



2023年9月23日 10:00 - 17:00





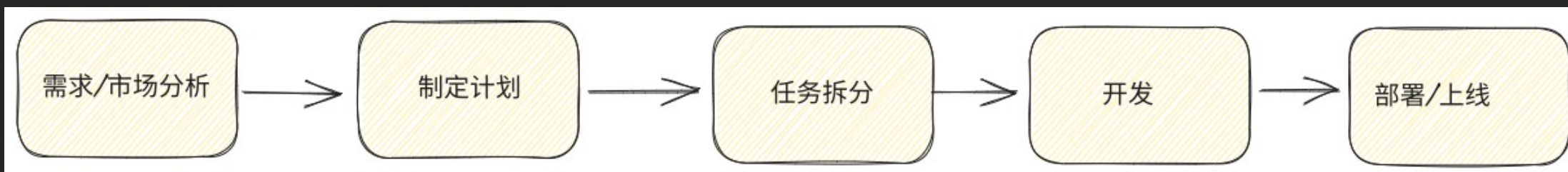
云原生场景下的 LLM 应用开发

张晋涛





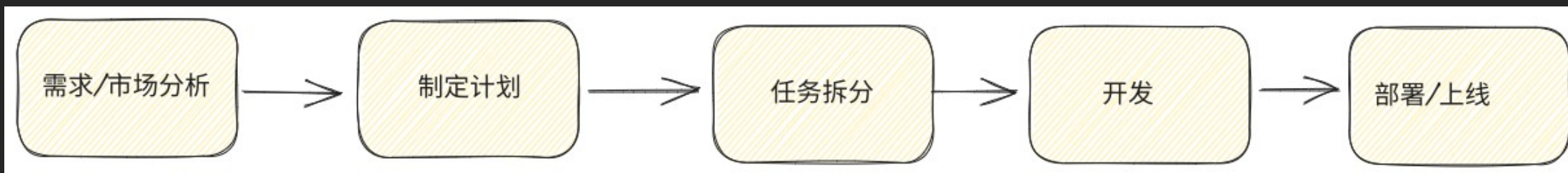
回顾：如何开发一个应用



- 目标是什么
- 需要做什么
- 如何去做
 - 技术选型
 - 方案设计
- 后续的迭代和维护



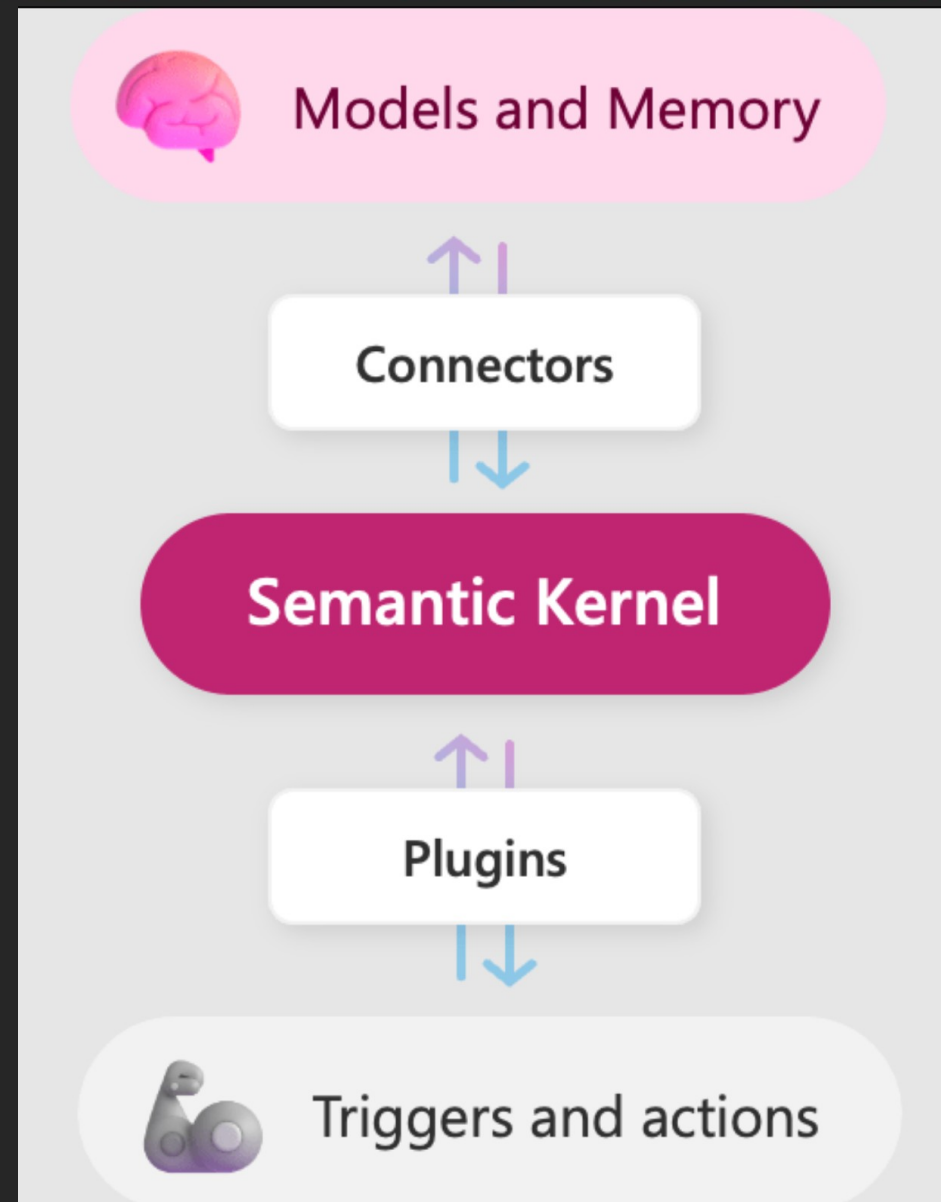
如何开发一个 LLM 应用

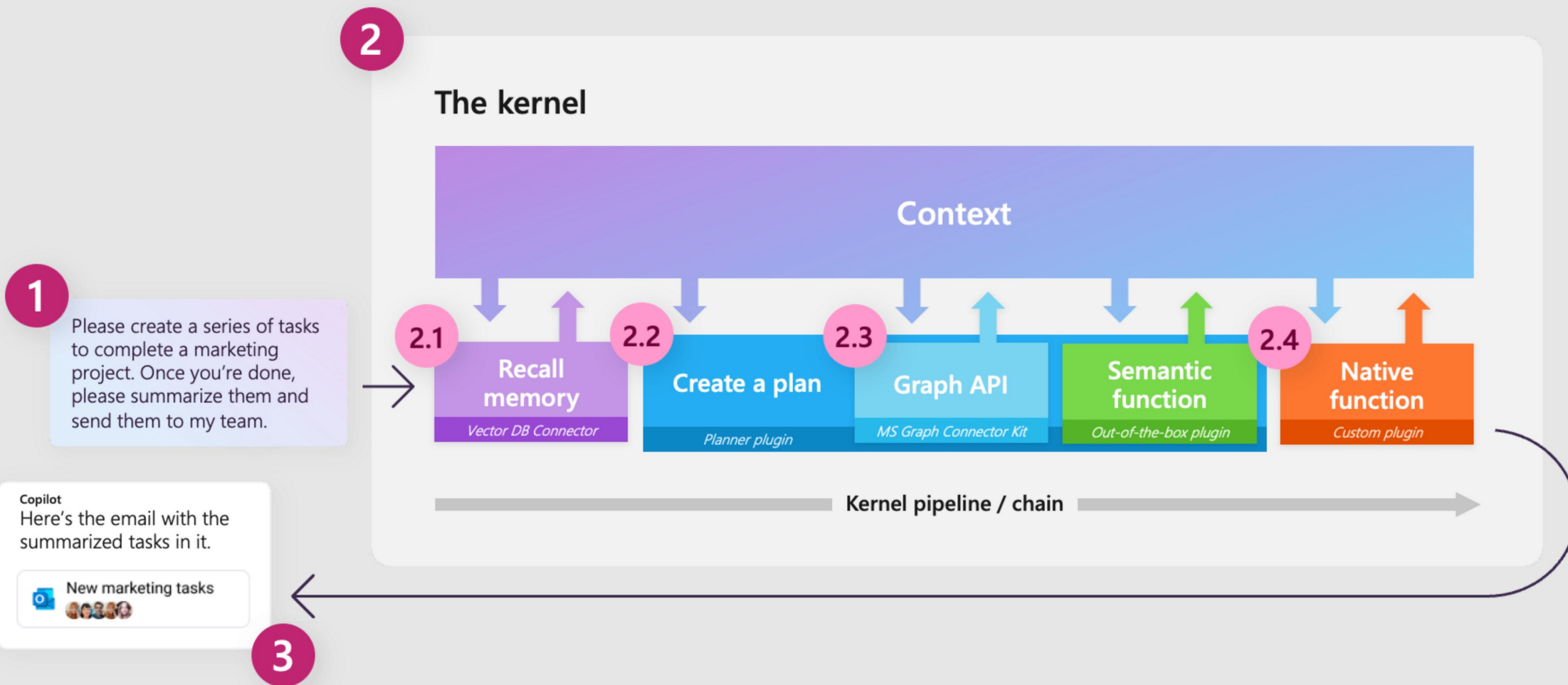


- 目标是什么
- 需要做什么
- 如何去做
 - 技术选型
 - 方案设计
