

# Serverless

## for data and AI

Avner Braverman

Binaris CEO

**The future is Serverless**

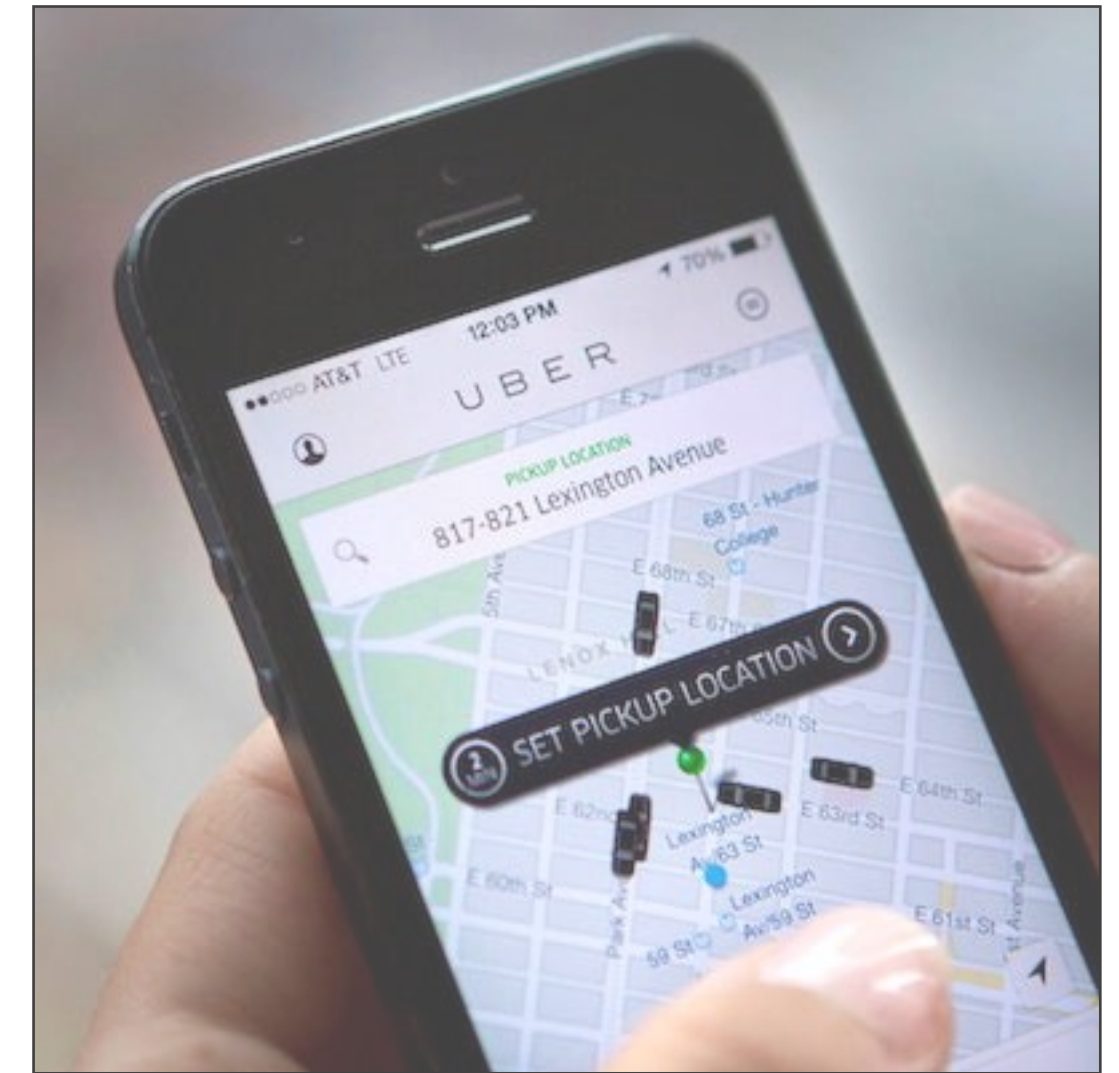
# On-Prem



# Cloud

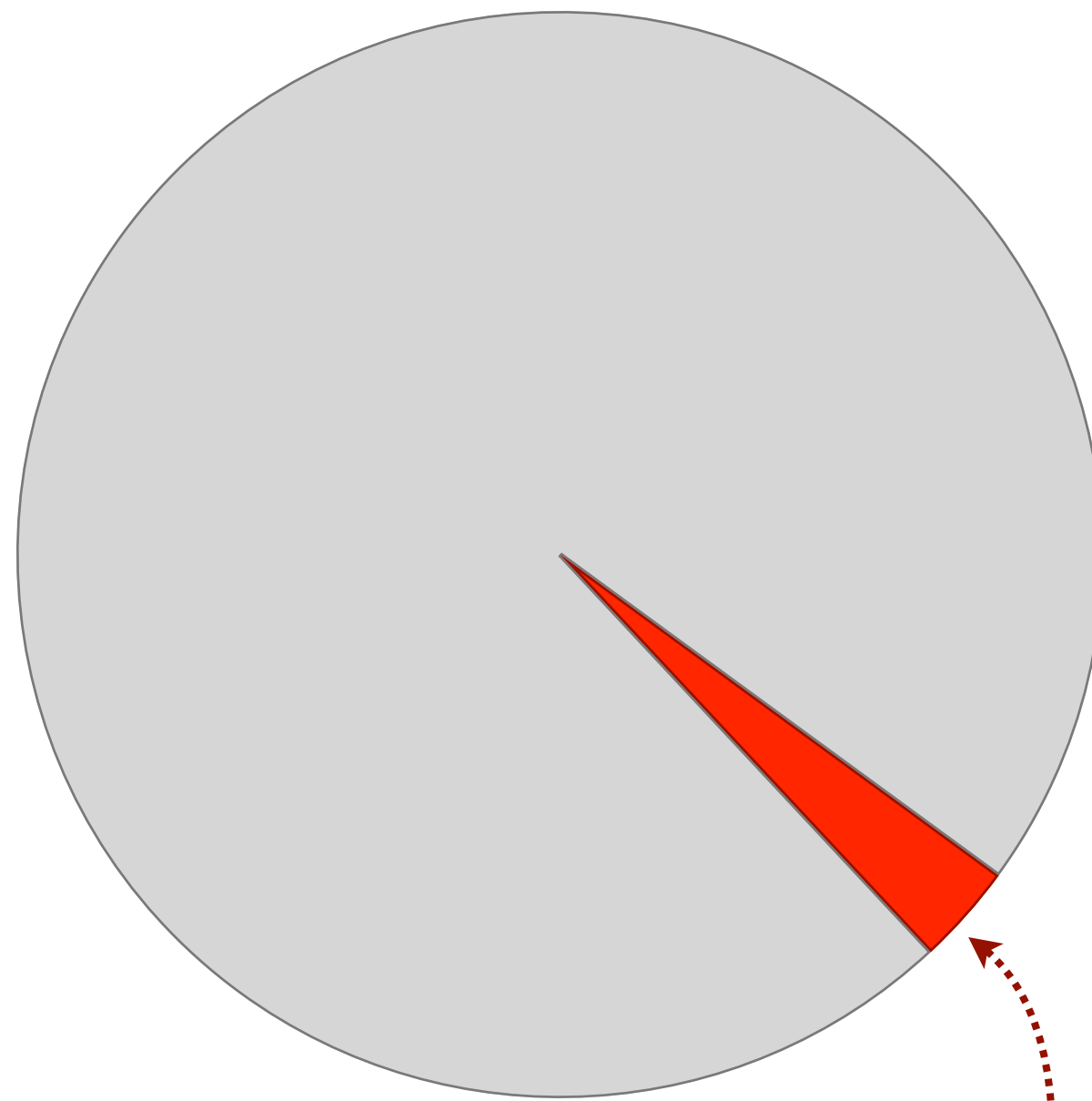


