DrupalCon
EUROPE2021
OCTOBER 4-7

Migrate for recurring data import

Lessons learned and tales from the trenches

arthur@webstanz.be

adelson@webstanz.be

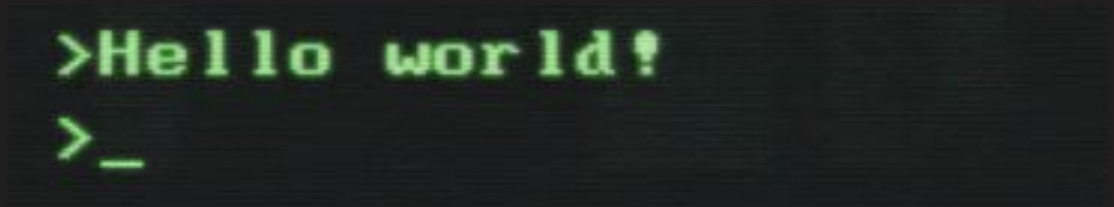
Speakers

Manages sync of a million records on a daily basis

arthur@webstanz.be

Deals with the unmanageable

adelson@webstanz.be

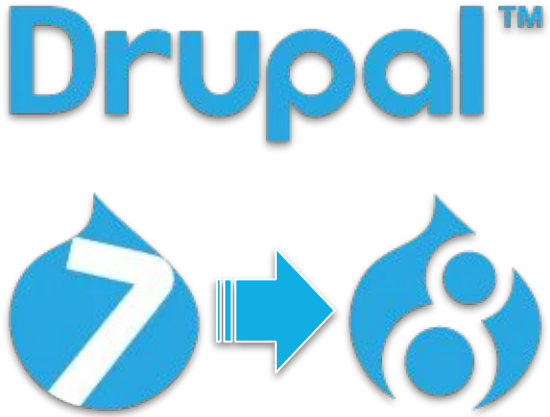


Drupal Migrate is a developer's tool !

Migrate Module

Originally developed to

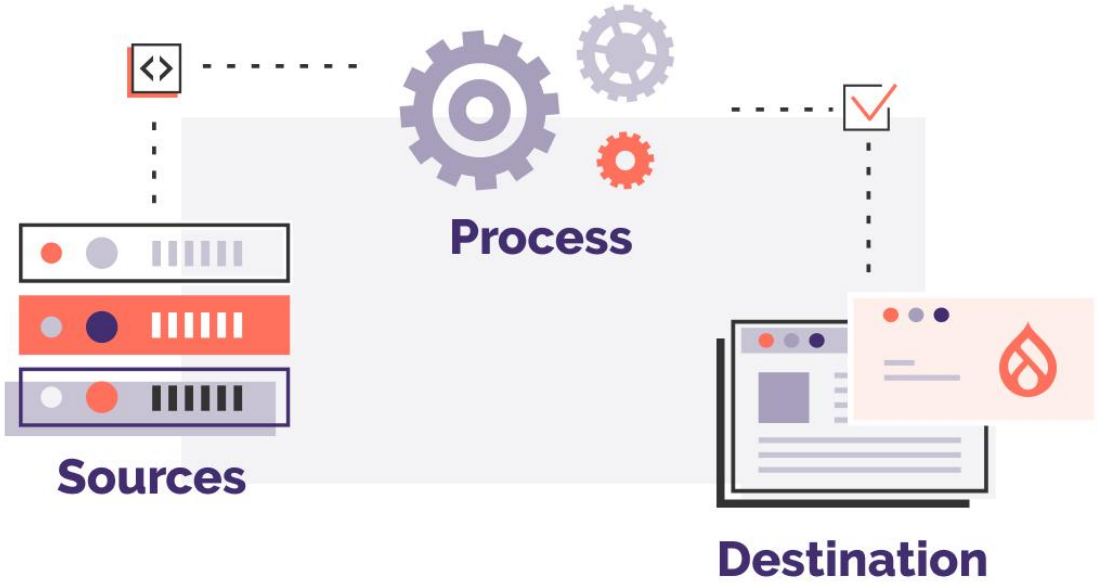
- perform complete site migration
- facilitate Drupal 7 upgrade to Drupal 8/9

