

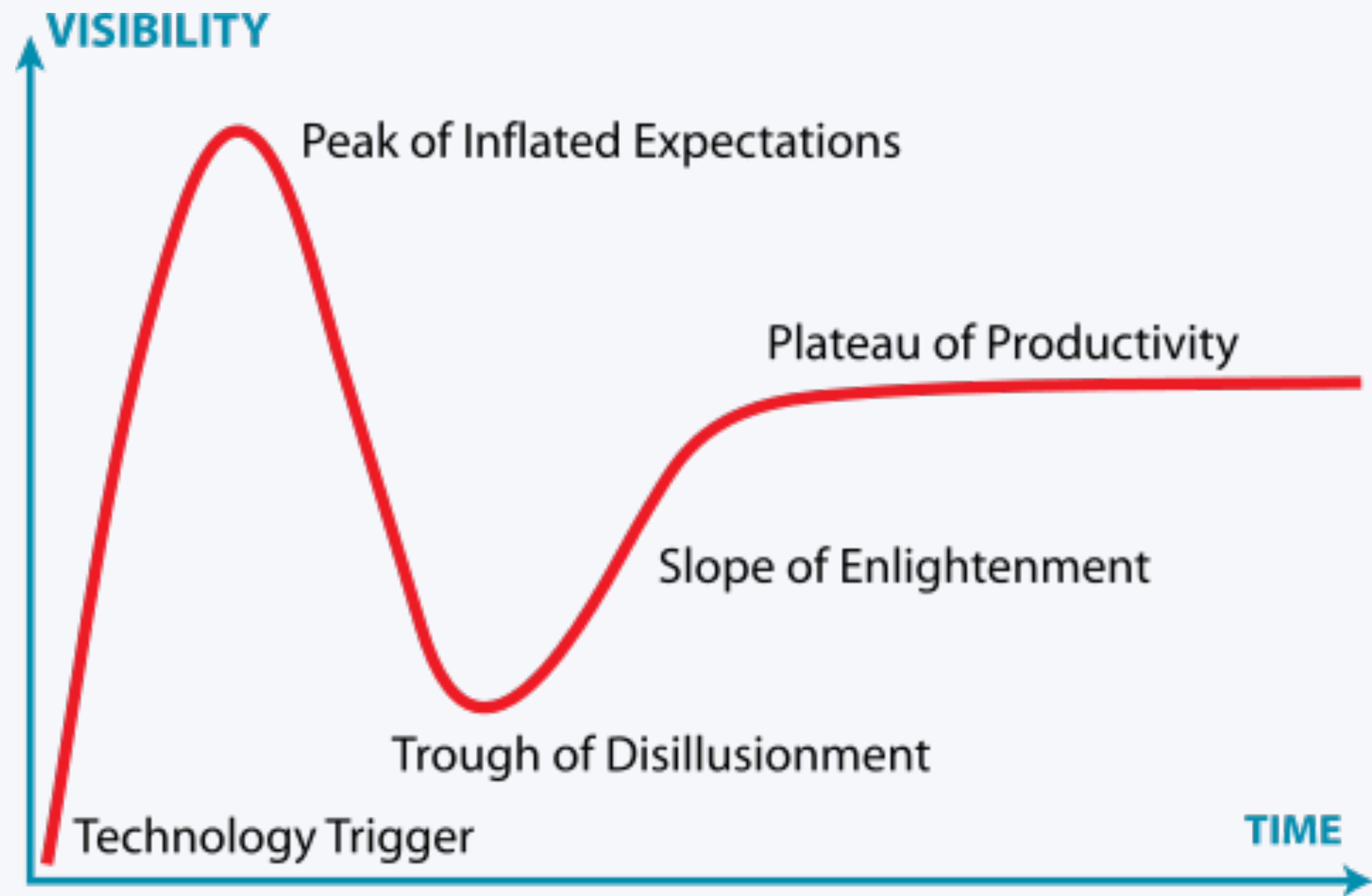
Serverless

Where Have We Come? Where Are We Going?

Stephen Fink
Distinguished Engineer
IBM Watson

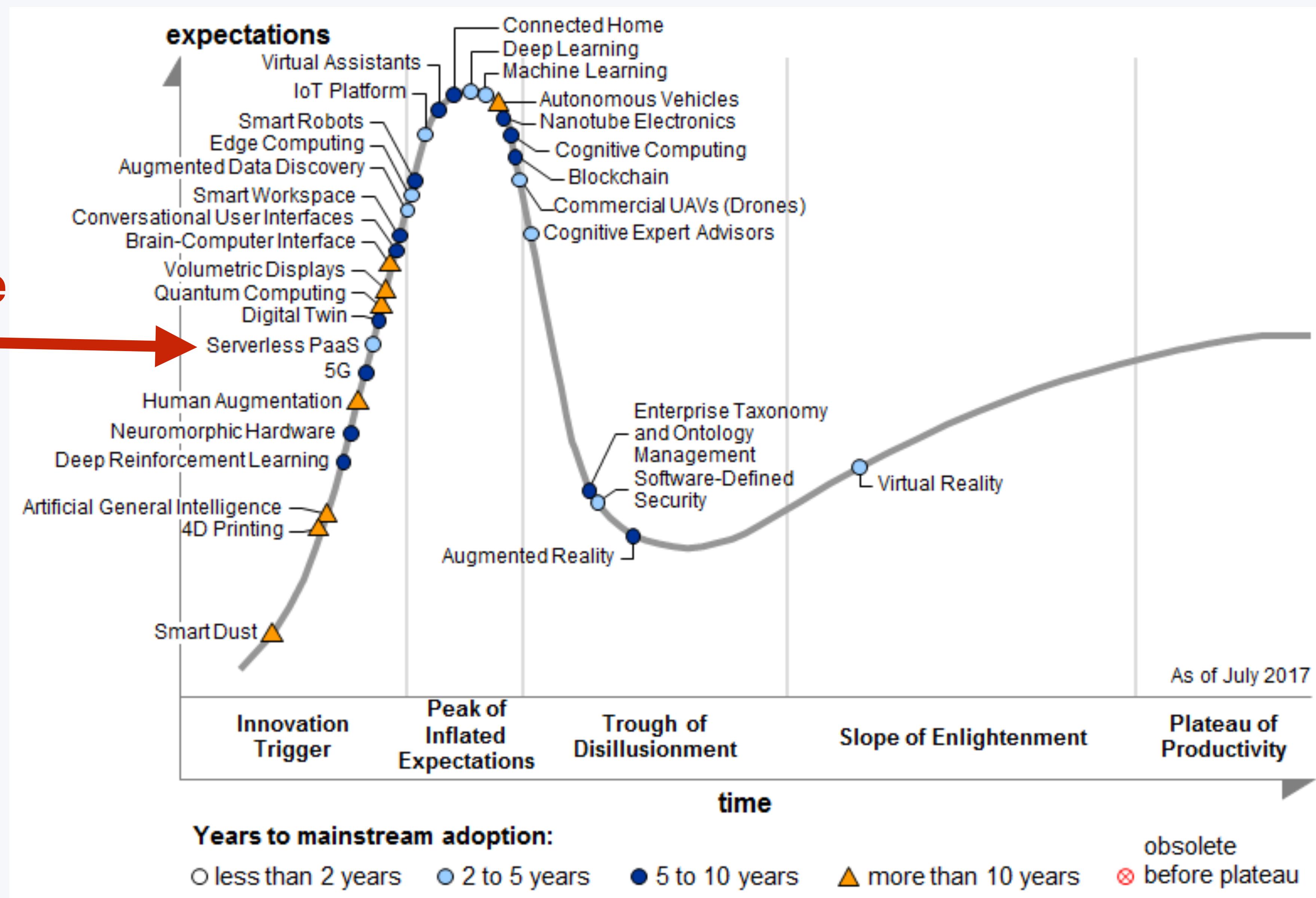
 @sjfink



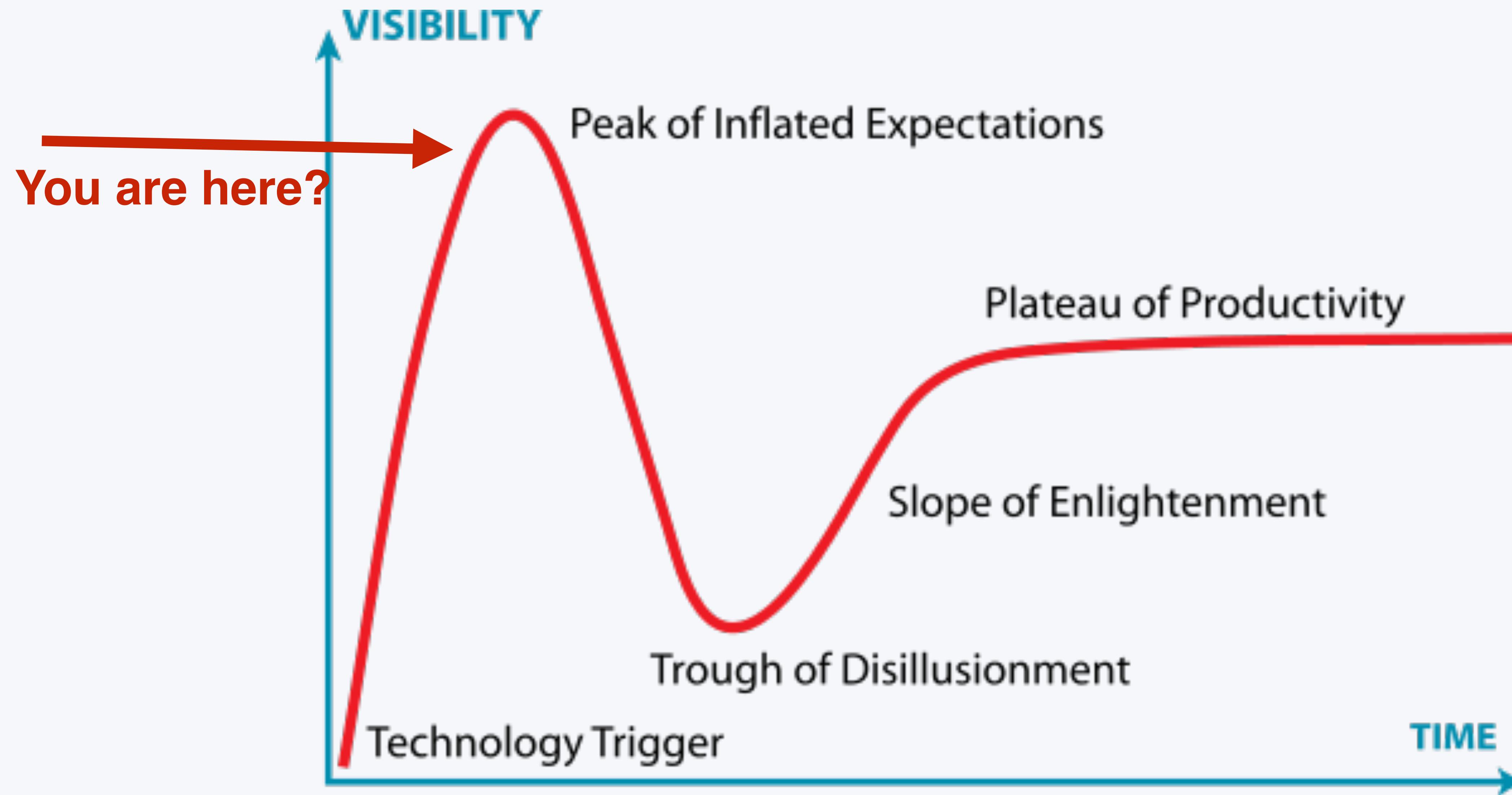


Gartner Hype Cycle for Emerging Technologies 2017

You are here



Serverless 2018?





Are you ready?

About your speaker



2014-2017 IBM's OpenWhisk/Cloud Functions



IBM Cloud Functions Apache OpenWhisk

About your speaker



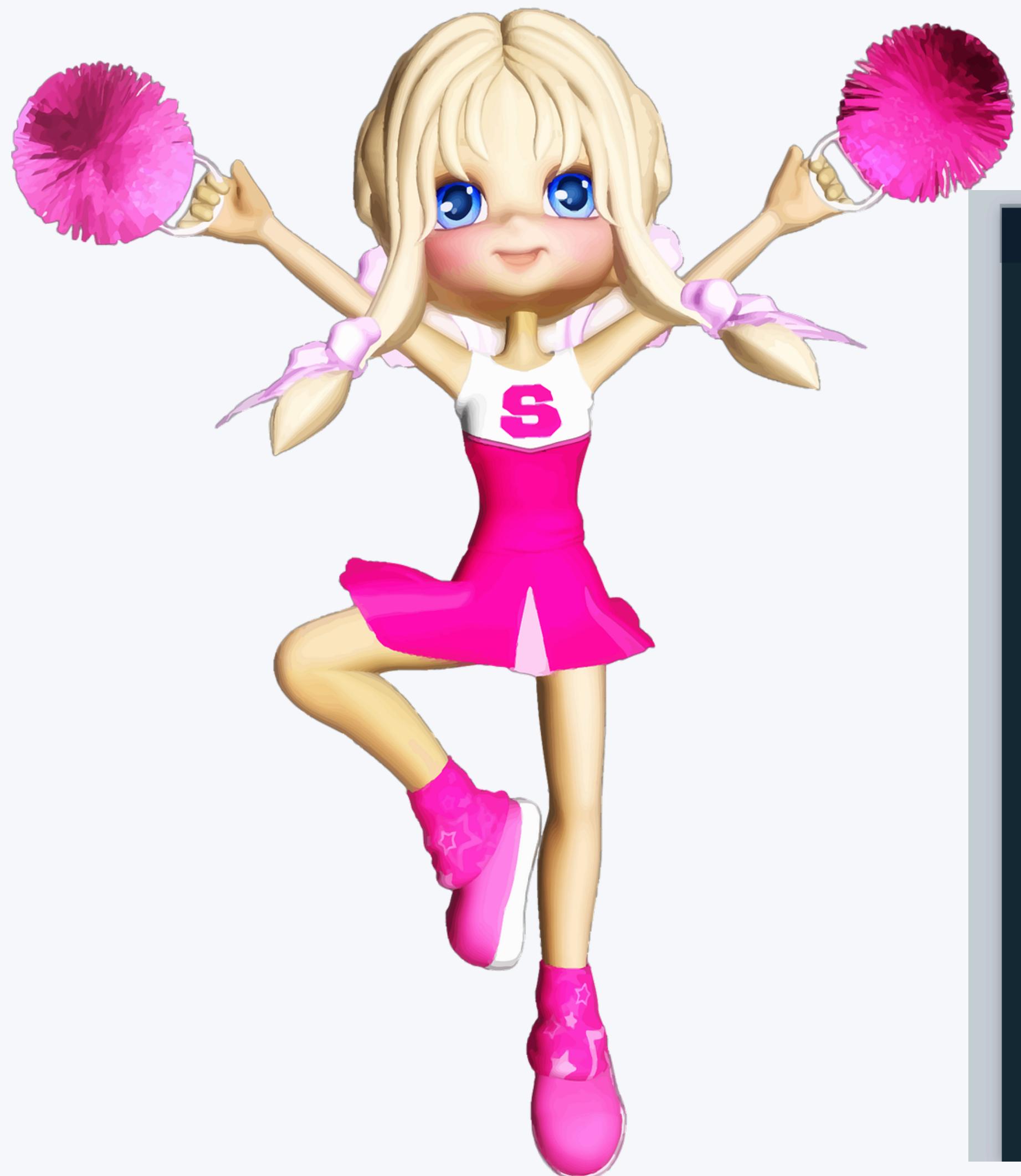
2014-2017 IBM's OpenWhisk/Cloud Functions
Serverless Advocate



IBM Cloud Functions



Apache OpenWhisk



Gave talks like

The slide features a dark background with a large, light blue gear icon on the right side. Above the gear is a teal satellite dish and a cluster of stars. The title 'The Serverless Revolution' is displayed in large, bold, teal letters. Below the title is a horizontal teal bar. The author's name, 'Stephen Fink', and title, 'Chief Architect IBM Watson Programming Models', are listed in white text. A Twitter handle '@sjfink' is shown below the author's information. At the bottom of the slide, there is a small copyright notice: '© 2017 IBM Corporation'.

