

AWS re:Invent

NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

COP343

Observability best practices at Amazon

David Yanacek (he/him)

Sr. Principal Engineer
AWS Monitoring & Observability

Ian McGarry (he/him)

Director of Engineering
AWS Monitoring & Observability

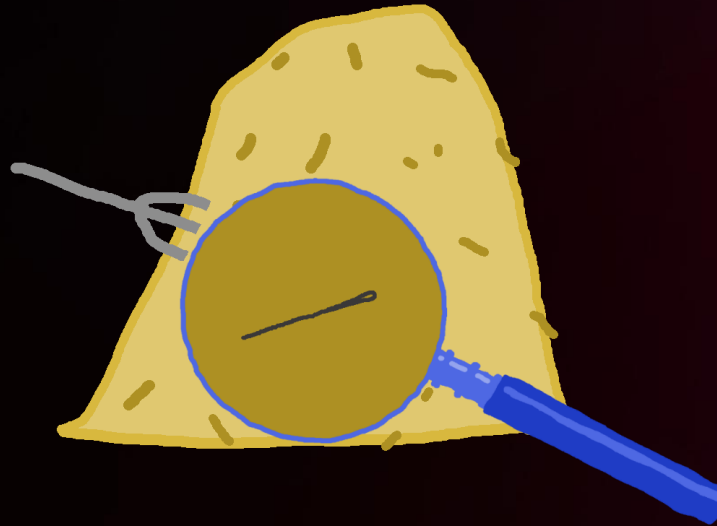


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Table of contents



The DevOps
flywheel at Amazon



Find the
root cause

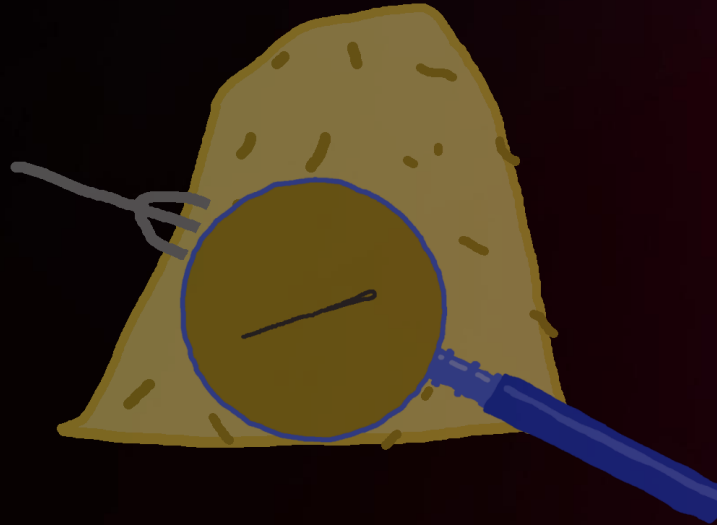


Measure from
everywhere

Chapter one



The DevOps
flywheel at Amazon

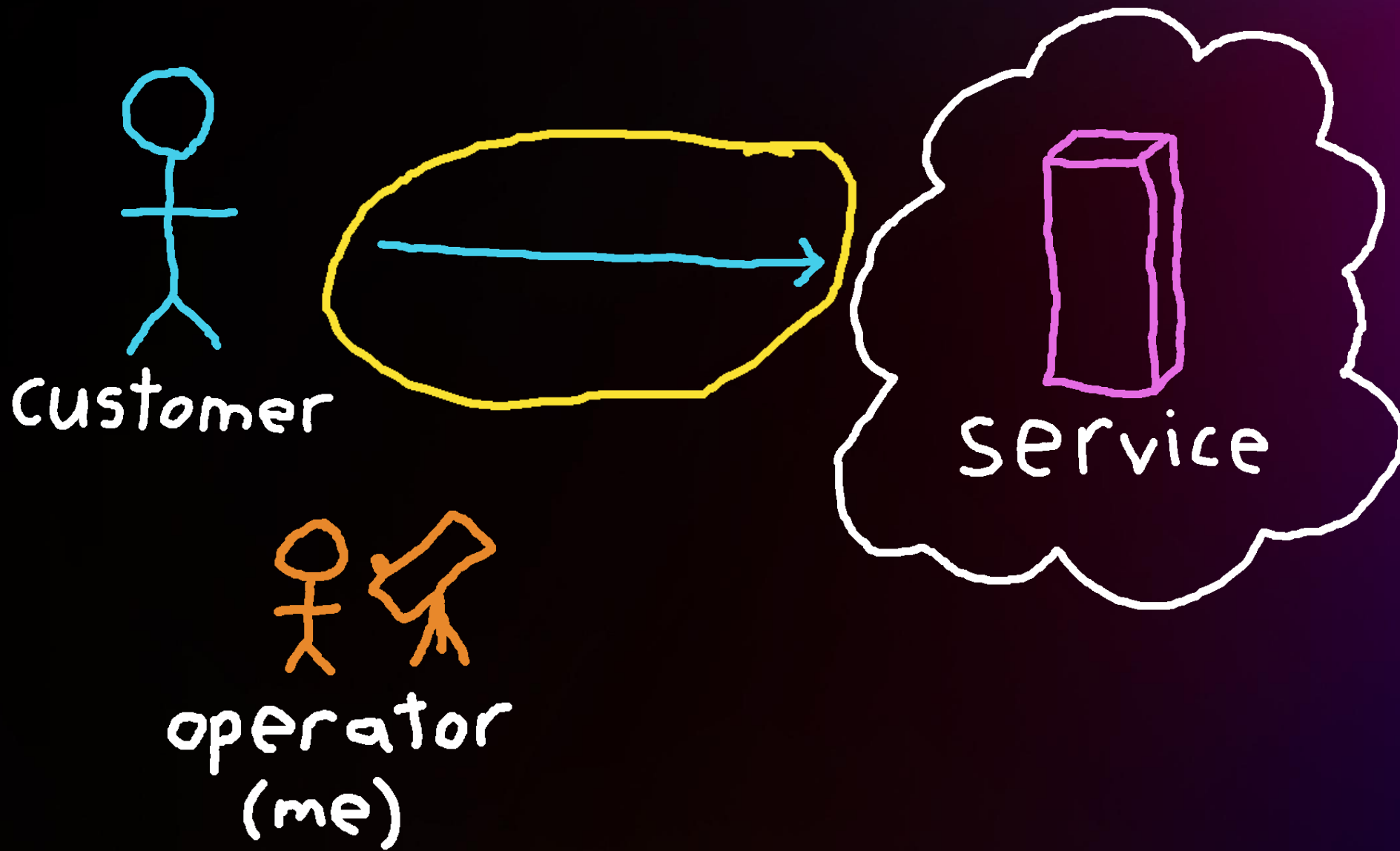


Find the
root cause

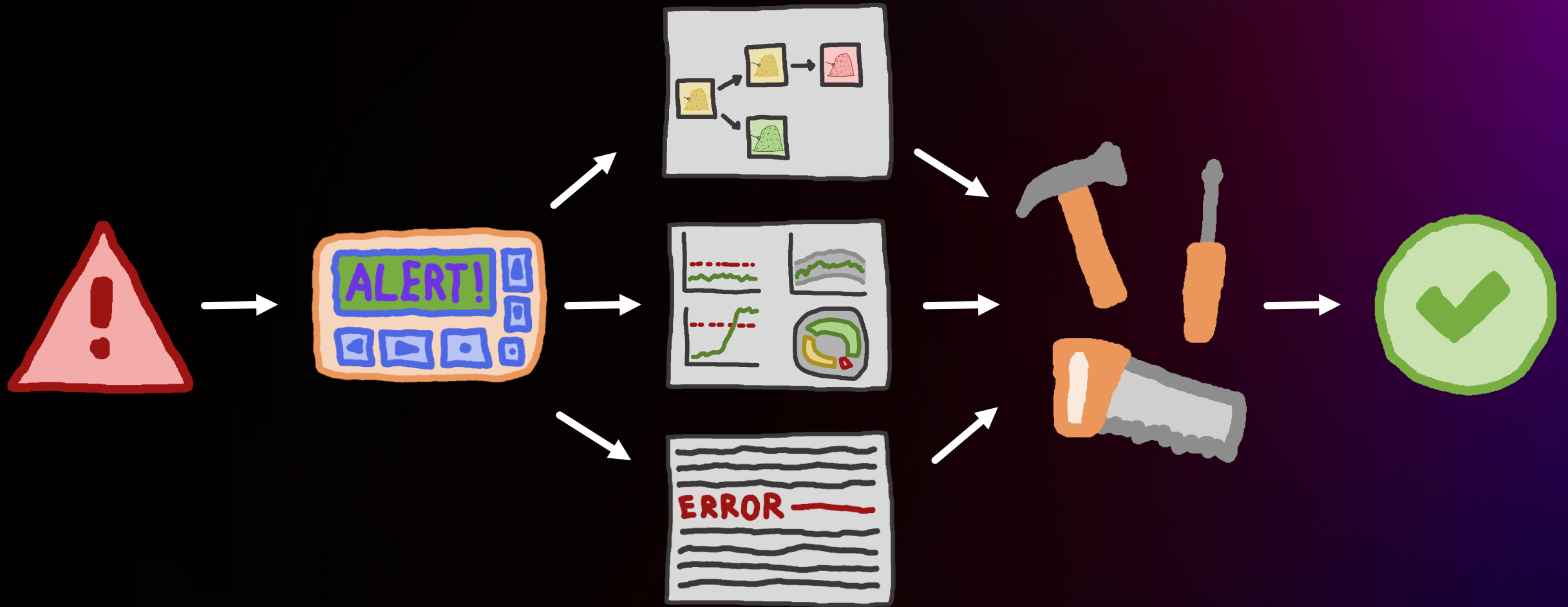


Measure from
everywhere

Observing customer experience



Responding to incidents



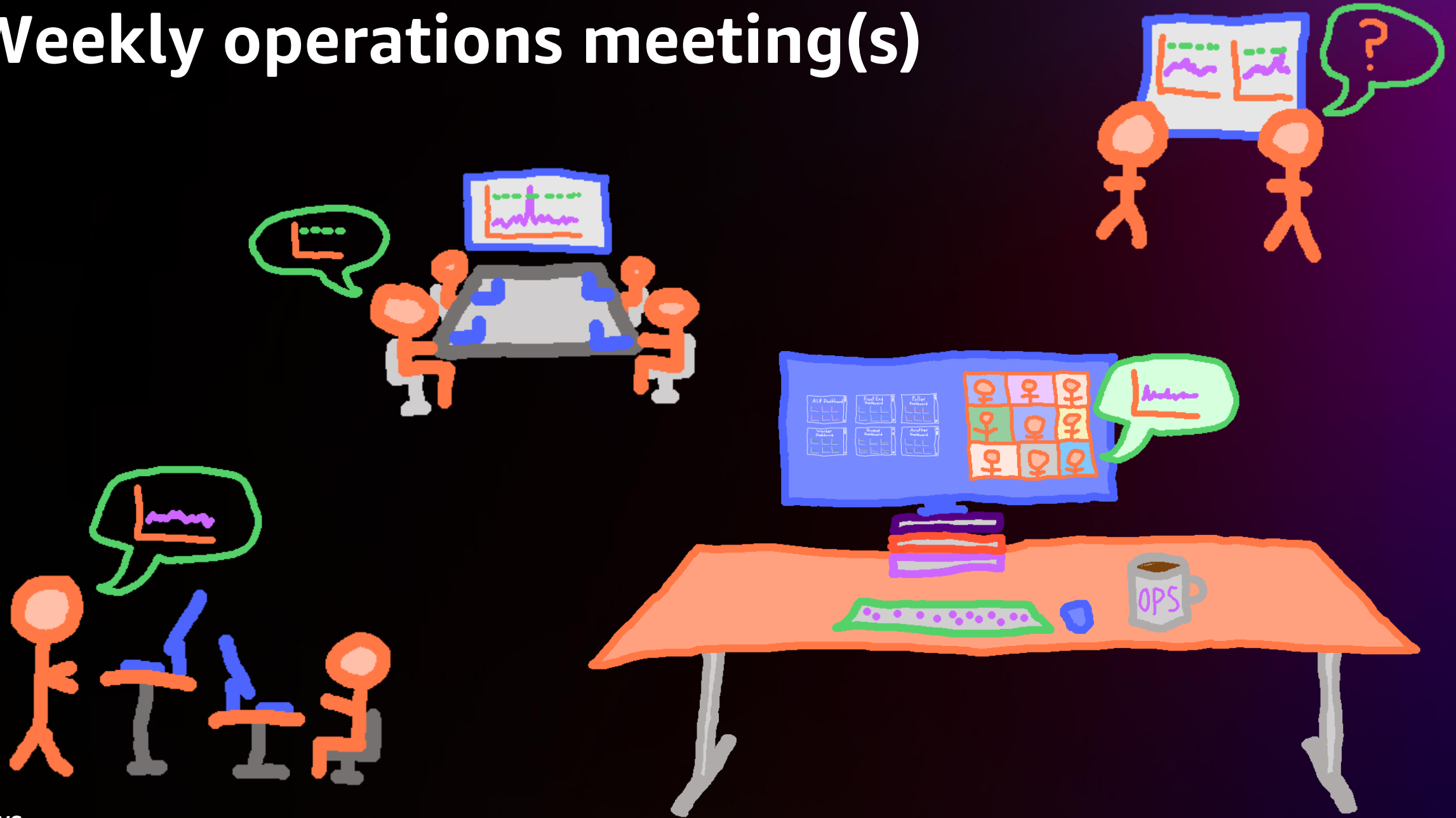
Post-incident retrospectives



A blue magnifying glass is positioned over a log entry. The log entry is as follows:

```
[WARN] Error talking to database  
[INFO] Retrying call after 5 seconds  
[WARN] Connect failed to retrieve data  
[ERROR] Error talking to database  
[INFO] Finished with call to database  
[ERROR] Error talking to database  
[WARN] Cache miss retrieving data  
[FATAL] Critical timeout talking to database  
[WARN] Retry after 5 seconds to retrieve data  
[ERROR] Error talking to database  
[WARN] Waiting 5 seconds before retrying  
[INFO] Making new connection to database  
[INFO] Removing connection from pool  
[WARN] Retry after 5 seconds to retrieve data  
[ERROR] Error talking to database  
[INFO] Making new connection to database
```

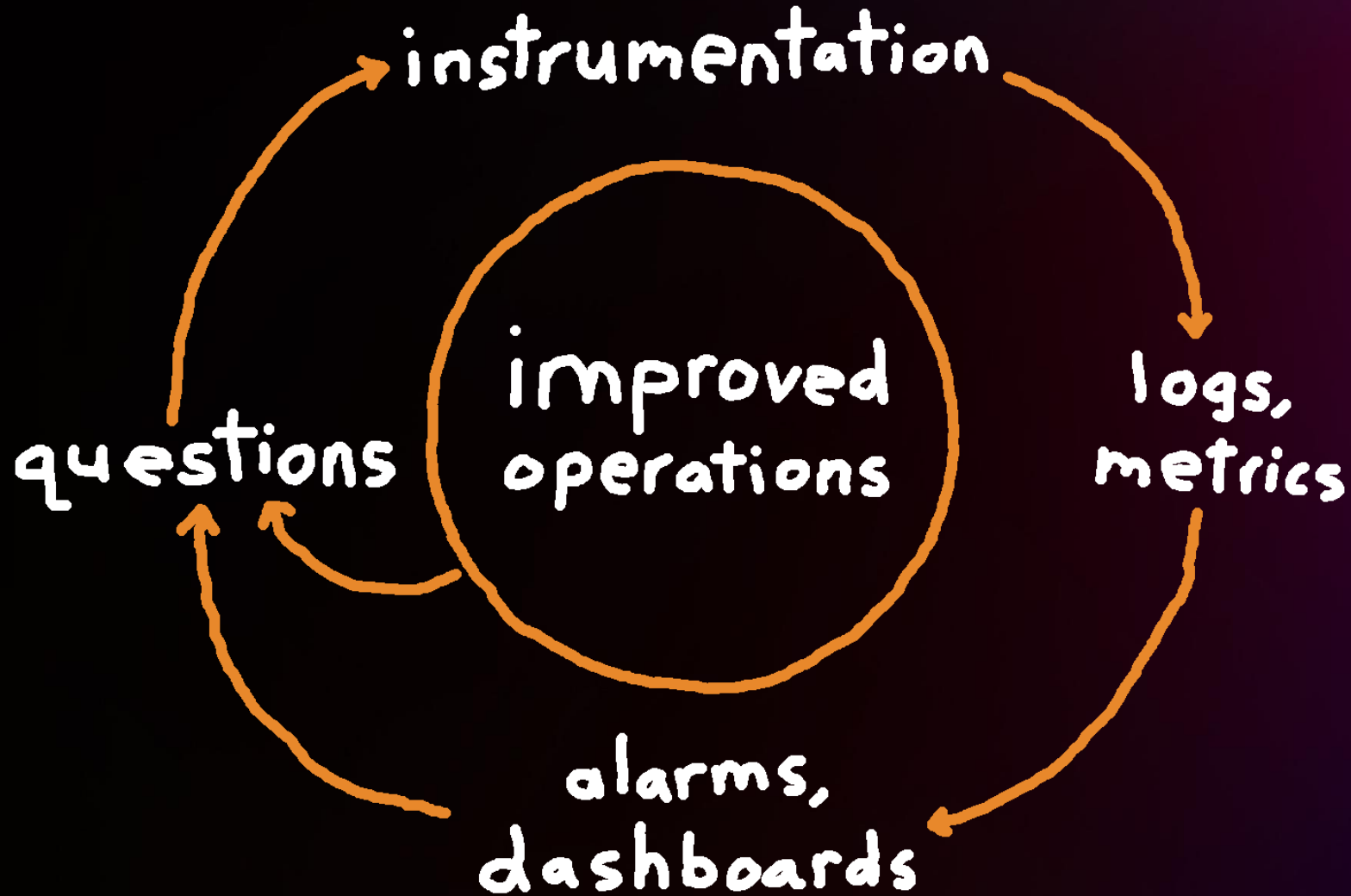
Weekly operations meeting(s)



Observability at scale

“Amazon CloudWatch is used to monitor more than 9 quadrillion metric observations, ingests more than 5 exabytes of logs per month.” As of November 2022

