AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

STG309

Integrating serverless applications with AWS storage services

Hunter Leath
Sr. Product Manager, Amazon EFS
Amazon Web Services

Rafael Koike
Principal Solutions Architect
Amazon Web Services

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Serverless object detection API



Web interface

Serverless API

Analytics

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Workshop agenda

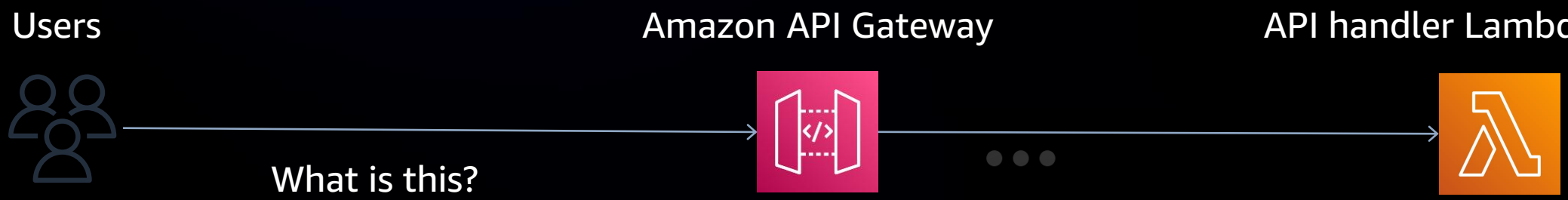
Amazon EFS + AWS Lambda
Module 1: Integrate a serverless API with Amazon EFS

Web hosting on Amazon S3
Module 2: Build a serverless web application with Amazon S3

Amazon S3 as a data lake
Module 3: Enhance application insights with an Amazon S3 data lake

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Module 1: A serverless API for ML powered object detection





```
def run_neural_net(image, config_path, weights_path, return_image):
    img_nparr = np.frombuffer(image, np.uint8)
    image = cv2.imdecode(img_nparr, cv2.IMREAD_COLOR)
    print(image)
    (H, W) = image.shape[:2]
    net = cv2.dnn.readNetFromDarknet(config_path, weights_path)
    ln = net.getLayerNames()
    ln = [ln[i[0] - 1] for i in net.getUnconnectedOutLayers()]
    blob = cv2.dnn.blobFromImage(imageswapRB)
    net.setInput(blob)
    detected_objects = net.forward(ln)
```

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

More parameters = Higher accuracy

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Model accuracy

Model	File size	Mean average precision	# of parameters
Yolo tiny	~43 MB	23.7	~ 119
Yolo big	~240 MB	60.6	~ 682

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

File system limitations

Lambda quotas

PDF | Kindle | RSS

/tmp directory storage	512 MB

Ephemeral, for example, must be downloaded for each invocation



Deployment package (.zip file archive) size	50 MB (zipped, for direct upload) 250 MB (unzipped) This quota applies to all the files you upload, including layers and custom runtimes. 3 MB (console editor)

Model size + required code exceeds limit



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The solution

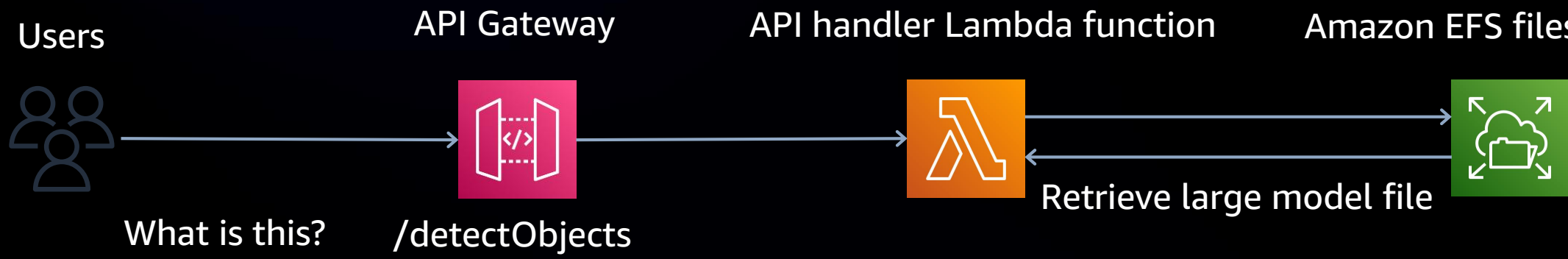


AWS Lambda

Amazon EFS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How it works





- Mounts filesystem when execution environment is prepared
- Minimal latency
- If Lambda is warm, mount is already available
- Scales to 25,000 concurrent connections

