vamos'24 | u^b

Feb 7-9, Bern, Switzerland

Variability in data transformation: towards data migration product lines

David Romero-Organvidez, David Benavides, Jose-Miguel Horcas, María Teresa Gómez-López

Roadmap

1. Introduction
2. Data migration background
3. Our solution: an SPL for data transformation
4. Transformation product line for relational databases
5. Proof of concepts
6. Implementation
7. Conclusions and future work

Roadmap

1. **Introduction**
2. Data migration background
3. Our solution: an SPL for data transformation
4. Transformation product line for relational databases
5. Proof of concepts
6. Implementation
7. Conclusions and future work

1. Introduction

Software modernization

Fundamental part of software modernization:

data migration

1. Introduction

Software modernization

but in many migration projects…

64% exceed **budget forecasts**

(according to Forbes)

1. Introduction

Software modernization

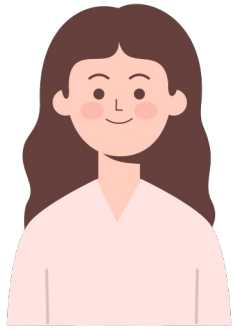
but in many migration projects…

54% exceed **time forecasts**

(according to Forbes)

1. Introduction

Motivating scenario



Ana

Andrea

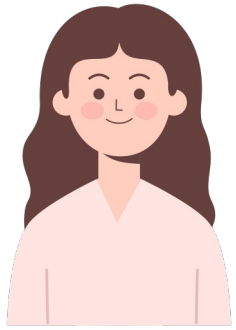
