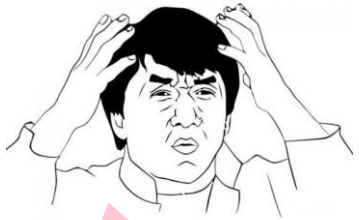
HUAWEI

HUAWEI CLOUD TechWave APAC 2024

Serverless & Event-Driven Architecture

APPLICATION MODERNIZATION

DEV



Too much costs and overhead!

My application only needs to run as and when it is needed. I do not want the overheads of underutilized resources that are always running, and do not want to add in extra resources or configuration just to start and stop resources..

Infrastructure?

Urgh - I only want to focus on development. I want to look for something that I can use quickly without thinking about infrastructure problems

- What is the right size and capacity for my server?
- How many servers?
- What OS?
- Should I scale up or down? Or in and out? When?
- How do I manage the security hardening and patching of the infra? How do I detect?
- What if my hardware or network fails?

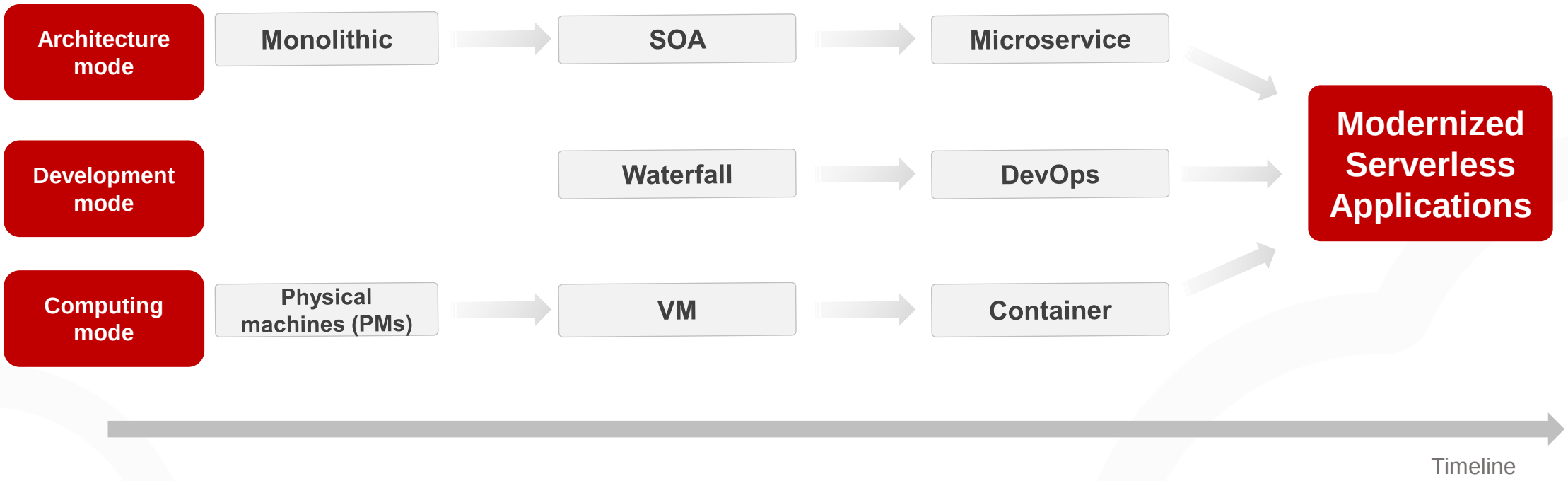
