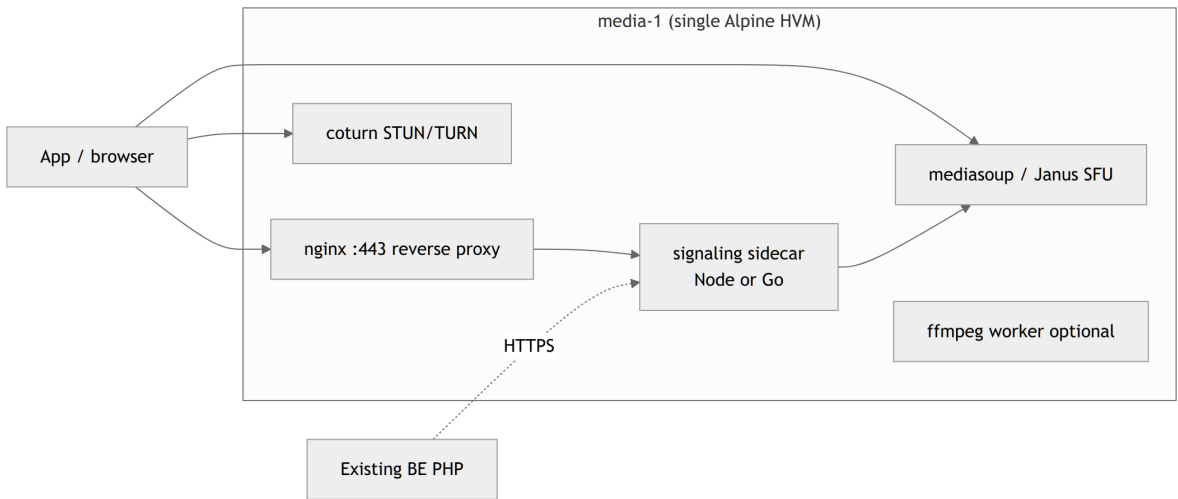
AI does **not** need MariaDB root on monolith BE if all integration is **HTTP APIs**.

2.5 Decisions PO should state (or defaults)

Decision	Default if silent
Topology code name	SPIDERWEB (SFU) first; MCU later
Engine	mediasoup (greenfield) — Janus on :8089 is other stack, avoid unless PO picks Janus on new host
Instance count	1 for lab; 3 for prod-shaped
Public path	Subdomain media.apps.f0xx.org → FE → media HVM (keeps TLS at FE)
TURN	coturn on same HVM (lab) or instance 2 (prod)

