Paws

An R Package for

Amazon Web Services

Agenda

- What can you use Amazon Web Services (AWS) to do
- How to use AWS from R
- Processing big jobs with R, Paws, and AWS Batch

What is AWS good for?

Process data on big hardware

- Servers
  (EC2)
- Clusters of servers
(Elastic MapReduce – EMR)
- Batch processing
  (Batch)

Host databases

- Relational databases
(Relational Database Service – RDS)
- Data warehouses
  (Redshift)
- NoSQL databases
  (DynamoDB)

Store files

- Simple Storage Service – S3

Other AWS services

- Machine learning
- Translation
- Speech-to-text
- Text-to-speech
- Image recognition
- Natural language processing
- Extract text from documents
- Mechanical Turk

- Queues
- Notifications
- Email
- Auto-scaling
- Functions-as-a-service
- Search
- Extract-transform-load
- etc.

How Paws works

- Paws allows you to use most AWS services from R

```
# Create a client for the S3 file storage service.*
s3 <- paws::s3()

# Show all files saved in my S3 bucket.
s3$list_objects(Bucket = "mybucket")
```

* You can provide your AWS credentials in environment variables, config files, and IAM roles

Another example with DynamoDB

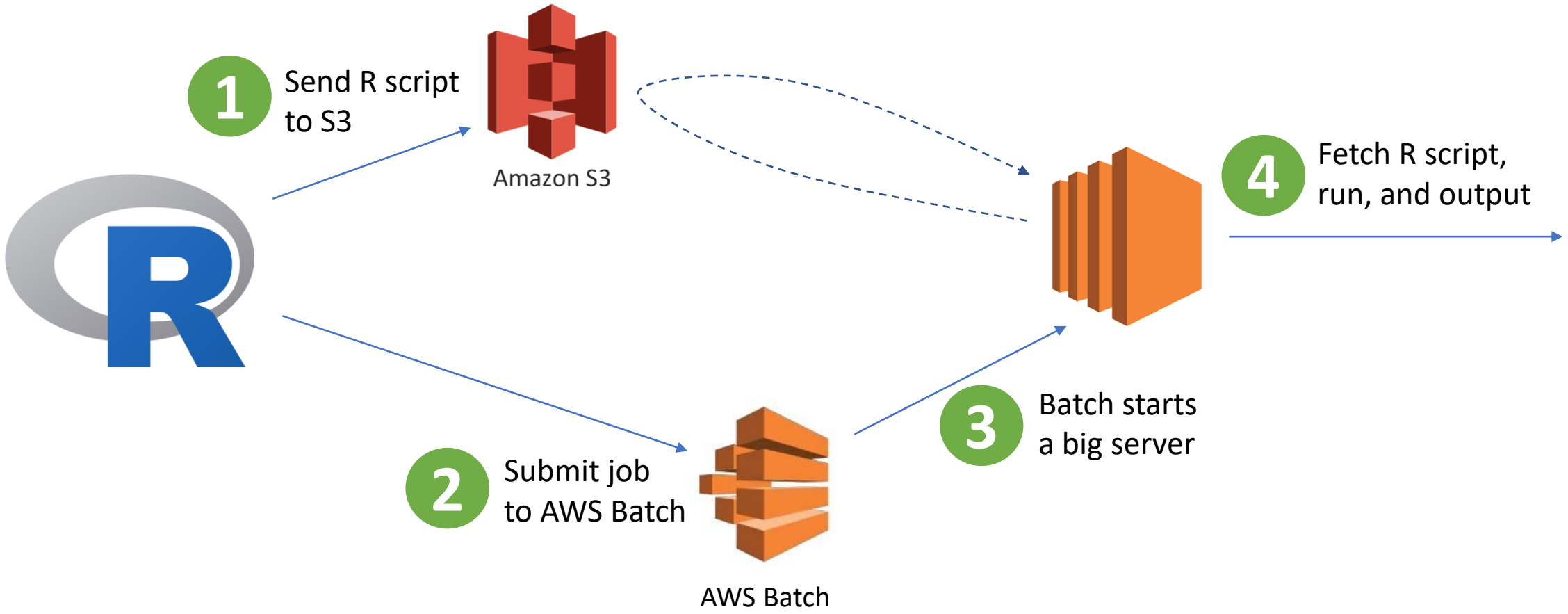
- DynamoDB is a database where you don’t have to manage servers.

```
# Write a new item to a DynamoDB table.
dynamodb$put_item(
  TableName = "Music",
  Item = list(
    Artist = c(S = "Queen"),
    SongTitle = c(S = "Bohemian Rhapsody")
  )
)
```

Batch processing with Paws

- **AWS Batch** runs jobs on big machines that stop when done
- Batch uses Docker containers to run your code
  - You'll need to create a container with your version of R and other software
- You tell Batch what kind of AWS instances to use (e.g. 72 cores)
- We're working on a library to do all this for you

Batch processing with Paws



Batch processing with Paws

1. Send R script to S3

```
s3$put_object(Body = file, Bucket = "mybucket", Key = "script.R")
```

2. Submit job to Batch

```
batch$submit_job(
    jobName = "my-job",
    jobDefinition = "run-r-script",
    jobQueue = "big-jobs",
    containerOverrides = list(command = "script.R")
)
```

3. Batch starts a big server (handled by Batch)

4. Fetch R script, run, and output (handled by Docker container)

See full code at https://github.com/davidkretch/paws_gasp_2019

Where to find out more

Install:

```
install.packages("paws")
```

Documentation:

https://paws-r.github.io

GitHub:

https://www.github.com/paws-r/paws