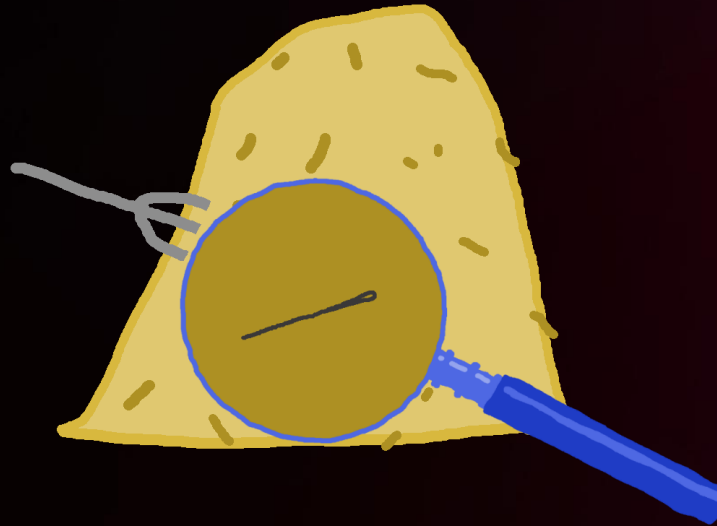
The cycle of observability



Chapter two



The DevOps
flywheel at Amazon

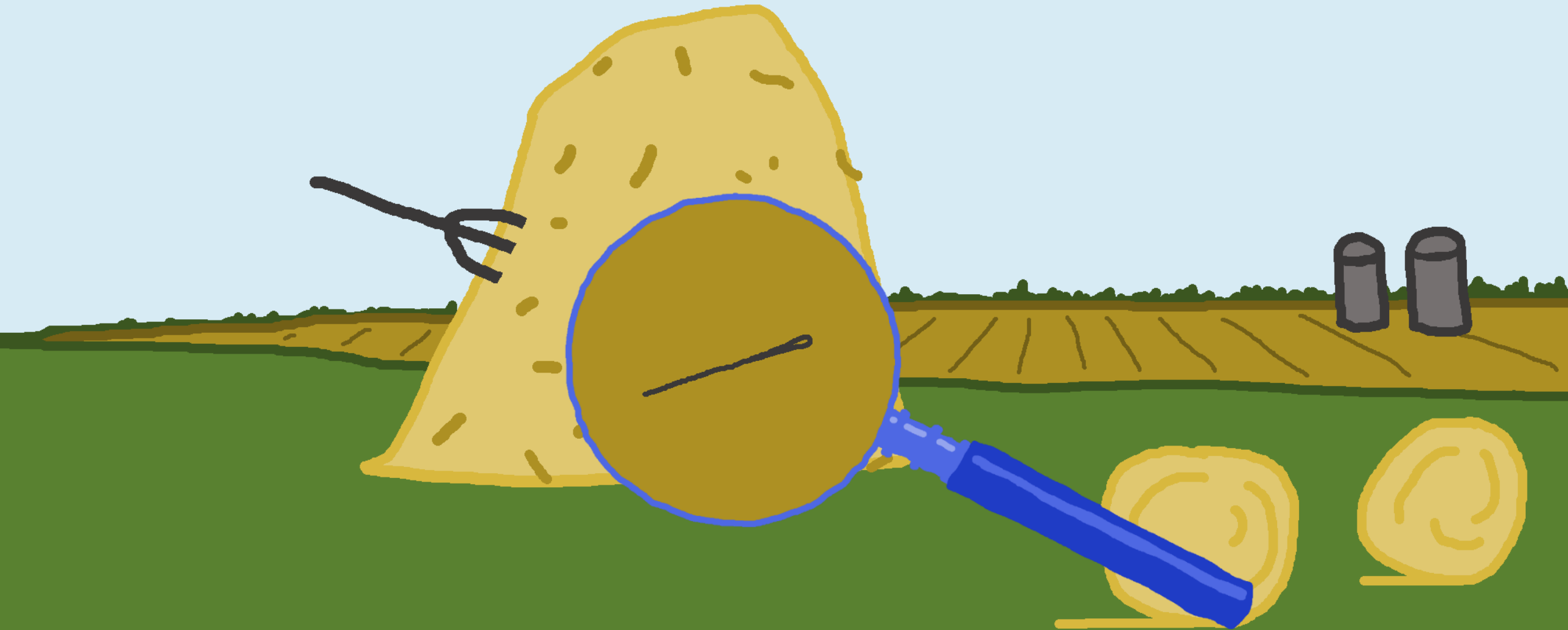


Find the
root cause

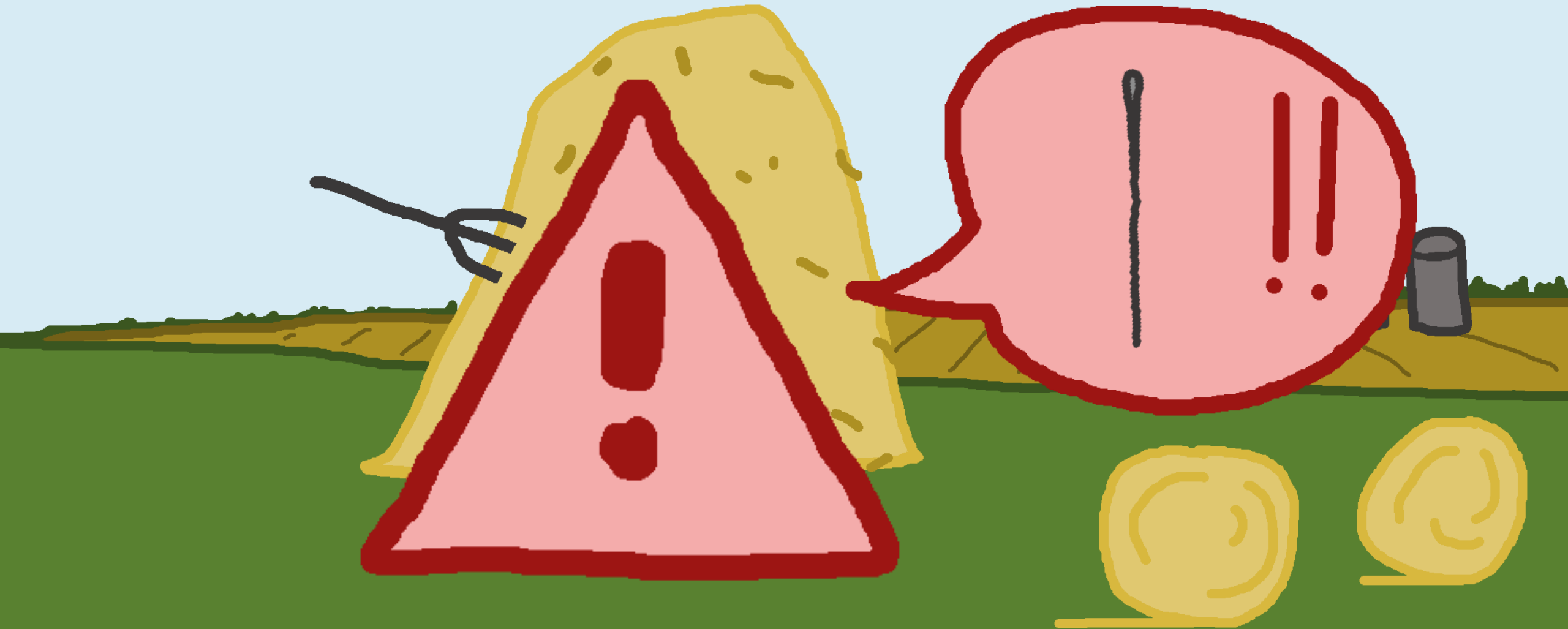


Measure from
everywhere

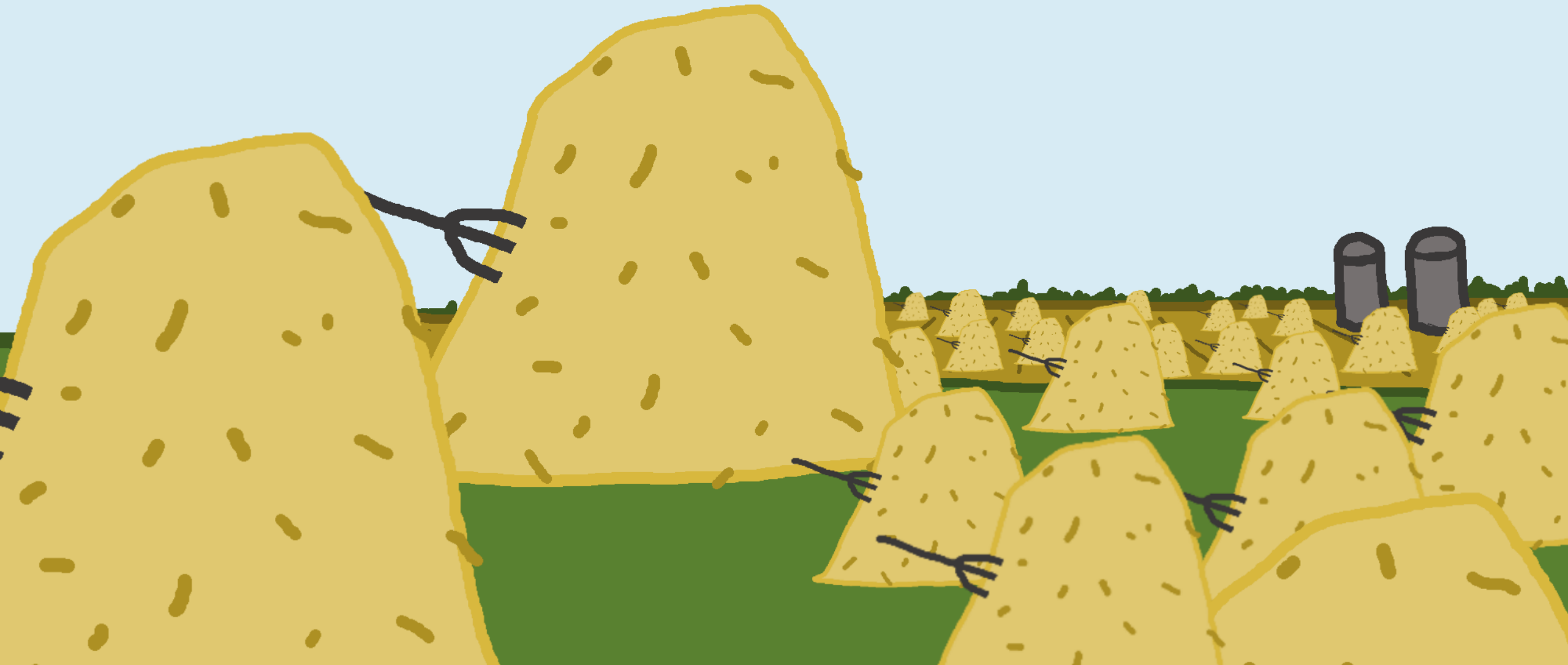
Finding the needle in the haystack



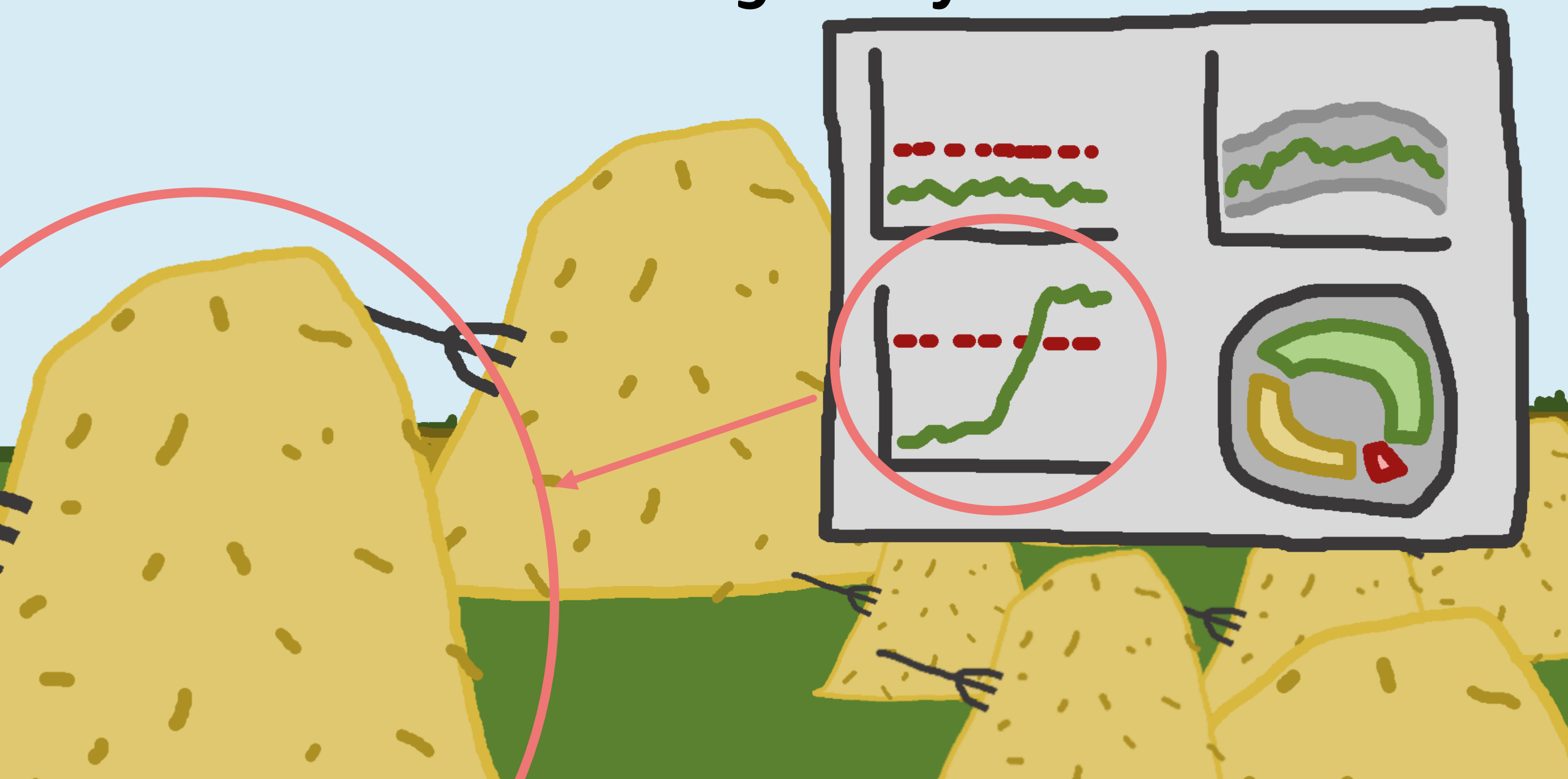
Alarm: Awareness of a needle



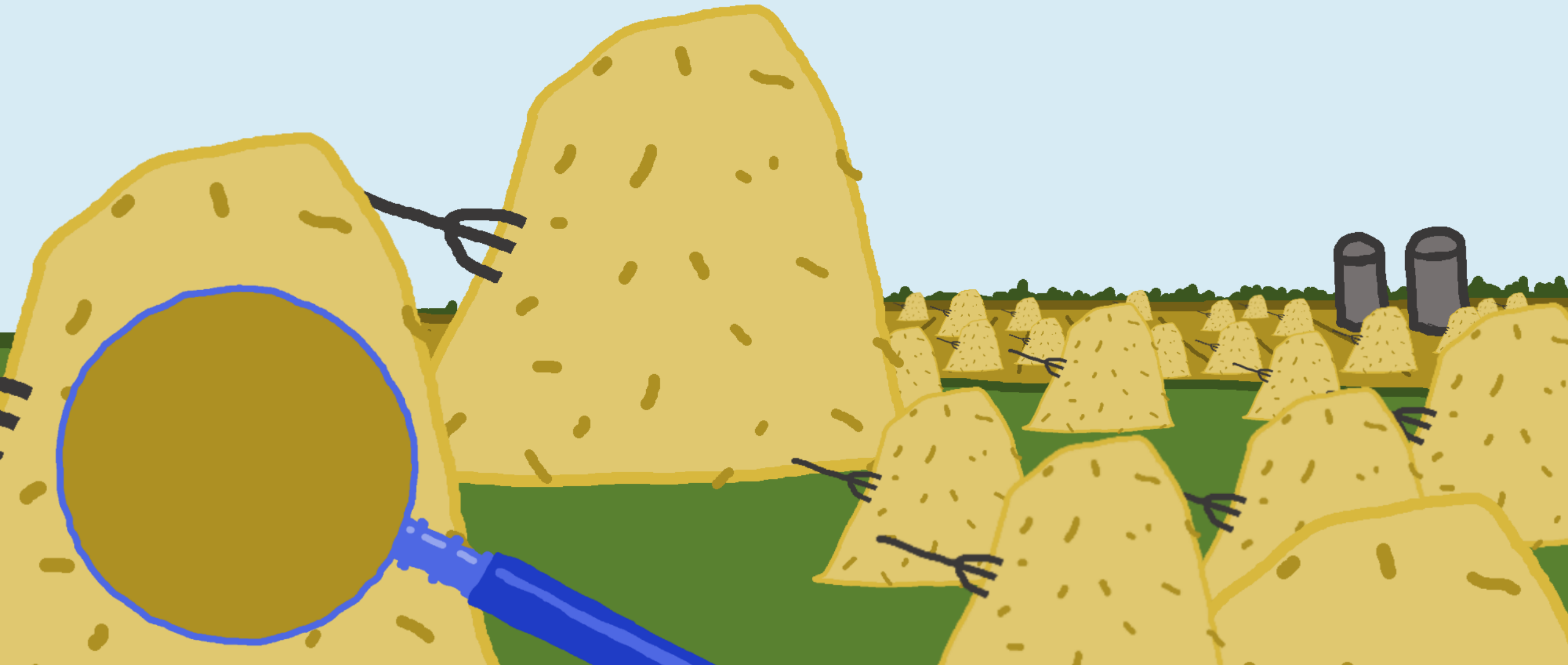
Which haystack?



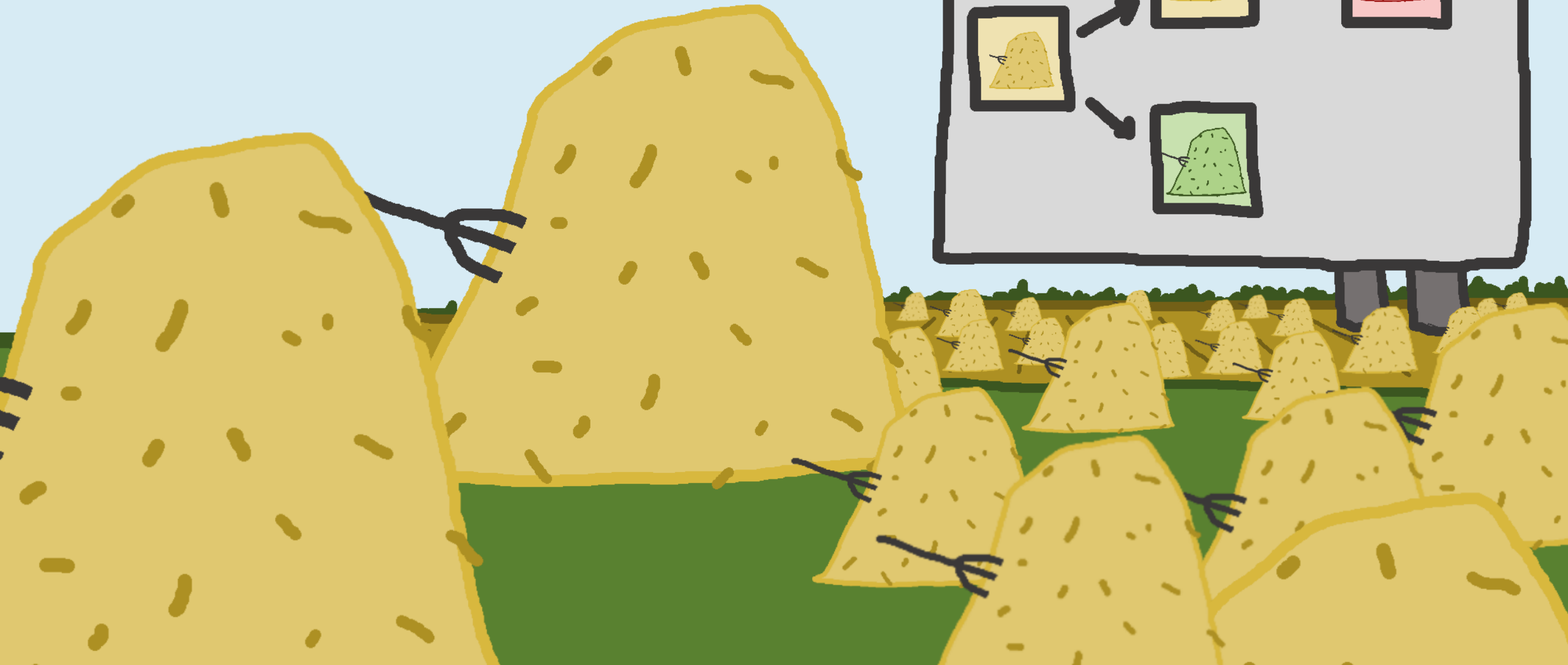
Dashboard: Find the right haystack



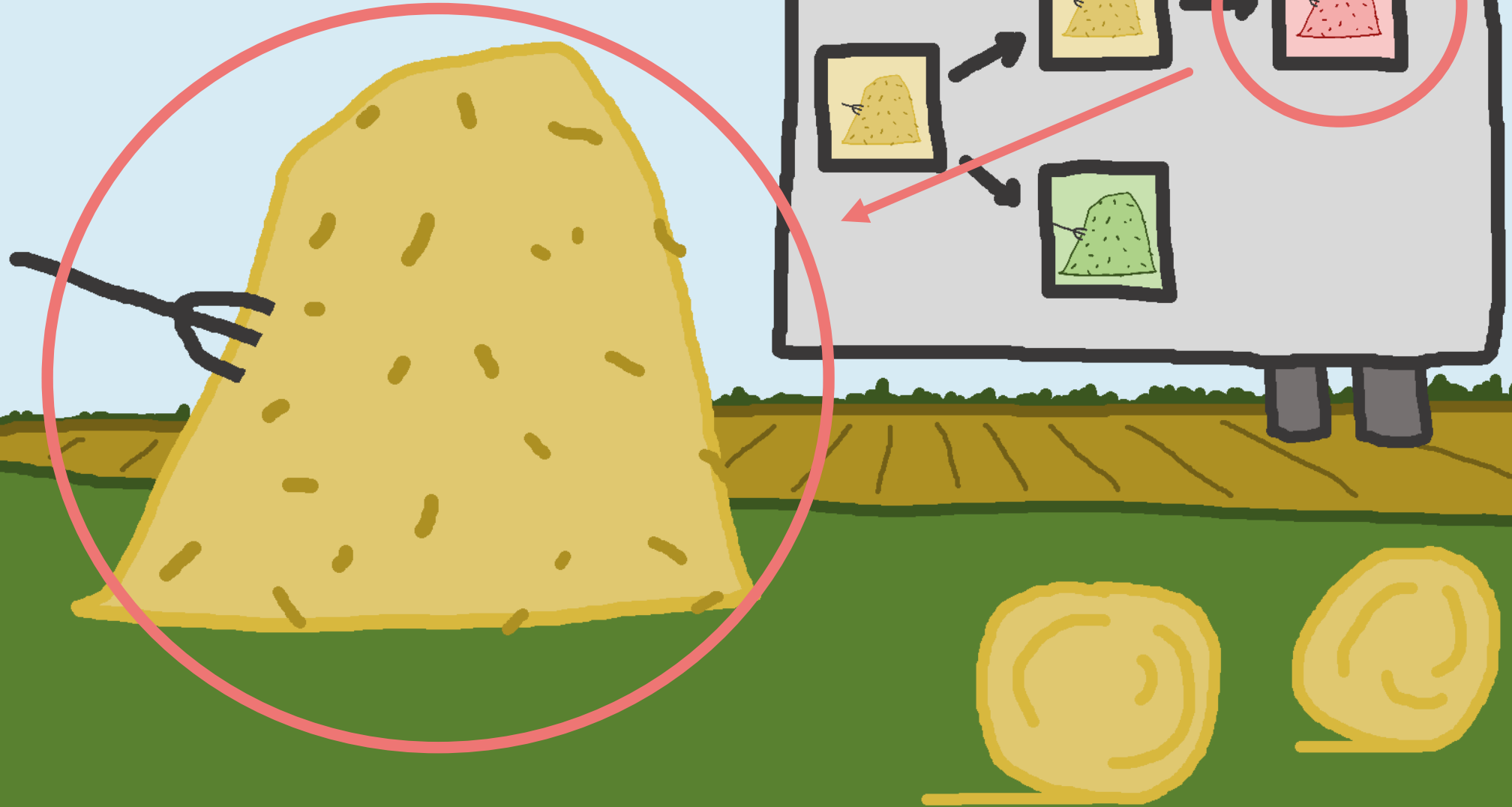
Wrong answer, no needle



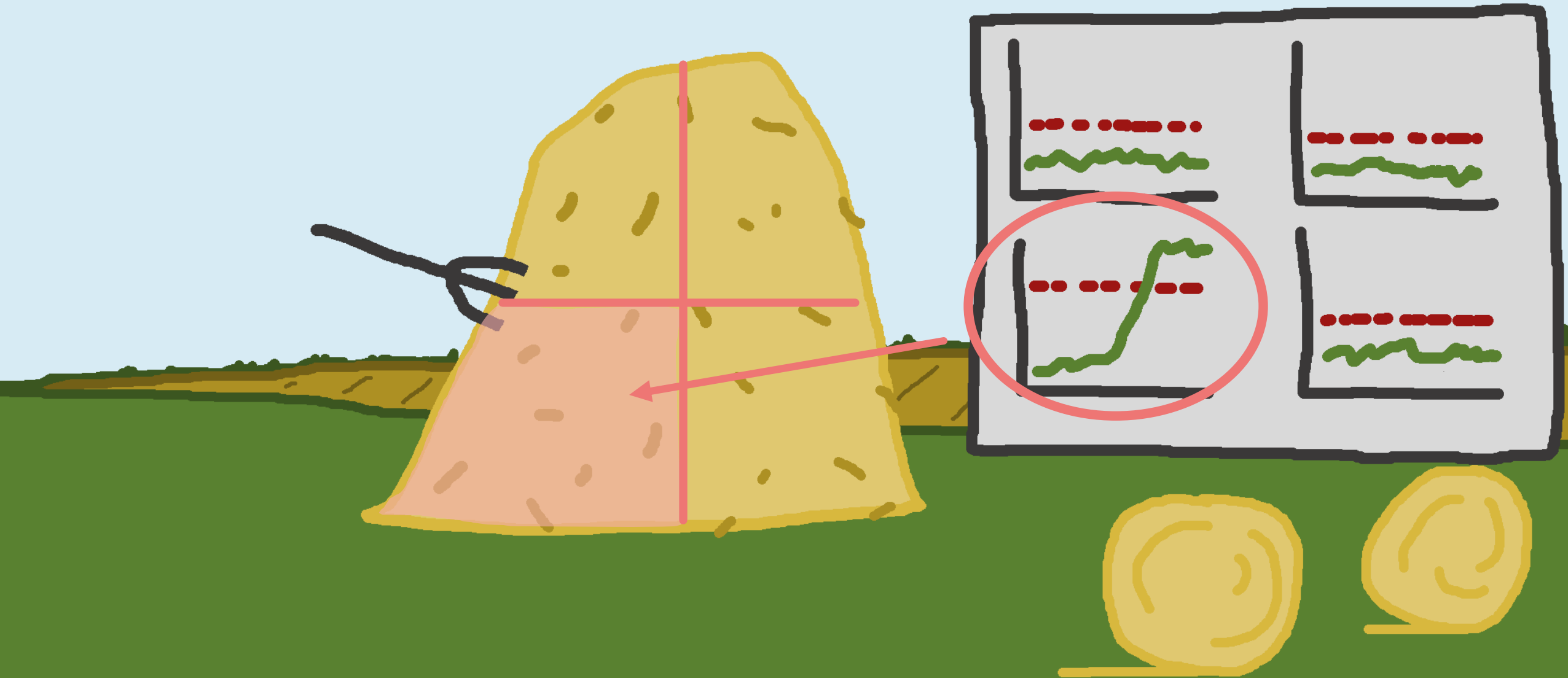
Traces: Following the clues



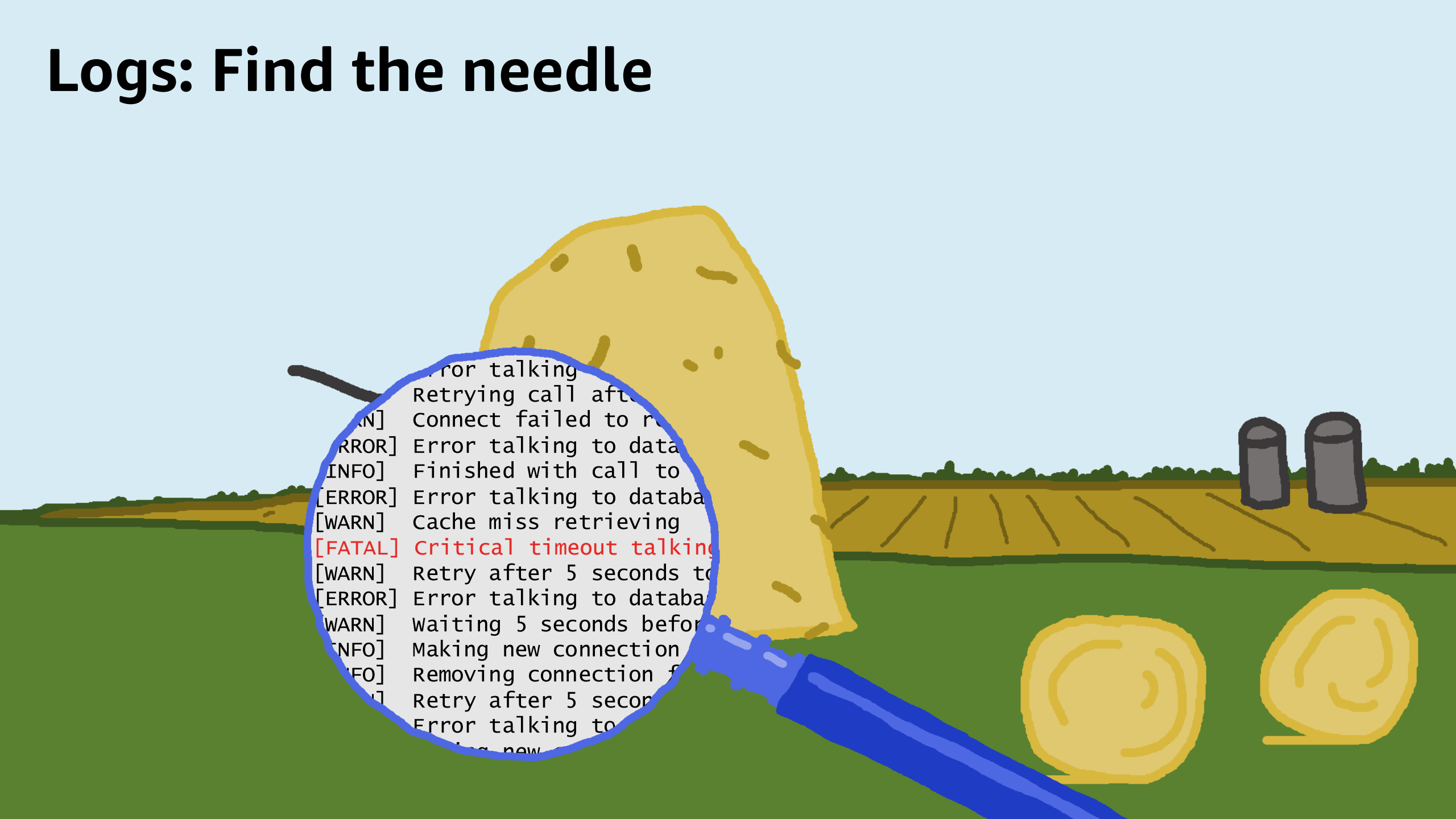
Traces: Following the clues



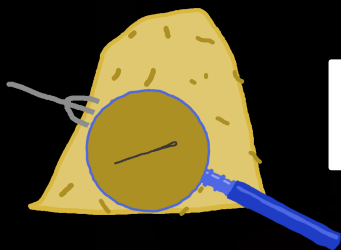
Metrics: Find the part of the haystack



Logs: Find the needle

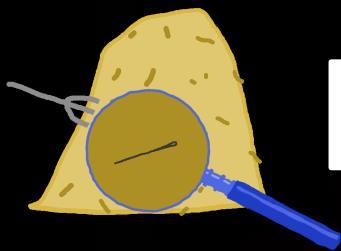


Error talking to database
Retrying call after 5 seconds
[WARN] Connect failed to remote database
[ERROR] Error talking to database
[INFO] Finished with call to database
[ERROR] Error talking to database
[WARN] Cache miss retrieving
[FATAL] Critical timeout talking to database
[WARN] Retry after 5 seconds to database
[ERROR] Error talking to database
[WARN] Waiting 5 seconds before retrying
[INFO] Making new connection to database
[INFO] Removing connection from pool
[WARN] Retry after 5 seconds to database
Error talking to database
log new



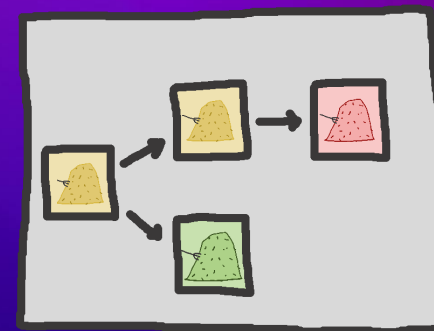
Finding the root cause

- Propagate trace information
- Build lots of dashboards
- Get multi-dimensional
- Use high cardinality metrics
- Dig even deeper with log analysis
- Look at the raw logs
- Analyze using profiler

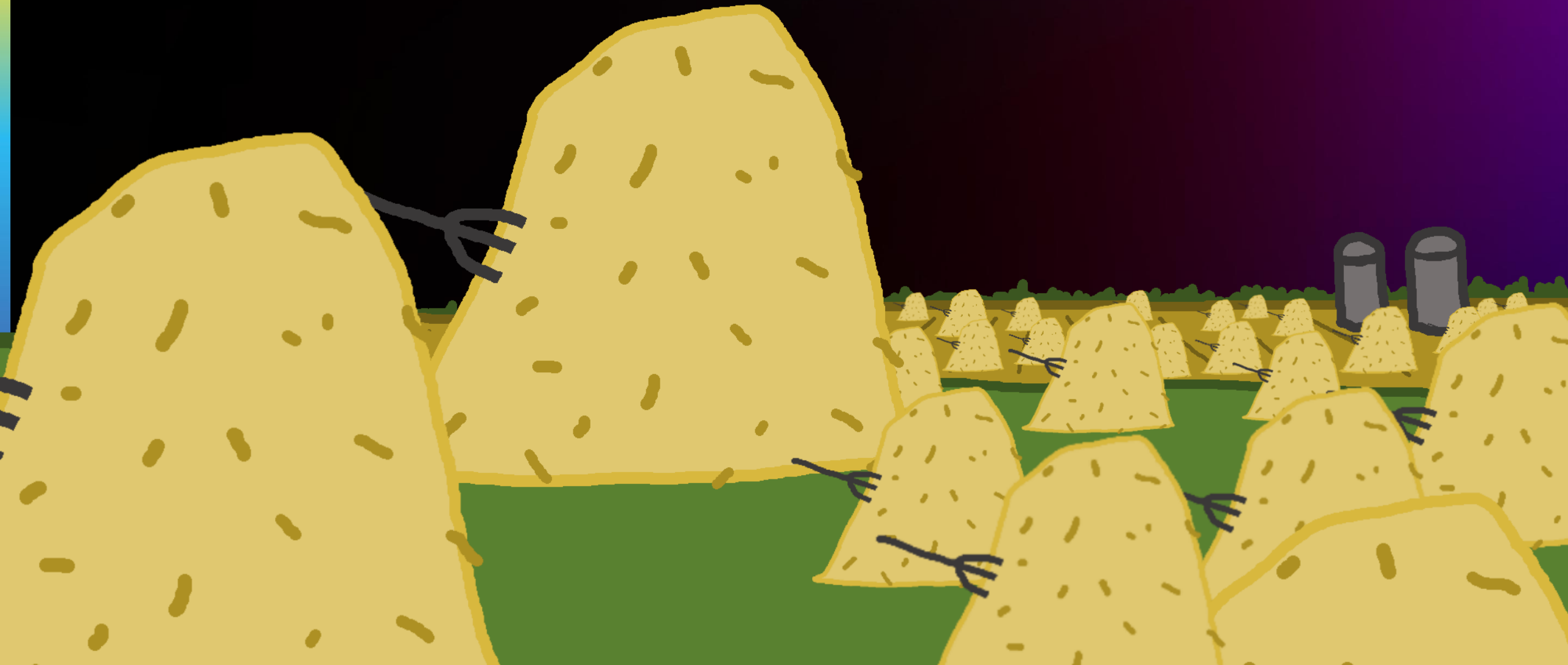


Finding the root cause

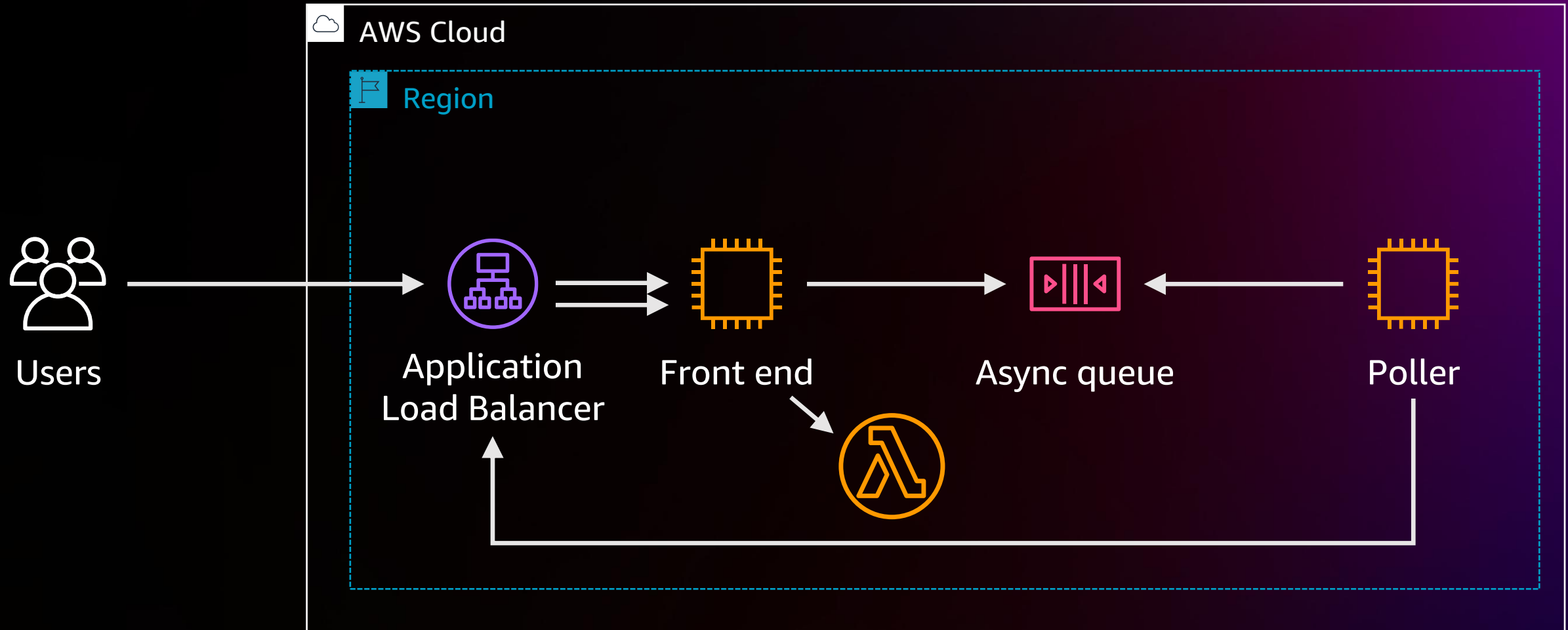
- **Propagate trace information**
- Build lots of dashboards
- Get multi-dimensional
- Use high cardinality metrics
- Dig even deeper with log analysis
- Look at the raw logs
- Analyze using profiler



