A serverless data pipeline

Julien Simon, Principal Technical Evangelist, AWS
julsimon@amazon.fr
@julsimon

Werner Vogels, CTO, Amazon.com
AWS re:Invent 2015

Managed services

+

AWS Lambda

=

Serverless architecture

AWS Lambda

- Deploy pure functions in Java, Python and Node.js

- Works nicely with AWS managed services:
  Amazon S3, Amazon DynamoDB, etc.

- Build event-driven applications

- Build RESTful APIs in conjunction with Amazon API Gateway

- Pay as you go: number of requests + execution time (100ms slots)

Another way to put it...

Tim Wagner,
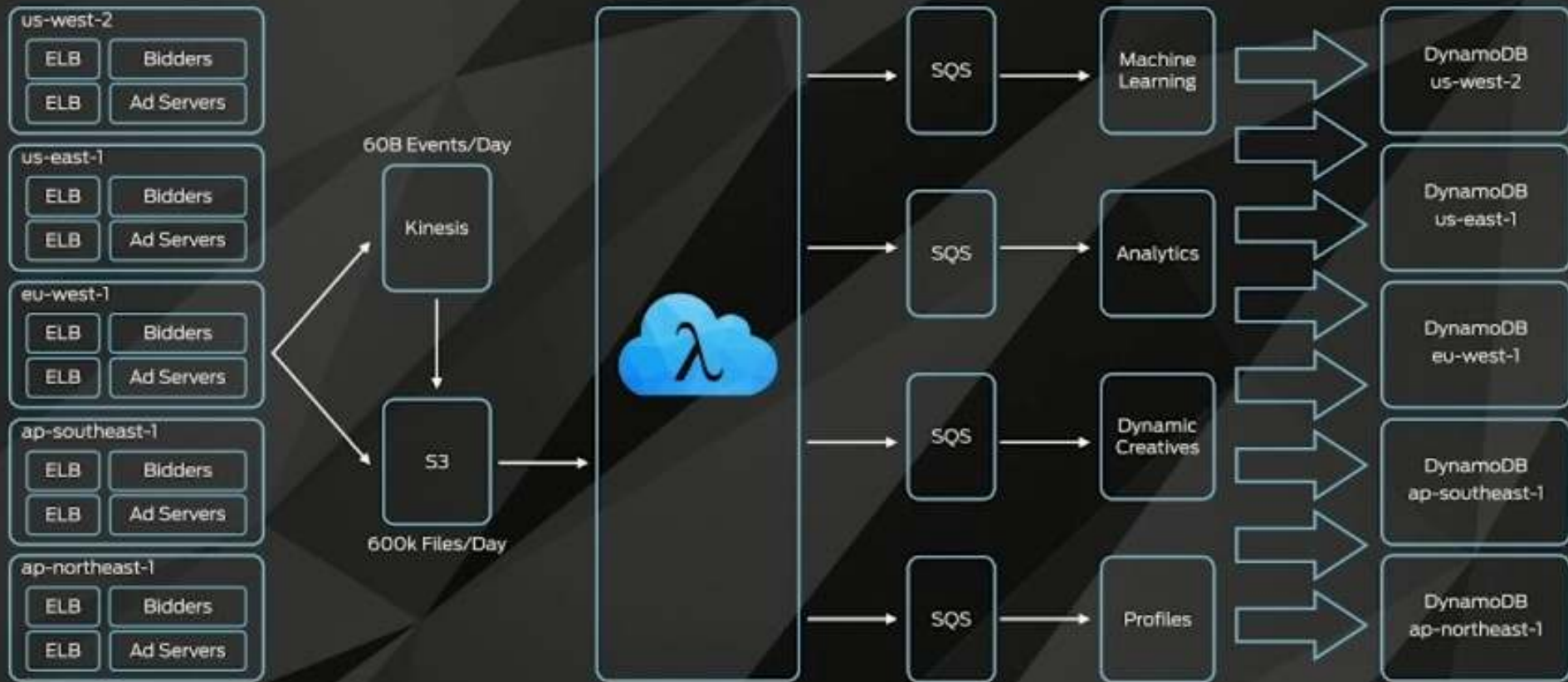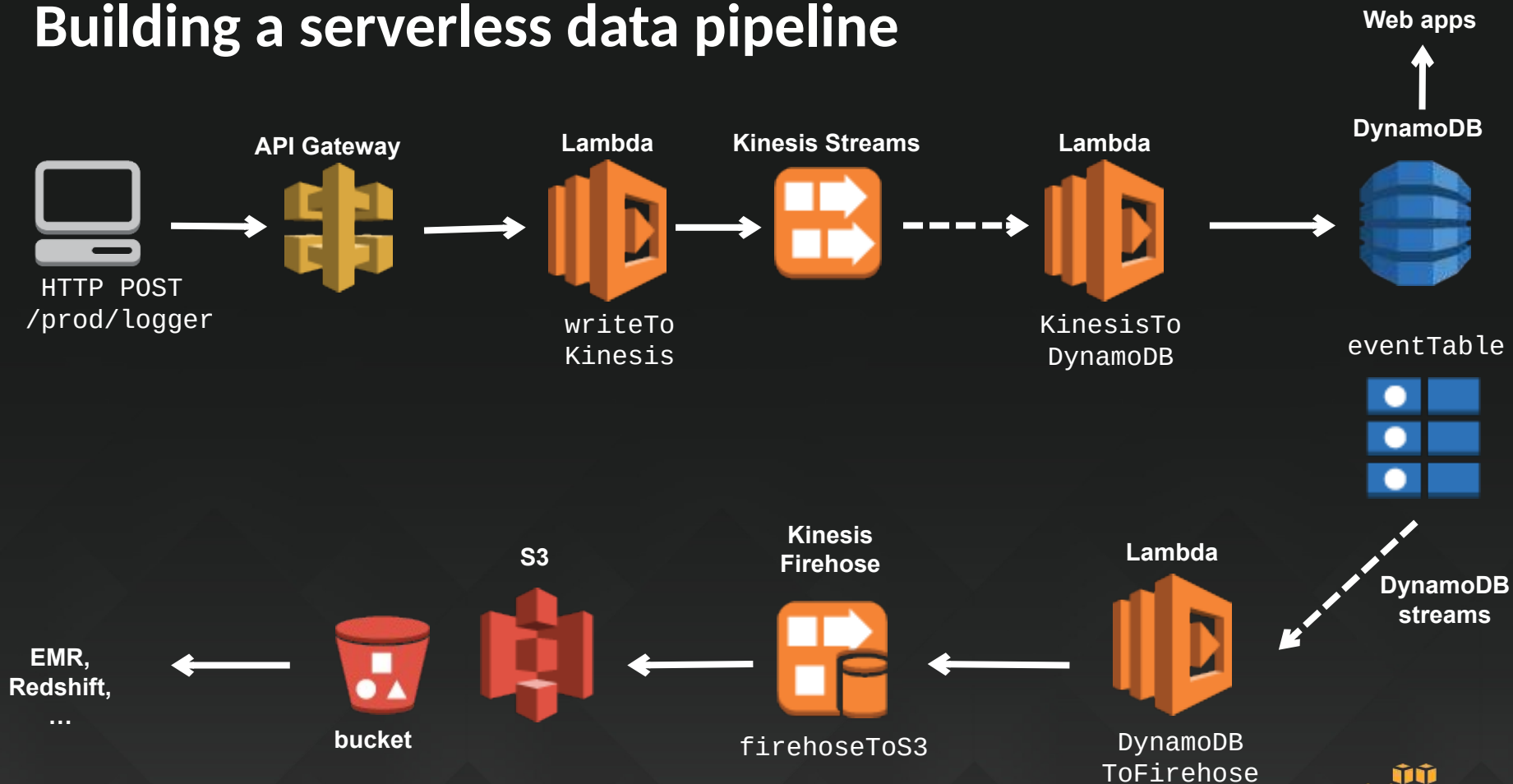General Manager,
AWS Lambda