Anthony

Aitana

Axel

1. Introduction

Motivating scenario



Ana

Andrea

Anthony

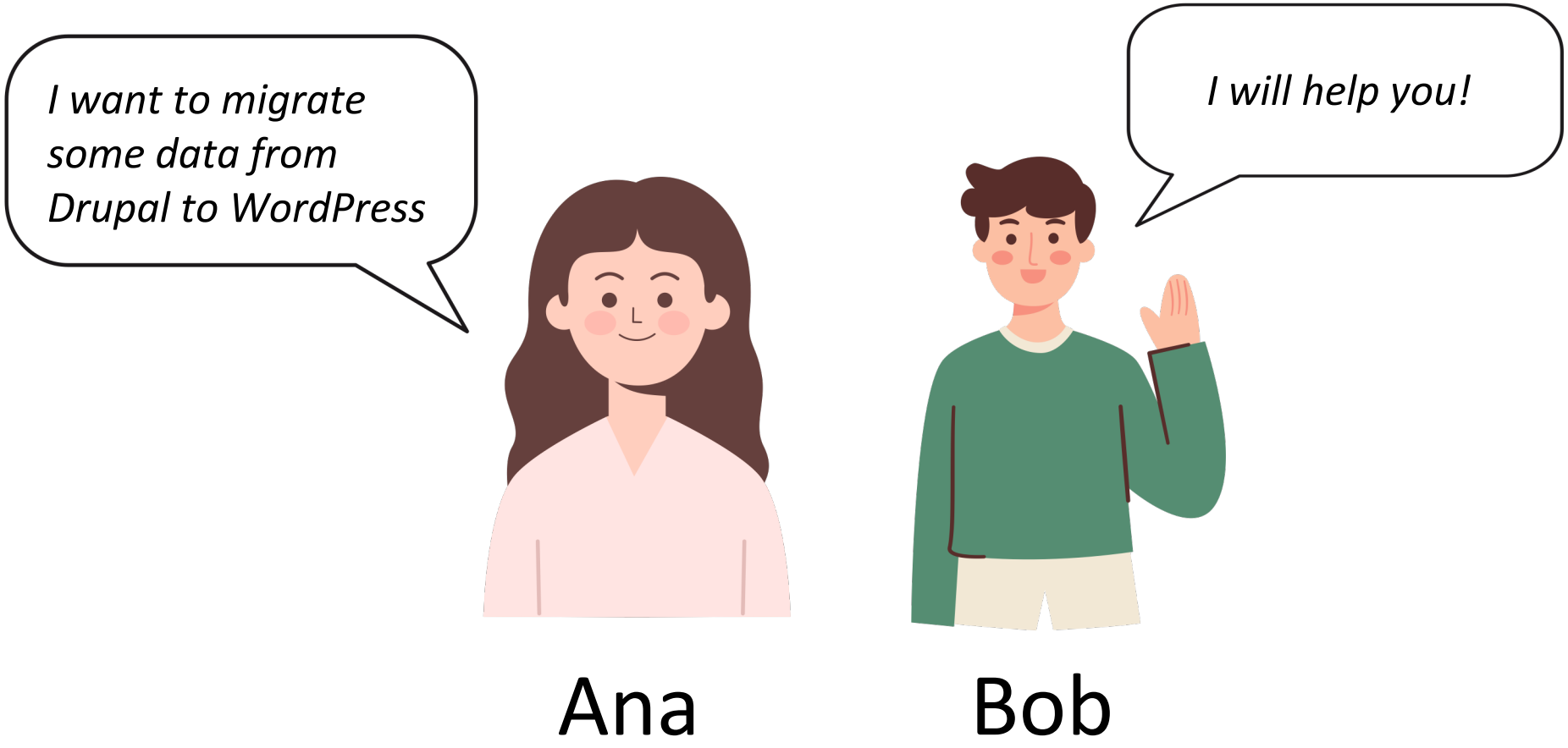
Bob

Aitana

Axel

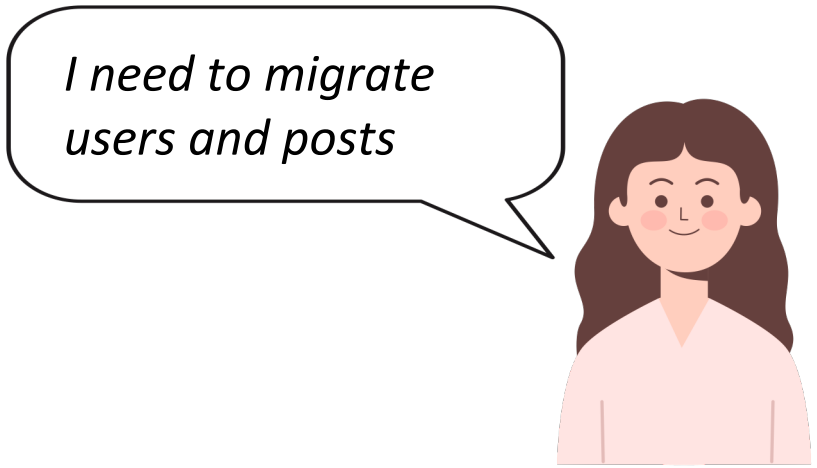
1. Introduction

Motivating scenario



1. Introduction

Motivating scenario

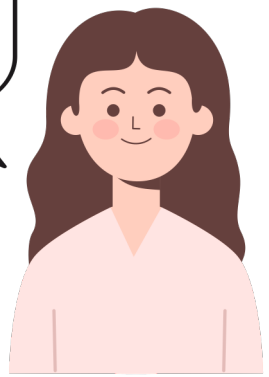


Ana

1. Introduction

Motivating scenario

I need to migrate users and posts



Ana

A little migration script...

Drupal

WordPress

1. Introduction

Motivating scenario

Hey, Bob! I don't need the "last connection" field for users

Andrea

1. Introduction

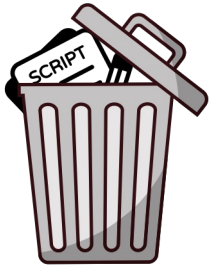
Motivating scenario

Hey, Bob! I don't need the "last connection" field for users

Andrea

clone & own…

Ehm… No problem!



1. Introduction

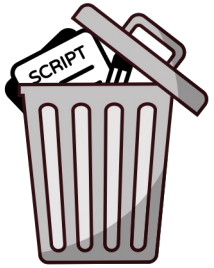
Motivating scenario

Hey, Bob! I don't need the "last connection" field for users

Andrea

clone & own…

Ehm… No problem!



Drupal

WordPress

1. Introduction

Motivating scenario

Bob, I also have to migrate the roles of the users and the articles.

Aitana

1. Introduction

Motivating scenario

Bob, I also have to migrate the roles of the users and the articles.