The Serverless
Revolution

—

Stephen Fink
Chief Architect
IBM Watson Programming Models

Twitter icon @sjfink

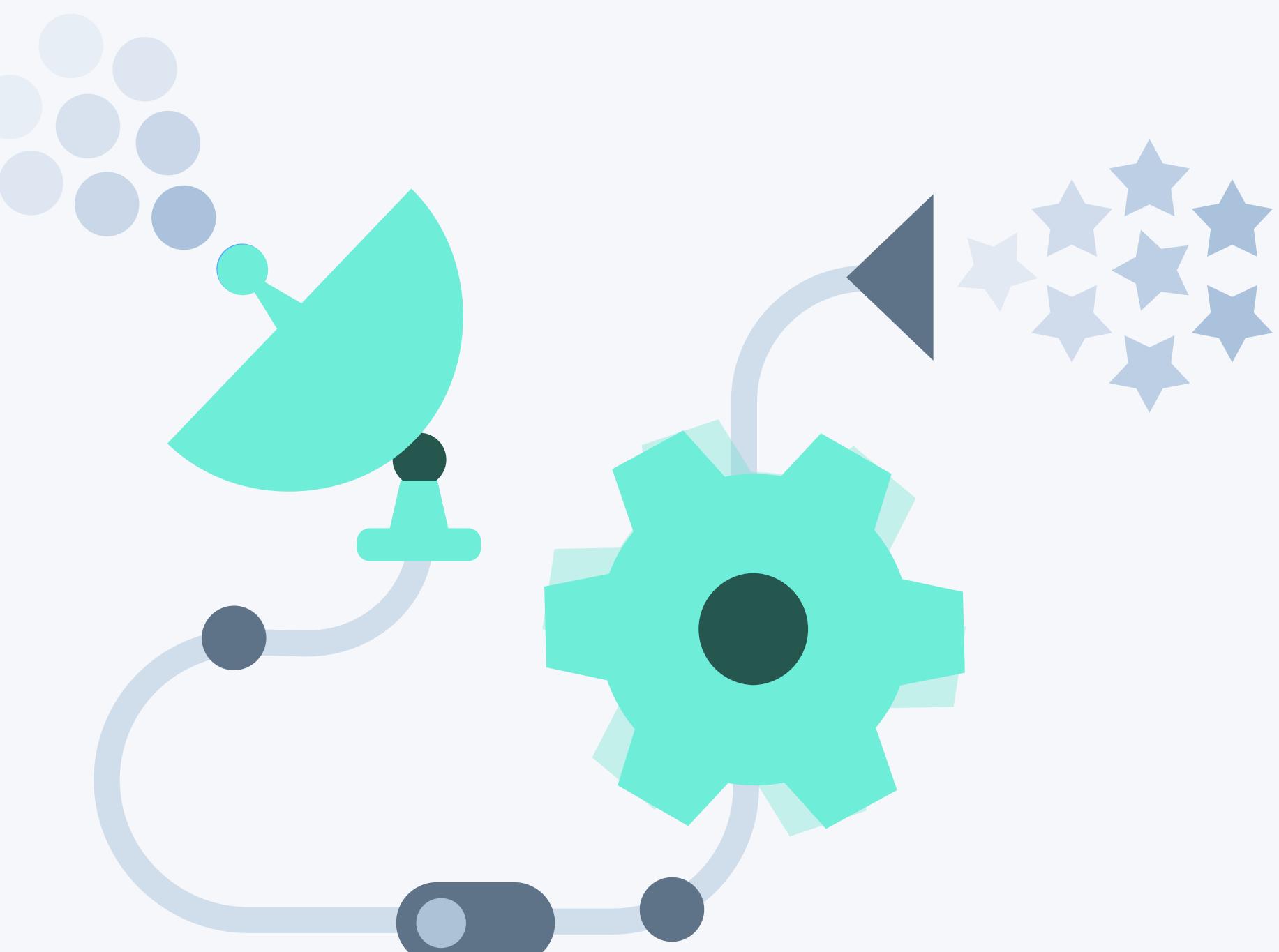
© 2017 IBM Corporation

With slides like ...



Serverless

“Making Development Fun Again”

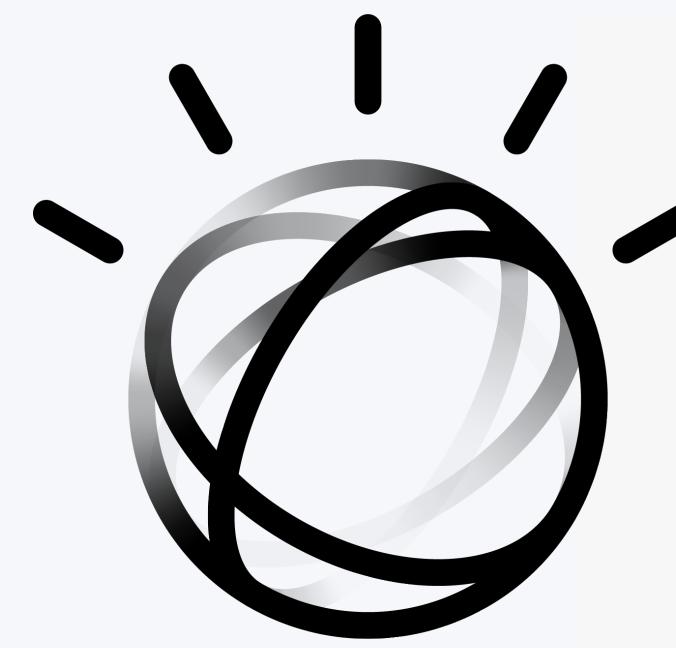


About your speaker

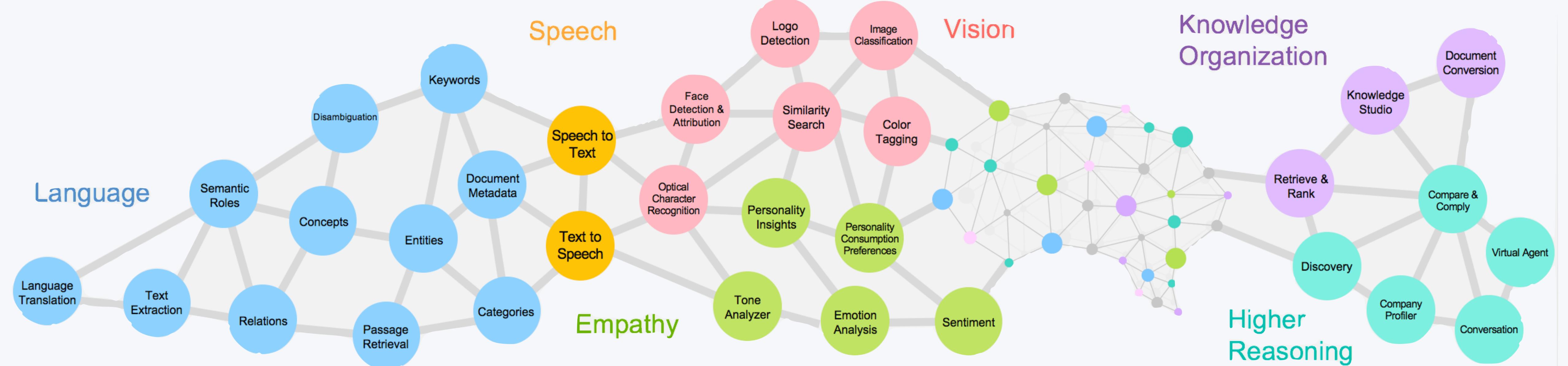


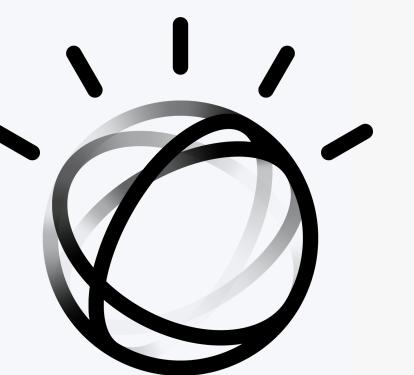
2014-2017 IBM's OpenWhisk/Cloud Functions
2017-2018 Architect for IBM Watson





Watson





Watson

large development organization



invested in
microservices



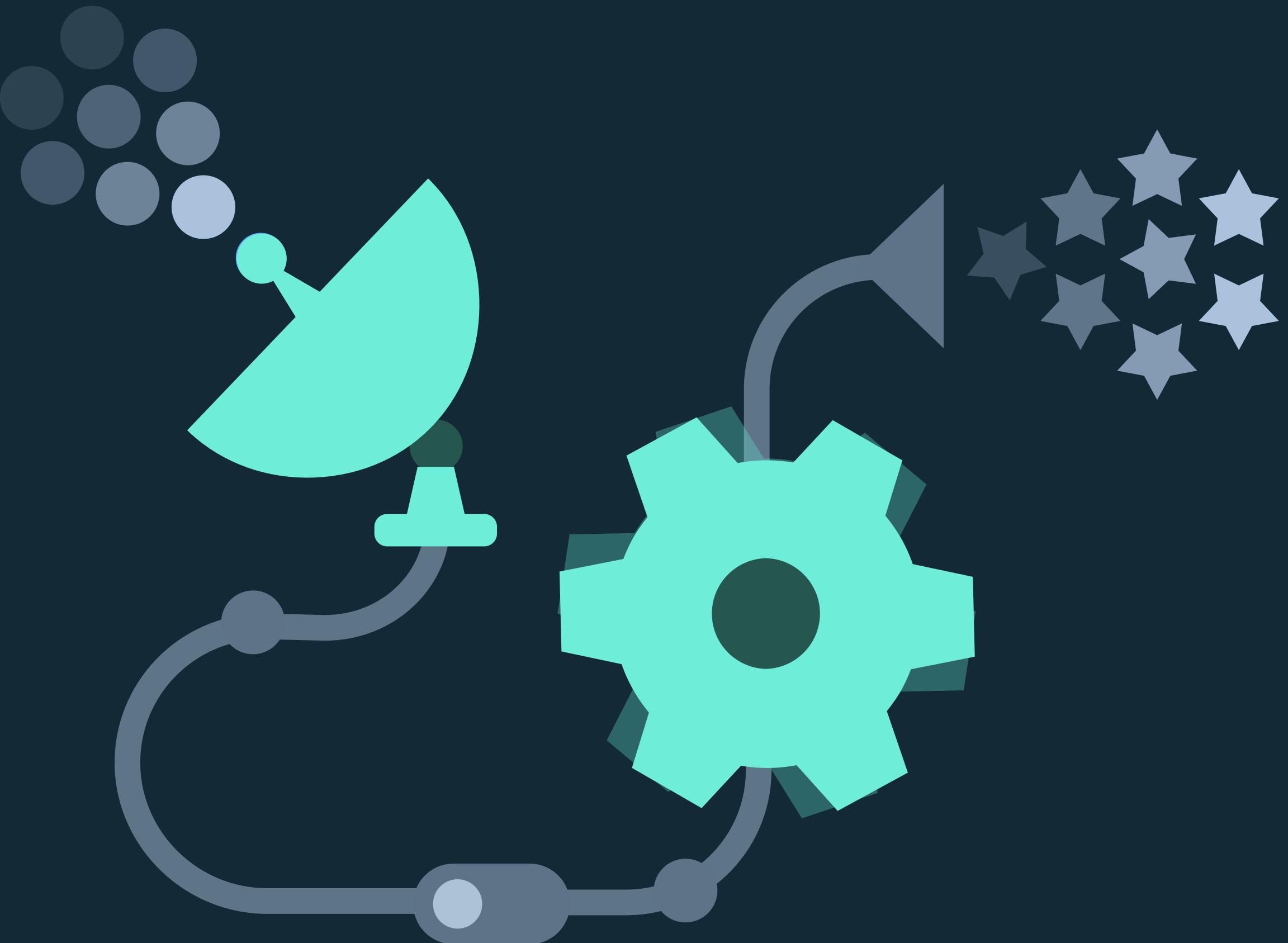
kubernetes

typical business
pressure



Serverless for Watson

Hype and Reality



Write functions.
Not plumbing.

Platform handles all
infrastructure transparently



No servers





conventional microservice architecture

Developer Responsibilities

business logic
service integrations
IaaS: programmatic
PaaS: declarative
containers
middleware
autoscaling
load balancing
service discovery
fault tolerance
logging
messaging
security patches



serverless architecture

Developer Responsibilities

business logic
service integrations (declarative)
REST API definition (declarative)

Platform Responsibilities

containers
middleware
autoscaling
load balancing
service discovery
fault tolerance
logging
messaging
security patches





Welcome
To Reactivity

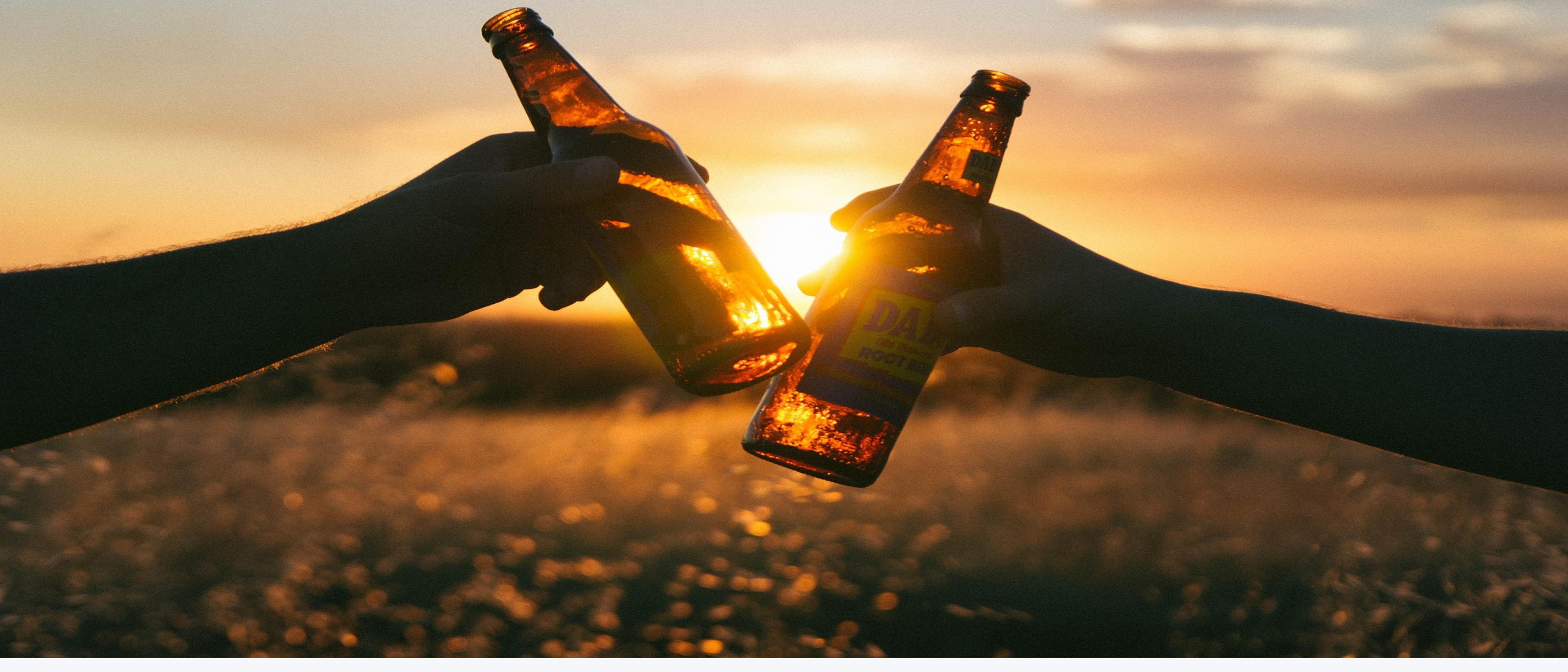


Time to Initial Value

build a robust,
fault-tolerant, scalable
microservice in minutes



Ship it!



