



twoday
capacity

ITPA
CONSULTING



Premium Sponsors



Raffle Prizes

Standard Sponsors

TIME **X** TENDER

[stoltze][it]

ORANGEMAN

 **exmon**


KPMG

Fellowwind



Data Platform Development The Git Way

Koos van Strien

Data Platform Next Step – 2023-06-09

LinkedIn 🖐️

koos.van.strien@wortell.nl

www.wortellsmartlearning.nl





TEAM

1

Business & Bricks Contest



Who's this guy?

- Koos van Strien
- Head of training, Solution Architect, Data Engineer, (yadayada)
- Co-owner Wortell Smart Learning

```
SELECT
  c.title,
  c.duration,
  ce.place,
  ce.starts_at
FROM courses AS c
INNER JOIN course_editions AS ce ON c.id = ce.course_id
```

Execute query



wortell

Session outline

- Setting up repositories
- Practices for branching
- Brave New World demo: ~~ADFGen2 Synapse Pipelines ADF~~
- Tips & Tricks & Questions

Session outline

- **Setting up repositories**
 - From TFVC/SVN/... to Git
- Practices for branching
- Brave New World demo: ~~ADFGen2 Synapse Pipelines~~ ADF
- Tips & Tricks & Questions

A dark stage with two bright spotlights shining down on a wooden floor. The spotlights create a V-shape of light in the dark background.

Setting the Stage

Data Platform Development – the Git way

wortell



123

SESAME STREET®

Use any Version Control?

Setting the stage

SONGS

wortell

123

SESAME STREET®

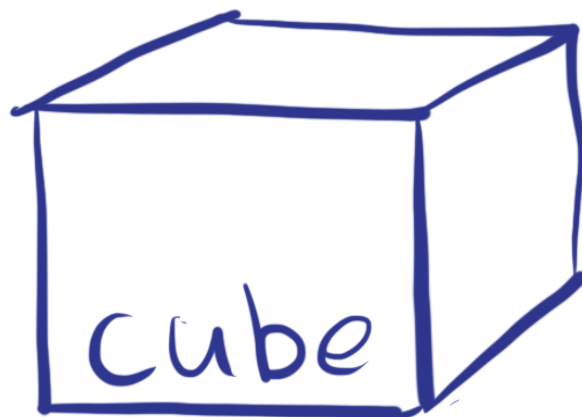
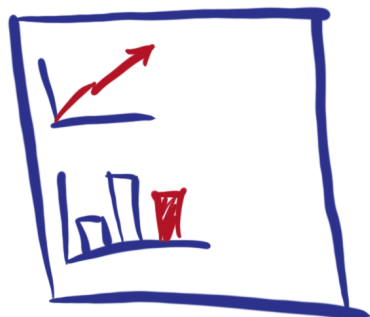
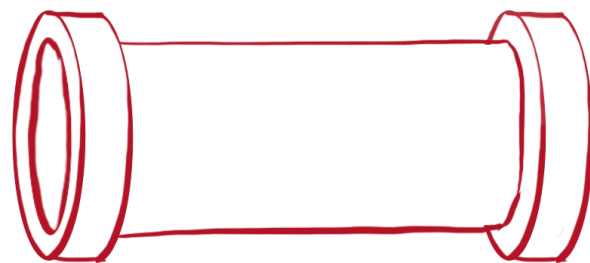
Use Git for Version Control?

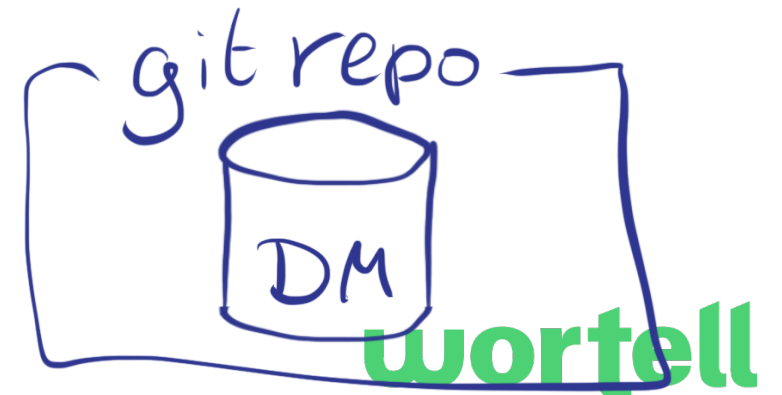
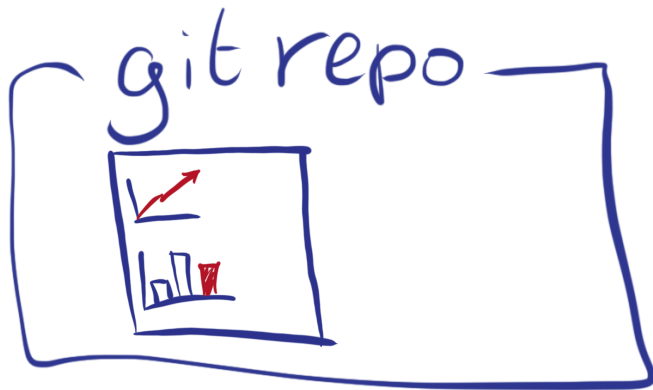
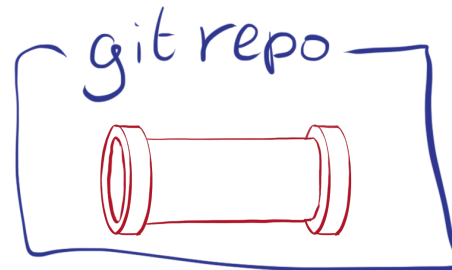
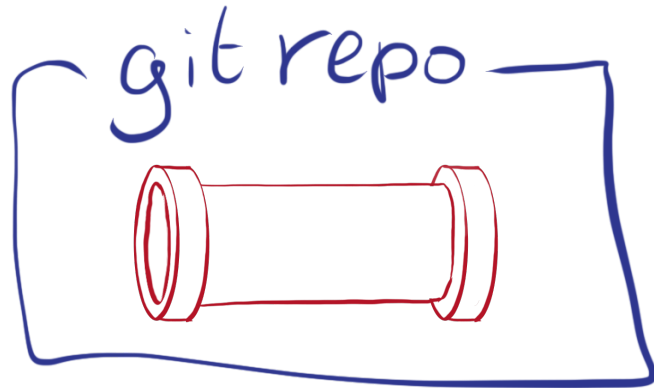
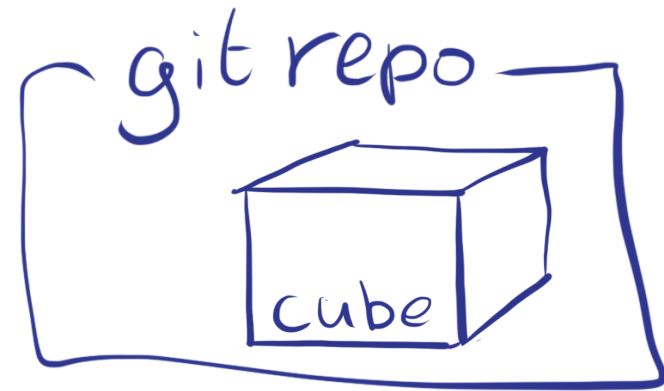
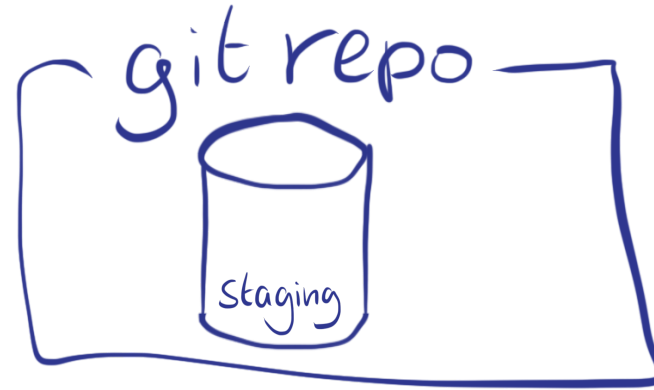
Setting the stage