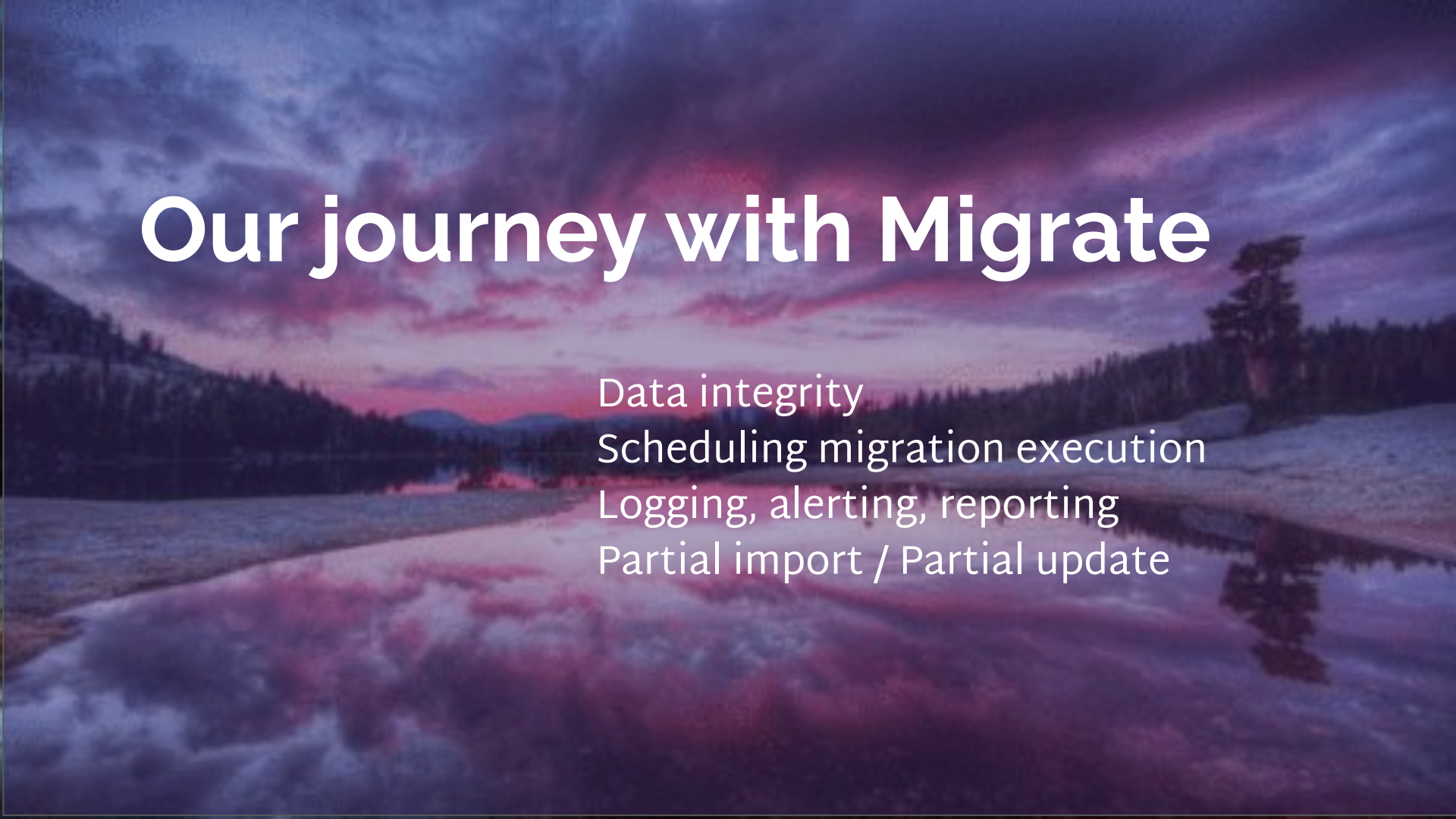


Migrate API

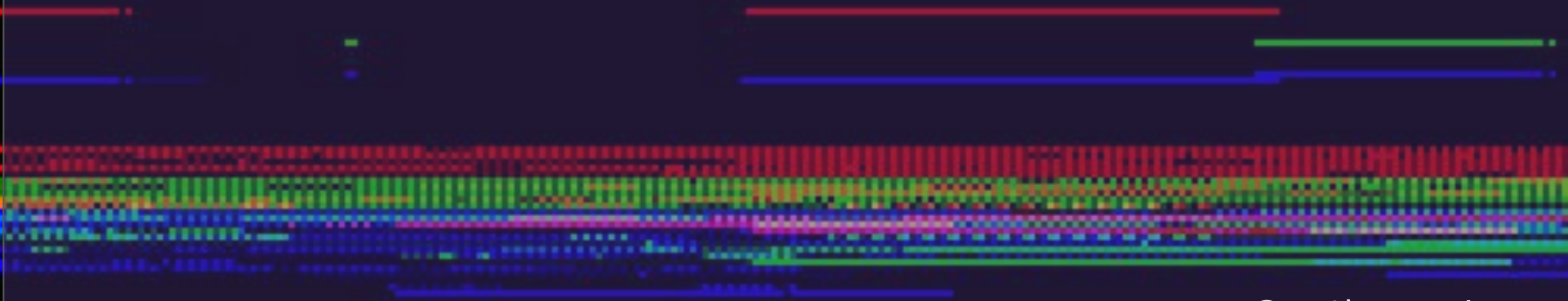
Flexible framework for migrating content into Drupal

from any sources




Our journey with Migrate
Data integrity
Scheduling migration execution
Logging, alerting, reporting
Partial import / Partial update

Data integrity



How to NOT import corrupted data ?

- Use Process Plugins to check the incoming data
- Try custom Process Plugins
- Use `MigrateSkipRowException`

Writing a process plugin | Migrate process plugins | Drupal Wiki guide on Drupal.org
https://www.drupal.org/docs/8/api/migrate-api/migrate-process/writing-a-process-plugin