Ship it!





Hype: Just deploy your code. The system handles logging and monitoring automatically!

Welcome to the **Cloud Functions** Dashboard

Thanks for checking us out. As a new user, you don't have any activity. Here is a screenshot of the Dashboard in action.

The screenshot shows the IBM Cloud Cloud Functions Dashboard. On the left, there's a sidebar with 'Getting Started', 'Actions', 'Triggers', **Monitor** (which is selected), and 'APIs'. The top navigation bar includes 'IBM Cloud', 'REGION US South', 'CLOUD FOUNDRY ORG eweiter@us.ibm.com', 'CLOUD FOUNDRY SPACE dev', 'Search platform resources', 'Catalog', 'Docs', 'Support', 'Manage', and a help icon.

Activity Summary: Invocation counts per action or rule.

Action	Invocation Count	Avg. Response Time
1234123412341234	13 activations	43.62ms
2missMachine	3	3.33ms on average
anActionwithaverylongname...	3	56ms on average
aVulFW	3	3.33ms on average
aVulFW_1234123412341234	3	3.67ms on average
backOffSlash	3	67ms on average
bubbleGum	3	2ms on average
bubbleGum_1234123412341...	3	3.67ms on average
buffalos	3	3.67ms on average
buffalos_1234123412341234	1	896ms

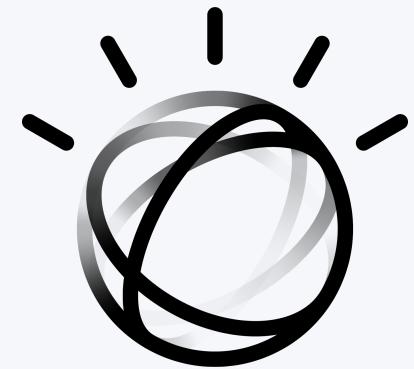
Activity Log: A list of log entries with timestamps and details.

Event	Timestamp	Details
changes	1/19/2018 7:44:13 PM	lccea711d64b4c1f8ea711d64bdc4f4a { "error": "error connecting to database asdf" }
2missMachine	1/19/2018 11:18:49 AM	509bc8abacd4470f9bc8abacd4870f6b { "message": "hello world" }
2missMachine	1/19/2018 10:20:46 AM	9303fb8782a1411183fb8782a1a11164 { "message": "hello world" }
2missMachine	1/19/2018 9:54:39 AM	7ed0adb82dd84ed90adb82dd8ceddd9

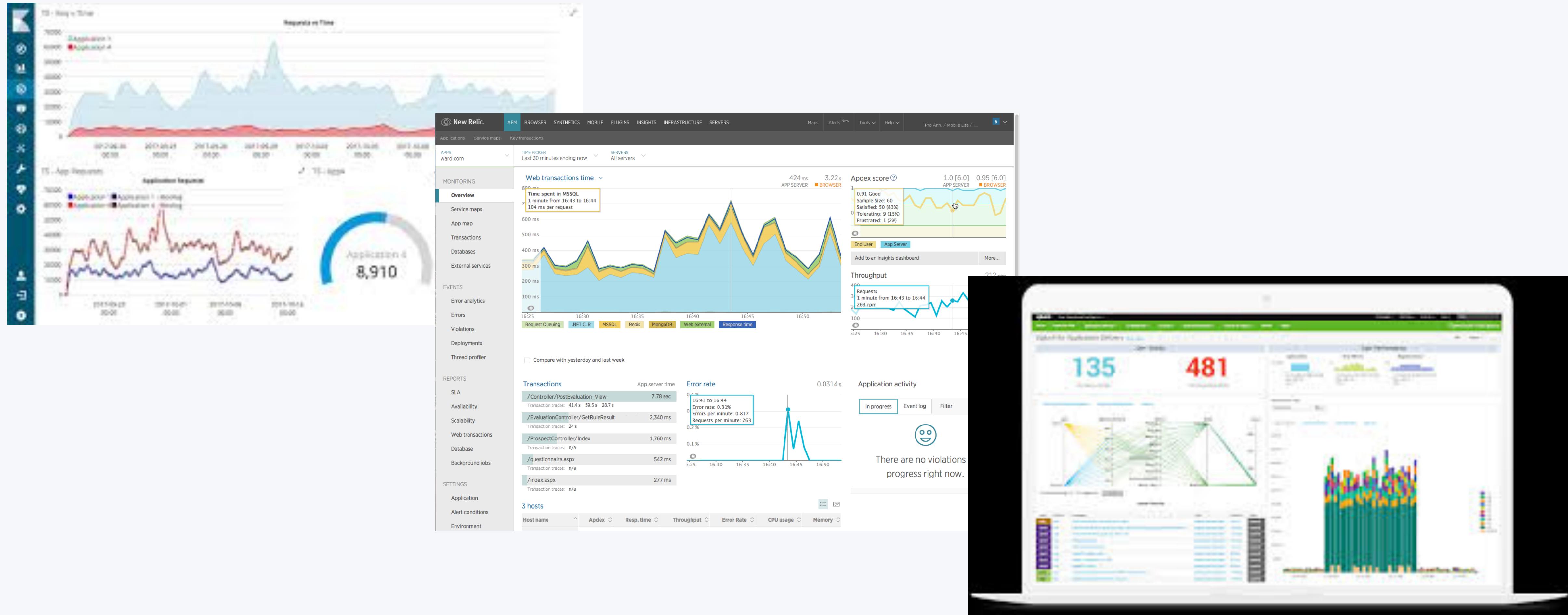
Activity Timeline: A timeline chart showing invocation counts over time.

Filtering Options:

- Time Frame: most recent 50
- Limit to: All Actions
- Exclude triggers from the views?

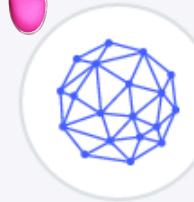


Reality: All Watson systems already have extensive management, logging, DevOps, and monitoring.
Serverless doesn't help integrate into existing production environment.





Hype: Just deploy your code. Specify an API declaratively. API Gateway does the rest!



Create API for Cloud Functions

API Info



API definition

Optionally import an Open API definition file that includes the required API configuration. The imported settings will replace any existing settings. You can export the current API definition to file. It is also possible to open the API definition in API Connect, IBM's premier API management platform. If API Connect is not present within your IBM Cloud space, we will first provision the service with the free Lite plan.

API definition ▾

API Basics *

First, specify a descriptive name for this API.

Next, accept the default domain for this API or select a custom domain. As a prerequisite, you will need to register a custom domain with IBM Cloud. This can be done from your organization settings page. For more information, please reference [the documentation](#).

Finally, specify a base path for this API.

API name *

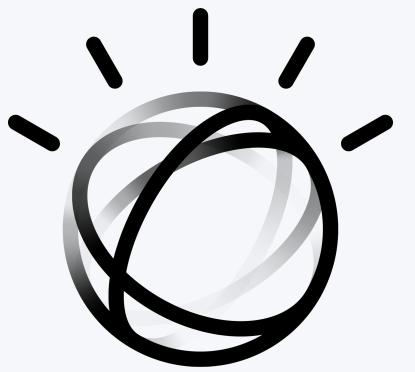
Descriptive name for API

Domain for API

Default domain

Base path for API *

Specify the base path for the API (e.g. /api)



Reality: Conventional web frameworks have tools, libraries, ecosystems tuned precisely for this problem. A serverless REST endpoint doesn't save that much code.

```
1 const express = require('express' 4.16.3 )
2 const app = express()
3
4 app.get('/', (req, res) => res.send('Hello World!'))
5
6 app.listen(3000, () => console.log('Example app listening on port 3000!'))
```

```
http.createServer(function (request, response) {
  // To Read a Cookie
  var cookies = parseCookies(request);

  // To Write a Cookie
  response.writeHead(200, {
    'Set-Cookie': 'mycookie=test',
    'Content-Type': 'text/plain'
  });
  response.end('Hello World\n');
}).listen(8124);

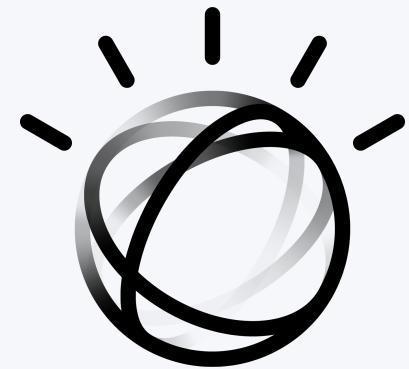
console.log('Server running at http://127.0.0.1:8124/');
```

```
function main() {
  return {
    headers: {
      'Set-Cookie': [
        'UserID=Jane; Max-Age=3600; Version=',
        'SessionID=asdfgh123456; Path = /'
      ],
      'Content-Type': 'text/html'
    },
    statusCode: 200,
    body: '<html><body><h3>hello</h3></body></html>' }
}
```



Hype: specify your service integrations declaratively.
Wrap services as functions, compose at will, and go!





Reality: Service integrations have many flavors and options.
Code and SDKs is still the easiest way to invoke Watson in realistic
use cases.

```
document.querySelector('#button').onclick = function () {  
  
    fetch('/api/speech-to-text/token')  
    .then(function(response) {  
        return response.text();  
    }).then(function (token) {  
  
        var stream = WatsonSpeech.SpeechToText.recognizeMicrophone({  
            token: token,  
            objectMode: true, // send objects instead of text  
            format: false // optional - performs basic formatting on the results such as capitals and periods  
        });  
  
        stream.on('data', function(data) {  
            console.log(data);  
        });  
  
        stream.on('error', function(err) {  
            console.log(err);  
        });  
  
        document.querySelector('#stop').onclick = stream.stop.bind(stream);  
  
    }).catch(function(error) {  
        console.log(error);  
    });  
};
```

A Mixed-Method Empirical Study of
Function-as-a-Service Software Development
in Industrial Practice

Philipp Leitner^{a,*}, Erik Wittern^b, Josef Spillner^c, Waldemar Hummer^b

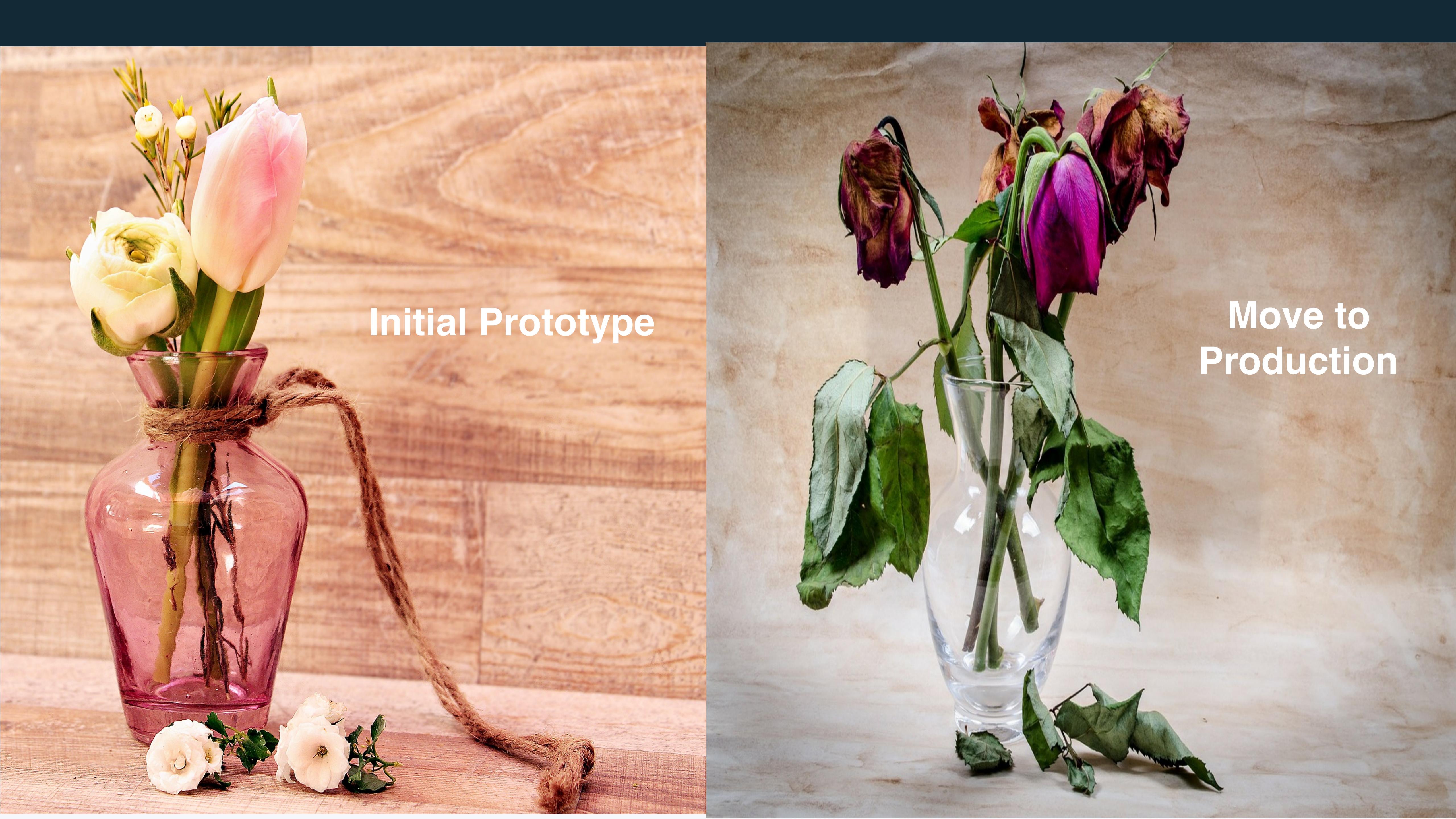
^aSoftware Engineering Division, Chalmers / University of Gothenburg, Sweden

^bIBM Research, Yorktown Heights, New York, USA

^cService Prototyping Lab, Zurich University of Applied Sciences, Switzerland

Which of the following do you consider significant challenges for using FaaS services?