3. Recommended instance layout (up to 3 HVMs)

3.1 Lab — single HVM (minimum handoff)

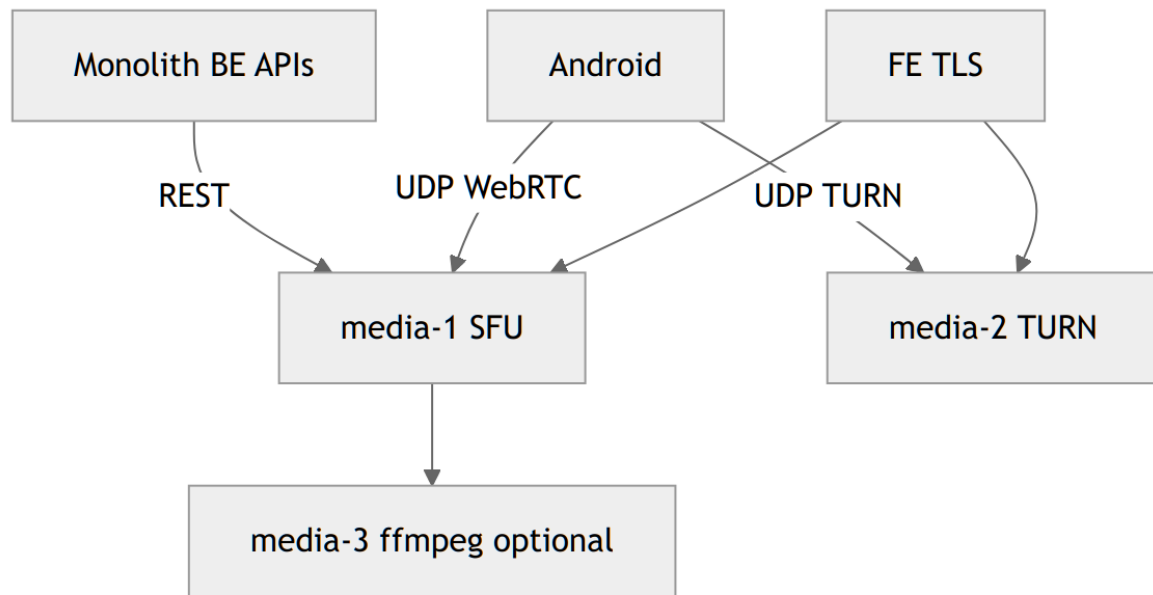


All services in **one docker-compose** stack; fine for stress and first WAN cast tests.

3.2 Prod-shaped — three HVMs

VM	Role	Containers / daemons
media-1	SPIDERWEB SFU	mediasoup (or Janus), nginx edge, signaling sidecar
media-2	TURN + signaling	coturn, optional dedicated signaling if scaled

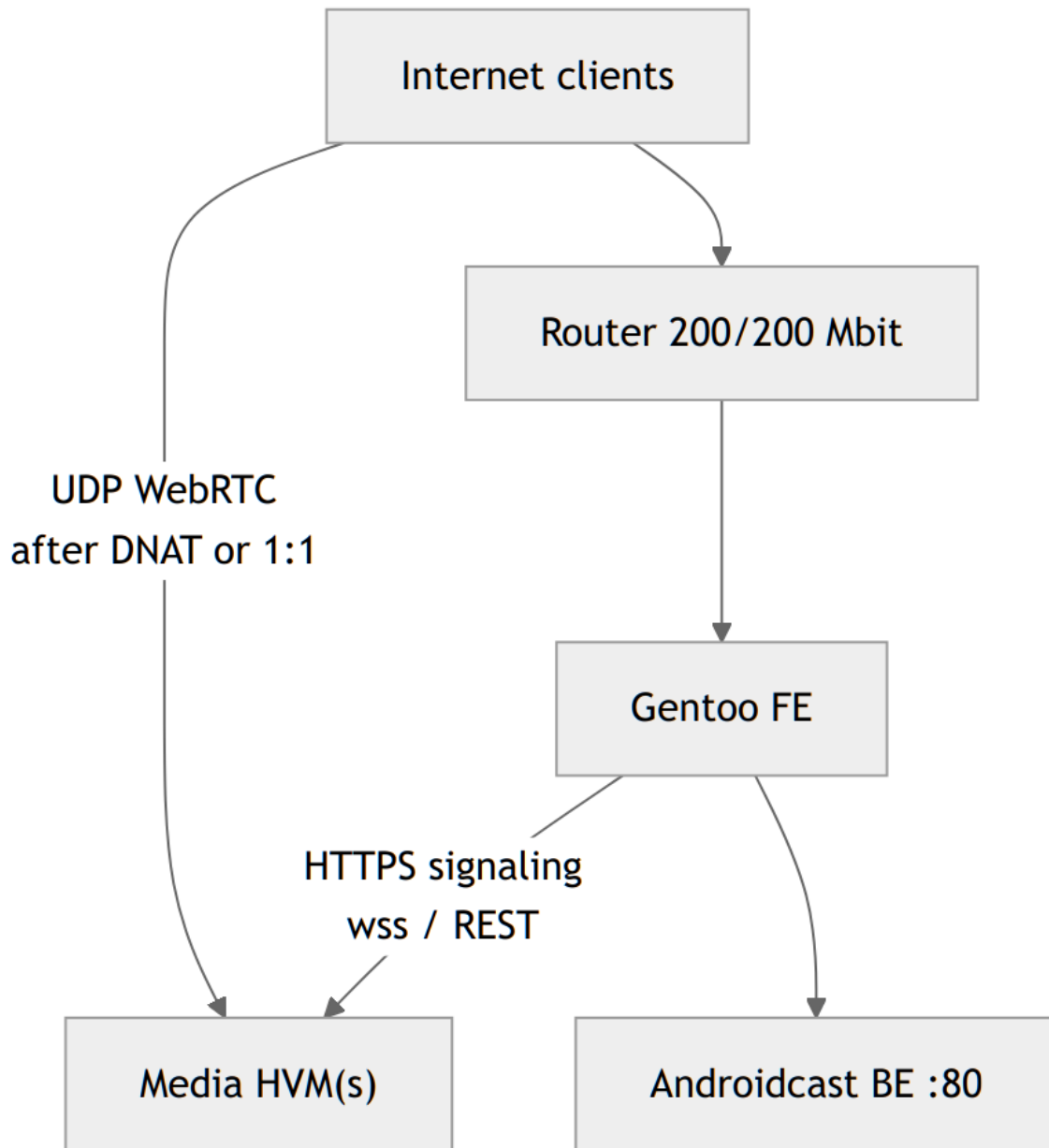
media-3	Transcode / MCU path	ffmpeg workers, optional MCU (SPIDER) if not SFU-only
---------	----------------------	---