**Also infrastructure / DevOps teams*

Serverless Computing: An Emerging Cloud Native Computing Paradigm

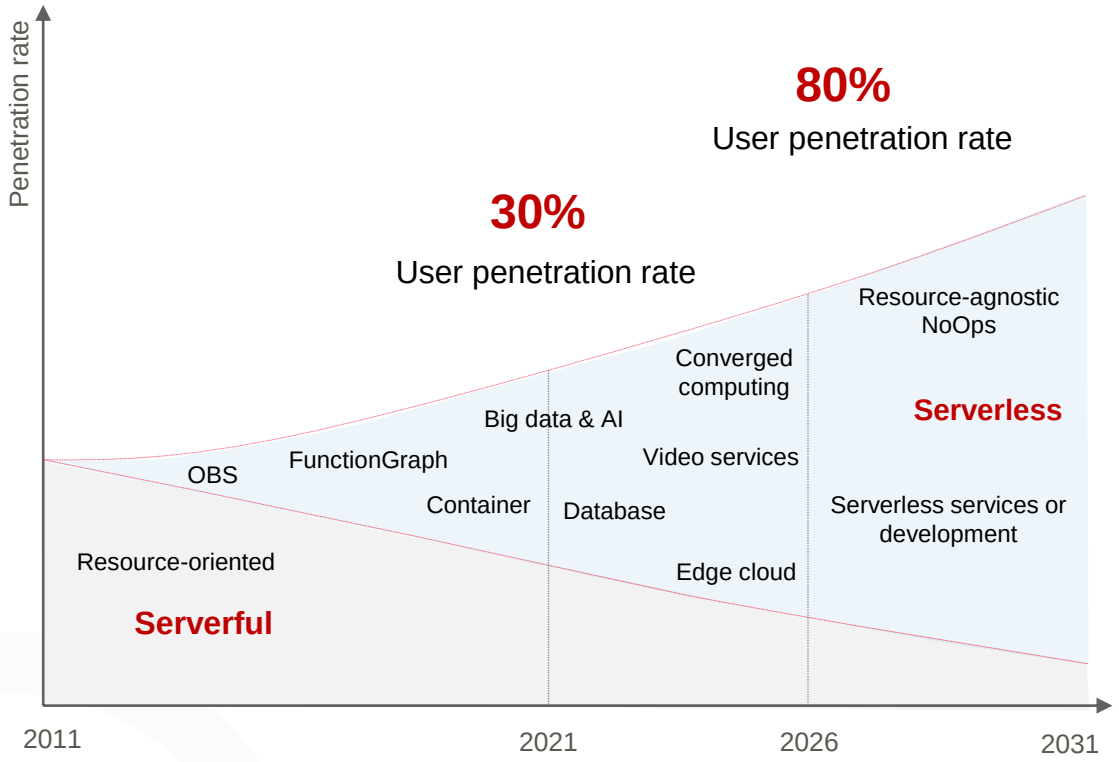
Serverful ---------- **Serverless**

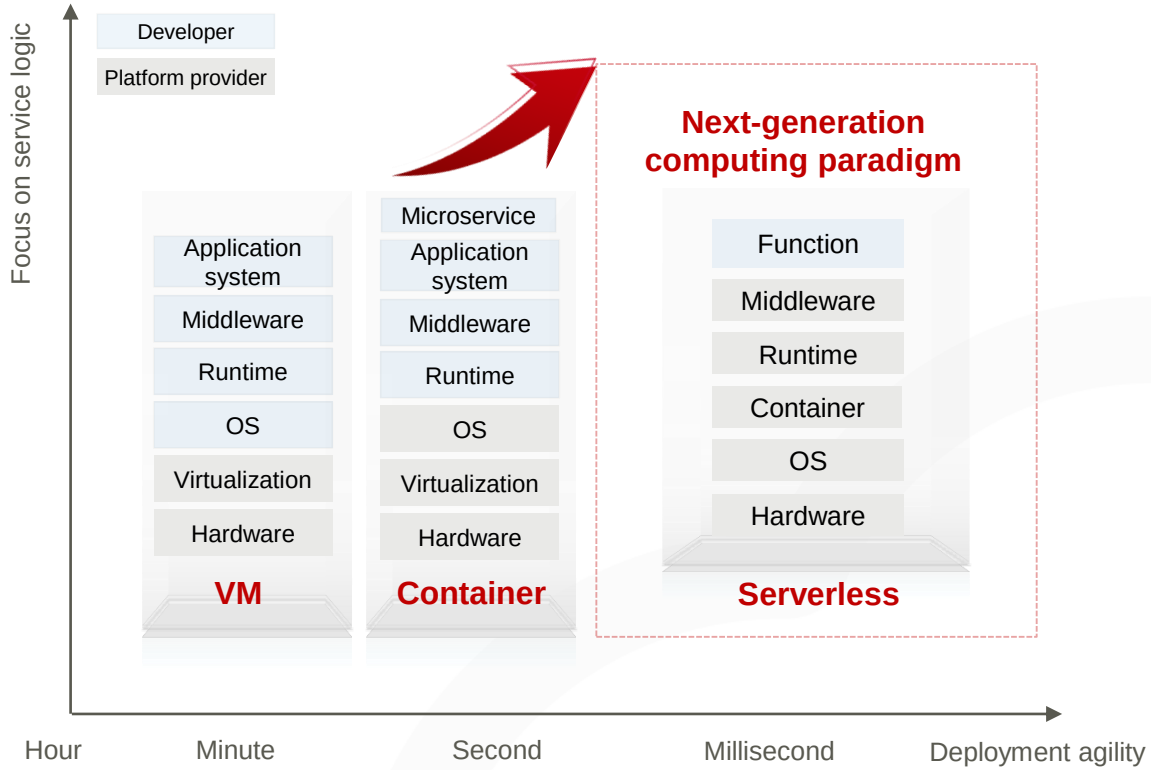
	Car purchase	Car rental	Ride hailing
Resource O&M	Server hardware O&M Software O&M	VM/container software O&M	**Resource-agnostic**
Business model	Ownership purchase	Pay by rental period	**On-demand** **Pay-per-use**
Service expansion	Planning Manual in days or hours	Resource utilization, periodical Semi-auto in minutes or seconds	**Request concurrency** **Fully auto in milliseconds**
Availability & Resiliency	Manually planned	Depends on choice	**Seamless**

Serverless Computing: An Emerging Cloud Native Computing Paradigm



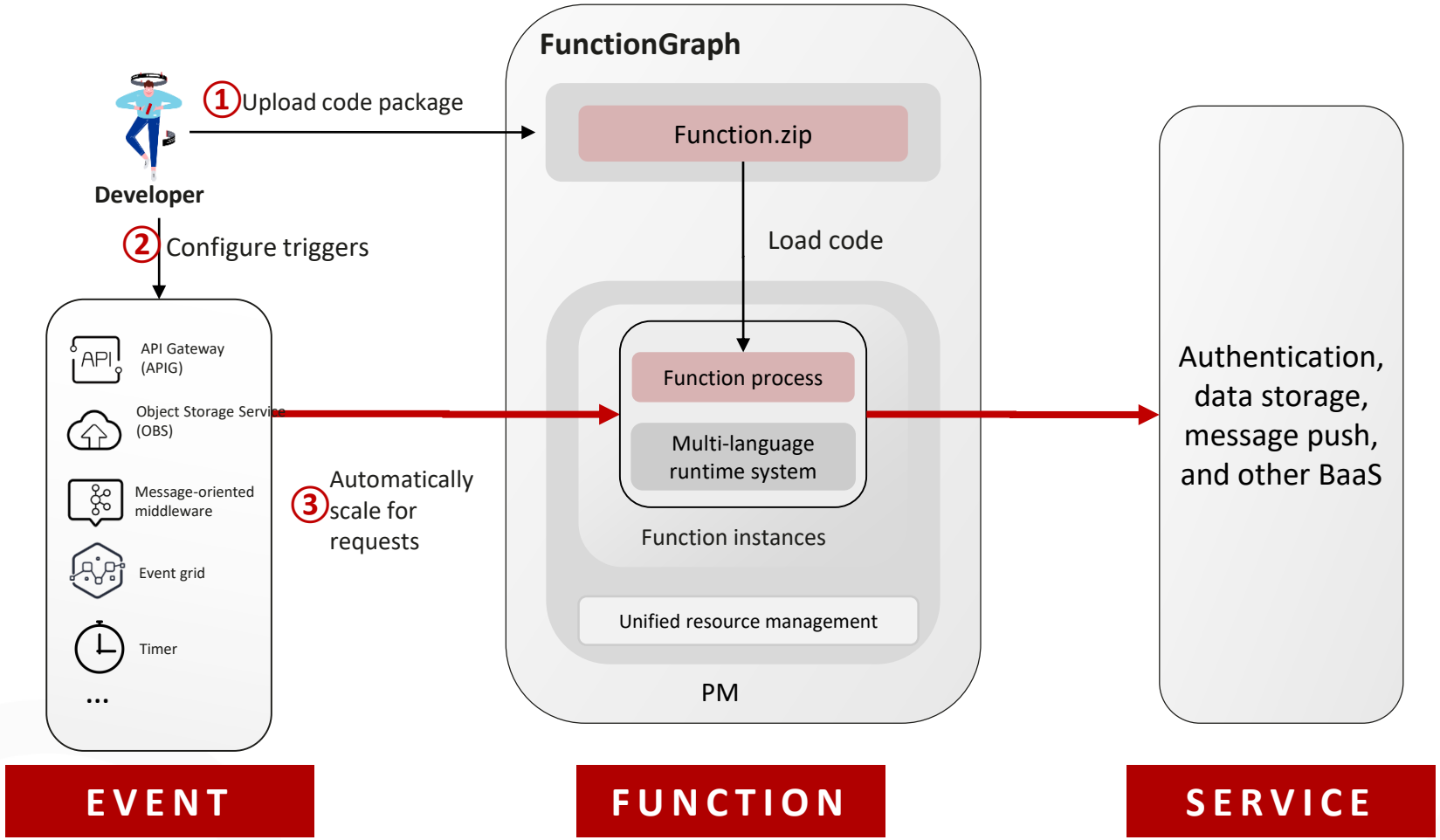
And We Think That Serverless Will Become the Primary Cloud Delivery Mode in Next 5 to 10 Years





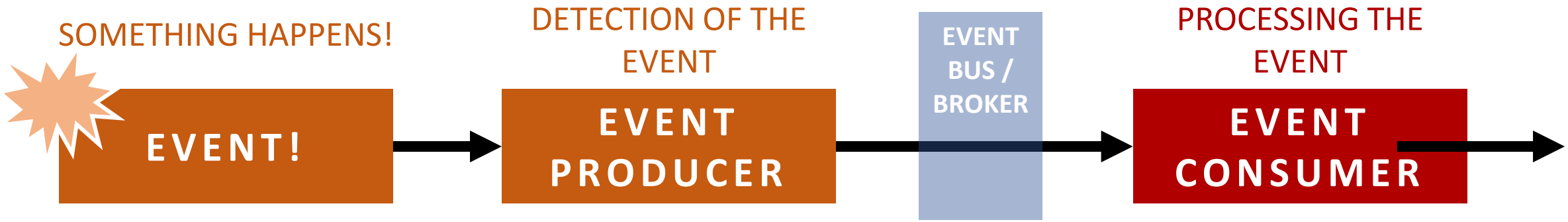
FunctionGraph

What Is FunctionGraph?