1: Lack of tooling (e.g., testing, deployment)	51 / 55%
2: Integration testing	37 / 40%
3: Vendor lock-in	30 / 32%
4: Container start-up latency	27 / 29%
5: Managing state in functions	25 / 27%
6: Unit testing	17 / 18%
7: Little support for reusing functions	13 / 14%
8: Lack of documentation	12 / 13%
9: Finding/hiring developers familiar with FaaS	11 / 12%
10: Little support for composition of functions	11 / 12%
11: CPU or processing limitations	8 / 9%
12: Memory limitation	5 / 5%
13: Other	3 / 3%



Initial Prototype

Move to
Production



SaaS



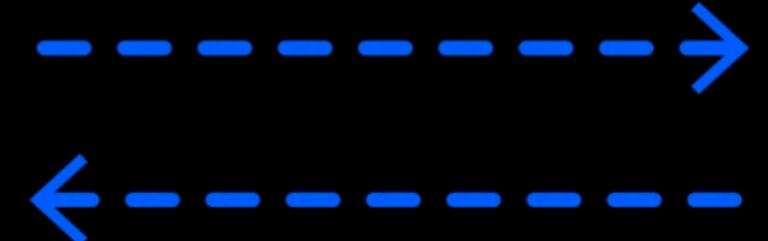
Software as a service (SaaS /sæs/)^[1] is a software licensing and delivery model in which software is licensed on a subscription basis and is centrally hosted.^{[2][3]} It is sometimes

Watson Assistant

Quickly build and deploy virtual assistants across a variety of channels

Get started free

→ View demo



Weather Bot with OpenWhisk



The Weather Bot provides forecasts for **U.S. cities** at this moment.

- asking for location

Hello there! I'm Watson weather bot. Let's talk about weather, shall we? To get started, tell me the name of your city. It'll help me send you weather forecasts for your area.

14:23:00

Type here

You need to add custom logic to a hosted SaaS chatbot.

Add node Add child node

| Google

Bienvenue

welcome

1 Response / 0 Context set

manger

#manger

0 Responses / 3 Context set / 3 Slots / Skip user input

#talk

1 Response / 0 Context set

location

#location

1 Response / 0 Context set / 1 Slot

prix

#combien

4 Responses / 0 Context set

météo

#meteo

0 Responses / 2 Context set / 1 Slot / Skip user input

Tout le reste

anything_else

0 Responses / 0 Context set / Skip user input

Tout le reste

Customize

If bot recognizes:

anything_else - +

Then respond with:

```
1 {
2   "output": {},
3   "actions": [
4     {
5       "name": "/vincent.perrin@fr.ibm.com_dev/vperrin/WDS_REST",
6       "type": "server",
7       "parameters": {
8         "url": "",
9         "input": "<? input.text ?>",
10        "authorization": "Basic MzI1YjA1YTktMTY4Yi00NzQwLw",
11        "collection_id": "",
12        "environment_id": "",
13      },
14      "credentials": "$mycreds",
```

And finally

Skip user input ▼ and evaluate child nodes -

SaaS tooling with
serverless
extension points.

Serverless:
perfect match
for extension points in
hosted solutions.

Write functions.
Not plumbing.

Hype or Reality?

Write functions.
Not plumbing.

Hype or Reality?

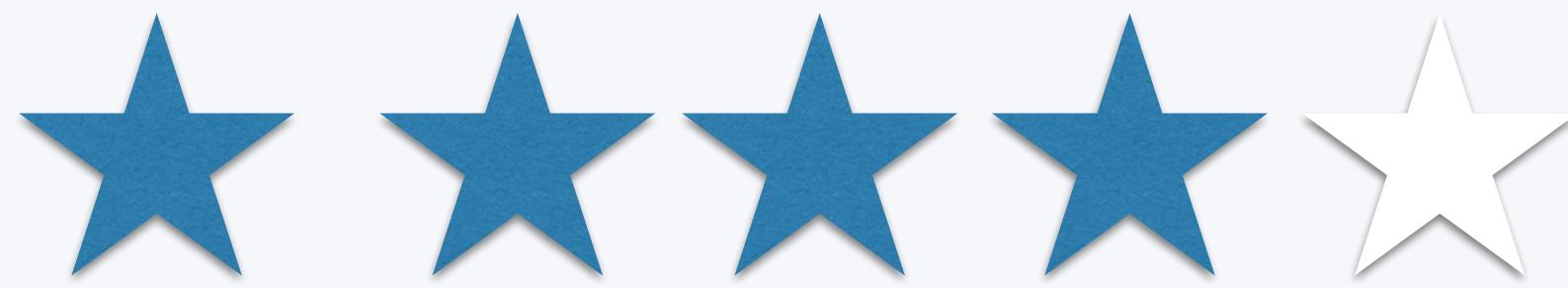
Initial Prototype



Write functions.
Not plumbing.

Hype or Reality?

Initial Prototype



Integration with Production
Systems



Write functions.
Not plumbing.

Hype or Reality?

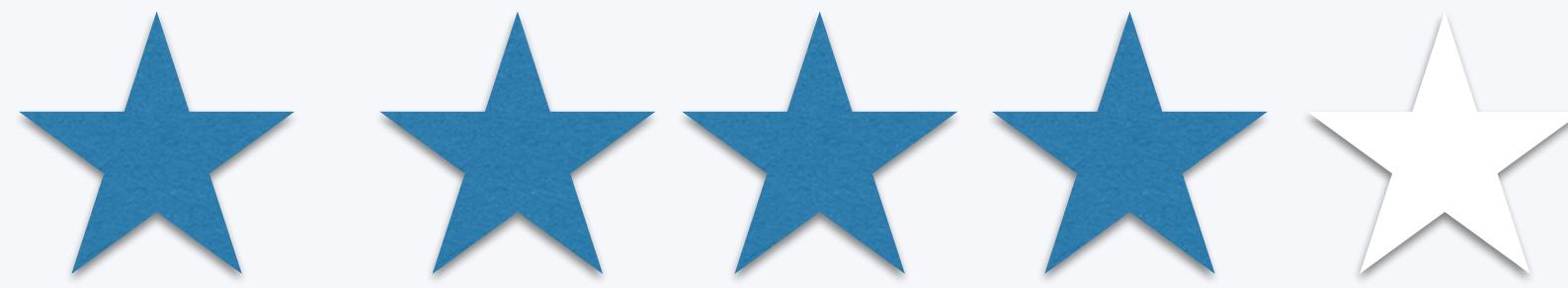
Initial Prototype



Integration with Production Systems

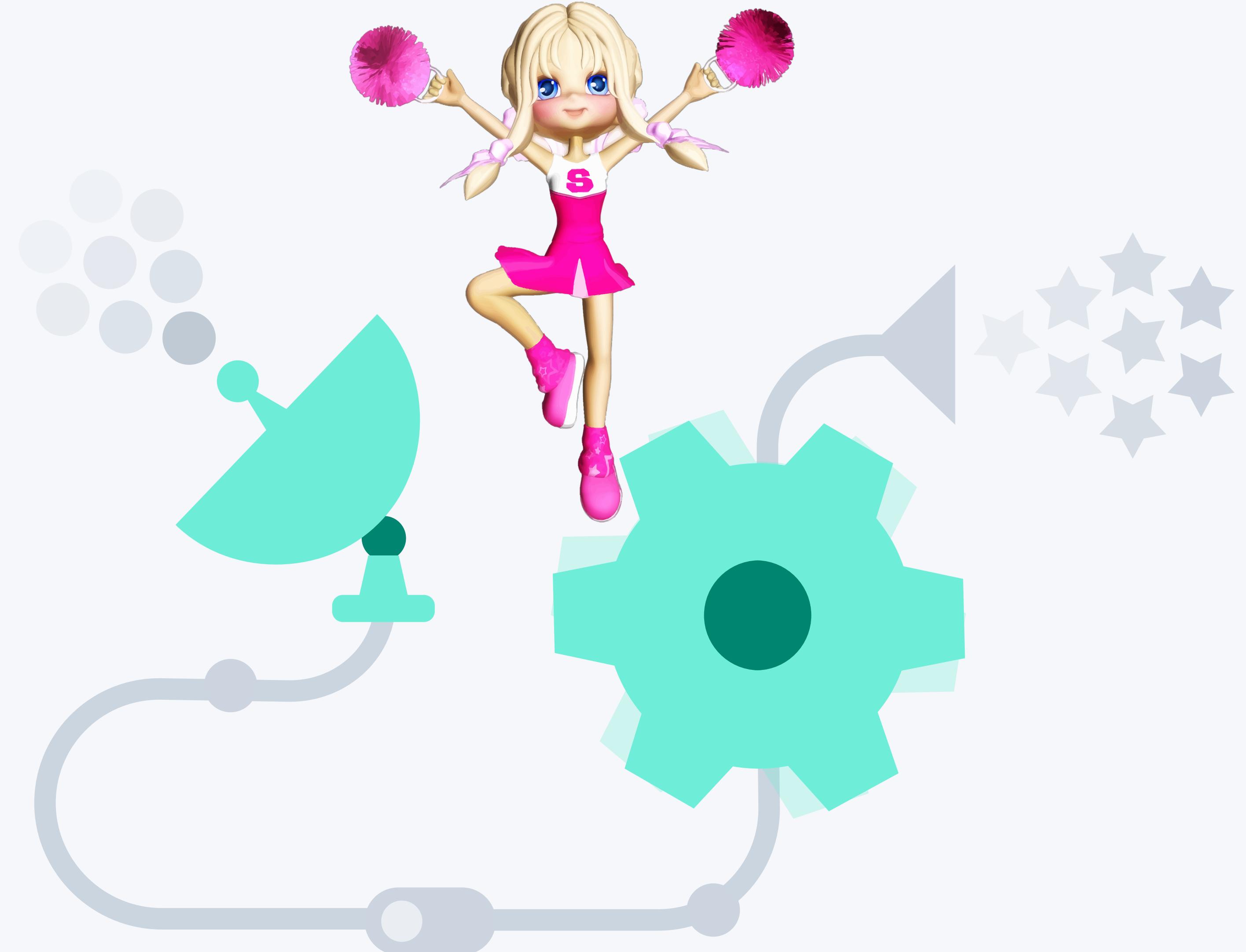


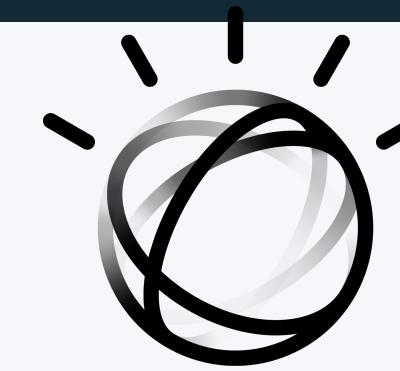
SaaS Extension Points



Event-driven Programming

Functions run in response to events

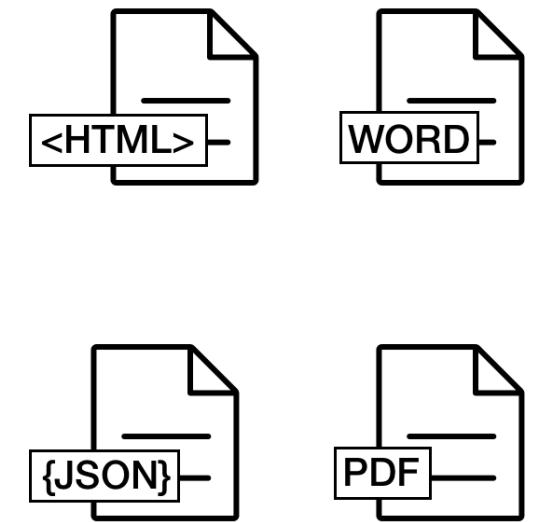




Event-driven programming in Watson AI

| Data

Private data



| Ingestion

Convert and enrich by leveraging Watson APIs to add NLP meta data to your content, making it easier to explore and discover insights

Clean and normalize through an automated processing of NLP results, improving data quality

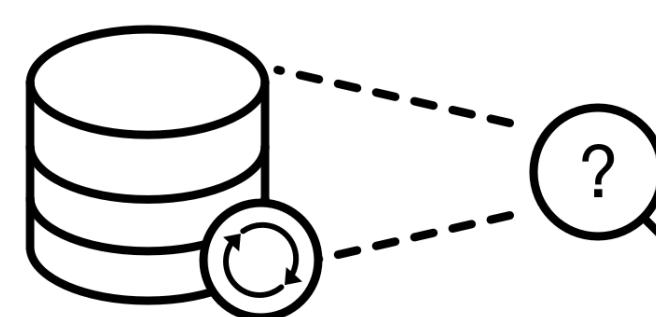
| Storage

Normalized data is **indexed** into a **collection** as part of your **environment** in the cloud



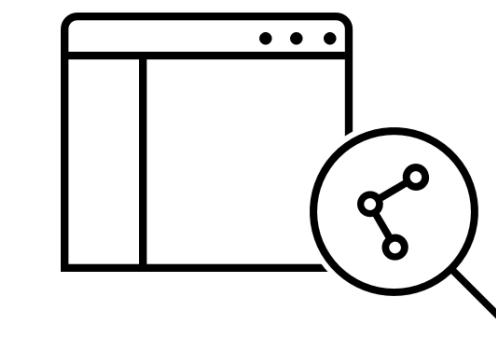
| Query

Understand data faster, create better hypothesis and deliver better outcomes



| Output

Actionable **insights** into your app

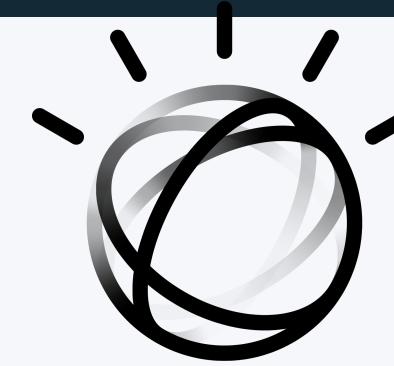


WHEN a document is uploaded, inject the document into the knowledge store

WHEN the logs grow to certain size gather the logs, perform ETL, retrain a new model