PO can hand **one** machine first; AI documents split boundaries so instances 2–3 are **additive**, not rewrite.

4. Network, DNS, and firewall prerequisites

4.1 Traffic paths



Do not attach media to androidcast :8089 vhost (INFRA.md §2).

4.2 Ports to plan (PO opens on FE/router → media HVM)

Port / proto	Service	Notes
443/tcp	HTTPS / WSS signaling	Via FE reverse proxy recommended
3478/udp,tcp	STUN/TURN (coturn)	Required for many NATs
49152–65535/udp	WebRTC media (ICE)	Range size depends on engine; document in SFU SPEC
8080/tcp	SFU HTTP API (internal)	Not public; FE or intra only
22/tcp	SSH	Intra or VPN only in prod

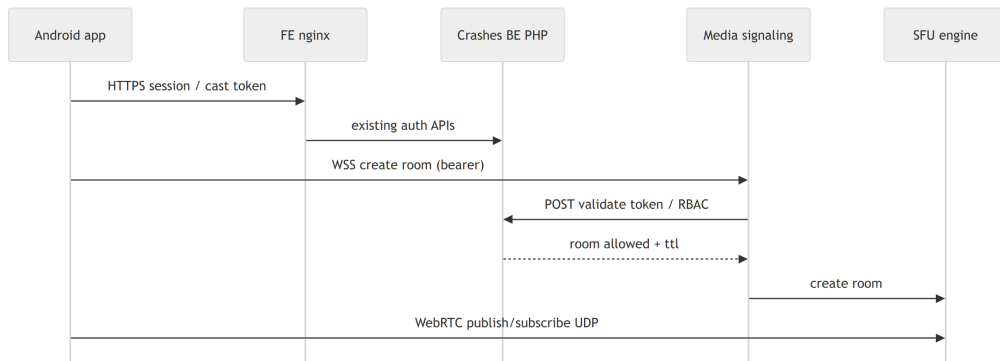
Exact ICE port range: set in mediasoup/Janus config during deploy; PO must allow **UDP range** on path to media HVM.