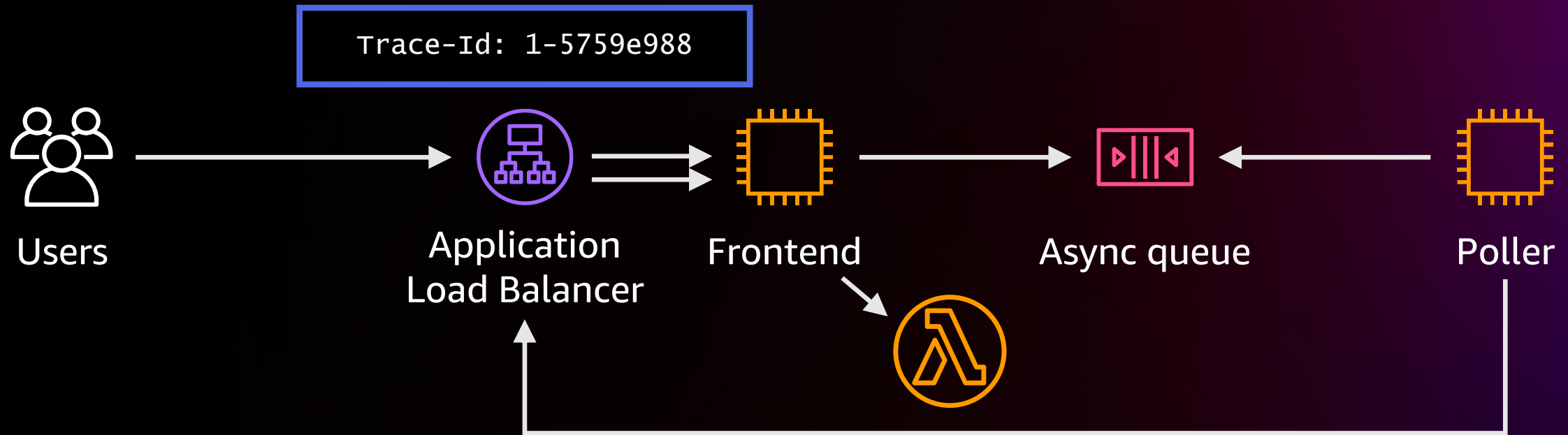
Finding the right haystack with traces



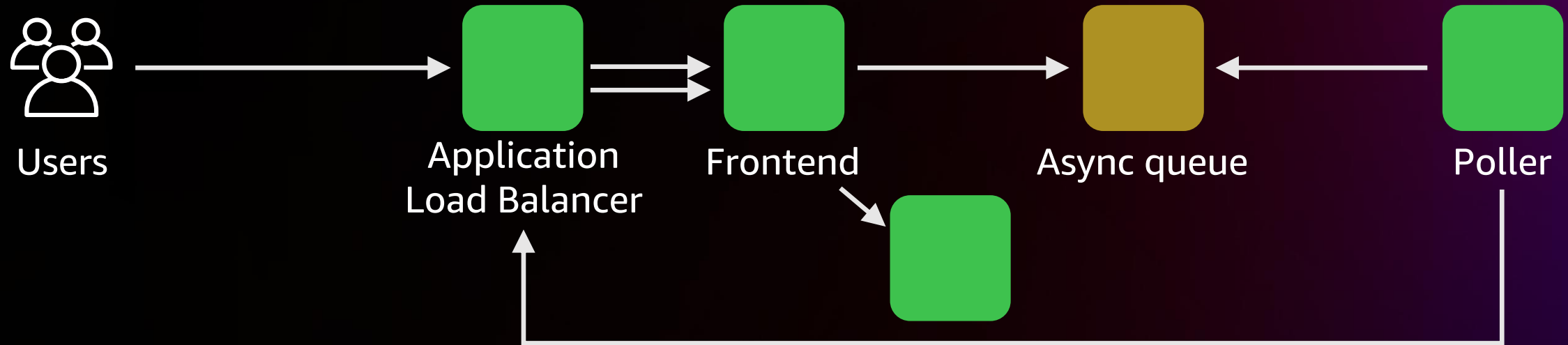
Reasoning about distributed systems

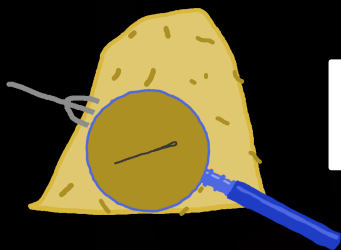


Distributed tracing



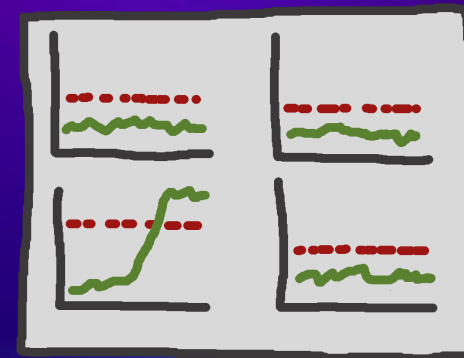
Distributed system-wide visibility



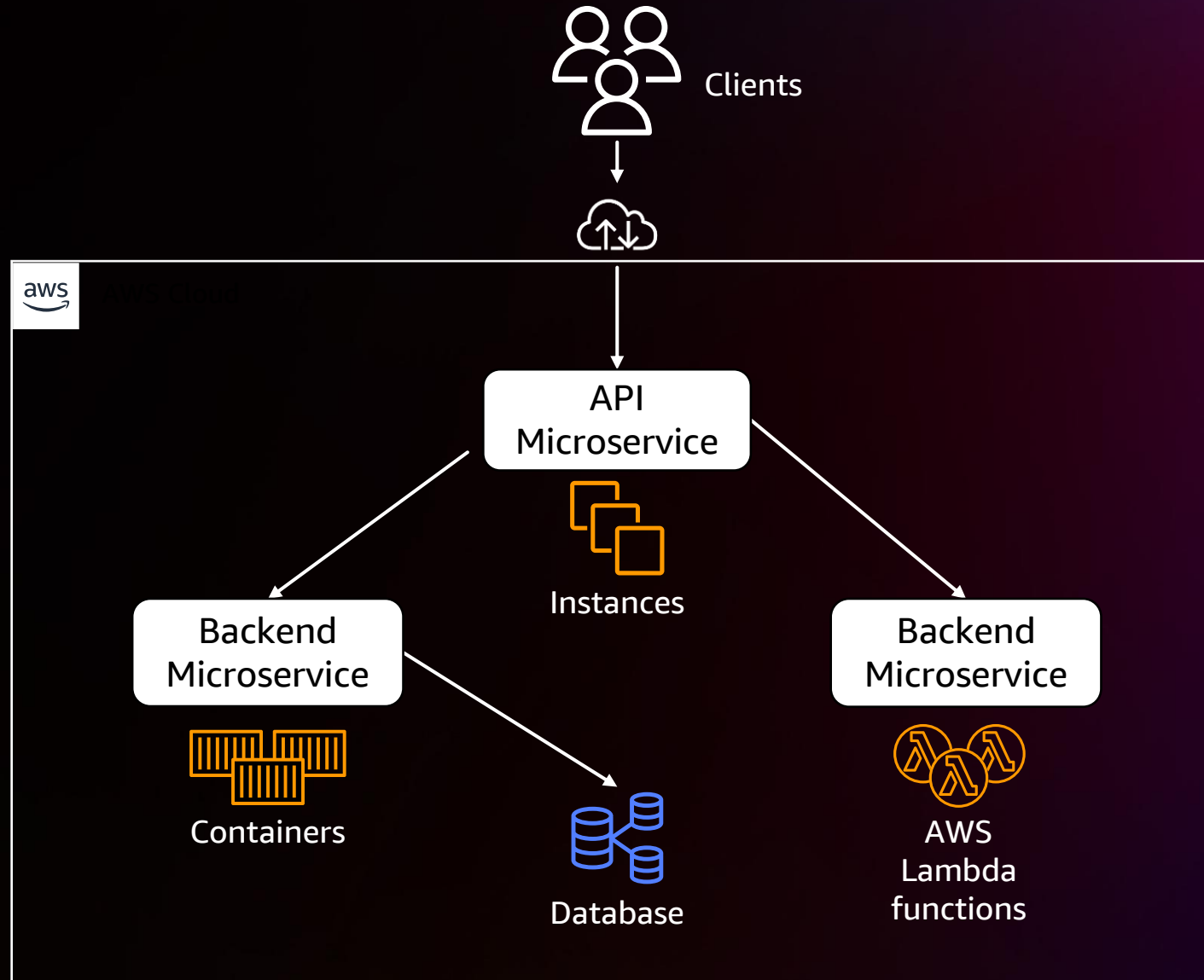


Finding the root cause

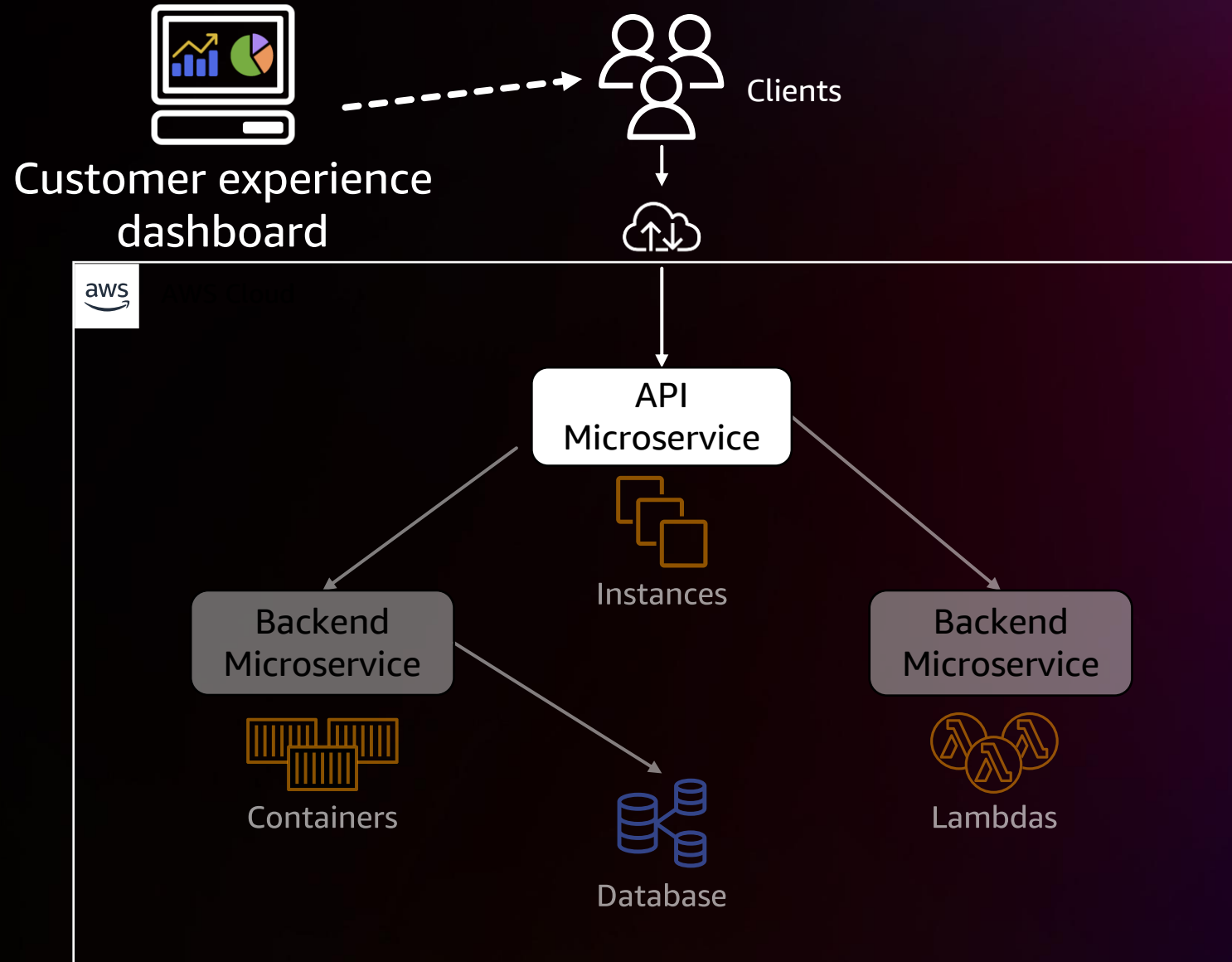
- Propagate trace information
- **Build lots of dashboards**
- Get multi-dimensional
- Use high cardinality metrics
- Dig even deeper with log analysis
- Look at the raw logs
- Analyze using profiler



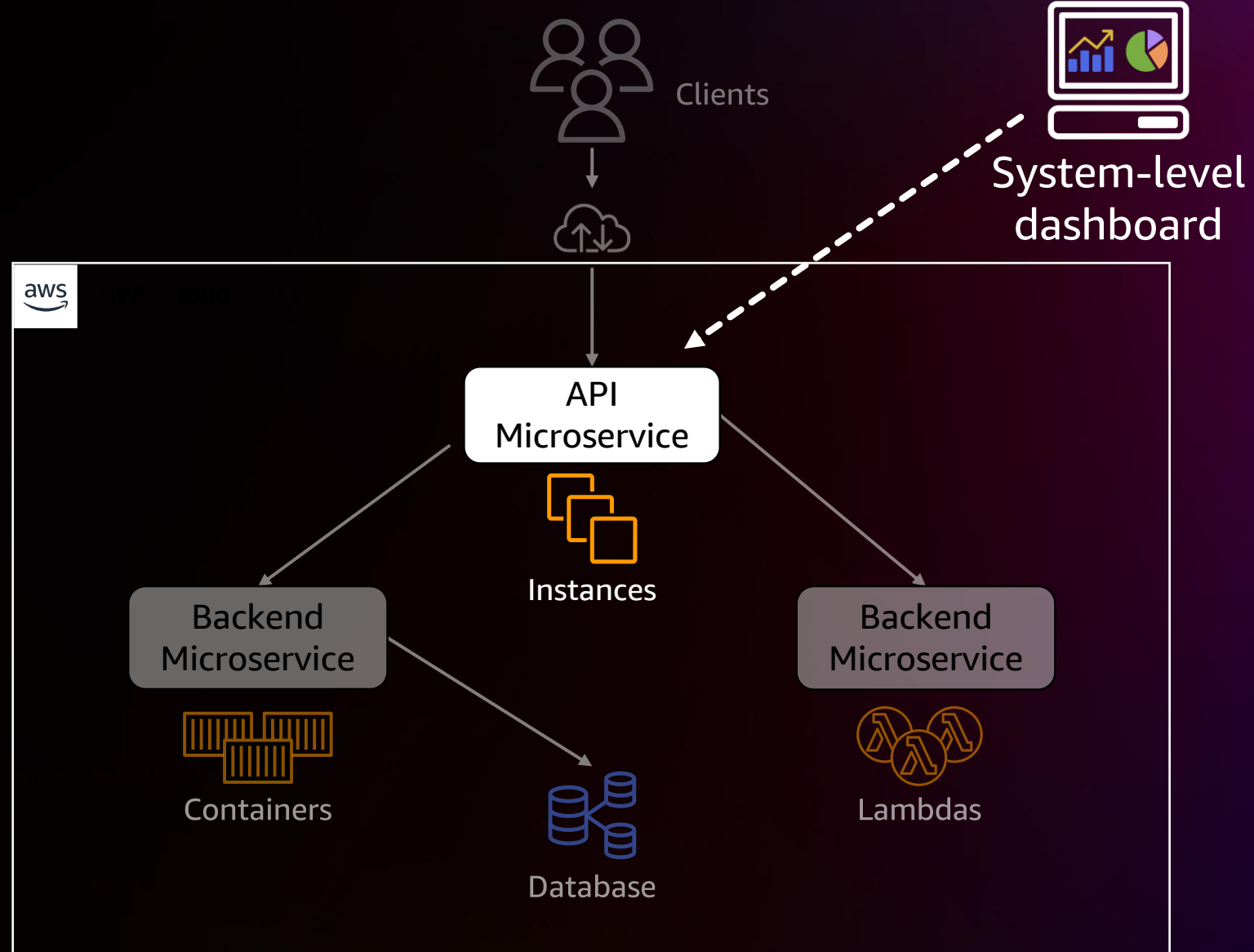
Hierarchy of dashboards



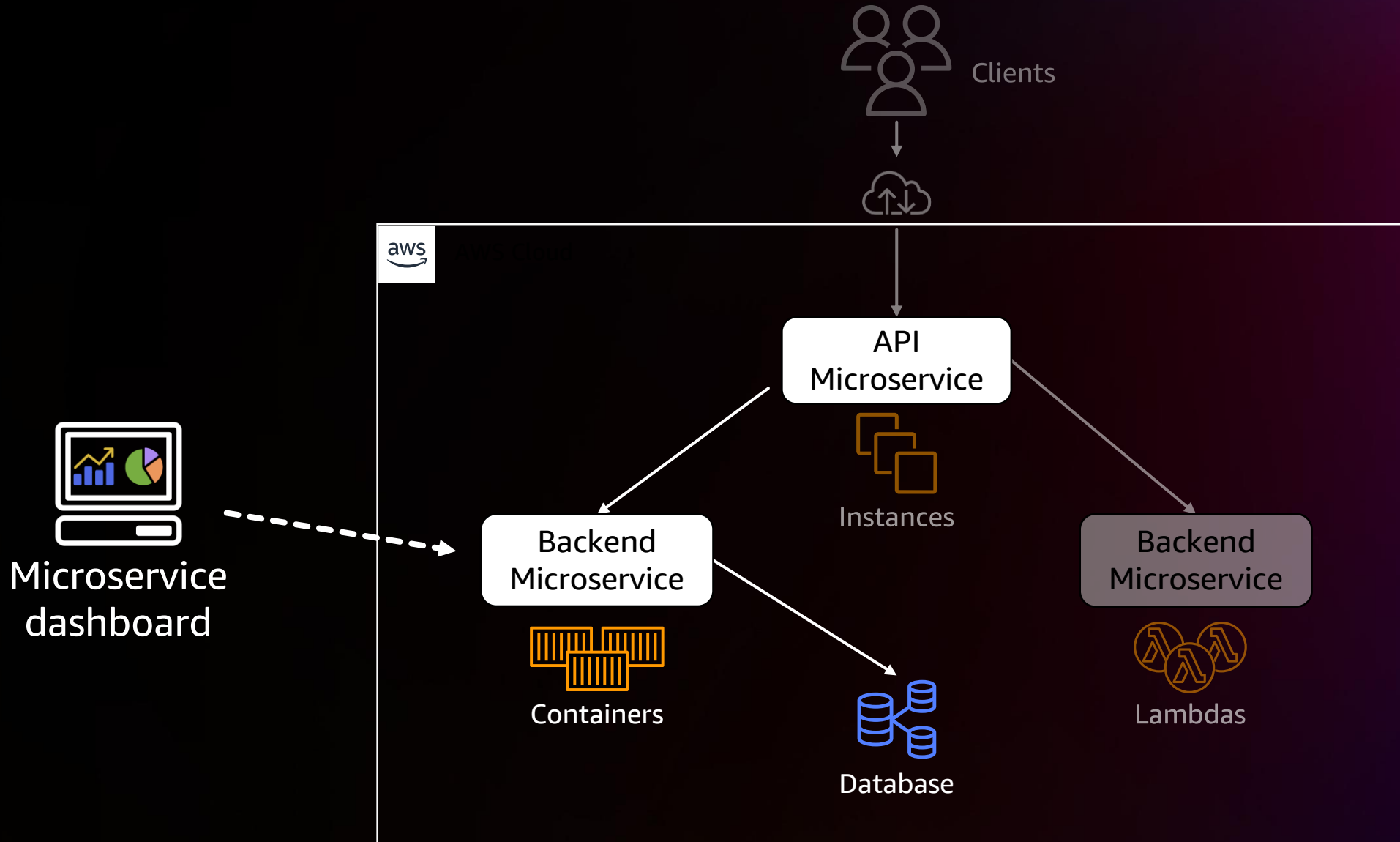
Customer experience dashboards



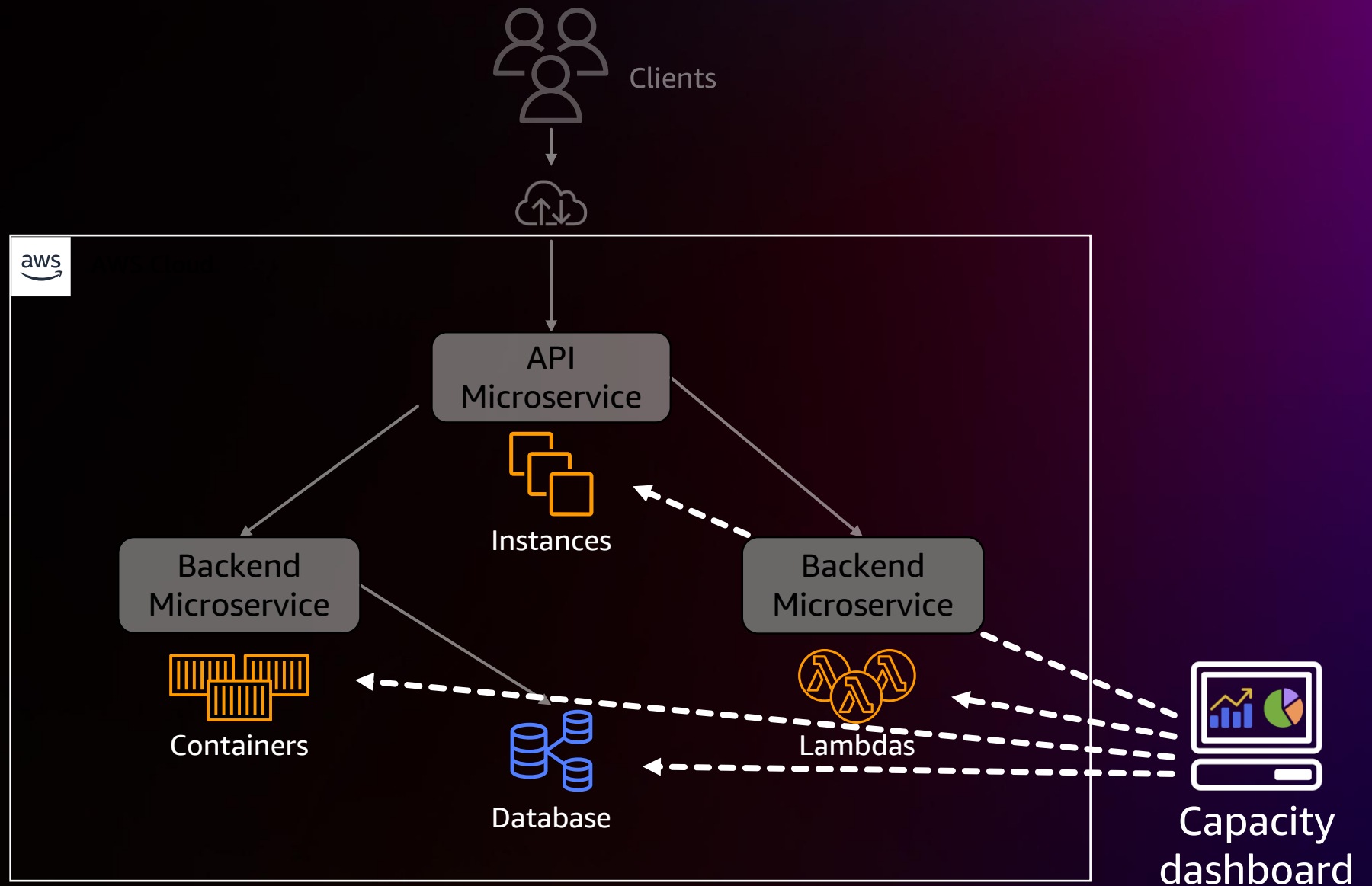
System-level dashboards



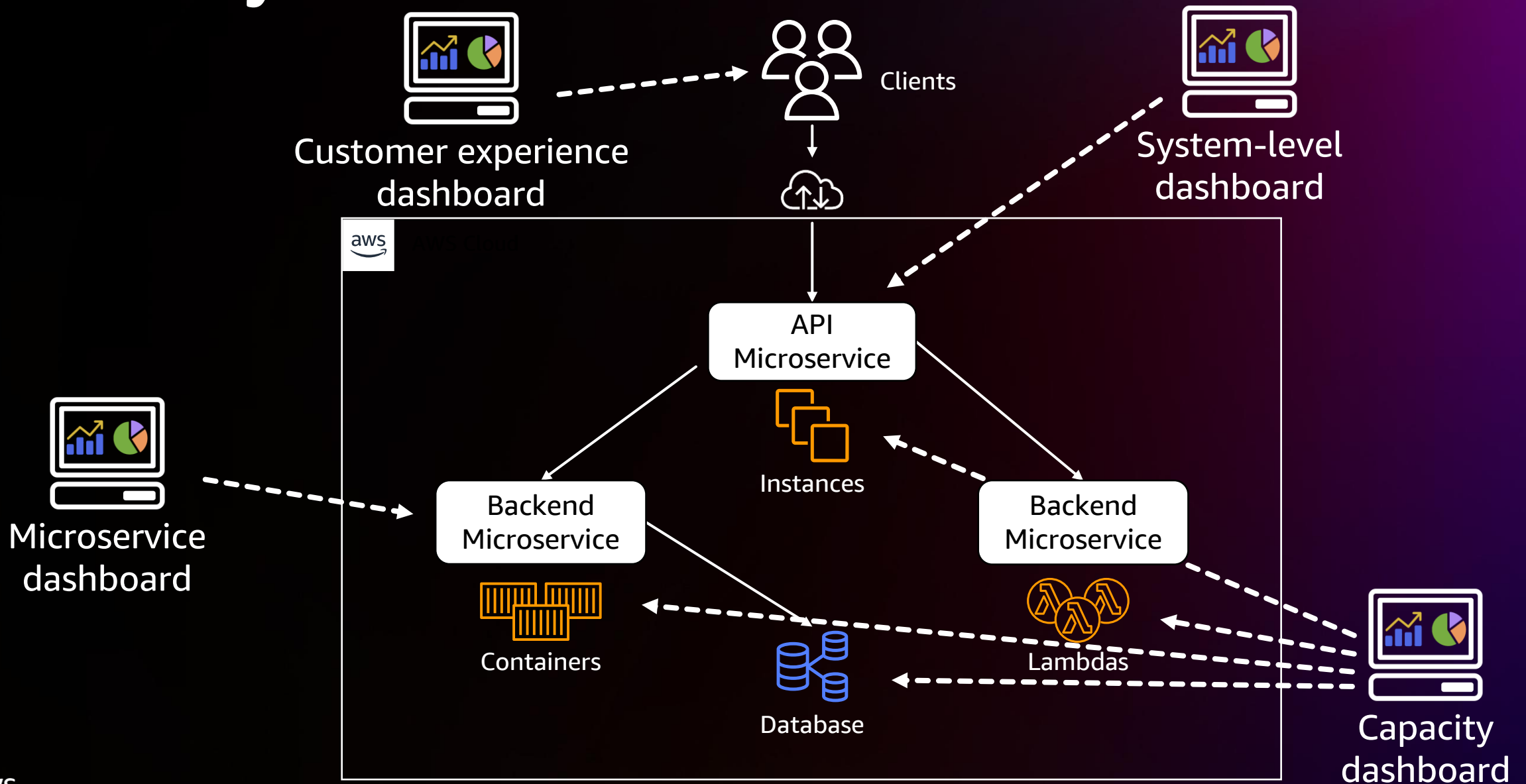
Microservice dashboards



Capacity dashboards



Hierarchy of dashboards



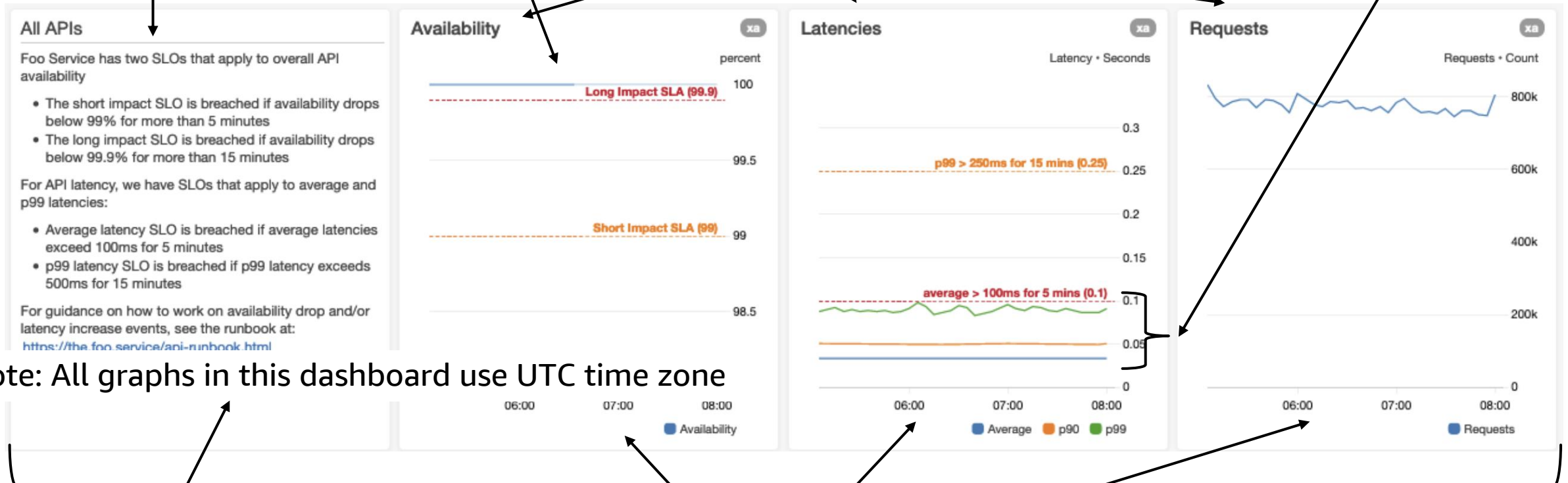
Dashboard design

Include descriptions

Avoid clutter

Key metrics at the top

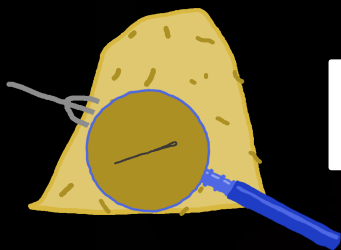
Narrow enough range



The dashboard is never done

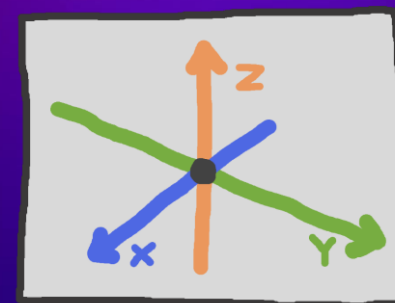


<https://aws.amazon.com/builders-library/building-dashboards-for-operational-visibility/>