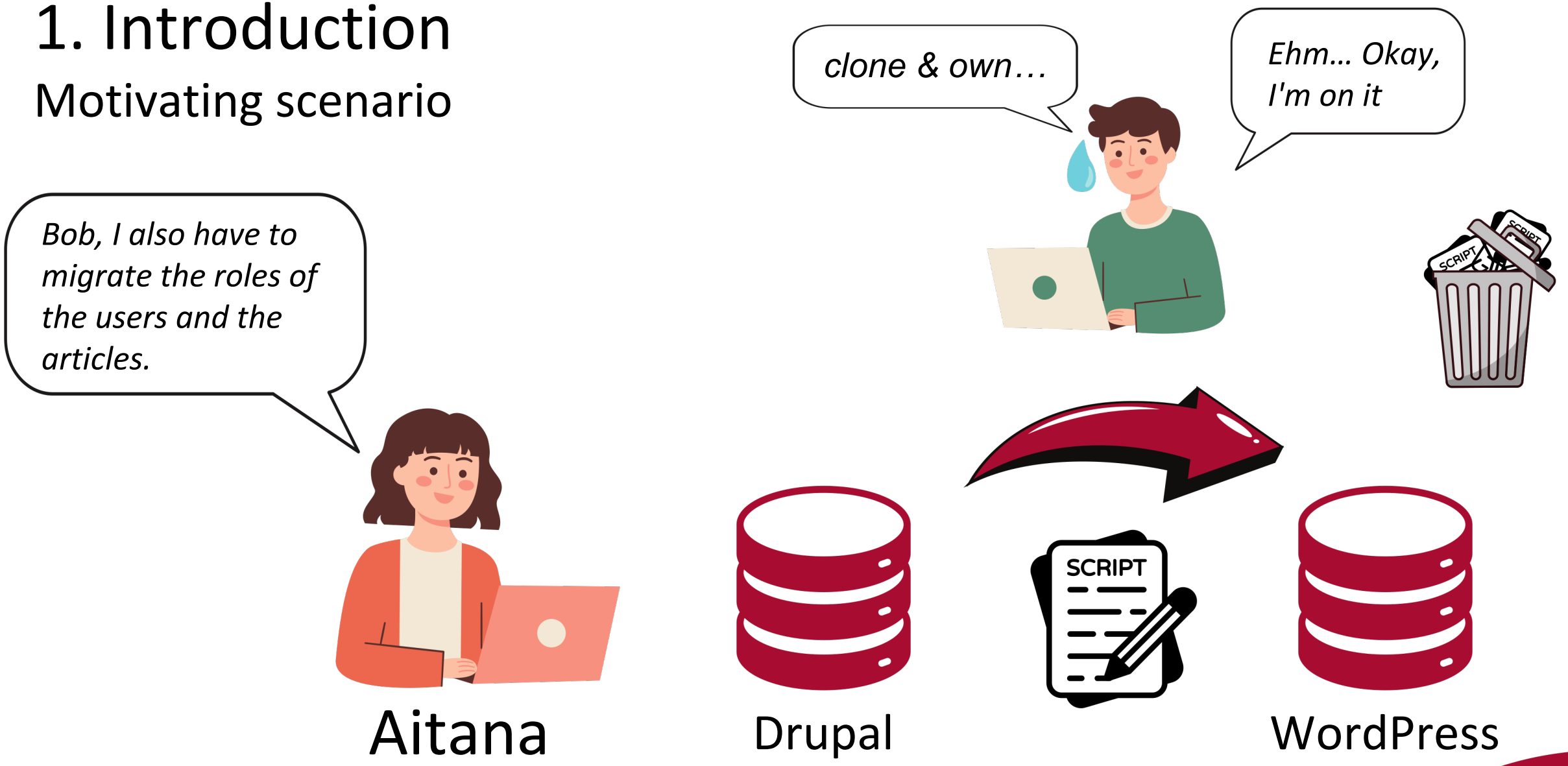
Aitana

clone & own…

Ehm… Okay, I'm on it

1. Introduction

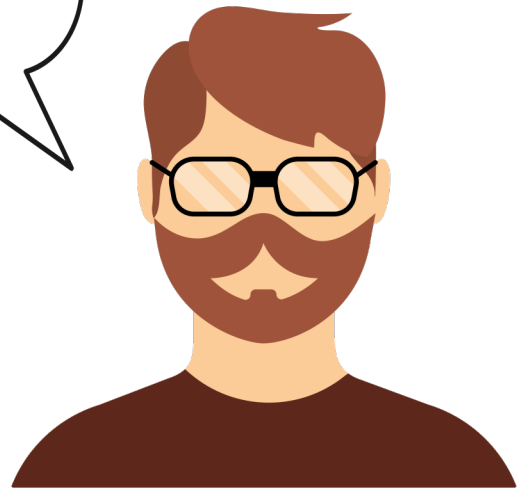
Motivating scenario



1. Introduction

Motivating scenario

While you're at it... can you migrate the comments on the posts? Bob?



Axel

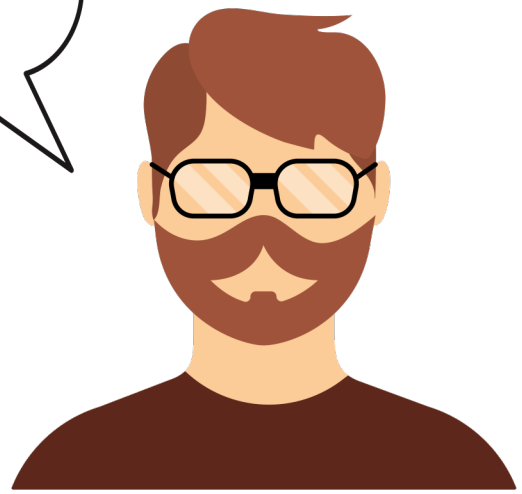
clone & own…

S-s-s-s-s-sure…

1. Introduction

Motivating scenario

While you're at it... can you migrate the comments on the posts? Bob?