4.3 DNS names (examples)

Name	Points to	Use
media.apps.f0xx.org	FE → proxy → media-1	Public signaling + health
turn.apps.f0xx.org	FE or media-2	TURN URI in app
media-lab.intra.raptor.org	10.7.x.x	AI + sysop intra testing

5. Integration with existing BE (isolation model)

Media HVM **must not** share MariaDB or PHP-FPM with crashes console. Integration is **API-only**:



Integration	Method	Existing artifact
Health gate	GET	api/sfu_health.php (stub today)
Feature flag	config	config.php → sfu.enabled, URLs
Room mint / revoke	New REST	OpenAPI under crashes or media-api submodule
Auth	Bearer + session RBAC	Same as remote access / short links pattern
Metrics	POST ingest	Optional → graph_upload.php or new series
No direct DB	—	Media DB (if any) is local to HVM (Redis/SQLite)

OpenAPI: extend examples/crash_reporter/backend or new backend/media-signaling repo later — AI needs **written** API contract before prod wire-up.

6. SFU vs MCU stack choice

	SPIDERWEB (SFU)	SPIDER (MCU)
Use	Multiparty cast; subscribers receive per-stream	Mix down; single output; higher CPU per room
Engine candidates	mediasoup (preferred greenfield), Janus VideoRoom	Janus MCU plugin, ffmpeg bridge, custom
CPU	Moderate per stream; scales with subscribers	High per room (encode)
Latency	Lower	Higher
First milestone	Yes	Phase 2 / media-3

Janus on existing :8089: treat as **foreign** — do not piggyback androidcast nginx. New Janus instance on **media HVM** only if PO rejects mediasoup.

Docker vs bare metal on HVM: use **Docker** for SFU/TURN/signaling; host only needs Alpine + docker + nginx + firewall tools.

7. Packages and software (installed later by AI)

PO delivers **almost empty** Alpine; AI installs when handoff complete.

7.1 Host (Alpine packages)

```

openssh sudo bash curl wget git ca-certificates
nginx openssl iptables ip6tables
docker docker-cli-compose
tcpdump bind-tools jq python3

```

7.2 Container images (typical compose)

Service	Image / build	Role
mediasoup or janus	Official or pinned Dockerfile in repo	SFU
coturn	coturn/coturn	STUN/TURN
signaling	Custom Node/Go from repo	Rooms, BE API client
nginx	nginx:alpine or host nginx	TLS termination at edge or intra
ffmpeg	jrottenberg/ffmpeg or alpine	Transcode sidecar (media-3)
redis	redis:alpine	Room state optional

7.3 Repo deliverables (AI produces after deploy)

Artifact	Location (proposed)
docker-compose.yml	examples/sfu/deploy/ or backend/media-pop/
nginx FE fragment	examples/sfu/deploy/nginx.fe-media.snippet

BE config keys	config.example.php → sfu.* (already stubbed)
Smoke script	examples/sfu/scripts/smoke_sfu.sh
OpenAPI	examples/sfu/openapi.yaml

8. Sizing and bandwidth (200/200 Mbit context)

WAN **200/200 Mbit** sits behind FE; media UDP may be **DNAT** to media HVM.