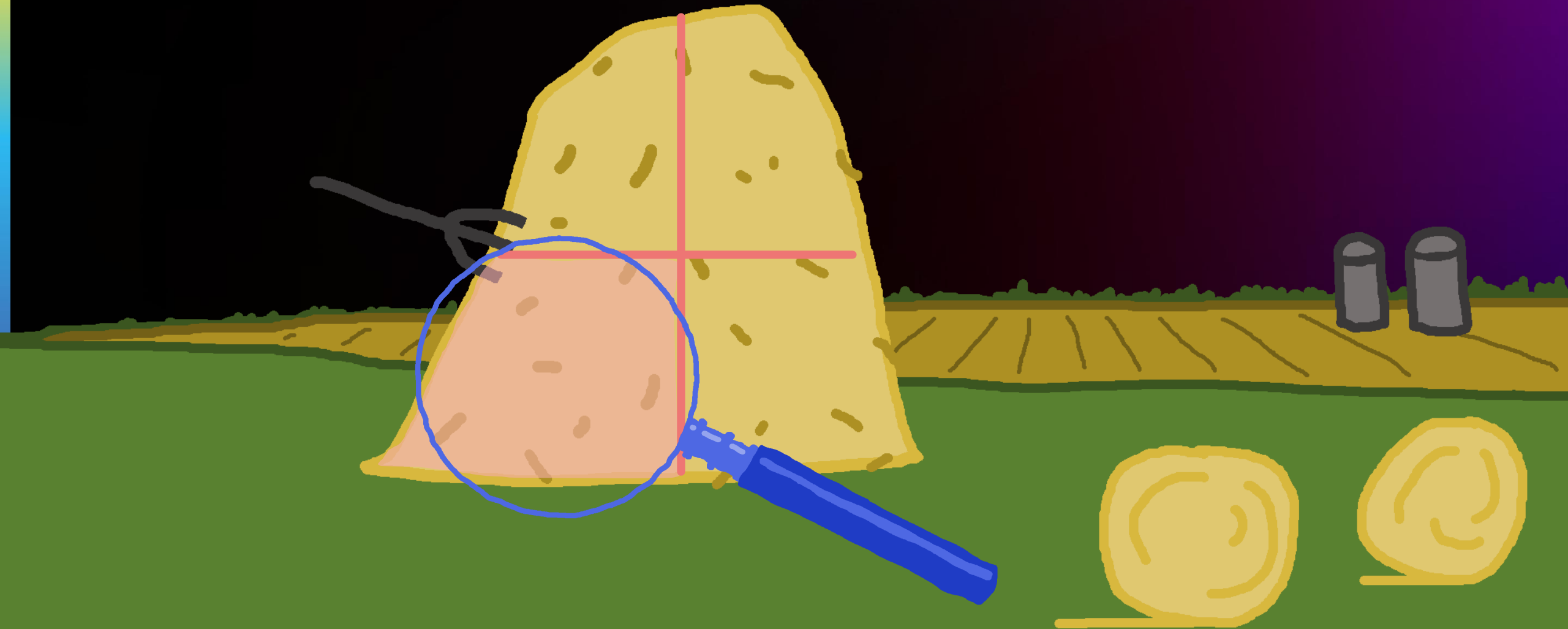


Finding the root cause

- Propagate trace information
- Build lots of dashboards
- **Get multi-dimensional**
- Use high cardinality metrics
- Dig even deeper with log analysis
- Look at the raw logs
- Analyze using profiler



Use metric dimensions to drill down



Log everything: Embedded Metrics Format

```
{
  "_aws": {
    "Timestamp": 1574109732004,
    "CloudWatchMetrics": [
      {
        "Namespace": "ProductInfoService",
        "Dimensions": [],
        "Metrics": [
          { "Name": "Time", "Unit": "Milliseconds" }
        ]
      }
    ]
  },
  "CustomerId": "user@example.com",
  "InstanceId": "us-west-2c",
  "Time": 100,
  "TraceId": "Root=1-5759e988-bd862e3fe1be46a994272793;Sampled=1"
}
```

Instrumentation

Configuration

Example: Simple web service code

```
public ProductInfo getProductInfo(String customerId, String productId) {  
    ProductInfo info = cache.get(customerId, productId);  
  
    if (info == null) {  
        info = db.query(customerId, productId);  
    }  
  
    return info;  
}
```

What product or customer was this for?

Was this a cache hit or miss?

If it failed, was it a timeout or did it return an error?

How long did the db query take? Did it return anything?

Example: Instrumenting code

```
public ProductInfo getProductInfo(String customerId, String productId,  
    Metrics metrics) {
```

```
    metrics.addProperty("CustomerId", customerId);  
    metrics.addProperty("ProductId", productId);
```

```
    ProductInfo info = cache.get(customerId, productId, metrics);
```

```
    if (info == null) {  
        info = db.query(customerId, productId, metrics);  
    }
```

```
    return info;
```

```
}
```

Record what we're working on

Propagate per-request
context object

Instrumentation in the log

```
{
```

```
...
```

```
"CustomerId": "user@example.com",  
"ProductId": "BrightOrangeSneakers",  
"ClientIp": "192.168.131.39",
```

Caller and request info

```
"InstanceId": "i-001234a4bf70dec41EXAMPLE",  
"AvailabilityZone": "us-west-2a",  
"Cache.Node": "10.0.2.182",
```

Infrastructure info

```
"Cache.Hit": 0,  
"Cache.Time": 1,  
"DB.Time": 10,  
"DB.Results": 1,  
"Time": 100.
```

Timing and info around dependencies

```
"TraceId": "Root=1-5759e988-bd862e3fe1be46a994272793;Sampled=1"
```

```
}
```

Log everything about a unit of work

{

...

```
"CustomerId": "user@example.com",  
"ProductId": "BrightOrangeSneakers",  
"ClientIp": "192.168.131.39",  
"InstanceId": "i-001234a4bf70dec41EXAMPLE",  
"AvailabilityZone": "us-west-2a",  
"Cache.Node": "10.0.2.182",
```

```
"Cache.Hit": 0,  
"Cache.Time": 1,  
"DB.Time": 10,  
"DB.Results": 1,  
"Time": 100,
```

Measurements

Attributes

```
"TraceId": "Root=1-5759e988-bd862e3fe1be46a994272793;Sampled=1"
```

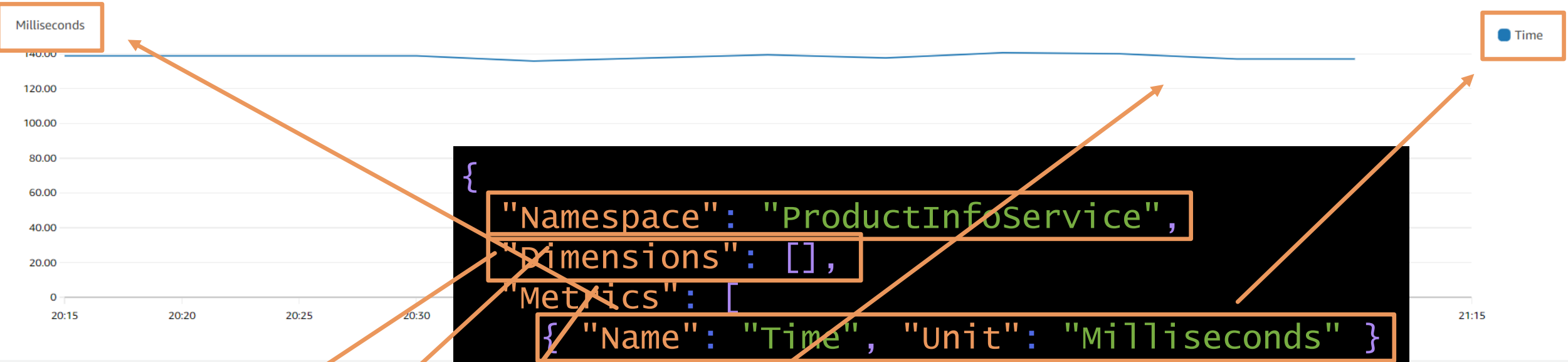
}

Log everything: Embedded Metrics Format

```
{
  "_aws": {
    "Timestamp": 1574109732004,
    "CloudWatchMetrics": [
      {
        "Namespace": "ProductInfoService",
        "Dimensions": [],
        "Metrics": [
          { "Name": "Time", "Unit": "Milliseconds" }
        ]
      }
    ]
  },
  "CustomerId": "user@example.com",
  "AvailabilityZone": "us-west-2c",
  "Time": 100,
  "TraceId": "Root=1-5759e988-bd862e3fe1be46a994272793;Sampled=1"
}
```

Metric definitions

Defines which data to turn into a metric

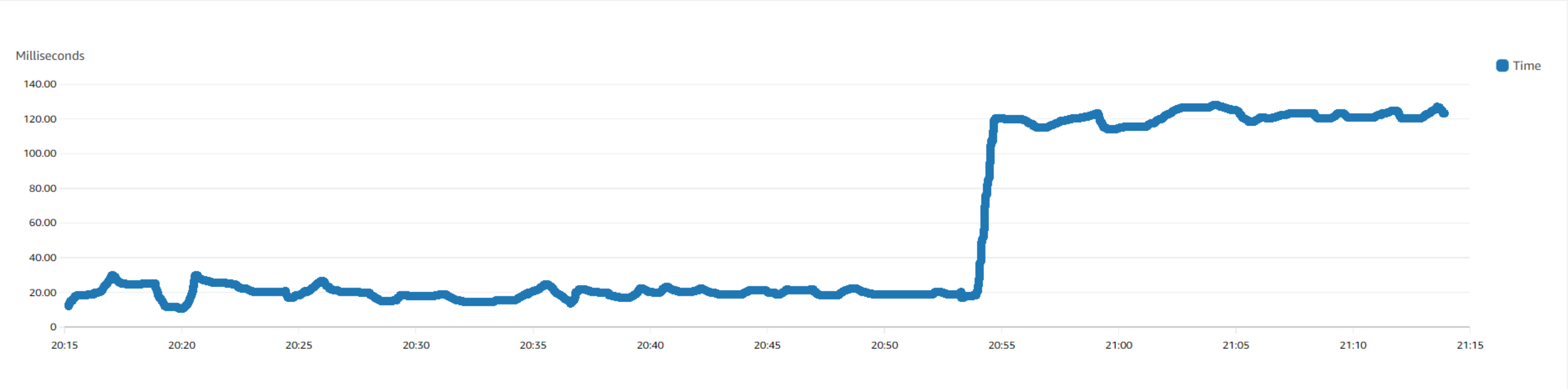


Browse Query Graphed metrics (1) Options Add query

Metrics (1) Info Oregon All > ProductInfoService > Metrics with no dimensions Search for any metric, dimension or resource id < 1 >

ProductInfoService

<input checked="" type="checkbox"/>	Metric name 1/1
<input checked="" type="checkbox"/>	Time



BrowseQueryGraphed metrics (1)OptionsSource

Add math▼Add query▼

Metrics (1) [Info](#)

Oregon▼All > ProductInfoService > Metrics with no dimensions

🔍 Search for any metric, dimension or resource id

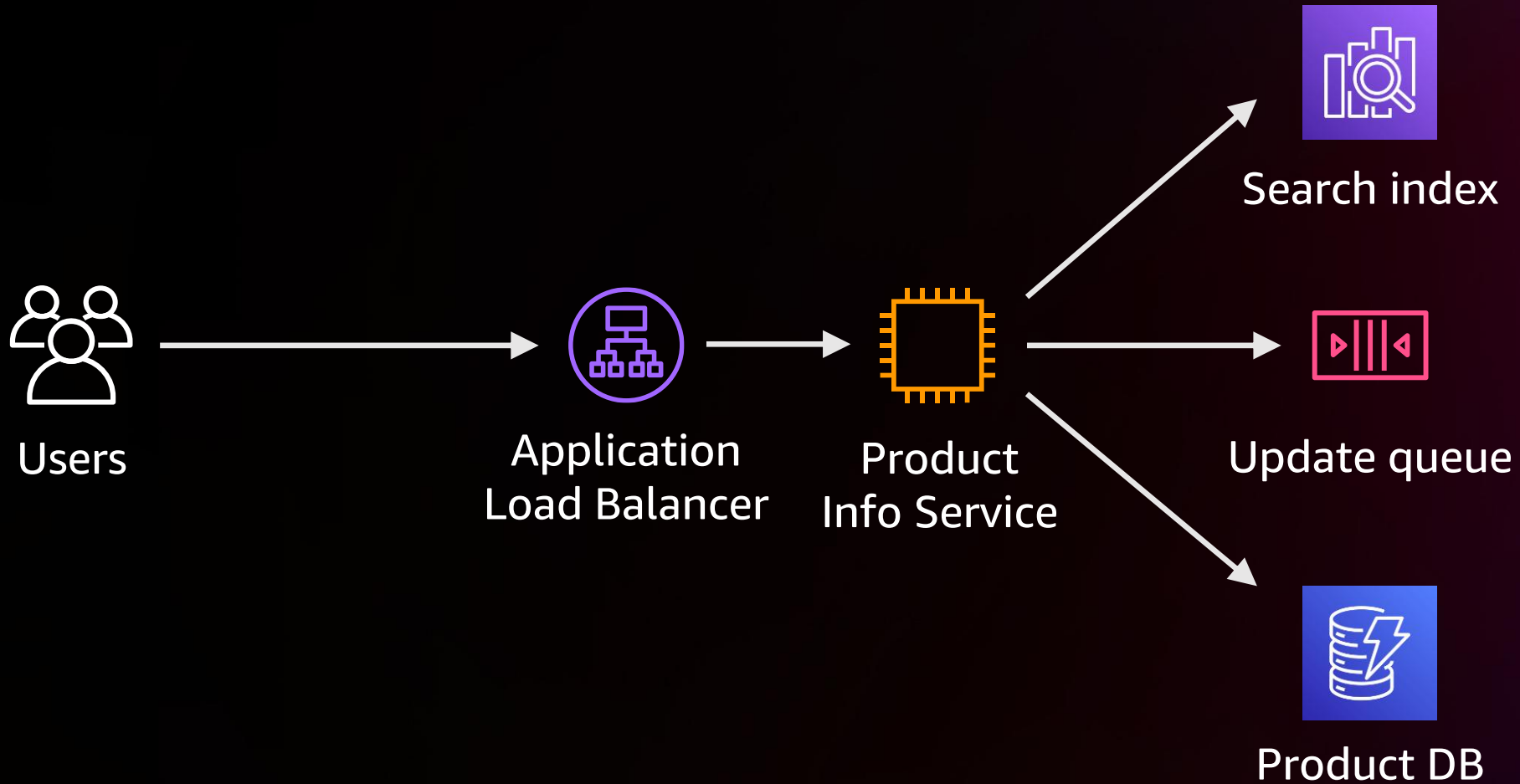
< 1 >

ProductInfoService [🔗](#) [✕](#)

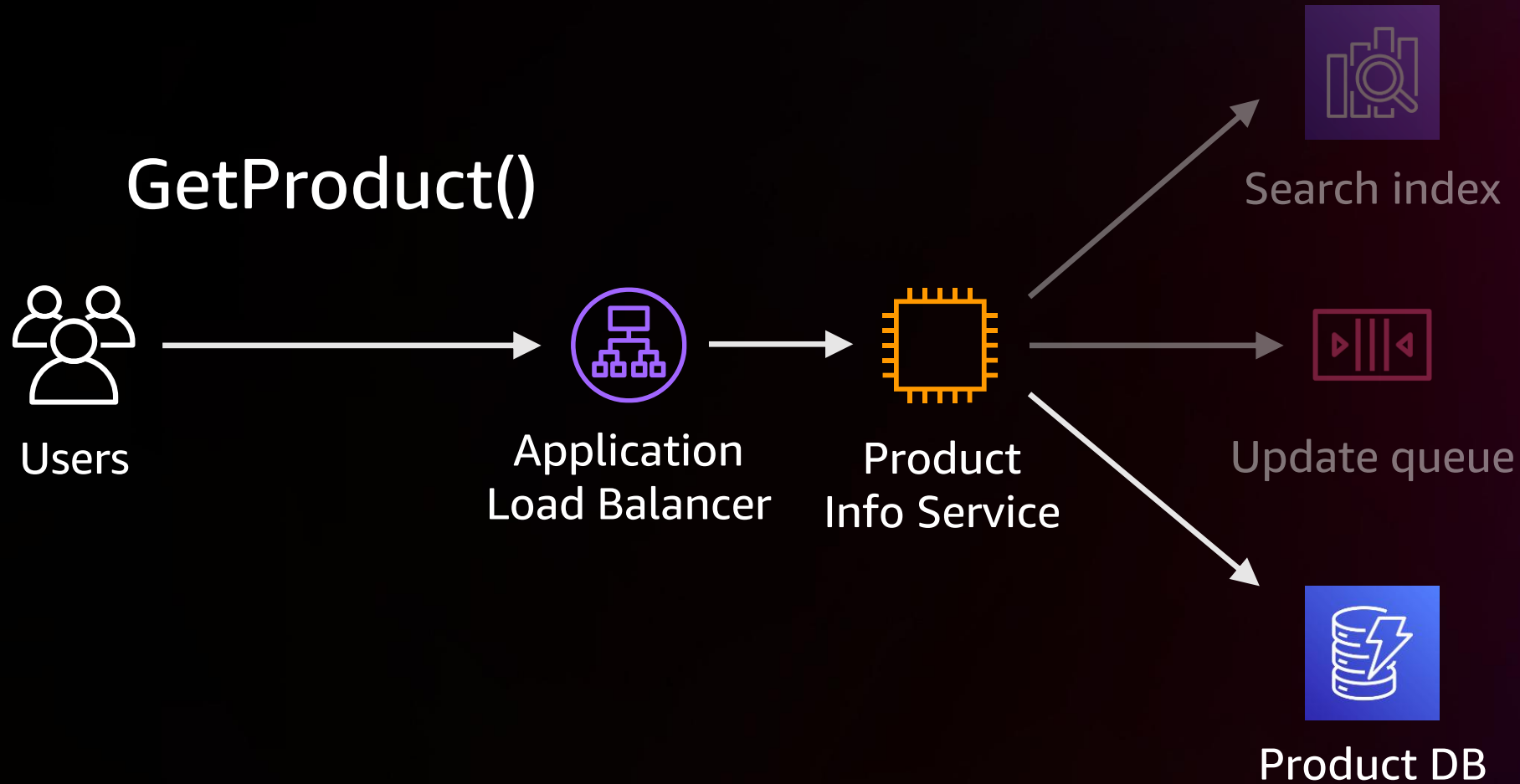
⚙️

<input checked="" type="checkbox"/>	Metric name 1/1	▲
<input checked="" type="checkbox"/>	Time	

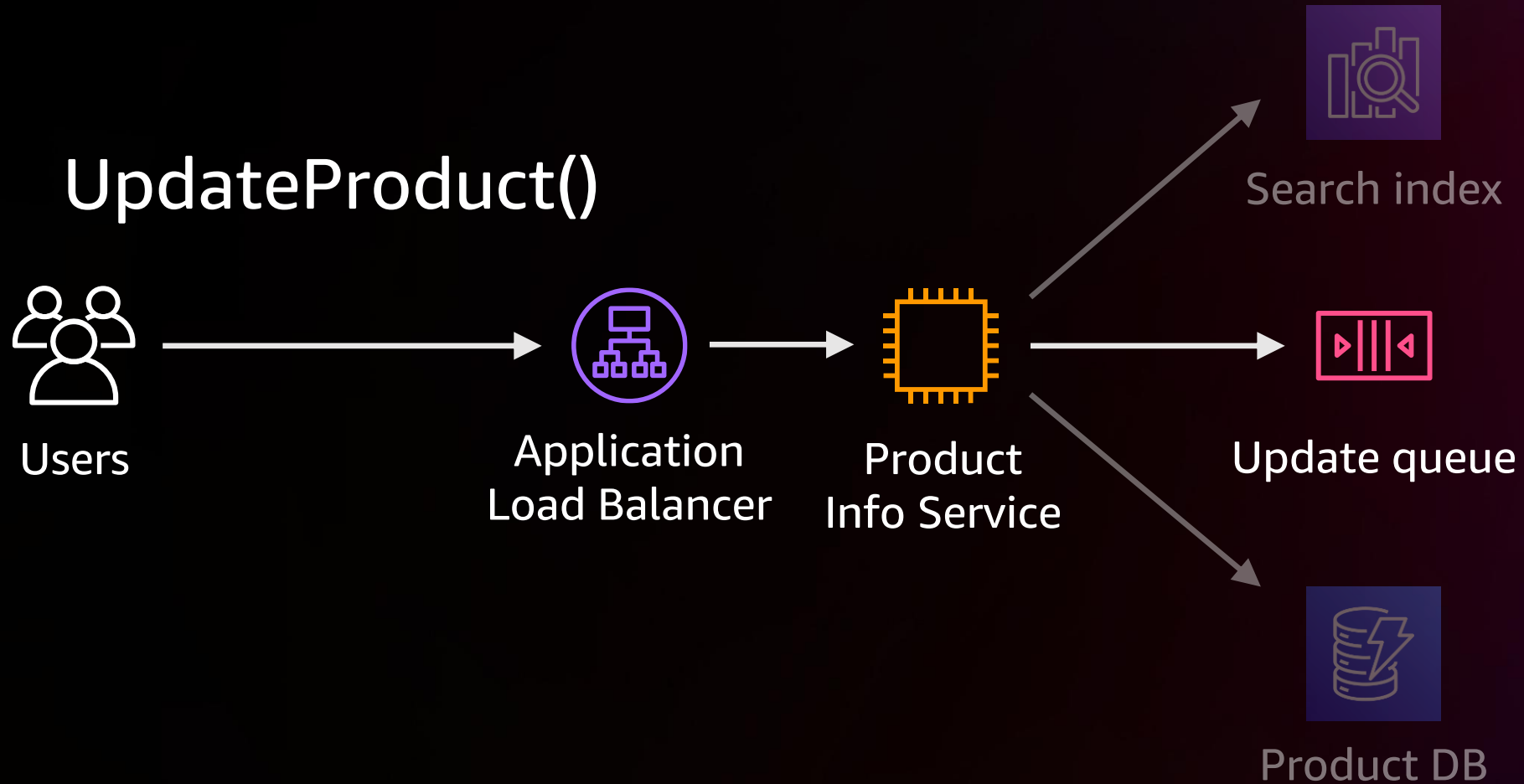
Product Info Service architecture



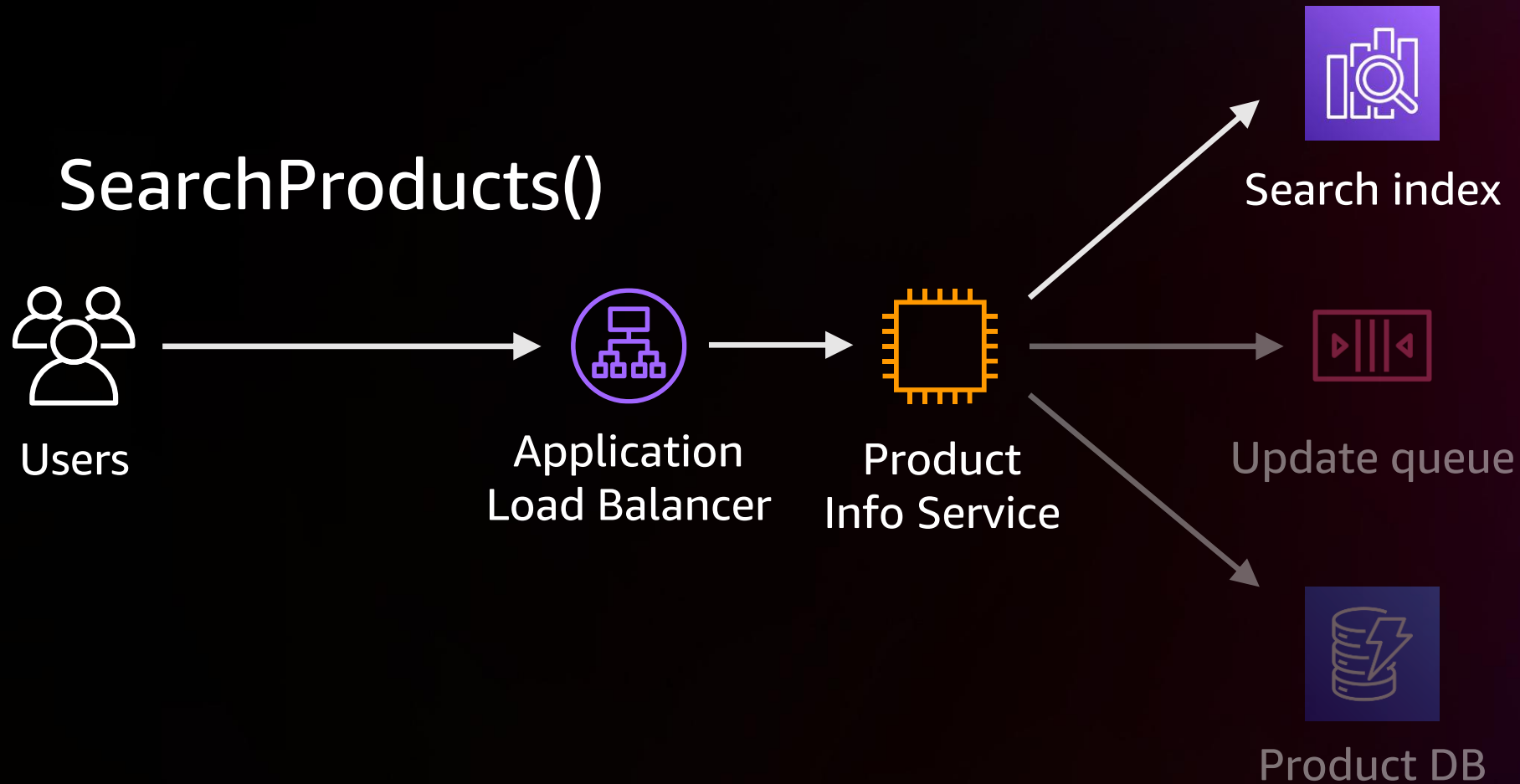
Product Info Service architecture



Product Info Service architecture



Product Info Service architecture



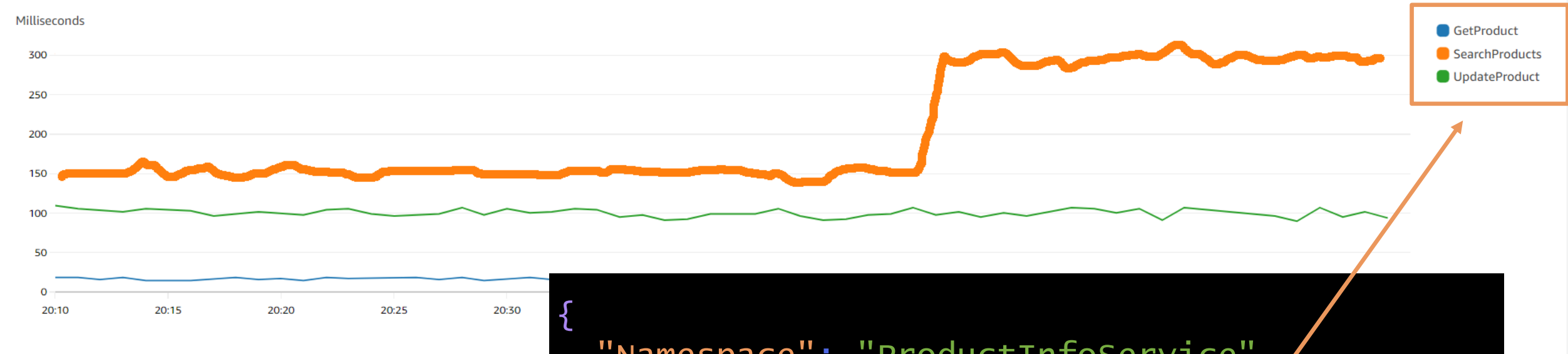
ProductInfoService Latency Per Operation

