

Lab 074: Foundry Agent Service ? Simulação de Operação em Produção

Nível: L300 Trilha: ? Microsoft Foundry Tempo: ~120 min ? Custo: Gratuito ? Usa dados simulados de agentes

O que Você Vai Aprender

O que é o Foundry Agent Service e como ele orquestra sistemas multi-agente em produção Como os tipos de agentes (especialista, orquestrador) trabalham juntos em uma arquitetura de produção Analisar a saúde da frota de agentes: volumes de requisições, latência, taxas de erro e status Identificar agentes degradados e riscos de configuração (ex.: filtros de conteúdo desativados) Construir um painel de saúde da frota a partir de dados simulados de monitoramento

Introdução

O Azure AI Foundry Agent Service fornece uma plataforma gerenciada para implantar, orquestrar e monitorar sistemas multi-agente em escala empresarial. Em vez de construir orquestração personalizada, você define agentes com ferramentas, memória e modelos específicos ? e o serviço cuida do roteamento, gerenciamento de estado e escalabilidade.

Tipos de Agentes

Tipo | Função | Exemplo | Orquestrador | Roteia requisições para especialistas, gerencia o fluxo de conversa | SupportRouter, Coordinator | Especialista | Lida com um domínio específico com ferramentas e memória dedicadas | ProductAdvisor, OrderProcessor |

O Cenário

Você é um SRE de Plataforma gerenciando uma implantação multi-agente para uma empresa de e-commerce. A frota tem 8 agentes ? 2 orquestradores e 6 especialistas ? rodando no Azure Container Apps. Você foi alertado de que um agente está degradado e precisa investigar.

Seu dataset (foundry_agents.csv) contém o status atual da frota. Sua tarefa: analisar métricas de saúde, identificar problemas e produzir um relatório de status da frota.

info "Dados Simulados" Este é um lab de análise operacional, não um lab de implantação real. Ele usa um CSV simulado de frota de agentes que espelha as métricas que você veria no painel de monitoramento do Azure AI Foundry. Os padrões (picos de latência, taxas de erro, status degradado) representam cenários comuns de produção.

Pré-requisitos

Requisito | Por quê | Python 3.10+ | Executar os scripts de análise | Biblioteca pandas | Manipulação de dados |

[diagram or code sample]

tip "Início Rápido com GitHub Codespaces" Open in GitHub Codespaces

Todas as dependências estão pré-instaladas no devcontainer.

? Arquivos de Apoio

note "Baixe estes arquivos antes de iniciar o lab" Salve todos os arquivos em uma pasta lab-074/ no seu diretório de trabalho.

Arquivo | Descrição | Download | broken_foundry.py | Exercício de correção de bugs (3 bugs + autotestes) | ? Download | foundry_agents.csv | Dataset | ? Download |

Etapa 1: Entenda a Arquitetura da Frota

Antes de analisar os dados, entenda como os agentes se encaixam:

[diagram or code sample]

Campos de Configuração Principais

Campo | Descrição | memory_type | Como o agente persiste estado: cosmos_db (durável), ai_search (vetorial), session_only (efêmero), none | deployment | Infraestrutura: container_apps (gerenciado) ou vm (auto-hospedado) | content_filter | Se a segurança de conteúdo do Azure AI está enabled ou disabled | status | Saúde do agente: active ou

degraded |

Etapa 2: Carregue e Explore os Dados da Frota

[diagram or code sample]

Saída esperada:

[diagram or code sample]

Etapa 3: Analise o Volume de Requisições e a Distribuição de Carga

Como o tráfego está distribuído pela frota?

[diagram or code sample]

Saída esperada:

[diagram or code sample]

Agente | Requisições | Participação | Coordinator | 3.200 | 34,1% | SupportRouter | 2.100 | 22,4% | ProductAdvisor | 1.250 | 13,3% | OrderProcessor | 890 | 9,5% | QualityReviewer | 780 | 8,3% | InventoryMonitor | 560 | 6,0% | AnalyticsAgent | 420 | 4,5% | LegacyBridge | 180 | 1,9% |

tip "Insight" O orquestrador Coordinator lida com 34% de todo o tráfego ? ele é o ponto de entrada para a maioria das requisições. Se ele cair, todo o sistema é afetado. O

SupportRouter é o segundo mais ocupado, roteando consultas de suporte ao cliente para especialistas.