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

https://dashboard.eventengine.run

Participant hash:
ef0a-1c7563eae4-32

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Module 2: Serverless web app with Amazon S3

```html
<!DOCTYPE html>
<html lang="en">

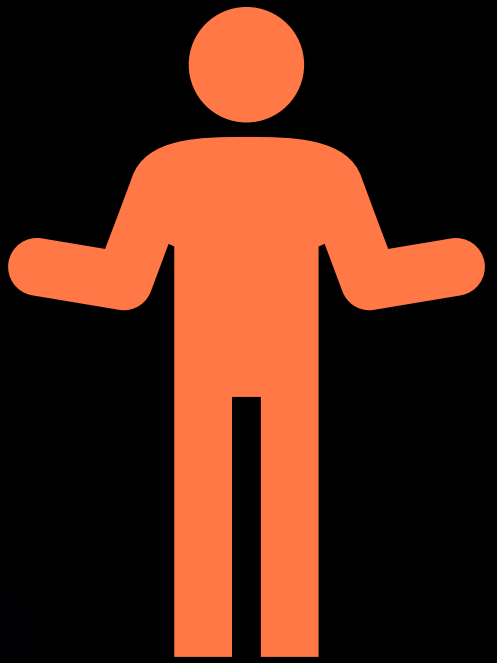
    <head>
        <meta charset="UTF-8">
        <title>Hello!</title>
    </head>

    <body>
        <h1>Hello World!</h1>
        <p>This is a simple paragraph.</p>
    </body>

</html>
```



Amazon Simple
Storage Service (Amazon S3)



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Traditional web server





Install OS

Updates

Security patches

Configure networking

Configure firewall

Secure the OS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Setup requires specific domain knowledge …

… Takes hours or even days to complete

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Writing code



Provisioning server



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Web apps are distinguished based on where the code is run

Modern web apps perform logic in a web browser and communicate with backend cloud resources, such as databases, through web APIs

A web app consists of a set of HTML, CSS, and JavaScript files

These files can be hosted by Amazon S3 and served to users with no need to provision or maintain a single server

Can scale to handle enterprise-level traffic with no additional work

Take advantage of AWS Amplify for fully managed hosting



Amazon Simple Storage Service (Amazon S3)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Hours or days to deploy web app