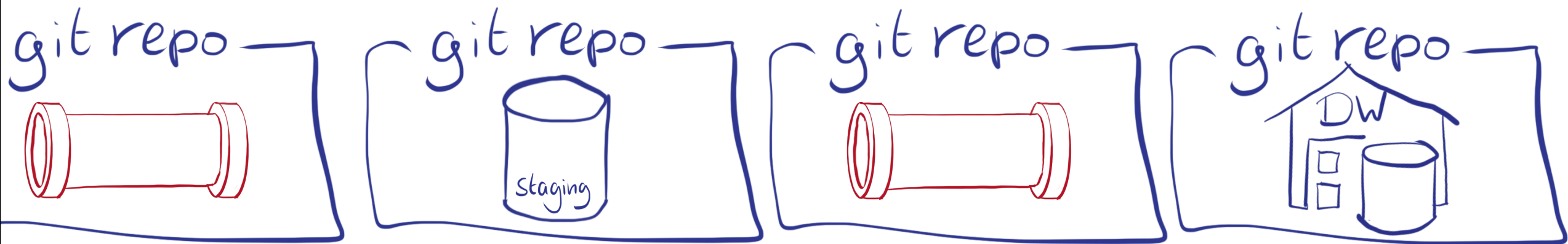
SONGS

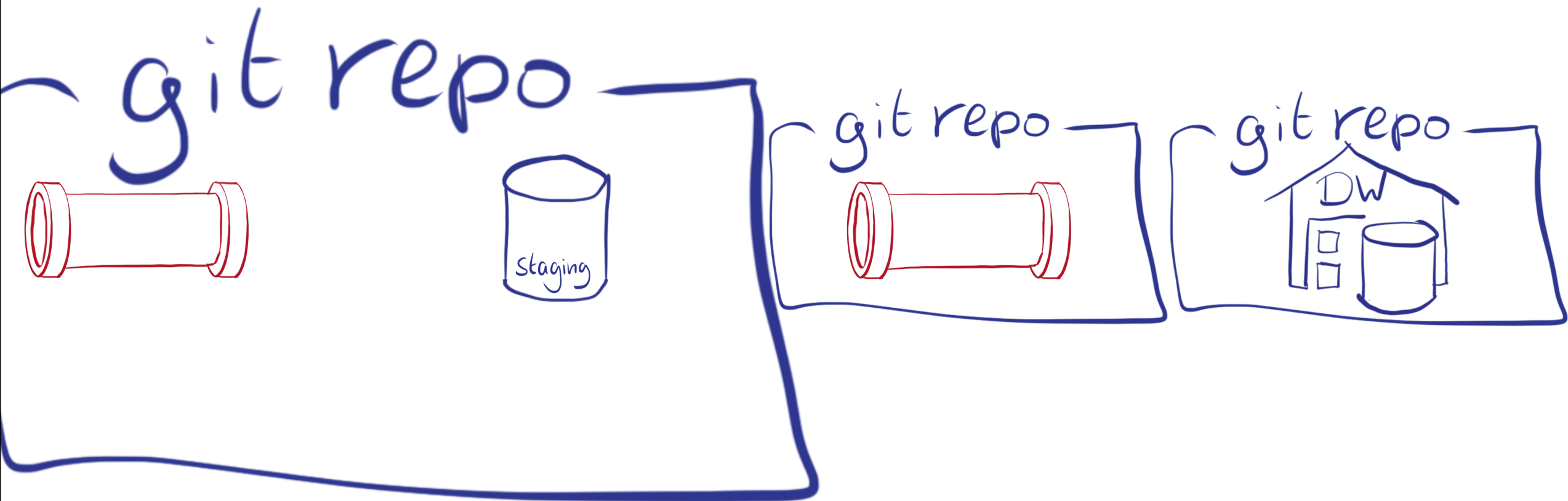
wortell

TFVC

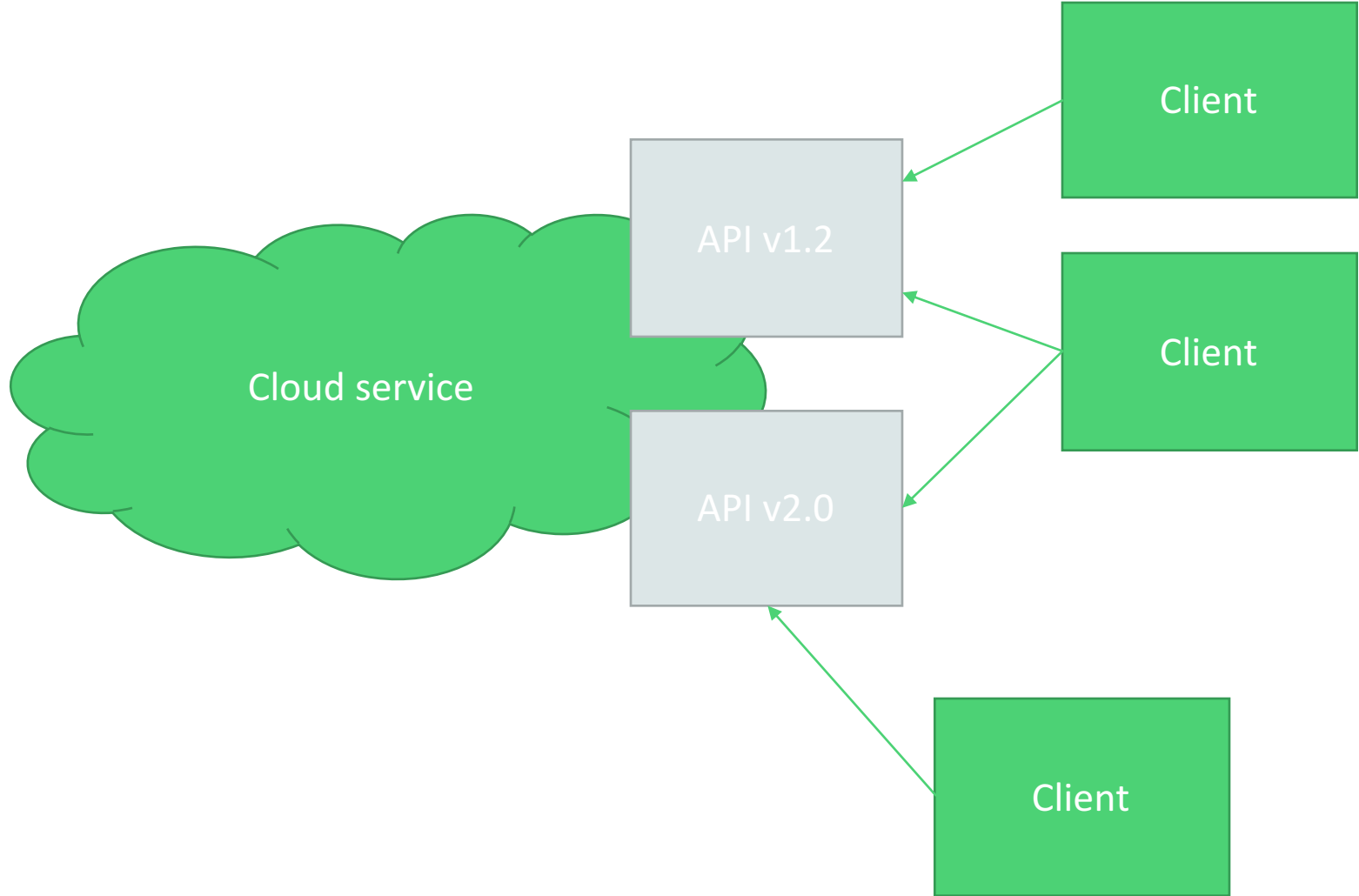




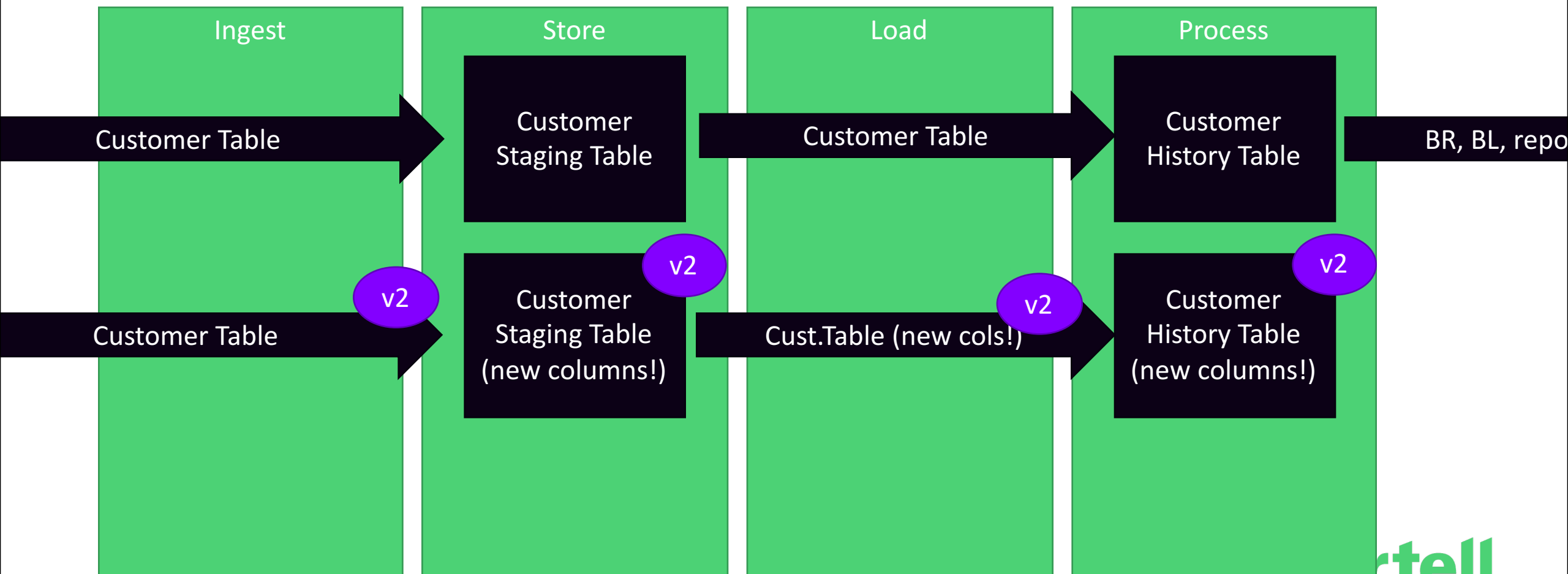




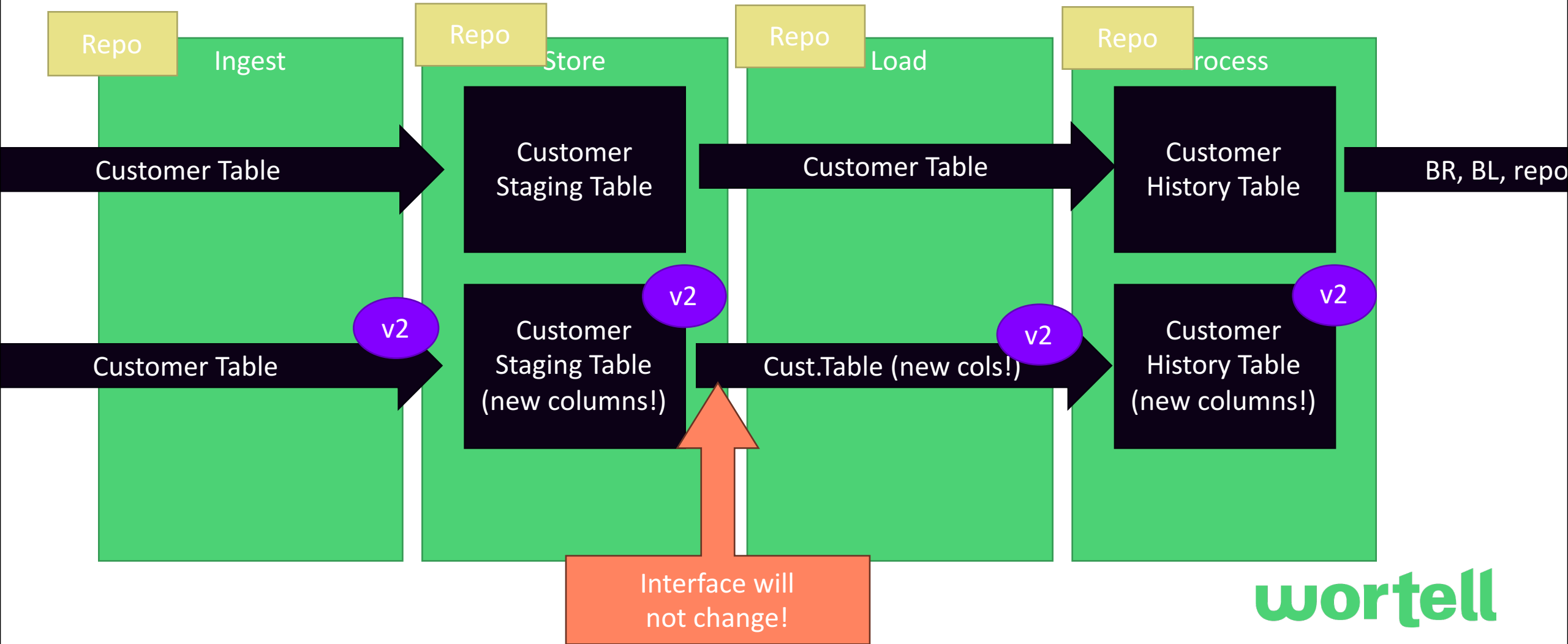
Non-destructive software updates



Non-destructive data software updates



Non-destructive data software updates



In comes Data Mesh...



123

SESAME STREET®

Familiar with Data Mesh?

Raise your hand!

SONGS

wortell



123
SESAME STREET®

“Doing” Data Mesh?

At your company, or at your customer's?