1h3h12h1d3d1wCustom

Line

Actions

1m



BrowseQueryGraphed metrics (1)OptionsSource

Metrics (3) Info

Oregon All > ProductInfoService > Operation

ProductInfoService

Operation 3/3

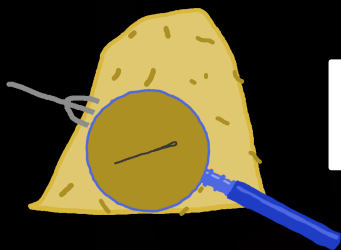
GetProduct

SearchProducts

UpdateProduct

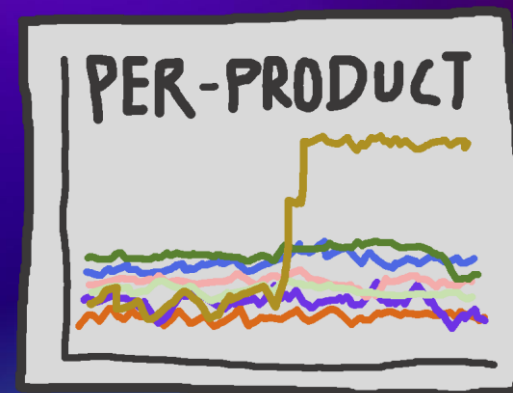
```
{  
  "Namespace": "ProductInfoService",  
  "Dimensions": [ ["Operation"] ],  
  "Metrics": [  
    { "Name": "Time", "Unit": "Milliseconds" }  
  ]  
}
```

```
...  
"Operation": "SearchProducts",  
"Time": 305,  
...
```



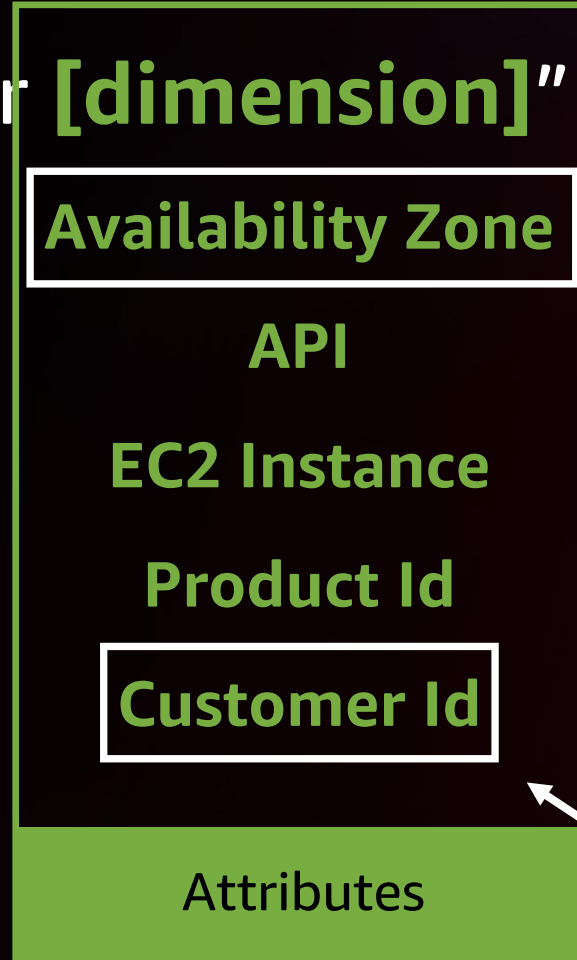
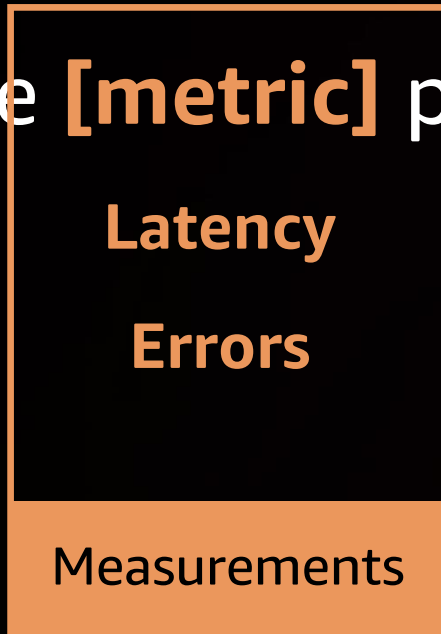
Finding the root cause

- Propagate trace information
- Build lots of dashboards
- Get multi-dimensional
- **Use high cardinality metrics**
- Dig even deeper with log analysis
- Look at the raw logs
- Analyze using profiler



Dimensionality

“Show me **[metric]** per **[dimension]**”



(Few of these)

Cardinality

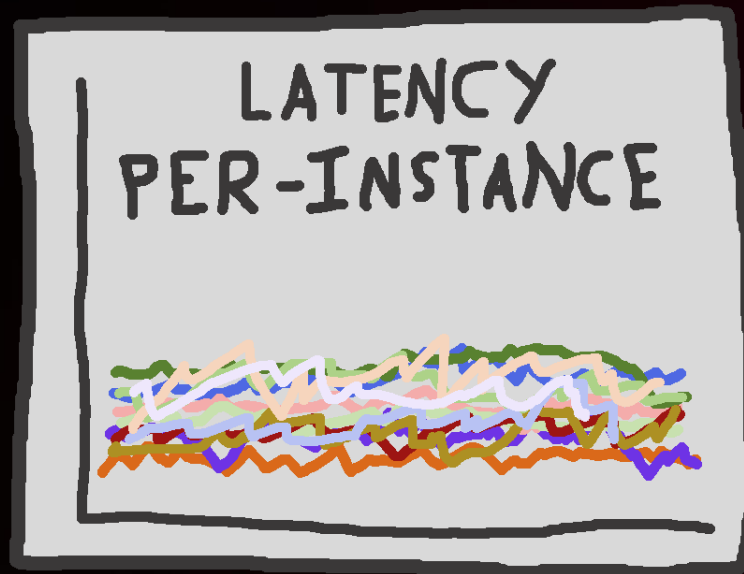
(Many of these)

Analyzing high cardinality metrics

“Are all of my Amazon EC2 instances returning errors?”

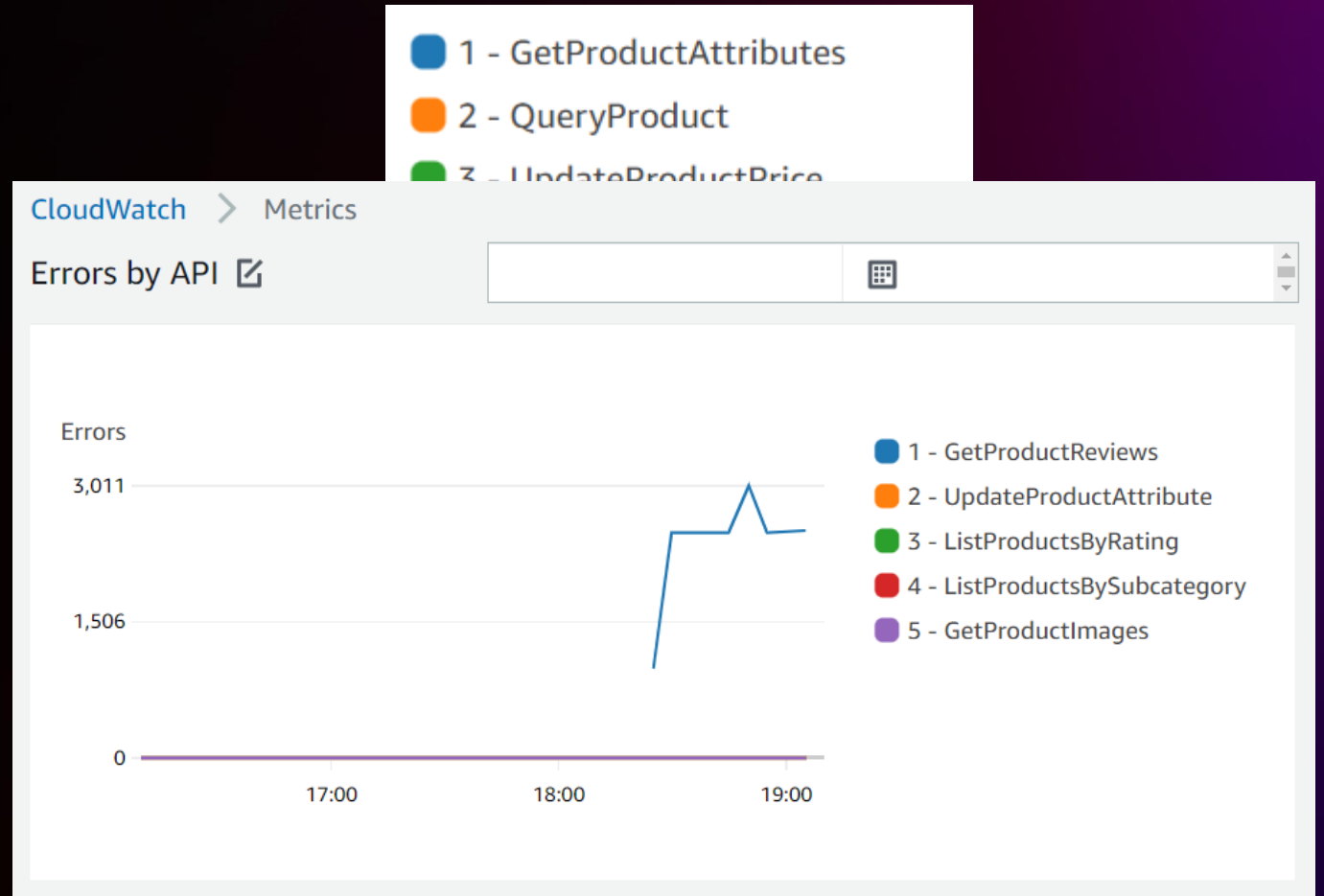
“Are different customer experience different impact?”

“Is just one Product Id having problems?”



CloudWatch Metric Insights

```
SELECT SUM(Errors)
FROM ProductInfoService
GROUP BY Operation
ORDER BY SUM() DESC
LIMIT 5
```



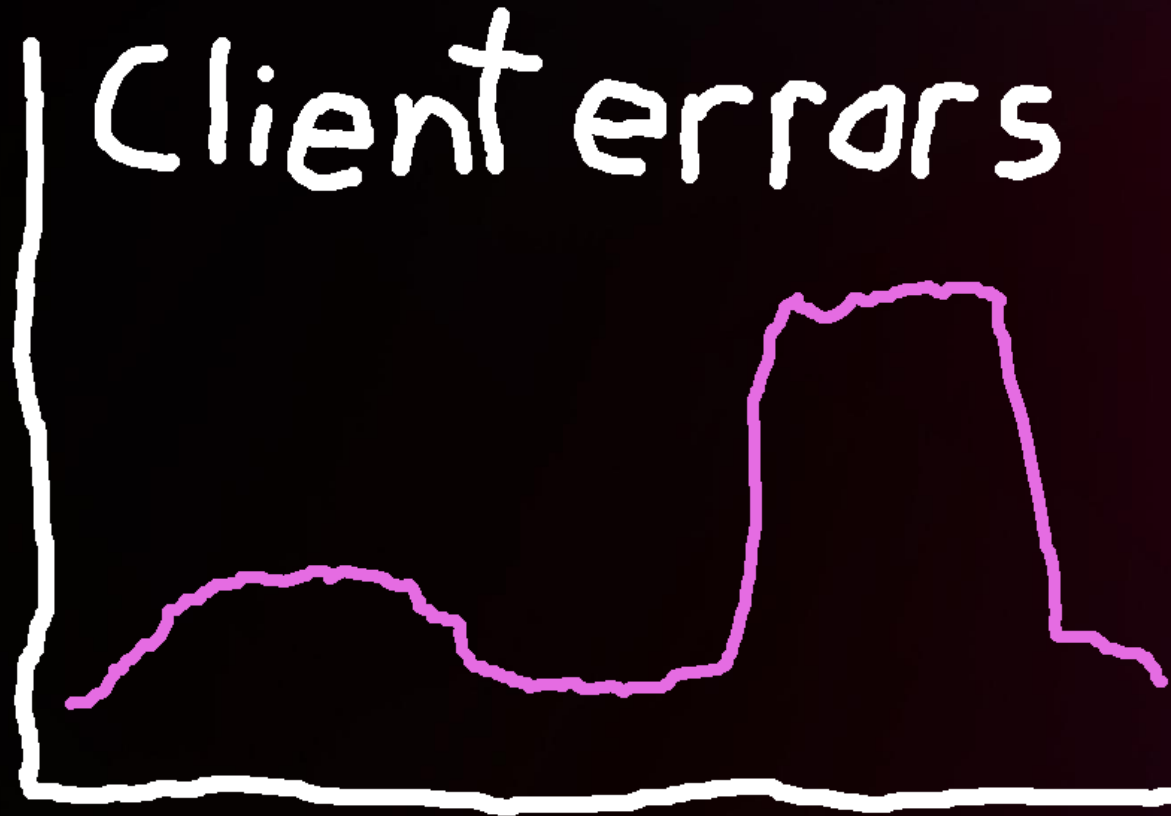
...

(many more)

CloudWatch Contributor Insights



Combining dimensionality and cardinality



Client vs. server fault

Server fault (5XX)

+	-----	+	-----	+
	500		Internal Server Error	
	501		Not Implemented	
	502		Bad Gateway	
	503		Service Unavailable	
	504		Gateway Timeout	
	505		HTTP Version Not Supported	
+	-----	+	-----	+


Client fault (4XX)

+	-----	+	-----	+
	400		Bad Request	
	401		Unauthorized	
	402		Payment Required	
	403		Forbidden	
	404		Not Found	
	405		Method Not Allowed	
	406		Not Acceptable	
	407		Proxy Authentication Required	
	408		Request Timeout	
	409		Conflict	
	410		Gone	
	411		Length Required	
	412		Precondition Failed	
	413		Payload Too Large	
	414		URI Too Long	
	415		Unsupported Media Type	
	416		Range Not Satisfiable	
	417		Expectation Failed	
	426		Upgrade Required	
+	-----	+	-----	+

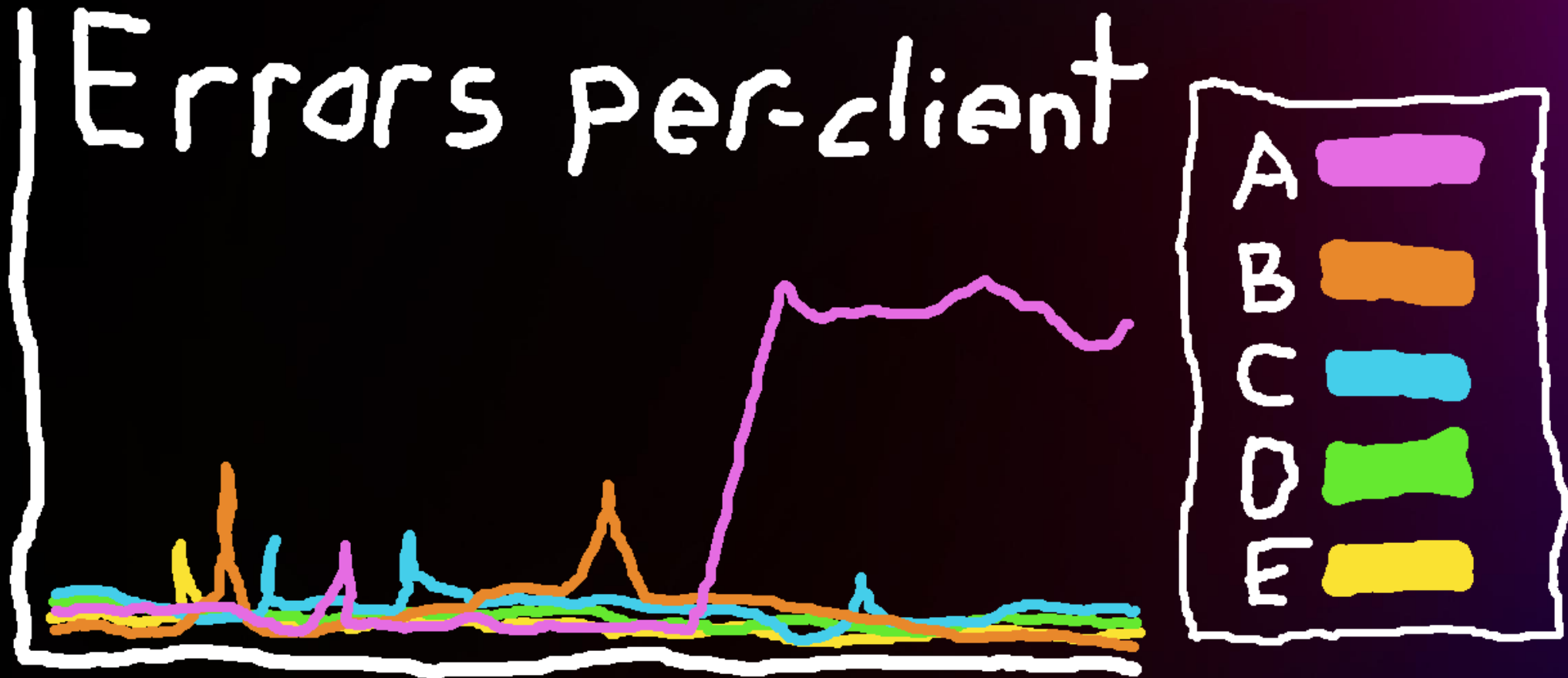
Input validation bugs

(Bug: Making input validation **more** strict)

```
public void validate() {  
-   if(input.length() > 140) {  
+   if(input.length() > 100) {  
        throw new ClientError();  
    }  
}
```



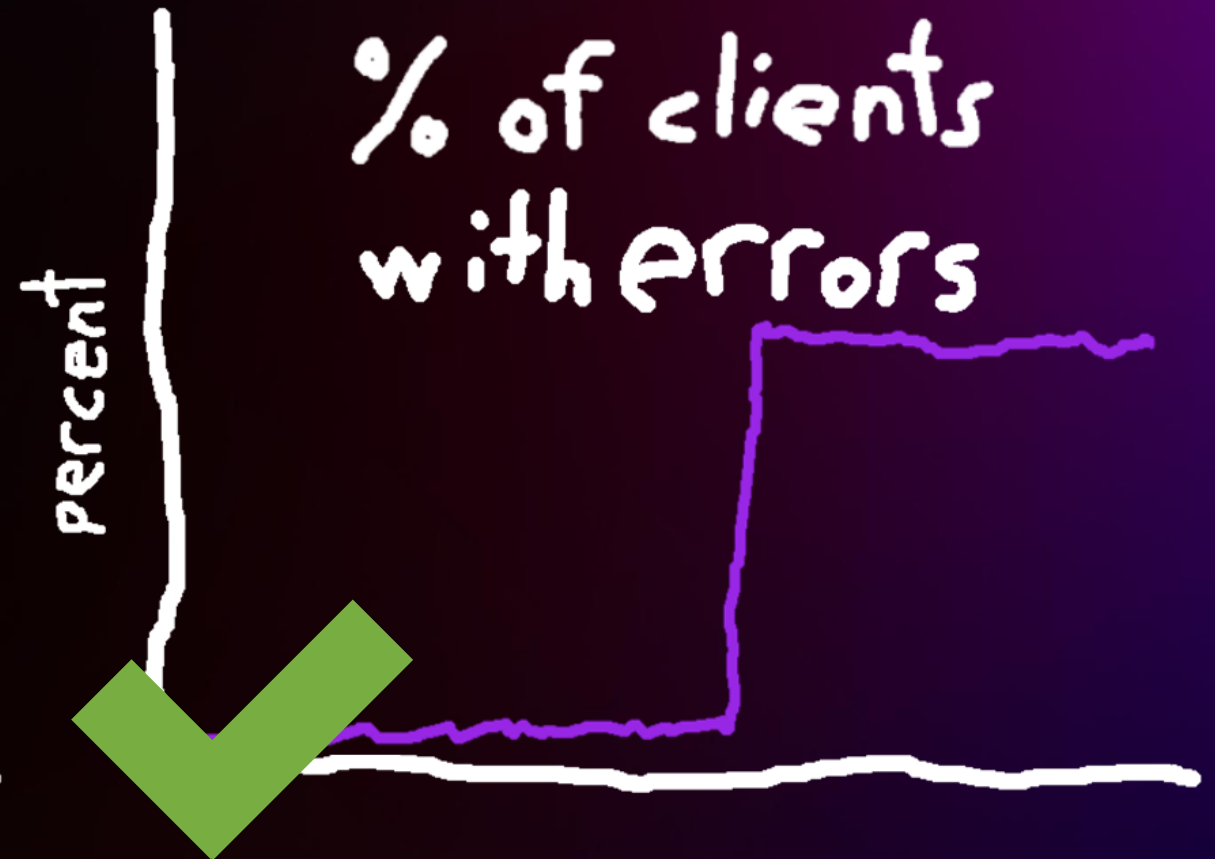
Graphing top contributors



Graphing top contributors



Metrics from another dimension



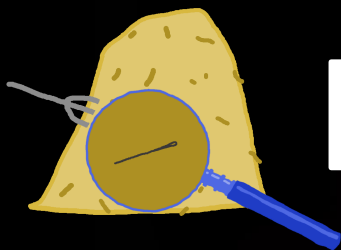
Dimensionality and cardinality together

$$\frac{\# \text{ of customers with errors}}{\# \text{ of customers}}$$

$$\frac{\text{INSIGHT_RULE_METRIC}(\text{ErrorsPerAccount}, \text{UniqueContributors})}{\text{INSIGHT_RULE_METRIC}(\text{RequestsPerAccount}, \text{UniqueContributors})}$$

