

Azure Synapse for Actuaries

Solvency II Use Case

October 28, 2021

Presenters:

Brent Carpenetti; Senior Actuary, Hannover Re
Hiram Fleitas; Senior Customer Engineer, Microsoft

SOA Antitrust Compliance Guidelines

Active participation in the Society of Actuaries is an important aspect of membership. While the positive contributions of professional societies and associations are well-recognized and encouraged, association activities are vulnerable to close antitrust scrutiny. By their very nature, associations bring together industry competitors and other market participants. The United States antitrust laws aim to protect consumers by preserving the free economy and prohibiting anti-competitive business practices; they promote competition. There are both state and federal antitrust laws, although state antitrust laws closely follow federal law. The Sherman Act, is the primary U.S. antitrust law pertaining to association activities. The Sherman Act prohibits every contract, combination or conspiracy that places an unreasonable restraint on trade. There are, however, some activities that are illegal under all circumstances, such as price fixing, market allocation and collusive bidding.

There is no safe harbor under the antitrust law for professional association activities. Therefore, association meeting participants should refrain from discussing any activity that could potentially be construed as having an anti-competitive effect. Discussions relating to product or service pricing, market allocations, membership restrictions, product standardization or other conditions on trade could arguably be perceived as a restraint on trade and may expose the SOA and its members to antitrust enforcement procedures.

While participating in all SOA in person meetings, webinars, teleconferences or side discussions, you should avoid discussing competitively sensitive information with competitors and follow these guidelines:

- **Do not** discuss prices for services or products or anything else that might affect prices
- **Do not** discuss what you or other entities plan to do in a particular geographic or product markets or with particular customers.
- **Do not** speak on behalf of the SOA or any of its committees unless specifically authorized to do so.
- **Do** leave a meeting where any anticompetitive pricing or market allocation discussion occurs.
- **Do** alert SOA staff and/or legal counsel to any concerning discussions
- **Do** consult with legal counsel before raising any matter or making a statement that may involve competitively sensitive information.

Adherence to these guidelines involves not only avoidance of antitrust violations, but avoidance of behavior which might be so construed. These guidelines only provide an overview of prohibited activities. SOA legal counsel reviews meeting agenda and materials as deemed appropriate and any discussion that departs from the formal agenda should be scrutinized carefully. Antitrust compliance is everyone's responsibility; however, please seek legal counsel if you have any questions or concerns.

Presentation Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented. Attendees should note that the sessions are audio-recorded and may be published in various media, including print, audio and video formats without further notice.

What is Solvency II?

Solvency II is a risk-based capital regime, similar in concept to Basel II, implemented for European insurers on 1 January 2016. The risk-based capital requirement, the Solvency Capital Requirement (SCR), is calculated using either the standard formula; an approved internal model; or a mixture of both. The calculation is calibrated to a 99.5 per cent confidence level over a 1-year period.

What is Synapse Analytics?

Synapse is a limitless analytics service that brings together data integration, enterprise data warehousing, and big data analytics. It gives you the freedom to query data on your terms, using either serverless or dedicated resources—at scale. Azure Synapse brings these worlds together with a unified experience to ingest, explore, prepare, and serve data for immediate BI and machine learning needs.

Why?

10M+ Simulations * 100s Perspectives

- Spreadsheets cannot manage the data size.
- Database servers lack adv. statistical functions.
- R-language and library versions hinder collaboration between actuaries.

Business processes often experience delays

- Excel VBA(s) fall short of real automation.
- Actuarial code is not programmatically DRY (Don't Repeat Yourself).
- Actuaries lack collaboration with IT.

Ref: <https://support.microsoft.com/office/excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3>

Why?

Legacy workflows do not drive performance

- Process owners are consumed with operations.
- Little availability to evaluate outcomes.
- Teams are unable to achieve new opportunities.

How?

Modernize intelligent R & SQL

- R is best used for model or statistics heavy workloads
- SQL is best used for data management workloads

Collaborate with Azure DevOps or GitHub

- Code management to decrease knowledge silos
- Improve activity visibility and documentation

Automate with Azure Data Factory or Synapse Pipelines

- Expose flows to teams
- Enable real-time monitoring and completion tracking

Unify the experience with Azure Synapse

AGENDA

1. Demo intelligent R & SQL
2. Demo Azure DevOps
3. Demo Azure Data Factory
4. Demo Azure Synapse
5. What's Next & Resources

DEMO 1

Intelligent R & SQL

- a) R Desktop: “Bad-Case Scenario”
- b) R on the Server: An interim solution
- c) R in the database & Azure Data Studio

ImpactSOA - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

cubist_example.R*

```
1 #required
2 install.packages('mlbench', type='binary')
3 install.packages('Cubist', type='binary')
4
5 #cubist example
6 library(Cubist)
7 library(mlbench)
8 data(BostonHousing) ←
9
10 ## 1 committee, so just an M5 fit:
11 mod1 <- cubist(x = BostonHousing[, -14], y = BostonHousing$medv)
12 mod1
13
14 ## Now with 10 committees
15 mod2 <- cubist(x = BostonHousing[, -14], y = BostonHousing$medv, committees = 10)
16 mod2
```

(Top Level)

Console Terminal Jobs

R 3.5.2 · ~/Desktop/ImpactSOA/

> mod1

```
Call:
cubist.default(x = BostonHousing[, -14], y = BostonHousing$medv)

Number of samples: 506
Number of predictors: 13

Number of committees: 1
Number of rules: 4
```

>

```
> ## Now with 10 committees
> mod2 <- cubist(x = BostonHousing[, -14], y = BostonHousing$medv, committees = 10)
> mod2
```

```
Call:
cubist.default(x = BostonHousing[, -14], y = BostonHousing$medv, committees = 10)

Number of samples: 506
Number of predictors: 13

Number of committees: 10
Number of rules per committee: 4, 6, 4, 6, 6, 7, 7, 7, 4, 5
```

Run Source

Environment History Connections Tutorial

Import Dataset 221 MiB Global Environment

Data

BostonHousing	506 obs. of 14 variables
mod1	List of 14
mod2	List of 14

List

Run

Source

Environment History Connections Tutorial

Import Dataset 221 MiB Global Environment

Data

BostonHousing	506 obs. of 14 variables
mod1	List of 14
mod2	List of 14

Files Plots Packages Help Viewer

Cubist

R: Fit a Cubist model Find in Topic

Quinlan. C4.5: Programs For Machine Learning (1993b) Morgan Kaufmann Publishers Inc. San Francisco, CA

Wang and Witten. Inducing model trees for continuous classes. Proceedings of the Ninth European Conference on Machine Learning (1997) pp. 128-137

<http://rulequest.com/cubist-info.html>

See Also

[cubistControl\(\)](#), [predict.cubist\(\)](#), [summary.cubist\(\)](#), [dotplot.cubist\(\)](#), [caret::train\(\)](#)

Examples

```
library(mlbench)
data(BostonHousing)

## 1 committee, so just an M5 fit:
mod1 <- cubist(x = BostonHousing[, -14], y = BostonHousing$medv)
mod1

## Now with 10 committees
mod2 <- cubist(x = BostonHousing[, -14], y = BostonHousing$medv, committees = 10)
mod2
```

[Package Cubist version 0.3.0 [Index](#)]

Diagram A

Desktop

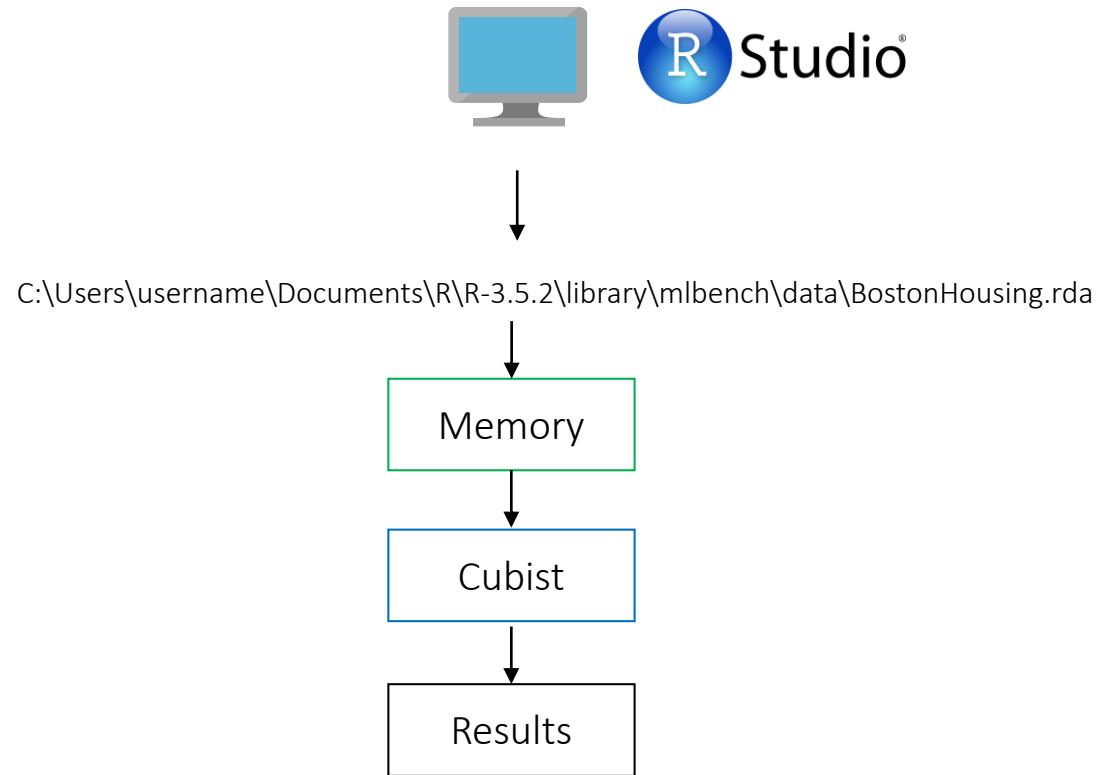
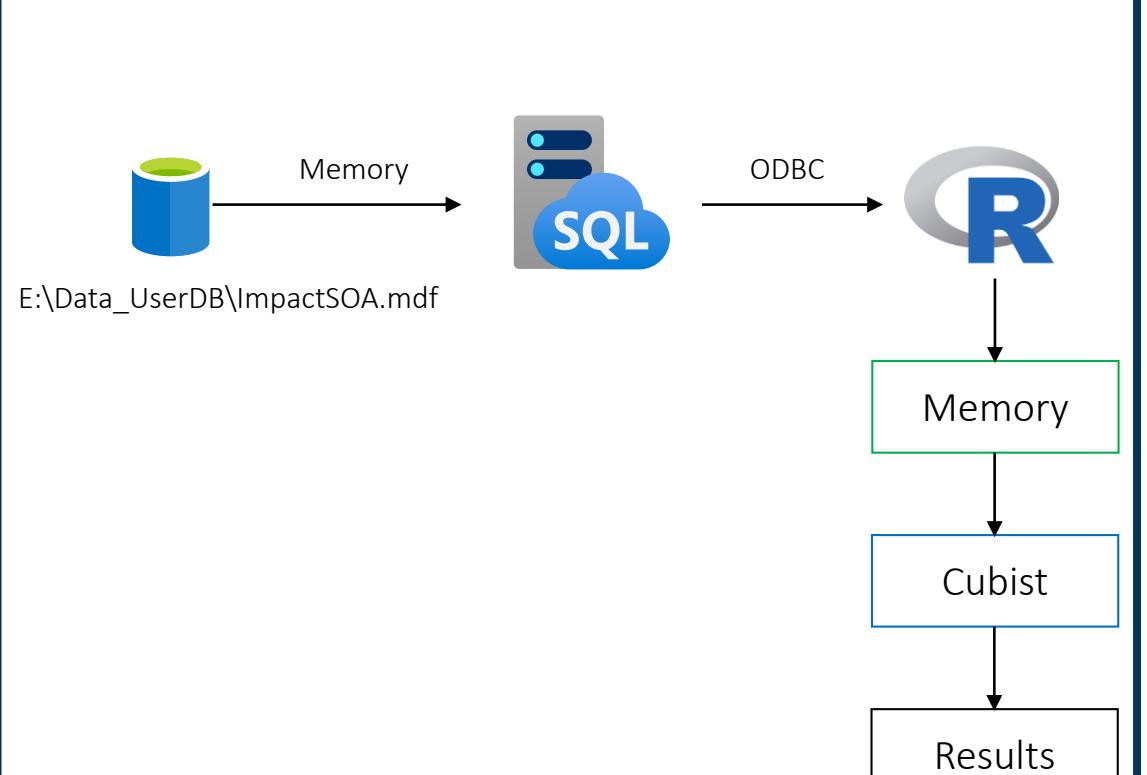


