

The background of the slide features a soft-focus image of numerous hot air balloons floating in a sky that transitions from a pale blue at the top to a warm orange and yellow near the horizon, suggesting a sunrise or sunset. The balloons are in various colors and sizes, scattered across the frame.

DESIGNING FOR THE SERVERLESS AGE

MindMup.com migration

Heroku Feb 2016 ⇒ Lambda Feb 2017

- **~ 1/2 hosting cost**
- **~ +50% active users**
- **estimated ~66% savings**

Serverless Computing: Economic and Architectural Impact

Gojko Adzic

Neuri Consulting LLP

Neuri Consulting LLP
100 Abchurch Lane
London, United Kingdom WC2A 1AL
gojko@neuri.co.uk

Robert Chatley

Imperial College London

Imperial College London
180 Queen's Gate
London, United Kingdom SW7 2AZ
rbc@imperial.ac.uk

<https://dl.acm.org/citation.cfm?id=3117767>

ABSTRACT

Amazon Web Services unveiled their 'Lambda' platform in late 2014. Since then, each of the major cloud computing infrastructure providers has released services supporting a similar style of deployment and operation, where rather than deploying and running monolithic services, or dedicated virtual machines, users are able to deploy individual functions, and pay only for the time that their code is actually executing. These technologies are gathered together under the marketing term 'serverless' and the providers suggest that they have the potential to significantly change how client/server applications are designed, developed and operated.

Lambda¹, which was first announced at the end of 2014 [7], and which saw significant adoption in mid to late 2016. All the major cloud service providers now offer similar services, such as Google Cloud Functions², Azure Functions³ and IBM OpenWhisk⁴. This paper primarily discusses AWS Lambda, as this was the first platform to launch and is the most fully featured.

Historically, application developers would procure or lease dedicated machines, typically hosted in datacentres, to operate their systems. The initial capital expenditure required to purchase new machines, and the ongoing operational costs, were high. Lead times loads in systems with varying demand required advance planning