Axel

clone & own…

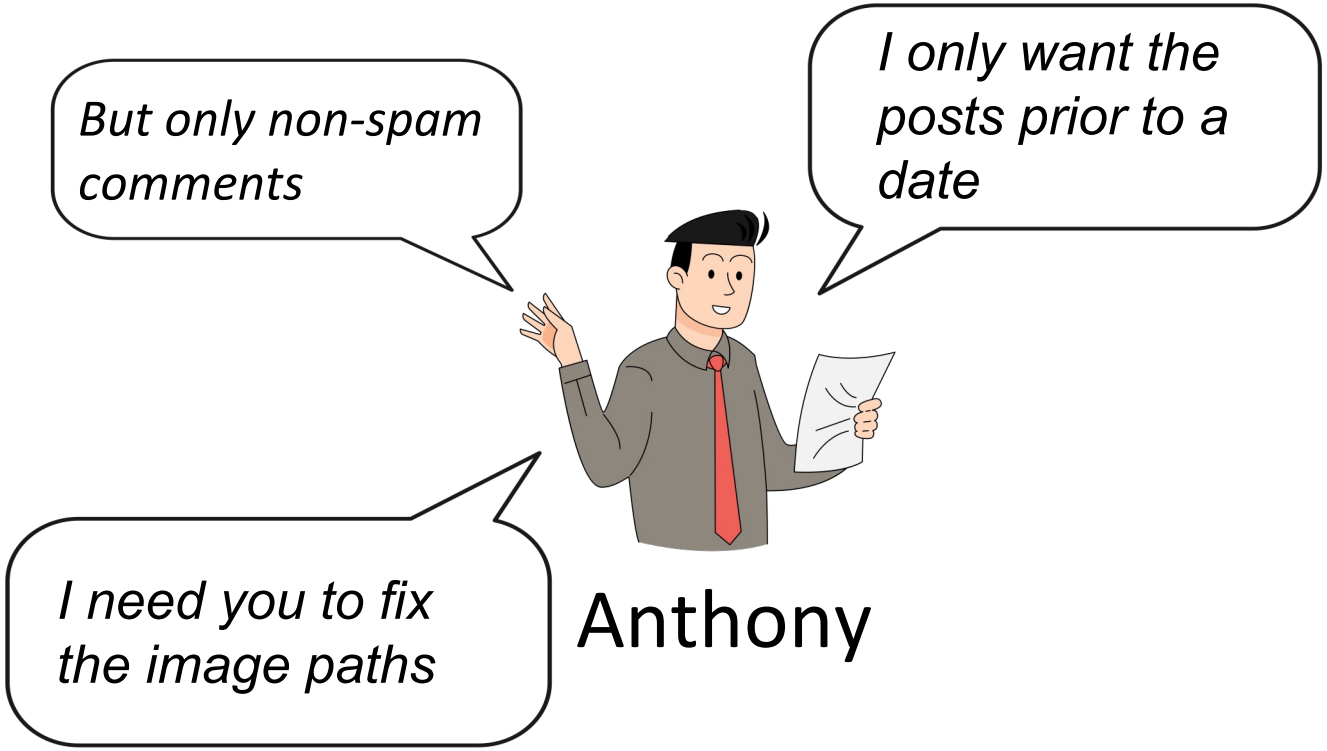
S-s-s-s-s-sure…

Drupal

WordPress

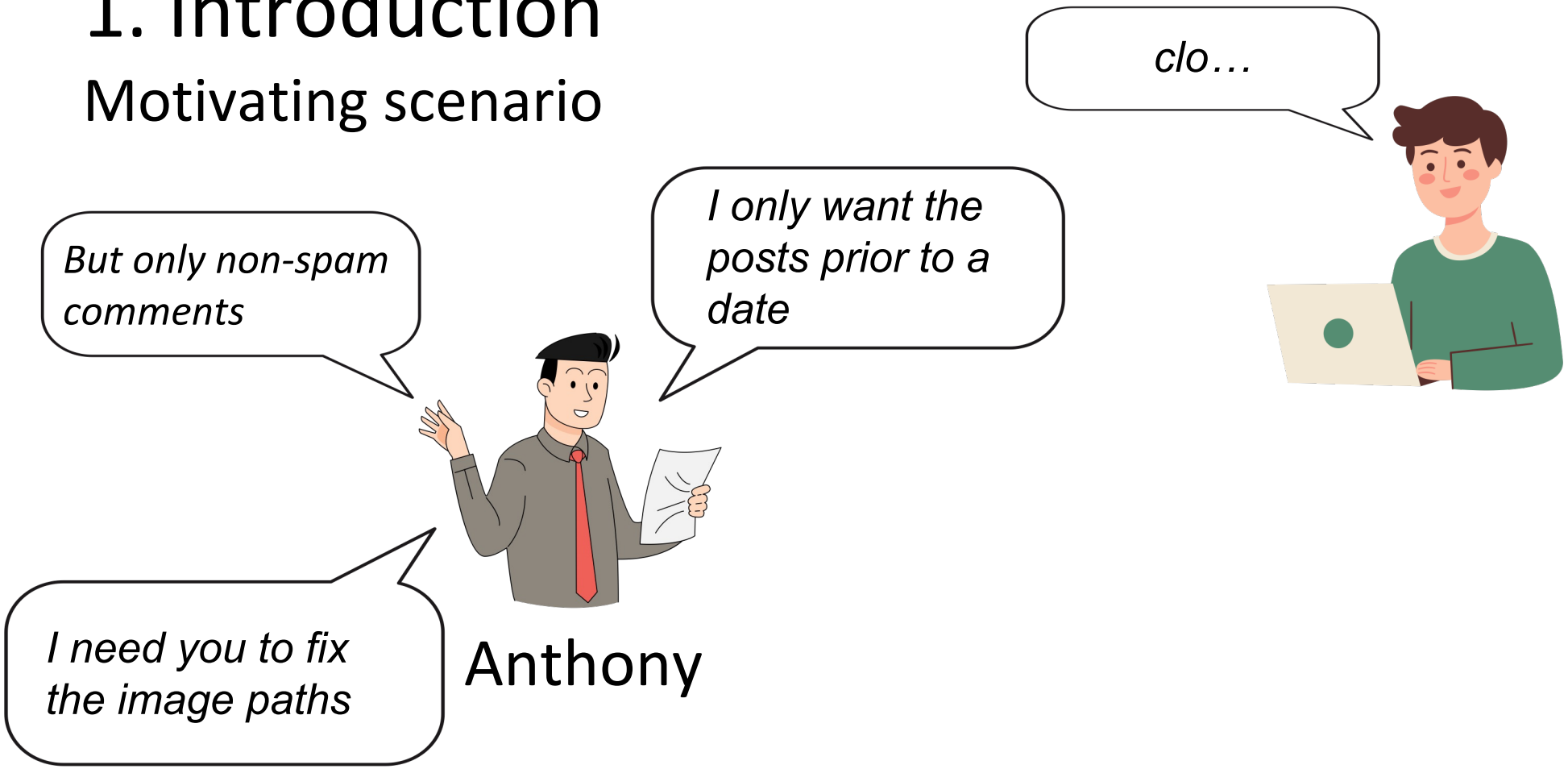
1. Introduction

Motivating scenario



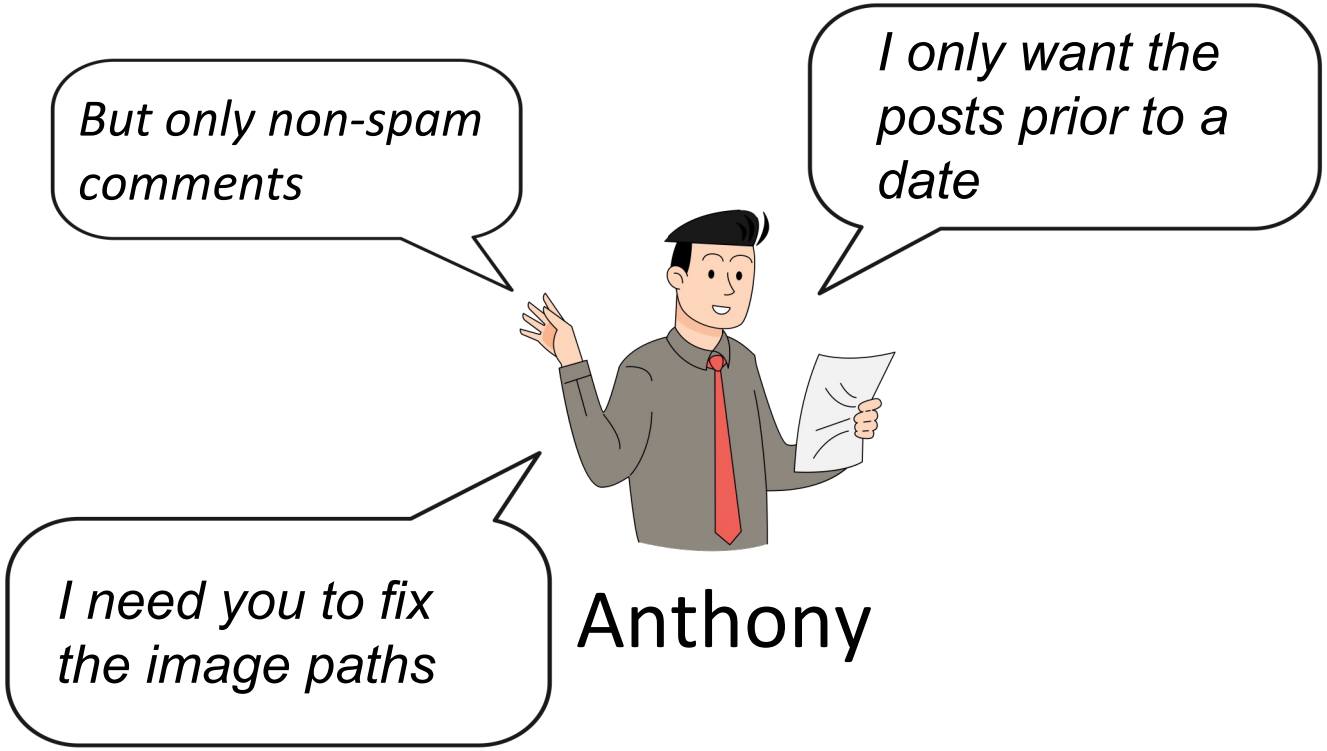