- A serverless **Function-as-a-Service (FaaS)** product
- Runtime hosting
- Event-driven
- Auto scaling (even to zero)
- Execution duration limit

What is Event-Driven?



- User clicks on something on the frontend client (API)
- New messages
- Change in operational state
- Change in time
- Change in data state (new file uploaded, deleted)

Pushed to consumer /
Polled by consumer

Event-Driven Architecture encourages the principle of Loose Coupling

Event-Driven + FunctionGraph provides you Just-In-Time Computing

Source Event Triggers for FunctionGraph

(A)SYNCHRONOUS	DATA EVENTS	MESSAGES & STREAM	OPERATIONAL	INTERVAL / FIXED	EVENT BUS
 APIG	 OBS	 DMS for Kafka	 CTS	 Timer	 EventGrid
	 DDS	 DMS for RabbitMQ	 LTS		
	 GaussDB(for Mongo)	 SMN			
		 DIS			

FunctionGraph Use Cases

Scenario 1: Web Applications

Use FunctionGraph together with other cloud services to quickly build cloud-native web applications by simply writing code. This improves rollout iteration efficiency and reduces O&M costs.

- **Applets**
- **Web pages/Apps**
- **Chatbot**
- **Backends for Frontends (BFF)**

Scenario 2: Event-driven Applications

Services are executed in event-driven mode and resources are provided based on demands. Developers need not be concerned about service peaks or troughs. Idle resources are not billed and O&M costs will be reduced.

- **Image processing**
- **Live streaming/transcoding**
- **Real-time stream processing**
- **IoT rule/event processing**

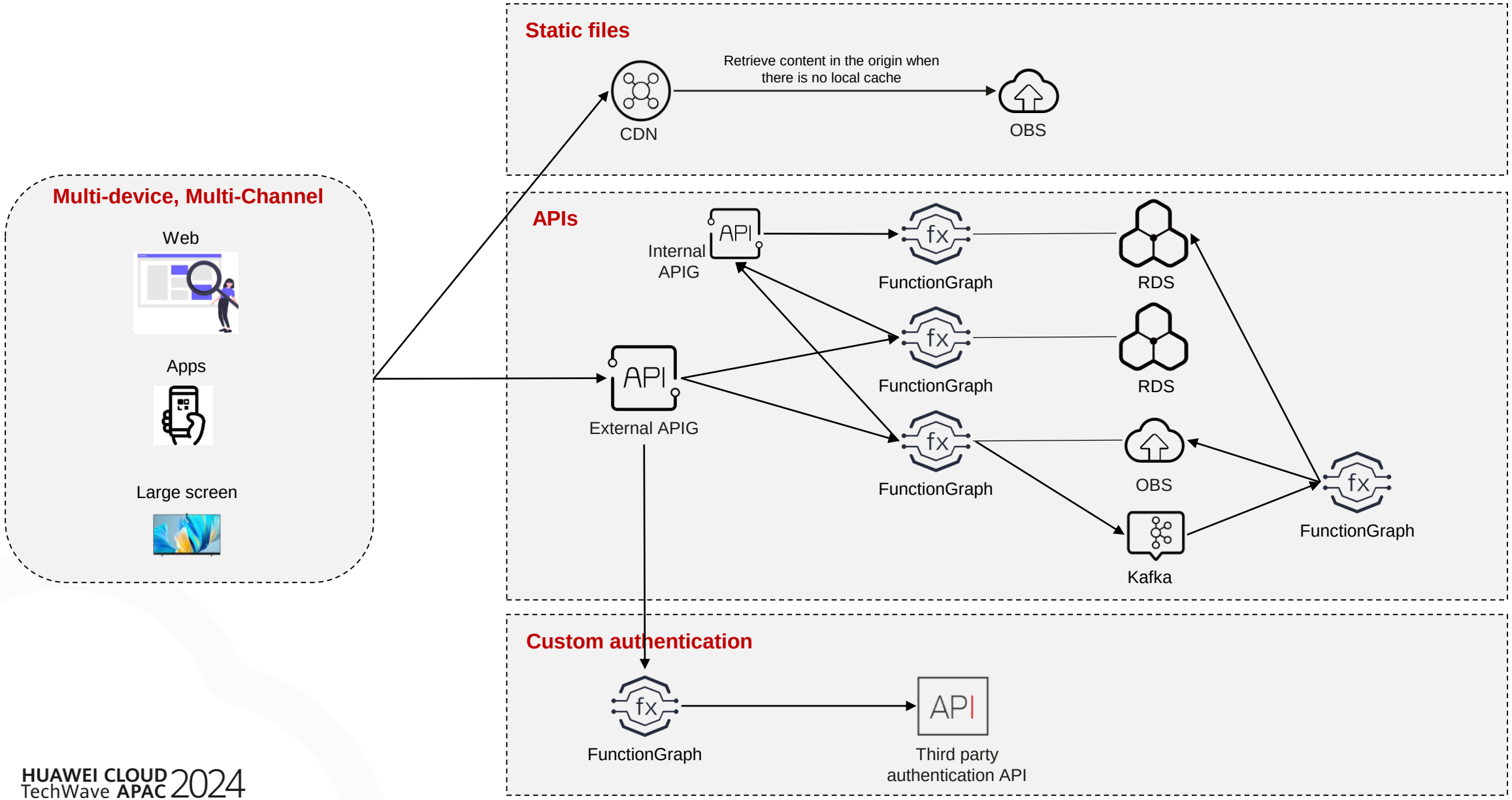
Scenario 3: AI Applications

Intelligence evolution requires various services to be integrated for quick rollout.