# Serverless



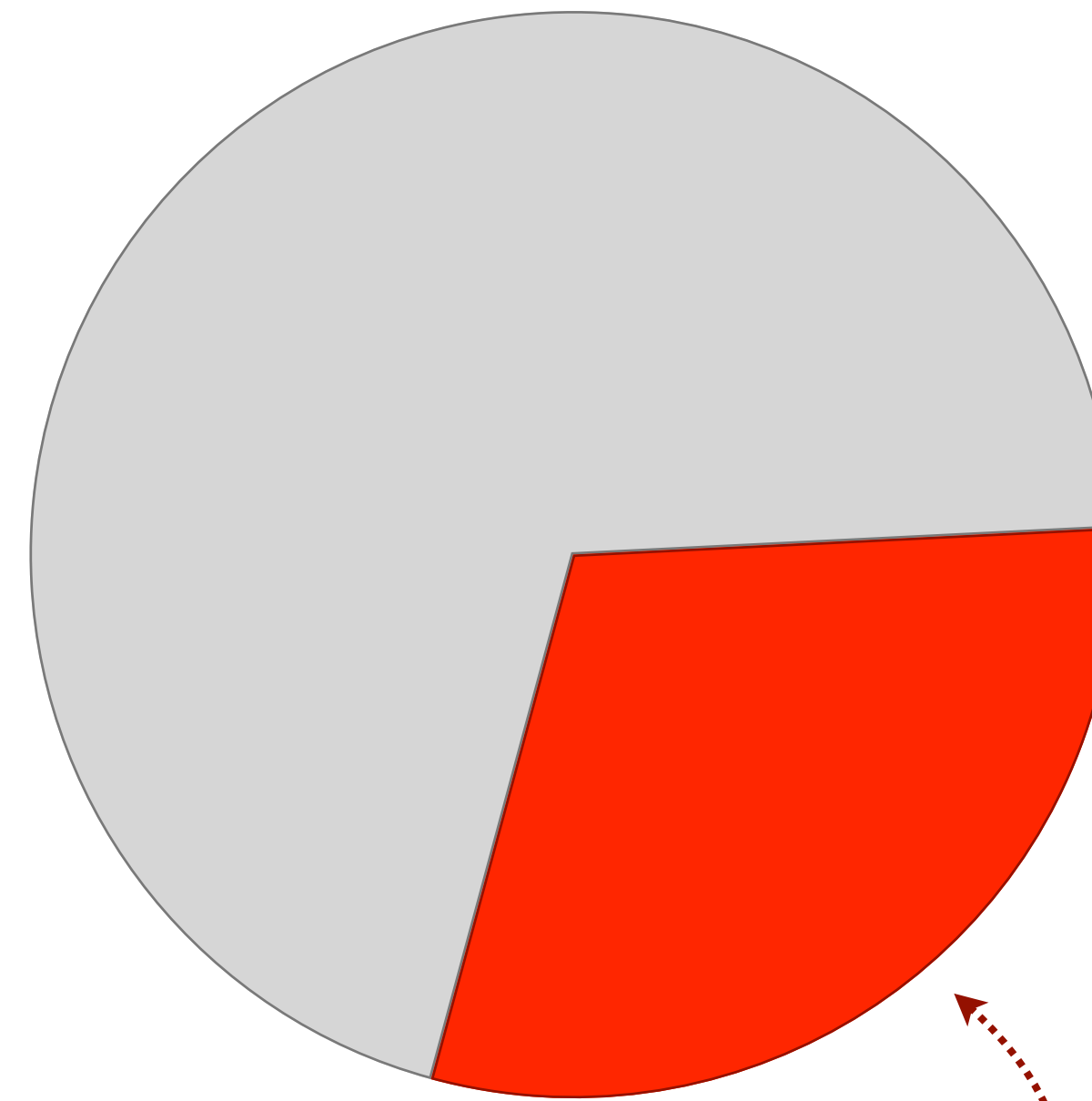
	<b>Ride sharing</b>	<b>Serverless</b>
Infrastructure	Someone else's car	Never manage servers
Cost	Pay per ride	Pay per CPU time

