

DATA SCIENCE DATAOPS IN THE MICROSOFT CLOUD



strategy



insight



engineering



cloud

endjin

work smarter



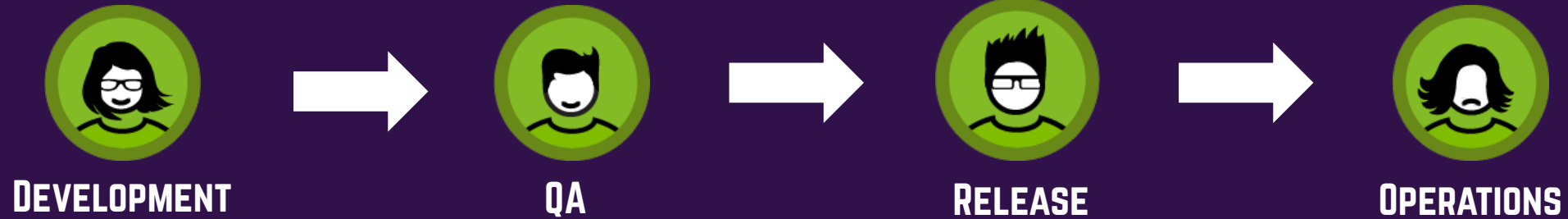
**ACCELERATE
BUSINESS INNOVATION
WITH THE
MICROSOFT CLOUD**

FOR EXAMPLE:

- INNOVATION AND NEW PRODUCT DEVELOPMENT CONSULTANCY