WHEN a new model is available evaluate the accuracy

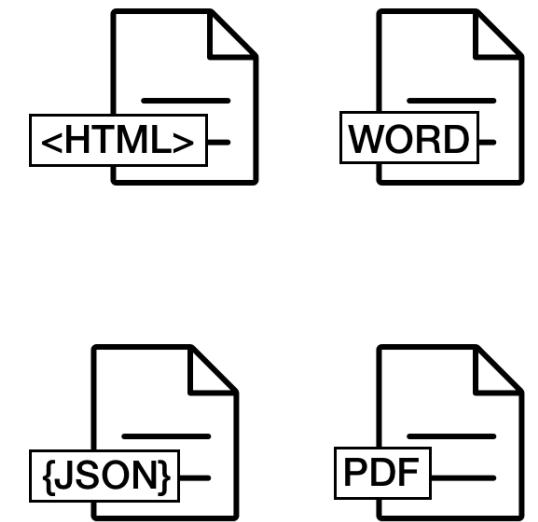
EVERY 15 minutes poll a stream for new data to inject



Event-driven programming in Watson AI

| Data

Private data



| Ingestion

Convert and enrich by leveraging Watson APIs to add NLP meta data to your content, making it easier to explore and discover insights

Clean and normalize through an automated processing of NLP results, improving data quality

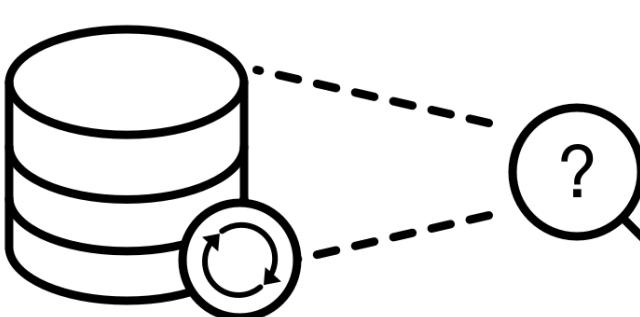
| Storage

Normalized data is **indexed** into a **collection** as part of your **environment** in the cloud



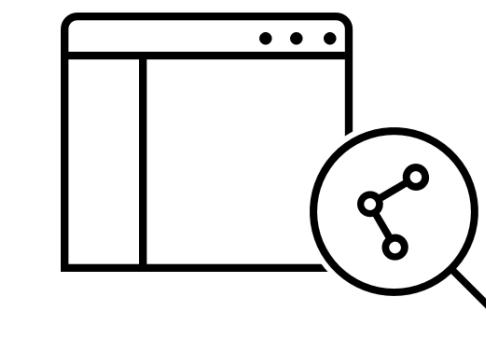
| Query

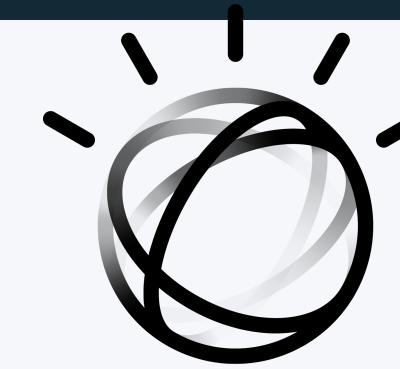
Understand data faster, create better hypothesis and deliver better outcomes



| Output

Actionable insights into your app

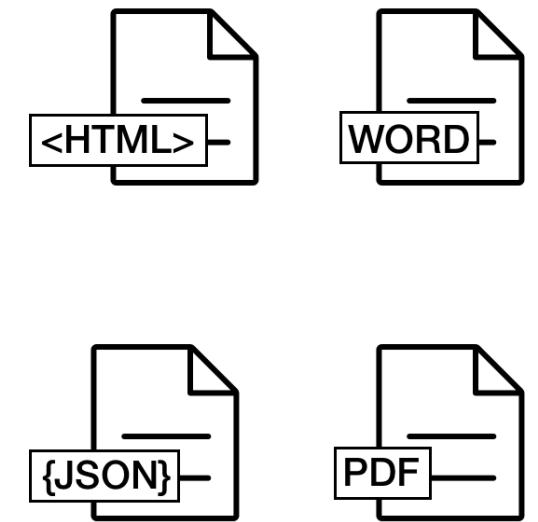




Event-driven programming in Watson AI

| Data

Private data



| Ingestion

Convert and enrich by leveraging Watson APIs to add NLP meta data to your content, making it easier to explore and discover insights

Clean and normalize through an automated processing of NLP results, improving data quality

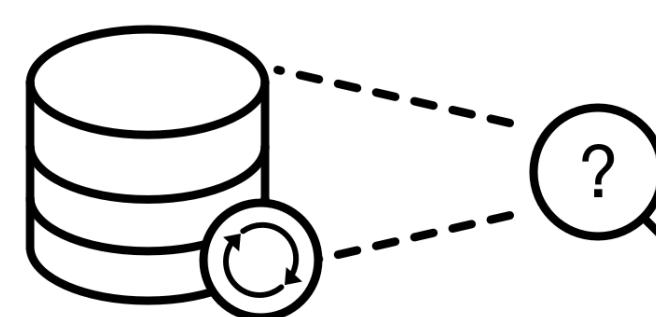
| Storage

Normalized data is **indexed** into a **collection** as part of your **environment** in the cloud



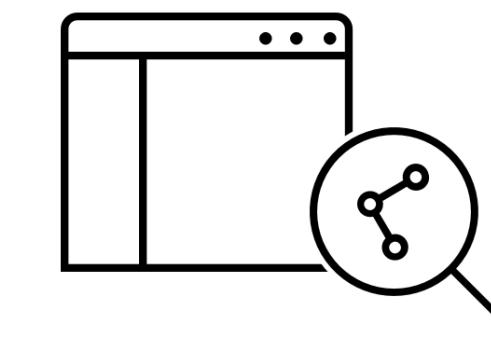
| Query

Understand data faster, create better hypothesis and deliver better outcomes

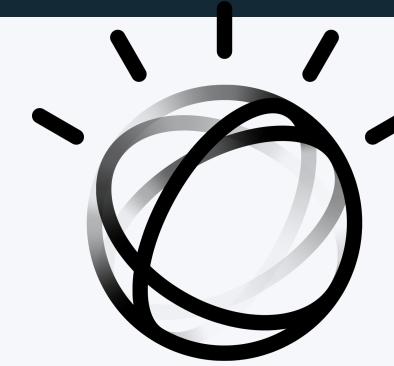


| Output

Actionable **insights** into your app



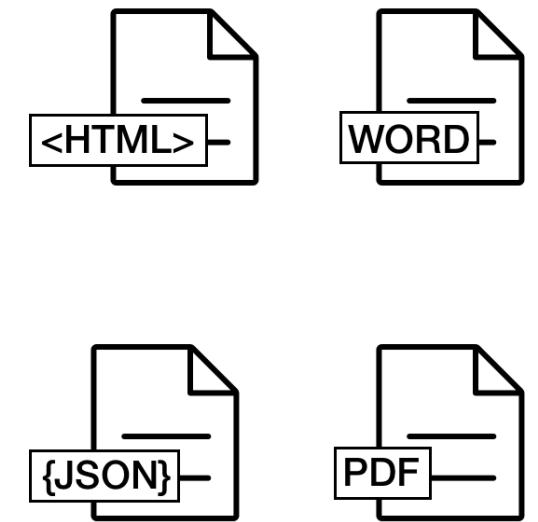
WHEN a document is uploaded, inject the document into the knowledge store



Event-driven programming in Watson AI

| Data

Private data



| Ingestion

Convert and enrich by leveraging Watson APIs to add NLP meta data to your content, making it easier to explore and discover insights

Clean and normalize through an automated processing of NLP results, improving data quality

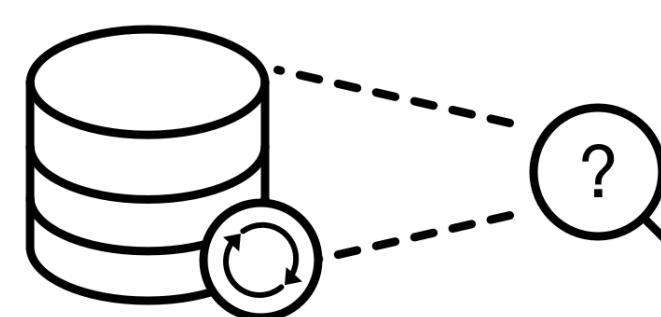
| Storage

Normalized data is **indexed** into a **collection** as part of your **environment** in the cloud



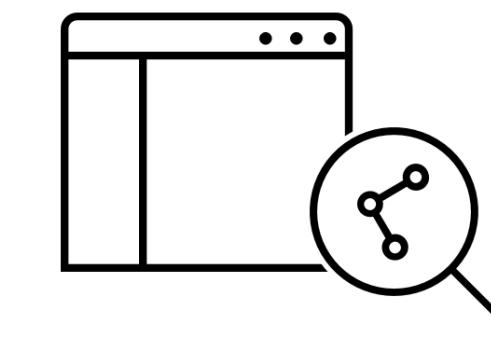
| Query

Understand data faster, create better hypothesis and deliver better outcomes



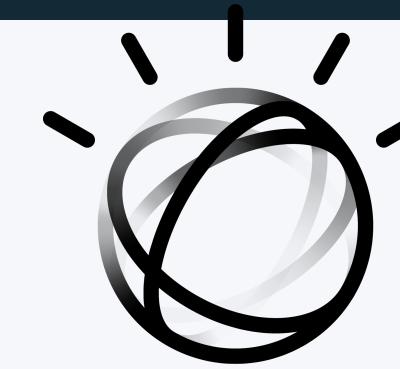
| Output

Actionable **insights** into your app



WHEN a document is uploaded, inject the document into the knowledge store

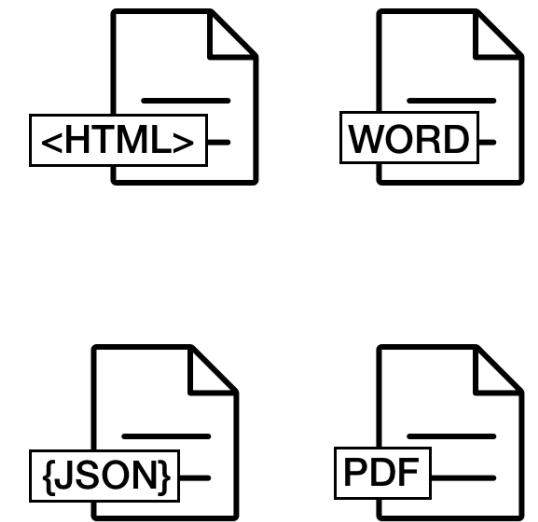
WHEN the logs grow to certain size gather the logs, perform ETL, retrain a new model



Event-driven programming in Watson AI

| Data

Private data



| Ingestion

Convert and enrich by leveraging Watson APIs to add NLP meta data to your content, making it easier to explore and discover insights

Clean and normalize through an automated processing of NLP results, improving data quality

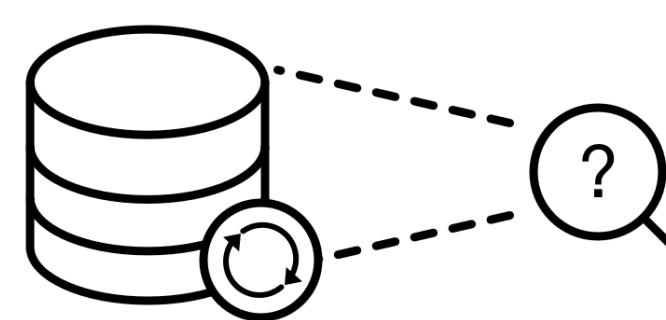
| Storage

Normalized data is **indexed** into a **collection** as part of your **environment** in the cloud



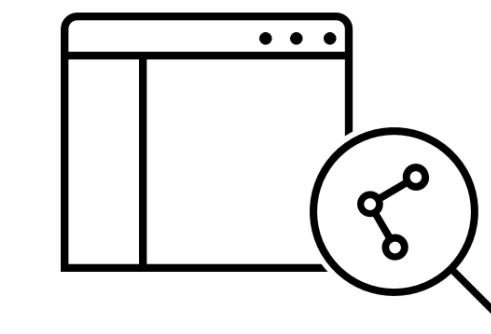
| Query

Understand data faster, create better hypothesis and deliver better outcomes



| Output

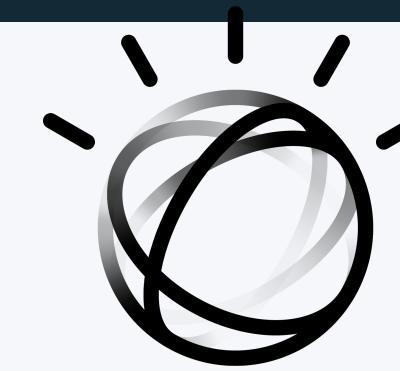
Actionable **insights** into your app



WHEN a document is uploaded, inject the document into the knowledge store

WHEN the logs grow to certain size gather the logs, perform ETL, retrain a new model

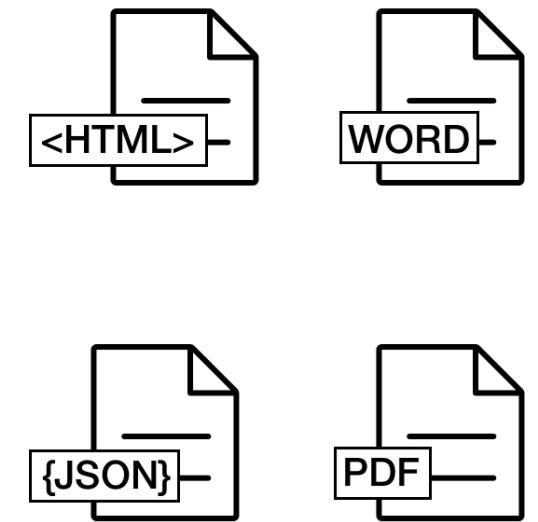
WHEN a new model is available evaluate the accuracy



Event-driven programming in Watson AI

| Data

Private data



| Ingestion

Convert and enrich by leveraging Watson APIs to add NLP meta data to your content, making it easier to explore and discover insights

Clean and normalize through an automated processing of NLP results, improving data quality

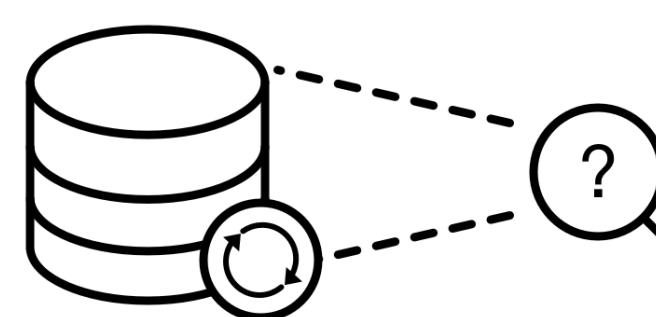
| Storage

Normalized data is **indexed** into a **collection** as part of your **environment** in the cloud



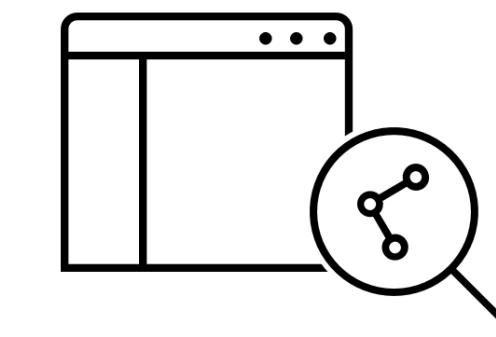
| Query

Understand data faster, create better hypothesis and deliver better outcomes



| Output

Actionable **insights** into your app



WHEN a document is uploaded, inject the document into the knowledge store

WHEN the logs grow to certain size gather the logs, perform ETL, retrain a new model

WHEN a new model is available evaluate the accuracy

EVERY 15 minutes poll a stream for new data to inject

Discovery

Unlock hidden value in data to find answers, monitor trends and surface patterns, with the world's most advanced cloud-native insight engine.