- 后续的迭代和维护
- 用什么模型（本地 /OpenAI/Azure OpenAI..）
- 需要哪些能力
- 目标应用的类型（ Chat/Agent/Copilot..）
- 如何组织 prompts
- 工程化实践



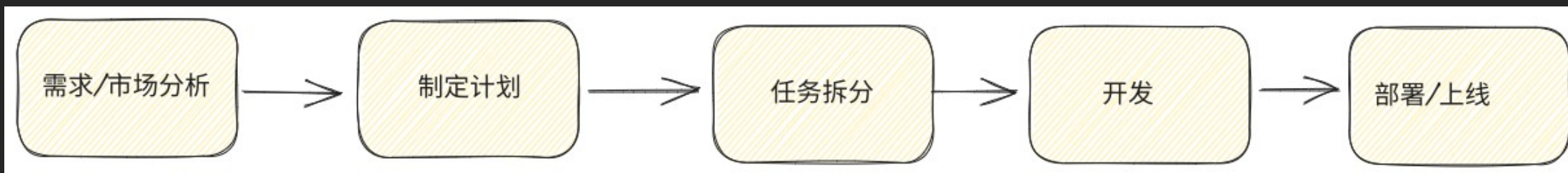
- SK 支持对接多种模型
- 对话的上下文存储
- 可扩展







云原生场景下如何开发一个 LLM 应用



- 技术栈选型
 - 开发语言（ Python/Java/.net/Go... ）
 - 运行环境（容器 /Kubernetes ）
- 扩缩容
- 迭代 / 更新 / 升级



Semantic Kernel
是更适合工程化开发
LLM 应用的框架