 - APPLICATION PROOF OF CONCEPT & PROTOTYPING
 - CLOUD SYSTEMS ARCHITECTURE GUIDANCE & SUPPORT
- ➔ **BIG DATA, ADVANCED ANALYTICS & VISUALIZATION**

endjin

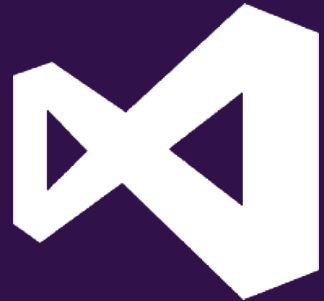
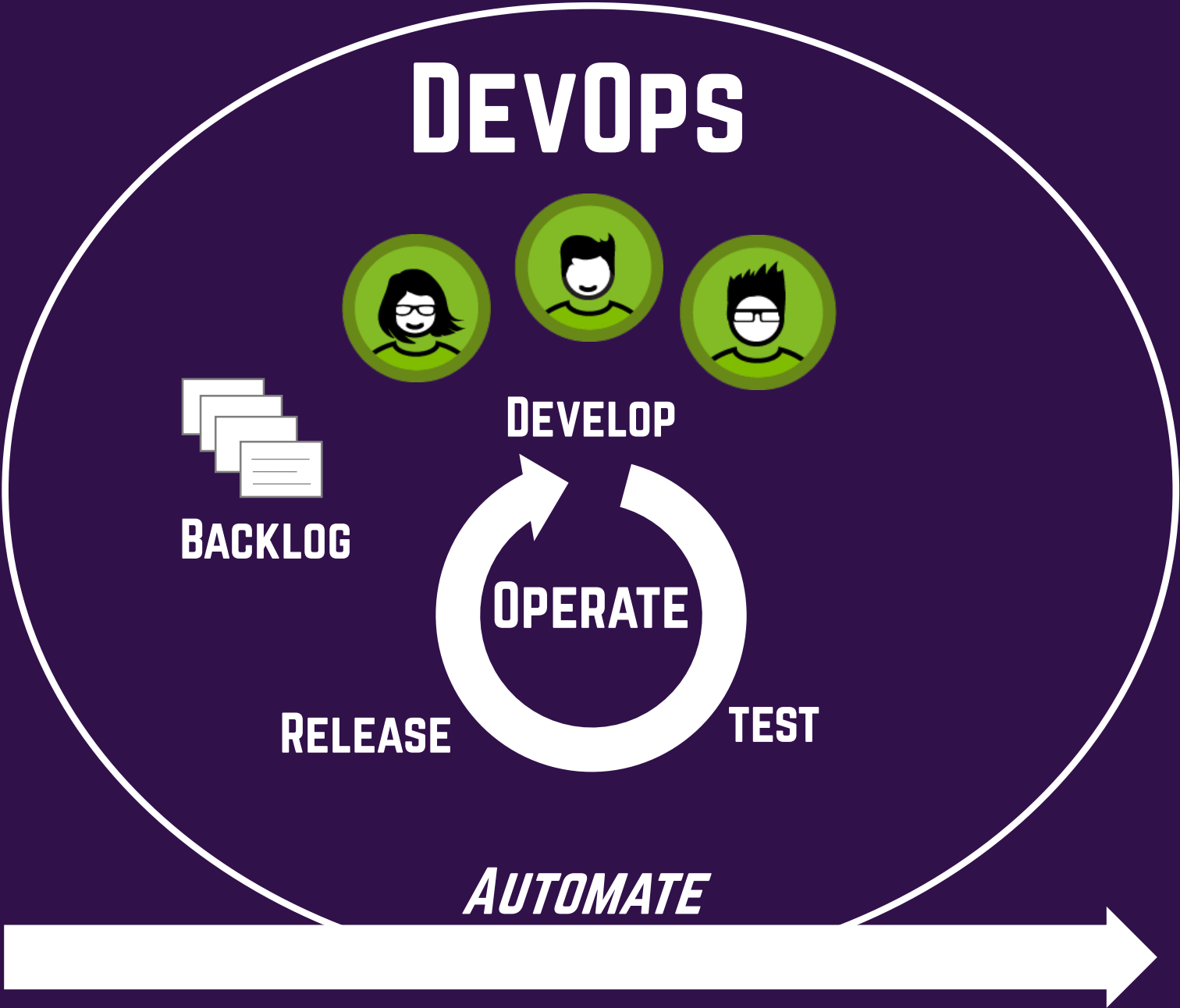
HOW IT WAS...