## Legacy applications



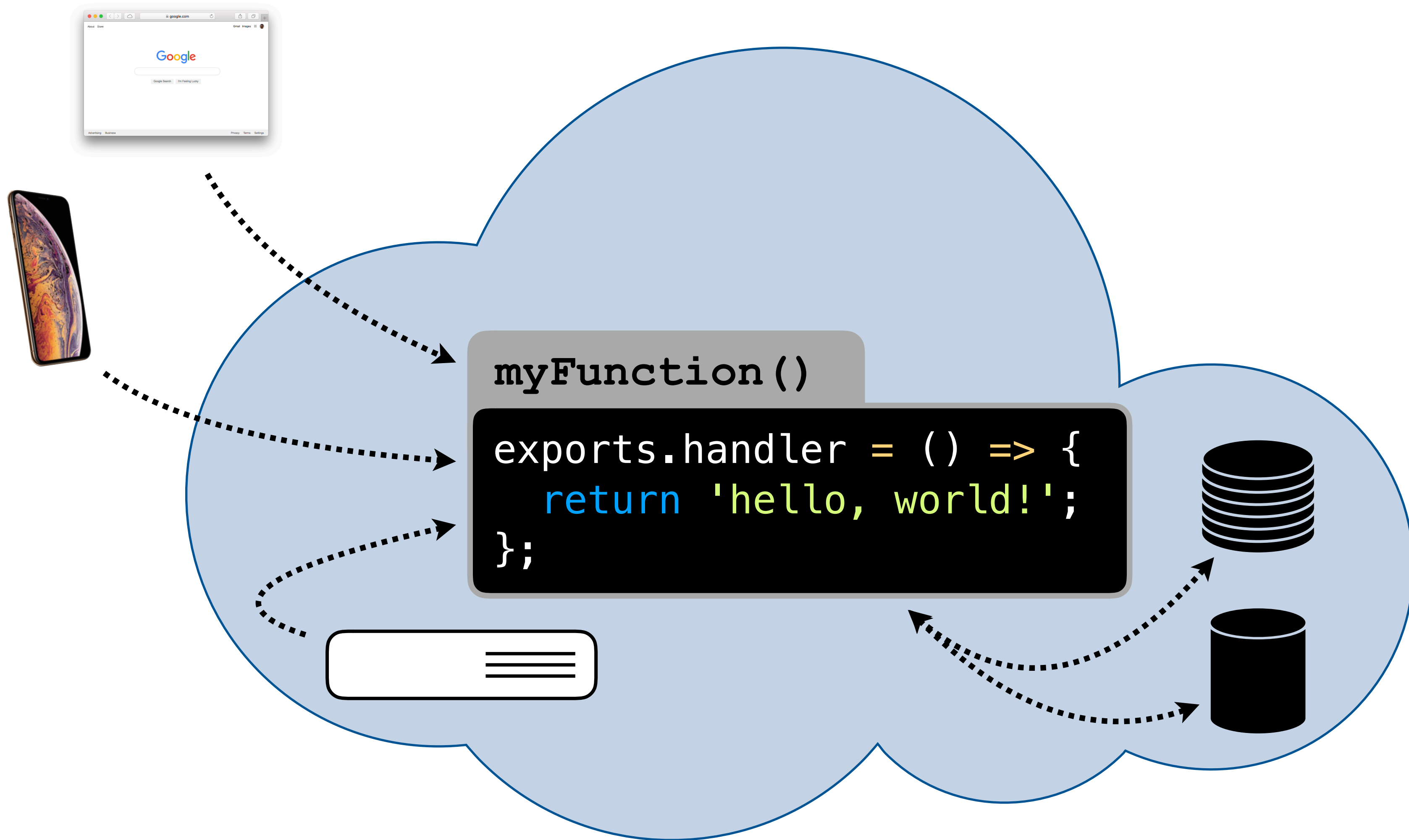
**3% utilization**

## Cloud-Native applications



**30% utilization**

	<b>Ride sharing</b>	<b>Serverless</b>
Infrastructure	Someone else's car	Never manage servers
Cost	Pay per ride	Pay per CPU time
Ease of use	Focus on your phone	Focus on business logic
Who can use?	No driver's license	No cloud expertise