Diagram B

Server



ImpactSOA - RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

on_server_example.R* cubist_example.R

Source on Save Run Source

1 #required
2 install.packages('RODBC', type='binary')
3 install.packages('Cubist', type='binary')
4
5 #cubist example
6 library(Cubist)
7 library(RODBC)
8 connect <- odbcDriverConnect('driver={SQL Server};server=localhost;database=ImpactSOA;trusted_connection=true')
9 BostonHousing <- sqlQuery(channel=connect,query='select * from dbo.BostonHousing')
10
11 ## 1 committee, so just an M5 fit:
12 mod1 <- cubist(x = BostonHousing[, -14], y = BostonHousing\$medv)
13 mod1
14
15 ## Now with 10 committees
16 mod2 <- cubist(x = BostonHousing[, -14], y = BostonHousing\$medv, committees = 10)
17 mod2
18

10:1 (Top Level) R Script

Console Terminal Jobs

R 3.5.2 ~/Desktop/ImpactSOA/ > mod1

Call:
cubist.default(x = BostonHousing[, -14], y = BostonHousing\$medv)

Number of samples: 506
Number of predictors: 13

Number of committees: 1
Number of rules: 4

> mod2

Call:
cubist.default(x = BostonHousing[, -14], y = BostonHousing\$medv, committees = 10)

Number of samples: 506
Number of predictors: 13

Number of committees: 10
Number of rules per committee: 4, 6, 4, 6, 6, 7, 7, 7, 4, 5

>

Environment History Connections Tutorial

Import Dataset 194 MiB Global Environment

Data

BostonHousing	506 obs. of 14 variables
mod1	List of 14
mod2	List of 14

Files Plots Packages Help Viewer

R: Query an ODBC Database Find in Topic

sqlQuery {RODBC}

R Documentation

Query an ODBC Database

Description

Submit an SQL query to an ODBC database, and retrieve the results.

Usage

```
sqlQuery(channel, query, errors = TRUE, ..., rows_at_time)
sqlGetResults(channel, as.is = FALSE, errors = FALSE,
max = 0, bufsize = 1000,
nullstring = NA_character_, na.strings = "NA",
believeNRows = TRUE, dec = getOption("dec"),
stringsAsFactors = default.stringsAsFactors())
```

Arguments

channel	connection handle as returned by odbcConnect .
query	any valid SQL statement.

Diagram B

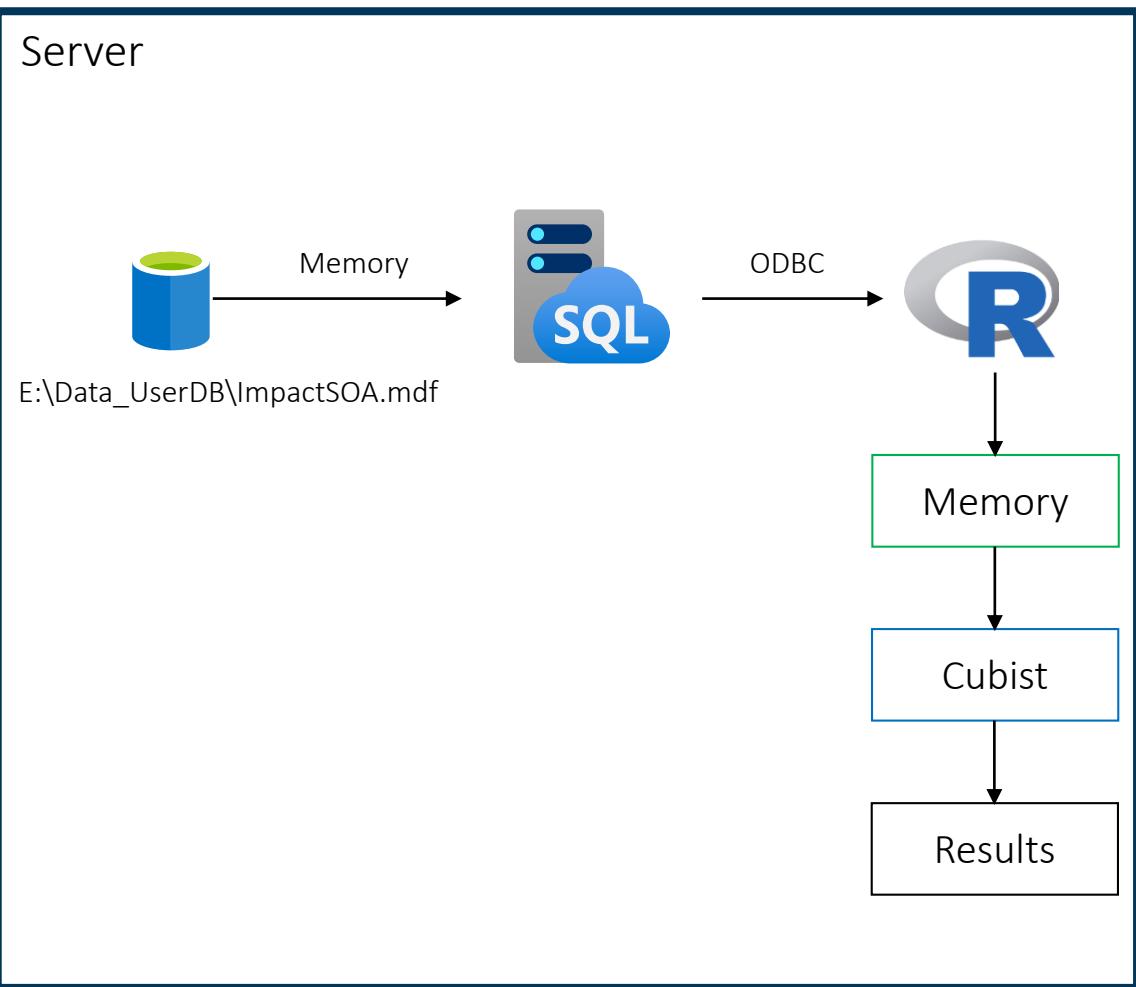
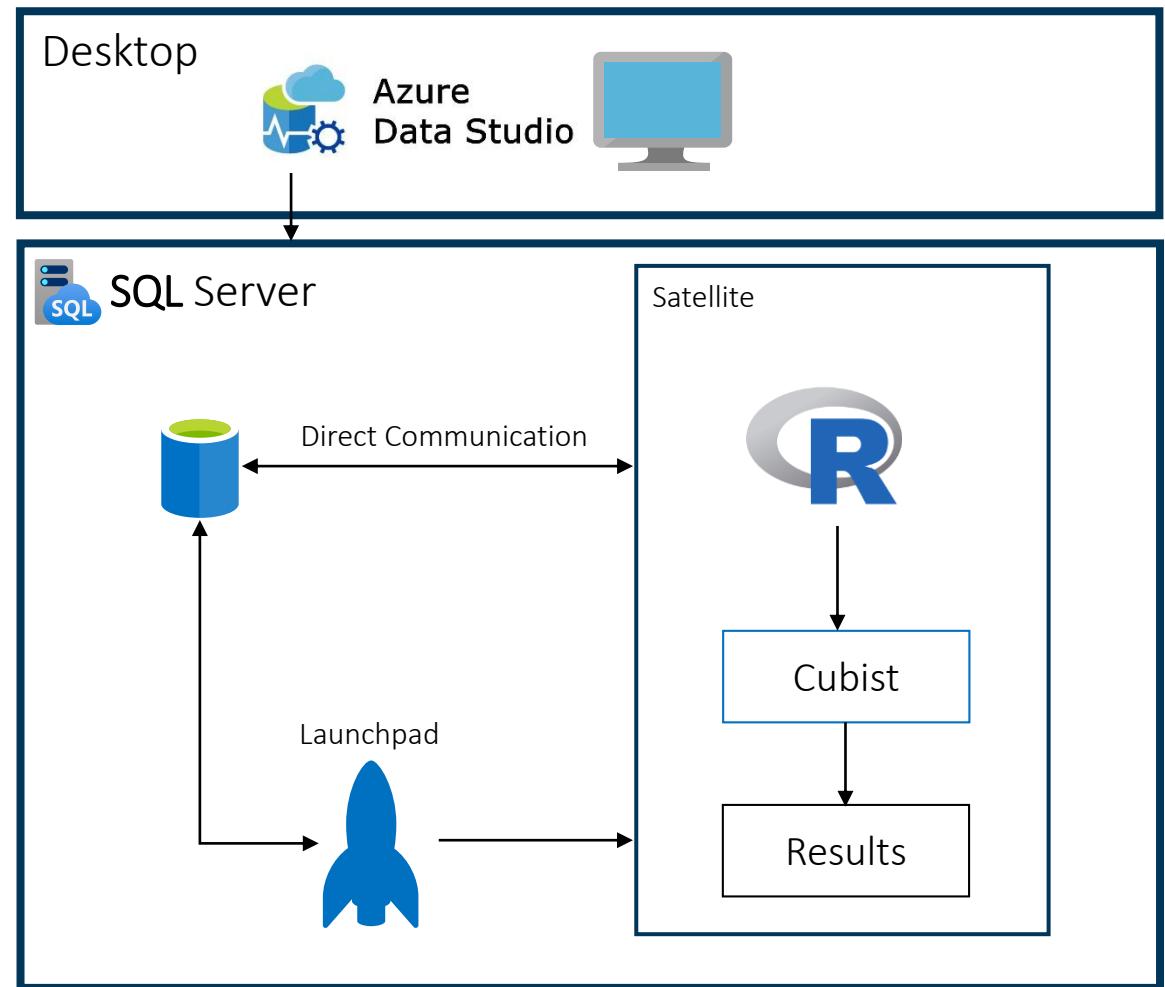