NO FEEDBACK
BIG RELEASES
NO ACCOUNTABILITY
LACK OF AGILITY
WASTE

SOLVED!

DEVOPS



VSTS

THE DATA SCIENCE PROBLEM TODAY....



R R R



DATA SCIENTISTS

R? R?



DEVOPS

BACKLOG?
VERSION CONTROL?
QUALITY?
PERFORMANCE?
RELEASE?
AUTH?

WHAT IF....


**R
DEVELOPMENT**



**VERSION CONTROL
TEST
RELEASE**




OPERATE

Azure Machine Learning
Microsoft R Server
R Tools for Visual Studio
R Studio



Azure Machine Learning
Deploy R
API Management
Active Directory

DEMO

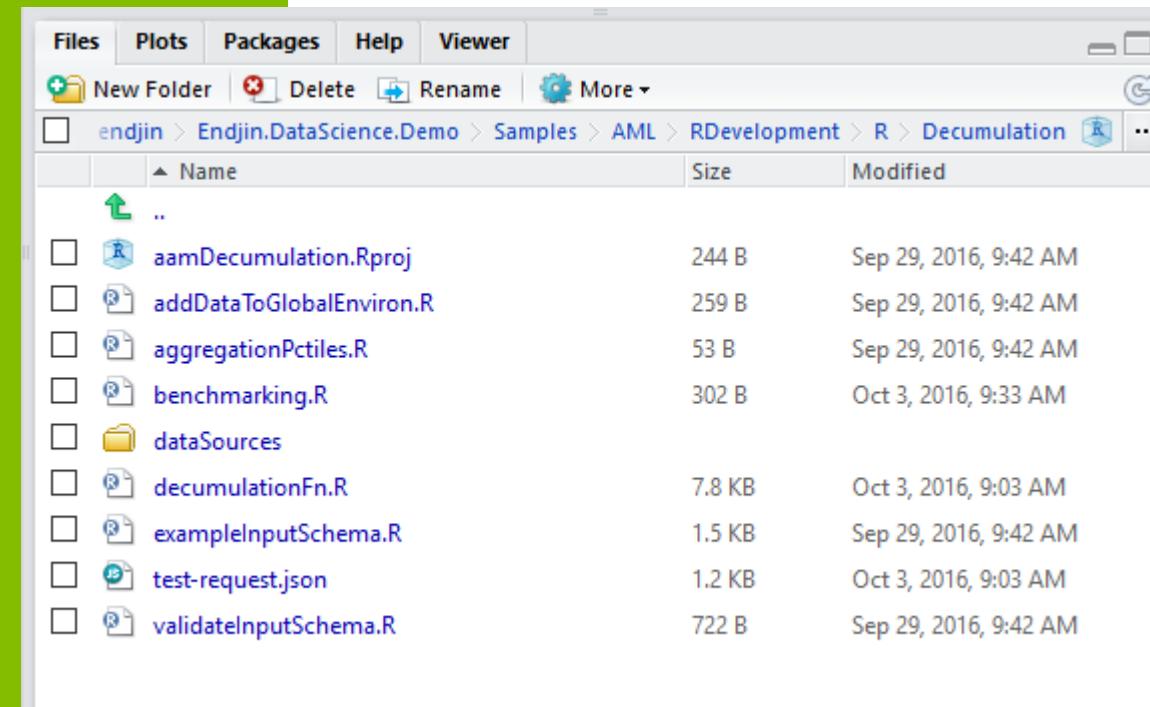
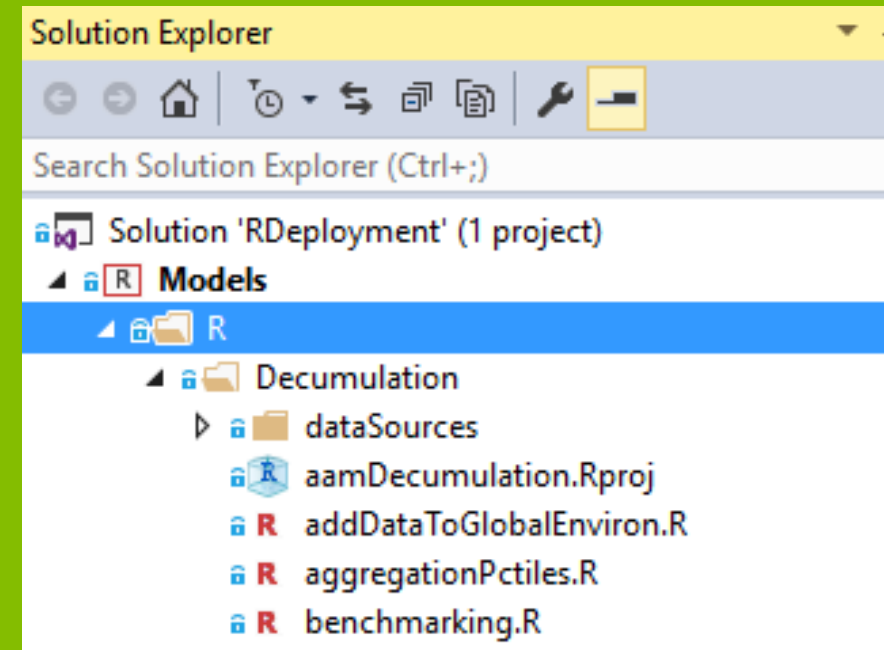
Automated R deployments with VSTS:

<https://vimeo.com/159813974>

Automated R unit tests with VSTS:

<https://vimeo.com/167279100>

TAKE AN R MODEL IN R STUDIO OR VISUAL STUDIO



ADD SOME TESTS...

```
test-sum.R  # X
context("sum")

test_that("sum adds positive numbers correctly", {
  expect_equal(sum(1, 1), 2)
  expect_equal(sum(1, 2), 3)
  expect_equal(sum(2, 2), 4)
  expect_equal(sum(2, 3), 5)
})
```