- **AI inference**
- **Facial recognition**
- **License plate recognition**
- **Optical Character Recognition (OCR)**

- **Automate operations / Fix operations**
- **+ Many more on your imagination**

Building Web and Mobile Applications with FunctionGraph



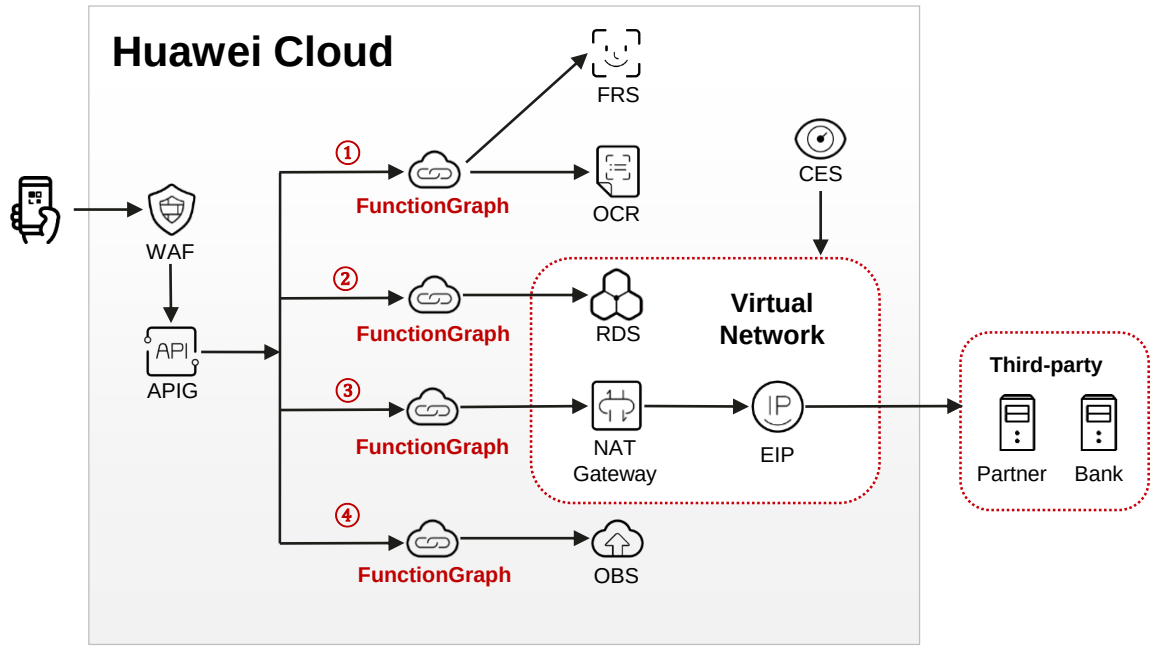
Thailand BSS: Building Mobile Application Backends with FunctionGraph with 15% Less Cost

Bangkok Smartcard System Company Limited (BSS) launched the Rabbit Card in 2012 for e-payment in local transportation, retailing, and tourism industries. It has its own membership service system and business ecosystem. BSS wants to establish a self-operated online platform MyRabbit to deliver e-KYC, recharge, bill query, and rebate, and develop an independent online ecosystem.

Pain points

- As required by local laws and regulations, BSS must provide e-KYC (real-name authentication), which depends on AI services such as facial recognition and text recognition. And it is beyond the capability of the customer.
- As the user conversion rate is uncertain, the top concern of this platform is cost-effectiveness, followed by architecture elasticity and reliability.

Solution



Benefits

- Quick integration and service compliance with local laws and regulations
- Auto scaling: Resources are automatically scheduled for functions to process requests during peak hours
- **15% less costs**

Huawei Cloud EventGrid Helps Automotive Enterprise Achieve Serverless IoV and Zero Event-Driven Data Loss

Customer profile

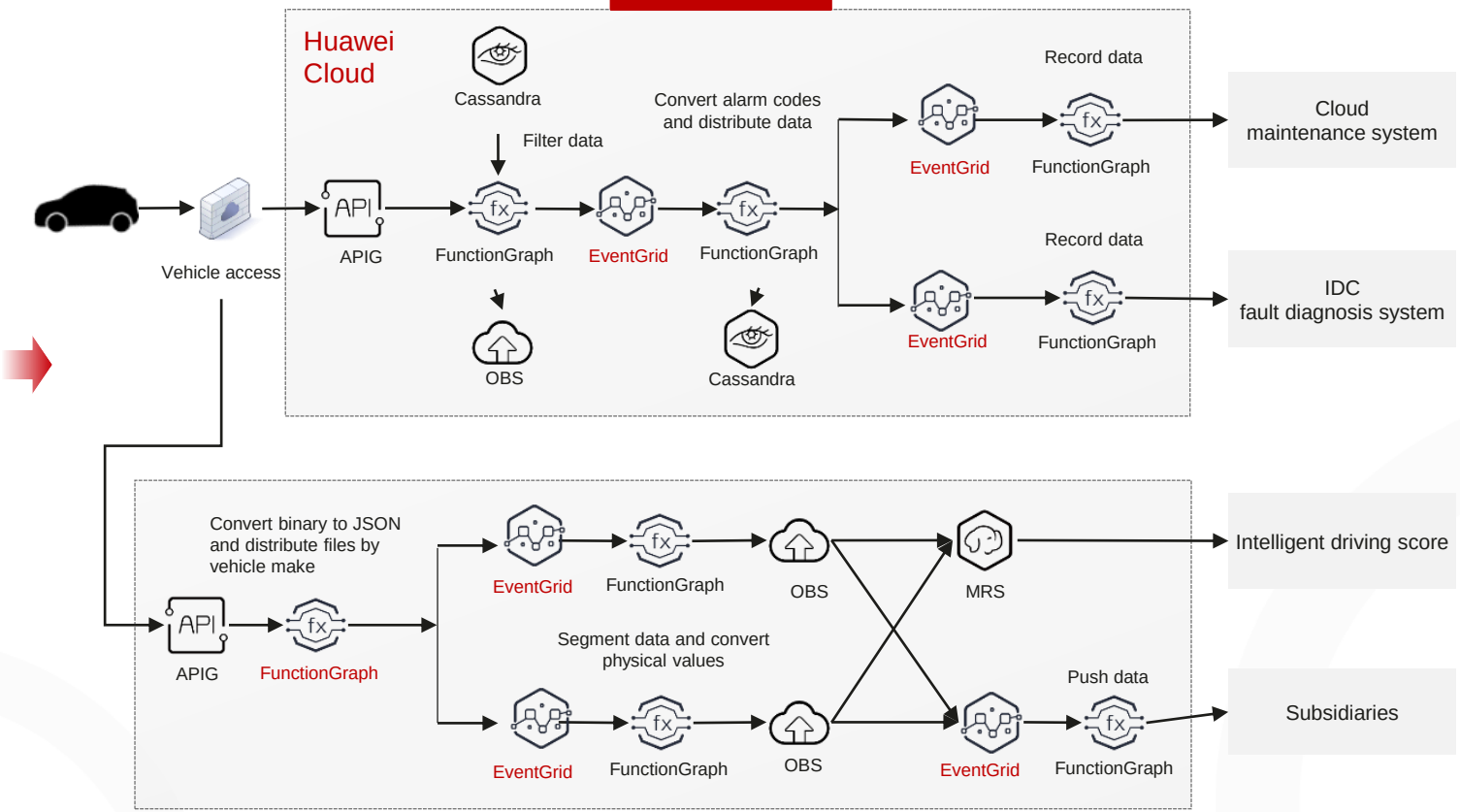
A global top vehicle vendor with 10 million-unit annual sales volume uses EG+FG for powerful, reliable serverless in IoV scenarios.