Diagram C



File Edit View Help • SQLEngine_AzureDataStudio_Sample.ipynb internal_model (Workspace) - Azure Data Studio

SQLEngine_AzureDataStudio_Sample.ipynb

ImpactSOA_repo > SOA-BasicsForActuaries > ImpactSOA_CodeSamples > SQLEngine_AzureDataStudio_Sample.ipynb

+ Cell ▾ Run all Kernel SQL Attach to [redacted]

[21] 1 CREATE EXTERNAL LIBRARY Cubist FROM (CONTENT = 'E:/ServerPackageDirectory/Cubist_0.2.3.zip') WITH (LANGUAGE = 'R');
2 CREATE EXTERNAL LIBRARY mlbench FROM (CONTENT = 'E:/ServerPackageDirectory/mlbench_2.1-1.zip') WITH (LANGUAGE = 'R');

Commands completed successfully.

Total execution time: 00:00:00.906

[23] 1 exec sp_execute_external_script
2 @language = N'R',
3 @script = N'
4 library(Cubist)
5 library(mlbench)
6 data(BostonHousing)
7
8 ## 1 committee, so just an M5 fit:
9 mod1 <- cubist(x = BostonHousing[, -14], y = BostonHousing\$medv)
10 print(mod1)
11
12 ## Now with 10 committees
13 mod2 <- cubist(x = BostonHousing[, -14], y = BostonHousing\$medv, committees = 10)
14 print(mod2)
15 ';

STDERR message(s) from external script: Warning messages: 1: package 'Cubist' was built under R version 3.5.3 2: package 'mlbench' was built under R version 3.5.3

STDOUT message(s) from external script: Call: cubist.default(x = BostonHousing[, -14], y = BostonHousing\$medv) Number of samples: 506 Number of predictors: 13 Number of committees: 1 Number of rules: 4 Call: cubist.default(x = BostonHousing[, -14], y = BostonHousing\$medv, committees = 10) Number of samples: 506 Number of predictors: 13 Number of committees: 10 Number of rules per committee: 4, 6, 4, 6, 6, 7, 7, 7, 4, 5

Total execution time: 00:00:02.440

master* 6 △ 0 Choose SQL Language

Resources Demo 1

- <https://github.com/hfleitas/SOA-BackToBasicsAnalyticsForActuaries>
- <https://rstudio.com/products/rstudio/download>
- [Reading data from Microsoft SQL Server into R - Stack Overflow](#)
- [Install on Windows - SQL Server Machine Learning Services | Microsoft Docs](#)
- [R language extension - SQL Server Machine Learning Services | Microsoft Docs](#)
- [CREATE EXTERNAL LIBRARY \(Transact-SQL\) - SQL Server - SQL Server | Microsoft Docs](#)

DEMO 2

Azure DevOps

- a) Boards
- b) Repos: enable more rapid actuarial development
- c) Pipelines: maturing to CI/CD

FabrikamFiber Board

New Active 5/5 Staging 15/15 Deployed 3/3

+ New item

Home page (selected room)
Kat Larson
Design

Home page (no room selected)
Carlos Slattery
Spike Xamarin

Mobile (Spike)
Celeste Burton
Design

Hotels filter page
Carlos Slattery
Xamarin

Top page controls
Celeste Burton
ML Xamarin

Entry + validations
Carole Poland

Footer
Cecil Folk
ML Xamarin

Guests page
Carole Poland
ML Xamarin

Navigation menu
Carlos Slattery
AL Xamarin

Code of Conduct
Celeste Burton
General Xamarin

NFC open door
Cecil Folk
Spike Xamarin

Login page
Celeste Burton
Blocked Xamarin

Images from api
Carole Poland
General

Ambient settings
Carlos Slattery
Xamarin

Room Tab
Celeste Burton
Rooms [Detail]

Adapt some parts of UI to UWP for Desktop
Carole Poland
Blocked Xamarin

Notifications list
Carole Poland
General

Map filter
Carole Poland
General Room [List]

Hotel reviews page
Celeste Burton
Rooms [Detail]

Project settings



AW AdventureWorks Mobile

- Overview
- Boards
- Repos**
- Files
- Commits
- Pushes
- Branches
- Tags
- Pull requests**
- Pipelines
- Test Plans
- Artifacts

Project settings

Pull requests

[+ New Pull Request](#)[Mine](#) [Active](#) [Completed](#) [Abandoned](#)

Created by me

- Initialize client with .client.init**
Celeste Burton requested #238 into master
- 0
3 minutes ago

- Use latest React version**
Celeste Burton requested #230 into features/react-update

0
1 hour ago

Assigned to me

- Check returned identity for null status**
Colin Ballinger requested #212 into master
- 0
new pull request • 15 minutes ago

- [WIP] Add tests for reticulating splines**
Robin Counts requested #221 into master
- 0
4 hours ago

Assigned to my team

- Add exception mappings for disconnect**
Colin Ballinger requested #249 into master
- 5
2 comments • 15 minutes ago

- Maintain folder structure when converting isomorphs**
Robin Counts requested #234 into master
- 0
1 hour ago

- Testing configuration settings**
Robin Counts requested #239 into master
- 2
3 hours ago

- Hotfix payload to Orion**
Robin Counts requested #201 into releases/orion
- 3
5 hours ago

```
251 export interface SaveInfoBannerProps {  
252   saveInfo: SaveInfo;  
253   versionSpec: VersionSpec;  
254   interface CommitLinkProps {  
255     url: string;  
256     text: string;  
257     tooltip: string;  
258     prefixMessage: string;  
259   }  
}
```

```
251 export interface SaveInfoBannerProps {  
252   saveInfo: SaveInfo;  
253   versionSpec: VersionSpec;  
254   interface CommitLinkProps {  
255     url: string;  
256     + targetHubId: string;  
257     text: string;  
258     tooltip: string;  
259     prefixMessage: string;  
260   }  
}
```



Liam Parker

@Chloe Clark `const CommitLink = ({ prefixMessage, url, tooltip, text }:`

Chloe Clark



Can we try what we did in [User Story 567435](#)



Liam Parker

Yeah, that would be a better pattern, I'll make the change in my next update

1

Chloe Clark



Awesome!

- AdventureWorks
- Overview
- Pipelines
- Builds
- Releases
- Library
- Deployment groups

Project settings <<

Enabling feature flags for Preview Attachment and Grid Views

AdventureWorks/PackageFramework ⚡ master ↗ #889

Summary Logs Tests YAML

Job Type	Status	Duration
Windows Job	Running	1m 53s
Linux Job	Running	3m 29s
macOS Job	Running	3m 07s