AdRoll

60 billion Real-Time Bidding events daily

Building a serverless data pipeline

CLI: create DynamoDB table

```
aws dynamodb create-table \
--table-name eventTable \
--attribute-definitions \
AttributeName=userId,AttributeType=N \
AttributeName=timestamp,AttributeType=N \
--key-schema \
AttributeName=userId,KeyType=HASH \
AttributeName=timestamp,KeyType=RANGE \
--provisioned-throughput ReadCapacityUnits=5,WriteCapacityUnits=5 \
--stream-specification StreamEnabled=true,StreamViewType=NEW_IMAGE
```

DynamoDB

eventTable

amazon
web services
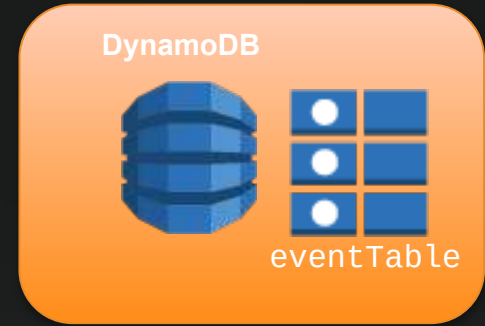
CLI: create Lambda function

```
aws lambda create-function \
--function-name DynamoDBToFirehose \
--role DYNAMODBTOFIREHOSE_ROLE_ARN \
--zip-file fileb://DynamoDBToFirehose.zip \
--handler DynamoDBToFirehose.lambda_handler \
--runtime python2.7 \
--memory-size 128 \
--description "Write DynamoDB stream to Kinesis Firehose"

aws lambda create-event-source-mapping \
--function-name DynamoDBToFirehose \
--event-source DYNAMODB_STREAM_ARN \
--batch-size 10 \
--starting-position TRIM_HORIZON
```
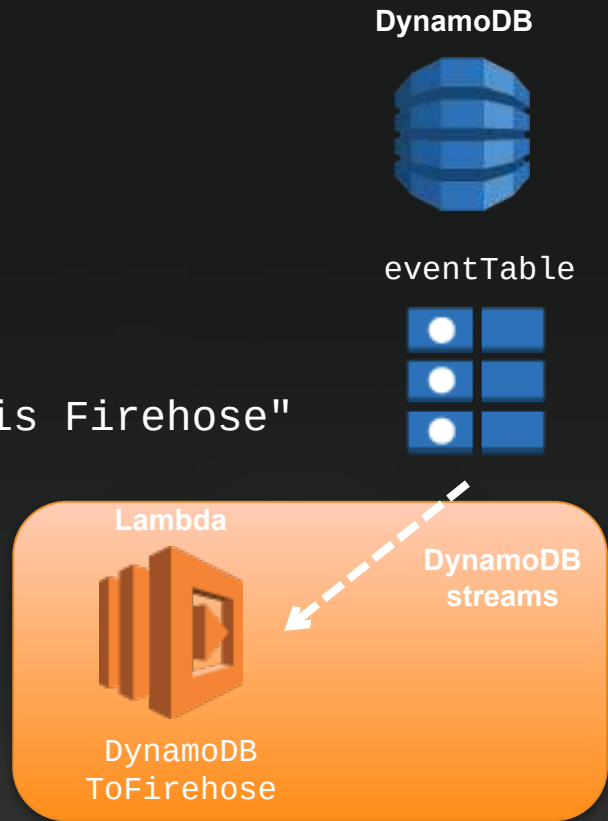
DynamoDB