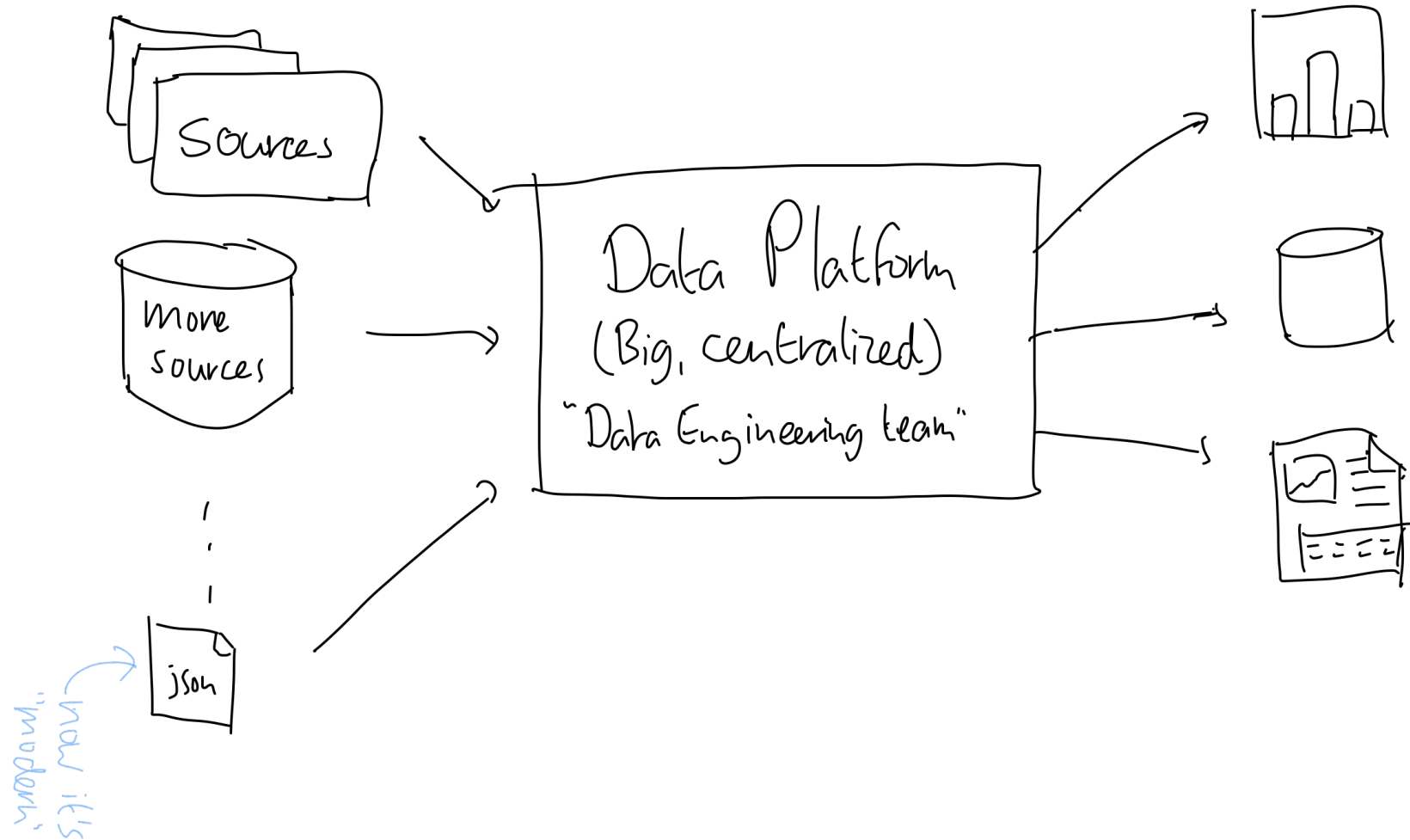


SONGS

wortell

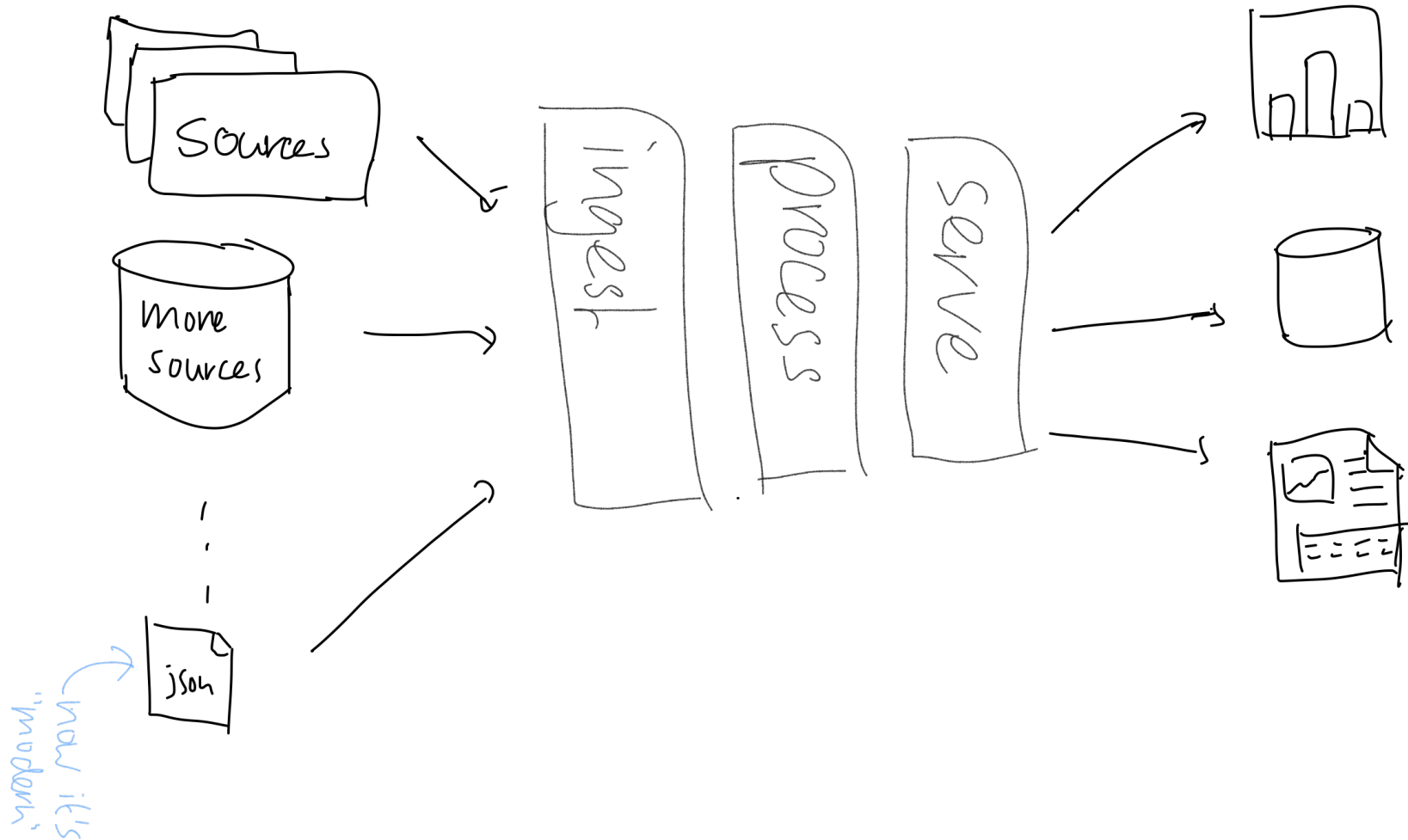
What is Data Mesh?

Evolution of your Data Platform (current situation)



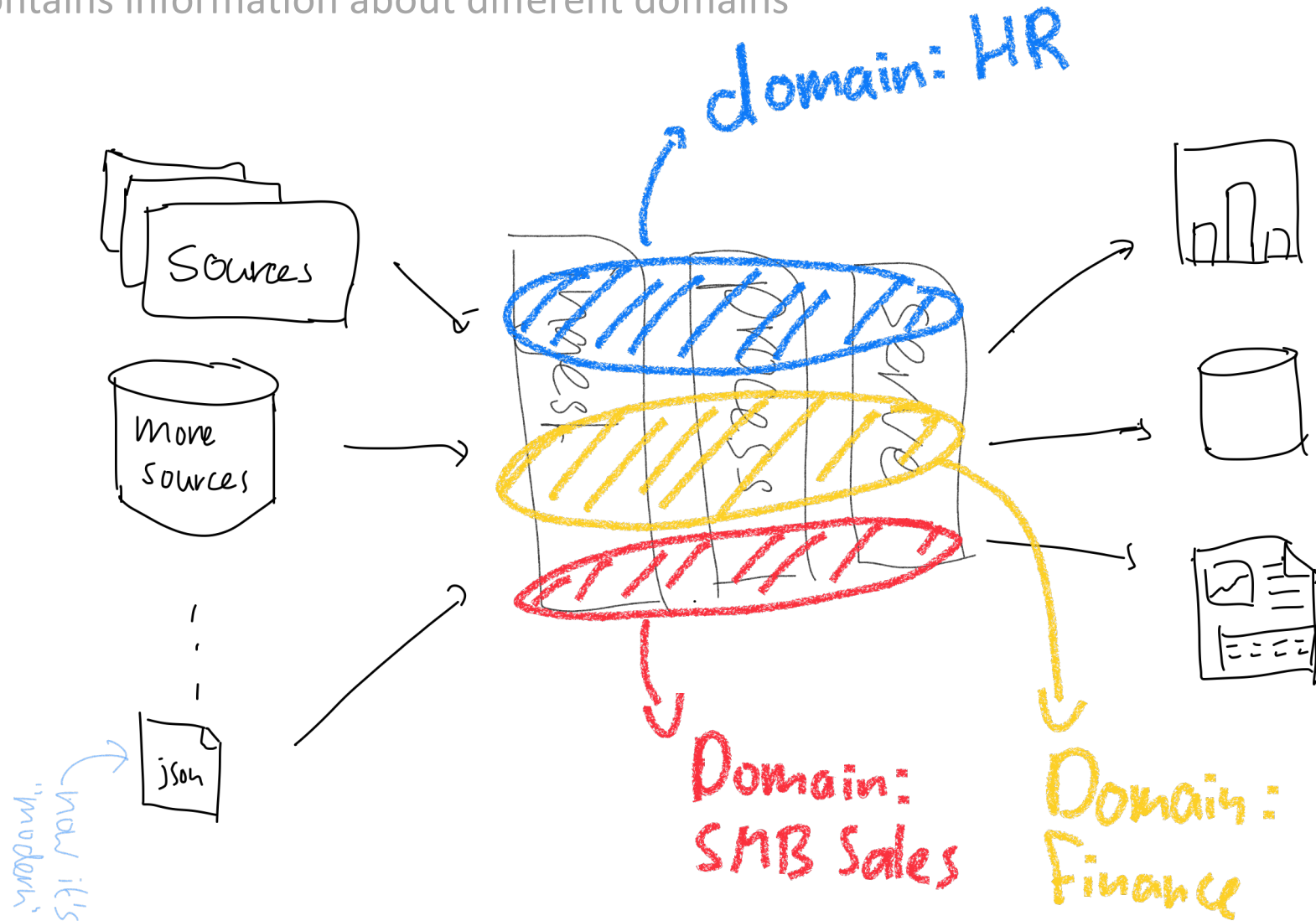
What is Data Mesh?