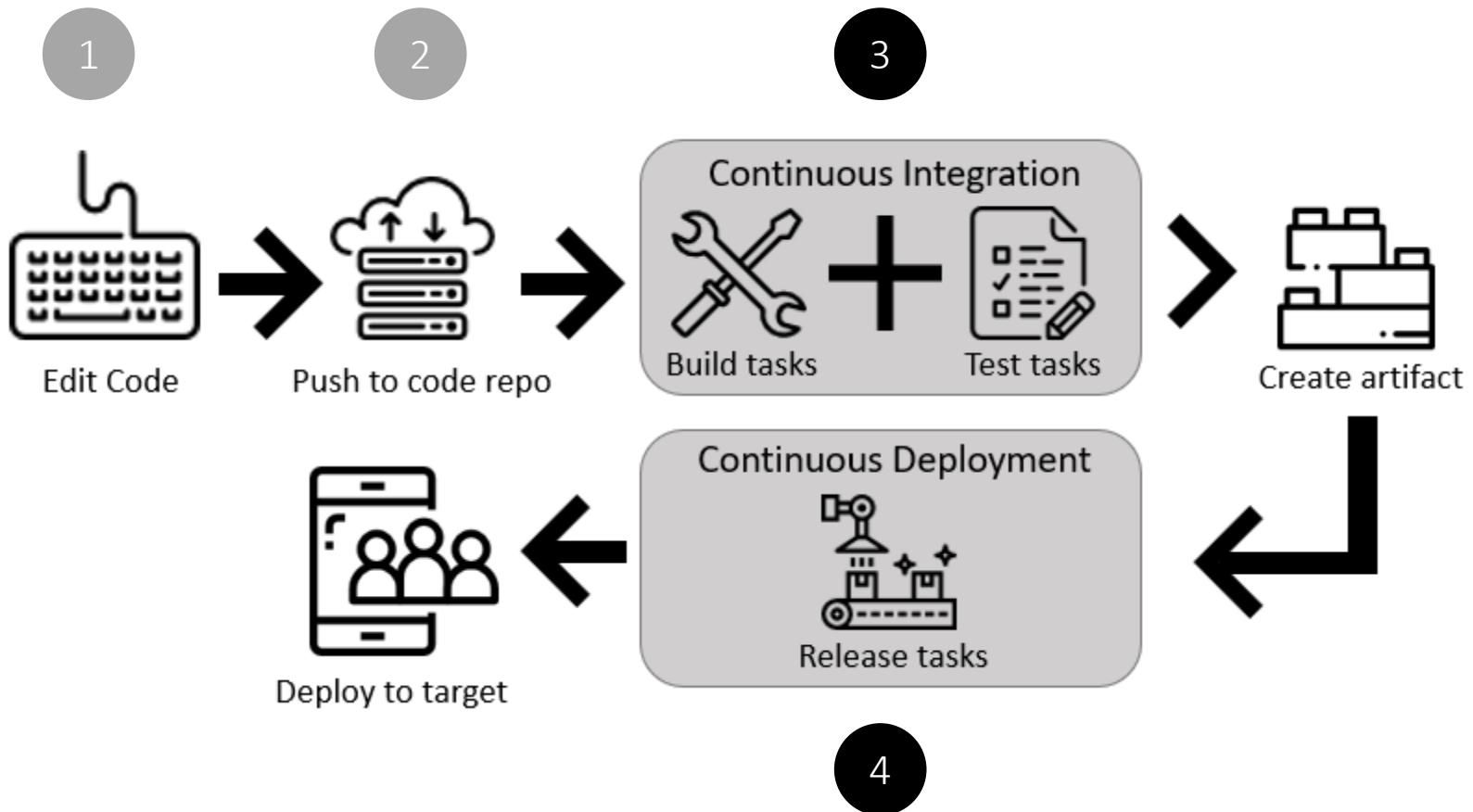
Linux Job

Agent: Hosted Linux

Task	Time
Prepare job	<1s
Initialize job	1s
Get sources	24s
Cmdline	28s
Nodetool	3s
Install dependencies	2m 31s

```
yarn install v1.7.0
$ node build/npm/preinstall.js
[1/4] 🏛 Resolving packages...
[2/4] 🚀 Fetching packages...
[3/4] ⚡ Linking dependencies...
[4/4] 🏭 Building fresh packages...
$ npm run compile
[########################################-----] 152/243
> code-oss-dev-build@1.0.0 compile ./adventureworks/build
> tsc -p tsconfig.build.json

⚡ Done in 4.89s.
$ node ./postinstall
[##] 2/2 removed './adventureworks/extensions/node_modules/typescript/lib/tsc.js'
removed './adventureworks/extensions/node_modules/typescript/lib/tsserverlibrary.d.ts'
removed './adventureworks/extensions/node_modules/typescript/lib/tsserverlibrary.js'
removed './adventureworks/extensions/node_modules/typescript/lib/typescriptServices.d.ts'
removed './adventureworks/extensions/node_modules/typescript/lib/typescriptServices_is'
```



Ref: <https://docs.microsoft.com/azure/devops/pipelines/get-started/pipelines-get-started>

Azure DevOps interface showing the Pipelines page for the dw-Cl pipeline.

The pipeline has four stages, all of which have passed successfully (indicated by green checkmarks). The stages are:

- #25 Merge branch 'main' of https://github.com/hfleitas/synapsedelta
- #24 fixed dacpac
- #7 added devops json pipelines
- #6 Merge branch 'main' of https://github.com/hfleitas/synapsedelta

The last stage was manually triggered. The pipeline was last run on Sep 15 at 43s.

Screenshot Details:
A large green arrow points to the "Pipelines" link in the left sidebar, highlighting it as the active section. The "dw-Cl" pipeline name is also highlighted with a green arrow above the pipeline table.

Description	Stages	Last Run
#25 Merge branch 'main' of https://github.com/hfleitas/synapsedelta ↳ Individual CI for main 5aeedc8c	✓	Sep 15 43s
#24 fixed dacpac ↳ Individual CI for main 07668a04	✓	Sep 15 1m 38s
#7 added devops json pipelines ↳ Individual CI for main 44bd0d62	✓	Sep 10 41s
#6 Merge branch 'main' of https://github.com/hfleitas/synapsedelta ↳ Manually triggered for main d8003ef8	✓	Sep 10 46s

Pipelines - Run 25 artifacts

https://dev.azure.com/hiramfleitas0814/dw/_build/results?buildId=25&view=artifacts&pathAsName=false&type=publishedArtifacts

Azure DevOps

hiramfleitas0814 / dw / Pipelines / dw-CL / 25 / Published artifacts

Search

dw

Overview

Boards

Repos

Pipelines

Pipelines

Environments

Releases

Library

Task groups

Deployment groups

Test Plans

Artifacts

Compliance

Project settings

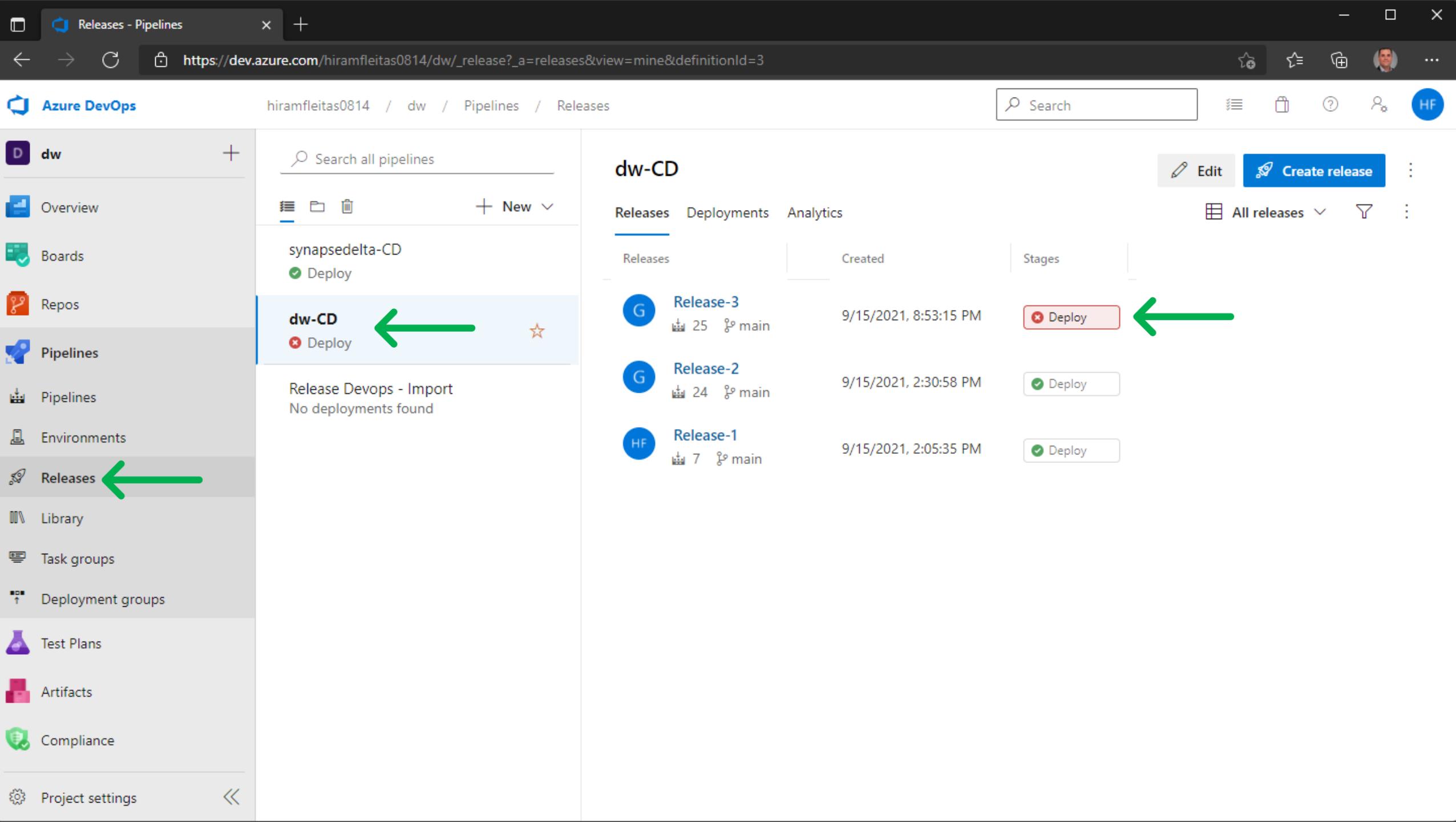
Artifacts

Published Consumed

Name Size

Name	Size
drop	18 KB
wplussynapsedw.dacpac	5 KB
wplussynapsedw.dll	4 KB
wplussynapsedw.pdb	10 KB

A green arrow points upwards from the artifact list towards the pipeline navigation bar. A blue arrow points to the 'wplussynapsedw.dacpac' file in the list.



dw-CD - Release-3 - Pipelines

https://dev.azure.com/hiramfleitas0814/dw/_releaseProgress?a=release-environment-logs&releaseId=22&environmentId=22

Azure DevOps hiramfleitas0814 / dw / Pipelines / Releases / dw-CD / Release-3 Search

D dw + dw-CD > Release-3 > Deploy Failed

Overview Boards Repos Pipelines Environments Releases Library Task groups Deployment groups Test Plans Artifacts Compliance Project settings

Pipeline Tasks Variables Logs Tests Deploy Cancel Refresh Download all logs Edit ...

Deployment process Failed

Agent job Failed · 3 errors

Agent job

Pool: Azure Pipelines · Agent: Hosted Agent Started: 9/15/2021, 8:53:39 PM ... 45s

Initialize job · succeeded 11s

Download artifact - _dw-CD - drop · succeeded 3s

Get Dacpac Script · 3 errors 30s

*** An error occurred during deployment plan generation. Deployment cannot continue.

Failed to import target model wplussynapsesdw. Detailed message Cannot connect to database when it is paused.

Cannot connect to database when it is paused.