Metrics from another dimension



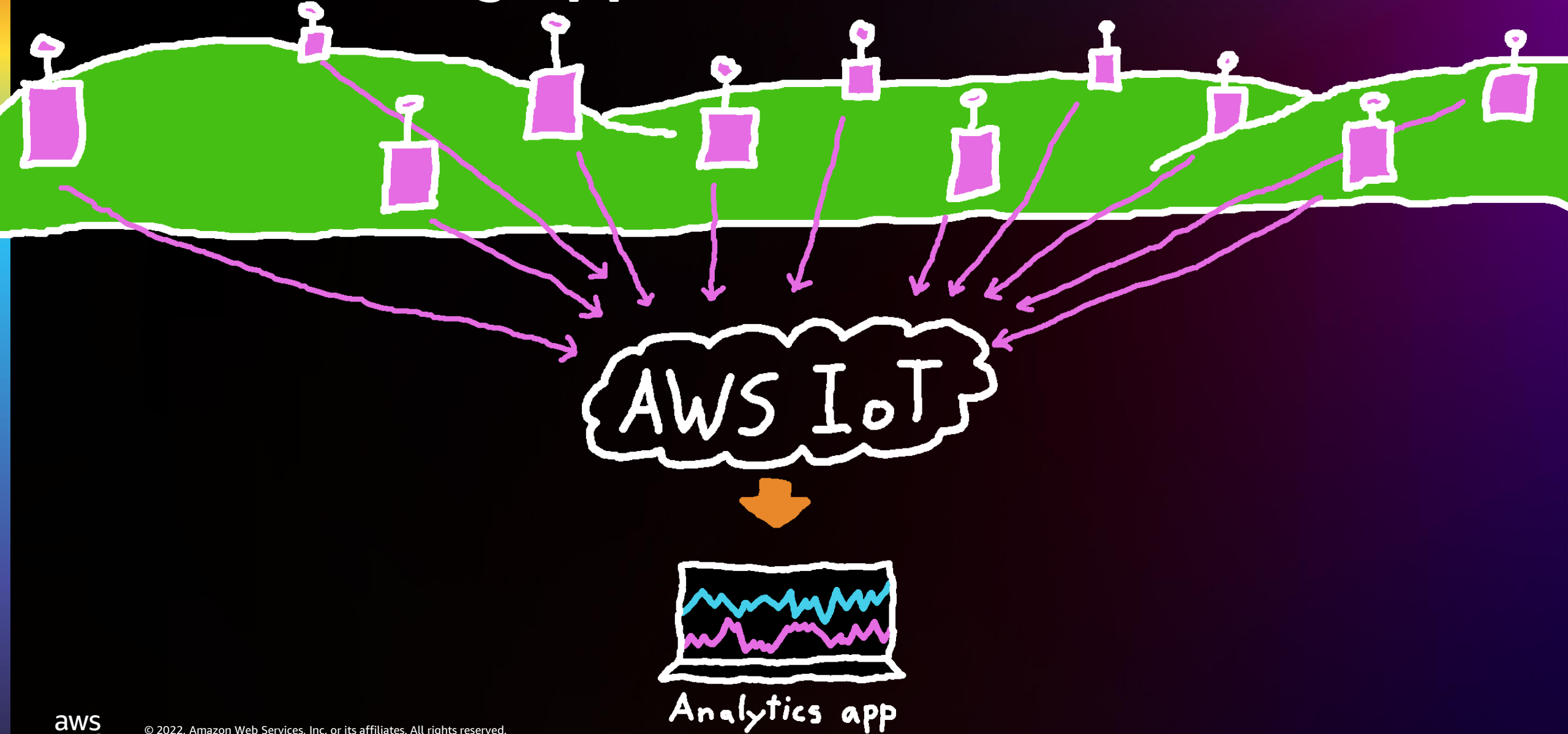


Finding the root cause

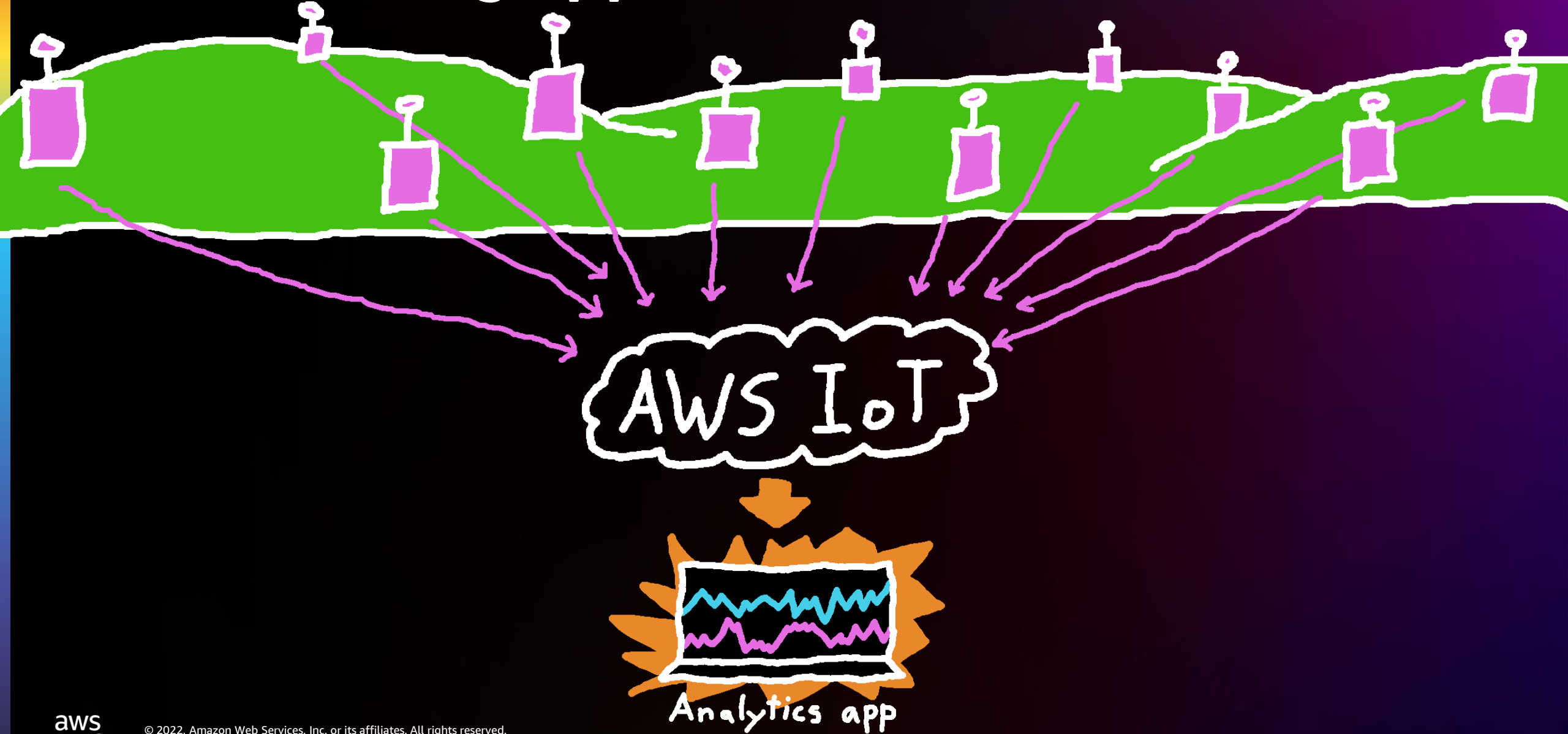
- Propagate trace information
- Build lots of dashboards
- Get multi-dimensional
- Use high cardinality metrics
- **Dig even deeper with log analysis**
- Look at the raw logs
- Analyze using profiler



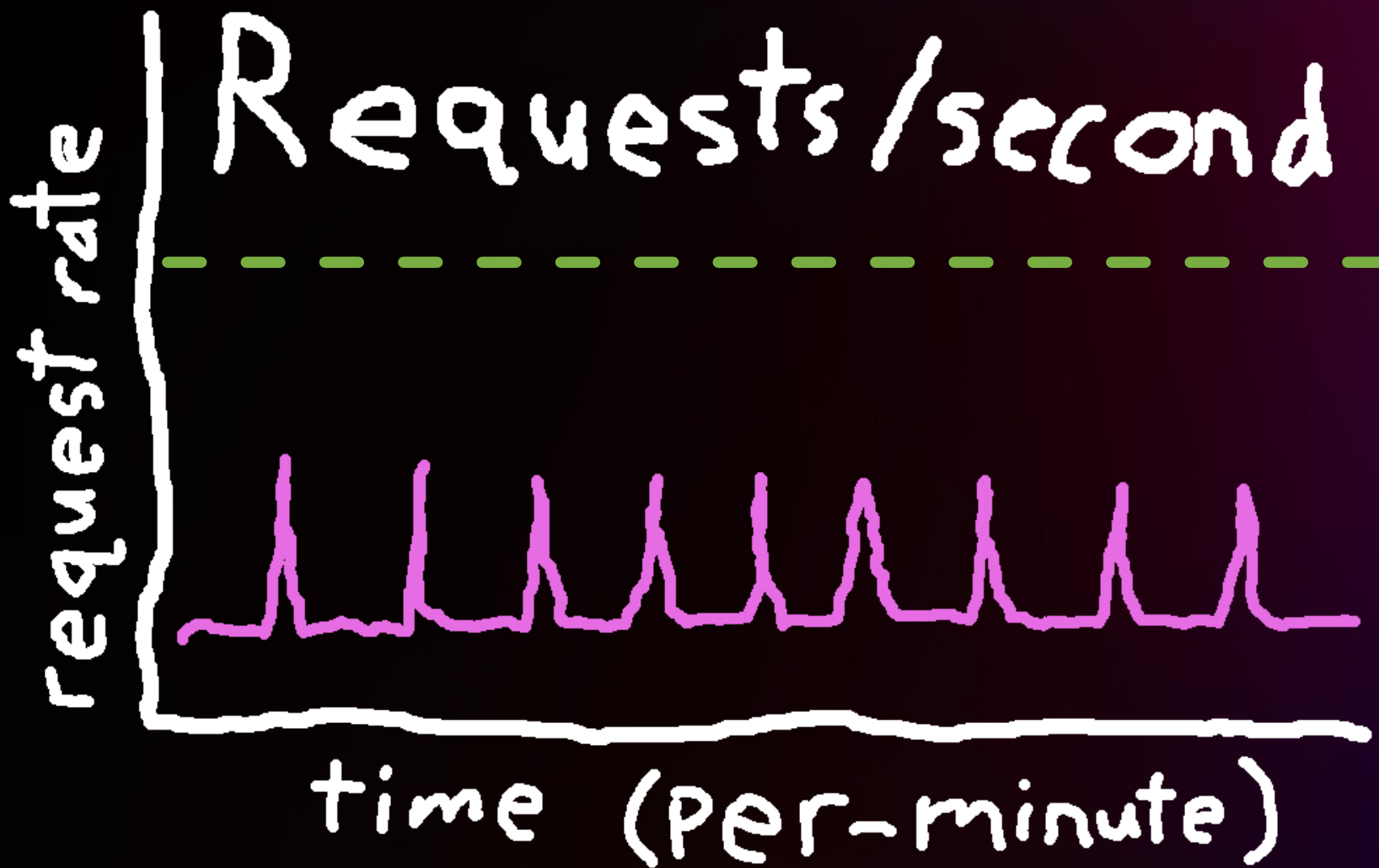
IoT monitoring application



IoT monitoring application



When request rates are a lie

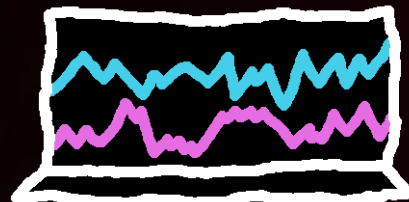


IoT monitoring application



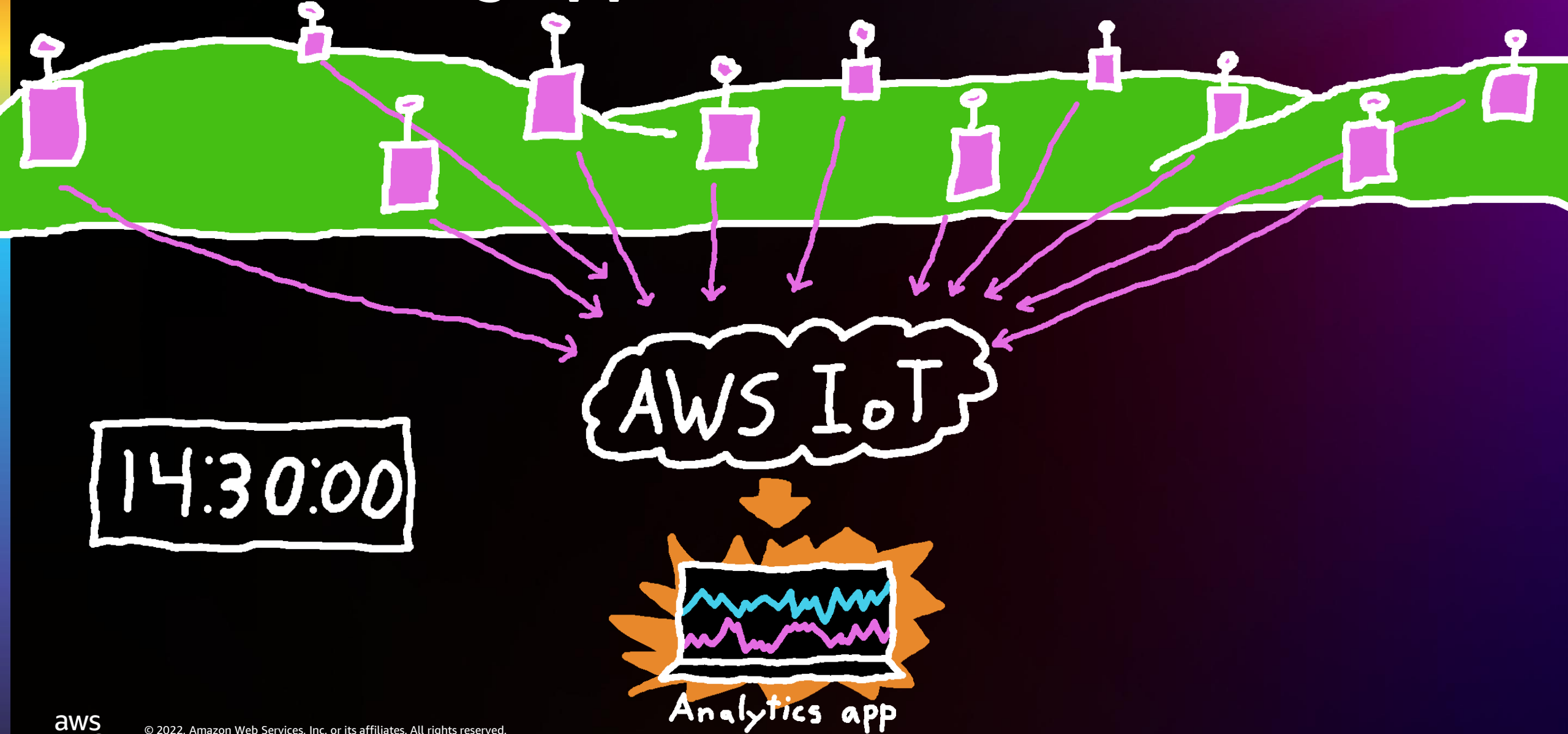
14:29:59

AWS IoT



Analytics app

IoT monitoring application

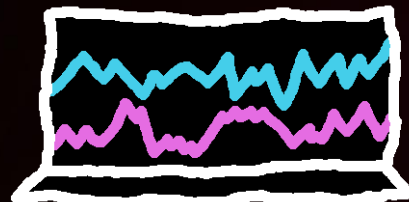


IoT monitoring application



14:30:01

AWS IoT



Analytics app

Slice and dice log analysis

```
1 fields @timestamp, @message
2 | sort @timestamp
3 | stats count(@message) by bin(1s)
4
```



Aggregation defined up front

```
{  
  "Operation": "GetProduct",  
  "CustomerId": "user@example.com",  
  "ProductId": "BrightOrangeSneakers",  
  "ClientIp": "192.168.131.39",  
  "InstanceId": "i-001234a4bf70dec41EXAMPLE",  
  "AvailabilityZone": "us-west-2a",  
  "Cache.Node": "10.0.2.182",  
  "Cache.Hit": 0,  
  "Cache.Time": 1,  
  "DB.Time": 10,  
  "Error": 0,  
  "Time": 100,  
  "TraceId": "Root=1-5759e988-bd862e3fe1be46a994272793;Sampled=1"  
}
```


Aggregation defined up front

```
{  
  "Operation": "GetProduct",  
  "CustomerId": "user@example.com",  
  "ProductId": "BrightOrangeSneakers",  
  "ClientIp": "192.168.131.39",  
  "InstanceId": "i-001234a4bf70dec",  
  "AvailabilityZone": "us-west-2a",  
  "Cache.Node": "10.0.2.182",  
  "Cache.Hit": 0,  
  "Cache.Time": 1,  
  "DB.Time": 10,  
  "Error": 0,  
  "Time": 100,  
  "TraceId": "Root=1-5759e988-bd862e3fe1be46a994272793;Sampled=1"  
}
```

"Show me latency per-operation"

```
"Dimensions": [["Operation"]],  
"Metrics": [  
  { "Name": "Time",  
    "Unit": "Milliseconds" }]
```

One-off aggregation

```
{  
  "Operation": "GetProduct",  
  "CustomerId": "user@example.com",  
  "ProductId": "BrightOrangeSneakers",  
  "ClientIp": "192.168.131.39",  
  "InstanceId": "i-001234a4bf70dec",  
  "AvailabilityZone": "us-west-2a",  
  "Cache.Node": "10.0.2.182",  
  "Cache.Hit": 0,  
  "Cache.Time": 1,  
  "DB.Time": 10,  
  "Error": 0,  
  "Time": 100,  
  "TraceId": "Root=1-5759e988-bd862e3fe1be46a994272793;Sampled=1"  
}
```

"Show me the top 20 products that are returning errors for requests from IP address 192.168.131.39"

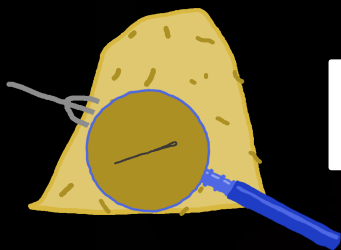
```
1 filter ClientIp='192.168.131.39' and Error > 0  
2 | stats count() as errorCount by ProductId  
3 | sort errorCount desc  
4 | limit 20
```

One-off aggregation

```
{  
  "Operation": "GetProduct",  
  "CustomerId": "user@example.com",  
  "ProductId": "BrightOrangeSneakers",  
  "ClientIp": "192.168.131.39",  
  "InstanceId": "i-001234a4bf70dec",  
  "AvailabilityZone": "us-west-2a",  
  "Cache.Node": "10.0.2.182",  
  "Cache.Hit": 0,  
  "Cache.Time": 1,  
  "DB.Time": 10,  
  "Error": 0,  
  "Time": 100,  
  "TraceId": "Root=1-5759e988-bd862e3fe1be46a994272793;Sampled=1"  
}
```

"Show me the products with the lowest cache hit rate, and which cache node they landed on"

```
1 stats · (sum(Cache.Hit) / count()) as cacheHitRate  
2 | · by ProductId, Cache.Node  
3 | · sort cacheHitRate  
4 | · limit 20
```

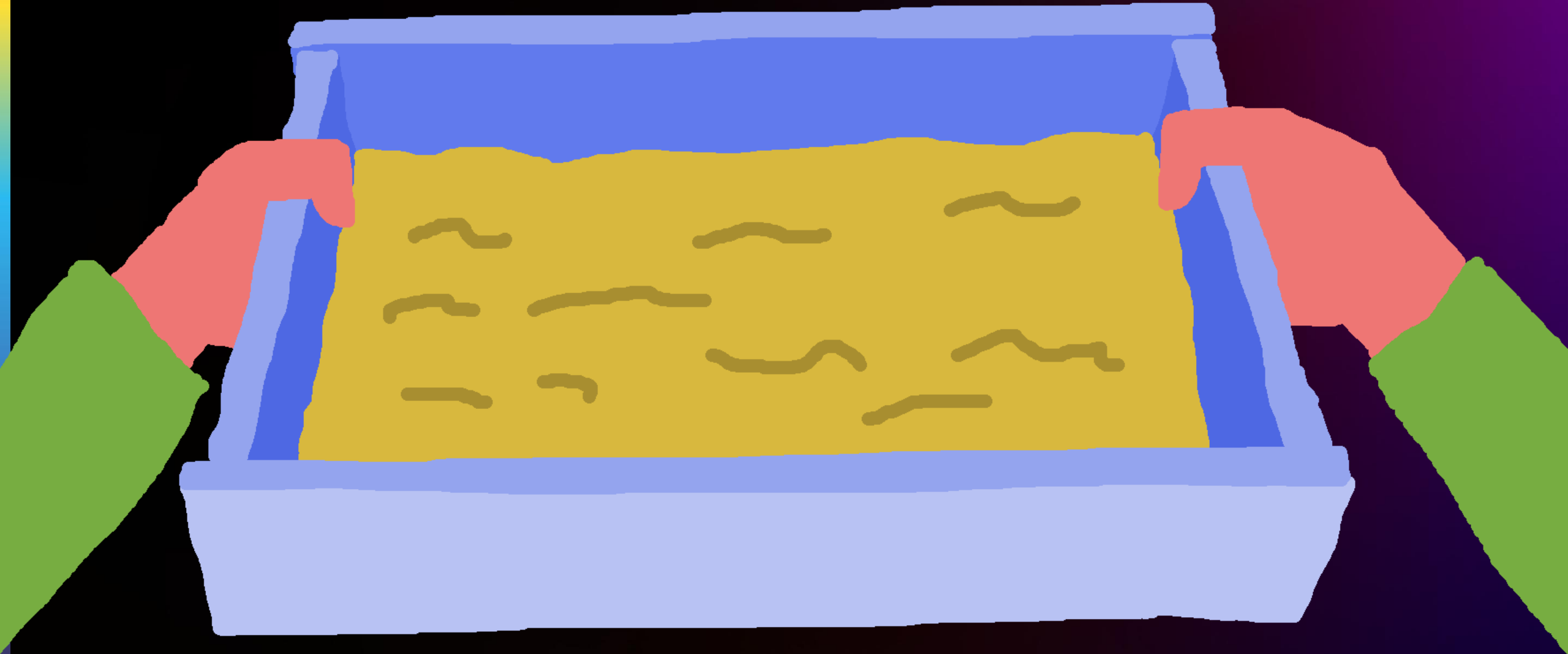


Finding the root cause

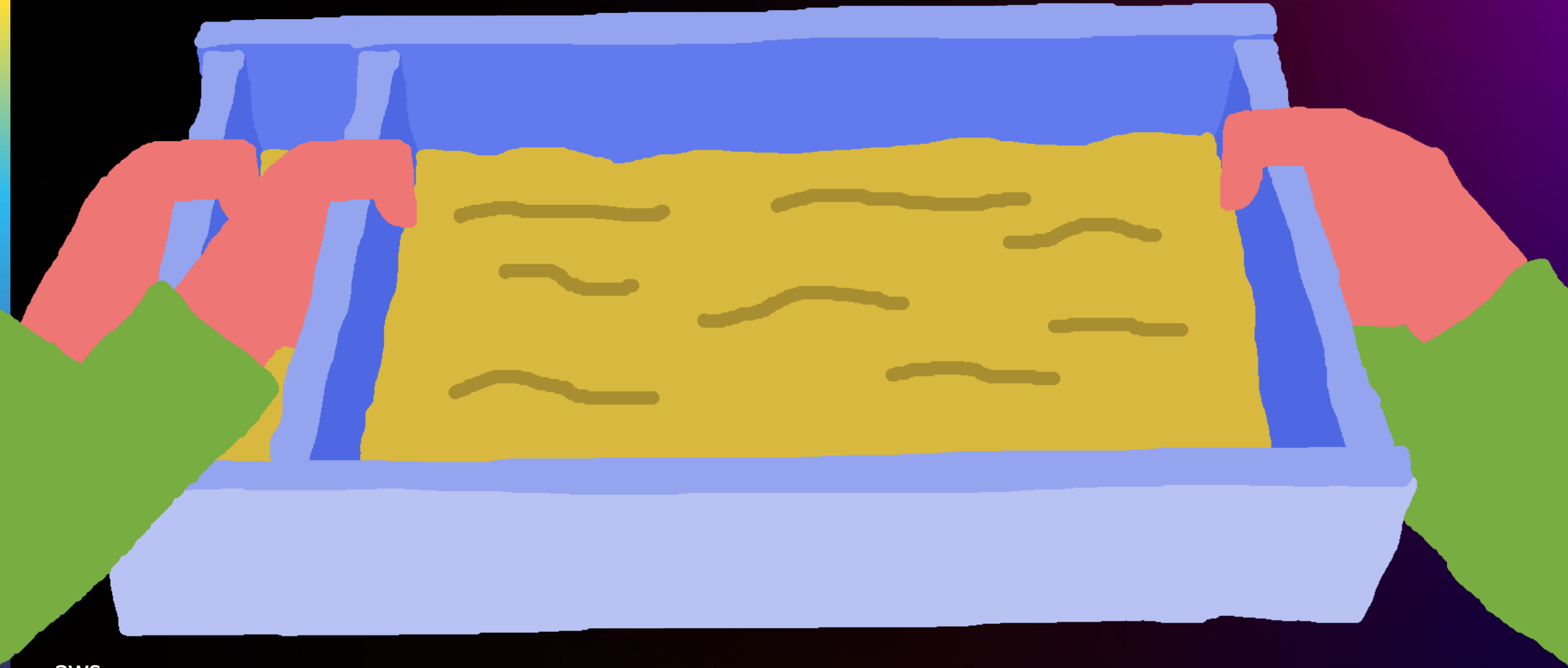
- Propagate trace information
- Build lots of dashboards
- Get multi-dimensional
- Use high cardinality metrics
- Dig even deeper with log analysis
- **Look at the raw logs**
- Analyze using profiler



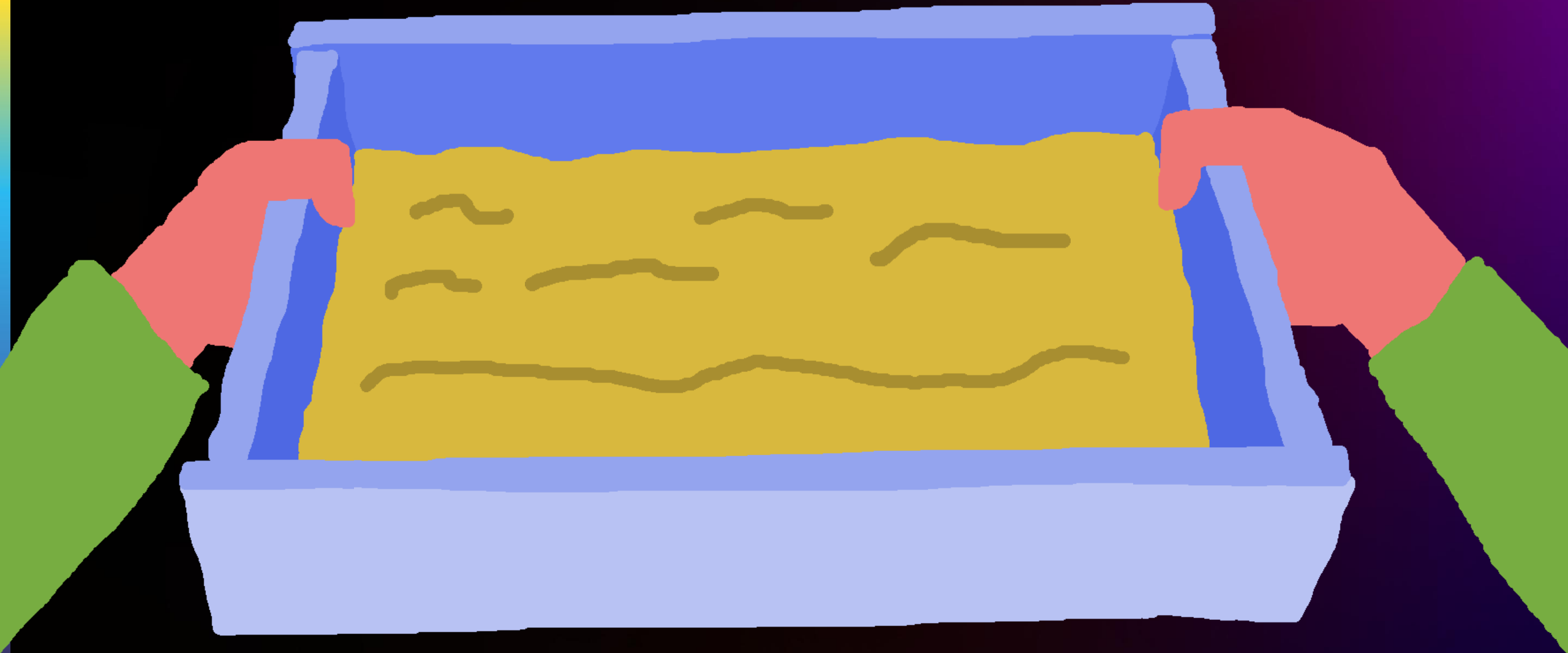
Sifting through logs



Sifting through logs



Sifting through logs



Raw logs

"I'm seeing errors talking to Amazon DynamoDB."

"Did the request make it to DynamoDB?"

```
java.net.UnknownHostException: dynamodb.us-east-1.amazonaws.com:  
unknown error
```

```
at java.net.Inet6AddressImpl.lookupAllHostAddr(Native Method)  
at java.net.InetAddress$2.lookupAllHostAddr(InetAddress.java:928)  
at java.net.InetAddress.getAddressesFromNameService(InetAddress.java:1323)  
at java.net.InetAddress.getAllByName0(InetAddress.java:1276)  
at java.net.InetAddress.getAllByName(InetAddress.java:1192)  
at java.net.InetAddress.getAllByName(InetAddress.java:1126)  
at com.amazonaws.SystemDefaultDnsResolver.resolve(SystemDefaultDnsResolver.j
```


Raw logs

“I’m seeing errors talking to Amazon DynamoDB.”

“What did DynamoDB say back?”

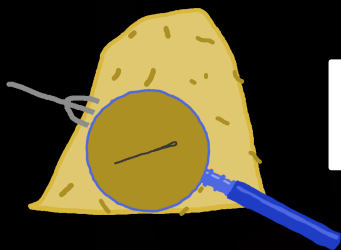
```
com.amazonaws.services.dynamodbv2.model.ConditionalCheckFailedException:  
The conditional request failed (Service: AmazonDynamoDBv2; Status Code: 400;  
Error Code: ConditionalCheckFailedException; Request ID: ...)  
    at com.amazonaws.http.AmazonHttpClient.handleErrorResponse(AmazonHttpClient.  
    at com.amazonaws.http.AmazonHttpClient.executeOneRequest(AmazonHttpClient.ja  
    at com.amazonaws.http.AmazonHttpClient.executeHelper(AmazonHttpClient.java:4  
    at com.amazonaws.http.AmazonHttpClient.execute(AmazonHttpClient.java:302)  
    at com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient.invoke(AmazonDynam  
    at com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient.updateItem(AmazonD
```

Raw logs

“I’m seeing errors talking to Amazon DynamoDB.”