 **Get Started**

[API Reference](#)

[Documentation](#)

[Fork on GitHub](#)

Start for free in IBM Cloud

What company are you interested in? 

Quickly find insights in the Watson Discovery News data collection of recent news articles. Easily explore a company's:

- Top stories over the last two months
- Top entities (people, topics, companies) mentioned in those articles
- Trend of public sentiment in news
- Anomalous periods of high press coverage
- Trend of most commonly paired entities (co-mentions)

Watson Discovery also lets you do the same analysis with your own data. Learn more [here](#).

Event-driven Programming

Functions run in response to events

Hype or Reality?

Event-driven Programming

Functions run in response to events

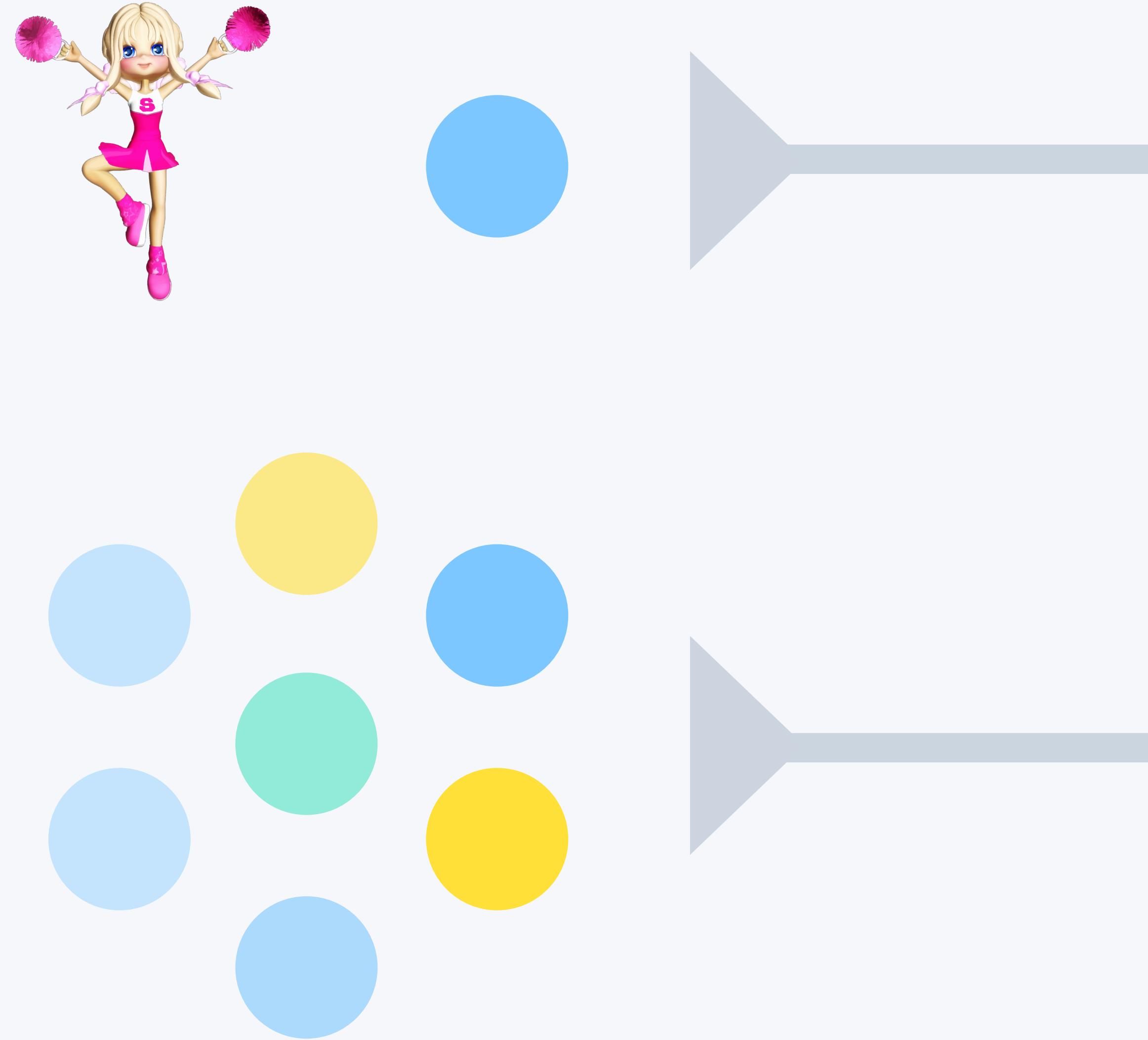
Hype or Reality?

- + not latency sensitive
- + highly elastic



Scale instantly. No provisioning.

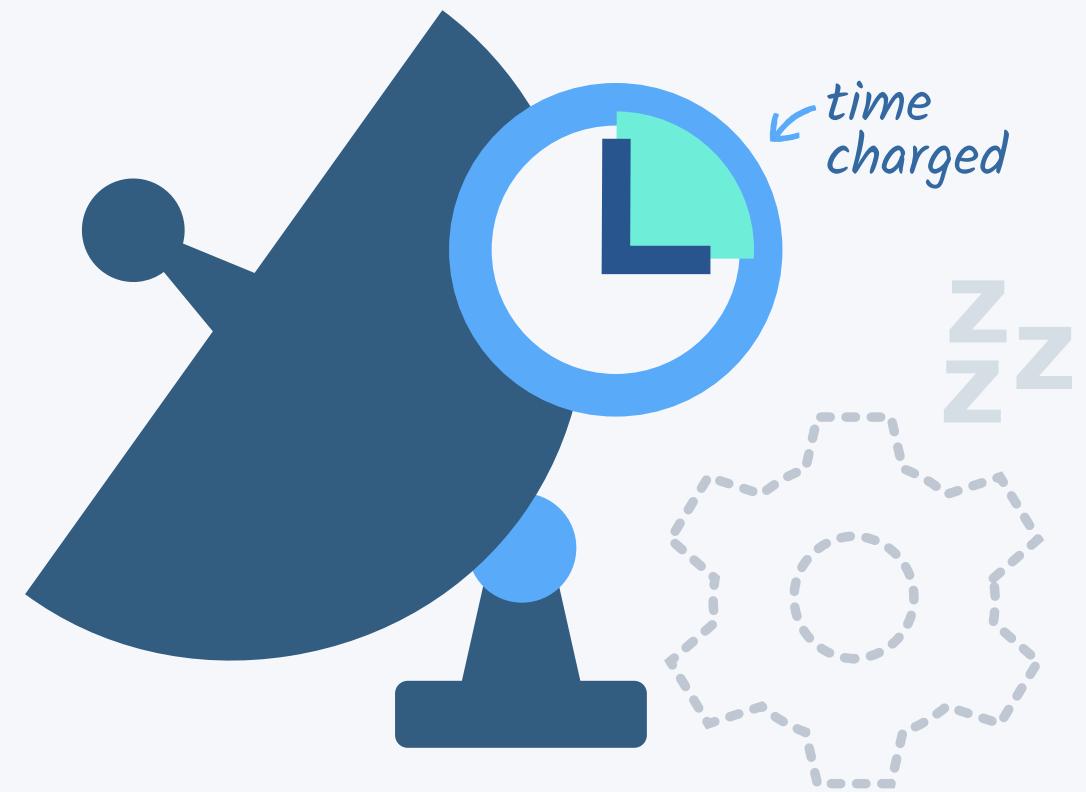
Enlist more resources
automatically based on
offered load





Fine-grain pricing.

Pay only for the exact time your actions run,
metered on the order of milliseconds



A Mixed-Method Empirical Study of Function-as-a-Service Software Development in Industrial Practice

Philipp Leitner^{a,*}, Erik Wittern^b, Josef Spillner^c, Waldemar Hummer^b

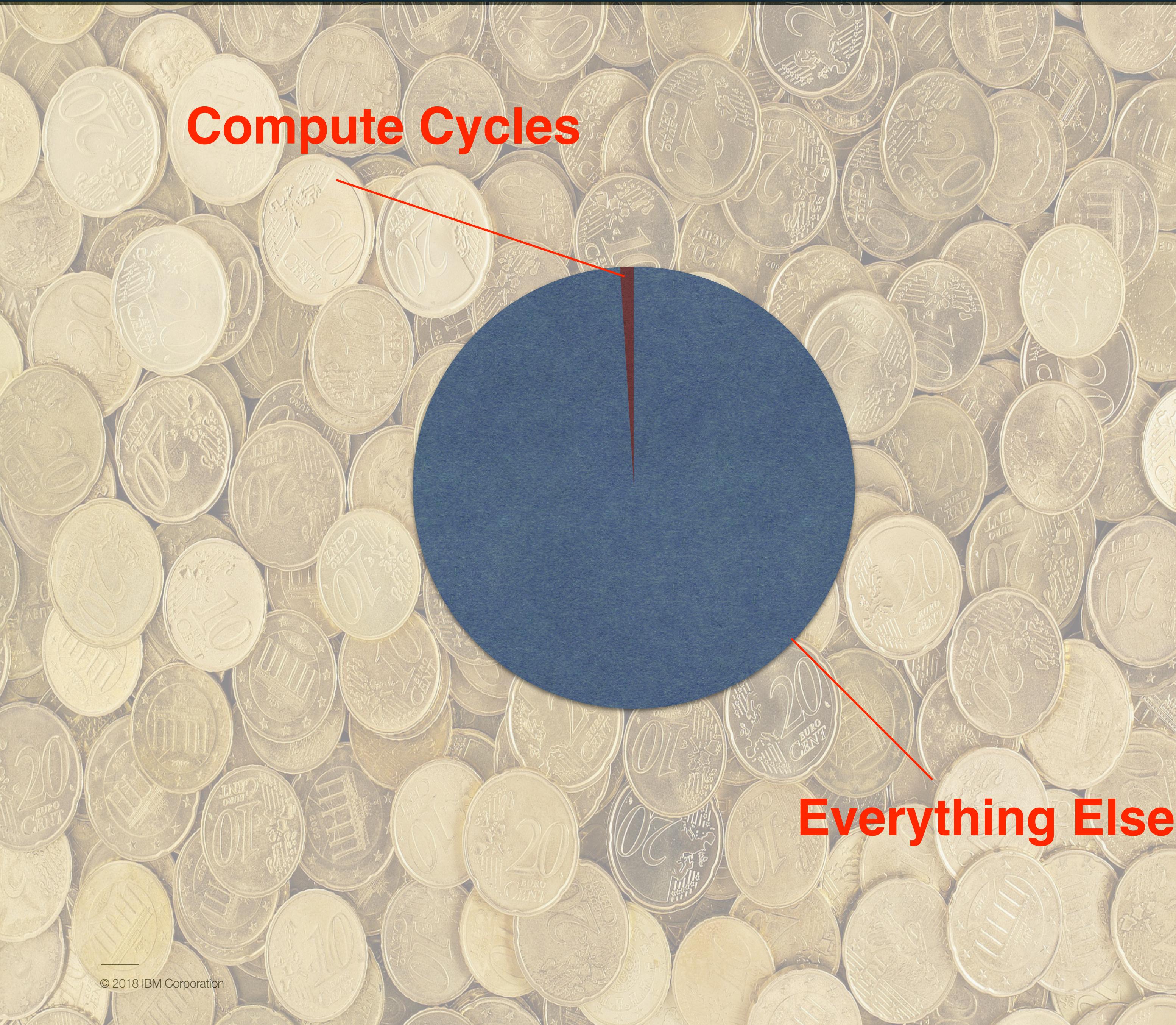
^aSoftware Engineering Division, Chalmers / University of Gothenburg, Sweden

^bIBM Research, Yorktown Heights, New York, USA

^cService Prototyping Lab, Zurich University of Applied Sciences, Switzerland

Do you think that using FaaS at the moment is cheap in terms of cloud hosting costs?

- | | |
|---------------------------------------------------------|-----------------|
| 1: Total costs of FaaS are lower than its alternatives | 65 / 71% |
| 2: Costs do not matter to us at this point | 20 / 22% |
| 3: Total costs of FaaS are higher than its alternatives | 3 / 3% |
| 4: Other | 3 / 3% |



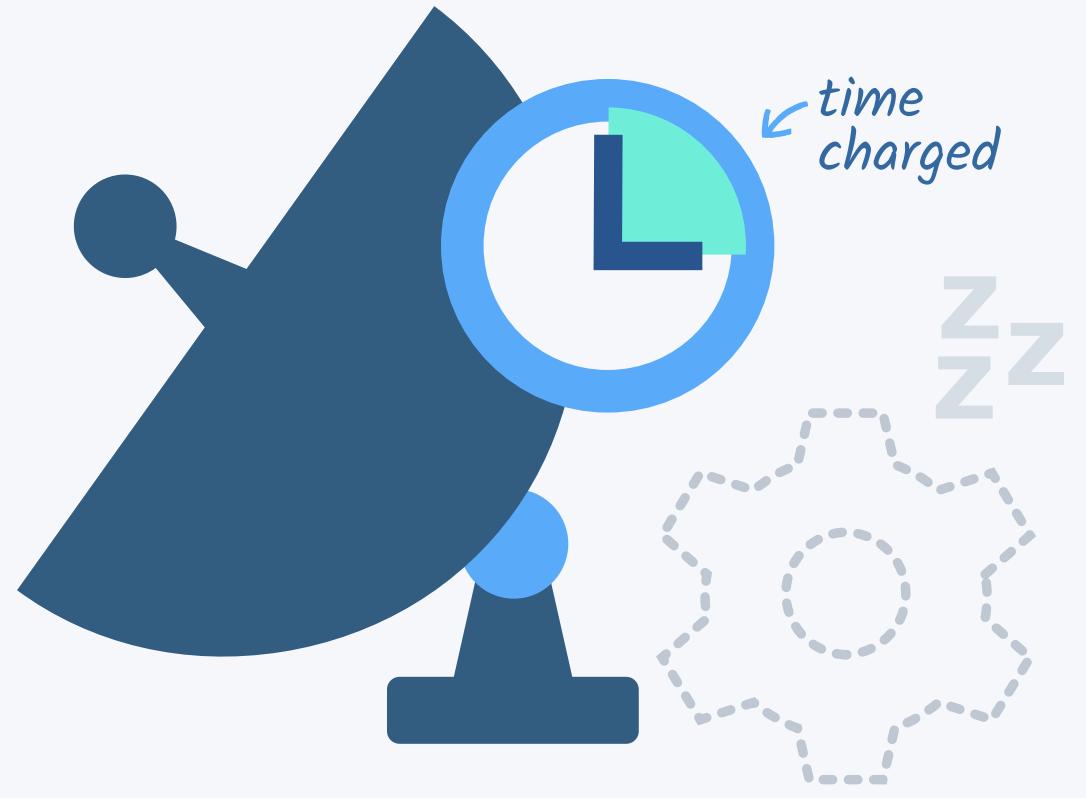
Total Cost of Solution (unscientific estimates)