The Azure SQL DACPAC task failed. SqlPackage.exe exited with code 1. Check out how to troubleshoot failures at <https://aka.ms/sqlazuredeployreadme#troubleshooting>

Show Dacpac Script · skipped ⓘ

Publish Dacpac (Deploy) · skipped ⓘ

Finalize Job · succeeded <1s

dw-CD - Release-2 - Pipelines

https://dev.azure.com/hiramfleitas0814/dw/_releaseProgress?a=release-environment-logs&releaseId=21&environmentId=21

Azure DevOps

dw

Overview Boards Repos Pipelines Pipelines Environments Releases Library Task groups Deployment groups Test Plans Artifacts Compliance Project settings

Publish Dacpac (Deploy)

1 2021-09-15T18:46:09.4872626Z ##[section]Starting: Publish Dacpac (Deploy)
2 2021-09-15T18:46:09.5051629Z =====
3 2021-09-15T18:46:09.5052192Z Task : Azure SQL Database deployment
4 2021-09-15T18:46:09.5052573Z Description : Deploy an Azure SQL Database using DACPAC or run scripts using SQLCMD
5 2021-09-15T18:46:09.5052865Z Version : 1.184.0
6 2021-09-15T18:46:09.5053151Z Author : Microsoft Corporation
7 2021-09-15T18:46:09.5053513Z Help : <https://docs.microsoft.com/azure/devops/pipelines/tasks/deploy/sql-azure-dacpac-deployment>
8 2021-09-15T18:46:09.5054292Z =====
9 2021-09-15T18:46:14.7066678Z Temporary inline SQL file: C:\Users\VssAdministrator\AppData\Local\Temp\tmp9349.tmp
10 2021-09-15T18:46:14.7426777Z Invoke-Sqlcmd -ServerInstance "hfpocts1.sql.azuresynapse.net" -Database "wpplussynapsedw" -Username "sqladminuser" -Pa
11 2021-09-15T18:46:26.1209698Z DACPAC file path: D:\a\r1\a_dw-CI\drop\wpplussynapsedw.dacpac
12 2021-09-15T18:46:26.3561949Z ##[command]"D:\a_tasks\SqlAzureDacpacDeployment_ce85a08b-a538-4d2b-8589-1d37a9ab970f\1.184.0\vswhere.exe" -version [1
13 2021-09-15T18:46:26.4508936Z "C:\Program Files\Microsoft SQL Server\150\DAC\bin\SqlPackage.exe" /Action:Publish /SourceFile:"D:\a\r1\a_dw-CI\drop\wpplussynapsedw.dacpac" /TargetConnectionString:"Data Source=hfpocts1.sql.azuresynapse.net;Initial Catalog=wpplussynapsedw;User Id=sqladminuser;Password=sqladminuser;Trusted_Connection=False;Encrypt=True;Connection Timeout=30"
14 2021-09-15T18:46:26.7992890Z Publishing to database 'wpplussynapsedw' on server 'hfpocts1.sql.azuresynapse.net'.
15 2021-09-15T18:46:30.7152019Z Initializing deployment (Start)
16 2021-09-15T18:47:39.4776453Z Initializing deployment (Complete)
17 2021-09-15T18:47:39.4794978Z Analyzing deployment plan (Start)
18 2021-09-15T18:47:39.5713872Z Analyzing deployment plan (Complete)
19 2021-09-15T18:47:39.5736200Z Updating database (Start)
20 2021-09-15T18:47:42.4104972Z Creating Table [dbo].[lastchanges]...
21 2021-09-15T18:47:43.1380807Z Update complete.
22 2021-09-15T18:47:45.2745342Z Updating database (Complete)
23 2021-09-15T18:48:07.3741358Z Successfully published database.
24 2021-09-15T18:48:07.3766129Z Time elapsed 0:01:40.62
25 2021-09-15T18:48:07.4838569Z ##[section]Finishing: Publish Dacpac (Deploy)
26

CONNECTIONS

...

SERVERS

- hfpocws1.database.windows.net, <default> (Hiram Fleitas - hir...
 - Databases
 - System Databases
 - hfpocws1p1
 - wplussynapsedw
 - Tables
 - dbo.Character
 - dbo.dateformats
 - dbo.dateformats2
 - dbo.devs
 - dbo.Episode
 - dbo.extDevs
 - dbo.heapTable
 - dbo.lastchanges
 - Columns
 - Constraints
 - Indexes
 - Statistics
 - dbo.Letters

AZURE

SQL SERVER BIG DATA CLUSTERS

CENTRAL MANAGEMENT SERVERS

master 0 0



Azure Data Studio

[New](#)[Open file...](#)[Open folder...](#)

Create a connection

Connect to a database instance through the connection dialog.



Run a query

Interact with data through a query editor.



Create a notebook

Build and run notebooks.



Deploy a server

Create a new instance of a

EXTENSIONS: MARKETPLACE ... Welcome Extension: SQL Database Projects X

project

SQL Database Projects 0.12.0
The SQL Database Projects extension for Az...
Microsoft

SQL Database Projects

microsoft.sql-database-projects

Microsoft | Repository | License | v0.12.0

The SQL Database Projects extension for Azure Data Studio allows users to develop ...

[Disable](#) [Uninstall](#) This extension is enabled globally.

This extension is recommended by Azure Data Studio.

[Details](#) [Feature Contributions](#) [Dependencies](#)

Microsoft SQL Server Database Projects for Azure Data Studio

Microsoft SQL Server Database Projects for Azure Data Studio includes:

Database Projects

The Database Projects extension provides a way to design, edit, and publish schemas for SQL databases from a source controlled project.

Please report issues and feature requests [here](#).

Getting Started with Database Projects

Resources Demo 2

- [Understand what you get with Azure Boards - Azure Boards | Microsoft Docs](#)
- [Azure Boards | Microsoft Azure](#)
- [About project boards - GitHub Docs](#)
- [Azure Repos – Git Repositories | Microsoft Azure](#)
- [Azure Pipelines | Microsoft Azure](#)
- [<https://microsoft.github.io/PartsUnlimited/iac/200.2x-1aC-CDAzureSQLdbwithVSTSandVS.html>](#)
- [<https://docs.microsoft.com/azure/devops/pipelines/targets/azure-sqldb>](#)
- [<https://docs.microsoft.com/sql/azure-data-studio/download-azure-data-studio>](#)

DEMO 3

Azure Synapse

- a) Pipelines for Orchestrating Automation
- b) Transformations: Data Flows (Spark in low-code)
- c) Data movement: Including on-premises sources

rig - Azure Data Factory

https://adf.azure.com/en-us/authoring/pipeline/Primer?factory=%2Fsubscriptions%2Feaab21d5-8ecd-4ef0-a0c4-92fac2e22875%2FresourceGroups%2Frig%2Fprovider...

Microsoft Azure | Data Factory > rig

Search

master branch | Validate all | Save all | Publish

Factory Resources | Primer | Trollhunters

Filter resources by name

Pipeline (10)

- notify
- Primer (selected)
- Scale
- TaxiDemo
- Trollhunters

50-powerbi (2)

Ext (3)

Dataset (20)

Data flows (4)

Power Query (Preview) (0)

Templates (0)

Primer Pipeline Diagram:

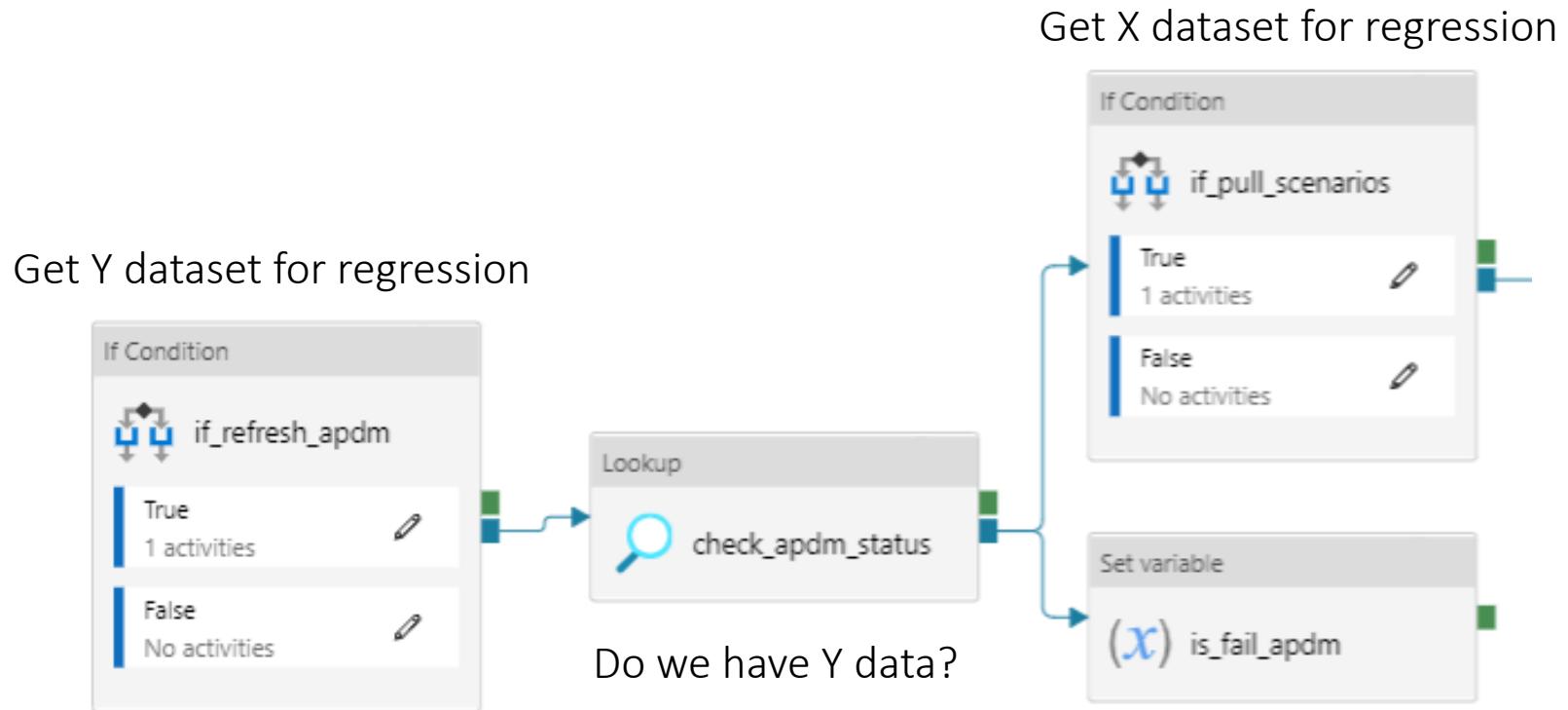
```
graph LR; GetStatus[GetStatus] --> IfCondition{If Condition: IfPausedThenOnline}; IfCondition -- True --> UntilOnline[Until: UntilOnline]; UntilOnline --> execTrollhunters[Execute Pipeline: execTrollhunters]; execTrollhunters --> GetStatusEnd[GetStatusEnd]; execTrollhunters --> Notify[Notify];
```