Sanitize your data prior to synchronization

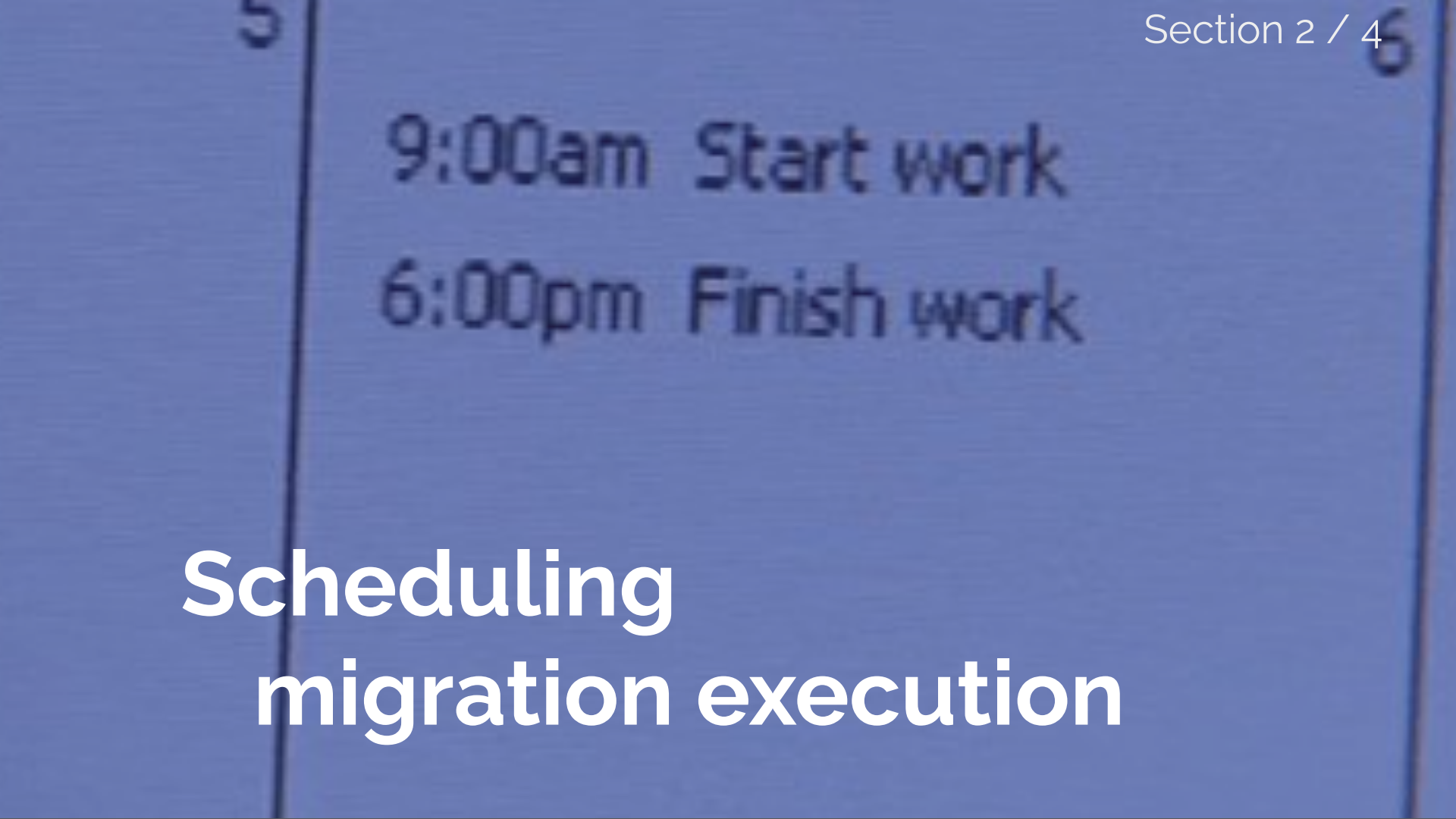
If you don’t, you should probably backup your database **before** running Migrate 🙃🙃

- entity validation on destination plugin
- `skip_on_value` & `skip_on_empty`
- `static_map` & `default_value`

  > failsafe against unexpected values

Good to know

If your endpoint data source is poor quality
you might consider extending the source plugin in order to filter out rows