Pain points

High downstream pressure
Uneven up/downstream pressure

- Heavy interconnection workload
  P2P cloud services
- Data loss and unreliability
  Abnormal flow control or downstream events
- Data cleaning rules
  Code change needed

Solution



Benefits

- Asynchronous decoupling, traffic balancing
  Event cache relieves downstream pressure
- Development workload **reduced by 50%**
  Event center connection with no code
- High reliability and **zero data loss;** Failed event redelivery and tracing
- Data cleaning rules changed with no code change
  Data cleaning rules configurable on the console

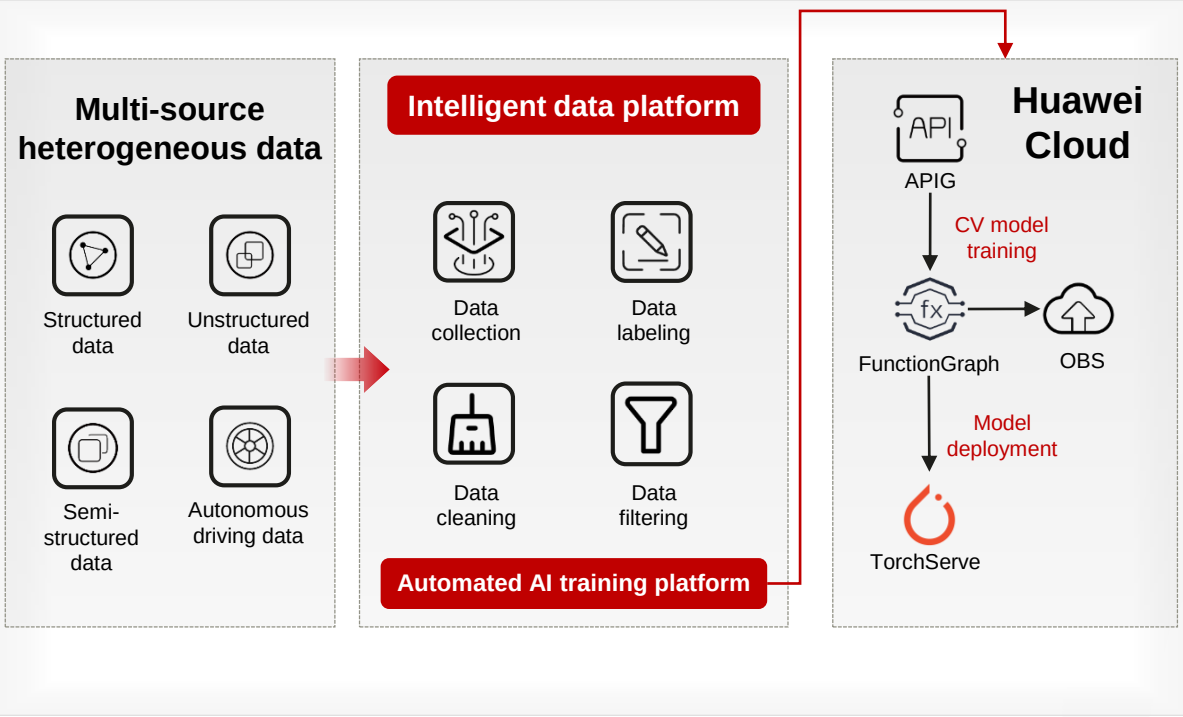
FunctionGraph Automatic AI Training Boosts Utilization by 30% and Savings by 40%

MindFlow: an industry-leading AI infrastructure provider of intelligent data platform for data lifecycle management, AI data middle-end, and basic data services (data labeling, data collection, and data cleaning).

Pain points

- Low GPU hardware automation and elasticity
- Low GPU resource utilization: triggered by service events, temporary stable running

Solution



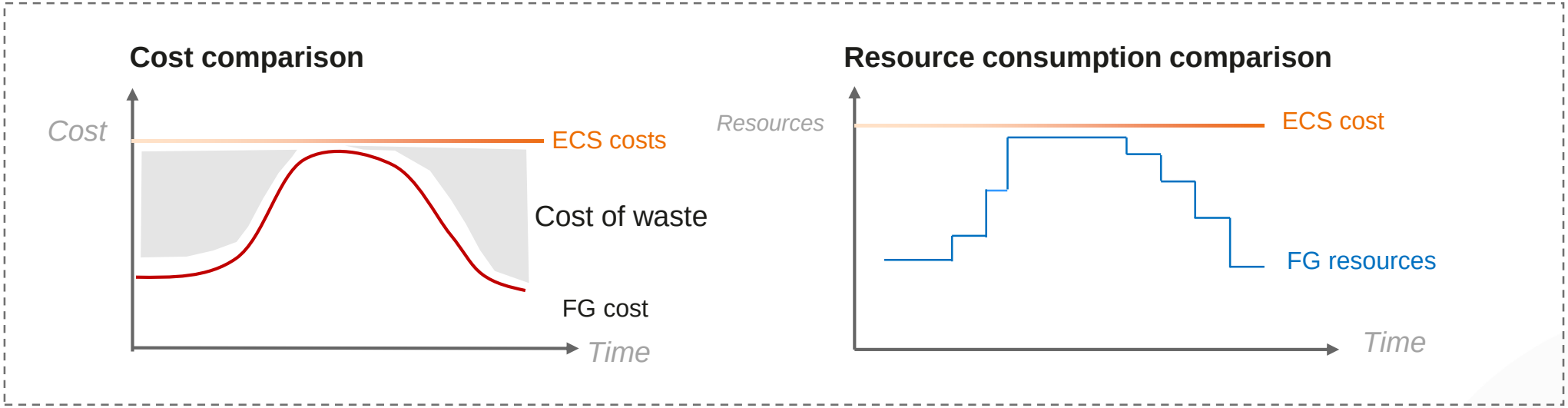
Benefits

- **Cost reduced by 40%**
  GPU sharding enables at least 1 GB GPU computing power to be configured for maximum resource usage.
- **0 idle resource fees**
  FunctionGraph automatically scales instances to service requests. Users are only billed for received requests.

Comprehensive Monitoring Capabilities are Included Natively

- **Metrics**: Interconnected with Cloud Eye
  - Invocations
  - Errors
  - Duration (maximum, average, and minimum durations)
  - Throttles
  - Instance statistics (concurrency, reserved instances, elastic instances)
- **Logs**: Interconnected with LTS
- **Tracing**: Function traces (only for Java)

Lower Costs and Higher Efficiency



Auto scaling

Users do not need to configure any resources. FunctionGraph schedules underlying compute resources based on the number of concurrent requests, and scale out CPUs and GPUs in milliseconds.

Extremely low cost

Users are billed based on the time and number of FunctionGraph API calls, reducing costs by 90% or more

How Low?