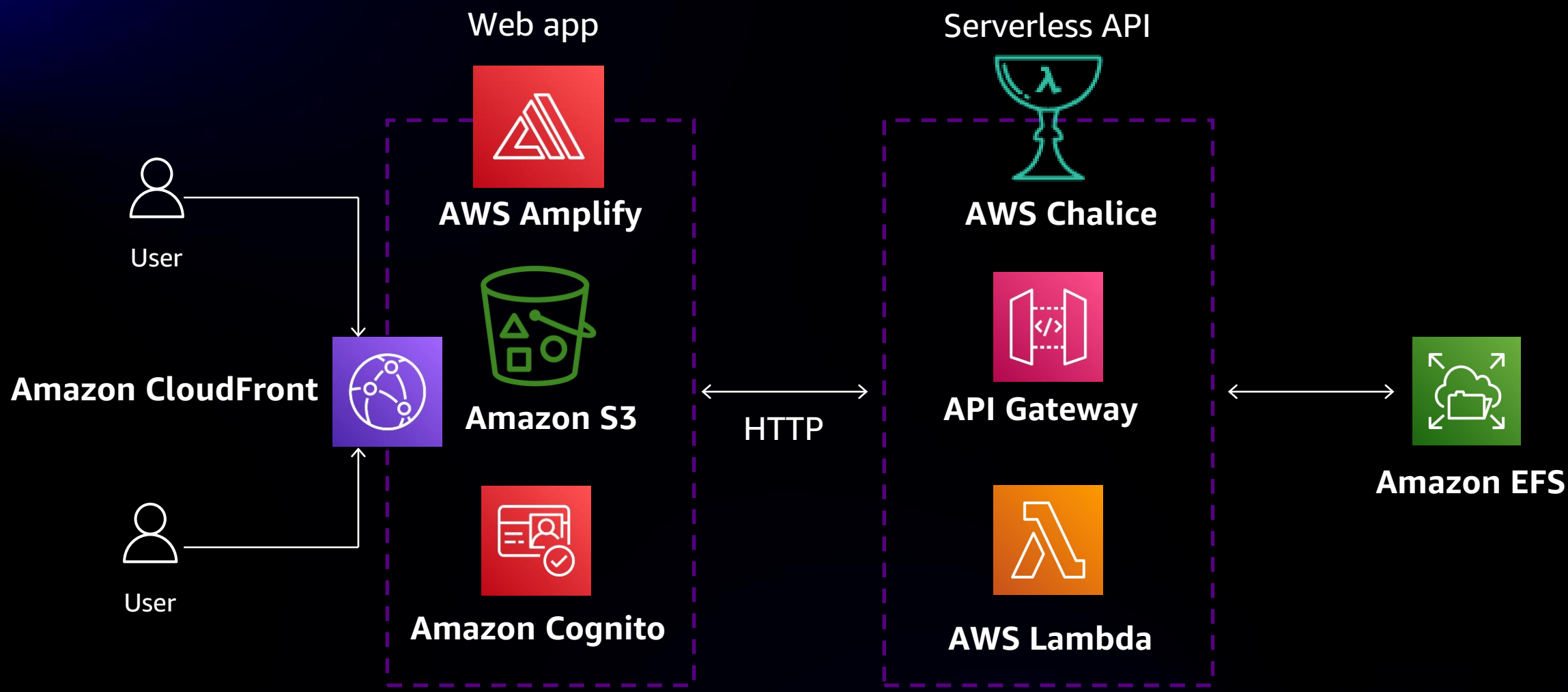


Deploying your web app in minutes



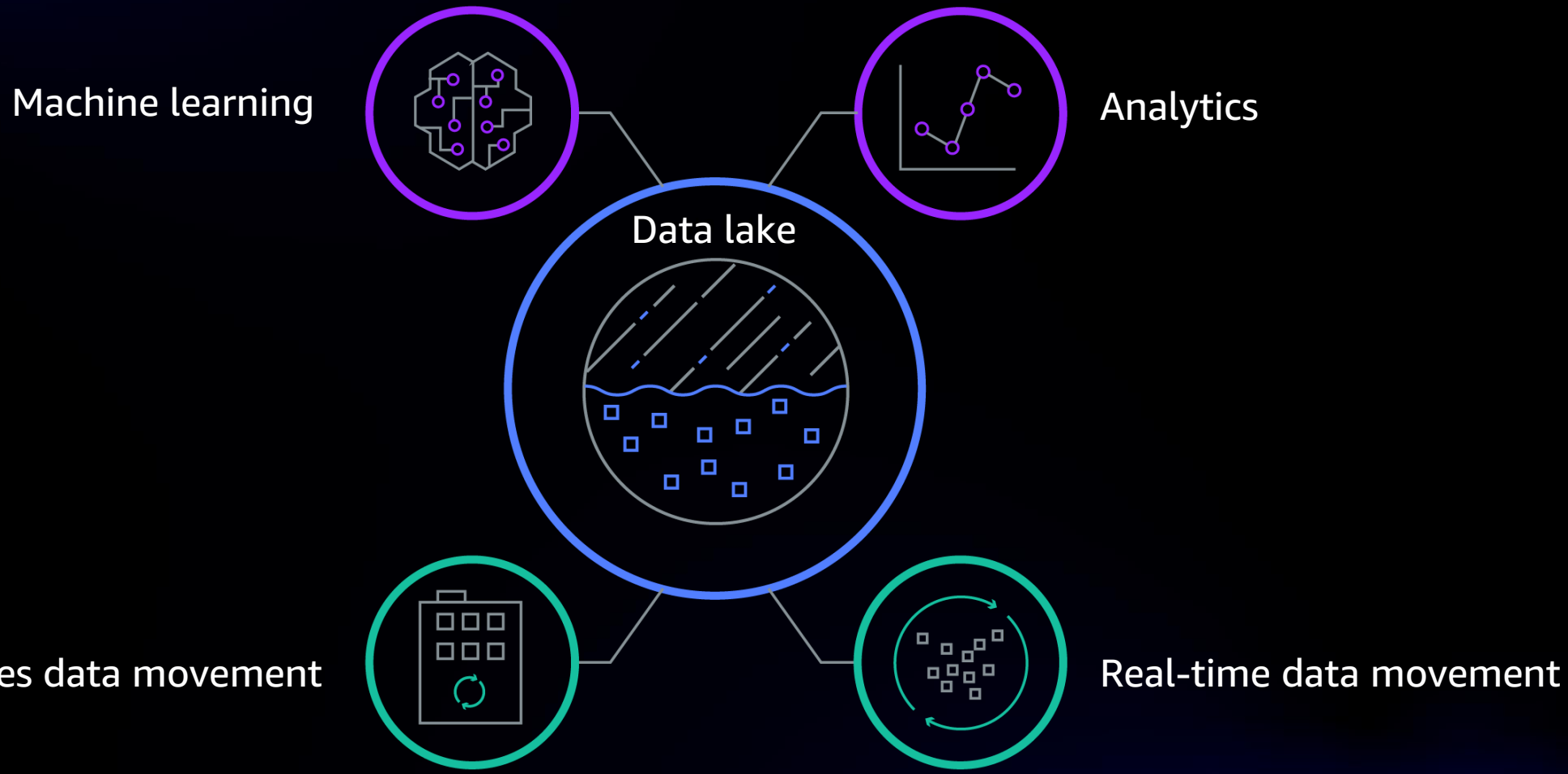
“Thank you S3”