Scheduling migration execution

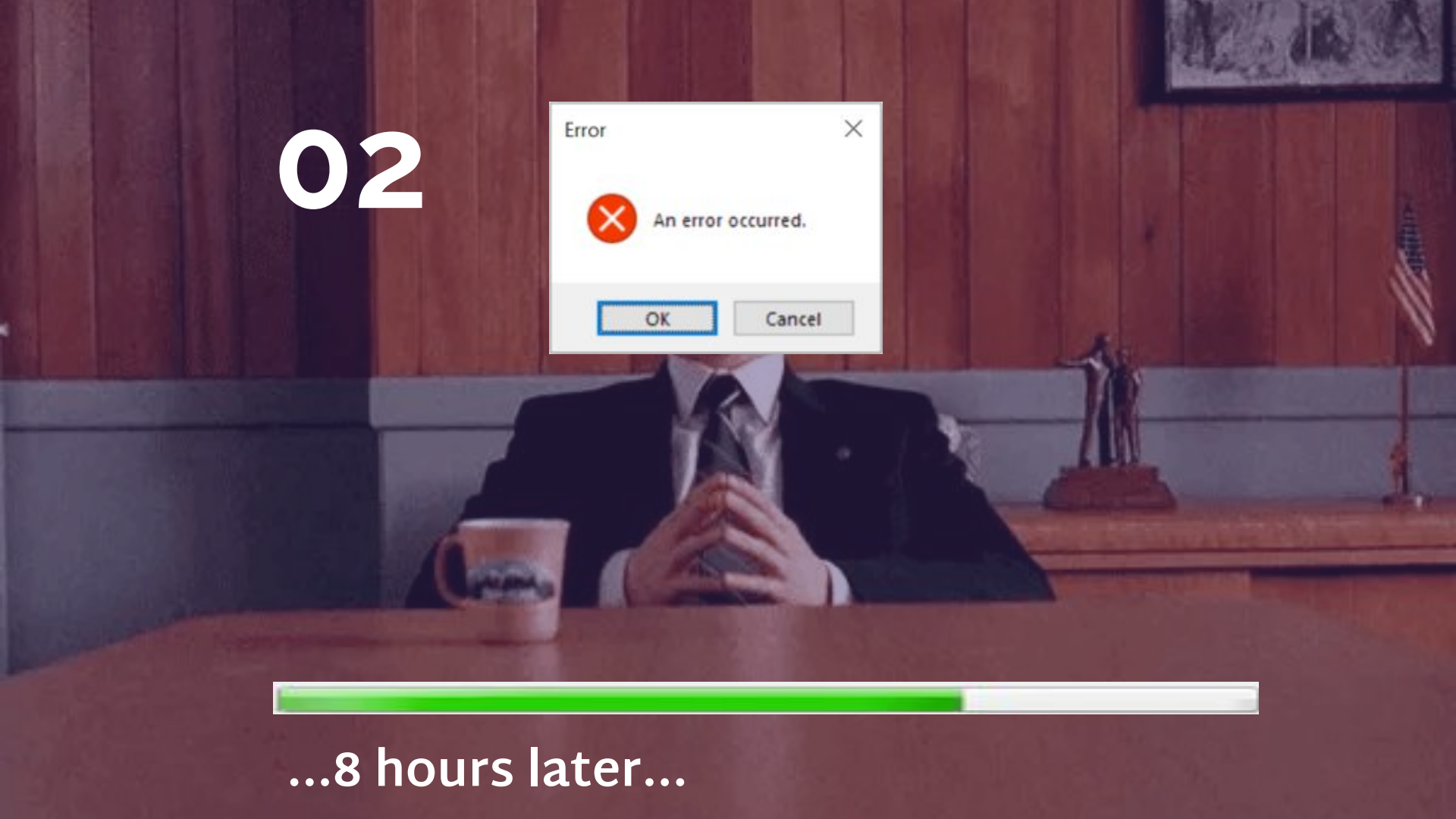
00

No time to waste ?

01

Launch migration over night

02
...8 hours later...


02
Error
An error occurred.
OK
Cancel
...8 hours later...

```
drush migrate:reset-status <migration_name>
drush mrs <migration_name>
```

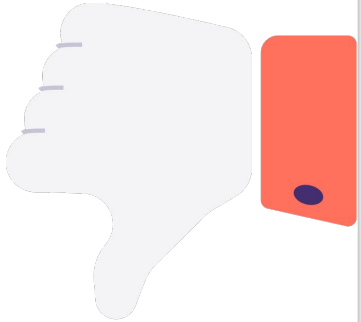
3 ways to execute scheduled migrations

1. use Drush with a regular cron job
   > preferred method
   > straightforward
2. use hook_cron or similar
3. use a module (e.g. migrate_scheduler)

Logging
Alerting
Reporting

How to be alerted when things go south ?

- Wait for client's panicked calls
- Custom Sensors (monitoring module)
- ?????



Garbage in - Garbage out

Because shit happens, be prepared to report !