Billing Item	Monthly Usage	Price (USD) Singapore Region	Pricing Basis
Requests	**≤ 1 million requests**	**0**	**Price per 1 million requests**
Requests	> 1 million requests	0.2	Price per 1 million requests
Metering duration	**≤ 400,000 GB-seconds**	**0**	**Price per GB-second**
Metering duration	> 400,000 GB-seconds	0.00001667	Price per GB-second

Competitive Pricing

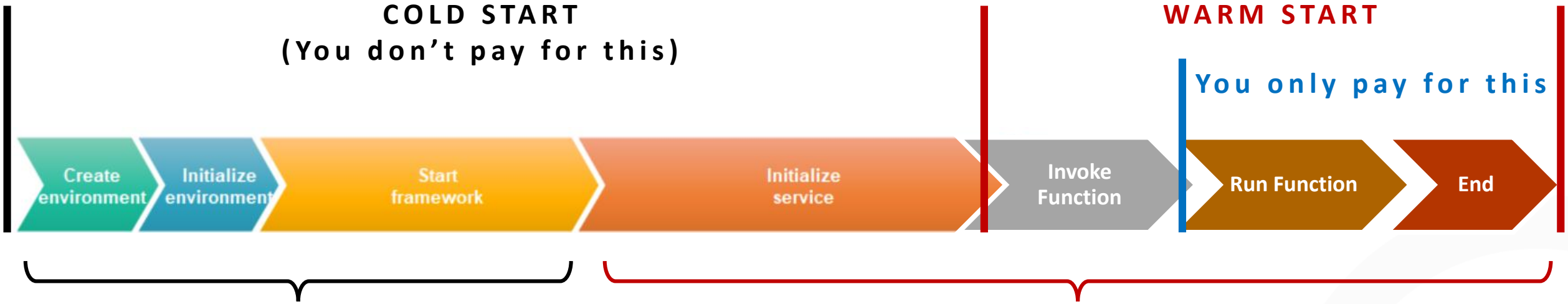
+

Only Pay when in Use

+

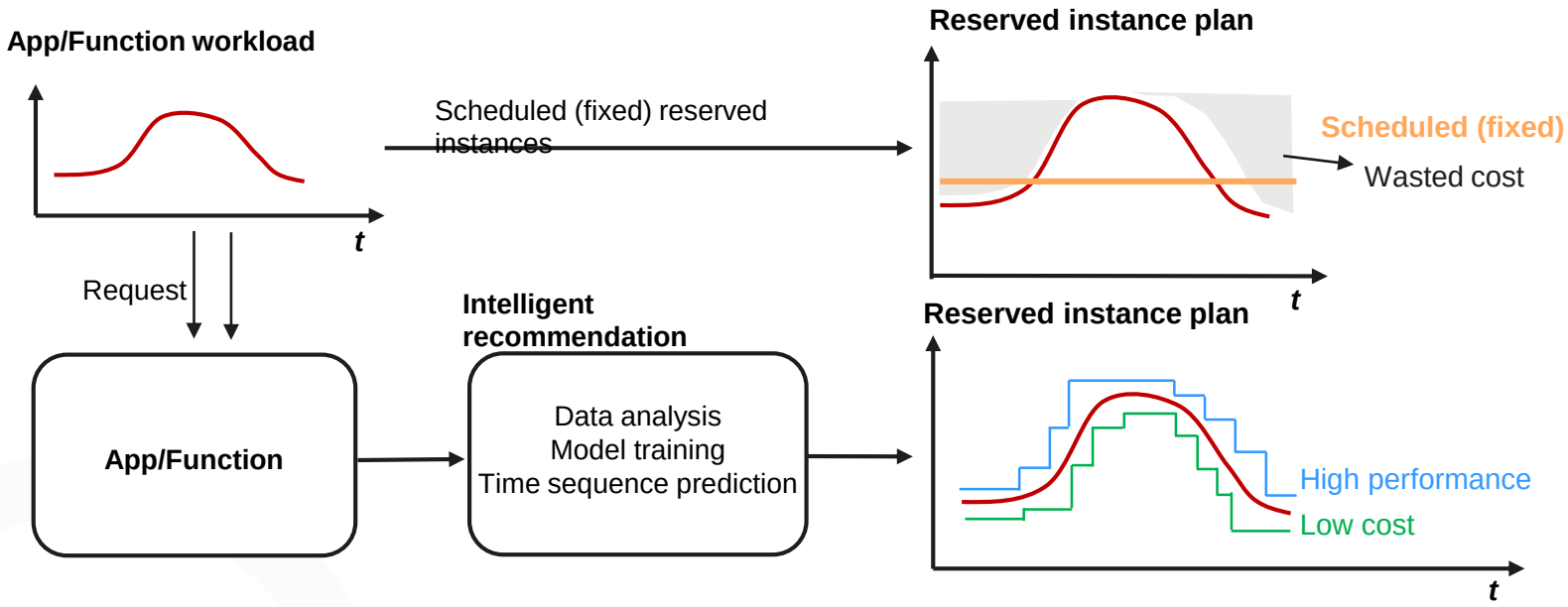
Free Tier!