Etapa 4: Identifique Agentes Degradados e em Risco

4a ? Agentes Degradados

[diagram or code sample]

Saída esperada:

[diagram or code sample]

4b ? Agentes com Alta Taxa de Erro

[diagram or code sample]

4c ? Status do Filtro de Conteúdo

[diagram or code sample]

warning "Risco de Segurança" LegacyBridge (FA08) tem seu filtro de conteúdo desativado e roda em uma VM auto-hospedada. Isso é um risco de conformidade ? todos os agentes em produção devem ter a segurança de conteúdo habilitada, especialmente aqueles que lidam com dados de clientes.

Etapa 5: Analise Padrões de Memória e Infraestrutura

[diagram or code sample]

[diagram or code sample]

Etapa 6: Construa o Relatório de Saúde da Frota

[diagram or code sample]

? Exercício de Correção de Bugs

O arquivo lab-074/broken_foundry.py contém 3 bugs que produzem métricas incorretas da frota. Você consegue encontrar e corrigir todos?

Execute os autotestes para ver quais falham:

[diagram or code sample]

Você deverá ver 3 testes falhando. Cada teste corresponde a um bug:

Teste | O que verifica | Dica | Teste 1 | Total de requisições em 24h | Deve somar as requisições, não calcular a média | Teste 2 | Contagem de agentes degradados | Deve contar o status degraded, não active | Teste 3 | Agentes sem memória durável | Deve contar none/session_only, não cosmos_db |

Corrija todos os 3 bugs e execute novamente. Quando você vir All passed!, está pronto!

? Verificação de Conhecimento

question "Q1 (Múltipla Escolha): Qual é o papel de um agente orquestrador em uma implantação multi-agente do Foundry?"

- A) Ele executa uma tarefa de domínio específico, como processamento de pedidos - B) Ele roteia requisições para agentes especialistas e gerencia o fluxo de conversa - C) Ele armazena a memória do agente no Cosmos DB - D) Ele monitora a saúde dos agentes e reinicia agentes com falha

??? success "? Revelar Resposta" Correta: B) Ele roteia requisições para agentes especialistas e gerencia o fluxo de conversa

Agentes orquestradores atuam como o "controlador de tráfego" em um sistema multi-agente. Eles recebem requisições de entrada, determinam qual(is) especialista(s) deve(m) tratá-las, roteiam a conversa adequadamente e gerenciam o fluxo geral. Os especialistas lidam com o trabalho específico do domínio.

question "Q2 (Múltipla Escolha): Por que um filtro de conteúdo desativado é um risco de segurança para agentes em produção?"

- A) Ele torna o agente mais lento - B) Ele permite que o agente gere conteúdo prejudicial, tendencioso ou que viole políticas - C) Ele impede que o agente acesse APIs externas - D) Ele aumenta os custos de tokens

??? success "? Revelar Resposta" Correta: B) Ele permite que o agente gere conteúdo prejudicial, tendencioso ou que viole políticas

Os filtros do Azure AI Content Safety detectam e bloqueiam conteúdo prejudicial (discurso de ódio, violência, autolesão, conteúdo sexual). Desativar o filtro significa que o agente pode produzir ou responder a esse tipo de conteúdo sem proteções ? um risco de conformidade e reputação em qualquer implantação em produção.

question "Q3 (Execute o Lab): Qual é o número total de requisições em toda a frota nas últimas 24 horas?"

Execute a análise da Etapa 3 no ? foundry_agents.csv e verifique os resultados.

??? success "? Revelar Resposta" 9.380 requisições

Soma de todos os valores requests_24h dos agentes: $1.250 + 890 + 2.100 + 560 + 3.200 + 780 + 420 + 180 = 9.380$.

question "Q4 (Execute o Lab): Quantos agentes estão em estado degradado?"

Execute a análise da Etapa 4a para descobrir.

??? success "? Revelar Resposta" 1 agente

Apenas o AnalyticsAgent (FA07) está em estado degraded, com uma taxa de erro de 8,5% e latência média de 850ms ? significativamente pior que os outros agentes. Isso provavelmente indica um problema de conectividade com o backend do seu armazenamento de memória AI Search.

question "Q5 (Execute o Lab): Quantos agentes têm o filtro de conteúdo desativado?"

Execute a análise da Etapa 4c para verificar o status do filtro de conteúdo.

??? success "? Revelar Resposta" 1 agente

Apenas o LegacyBridge (FA08) tem content_filter=disabled. Ele também é o único agente implantado em uma VM auto-hospedada em vez de Container Apps, e possui a maior taxa de erro (12,0%) da frota. Este agente precisa de atenção imediata.