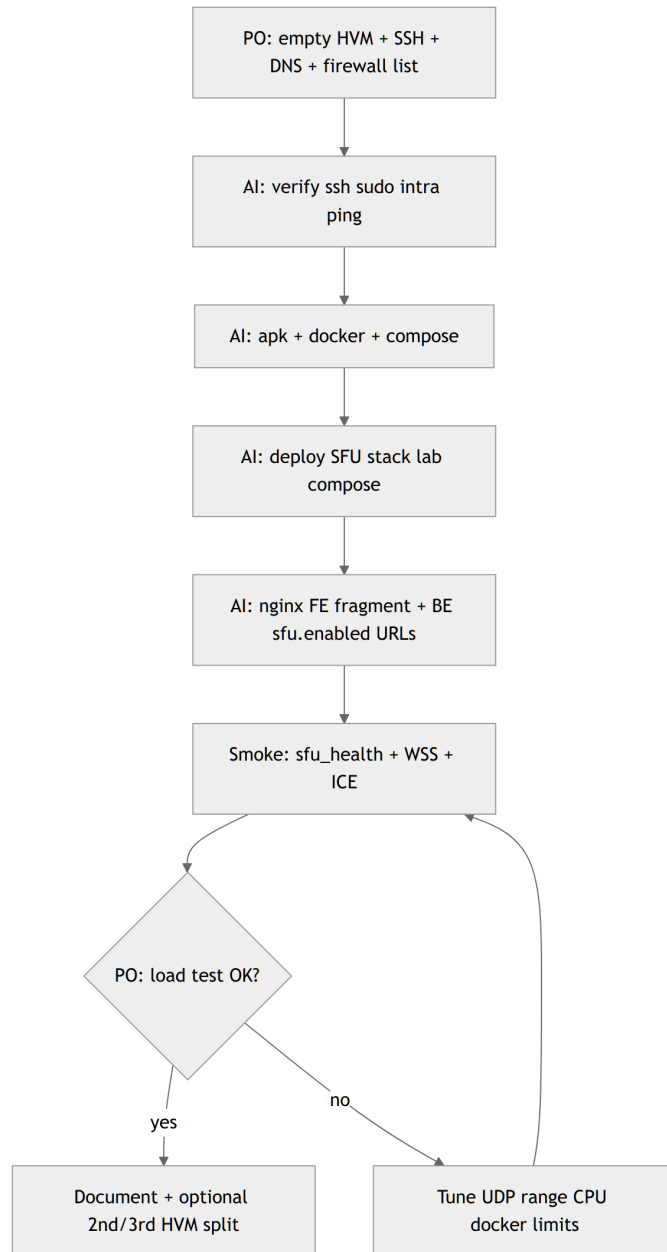
Profile	vCPU	RAM	Notes
Lab / stress (1 HVM)	4–8	8–16 GiB	SFU + TURN + signaling; ~10–30 parallel WebRTC streams
Prod SFU node	8+	16 GiB	Dedicated NIC budget; see SPEC §12.3
TURN only	2	4 GiB	Mostly relay bandwidth
ffmpeg node	8+	16–32 GiB	CPU-bound transcodes

Bandwidth rule of thumb (SFU):

Streams	Approx WAN need
10 × 1 Mbps video	~10–15 Mbps + overhead
50 × 1 Mbps	~50–80 Mbps — approaches 200 Mbit shared with rest of estate
100+ viewers	CDN/SFU PoP or second site — not single 200 Mbit

Stress test: measure on **intra** first, then **public DNS** through FE.

9. AI workflow after prerequisites are met



Step	AI action	PO action
0	—	Snapshot empty VM
1	Probe CPU/RAM/NIC, apk update	Confirm DNS/firewall
2	Install docker, compose, pin versions	—
3	Bring up mediasoup + coturn + signaling	—
4	FE nginx location /sfu/ or subdomain proxy	Reload FE nginx
5	Enable sfu.enabled + URLs on monolith BE only	Sync config.php
6	Run smoke + optional load/WebRTC client	Review metrics
7	Snapshot “golden” media PoP	—

10. Out of scope for first media PoP

- Replacing LAN UDP cast (alpha path unchanged)
- Moving crashes/tickets/Gitea to media HVM
- Full GEO policy (SPEC §14) — later with F3/F10
- Production payments / recording storage
- Merging with Janus stack on BE :8089

11. PO checklist (copy before handoff)

[] Alpine 3.x HVM provisioned (1-3); specs noted
[] SSH + sudo for AI; key installed
[] Intra IP 10.7.x.x stable
[] External DNS name for test (or FE-only proxy agreed)
[] Firewall: UDP 3478 + WebRTC range → media HVM
[] Snapshot taken (empty baseline)
[] Decision: SPIDERWEB (SFU) first — mediasoup default
[] Crashes BE base URL + API token for room validation
[] Confirm: do NOT use androidcast :8089 / monolith BE for SFU
[] Optional: 2nd/3rd HVM reserved for TURN + ffmpeg

12. Changelog

Date	Change
2026-06-12	Initial prerequisites — HVM vs Docker, 3-instance layout, BE API isolation, PO checklist