Detailed project plan

Project Plan (2-Week Timeline)

Team of 4 (A, B, C, D)

Day	Dates	Tasks	Assigned To	Deliverable
1	**Week 1**	**Kick-off meeting: confirm roles, repo, environments. Finalize Phase 1 docs.**	**All**	**Phase 1 Docs ready, Kickoff Notes**
Milestone 1	**Week 1**	**Environment setup validated (Python, Queue running, Git repo ready)**	**All**	**"Ready to Code" checklist passed**
2	**Week 1**	**Master: implement task enqueue in `master.py`** **Crawler: stub `fetch_and_parse()`**	**A, B**	**Basic message passing tested**
3	**Week 1**	**Indexer: stub Whoosh indexer in `indexer.py`** **Client: scaffold CLI in `client.py`**	**C, D**	**Basic Whoosh schema created, CLI initialized**

Milestone 2	**Week 1**	**All components can start and stop without crashing (empty pipelines)**	**All**	**Initial code skeleton verified**
4	**Week 1**	**Integrate A → B: master → crawler passes tasks, B logs fetched HTML** **B → C: write JSON**	**A, B, C**	**JSON crawl records saved locally**
5	**Week 1**	**Integrate C: read JSON, index text → Whoosh** **D: connect CLI search to Whoosh**	**C, D**	**Search finds URL for a keyword**
Milestone 3	**End of Week 1**	**End-to-end basic flow completed (crawl, store, index, search)**	**All**	**MVP v0 Ready (Minimum Viable Product)**
6	**Week 2**	**B: Add link extraction; Enqueue new links into queue**	**B**	**Recrawling depth-1 operational**
7	**Week 2**	**A: implement retry logic; B/C: add logging everywhere**	**A, B, C**	**Error handling + Monitoring logs ready**
8	**Week 2**	**D: enhance CLI – seed list, crawl start, progress monitor**	**D**	**CLI fully functional for crawl and search**

9	Week 2	Full end-to-end test (10 seeds, depth=2), debugging	All	Final Test Report
Milestone 4	End of Week 2	Full system operational; bugs resolved	All	Ready for Phase 2 Submission
10	Week 2	Final fixes, polish, README update, prepare demo	All	Final Code + Documentation submission

---

Role Key

- **A (Master Lead)**: Implements and tests task scheduling, retry logic.

- **B (Crawler Lead)**: Fetches pages, extracts text & links, writes JSON.

- **C (Indexer Lead)**: Builds Whoosh index, ingests JSON, handles searches.

- **D (Client & Testing Lead)**: Develops CLI/UI, orchestrates tests, demo prep.

Notes

- **Git Workflow**: Feature branches for each component, merge into `main` by end of day.
- **Code Testing Requirement**: Every push to `main` branch must run basic test cases. Unit tests are preferred, but at least functional tests for crawler, master, indexer.