# Demo



# Key providers



**AWS**  
**Lambda**



**Azure**  
**Functions**



**Binaris**



**Google**  
**Cloud Functions**

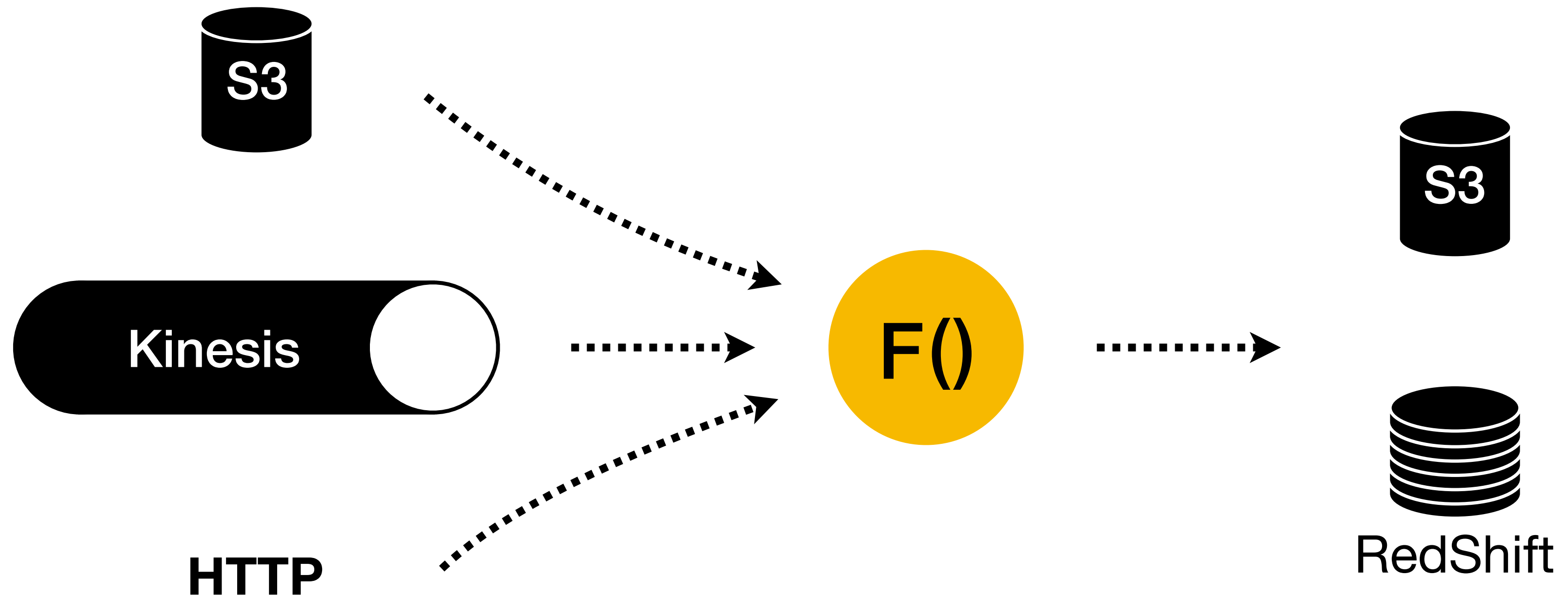


**IBM**  
**Cloud Functions**

- ① **Serverless ETL**
- ② **Serverless MapReduce**
- ③ **Serverless streaming**
- ④ **Serverless training**
- ⑤ **Serverless inference**

# Serverless ETL

---



# Serverless ETL

---

## Upsides

Easy to build

Elastic

Usually cheaper to run

## Challenges

Micro-batching

Run duration limits (15 min on AWS Lambda)

Memory limits (up to 3GB on AWS Lambda)