1. Introduction

Motivating scenario



1. Introduction

Motivating scenario



1. Introduction

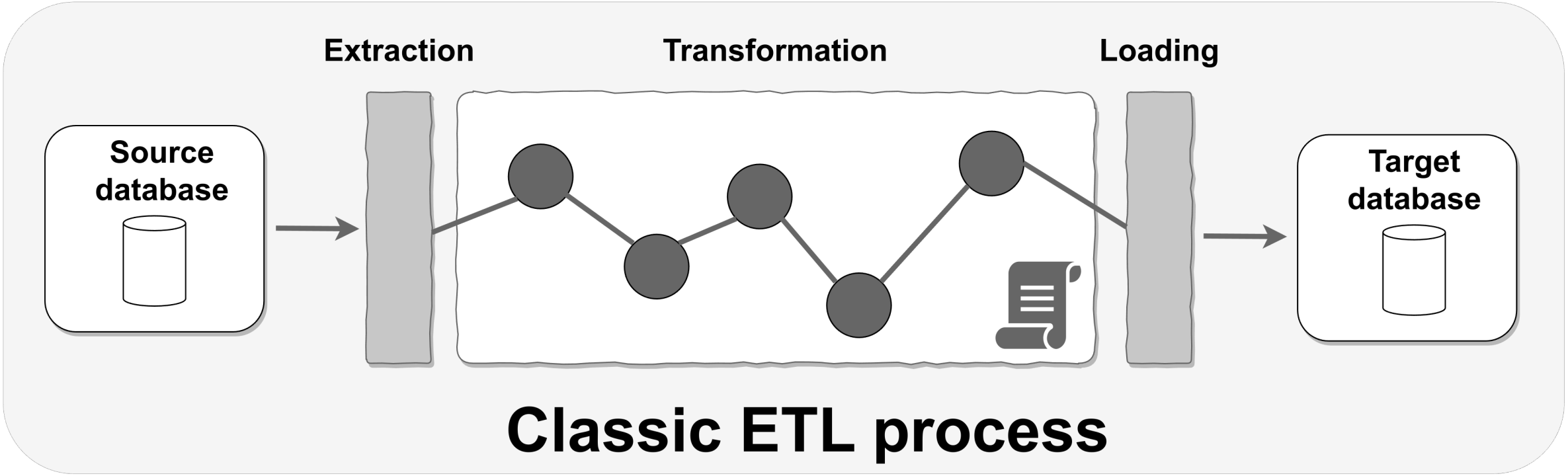
Motivating scenario

Roadmap

1. Introduction
2. **Data migration background**
3. Our solution: an SPL for data transformation
4. Transformation product line for relational databases
5. Proof of concepts
6. Implementation
7. Conclusions and future work