**Lambda fundamentally
changes financial
incentives for
deployment
architectures**

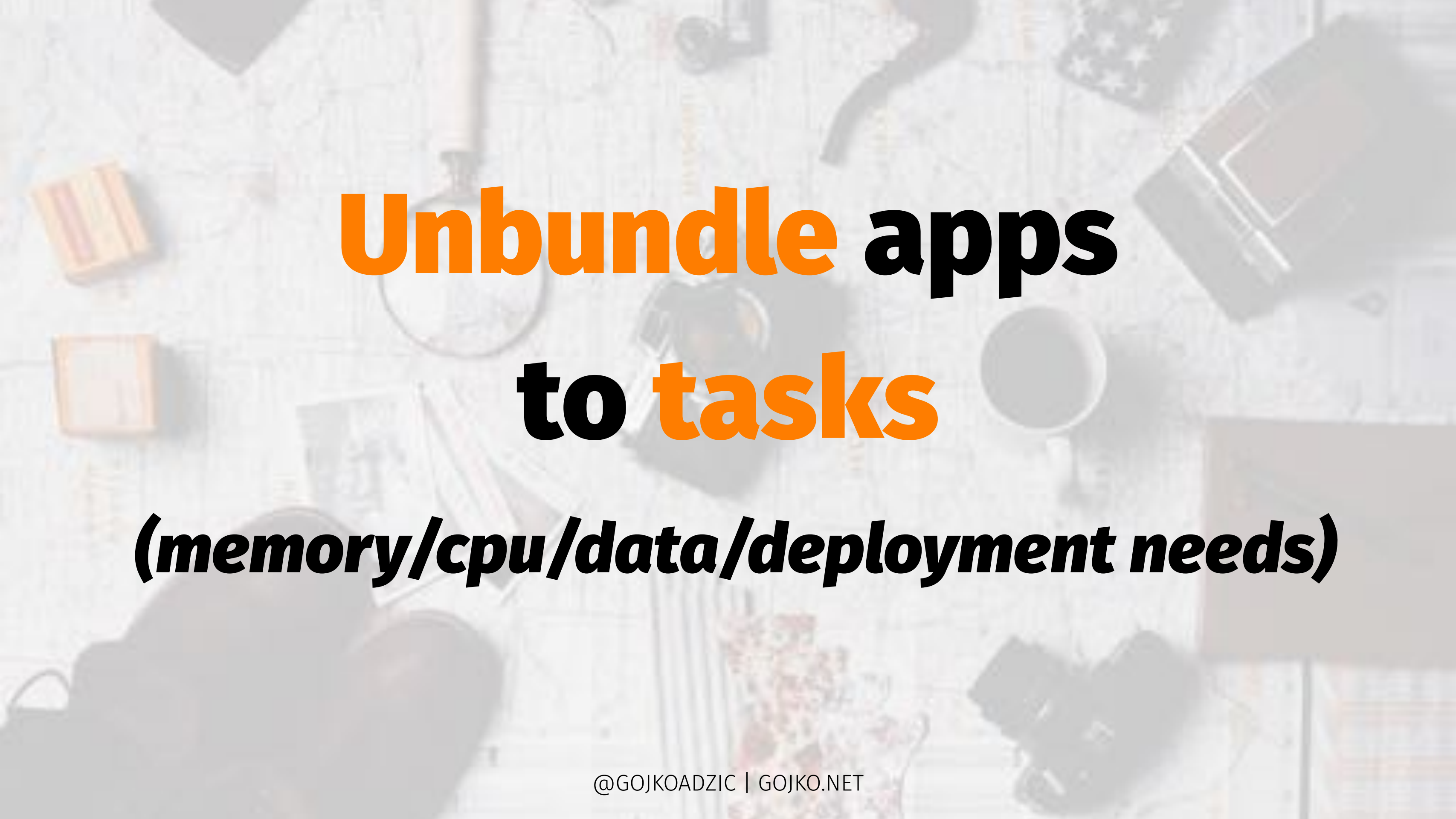
Pricing

- \$0.00000002 **per request**
- \$0.0000000834 **for 100ms @ 512MB**
- First **1 million** requests per month are **free**

**Historically, good
architecture optimised
for reserved resources**

⇒ **Bundling into apps**

Reserved ⇒ **Utilised** **Capacity**



Unbundle apps **to tasks**

(memory/cpu/data/deployment needs)