Development
Operations
Support

Storage
Networking

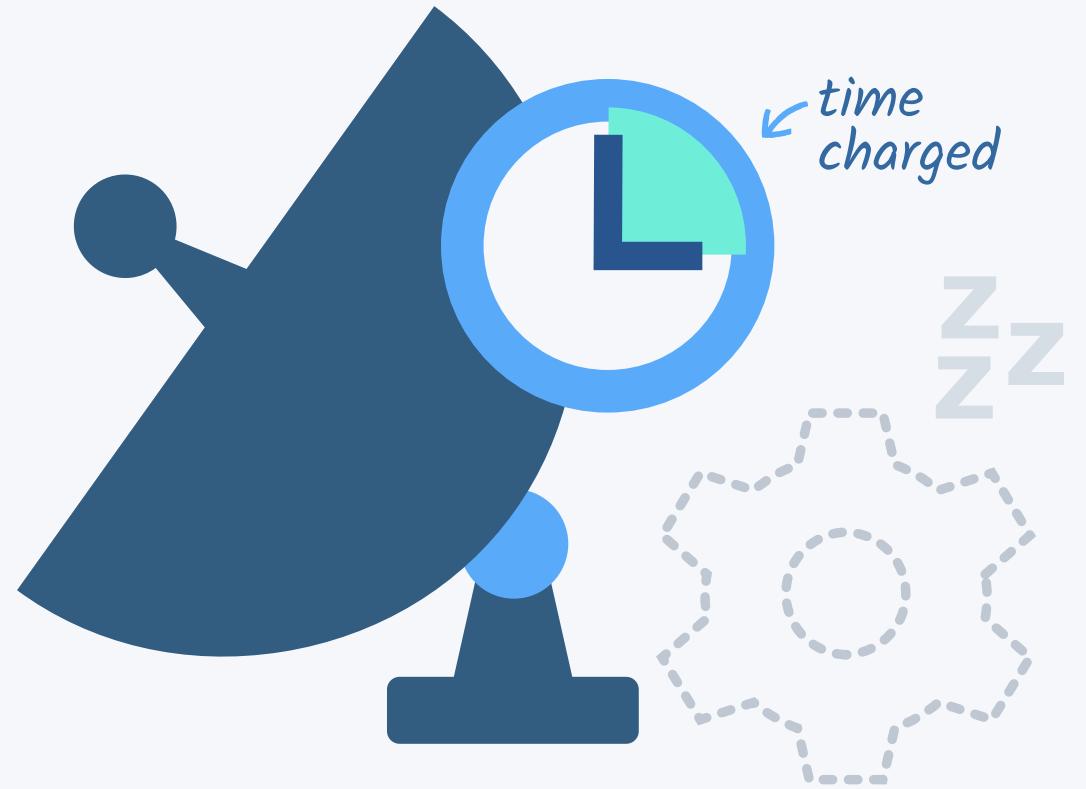
Compute

Scale instantly. No provisioning.
Fine-grain pricing.



Hype or Reality?

Scale instantly. No provisioning.
Fine-grain pricing.

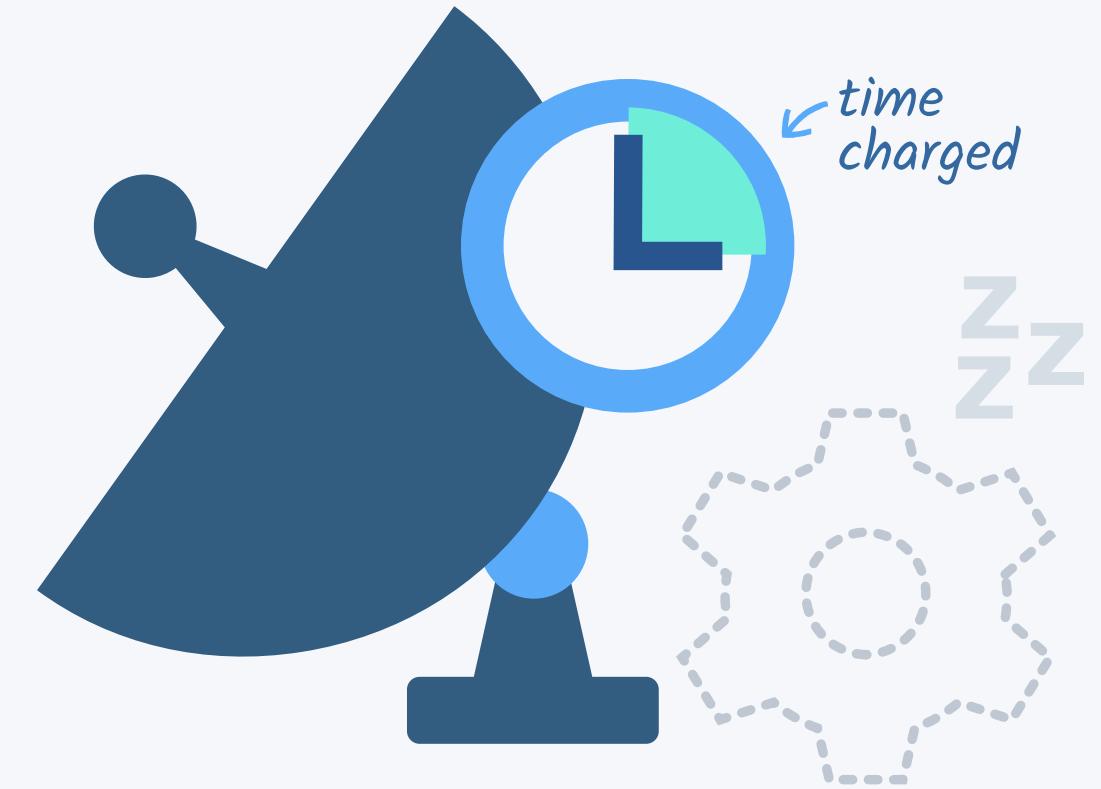


Hype or Reality?

Startup or hobby-ist



Scale instantly. No provisioning.
Fine-grain pricing.



Hype or Reality?