eventTable

Lambda

DynamoDB streams

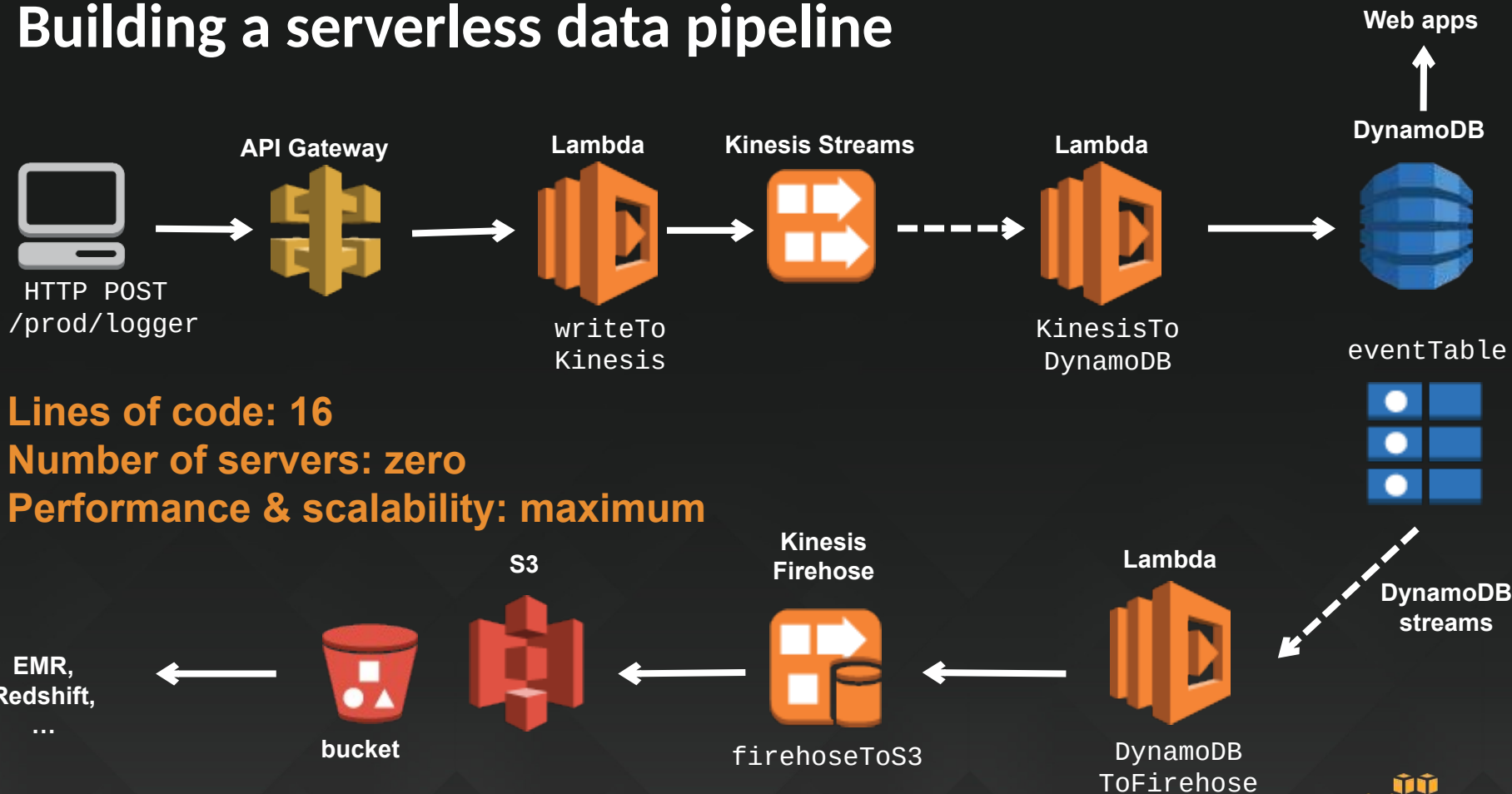DynamoDB
ToFirehose

amazon
web services

CLI: create Kinesis Firehose stream

```
aws firehose create-delivery-stream \
--delivery-stream-name firehoseToS3 \
--s3-destination-configuration \
RoleARN=FIREHOSETOS3_ROLE_ARN, \
BucketARN="arn:aws:s3:::jsimon-public", \
Prefix="firehose", \
BufferingHints=\{SizeInMBs=1,IntervalInSeconds=60\}, \
CompressionFormat="GZIP", \
EncryptionConfiguration={NoEncryptionConfig="NoEncryption"}
```
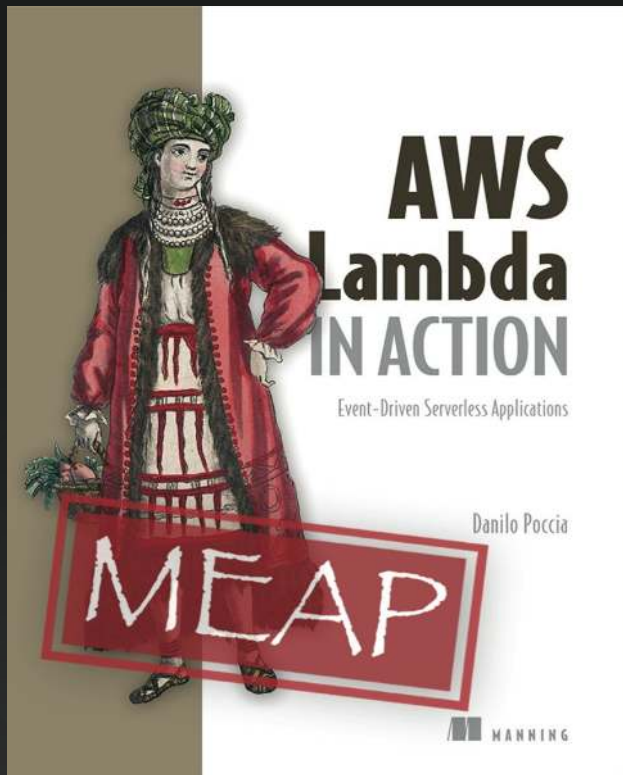
Kinesis Firehose

firehoseToS3

Lambda

DynamoDB
ToFirehose

eventTable

DynamoDB streams

DynamoDB

amazon
web services

Building a serverless data pipeline

Web apps

DynamoDB

API Gateway

Lambda
writeTo
Kinesis

Kinesis Streams

Lambda
KinesisTo
DynamoDB

HTTP POST
/prod/logger

eventTable

Lines of code: 16
Number of servers: zero
Performance & scalability: maximum

EMR,
Redshift,
...

S3
bucket

Kinesis Firehose
firehoseToS3

Lambda
DynamoDB
ToFirehose

DynamoDB
streams

amazon
web services

Ready for some testing?

http://api.julien.org

Upcoming book on AWS Lambda

Written by AWS Technical Evangelist Danilo Poccia

Early release available at: https://www.manning.com/books/aws-lambda-in-action

To go further

http://aws.amazon.com/lambda/

https://aws.amazon.com/blogs/compute/

AWS User Group Turkey
http://www.meetup.com/AWS-User-Group-Turkey/

Qwiklabs: https://qwiklabs.com

Serverless framework: https://github.com/serverless/

Teşekkür ederim ☺

Julien Simon, Principal Technical Evangelist, AWS
julsimon@amazon.fr
@julsimon

amazon
webservices