2. Data migration background

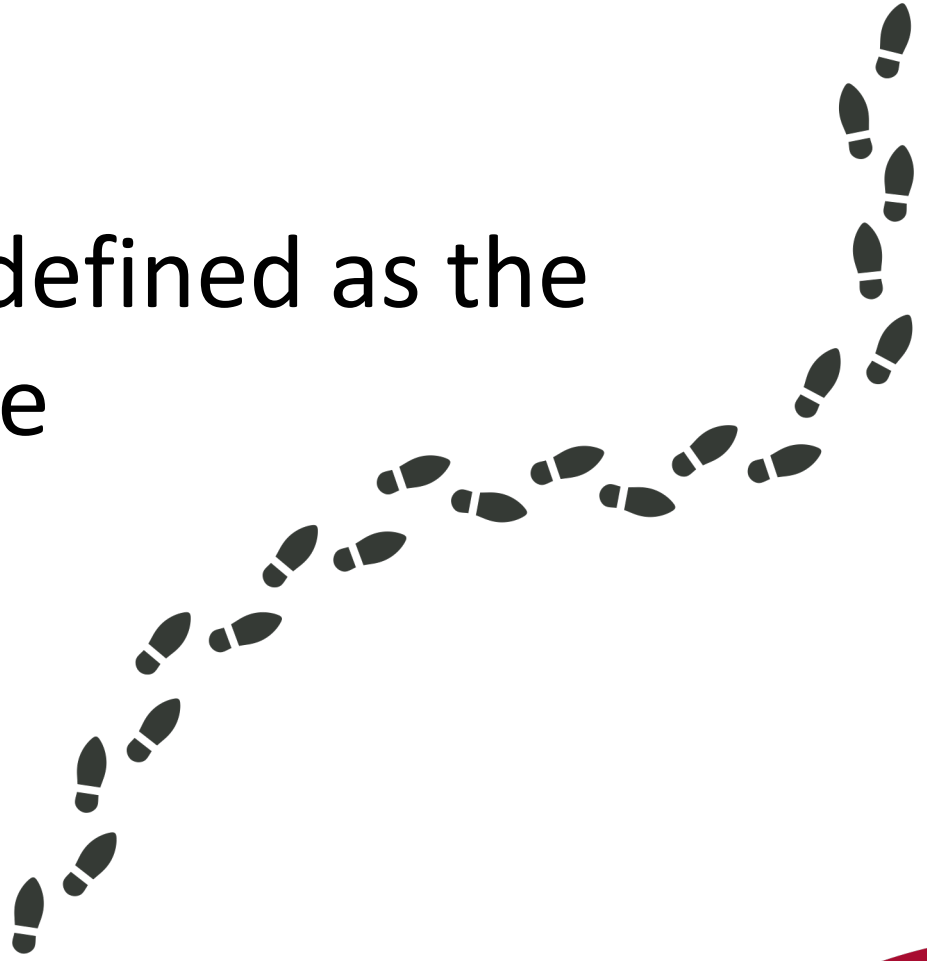
Data migration phases



2. Data migration background

Data migration strategies

A ***migration transformation*** is defined as the composition of a finite sequence of **transformation steps**



2. Data migration background

Data migration strategies

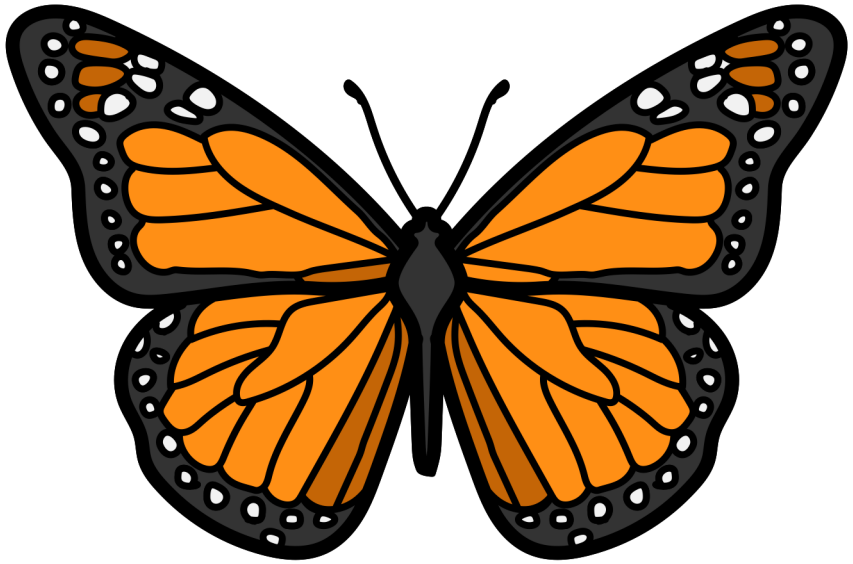
Big Bang

- ¡One step!