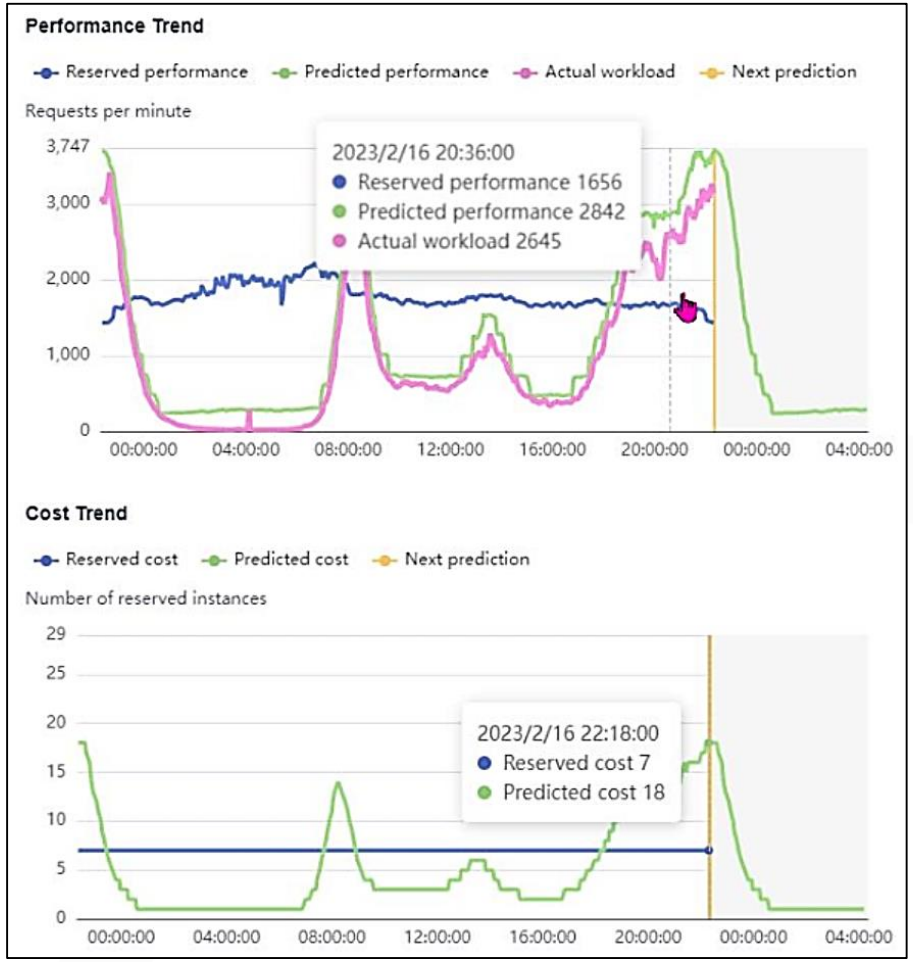
Ultra-fast Optimized Cold Start



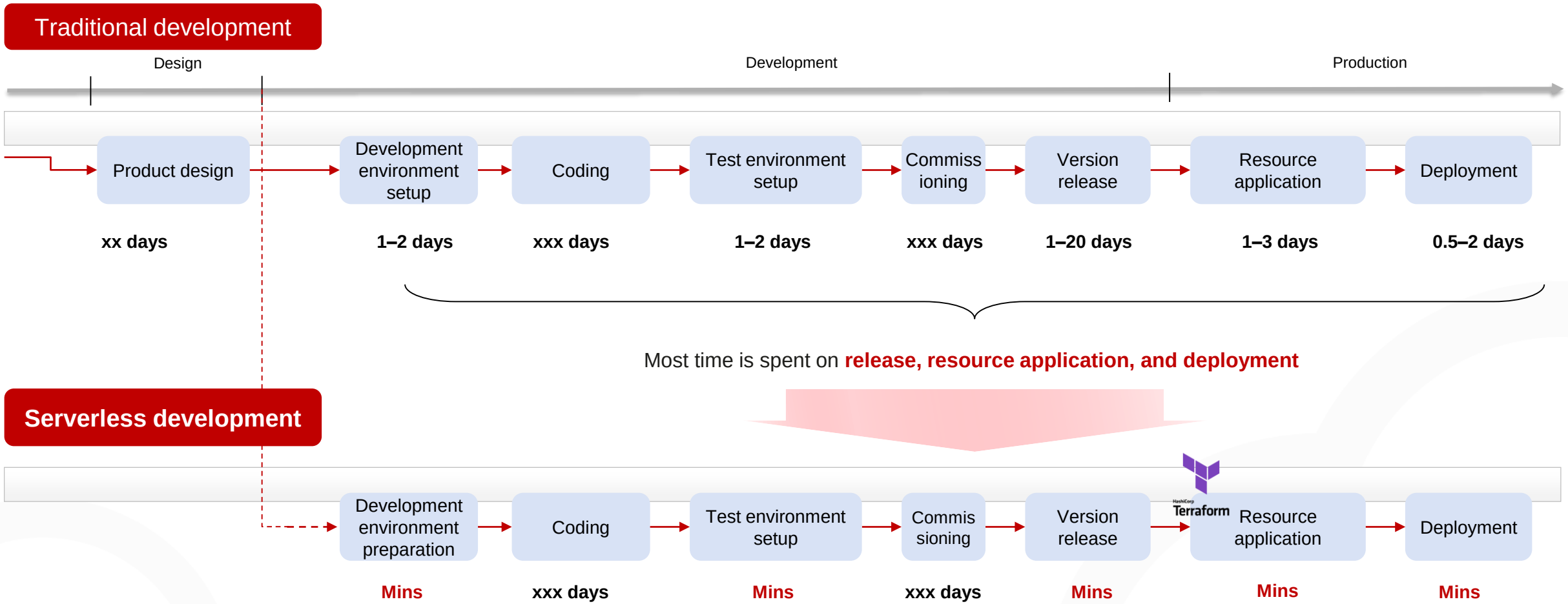
Reserved Instance & Intelligent Recommendation Policy

- Fixed number of scheduled scaling policy
- Intelligent recommendation policy:
  High performance, balance, low cost





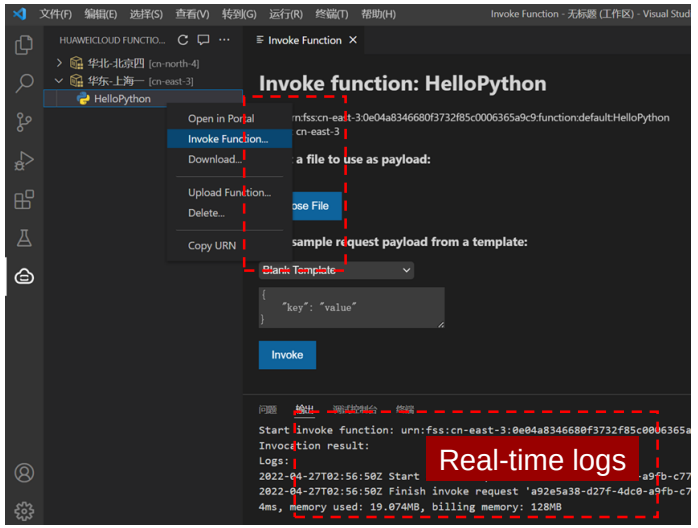
Improving Developers and DevOps Experience



Various Tools to Lower Learning Cost, Strengthen Engineering Efficiency, and Shorter TTM

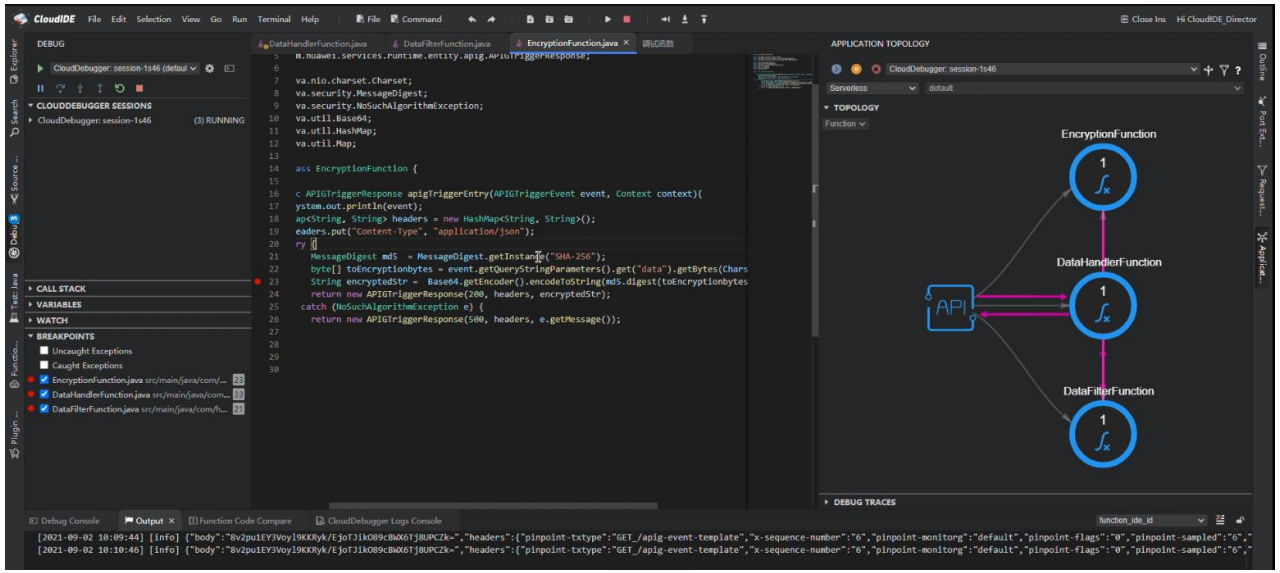
(Off-cloud) VSCode Extension and CLI

1. Create and deploy function with a wizard
2. Test function on the cloud, download code for debugging, and update it to the cloud