Parameters

New | Delete

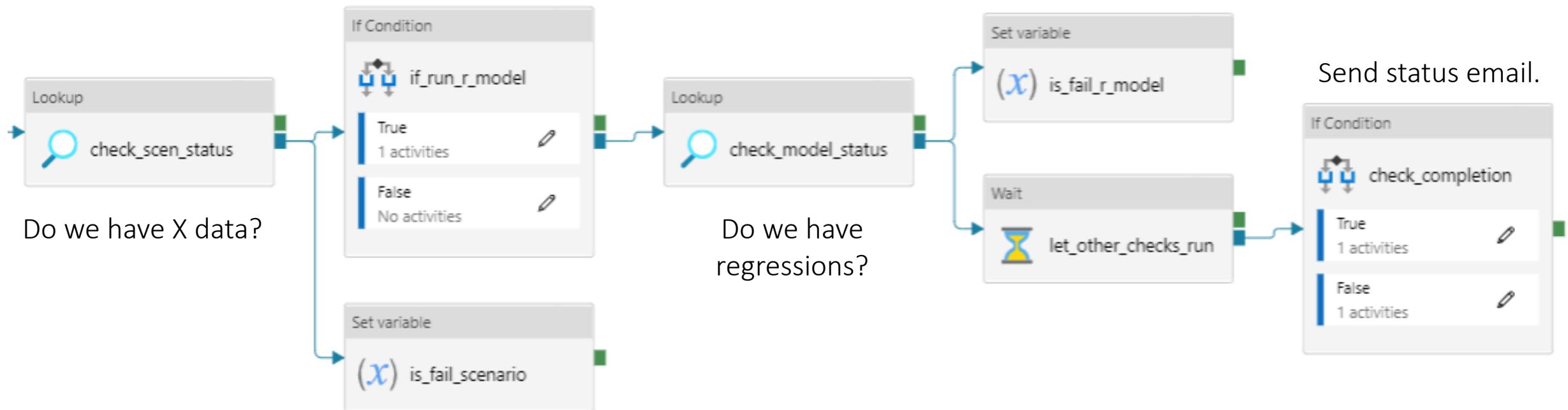
Name	Type	Default value
SubscriptionID	String	eaab21d5-8ecd-4ef0-a0c4-92fa
ResourceGroup	String	dw
Server	String	hiramdw
DW	String	dw
SLO	String	DW1000c
Notify	String	azure@fleitasarts.com

Pipeline: Publish Solvency II Model

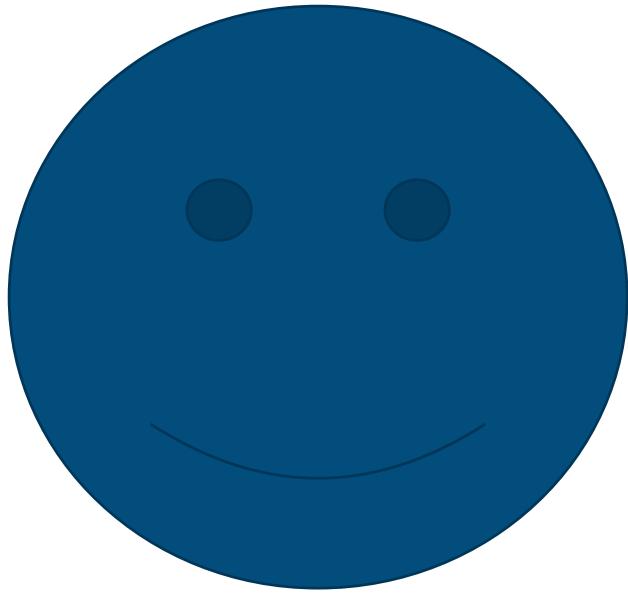


Cont'd

Run R regression models



Solvency II Report



Microsoft Azure | Synapse Analytics ▶

Brent.Carpentier

Pipeline runs

Triggered Debug

 Rerun

 Cancel 

 Refresh

[Edit column](#)

Lis

Gant

 Search by run ID or name

Local time : Last 30 days

Pipeline name : A

Status : A

Runs : Latest runs

Triggered by : All

 Add filter

 Copy filters

Pipeline name	Run start ↑	Run end	Duration	Triggered by	Status	Error	Run	Parameters	Actions
> enterprise	9/23/21, 1:44:17 PM	9/23/21, 2:26:52 PM	00:42:35	01135968-0582-40da-90	✓ Succeeded		Rerun (Latest)	[@]	
> general	9/23/21, 11:58:34 AM	9/23/21, 2:27:56 PM	02:29:22	Manual trigger	✓ Succeeded		Rerun (Latest)	[@]	
> run	9/20/21, 7:21:44 PM	9/21/21, 3:21:54 AM	08:00:09	0f09b118-6d92-49c7-b71	✗ Failed	✖	Rerun (Latest)	[@]	
> pipeline	9/20/21, 4:51:33 PM	9/21/21, 3:23:01 AM	10:31:28	Manual trigger	✓ Succeeded		Rerun (Latest)	[@]	
> test	9/17/21, 10:49:53 AM	9/17/21, 12:45:41 PM	01:55:48	Manual trigger	✓ Succeeded		Rerun (Latest)		
test	9/16/21, 9:31:26 AM	9/16/21, 11:02:34 AM	01:31:08	Manual trigger	✗ Cancelled	✖	Original		
test	9/15/21, 11:55:55 AM	9/15/21, 1:50:43 PM	01:54:48	Manual trigger	✓ Succeeded		Original		
test	9/15/21, 9:13:33 AM	9/15/21, 11:13:47 AM	02:00:14	Manual trigger	✓ Succeeded		Original		
relabel	9/1/21, 2:24:59 PM	9/1/21, 9:59:07 PM	07:34:07	ce012cf7-7c18-402f-834f	✓ Succeeded		Original	[@]	
enterprise	9/1/21, 12:07:27 PM	9/1/21, 2:23:29 PM	02:16:02	872a56e5-7c14-44a1-8a:	✓ Succeeded		Original	[@]	
general	9/1/21, 12:04:00 PM	9/1/21, 2:24:39 PM	02:20:38	e0dd1d61-e84c-4d05-92	✓ Succeeded		Original	[@]	
pipeline	9/1/21, 12:03:57 PM	9/1/21, 10:00:20 PM	09:56:22	Manual trigger	✓ Succeeded		Original	[@]	
> run	9/1/21, 9:56:44 AM	9/1/21, 10:47:12 AM	00:50:28	Manual trigger	✓ Succeeded		Rerun (Latest)	[@]	
pipeline	9/1/21, 12:14:20 AM	9/1/21, 9:35:38 AM	09:21:17	Manual trigger	✓ Succeeded		Original	[@]	
relabel	8/31/21, 10:26:43 PM	9/1/21, 12:07:20 AM	01:40:36	abe67b67-85c6-45a6-97:	✗ Cancelled	✖	Original	[@]	



Microsoft Azure | Synapse Analytics > https://web.azuresynapse.net/en-us/authoring/orchestrate/pipeline

Search

Add trigger

If Condition

if_refresh_apdm

True
1 activities

False
No activities

Lookup

check_apdm_status

If Condition

if_pull_scenarios

True
1 activities

False
No activities

Lookup

check_scen_status

If Condition

if_is_fail_andm

True
1 activities

False
No activities

Set variable

is fail andm

Parameters

New Delete

Name	Type	Default value
RunYear	Int	0
RunNumber	Int	0
UserEmail	String	Value
pull_apdm	Bool	false
pull_scenarios	Bool	false
run_r_model	Bool	false
projection_set	String	c(3,4,5)
skip_agent_jobs	Bool	false
full_erm_refresh	String	false

New trigger

Name * Weekly

Description

Type * Schedule

Start date * 09/03/2021 8:00 PM

Time zone * Coordinated Universal Time (UTC)

Recurrence * Every 1 Week(s)