“Architectural quants” (Fowler)



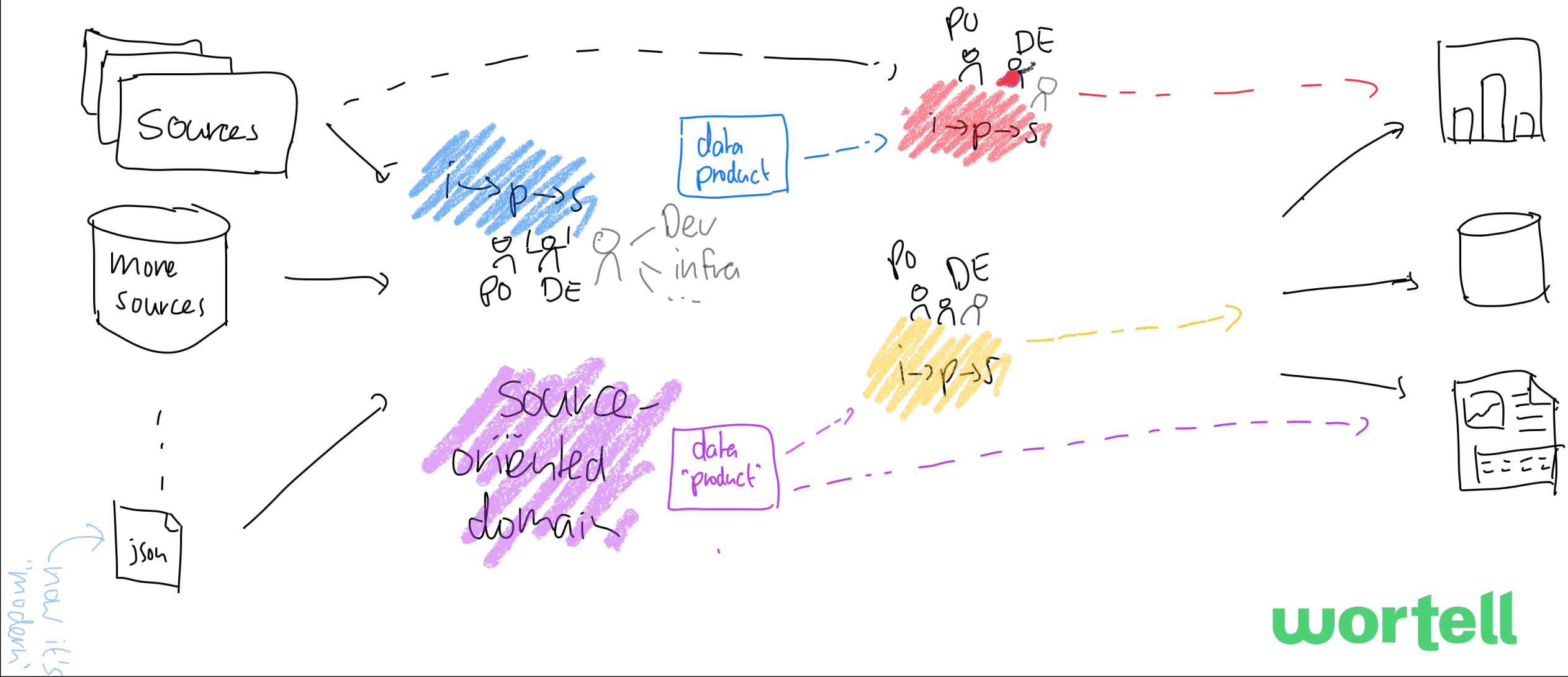
What is Data Mesh?

Pipeline contains information about different domains



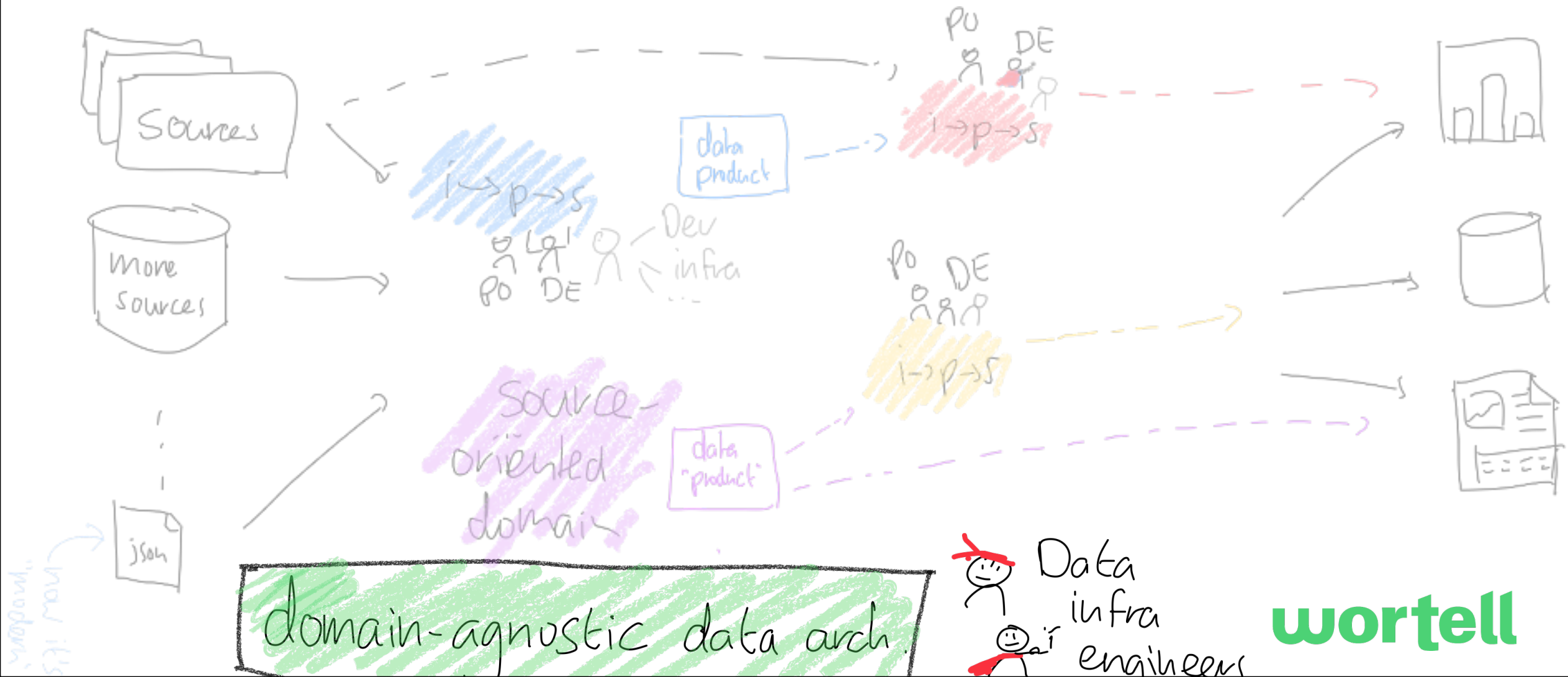
What is Data Mesh?

