

Fable 5

Orchestrator Workflow

Let Fable 5 plan and delegate while Sonnet 5 handles the token-heavy execution work.

by John Savage AI

● Best AI Tools I Use Daily

<https://tools.johnsavageai.com/>

● All My Socials

<https://linktr.ee/johnsavage14>

COPY-READY PROMPT

```
https://github.com/anthropics/claude-cookbooks/blob/main/managed_agents/CMA_plan_big_execute_small.ipynb
Implement this strategy for every Claude session.
Use Claude Fable 5 as the orchestrator for planning, delegation, and final synthesis.
Delegate token-heavy, well-scoped execution work to Claude Sonnet 5 workers, using
parallel delegation whenever subtasks are independent. Have Fable 5 verify and
synthesize the final result.
```

● Open Claude

<https://claude.ai>

● GitHub cookbook

https://github.com/anthropics/claude-cookbooks/blob/main/managed_agents/CMA_plan_big_execute_small.ipynb

● Managed Agents docs

<https://platform.claude.com/docs/en/managed-agents/multi-agent>

THE BENCHMARK

Anthropic's reported BrowseComp result was about 96% of all-Fable performance at about 46% of the cost. Treat it as a reference point, not a guarantee.

What this workflow does

A simple model split for complex Claude sessions that involve a lot of reading, browsing, or mechanical work.

THE IDEA

Turn one expensive, monolithic session into a coordinated team.

Fable 5 stays focused on planning, judgment, and final synthesis. Sonnet 5 workers handle narrow, repeatable, or independent subtasks in their own threads, then return distilled results for Fable 5 to verify.

Fable 5

COORDINATOR

Breaks down the goal, routes work, and checks the final answer.

Sonnet 5

EXECUTION LAYER

Reads, searches, edits, or completes well-scoped work at a lower rate.

Parallel work

WHERE IT SHINES

Independent subtasks can run at the same time instead of serially.

THE MODEL SPLIT

Think -> delegate -> execute -> verify



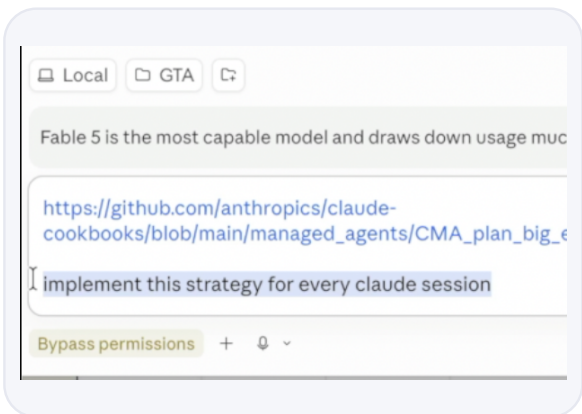
BEST FITS

Research, codebase sweeps, document review, log analysis, repetitive browser tasks, and other workflows where the token-heavy leg can be split into smaller jobs.

Reference setup

These are the supplied Claude screenshots: link first, instruction second.

REFERENCE 01



Paste the cookbook link, then add the instruction.

The key wording is: *implement this strategy for every Claude session.*

REFERENCE 02

https://github.com/anthropics/claude-cookbooks/blob/main/managed_agents/CMA_plan_big_execute_small.ipynb

Use the exact notebook URL.

This gives Claude the source material for the Managed Agents plan-big / execute-small pattern.

https://github.com/anthropics/claude-cookbooks/blob/main/managed_agents/CMA_plan_big_execute_small.ipynb

WHAT TO LOOK FOR

1. The GitHub cookbook is visible in the message.
2. The instruction tells Claude to use the strategy for every session.
3. The opening-page prompt makes the Fable 5 / Sonnet 5 model split explicit.
4. Save it as a Project or persistent instruction if you want it to carry across chats.

Why the split works

The opening-page prompt gives each model one clear job.

Fable 5

PLAN + VERIFY

Keeps the highest-value reasoning, routing, and final quality check in the coordinator.

Sonnet 5

READ + EXECUTE

Handles token-heavy browsing, editing, searching, and other well-scoped work.

Parallel workers

SPEND LESS AT SCALE

Independent subtasks can run in parallel, reducing serial frontier-model work.

THE MODEL SPLIT

Think -> delegate -> execute -> verify



WHY THIS WORDING

It keeps the model roles clear: Fable 5 plans and verifies; Sonnet 5 handles the high-volume execution leg; parallel delegation is used when it actually fits the task.

PERSISTENCE NOTE

A normal chat only affects the current session. To make it your default, save the prompt as a Project instruction or persistent workspace instruction when available.

How to use it

Five quick steps to turn the prompt into a repeatable Claude setup.

1

Open Claude

Start at claude.ai and open a new chat, Project, or workspace.

2

Enable the tools

Turn on Computer Use and Browser Use in the tools or permissions menu when available.

3

Paste the source

Paste the GitHub cookbook URL so Claude can inspect the strategy.

4

Paste the prompt

Add the Main Prompt from page 1 after the link.

5

Save for reuse

If you want the behavior across future chats, save the prompt as a Project or persistent workspace instruction.

BATCH DELEGATION TIP

For independent subtasks, add: Break the work into focused subtasks, delegate independent subtasks in parallel, wait for the workers, then verify and synthesize the final result.

Useful links



Best AI Tools I Use Daily

My updated list of AI tools for images, video, workflows, and content creation.
<https://tools.johnsavageai.com/>



All My Socials

Follow for more AI prompts, workflows, tutorials, and experiments.
<https://linktr.ee/johnsavage14>



Open Claude

<https://claude.ai>



Managed Agents docs

<https://platform.claude.com/docs/en/managed-agents/multi-agent>