Advanced recurrence options

Run on these days

Sun Mon Tue Wed Thu Fri Sat

Execute at these times

Hours 20 X

Minutes 0 X

Schedule execution times 20:00

Specify an end date

Annotations

+ New

Start trigger

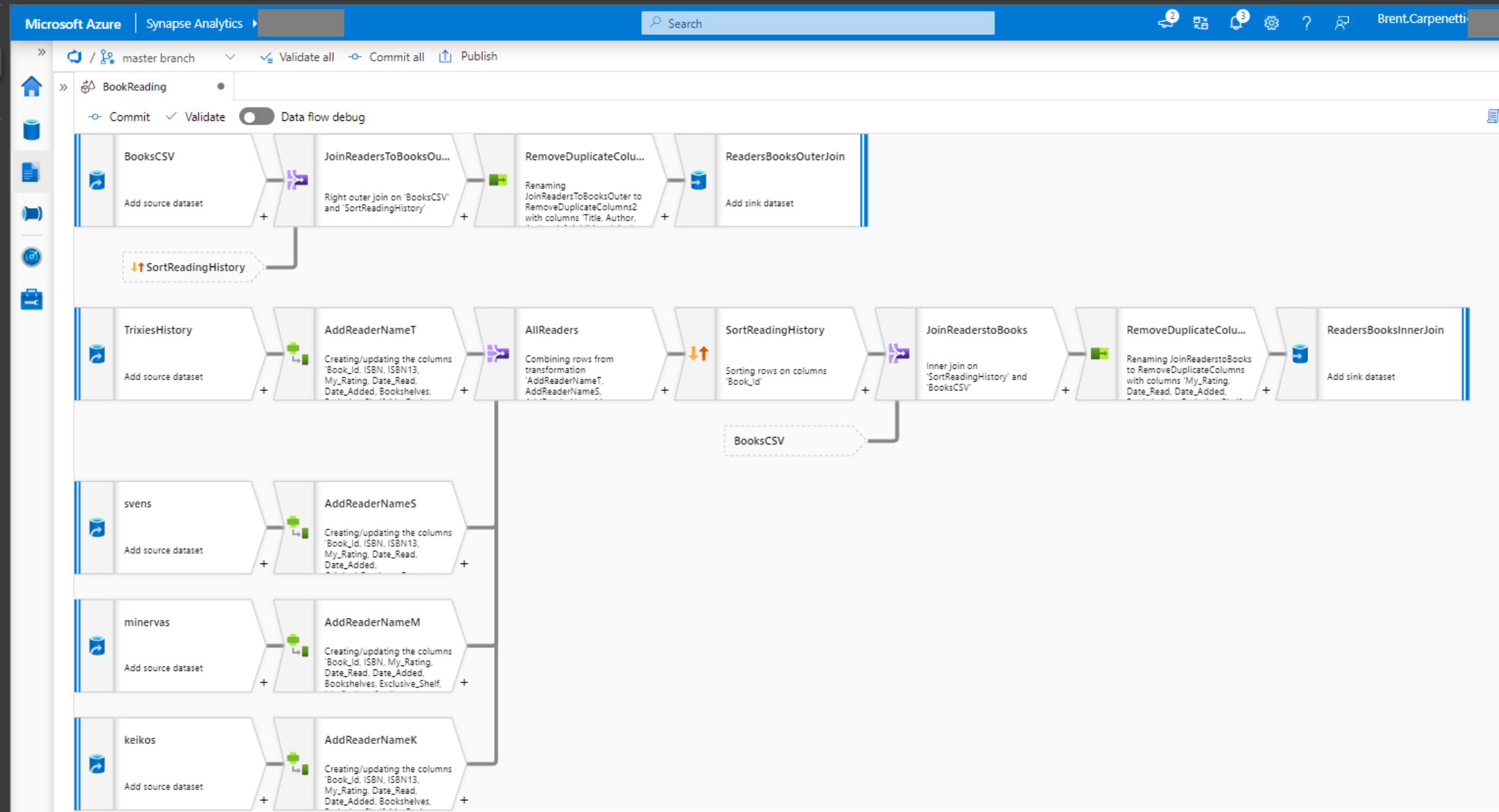
Start trigger on creation

Commit Cancel

```
graph TD; subgraph Pipeline [Pipeline Designer]; direction TB; A[If Condition if_refresh_apdm] --> B[Lookup check_apdm_status]; B --> C[Set variable is fail andm]; C --> D[If Condition if_pull_scenarios]; D -- True --> E[Lookup check_scen_status]; D -- False --> F[Set variable is fail andm]; end; subgraph TriggerForm [Trigger Form]; G[New trigger]; G --> H[Name: Weekly]; H --> I[Type: Schedule]; I --> J[Start date: 09/03/2021 8:00 PM]; J --> K[Time zone: Coordinated Universal Time (UTC)]; K --> L[Recurrence: Every 1 Week(s)]; end;
```

Tips and Tricks

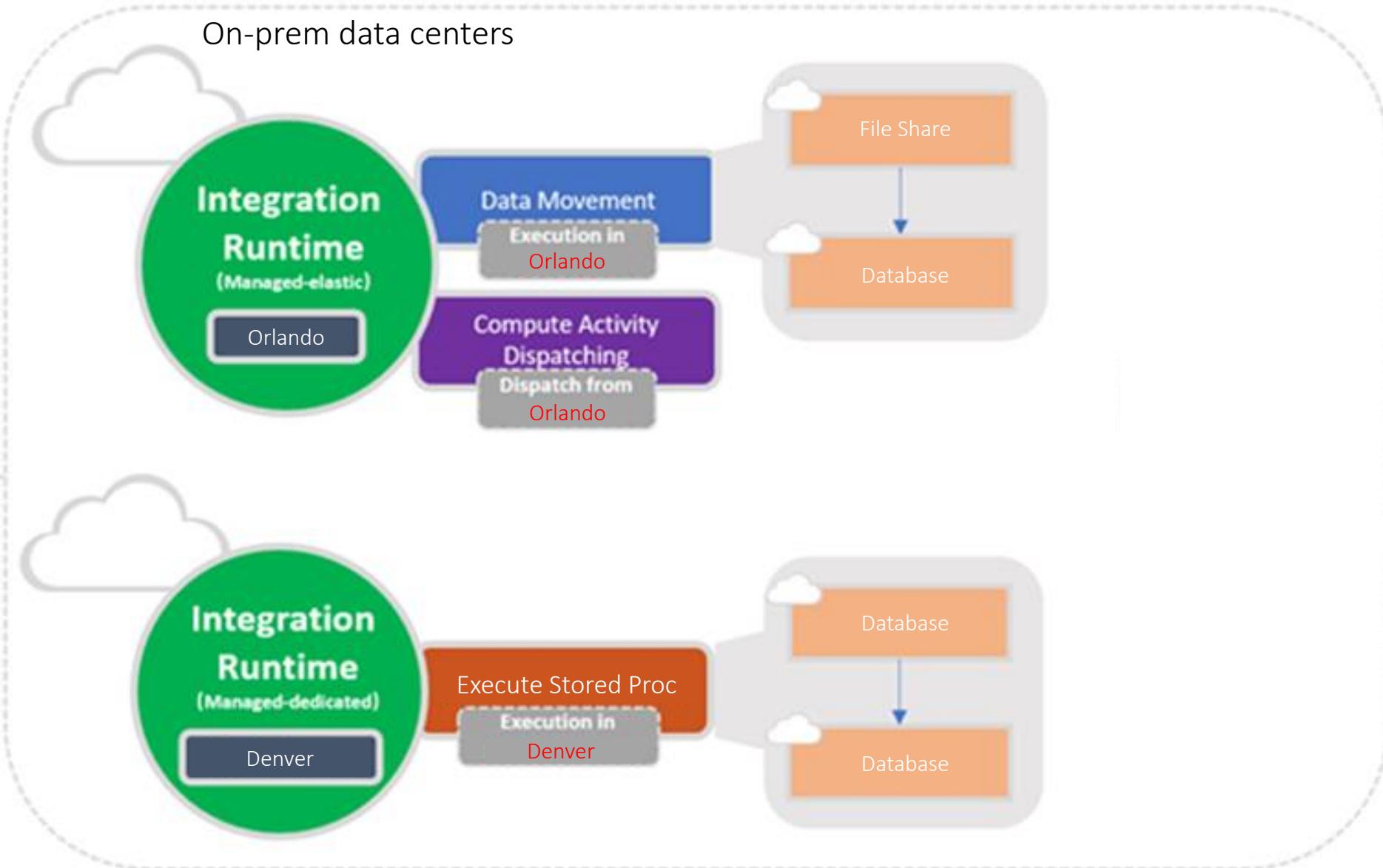
- Diagram your flow on a whiteboard before building the application
- Identify what data elements are dimensions or facts early on
- Sink/Stage/Write data at key process steps to create “checkpoints”
- Normalize your data before running machine learning or regressions
 - Data science refers to this as data cleansing
 - Connect with IT and check if a data model already exists (it likely does!)
- If you work on big data (Petabyte - structured/unstructured)
 - Consider defining unique, indexed, primary keys to improve join performance
 - Consider Spark compute resources (ADB/SynapseSpark) for transformations





East US

meta-data &
pipeline
trigger in East
US



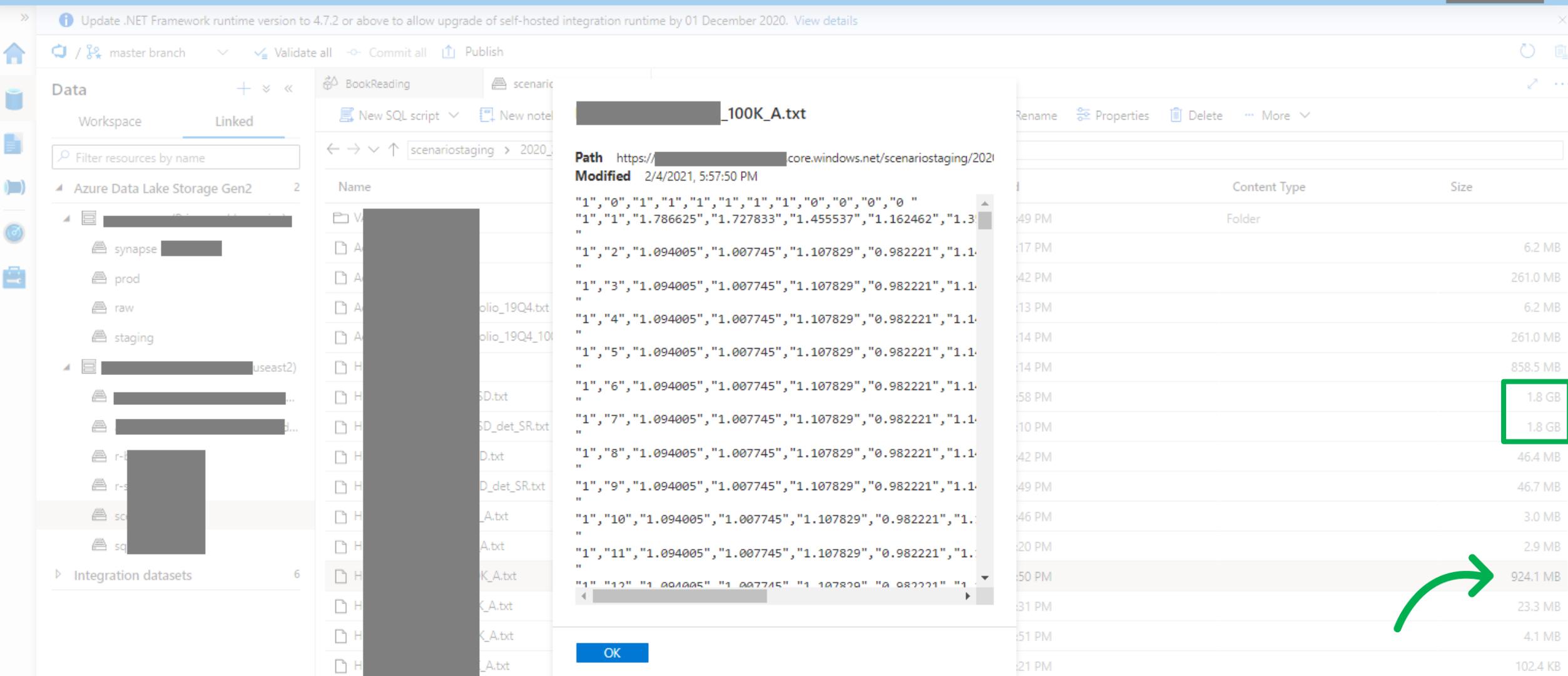
Resources Demo 3

- <https://github.com/hfleitas/rig/blob/master/pipeline/Primer.json>
- <https://docs.microsoft.com/azure/data-factory/solution-templates-introduction>
- [Azure Synapse Paths | Microsoft Learn](#)
- [Data integration at scale with Azure Data Factory or Azure Synapse Pipeline](#)
- [Use stored procedures - Azure Synapse Analytics | Microsoft Docs](#)

BONUS

Azure Synapse

- a) Data exploration
- b) Notebook development
- c) Server-less pools
- d) Purview: Data Dictionary & Process Flows



master branch Validate all Commit all Publish

Boston house price ...