Domain teams own their “data products” (reports, sets, ...)



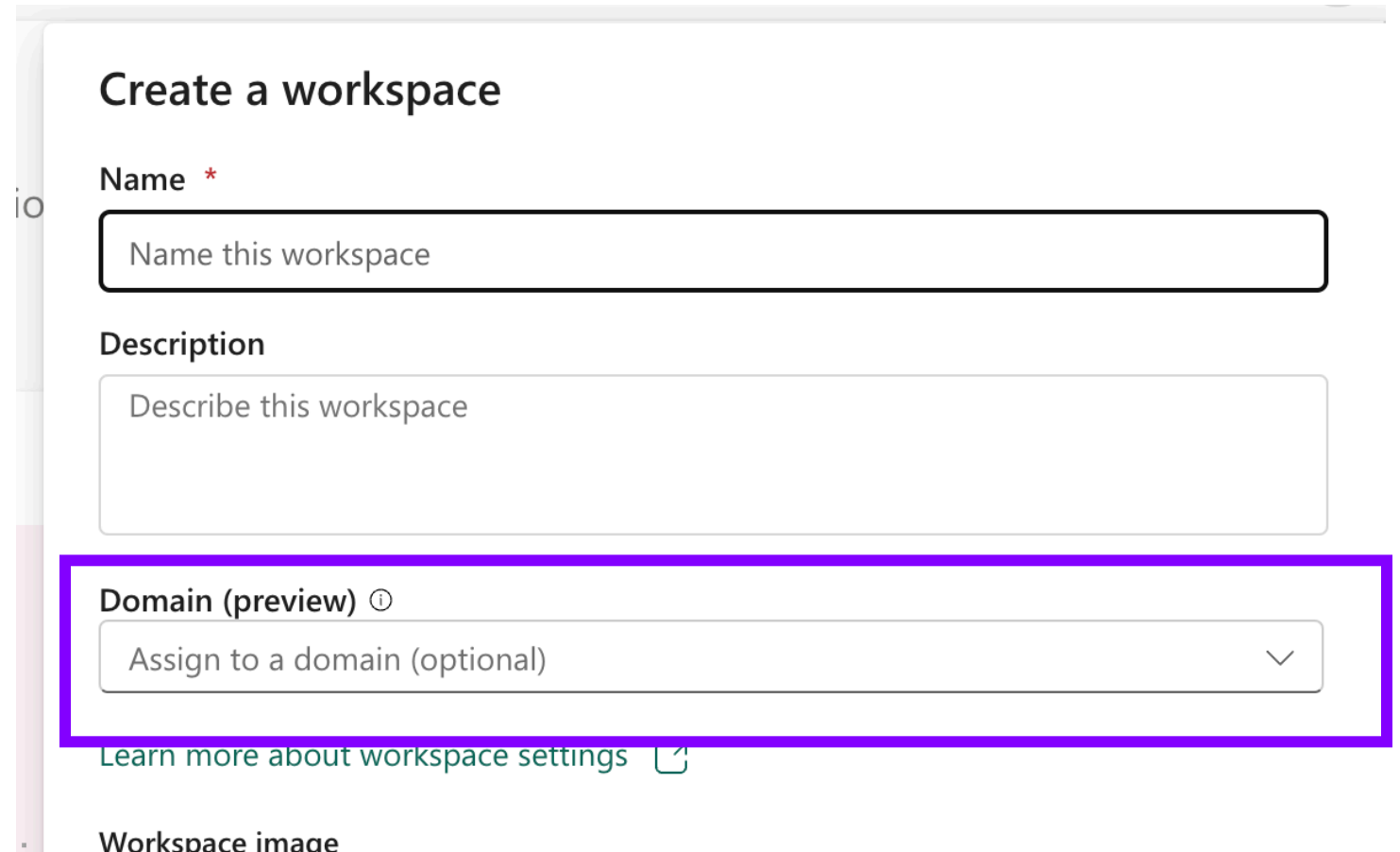
What is Data Mesh?

Domain-agnostic platform



Fabric is aligning with Data Mesh

- Workspace <-> “data product”
 - Coupled to Git
 - Dataflows, Data Engineering, Data Science → happening inside workspace
 - Workspaces can be coupled to “domains”
 - **Shortcuts!**



Create a workspace

Name *

Name this workspace

Description

Describe this workspace

Domain (preview) ⓘ

Assign to a domain (optional) ▼

[Learn more about workspace settings](#) ↗

Workspace image

Azure DevOps <-> Fabric?

- Repos: Currently limited
 - Git integration still in preview
- Pipelines: TBD
 - Within Fabric: Deployment Pipelines

<div><div>+ New ▾</div><div>↑ Upload ▾</div><div> Create deployment pipeline ...</div><div> Source control 0</div><div><input type="text" value="Filter by keyw"/></div></div>					
	Name	Git status	Type	Owner	Refr
	Dataflow 2	Unsupported	Dataflow Gen2	Koos van Strien	
	DataflowsStagingLakehouse	Unsupported	Dataset (default)	demo-workspace-fab...	
	DataflowsStagingLakehouse	Unsupported	SQL endpoint	demo-workspace-fab...	
	DataflowsStagingLakehouse	Unsupported	Lakehouse	Koos van Strien	
	git-demo-test	Synced	Report	demo-workspace-fab...	

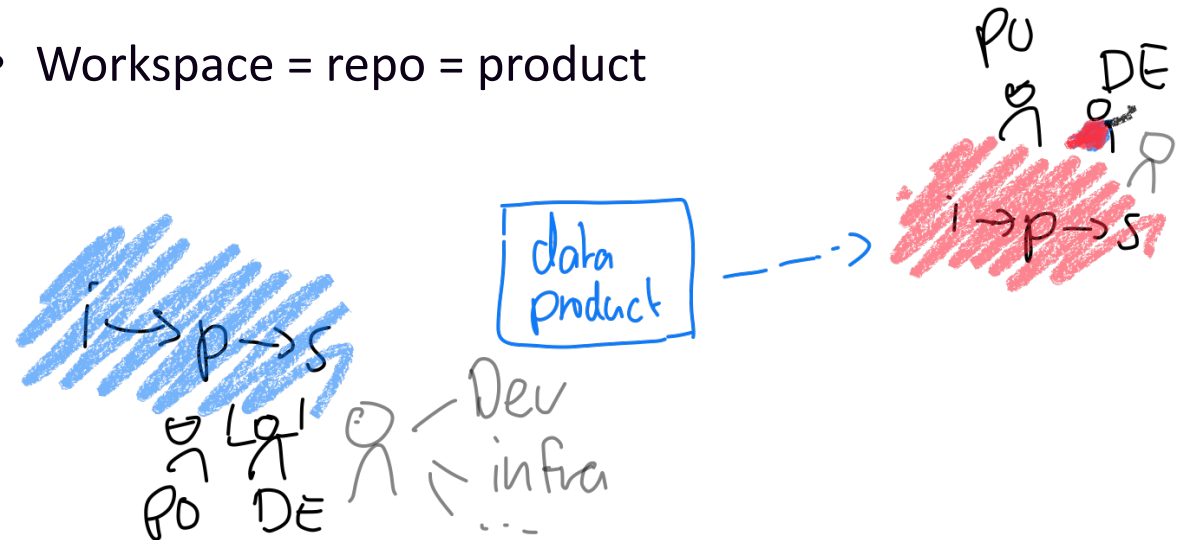
Domain-agnostic vs. domain-specific

domain-agnostic data arch.