**Historically, good
architecture optimised
for quick fail-over**

⇒ **complex state
management**



Time to Recover $\Rightarrow$ **Time to Start**

Empirical numbers

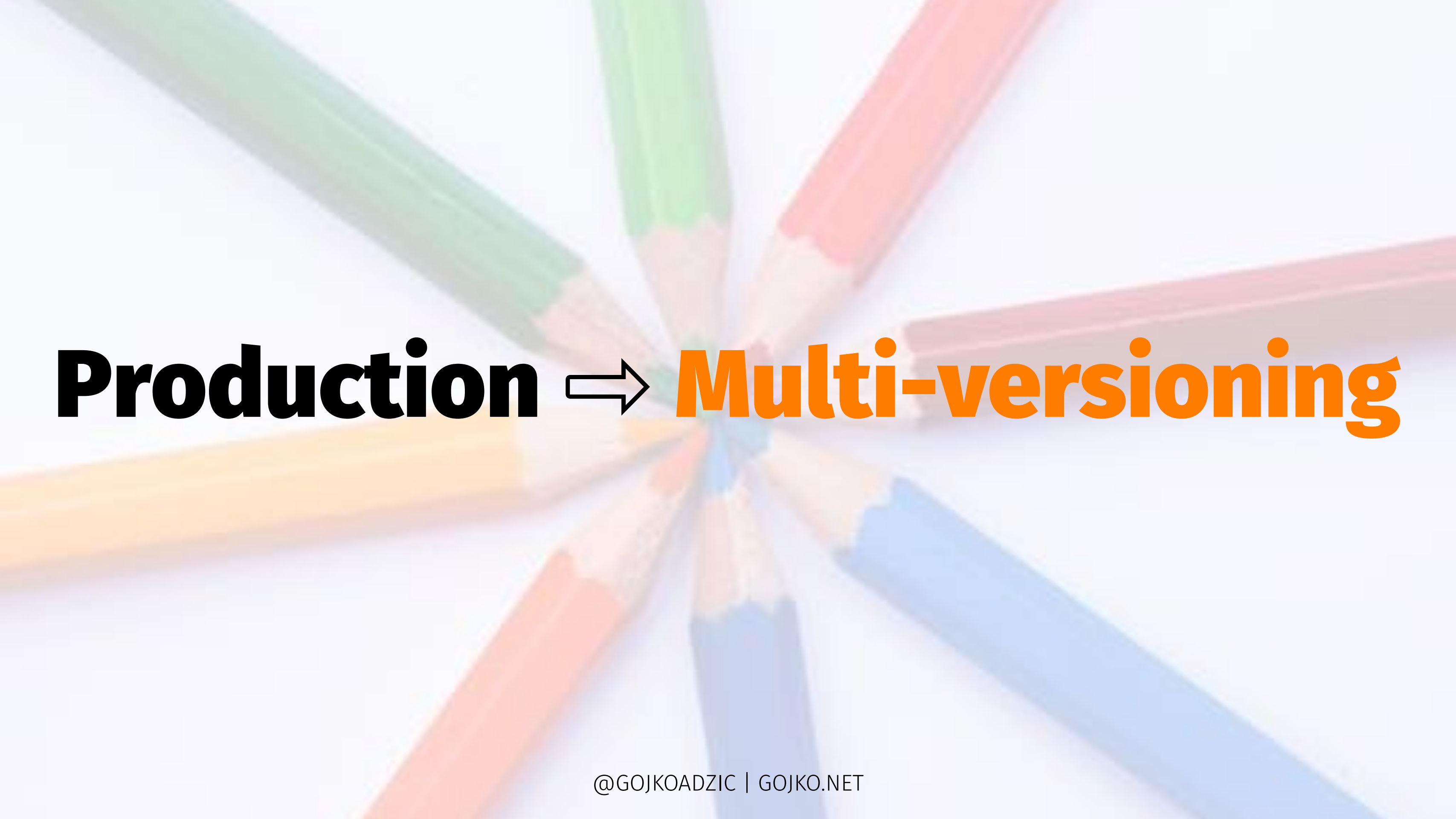
(not confirmed by AWS)

- **New instance (js) $< 1s$**
- **New version (js) $< 3-4s$**

Design for share-nothing *(not for stateless)*

**Historically, it was
difficult to replicate
"production"**

⇒ **expensive
experiments**



Production ⇒ **Multi-versioning**

Bundle tasks
around aggregates
(and invest in designing protocols)

Play **arbitrage** with different **charging models**

- **Lambda**: #requests, time, memory
 - **API GW**: #requests, transfer
 - **S3**: transfer (bytes)
 - **Cognito**: #users
 - **IOT GW**: #messages

