**🔹 Phase 1 Breakdown (10 hours)**

**🎯 Goals**

* Understand DDIA **fundamentals** (data models, storage engines, replication basics).
* Map DDIA concepts → **GenAI Pyramid Layer 1 (Infra)**.
* Apply mappings to **LexiFlow infra design** (your repo).
* Produce 2 artifacts:
  1. **Mapping table** (DDIA → Pyramid → LexiFlow).
  2. **HLD diagram** for LexiFlow infra.

**⏱️ Suggested Time Split**

* **4h** → DDIA Ch. 1–3 reading & notes.
* **3h** → AWS Infra deep dive (VPC, ECS, ALB/WAF).
* **3h** → Mapping + diagrams.

**📖 Step 1 — DDIA Concepts (Ch. 1–3)**

Focus on these:

1. **Data Models** → relational, document, graph. Think: Which fits vector DB use cases in RAG?
2. **Storage Engines** → B-trees vs LSM trees. Relate to FAISS, Weaviate, DynamoDB.
3. **Replication** → Leader-based replication. Map to ECS tasks across AZs, S3 replication.

**🔄 Step 2 — Map to Pyramid (Infra layer)**

Examples:

* **Replication (DDIA)** → ECS Fargate tasks across AZs + Weaviate replication.
* **Storage Engines (DDIA)** → DynamoDB partitioning (range/hash), FAISS ANN indexes.
* **Logs (DDIA)** → SQS/EventBridge as append-only event logs.
* **Failover (DDIA)** → ALB health checks + auto-restart ECS tasks.

**⚙️ Step 3 — Apply to LexiFlow**

LexiFlow Infra Pipeline (current state from your repo context):

* Route53 → WAF → ALB → ECS (FastAPI container).
* Vector DB (FAISS/Weaviate) + S3 for docs.
* EventBridge for async jobs (optional).
* CloudWatch + Secrets Manager.

Questions to answer in your notes:

* Where should replication be enforced? (ECS, Vector DB, S3 cross-region).
* How should partitioning work? (LexiFlow supports multi-tenant doc ingestion).
* What log-based design can be added? (Audit logs via SQS/EventBridge).

**📝 Step 4 — Produce Artifacts**

1. **Mapping Table (DDIA → Pyramid → LexiFlow)**

| **DDIA Concept** | **Pyramid Layer** | **LexiFlow Application** |
| --- | --- | --- |
| Replication | Infra | ECS Fargate tasks in multiple AZs, S3 cross-region replication, Vector DB replicas |
| Partitioning (hash/range) | RAG layer | Sharding docs in FAISS/Weaviate for balanced retrieval load |
| Logs as source of truth | Infra + Orchestration | Use SQS/EventBridge for ingestion logs & async indexing |
| Failover & Fault Tolerance | Infra | ALB + ECS health checks, auto-restarts, multi-AZ deployment |
| Storage Engines (LSM/B-tree) | Infra | DynamoDB choice for metadata vs FAISS ANN for vectors |

1. **HLD Diagram (Infra)**

* Client → Route53 → WAF → ALB → ECS Fargate (API + Worker).
* ECS connects to: S3 (raw docs), Vector DB (indexes), DynamoDB (metadata).
* EventBridge/SQS buffers ingestion.
* CloudWatch monitors all layers.
* Secrets Manager secures keys.

**✅ Your Action This Week**

* Read DDIA Ch. 1–3 and extract **concepts + relevance to Pyramid**.
* Draft **LexiFlow HLD diagram** (we can refine together).
* Build the **first mapping table** (I’ll help validate).