Data
infra
engineers

- Focus on infrastructure
- Domain-agnostic
- Azure tools
- Contracts / interfaces

- Focus on data product (in workspace)
- Within domain
- Less “strict” interface
- Workspace = repo = product



Session outline

- Setting up your repos
 - One per *independently deployable product*
 - Deployments can't "just" destroy existing interfaces
 - Interfaces *can* deprecate of course! Design a standard procedure for doing & communicating so
- Practices for Branching
- Brave New World demo: ~~ADFGen2 Synapse Pipelines ADF~~
- Tips & Tricks & Questions

Session outline

- Setting up your repos
- **Practices for Branching**
- Enforcing policies
- Brave New World demo: ~~ADFGen2 Synapse Pipelines ADF~~
- Tips & Tricks & Questions

Features
bug fixes
Use Feature branches

Merge using PRs
into main

Microsoft about Branch Strategy

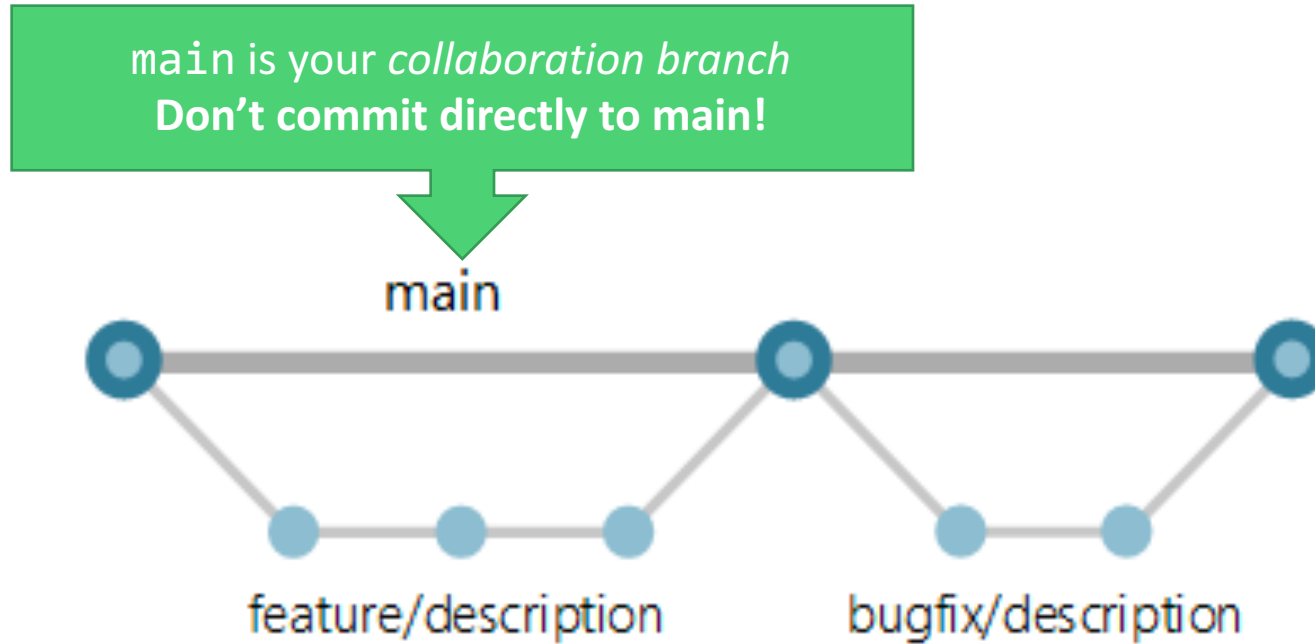
<https://docs.microsoft.com/en-us/azure/devops/repos/git/git-branching-guidance>

Keep your main branch
high quality
up-to-date

release branches

wortell

Use Feature Branches



Features
bug fixes
Use Feature branches

wortell

Merge using Pull Request into Main

- Dev A:
 - “Hey team, please merge this awesome new code into master?”
- Reviewer (Dev B):
 - ~~“Sure, once you fix this ugly piece of shit”~~
 - “Great work A! One small note: please fix this!”
 - Approves (after A’s fix)
- Reviewer (Dev C):
 - No problems found
 - Approves
- **Merge system takes over and merges feature branch to master**

*Merge using PR
into main*

Keep Main High-Quality

high quality
Keep your main branch
up-to-date

N

nextstep-2023

+

Overview

Boards

Repos

Files

Commits

Pushes

Branches

Tags

Pull requests

Pipelines

Test Plans

Artifacts

Branches

Mine

All

Stale

New branch

Search branch name

Branch	C..	Author	Aut...	Behind Ahead	S..	P..
<div><div>🔗</div><div>main</div><div>Default</div><div>Compare</div></div>	d04	<div>K</div> koos....	Yes...		★	⋮

Keep Main High-Quality

Keep your main branch
high quality
up-to-date

Protect this branch

- Setting a Required policy will enforce the use of pull requests when updating the branch
- Setting a Required policy will prevent branch deletion
- Manage permissions for this branch on the [Security page](#)

☒ Require a minimum number of reviewers

Require approval from a specified number of reviewers on pull requests.

Minimum number of reviewers

- ☐ Requesters can approve their own changes
- ☐ Allow completion even if some reviewers vote to wait or reject
- ☐ Reset code reviewer votes when there are new changes

☐ Check for linked work items

Encourage traceability by checking for linked work items on pull requests.

☐ Check for comment resolution

Check to see that all comments have been resolved on pull requests.

☒ Limit merge types

Control branch history by limiting the available types of merge when pull requests are completed.

Allowed merge types:

Creates a semi-linear history by replaying the source branch commits onto the target and then creating a merge commit.

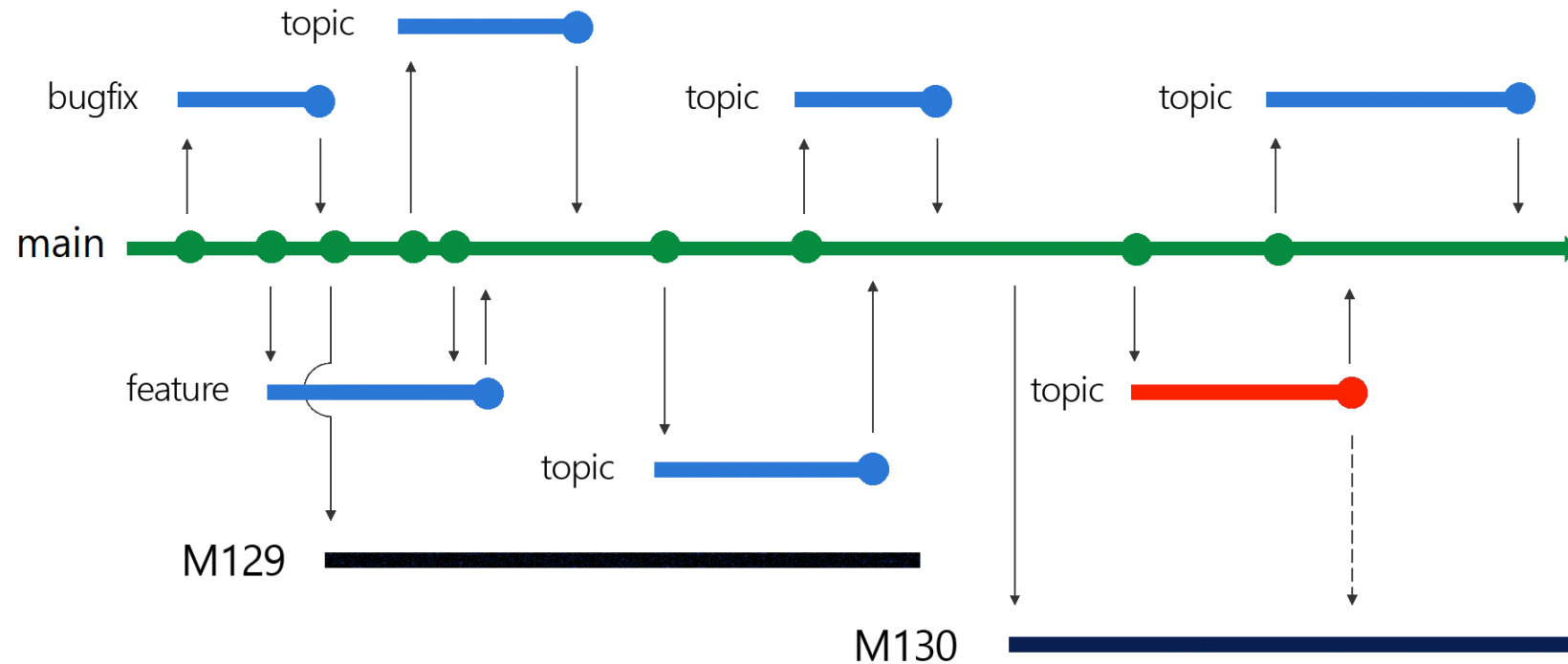
Build validation

Validate code by pre-merging and building pull request changes

+ Add build policy

wortell

Release using Release Branches

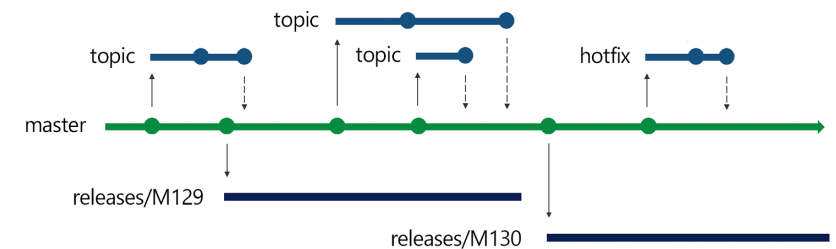


release branches?