Resumo

Tópico | O que Você Aprendeu | Foundry Agent Service | Plataforma gerenciada para orquestração e implantação multi-agente | Tipos de Agentes | Orquestradores roteiam; especialistas executam tarefas de domínio | Monitoramento da Frota | Acompanhe requisições, latência, taxas de erro e status por agente | Detecção de Degradação | Identifique agentes com taxas de erro ou latência elevadas | Segurança de Conteúdo | Todos os agentes em produção devem ter filtros de conteúdo habilitados | Padrões de Memória | Cosmos DB para

durável, AI Search para vetorial, session_only para efêmero |

Próximos Passos

Lab 034 ? Multi-Agente com Semantic Kernel (construindo os agentes) Lab 033 ?

Observabilidade de Agentes com Application Insights (monitoramento aprofundado) Lab 030 ?

Foundry Agent + MCP (conectando agentes a ferramentas externas) Lab 075 ? Power BI Copilot (visualizando dados da frota com dashboards assistidos por IA)

Modernização da Fase 3

warning "Verified 2026-10-27" Use o nome Microsoft Foundry e o Agent Service GA baseado na Responses API. Avaliações estão GA; hosted agents continuam preview e devem ser identificados assim. Não crie dependência nova de Workflows do portal, descontinuados em 01/12/2026.

[diagram or code sample]

Foundry Agent Service GA announcement Foundry Agent Service Foundry evaluation Foundry Workflows retirement

Phase 3 · Simulação / Caminho Real

Simulação Caminho Real

Preserve the existing free simulation and its fixtures. Estimated cost: US\$0.

Reuse the same acceptance criteria against the documented hosted service. Before running, record a service-specific estimate and a hard spending cap. After validation: Exclua implantações, endpoints, índices, traces e dados de teste após a validação.

Kit de apoio

O build do site gera downloads estáveis em PDF e ZIP. O ZIP contém o guia do lab e todos os arquivos de apoio declarados para este lab.

Baixar PDF gerado Baixar ZIP gerado

Supporting file: broken_foundry.py

```
/usr/bin/env python3 """Bug-fix: Broken Foundry Analyzer - Lab 074. Fix 3 bugs.""" import
io, pandas as pd def total_requests(agents): # Bug #1: Returns average instead of sum
return int(agents["requests_24h"].mean()) def count_degraded(agents): # Bug #2: Counts
'active' instead of 'degraded' return (agents["status"] == "active").sum() def
agents_without_memory(agents): # Bug #3: Counts cosmos_db (has memory) instead of
none/session_only return (agents["memory_type"] == "cosmos_db").sum() def run_tests():
csv="agent_id,status,requests_24h,memory_type\nF1,active,1000,cosmos_db\nF2,degraded,500,ses
sion_only\nF3,active,800,none\nF4,active,1200,cosmos_db"
a=pd.read_csv(io.StringIO(csv));p,f=0,0 t=total_requests(a);exp=3500 if t==exp:print(f"OK
T1: {t}");p+=1 else:print(f"FAIL T1: {t} (exp {exp})");f+=1 d=count_degraded(a);exp=1 if
d==exp:print(f"OK T2: {d}");p+=1 else:print(f"FAIL T2: {d} (exp {exp})");f+=1
n=agents_without_memory(a);exp=2 if n==exp:print(f"OK T3: {n}");p+=1 else:print(f"FAIL T3:
{n} (exp {exp})");f+=1 print(f"\n{'All passed!' if f==0 else f'{f} failed'}") if
__name__=="__main__": print("Tests:\n");run_tests()
```

Supporting file: foundry_agents.csv

```
agent_id,agent_name,agent_type,model,tools_count,memory_type,deployment,status,requests_24h,
avg_latency_ms,error_rate_pct,content_filter
FA01,ProductAdvisor,specialist,gpt-4o,5,cosmos_db,container_apps,active,1250,320,1.2,enabled
FA02,OrderProcessor,specialist,gpt-4o,8,cosmos_db,container_apps,active,890,450,2.1,enabled
FA03,SupportRouter,orchestrator,gpt-4o-mini,3,session_only,container_apps,active,2100,180,0.
5,enabled FA04,InventoryMonitor,specialist,gpt-4o-mini,4,ai_search,container_apps,active,560
```

,220,0.8,enabled FA05,Coordinator,orchestrator,gpt-4o,2,cosmos_db,container_apps,active,3200
,280,1.0,enabled FA06,QualityReviewer,specialist,gpt-4o,3,session_only,container_apps,active
,780,350,1.5,enabled FA07,AnalyticsAgent,specialist,gpt-4o,6,ai_search,container_apps,degrad
ed,420,850,8.5,enabled FA08,LegacyBridge,specialist,gpt-4o-mini,10,none,vm,active,180,1200,1
2.0,disabled