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Module 3: Amazon S3 data lake



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Traditional data analytics

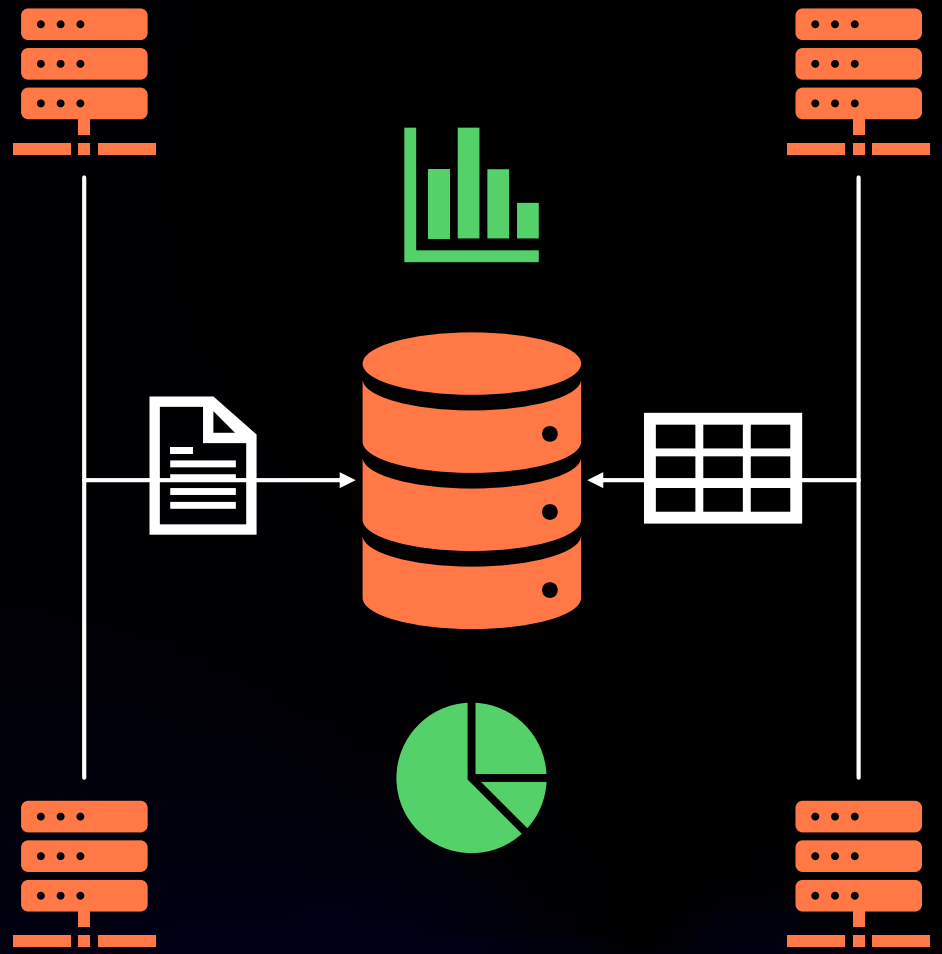
Can have similar issues to those described earlier, such as OS/software configuration, network/security configuration