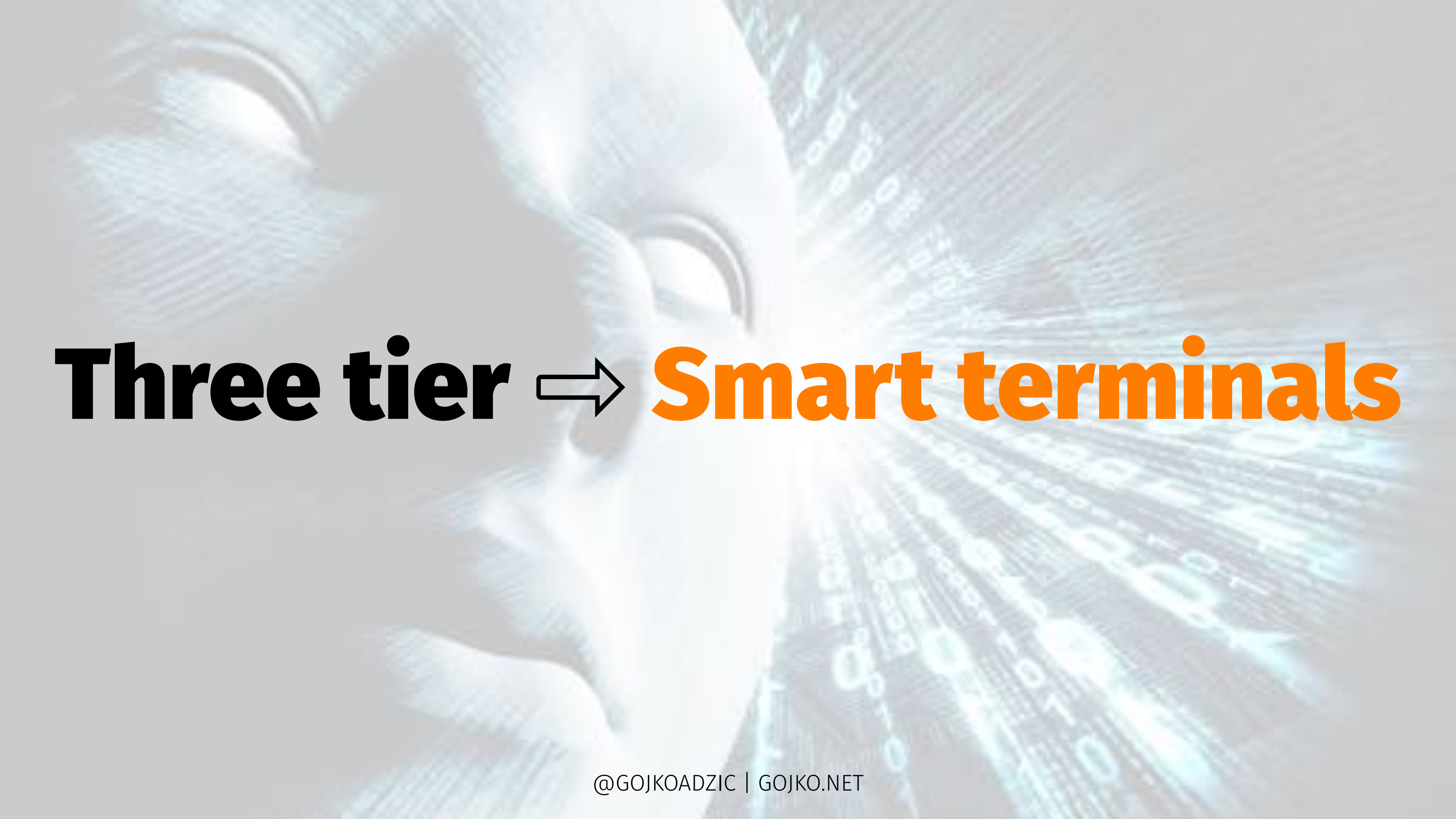


Managed apps $\Rightarrow$ **Platform glue**

```
api.get('/upload-file', request => {  
    'use strict';  
    const params = {  
        Bucket: request.env.bucketName,  
        Fields: {  
            key: 'upload/' + request.lambdaContext.awsRequestId  
        }  
        // + other limits, eg max size, file type...  
    };  
    return S3.createPresignedPost(params);  
});
```