**ETL accounts for 90% of Lambda CPU cycles**

① Serverless ETL

② **Serverless MapReduce**

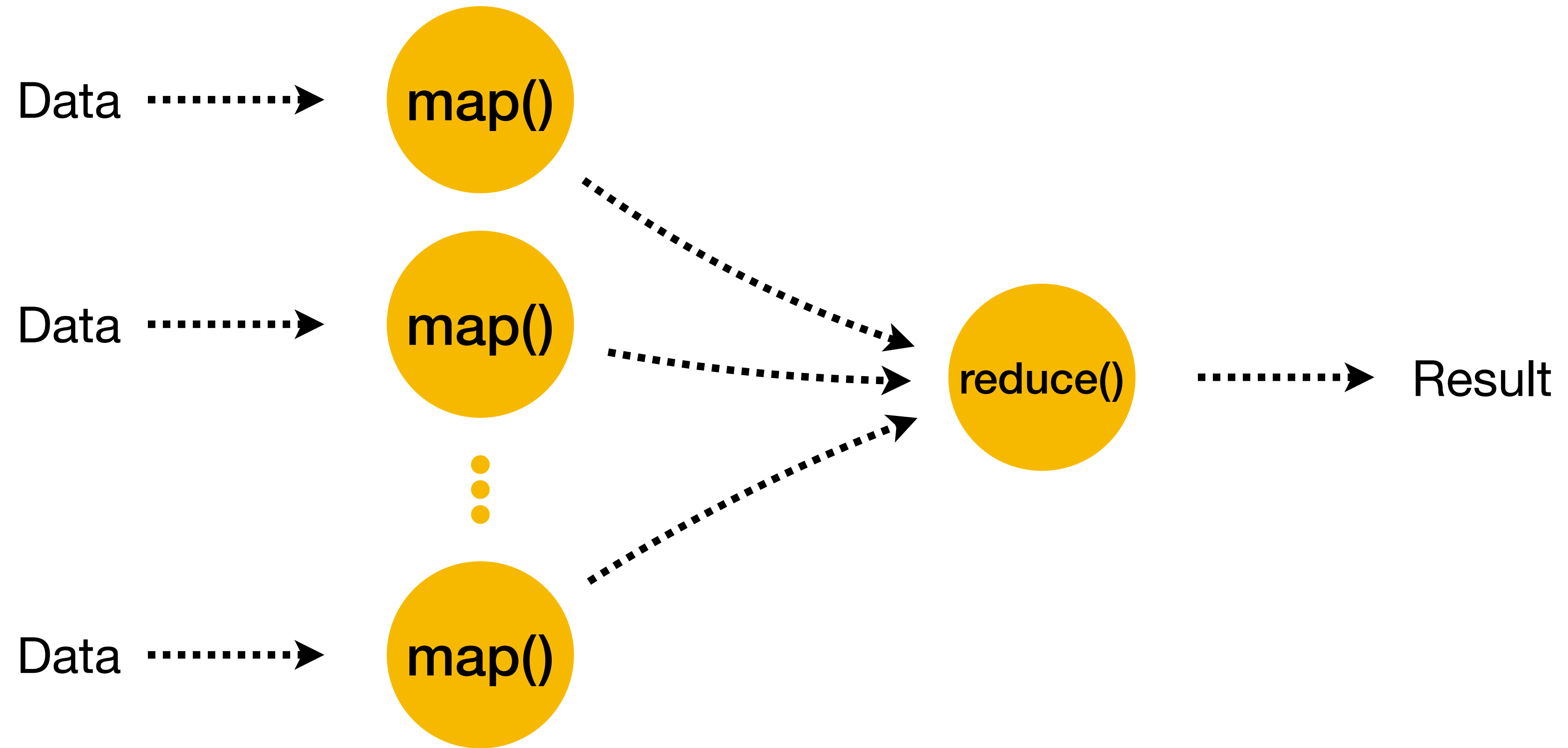
③ **Serverless streaming**

④ **Serverless training**

⑤ **Serverless inference**

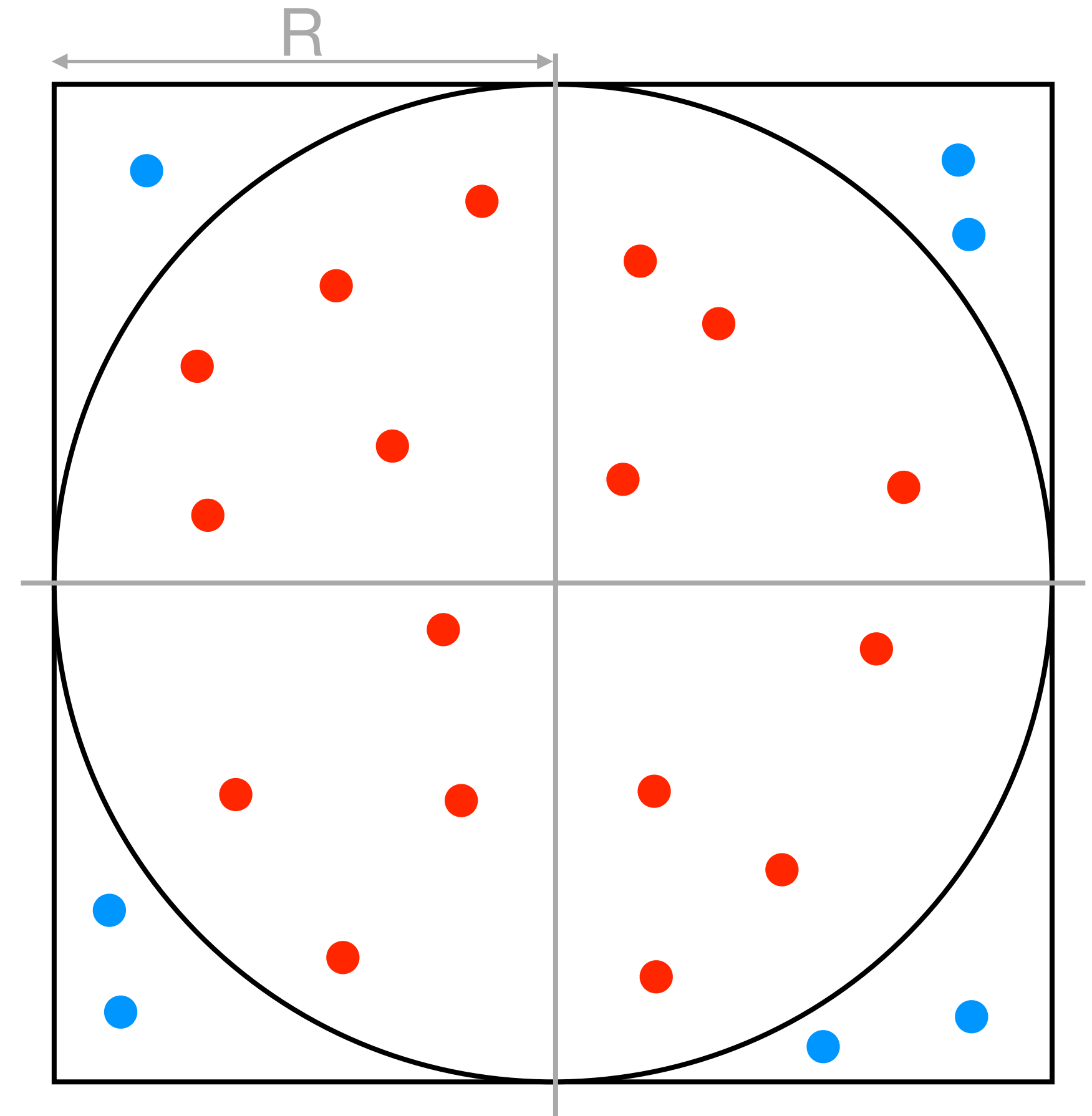
# Serverless MapReduce

---



## Calculating $\pi$ using the Monte Carlo method:

1. Spread random points inside a square
2. Geometry tell us that  
Circle area  $\sim$  # red points  $\sim \pi \cdot R^2$   
Square area  $\sim$  # points  $\sim 4 \cdot R^2$
3. Compute:  
$$\pi = 4 * (\text{\# red points}) / (\text{\# points})$$



# Serverless MapReduce

---

```
function computePi(points) {  
  let inside = 0;  
  
  // repeat points times:  
  for (let i = 0; i < points; i++) {  
  
    // random point (use R = 1)  
    const x = Math.random() * 2 - 1;  
    const y = Math.random() * 2 - 1;  
  
    // is it inside the circle?  
    if (x * x + y * y < 1) {  
      inside++;  
    }  
  }  
  return inside;  
  
  // use circle to square area ratios to compute  $\pi$   
  return 4 * inside / points;  
}
```