Often proprietary systems

Intricate and involves many components

Requires data model to be defined in advance

Data must be structured in a tabular format, for instance



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Some use cases do require that sort of data analysis architecture …

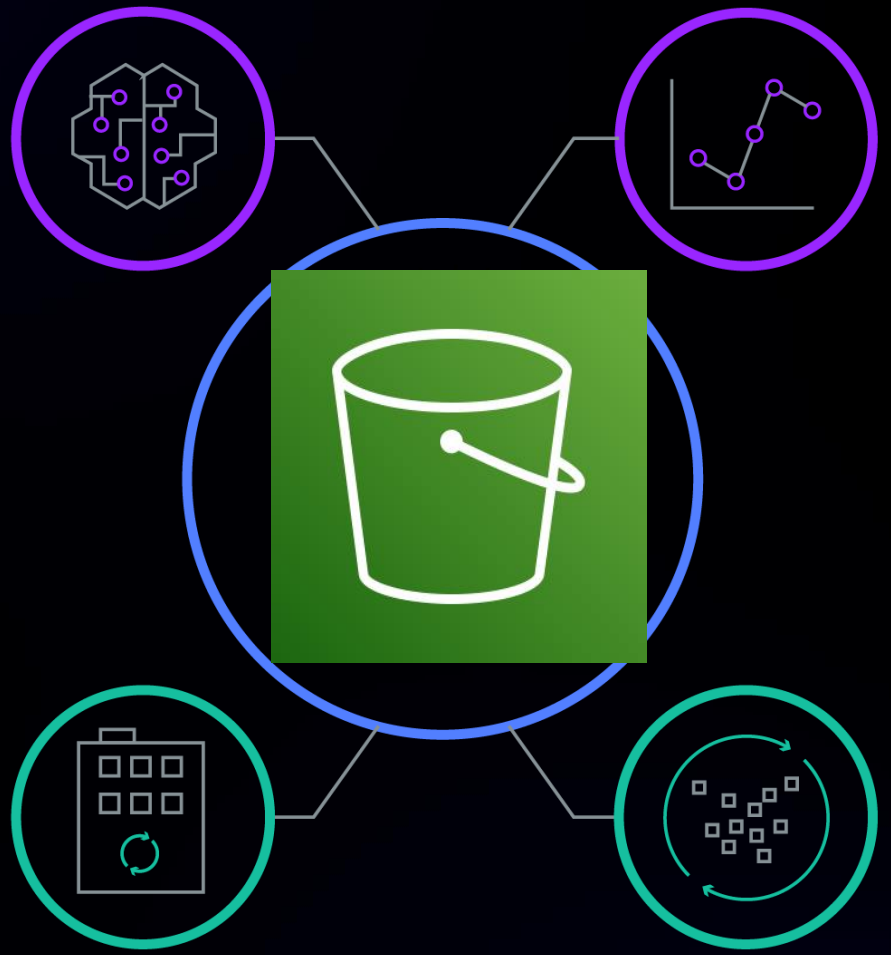
… but it’s not needed for analyzing unstructured or semi-structured ML data

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Central repository for all data, structured or unstructured

Schema written at the time of analysis (schema-on-read)

