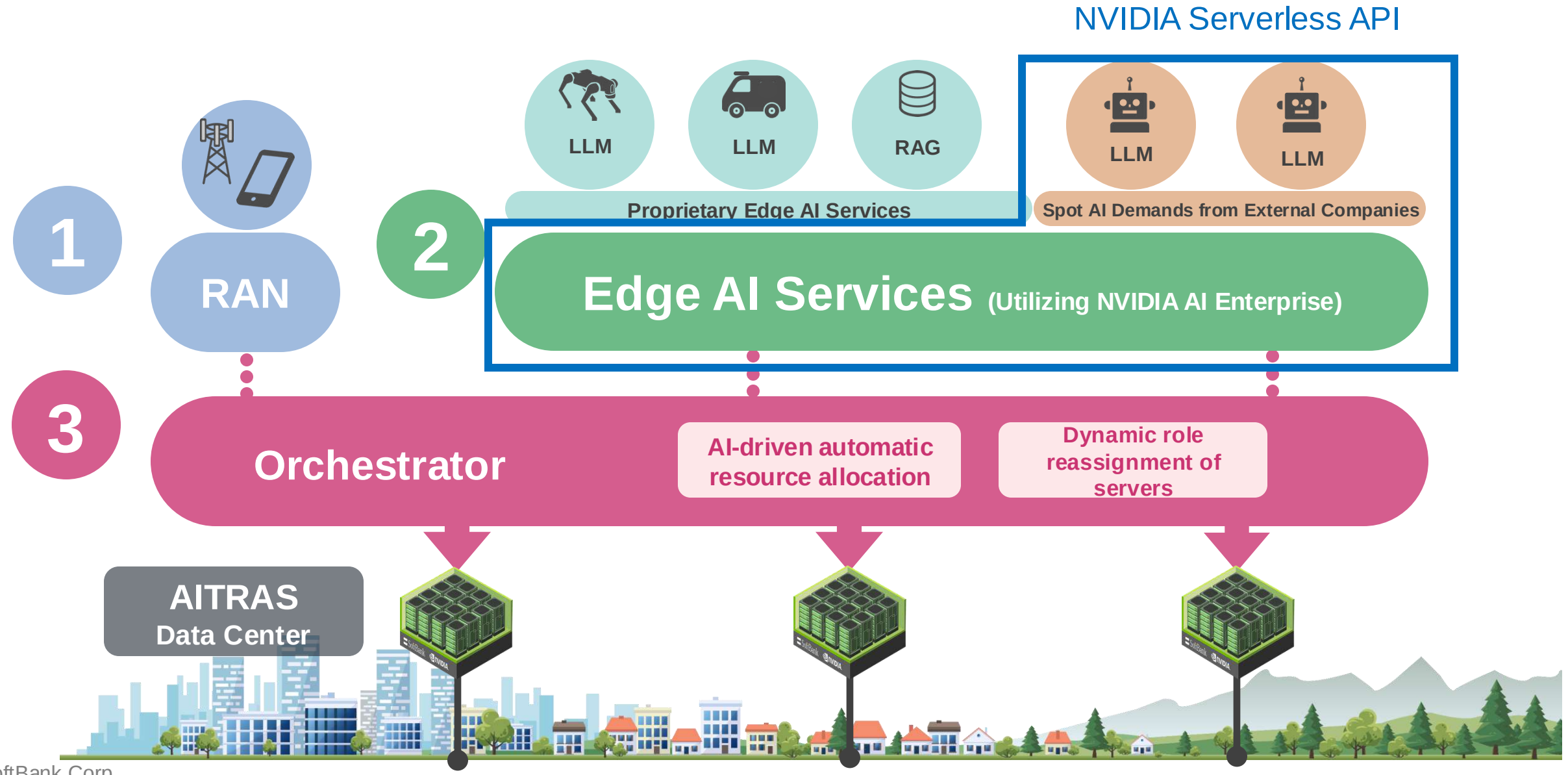
SoftBank

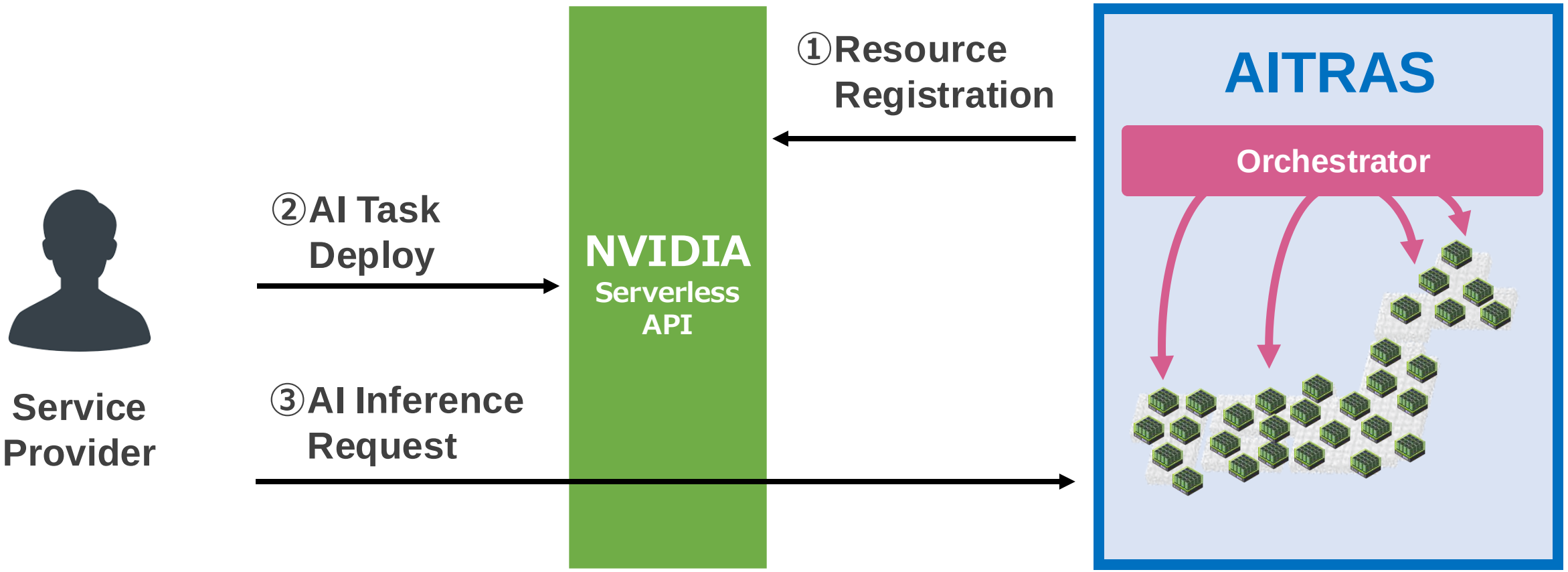
Utilization of Idle Resources by AITRAS

～Leveraging Serverless API powered by NVIDIA AI Enterprise～

Key Technologies of AITRAS



AITRAS × NVIDIA Serverless API

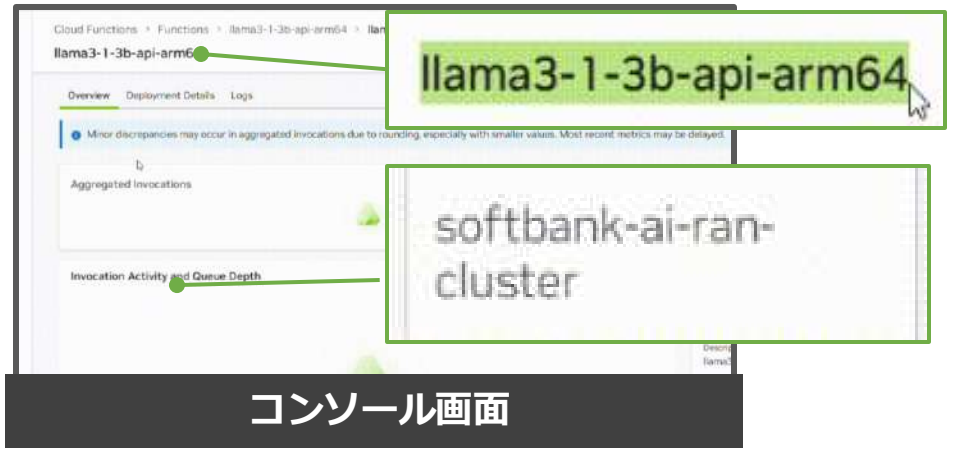


Providing Idle Resources Managed by AITRAS Across Japan to Customers via NVIDIA Serverless API