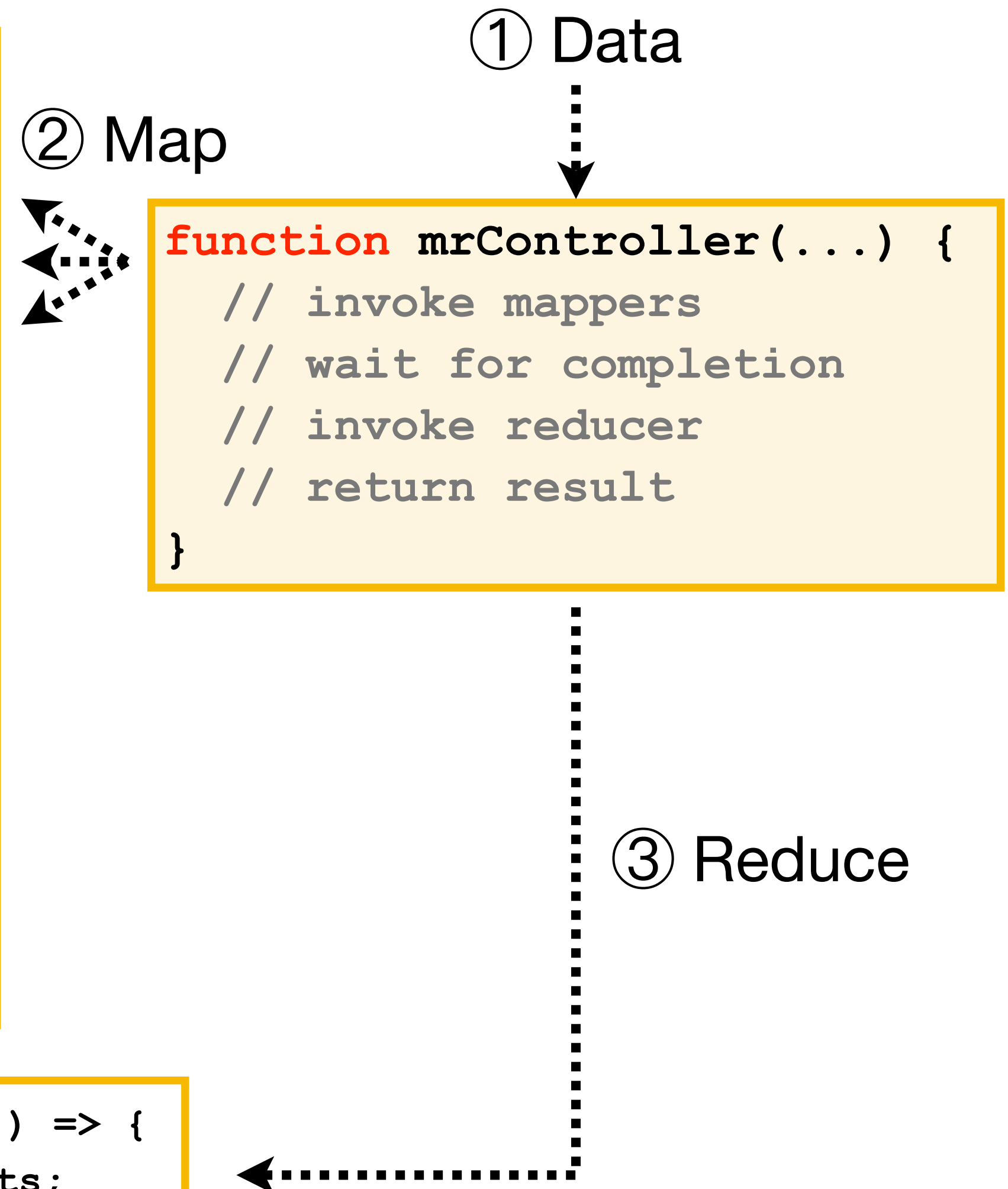
More points == Better  $\pi$



# Serverless MapReduce

```
exports.computePiMapper = ({ points }) => {  
  let inside = 0;  
  
  // repeat points times:  
  for (let i = 0; i < points; i++) {  
  
    // random point (use R = 1)  
    const x = Math.random() * 2 - 1;  
    const y = Math.random() * 2 - 1;  
  
    // is it inside the circle?  
    if (x * x + y * y < 1) {  
      inside++;  
    }  
  }  
  
  return inside;  
}
```

```
exports.computePiReducer = ({ inputs, points }) => {  
  return 4 * inputs.reduce((a, b) => a + b) / points;  
};
```