Supports different types of analytics



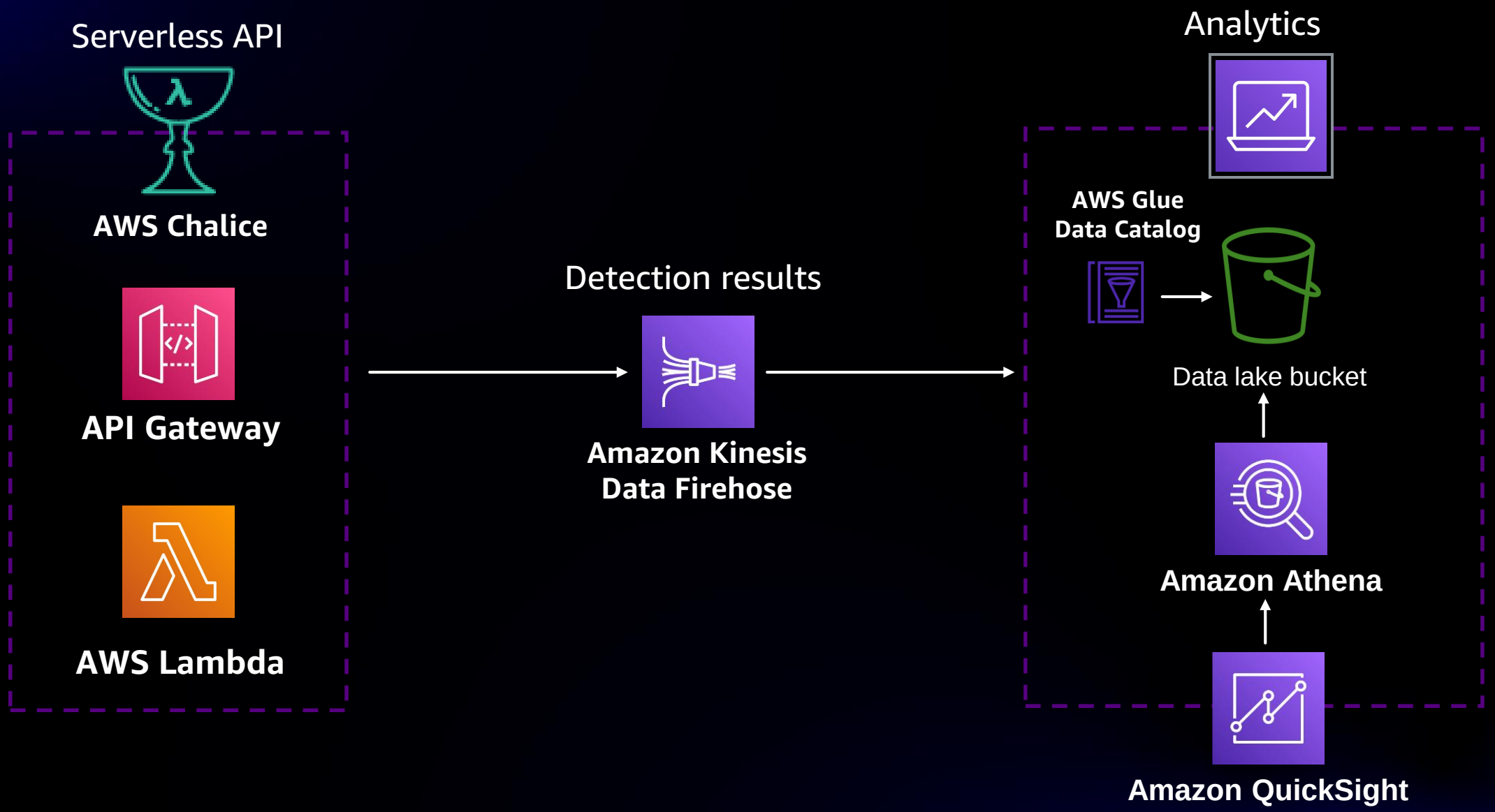
Automatically scales

No need to carefully define data models

Not a single server to manage

Amazon S3 data lake

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Hunter Leath
hleath@amazon.com

Rafael Koike
koiker@amazon.com

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.