- Output of the cron jobs and migrate messages
- Log process plugin
- Keep a copy of the source data for debugging

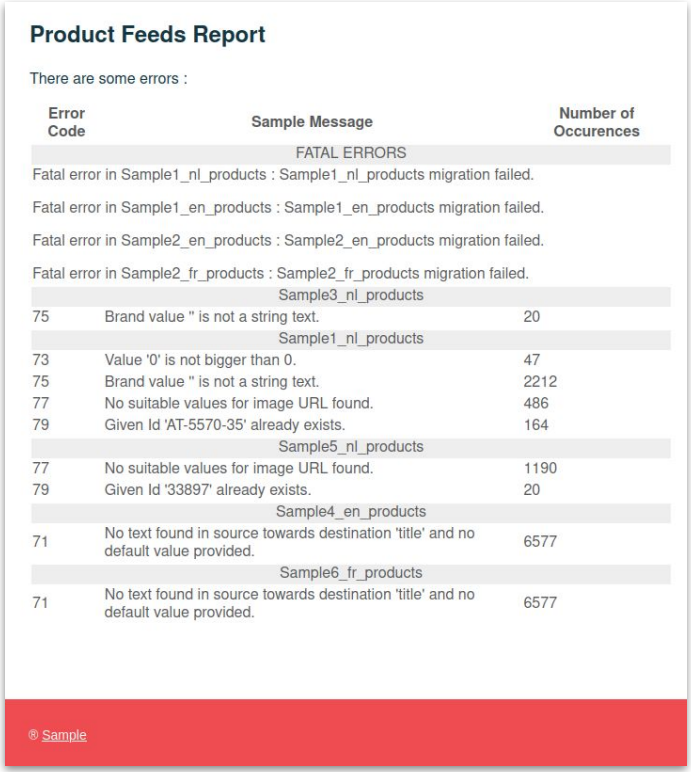
Error reporting example

Adding a custom feature : email report

Needs

- A simple emailable report
- Every type of error
- Explicit error messages
- One email for every daily imports

Result



Product Feeds Report

There are some errors :

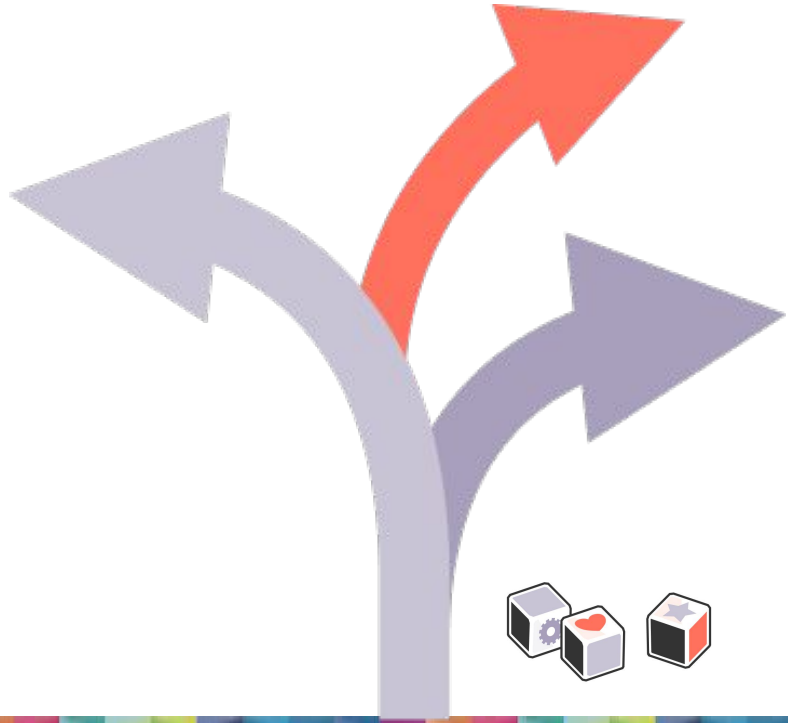
Error Code	Sample Message	Number of Occurences
	FATAL ERRORS	
Fatal error in Sample1_nl_products : Sample1_nl_products migration failed.		
Fatal error in Sample1_en_products : Sample1_en_products migration failed.		
Fatal error in Sample2_en_products : Sample2_en_products migration failed.		
Fatal error in Sample2_fr_products : Sample2_fr_products migration failed.		
	Sample3_nl_products	
75	Brand value '' is not a string text.	20
	Sample1_nl_products	
73	Value '0' is not bigger than 0.	47
75	Brand value '' is not a string text.	2212
77	No suitable values for image URL found.	486
79	Given Id 'AT-5570-35' already exists.	164
	Sample5_nl_products	
77	No suitable values for image URL found.	1190
79	Given Id '33897' already exists.	20
	Sample4_en_products	
71	No text found in source towards destination 'title' and no default value provided.	6577
	Sample6_fr_products	
71	No text found in source towards destination 'title' and no default value provided.	6577

® Sample

Partial import
Partial update

How to update conditionally or a subset of records only ?