# Demo

# Serverless MapReduce

---

## Upsides

Easy to build

Elastic

Scale at your fingertips

## Challenges

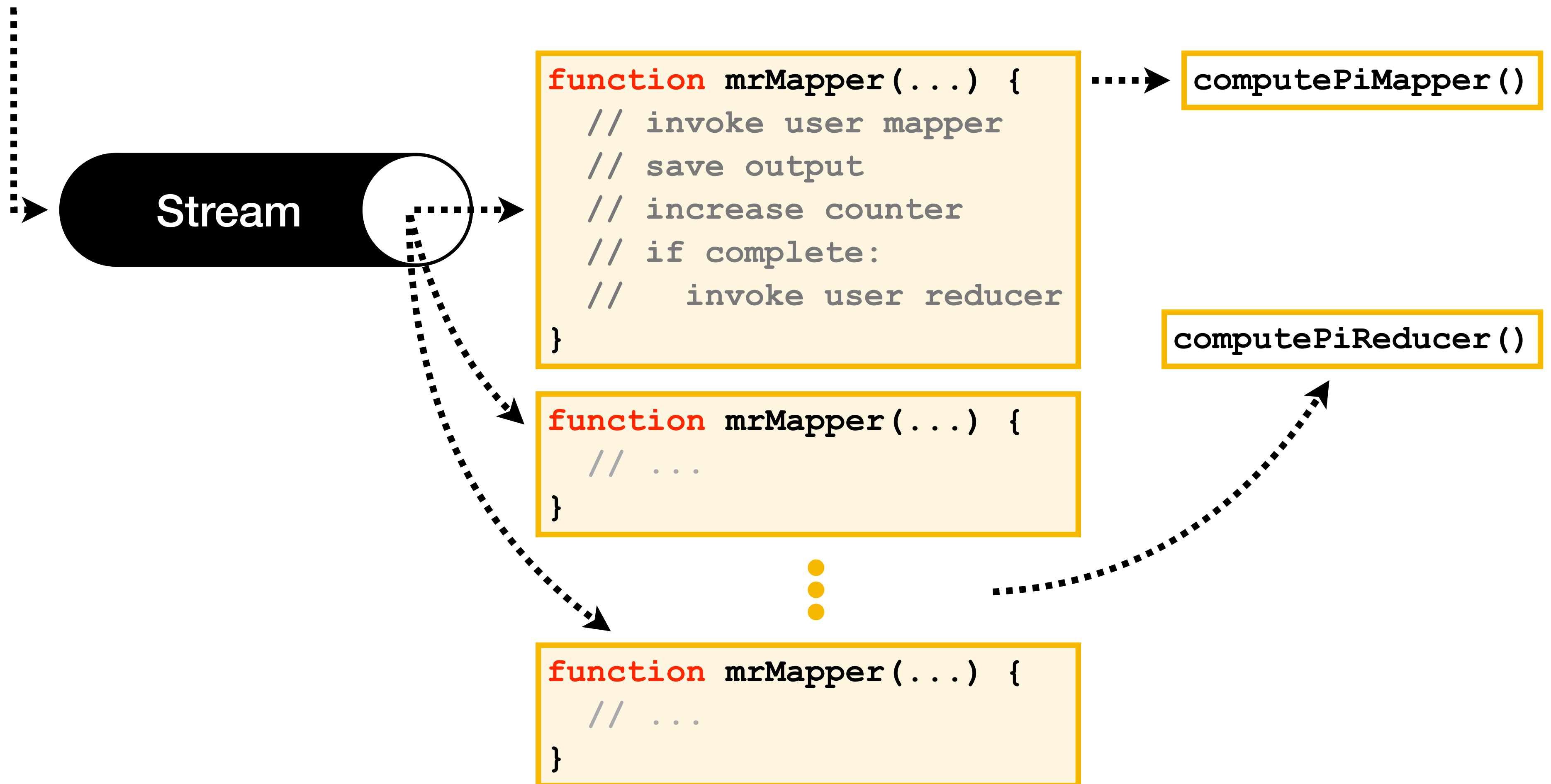
Failures?

Retries?

Laggards?

# Serverless MapReduce

```
function mrAsyncController(...) {  
  // add map commands to stream  
}
```



# Serverless MapReduce

---



[github.com/binaris/functions-examples](https://github.com/binaris/functions-examples)

# Serverless MapReduce

---

## Upside

Easy to build (with framework)