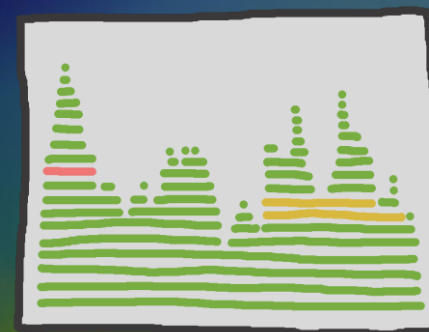
“What happened right before that?”

```
[WARN] 2022-11-14T20:37:58.617Z    DynamoDB call timed, out; retrying 1/3
```



Finding the root cause

- Propagate trace information
- Build lots of dashboards
- Get multi-dimensional
- Use high cardinality metrics
- Dig even deeper with log analysis
- Look at the raw logs
- **Analyze using profiler**



Example: Inefficient code

```
public ProductInfo getProductInfo(String customerId, String productId) {
```

```
    ProductInfo info = cache.get(customerId, productId);
```

```
    // compute pi just to make sure  
    // TODO refactor or optimize  
    for (int i = 0; i < 10000; i++) {  
        calculatePI();  
    }
```

No timer around this part of
the code

```
    if (info == null) {  
        info = db.query(customerId, productId);  
    }
```

```
    return info;
```

```
}
```

{CodeGuru} DemoProfilingGroup-WithIssues DEMO

Data

View

Time range

CPU

Overview

2021-07-16 @ 16:30 – 17:30 PDT latest 12h

Actions

Q

4 Recommendations

▼ Legend

My code

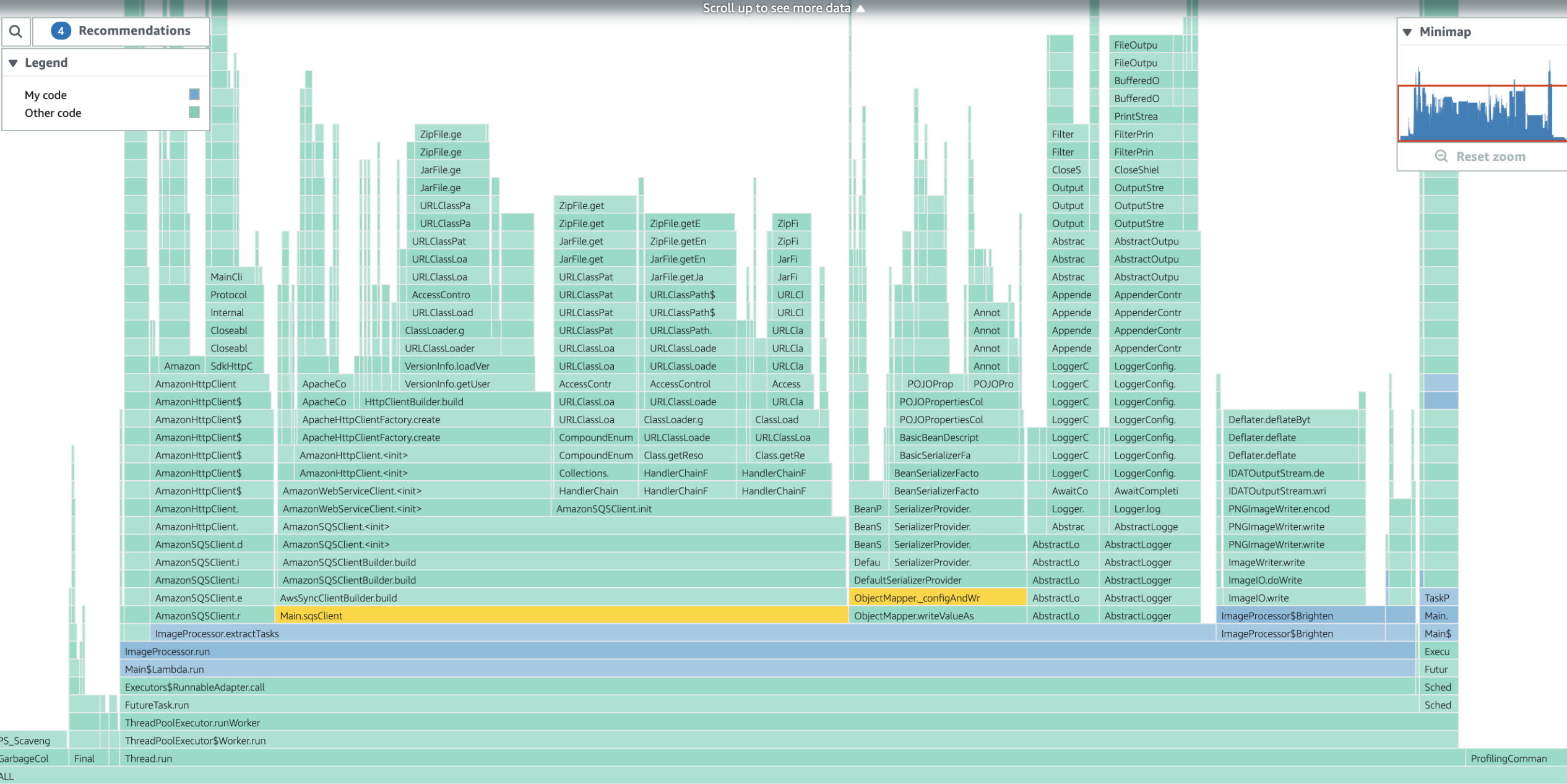
Other code

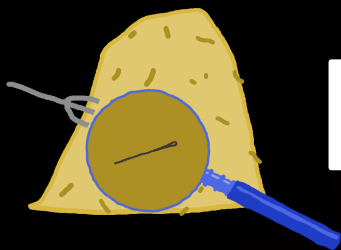
Scroll up to see more data ▲

▼ Minimap

Q

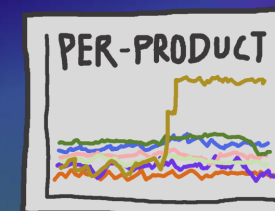
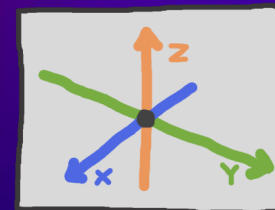
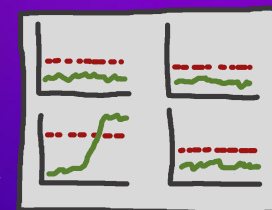
Reset zoom





Finding the root cause

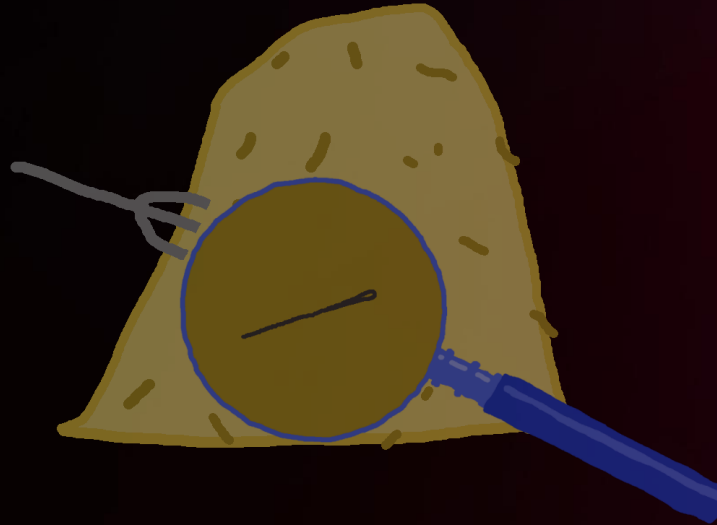
- Propagate trace information
- Build lots of dashboards
- Get multi-dimensional
- Use high cardinality metrics
- Dig even deeper with log analysis
- Look at the raw logs
- Analyze using profiler



Chapter three



The DevOps
flywheel at Amazon



Find the
root cause



Measure from
everywhere



Measure from everywhere

- Be your own customer
- Instrument close to the customer
- Alarm on everything

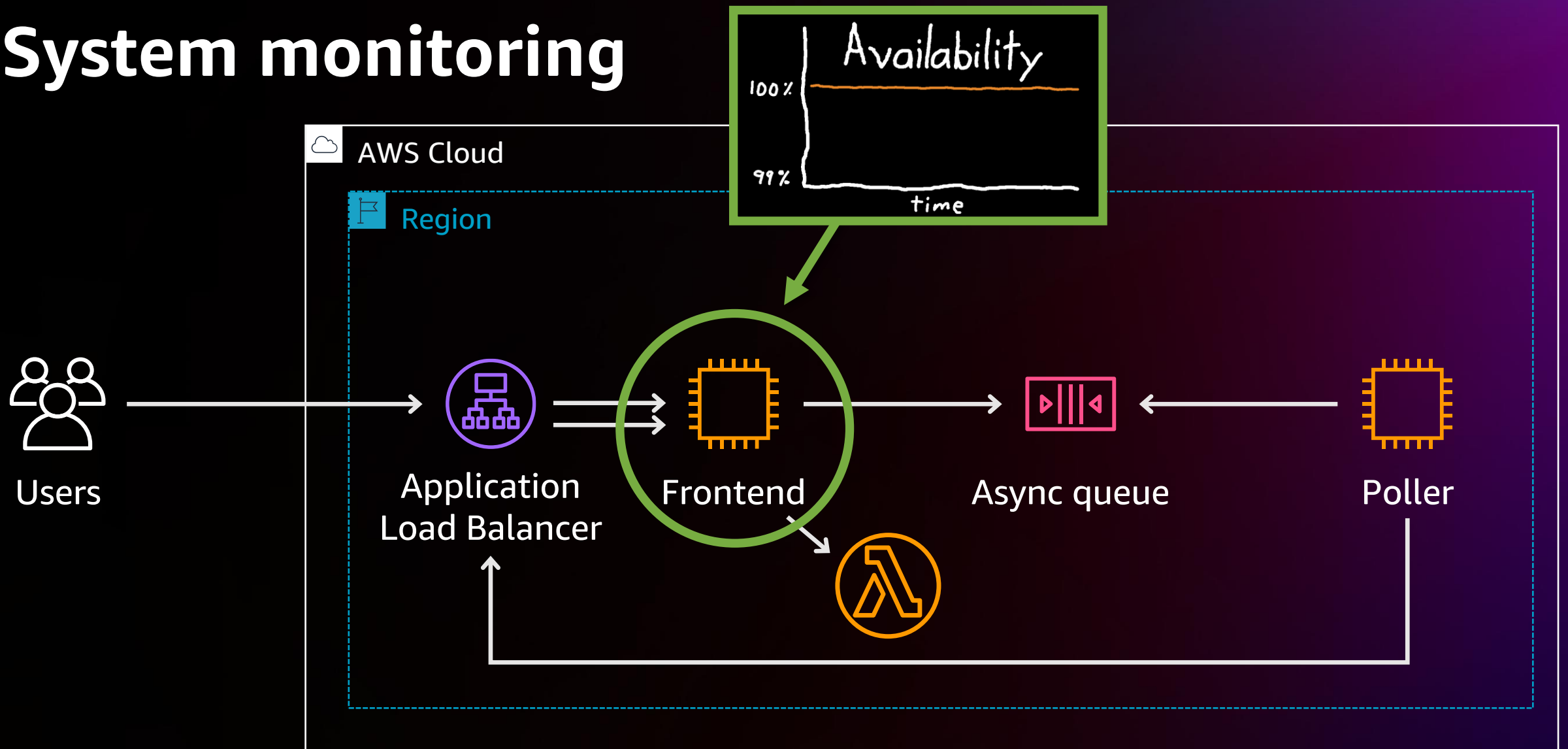


Measure from everywhere

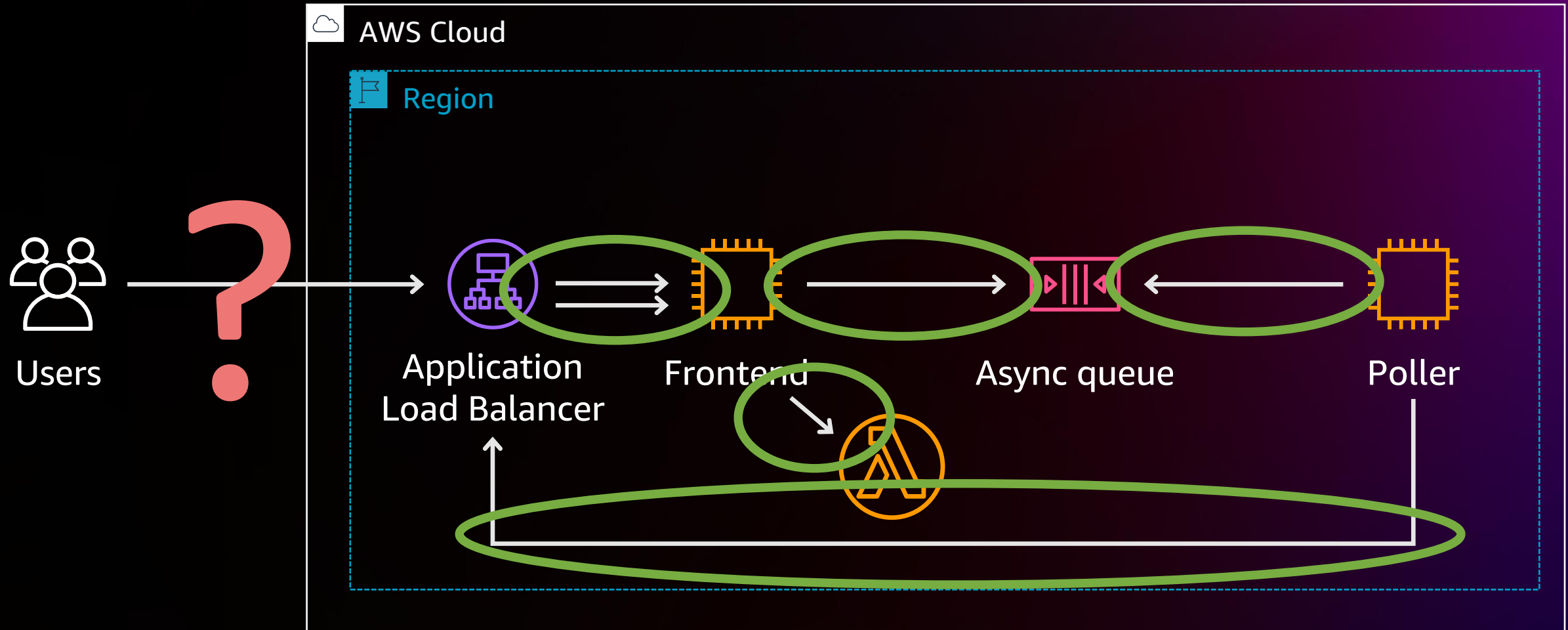


- **Be your own customer**
- Instrument close to the customer
- Alarm on everything

System monitoring



System monitoring

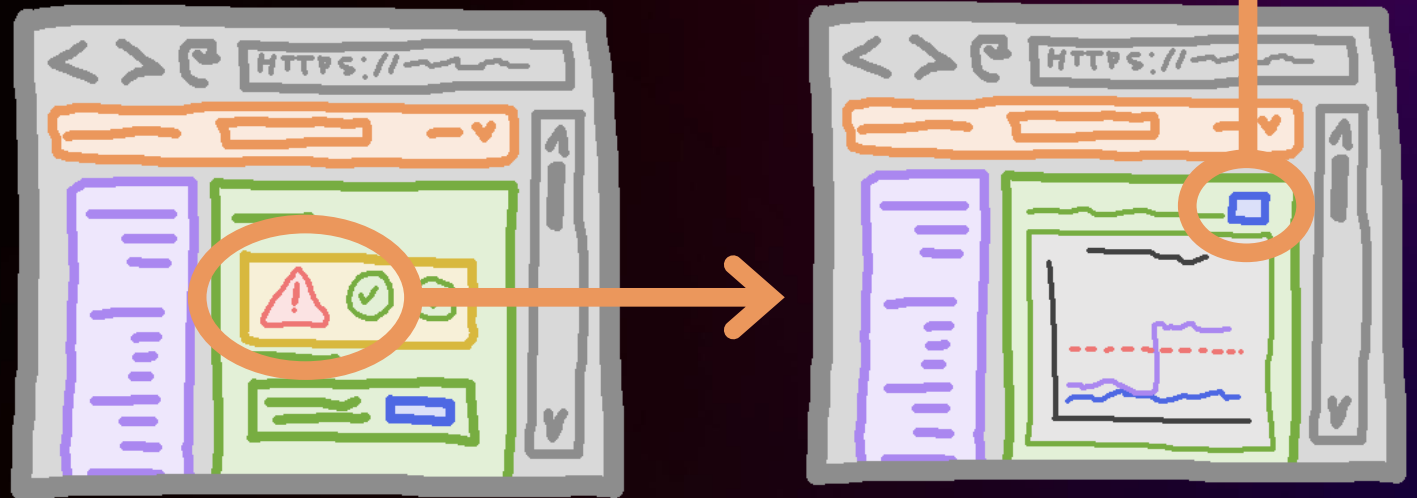


Solving browser testing

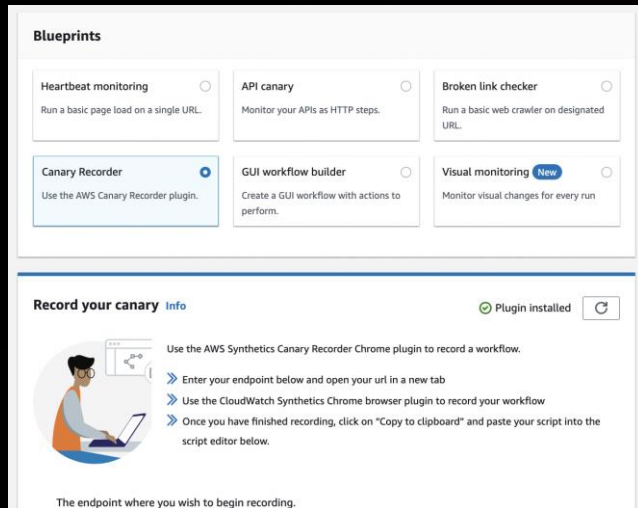


Diverse device types,
browsers

Testing flows of
human interaction

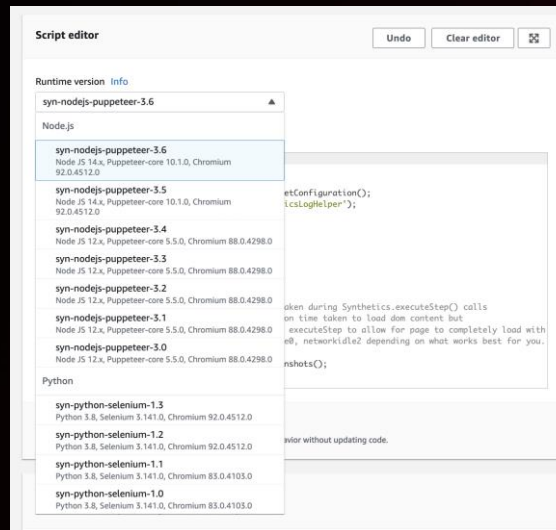


CloudWatch Synthetics

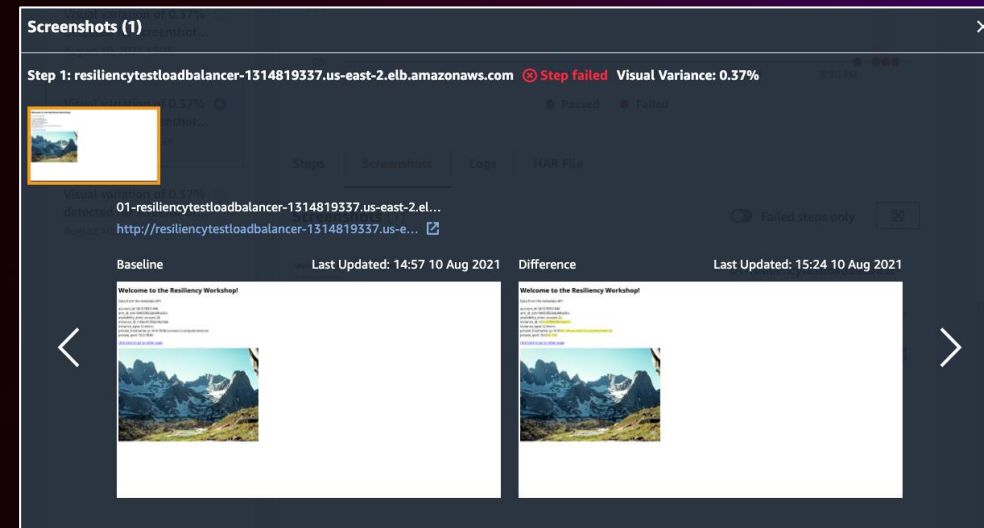


Pre-built blueprint scripts
& visual script reorder
(chrome plugin)

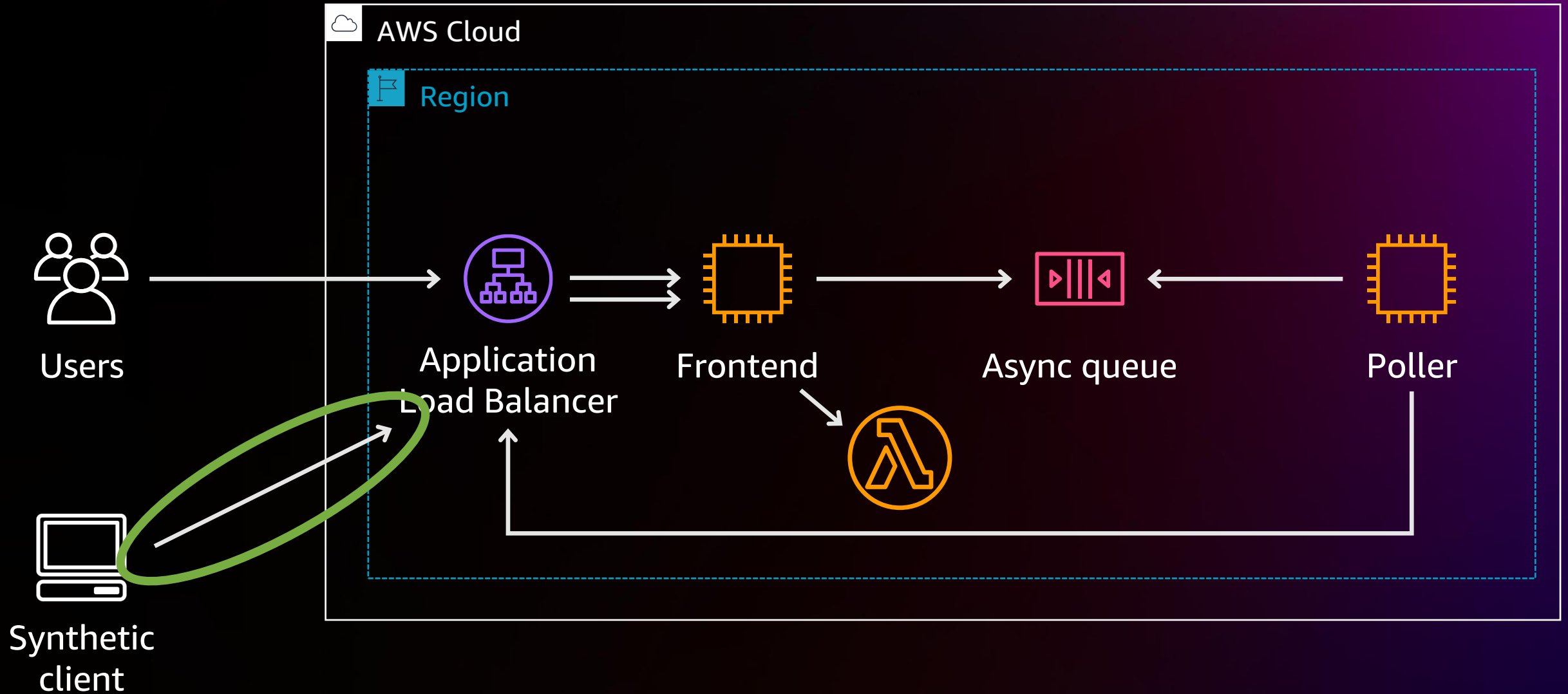
Fully customizable scripting in
NodeJS and Python using
Selenium & Puppeteer



Visually inspect
synthetic failures



System monitoring



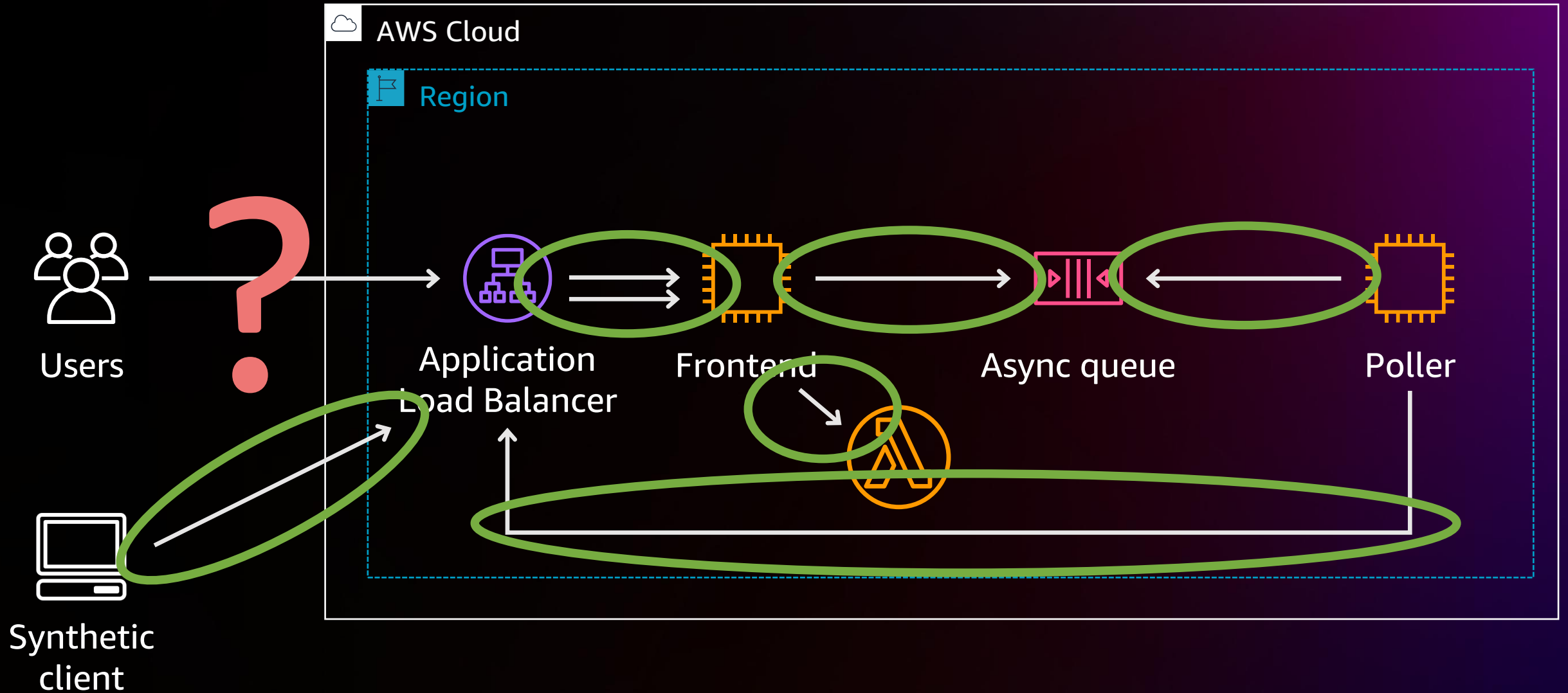


Measure from everywhere

- Be your own customer
- **Instrument close to the customer**
- Alarm on everything