Startup or hobby-ist

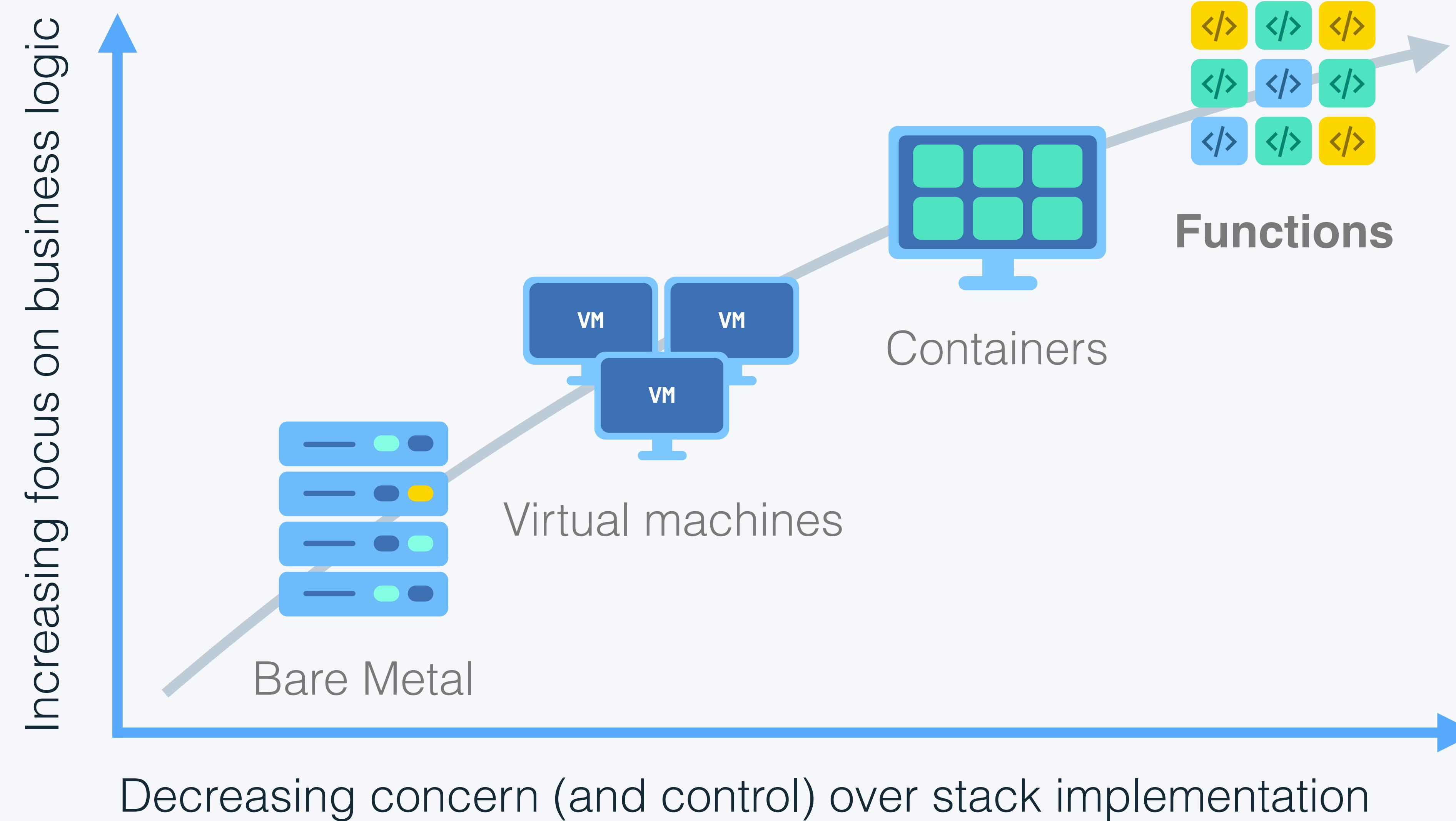


Large Enterprise

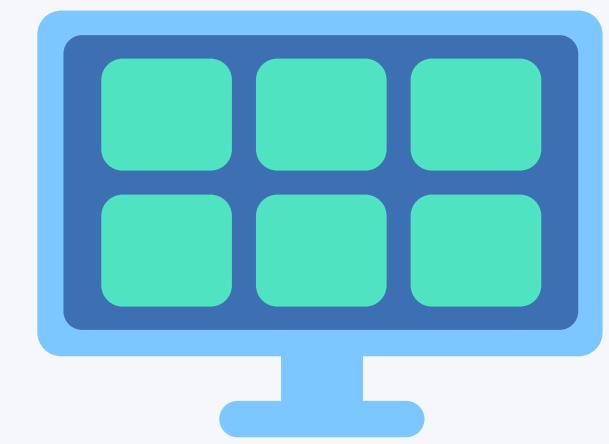


often irrelevant







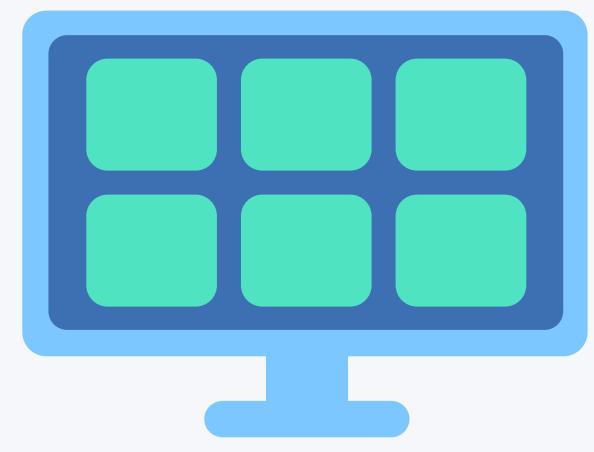


Containers

- + Fine-Grain Metering
- + Faster Autoscaling
- + Event-driven Programming



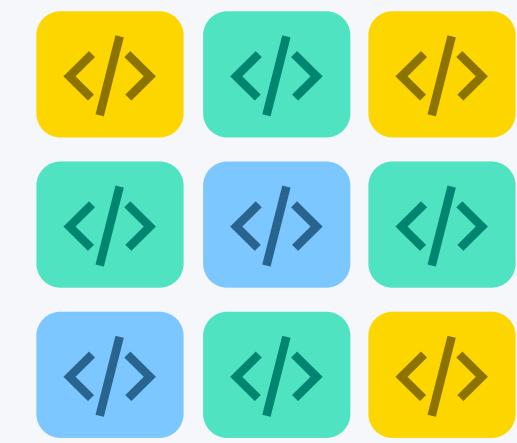
Functions



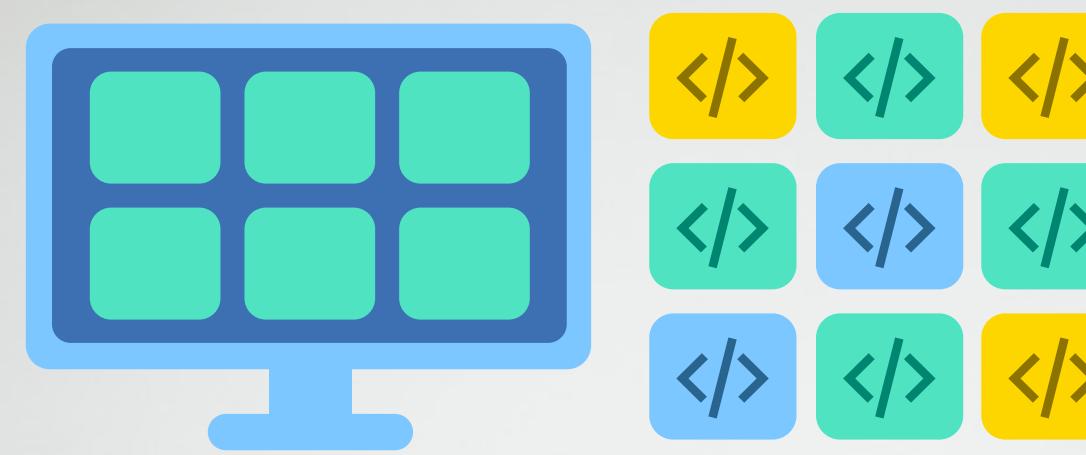
Containers



**+ Tools
+ Control and Flexibility
+ De Facto Standards**



Functions



Containers **Functions**

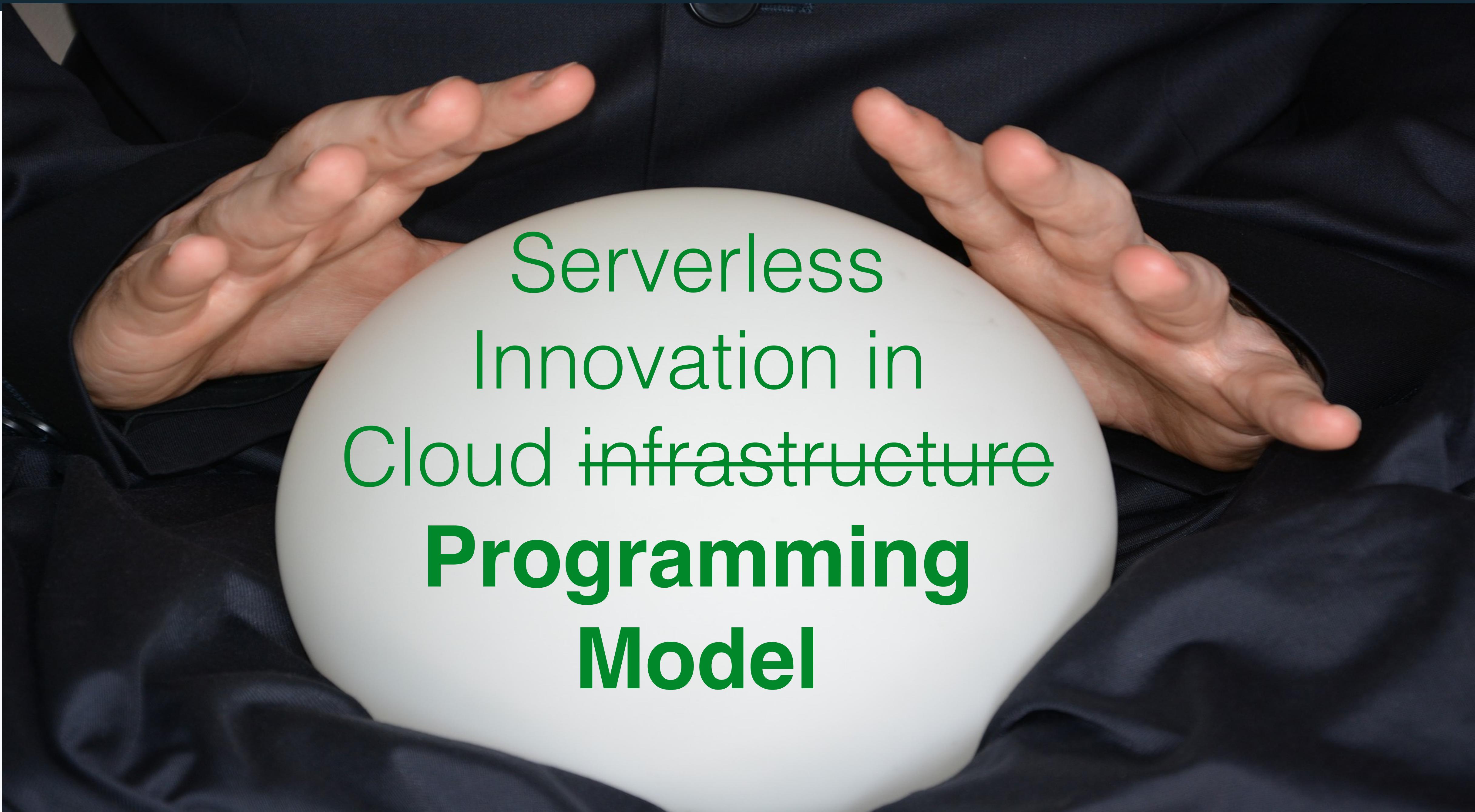
Convergence



Infrastructure Convergence

The End of the Road for Serverless?





Serverless
Innovation in
Cloud ~~infrastructure~~
Programming
Model

Composing IBM Cloud Functions



<https://ibm.biz/serverless-research>

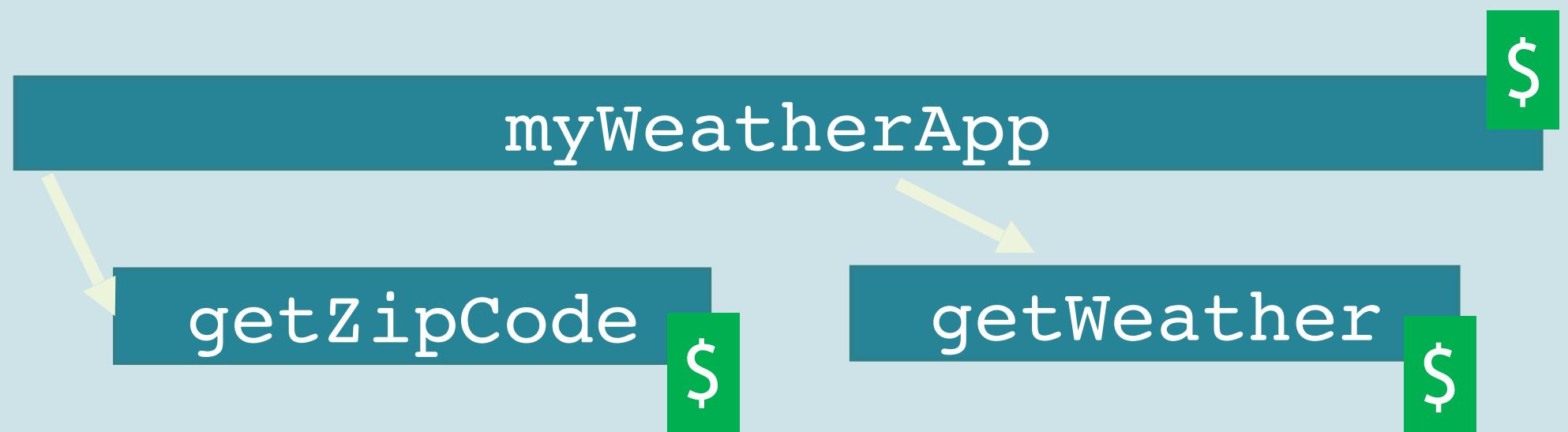
Kerry Chang, Olivier Tardieu
IBM Research

Apps are Compositions of Functions

```
try {  
  let zipCode = getZipCode(location);  
  return getWeather(zipCode);  
} catch(err) {  
  return { message: `Unable to retrieve weather info: ${err}` };  
}
```

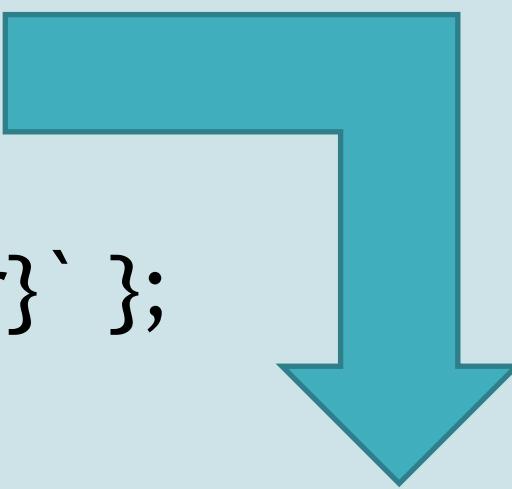
Does not work as a serverless app

- Time limit, double billing



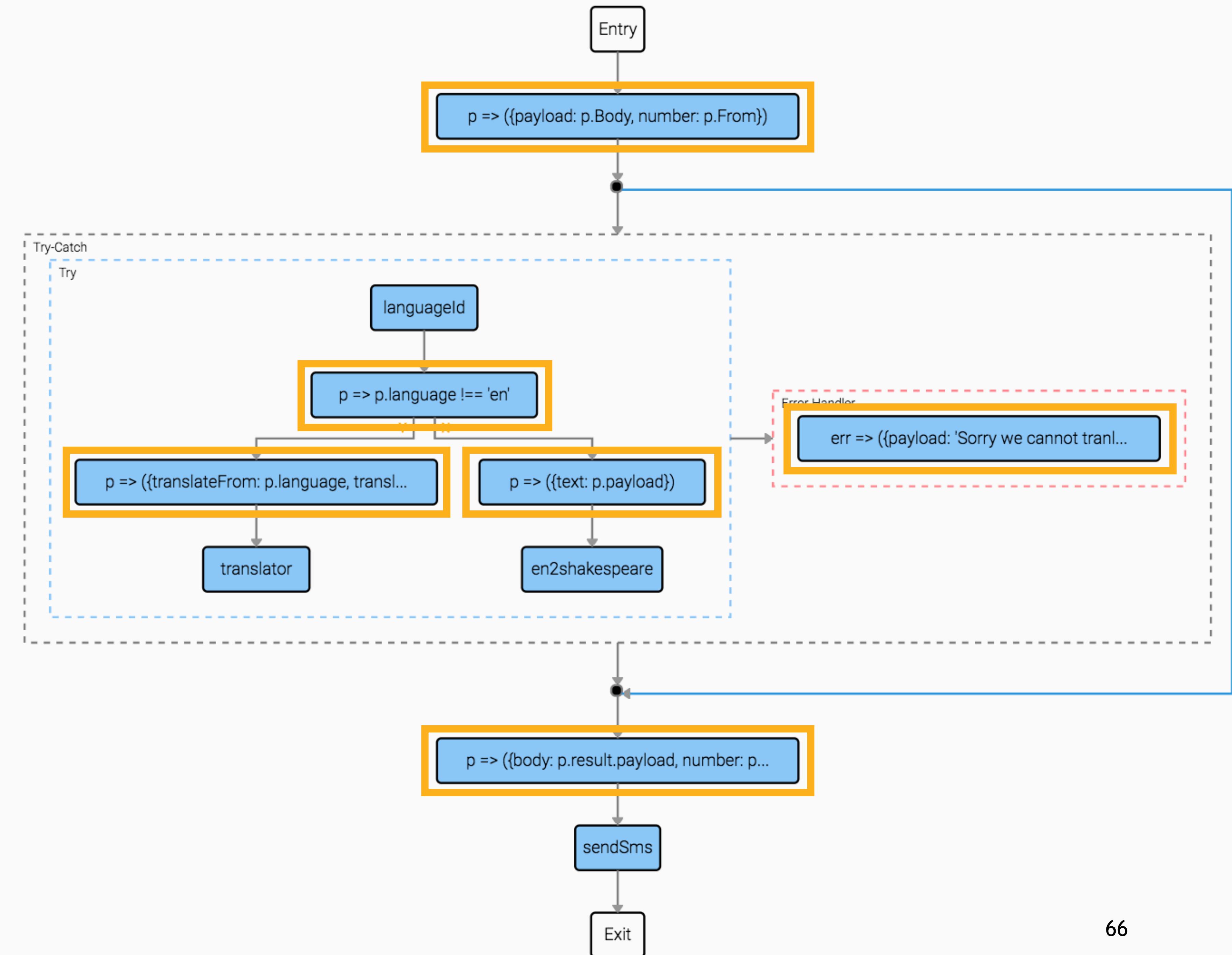
Serverless App using Composer

```
try {  
  let zipCode = getZipCode(location);  
  return getWeather(zipCode);  
} catch(err) {  
  return { message: `Unable to retrieve weather info: ${err}` };  
}
```

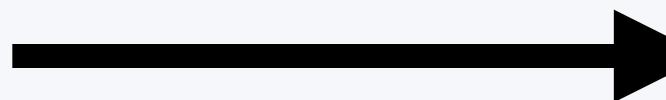


```
composer.try(  
  composer.sequence(  
    'getZipCode',           // cloud function  
    'getWeather'),         // cloud function  
    (err) => ({ message: `Unable to retrieve weather info: ${err}` })  
);
```

- if/else
- try/catch
- data forwarding
- inline functions for simple operations



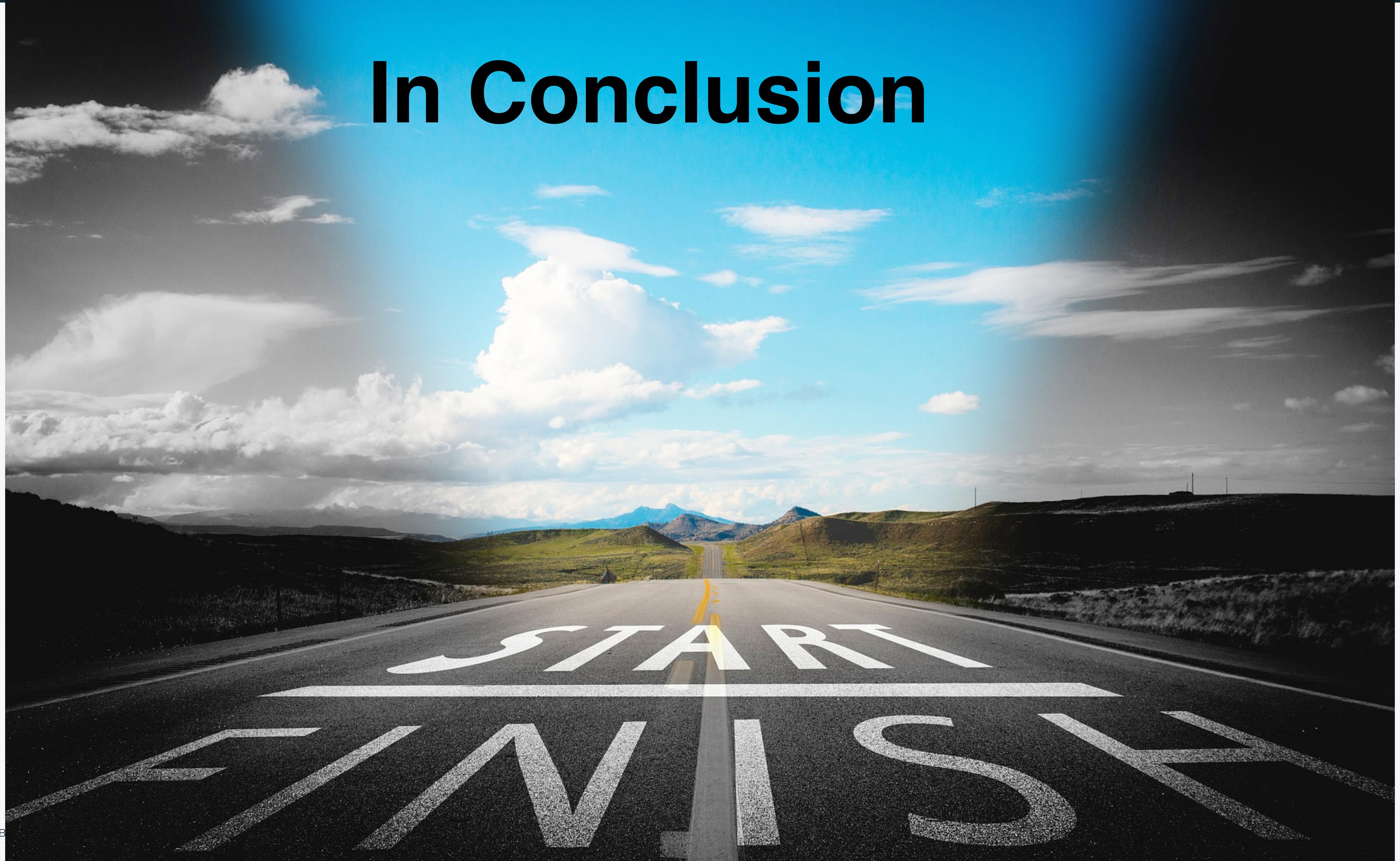
Event-driven Programming



Serverless Event-Driven Workflow



In Conclusion



In Conclusion

- 1. Integration: You don't own main ()**
- 2. Embrace containers - infrastructure convergence is coming**
- 3. Innovate on event-driven programming model**

Q & A