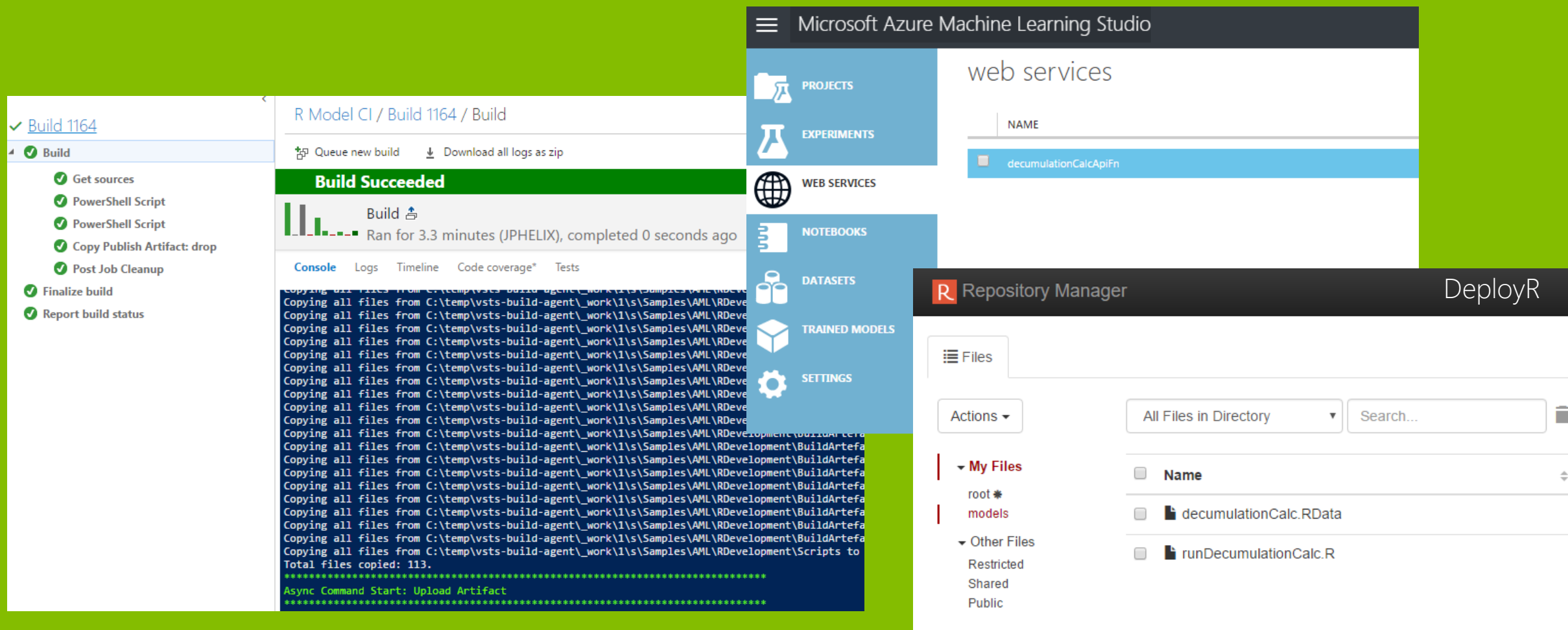
ADD A DEPLOYMENT DEFINITION FILE...

```
deploy.json  X
Schema: <No Schema Selected>
1  [
2  {
3    "script": "R/Decumulation/decumulationFn.R",
4    "function": "decumulationCalcApiFn",
5    "inputSchema": {
6      "input": "string"
7    },
8    "outputSchema": {
9      "result": "string"
10   },
11   "packages": [ "dplyr" ]
12 }
13 ]
```

ADD MODEL TO VERSION CONTROL...

```
> git add -A  
> git commit -m "Adding model"  
> git push
```

CONTINUOUS INTEGRATION AND DEPLOYMENT FOR R WITH VSTS...



The image displays a comprehensive CI/CD pipeline for R, featuring VSTS for build management, Azure ML Studio for service deployment, and DeployR for model hosting.

VSTS Build Log (Build 1164):

- Build:** Succeeded
- Steps:** Get sources, PowerShell Script, Copy Publish Artifact: drop, Post Job Cleanup, Finalize build, Report build status.
- Build Duration:** 3.3 minutes (JPHELIX), completed 0 seconds ago.
- Console Output:** Shows multiple 'Copying all files from...' messages and a total of 113 files copied. Ends with 'Async Command Start: Upload Artifact'.

Microsoft Azure Machine Learning Studio (web services):

- Service Name:** decumulationCalcApiFn

DeployR Repository Manager:

- Files:** My Files (root, models), Other Files (Restricted, Shared, Public).
- Actions:** All Files in Directory, Search...
- Model List:** decumulationCalc.RData, runDecumulationCalc.R

DATA SCIENCE DATAOPS

PROJECT MANAGEMENT

MANAGE PRIORITIES
WORK ITEM TRACKING
PROGRESS REPORTING

DEVELOPMENT

R MODELS
R PACKAGES
AML EXPERIMENTS

QUALITY ASSURANCE

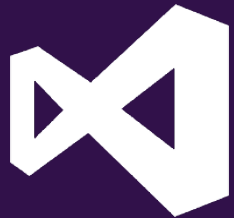
UNIT TESTING
MODEL VALIDATION
PERFORMANCE TESTING

RELEASE MANAGEMENT

MODEL DEPLOYMENT
PACKAGE MANAGEMENT
POLICIES
AUDIT

OPERATIONS

HOSTING
SCALE
AUTHENTICATION
METERING
THROTTLING
MONITORING



VSTS

+



Azure Machine
Learning

DeployR

+



API Management

+

endjin