System monitoring



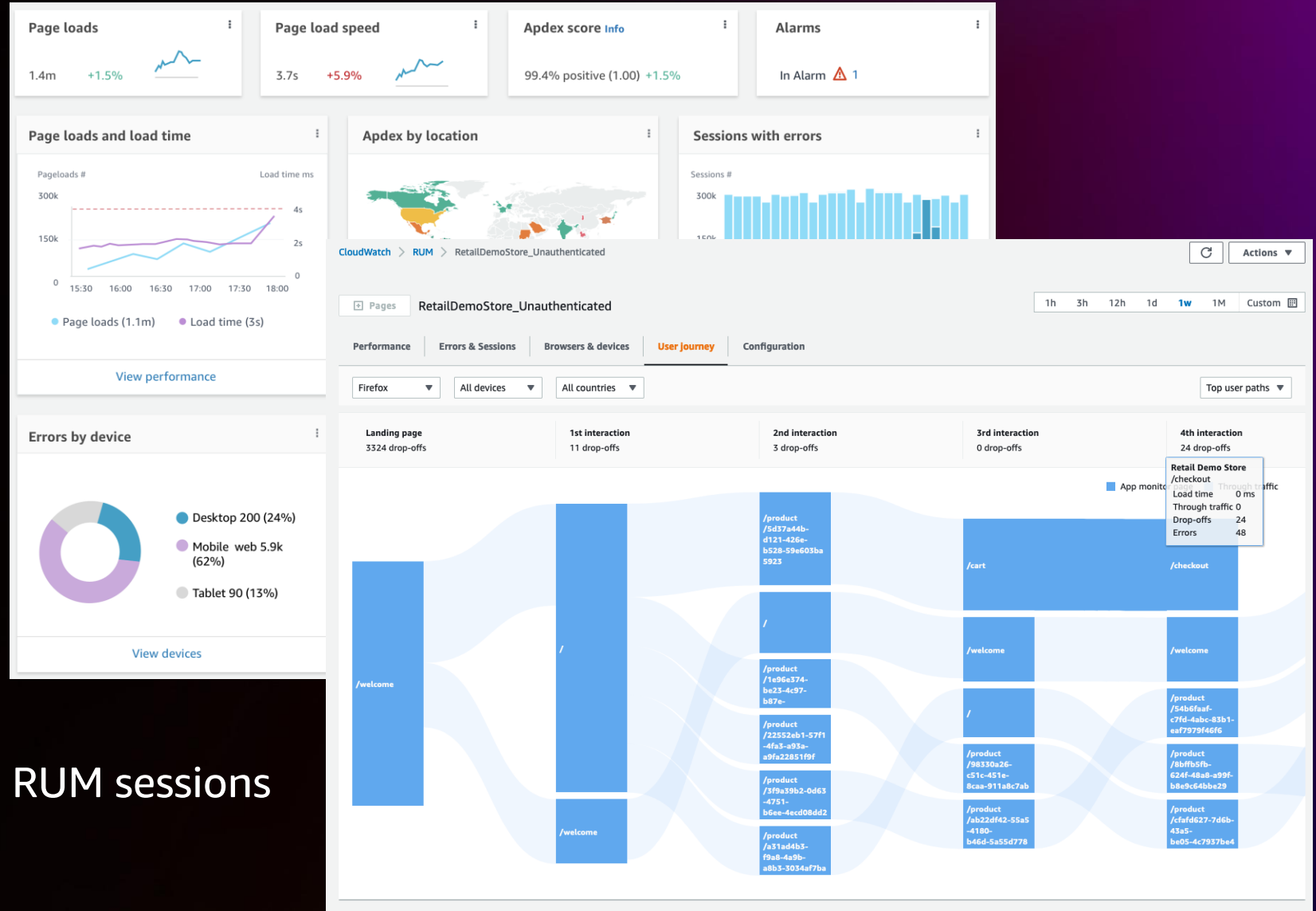
CloudWatch RUM

Quantifying slowness

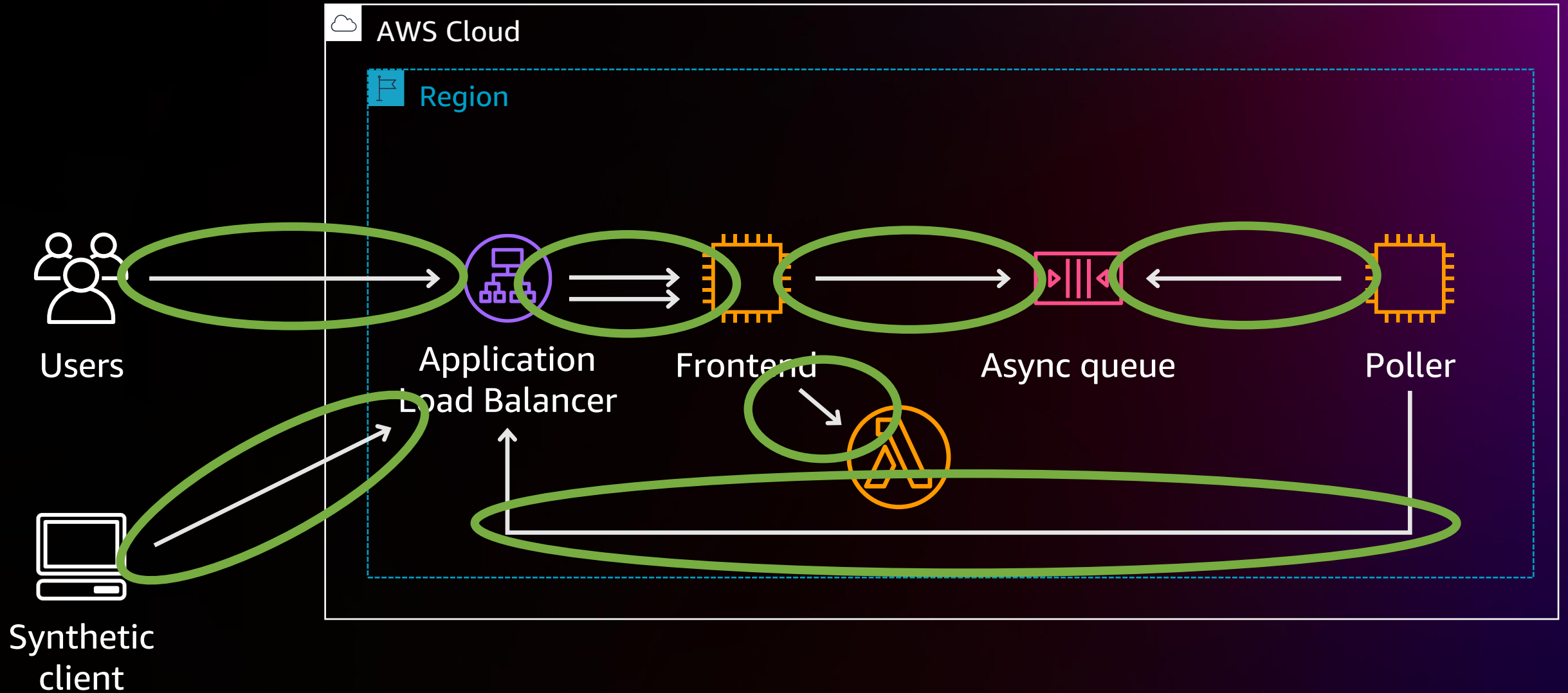
- Apdex scores
- Errors by sessions, devices, canaries, regions, pages
- Core web vitals
- Largest contentful paint
- First input delay
- Cumulative layout shift

Debugging slowness

- Page load steps
- Resource requests
- View X-Ray traces for errors in RUM sessions
- JavaScript, HTTP errors



System monitoring



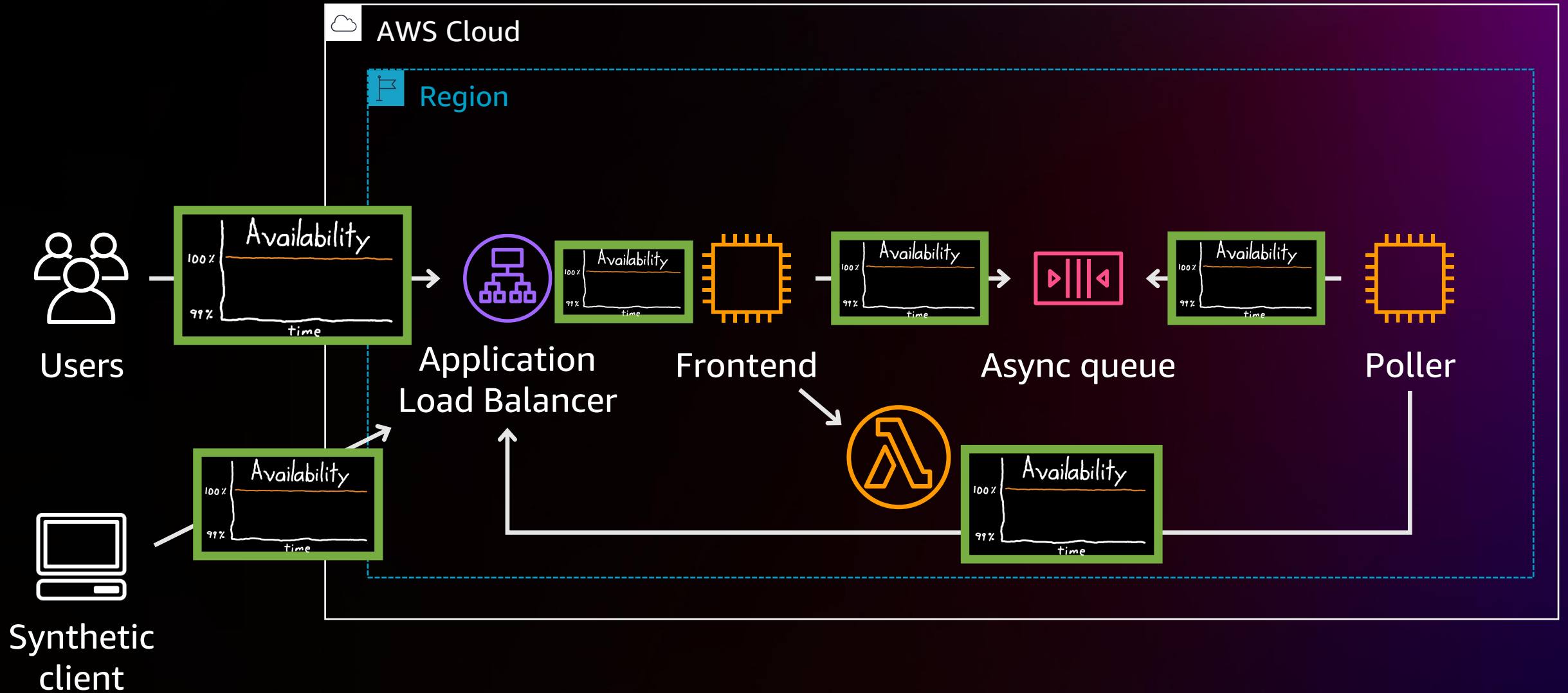


Measure from everywhere

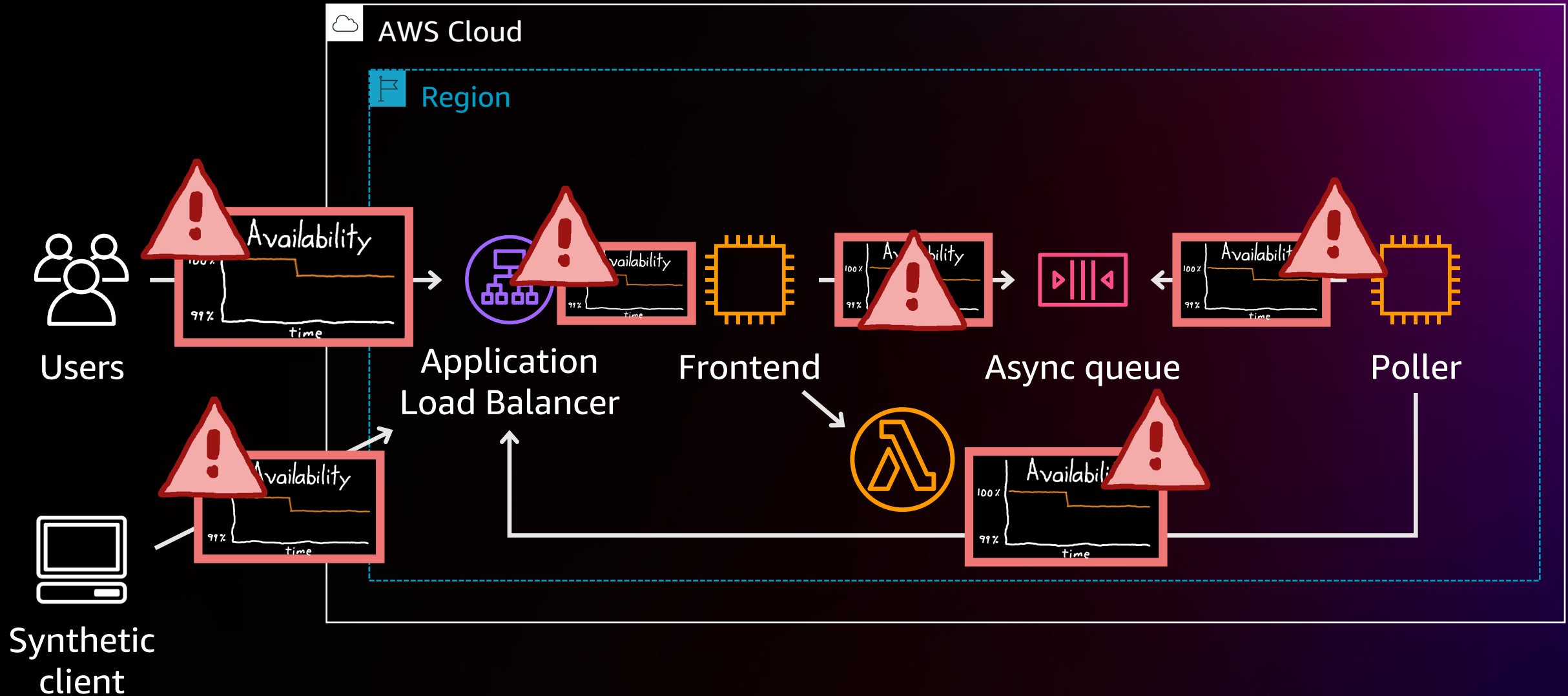
- Be your own customer
- Instrument close to the customer
- **Alarm on everything**



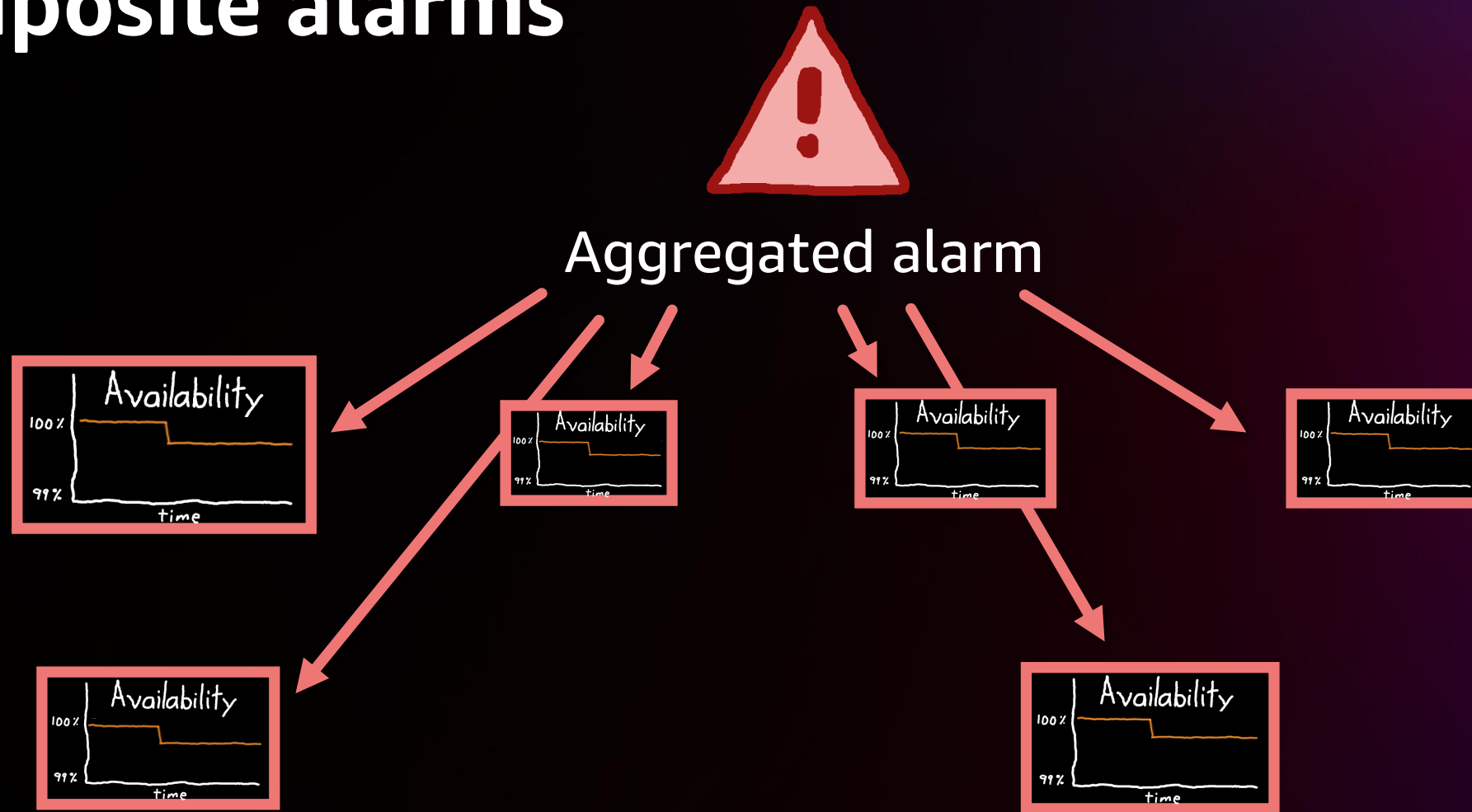
System monitoring



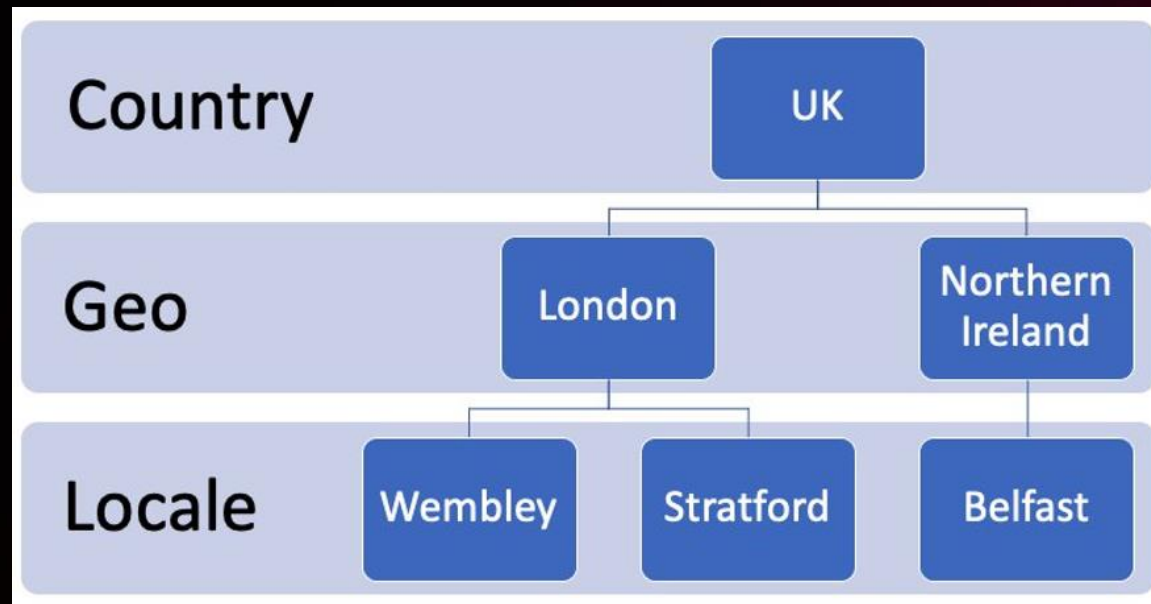
System monitoring



Composite alarms



Reducing noise at scale – BT



<https://aws.amazon.com/blogs/mt/how-bt-uses-amazon-cloudwatch-to-monitor-millions-of-devices/>



Measure from everywhere

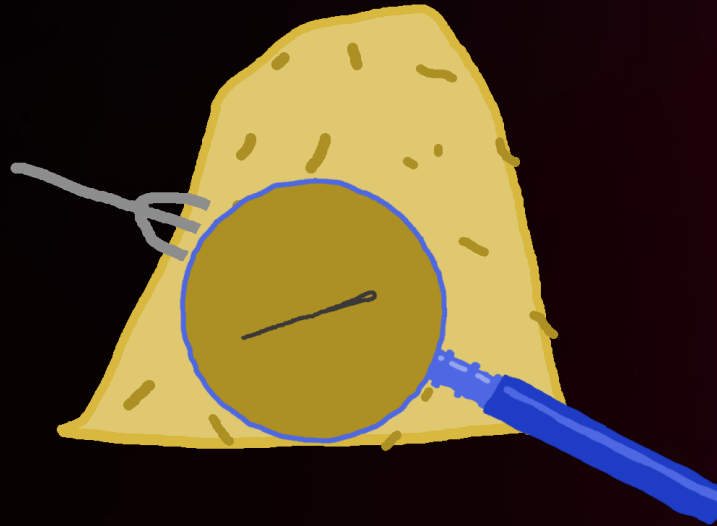
- Be your own customer
- Instrument close to the customer
- Alarm on everything



Key takeaways



The DevOps
flywheel at Amazon

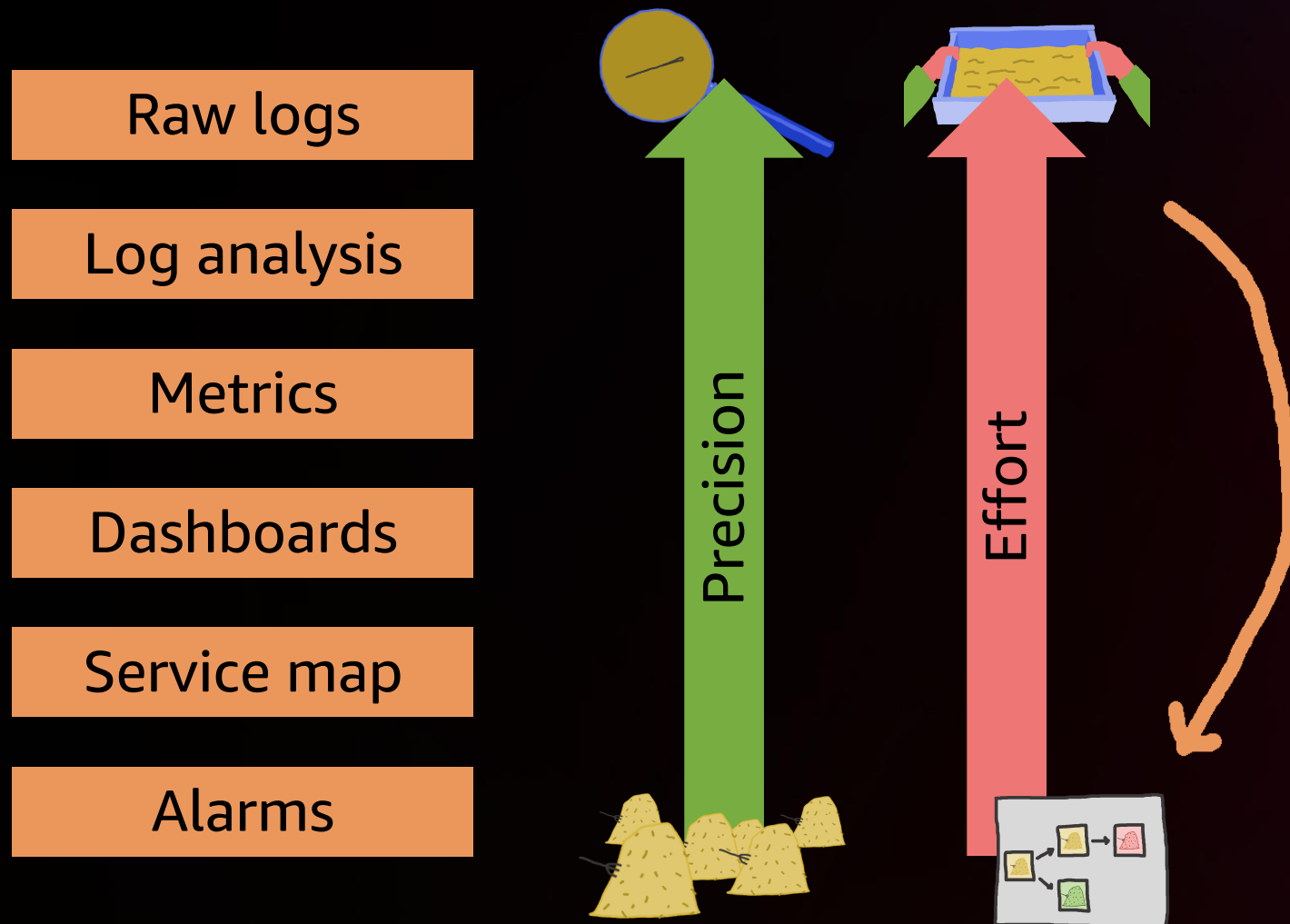


Find the
root cause

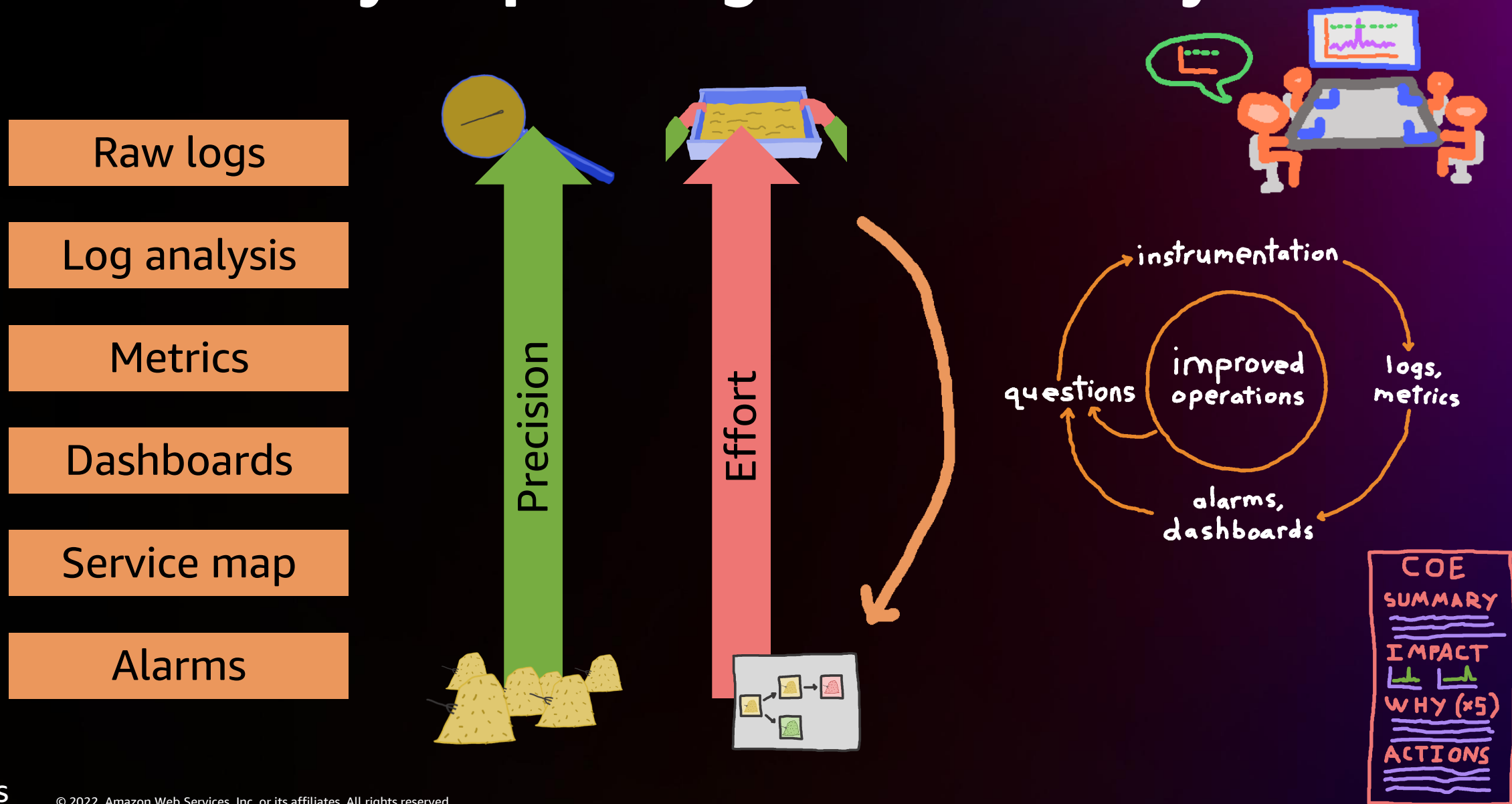


Measure from
everywhere

Continuously improving observability



Continuously improving observability



Thank you!

David Yanacek

 @dyanacek

Ian McGarry

 ianmcgarry

<https://aws.amazon.com/builders-library/>



Please complete the session survey in the **mobile app**



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.