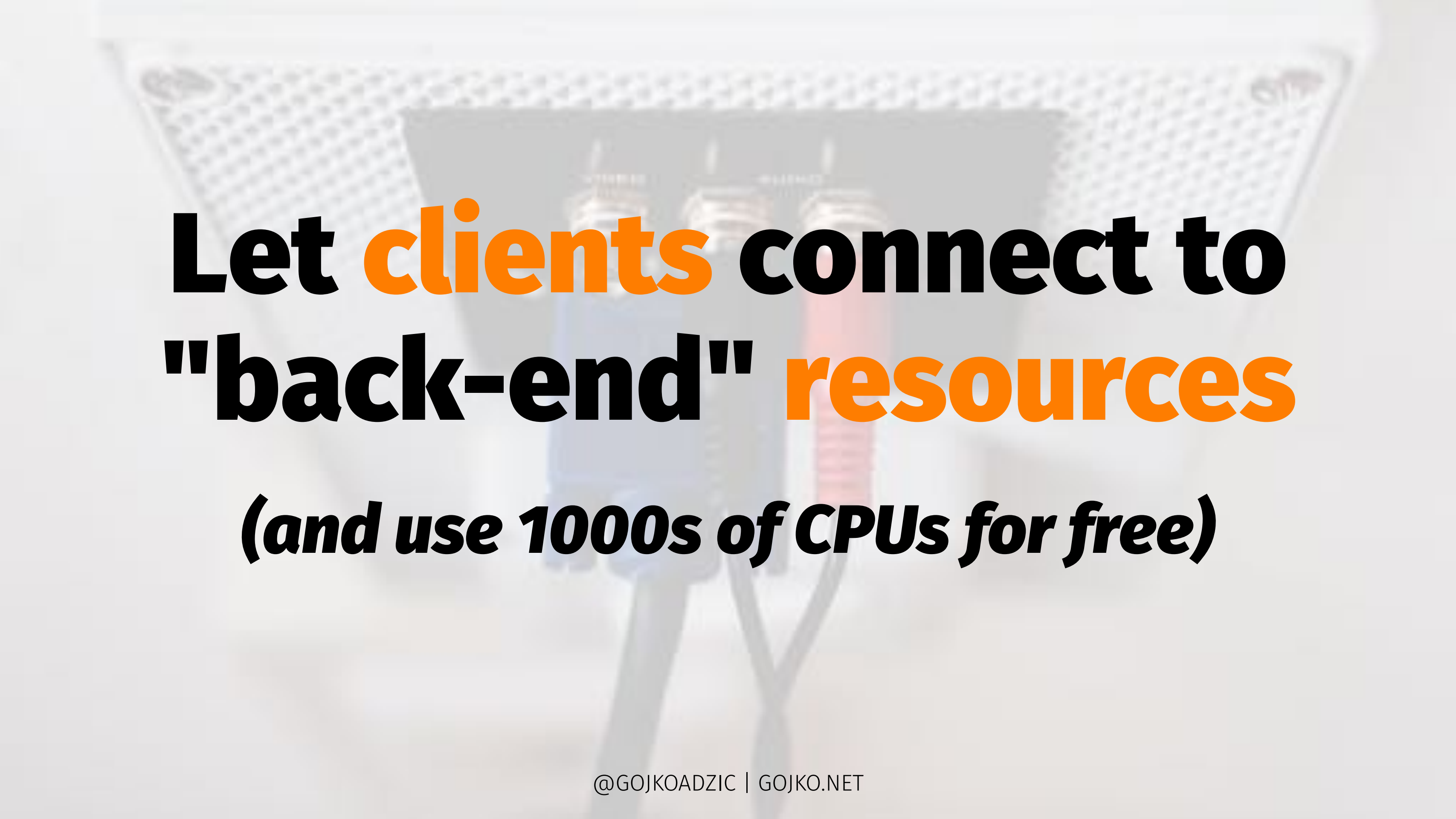
**Give the platform
usual server roles
(*gatekeeper, orchestration, sessions*)**



Three tier ⇒ **Smart terminals**

A background image showing three young women sitting outdoors, laughing and talking. The woman on the left is wearing a yellow dress and a white hat. The woman in the middle is wearing a pink dress. The woman on the right is wearing a white dress. They are all smiling and appear to be in a joyful conversation.

<https://serverless-chat.claudiajs.com/>



Let **clients connect to
"back-end" **resources****

(and use 1000s of CPUs for free)

A background image of three young women laughing and posing together. The woman on the left is wearing a yellow dress and a white hat. The woman in the middle is wearing a pink dress. The woman on the right is wearing a white dress. They are all smiling and laughing, creating a joyful and social atmosphere.

gojko.net
claudiajs.com