Demonstration

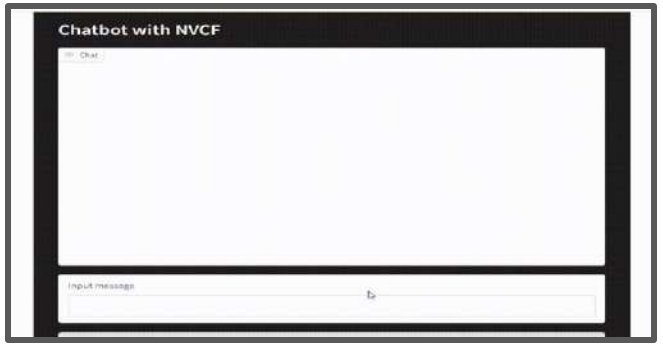
SoftBank

②AI Task Deploy



コンソール画面

③ AI Inference Request and Execution

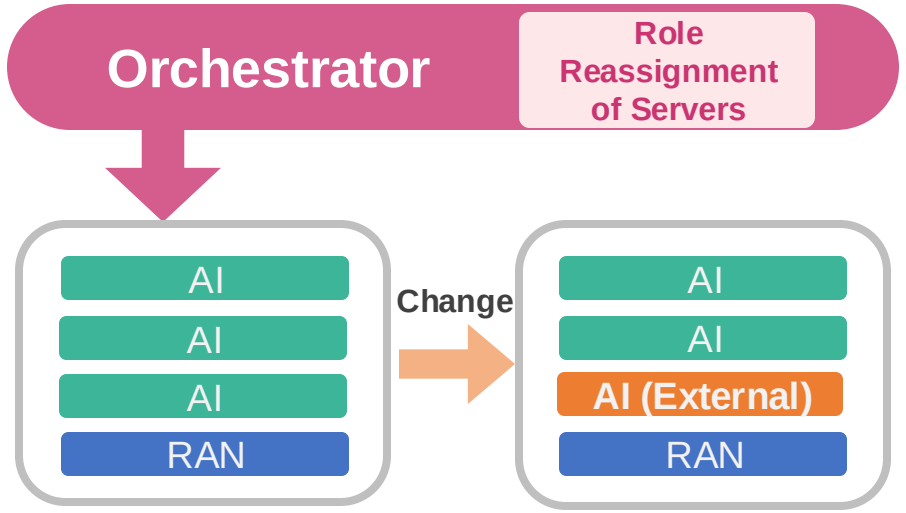


Chatbot Application

Chatbot
powered by
llama3

NVIDIA
Serverless
API

①Resource Registration



Orchestrator can absorb external AI demand when there is a resource surplus.


SoftBank