

USER: "give me spend analysis report" (50 tokens)  
ORCHESTRATOR: Plan agent sequence (2,800 tokens)  
Report schema (compressed): 1,500  
Instructions: 800  
Recent context: 500

SQL AGENT: 5 queries  $\times$  2,150 = 10,750 tokens  
Per query:  
DB schema: 2,000  
Task: 100  
Auth filter: 50

ANALYSIS AGENT: Generate insights (2,900 tokens)  
SQL results: 1,500  
Instructions: 500  
Task: 100

CANVAS AGENT: Generate JSON (5,500 tokens)  
canvas\_inputs: 2,000  
Report template: 310  
Instructions: 1,000  
JSON schema: 500  
(78.7% cache hit saves ~8,000 more!)

ORCHESTRATOR: Final message (2,089 tokens)  
Canvas Report Rendered : TOTAL EFFICIENT: 25,939 tokens