(On-cloud) CloudIDE

1. Create and deploy function with a wizard
2. Test function on the cloud, download code for debugging, and update it to the cloud
3. Debug function **online** (Java and Node.js)



Command Line tools

- Huawei KooCLI
- HashiCorp Terraform
- SERVERLESS DEVS
- serverless

Use What You Like or are Familiar With

8, 11

6, 8, 10, 12, 14, 16, 18

2.7, 3.6, 3.9, 3.10

1.x

.Net Core 2.1, 3.1, 6

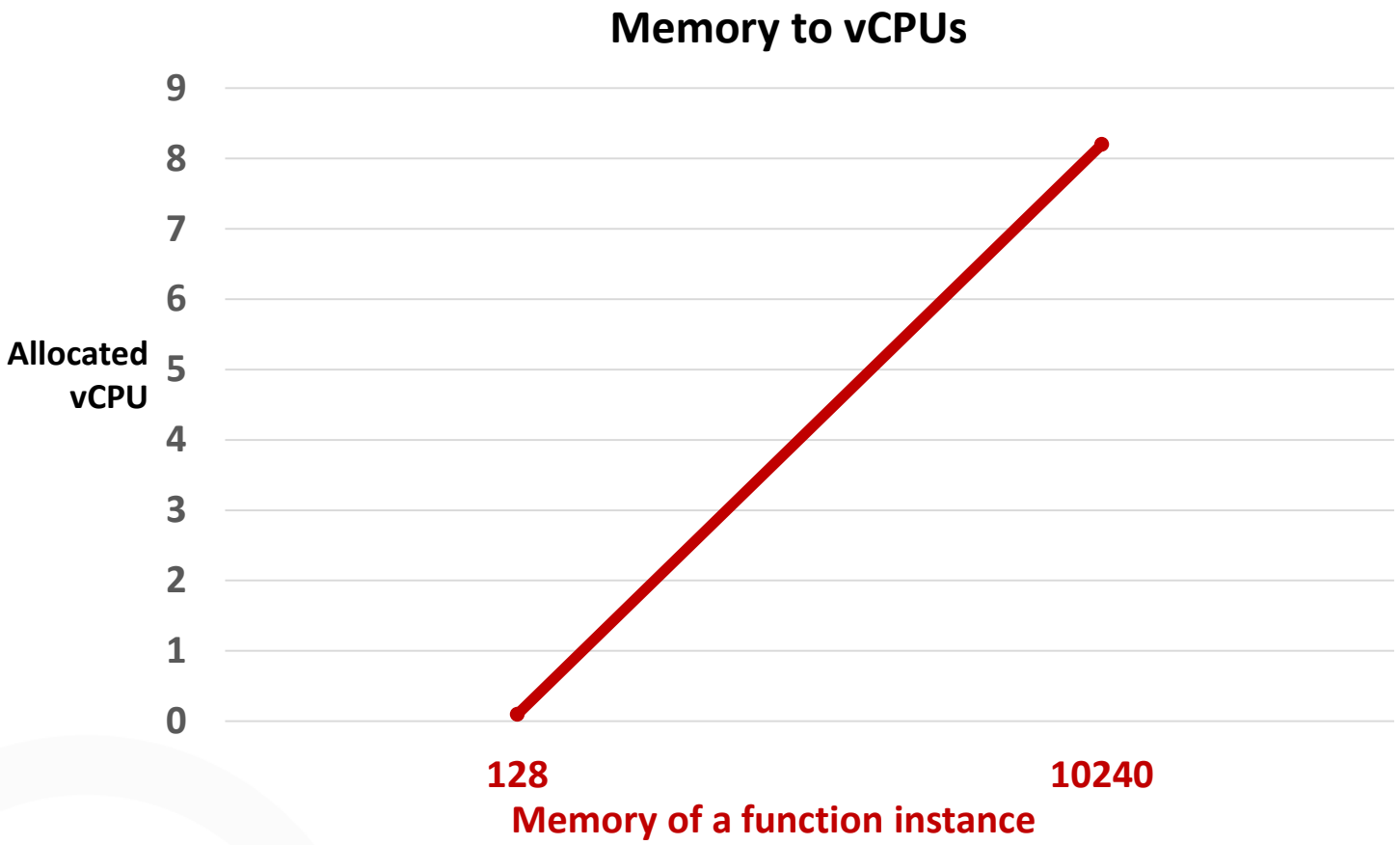
7.3

Shell scripts or Linux executables

Custom Runtime

Container Image

The Only Key Configuration is How Much Memory You Need



- Configure up to **10GB of memory**
- Larger memory → faster processing
- However, processing time is non-linear
- Thus, costs not always cheaper if faster
- Monitor and understand the trade-offs
- **Dynamic memory allocation is supported**

Some Important Things to Note for FunctionGraph

	Synchronous	Asynchronous
Max execution duration	900 secs	**72 hours**
Payload size	6MB for both request and response	256KB for both request and response
Max size of code deployment package	40 MB	
Max size of original code deployment package allowed during function API invocation	Zip 1500MB after decompression, OBS 300MB after compression	

Pitfalls when Using FunctionGraph

1. Monolith serverless?!
2. High number of concurrent DB connections (go through a database proxy instead)
3. Large volume of data transfers
4. Chattiness
5. Persistent or High consistent usage

Quick Summary - What Can Serverless Do For You?

1. Less Overhead, Better Focus
2. Extremely High Scalability
3. Greater Agility & Flexibility
4. Faster Time to Market
5. Better Cost Efficiency & Resource Utilization
6. Better Maintainability & Sustainability

Adding FunctionGraph to Your Architecting Arsenal Can Be Powerful

APPLICATION MANAGED SERVICES

ServiceStage
(computing on VMs and containers)

CAE
Cloud Application Engine
(computing on serverless containers)

COMPUTE SERVICES

CCE
Cloud Container Engine
Managed K8s

CCI
Cloud Container Instance
Serverless K8s

FunctionGraph
Function-as-a-Service

FunctionGraph,

a (*better*) alternative for Architects,
Developers, DevOps and Enterprises

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home and organization for a fully connected, intelligent world.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

HUAWEI