<https://learn.microsoft.com/en-us/devops/develop/how-microsoft-develops-devops>

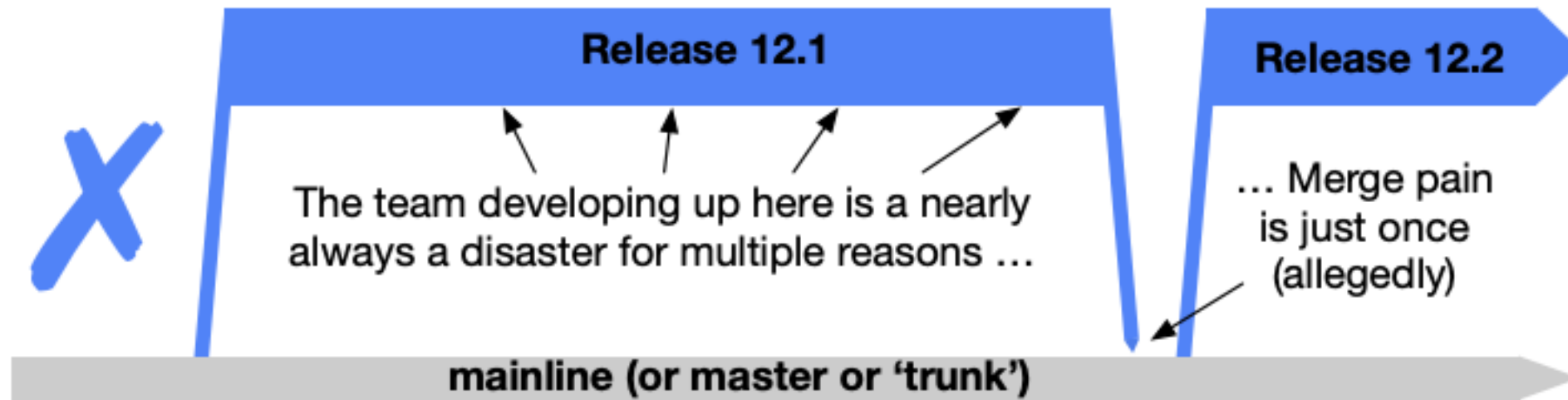
Release Flow Branching Structure

- Releases are *sprintly* (every X weeks)
- Features / topics within topic branches
- Bugfixes merged into main first
 - Then cherrypicked into release
- Every release goes through main first

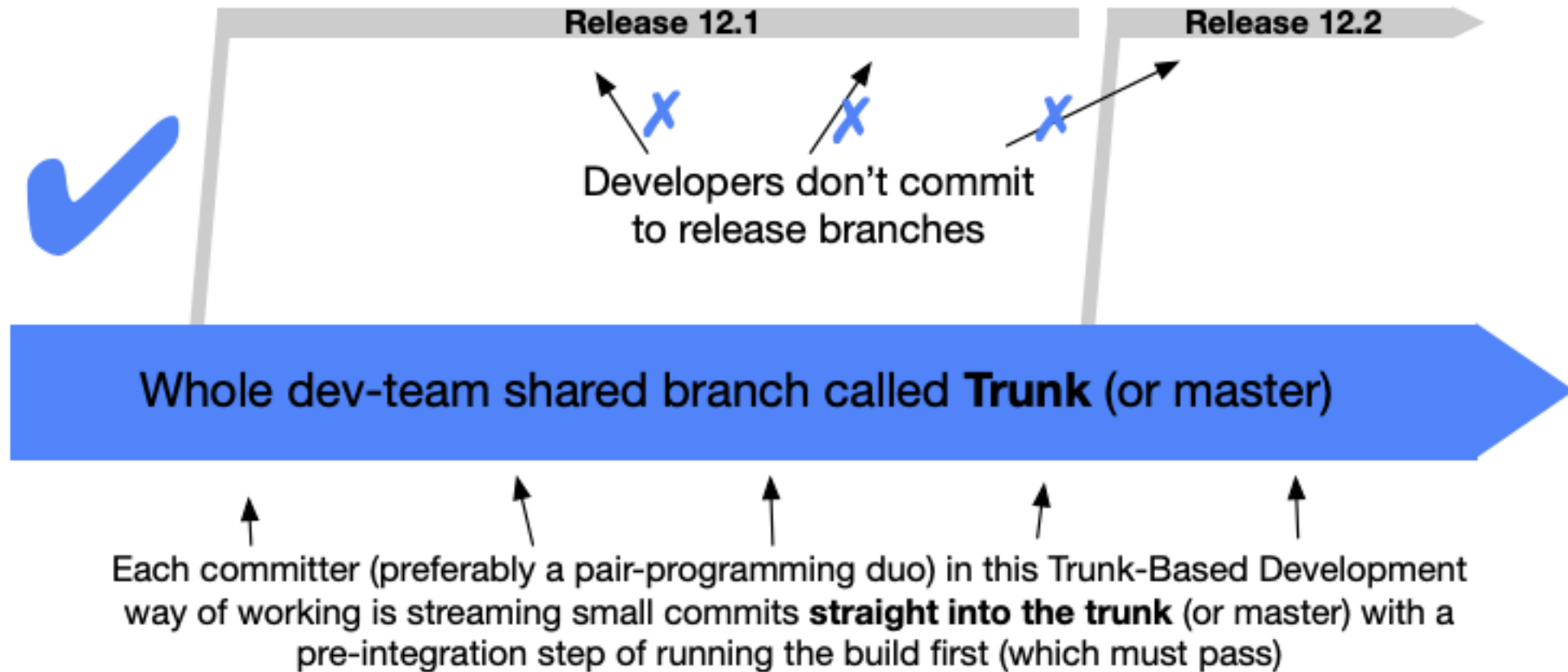


<https://devblogs.microsoft.com/devops/release-flow-how-we-do-branching-on-the-vsts-team/>

Don't: develop in release branches, then merge back



Do: *trunk-based development*



Session outline

- Setting up your repos
- Practices for Branching
 - Don't push to master
 - Enforce feature branches
 - Integrate ASAP
 - Use **feature toggles**
- Brave New World demo: ~~ADFGen2 Synapse Pipelines ADF~~
- Tips & Tricks & Questions

Session outline

- Setting up your repos
- Practices for Branching
- **Brave New World demo: ADFgen2 Synapse Pipelines ADF**
- Tips & Tricks & Questions

ADF Demo

Session outline

- Setting up repositories
- Practices for branching
- Brave New World demo: ADF
 - Development in Feature Branches
 - Trunk-based (merge to master before publishing)
 - Merge using Pull Request!
 - Results in an “artifact” you can deploy on any environment
- Tips & Tricks & Questions

Tips & Tricks

- Branch names: users/username/description
- Feature Toggles / Feature Flags (<https://martinfowler.com/articles/feature-toggles.html>)
- Don't try to boil the ocean
- *Any standard is better than your personal standard*



LinkedIn 🖐️

Data Platform Development The Git Way

Koos van Strien

Data Platform Next Step – 2023-06-09

koos.van.strien@wortell.nl
www.wortellsmartlearning.nl

wortell