Elastic

Scale to your fingertips

## Challenges

Need cache (e.g. Redis) to hold state

Need storage (e.g. S3) to handle data

① Serverless ETL

② Serverless MapReduce

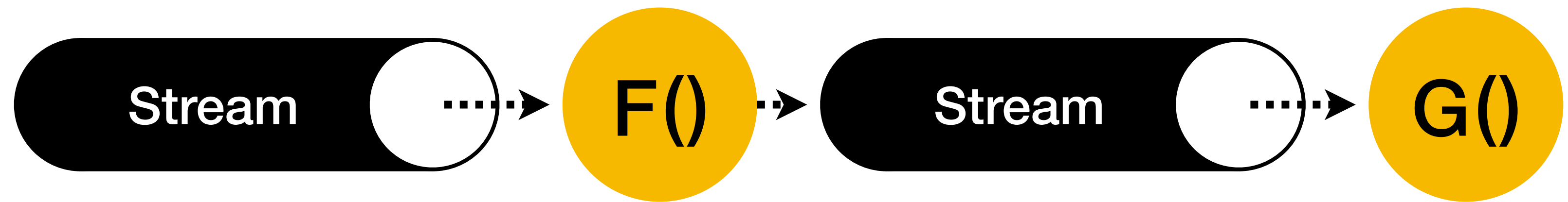
③ **Serverless streaming**

④ **Serverless training**

⑤ **Serverless inference**

# Serverless streaming

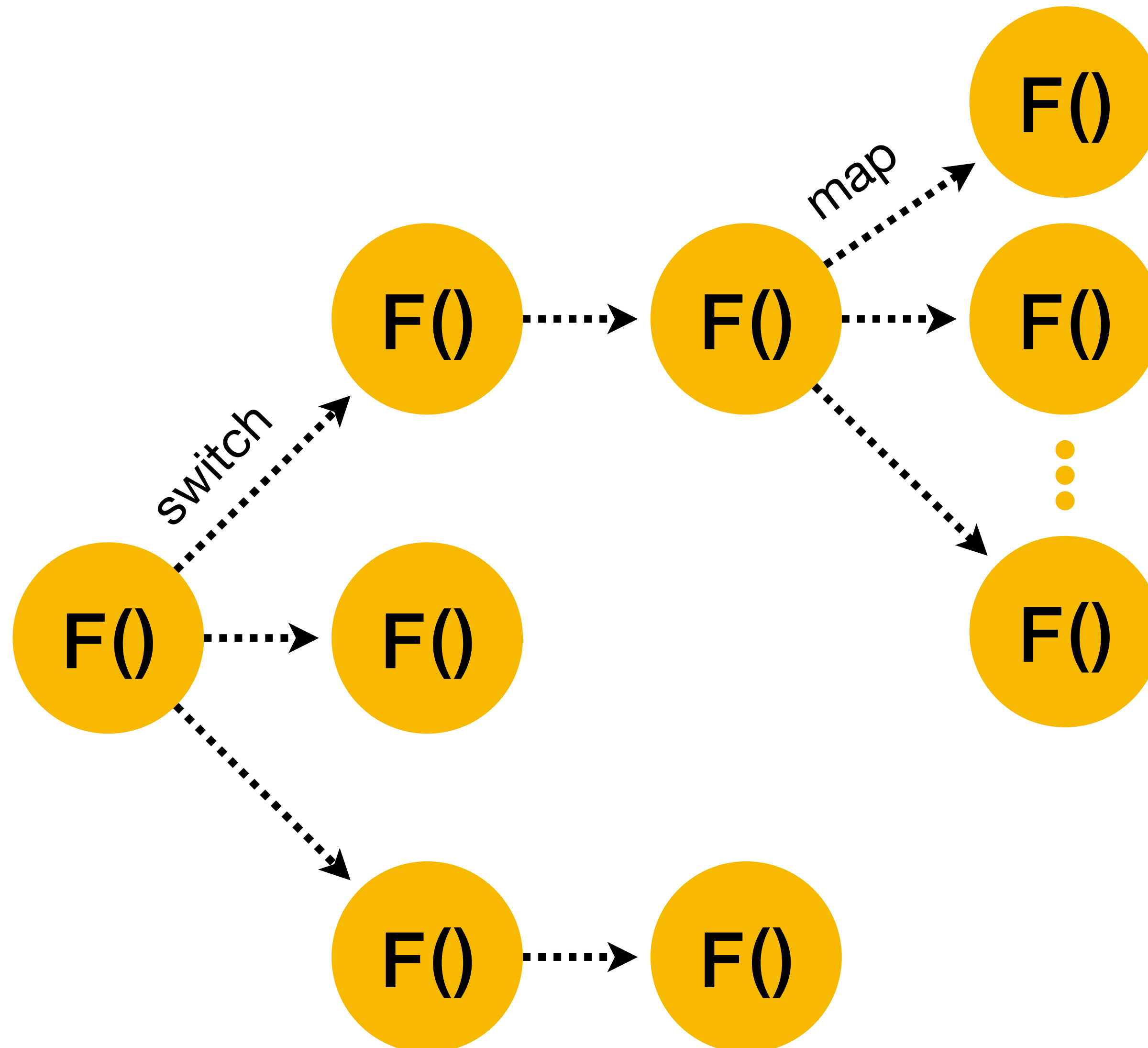
---





# Serverless streaming

---



# Serverless streaming

---

## Upsides

Easy to build

Elastic

Scale to your fingertips

Real-time

**Shameless plug  
for Binaris**



## Challenges

Need a smarter framework

- ① Serverless ETL
- ② Serverless MapReduce
- ③ Serverless streaming
- ④ Serverless training**
- ⑤ Serverless inference**

# Serverless training

---

Today, serverless is not a good fit for training 😞

**But, if**

- ① Your data is small (few GB)
- ② You need hyperparameter optimization

**Then you can leverage serverless scale!**

- ① Serverless ETL
- ② Serverless MapReduce
- ③ Serverless streaming
- ④ Serverless training
- ⑤ Serverless inference**



# Serverless inference

---



k-means



(Yeah, we know this is not really inference, but same computational pattern)



# Demo

# Serverless inference

---

## Upsides

Easy to build

Use standard tools (py, sklearn, TF)

Elastic

Scale to your fingertips

Real-time (no more plugs)

## Challenges

No GPU

Might be limited CPU/memory  
for some TF models



- ① **Serverless ETL**
- ② **Serverless MapReduce**
- ③ **Serverless streaming**
- ④ **Serverless training**
- ⑤ **Serverless inference**

# Serverless

## for data and AI

Serverless is easy to use

Cosmic ✨ scale without the rocket 🚀 science

Pay as you go

[avner@binaris.com](mailto:avner@binaris.com)

[@avnerbraverman](https://twitter.com/avnerbraverman)

# Rate today's session

## Cyberconflict: A new era of war, sabotage, and fear

**See passes & pricing**

[David Sanger](#) (The New York Times)  
9:55am-10:10am Wednesday, March 27, 2019  
Location: Ballroom  
Secondary topics: [Security and Privacy](#)

[Add to Your Schedule](#)  
[Add Comment or Question](#)

**Rate This Session**

We're living in a new era of constant sabotage, misinformation, and fear, in which everyone is a target, and you're often the collateral damage in a growing conflict among states. From crippling infrastructure to sowing discord and doubt, cyber is now the weapon of choice for democracies, dictators, and terrorists.

David Sanger explains how the rise of cyberweapons has transformed geopolitics like nothing since the invention of the atomic bomb. Moving from the White House Situation Room to the dens of Chinese, Russian, North Korean, and Iranian hackers to the boardrooms of Silicon Valley, David reveals a world coming face-to-face with the perils of technological revolution—a conflict that the United States helped start when it began using cyberweapons against Iranian nuclear plants and North Korean missile launches. But now we find ourselves in a conflict we're uncertain how to control, as our adversaries exploit vulnerabilities in our hyperconnected nation and we struggle to figure out how to deter these complex, short-of-war attacks.

**David Sanger**  
The New York Times

David E. Sanger is the national security correspondent for the *New York Times* as well as a national security and political contributor for CNN and a frequent guest on *CBS This Morning*, *Face the Nation*, and many PBS shows.


Session page on conference website

✓ **Attending** [Notes](#) [Remove](#)

## Cyberconflict: A new era of war, sabotage, and fear

🕒 9:55 AM - 10:10 AM, Wed, Mar 27, 2019

### Speakers



**David Sanger**  
National Security Correspondent  
The New York Times

📍 Ballroom

## Keynotes

David Sanger explains how the rise of cyberweapons has transformed geopolitics like nothing since the invention of the atomic bomb. From crippling infrastructure to sowing discord and doubt, cyber is now the weapon of choice for democracies, dictators, and terrorists.

[SESSION EVALUATION](#)

O'Reilly Events App