- **--update drush flag**
  - when you want to update everything
- `high_water_property` in the source plugin
  - when you have a “last modified” field
- `track_changes` in the source plugin
  - when you don’t have it or it’s not applicable
- **--sync drush flag**
  - when you want to rollback items not in the source anymore

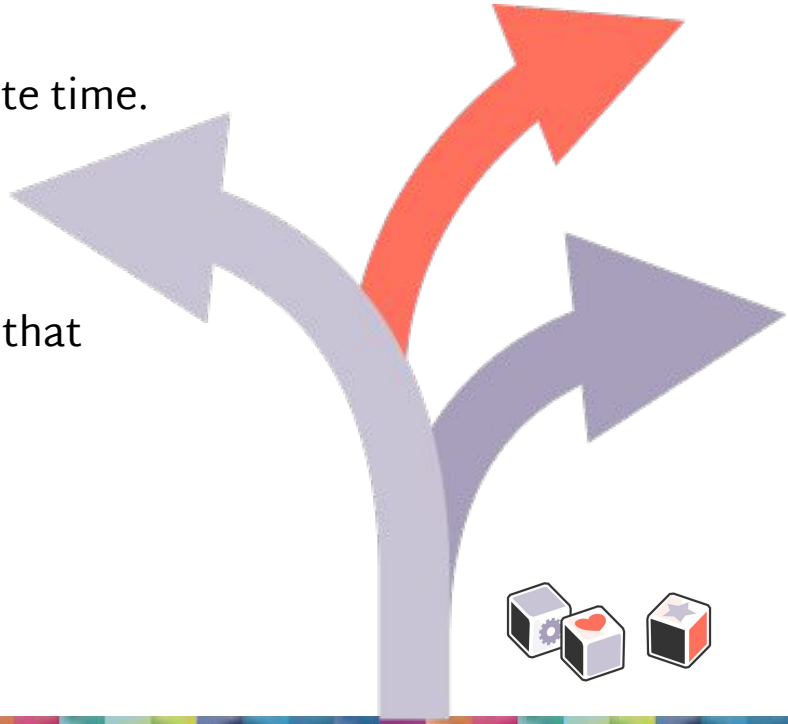


Avoid importing data that already exists!

When importing millions, you can't afford to waste time.

Use the `track_changes` feature to exclude data that has already been imported

- Implement skip using source hash
- Compatible with sync



You probably should use migrate if...

- Having regular entities is important
- The requirements are somewhat open ended and you plan on leveraging the migrate ecosystem (tooling, community support, ...)
- The asynchronous nature (no real time data fetching) of Migrate is an advantage or you don't care

Questions ?

arthur@webstanz.be

adelson@webstanz.be

DrupalCon
EUROPE2021
OCTOBER 4-7

Join us for contribution opportunities

Online event

https://drupalcontributions.opensocial.site

General, Mentored & Firstime Contributor Workshop

4th of October: 9:00 - 12:00
5th of October: 14:00 - 17:00
6th of October: 9:00 - 12:00
7th of October: 14:00 - 17:00

#DrupalContribution

Want to know more about Migrate ?

Get involved & contribute !

Join the Drupal Migrate' Slack channel

Big shout out
to all contributors !

bit.ly/3mmqGLt
Thank you !

DrupalCon
EUROPE2021
OCTOBER 4-7

What did you think?

Locate this session at the DrupalCon Europe 2021 website
and take the survey for each session:

https://events.drupal.org/europe2021/schedule/all