Run all Undo Commit Outline

Attach to ecmSpark

Language

PySpark (Python)

Variables

Preview Features

...

Not started

Table of contents

Boston house price prediction with LightGBM

- Prepare Dataset
- Baseline - Spark MLlib Linear Regressor
- Vowpal Wabbit
- LightGBM
- Clean up resources
- Next steps

+ Code

+ Markdown

```
1 lgr = LightGBMRegressor(  
2     objective='quantile',  
3     alpha=0.2,  
4     learningRate=0.3,  
5     numLeaves=31,  
6     labelCol='target',  
7     numIterations=100,  
8 )  
9  
10 # Using one partition since the training dataset is very small  
11 repartitioned_data = lr_train_data.repartition(1).cache()  
12 print(repartitioned_data.count())  
13 lg_model = lgr.fit(repartitioned_data)  
14 lg_predictions = lg_model.transform(lr_test_data)  
15  
16 display(lg_predictions.limit(10).toPandas())
```

[12] ✓ - Command executed in 2 sec 80 ms on 2:12:37 AM, 5/25/21

377

View

Table

Chart

Export results

7.8555181506504566

9.408284298711045

10.125545304878454

11.009850748168407

11.378612631892928

11.705711758454298

11.789698630551646

target

+



Query folders and multiple files

Description

Serverless SQL pool supports reading multiple files/folders by using wildcards, which are similar to the wildcards used in Windows OS. However, greater flexibility is present since multiple wildcards are allowed.

Tags

- [Serverless SQL pool](#)
- [Sample](#)
- [Sample Data](#)
- [Folder](#)
- [Query](#)
- [Files](#)

Preview

```
-- /* Multiple wildcards */
70
71  /* Read all files from multiple folders */
72  SELECT
73      YEAR(pickup_datetime) AS [year],
74      SUM(passenger_count) AS passengers_total,
75      COUNT(*) AS [rides_total]
76  FROM OPENROWSET(
77      BULK 'https://sqlondemandstorage.blob.core.windows.net/public-csv/t\*i/',
78      FORMAT = 'CSV', PARSER_VERSION = '2.0',
79      FIRSTROW = 2
80  )
81  WITH (
82      vendor_id VARCHAR(100) COLLATE Latin1_General_BIN2,
83      pickup_datetime DATETIME2,
84      dropoff_datetime DATETIME2,
85      passenger_count INT,
86      trip_distance FLOAT,
87      rate_code INT,
88      store_and_fwd_flag VARCHAR(100) COLLATE Latin1_General_BIN2,
89      pickup_location_id INT,
90      dropoff_location_id INT,
91      payment_type INT,
92      fare_amount FLOAT,
93      extra FLOAT,
94      mta_tax FLOAT,
95      tip_amount FLOAT,
96      tolls_amount FLOAT,
97      improvement_surcharge FLOAT,
98      total_amount FLOAT
99  ) AS nyc
100 GROUP BY
101     YEAR(pickup_datetime)
102 ORDER BY
103     YEAR(pickup_datetime);
104
105 /* Multiple wildcards */
106 SELECT
107     YEAR(pickup_datetime) AS [year],
```

You will be charged \$5/TB of data processed. The minimum charge is for 10MB with 1MB increments. [Learn more](#)



Adatum Corp

13,020 users

1,134 sources

76,392,971 assets

312 terms

Revenue

Search suggestions

revenue **customer**revenue **finance**revenue **glossary**revenue **information**

/view insights

Get insights on your data.



Knowledge center
Discover learning and tutorials.

Recently accessed

My items

Name

Order

Percent Sales

SalesOrderHeader

ProductCategory

SSNNumber

SalesLT.SalesOrderHeader.csv

Asset suggestions

TaxRevenue

teradata://adatumteradata.retail.com/DB/TaxRevenue

CustomerRevenue_{N}

https://adatumcostorage.blob.core.windows.net/CUST/CustomerRevenue_{N}.csv

Revenue

https://adatumco.dfs.core.windows.net/incr/raw/revenue

Revenue2020.tsv

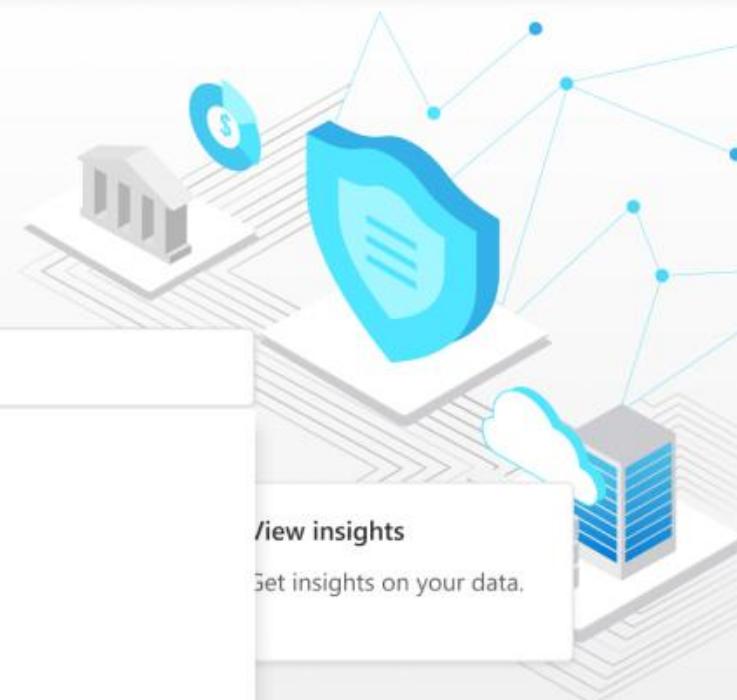
https://adatumco.dfs.core.windows.net/sales/Revenue2020.tsv

[View search results](#)

10 days ago

references, etc.

1 month ago

[View all recently accessed](#)



Campaign Analytics

Confidential

Certified

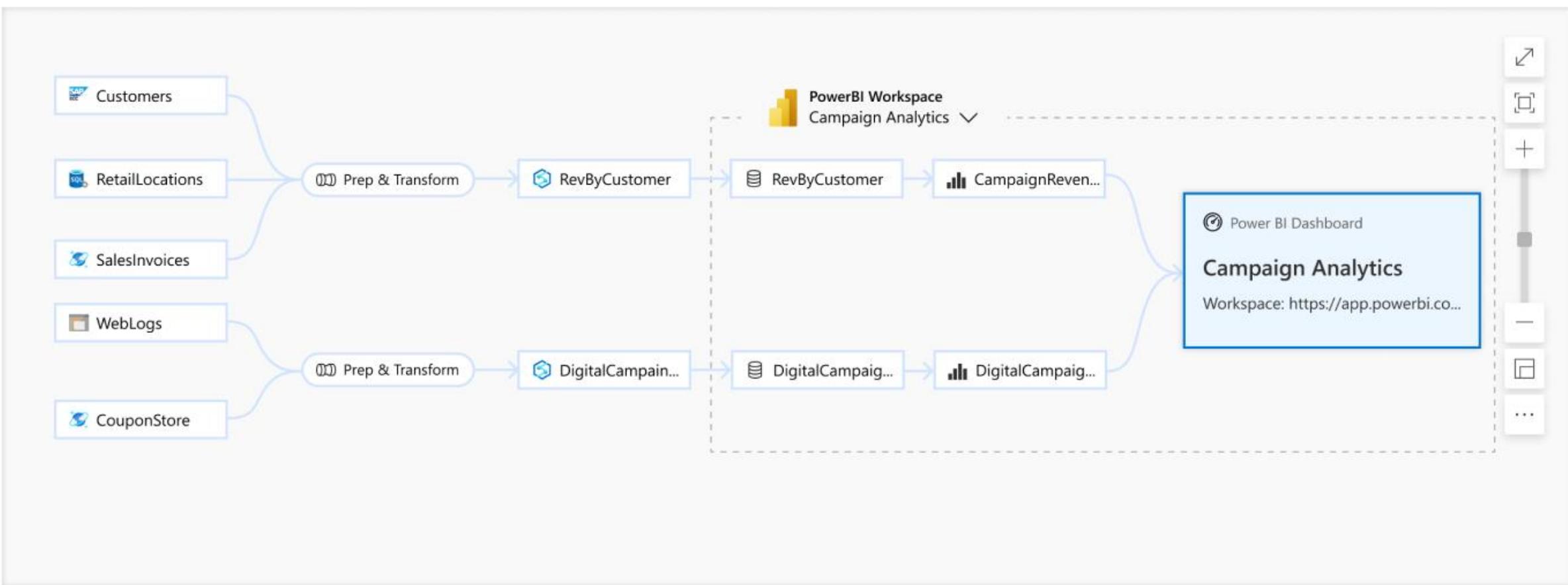
Power BI Dashboard

Edit Refresh

Open in Power BI

Overview Lineage Contacts Related

Search for assets or processes





Fill Out the
Evaluation and
Claim Your CPD
Credits

Download Slides

- (PDF LINK)

Thank you!

(PDF URL)

<https://linkedin.com/in/brentcarpenetti>

<https://aka.ms/hiram>

Presenters:

Brent Carpenetti; Senior Actuary, Hannover Re
Hiram Fleitas; Senior Customer Engineer, Microsoft