- Risk of needing a high process preparation
- Problematic

2. Data migration background

Data migration strategies

Butterfly



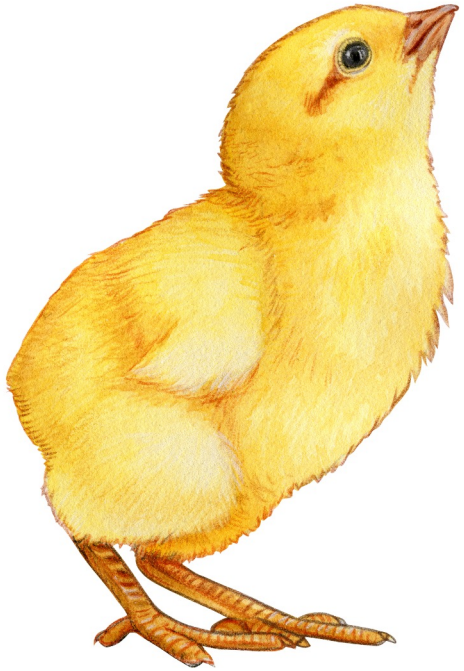
Old ⟷ **New**

- The data of a legacy system is the most important part of the system!
- It separates the target system development and data migration phases

2. Data migration background

Data migration strategies

Chicken little



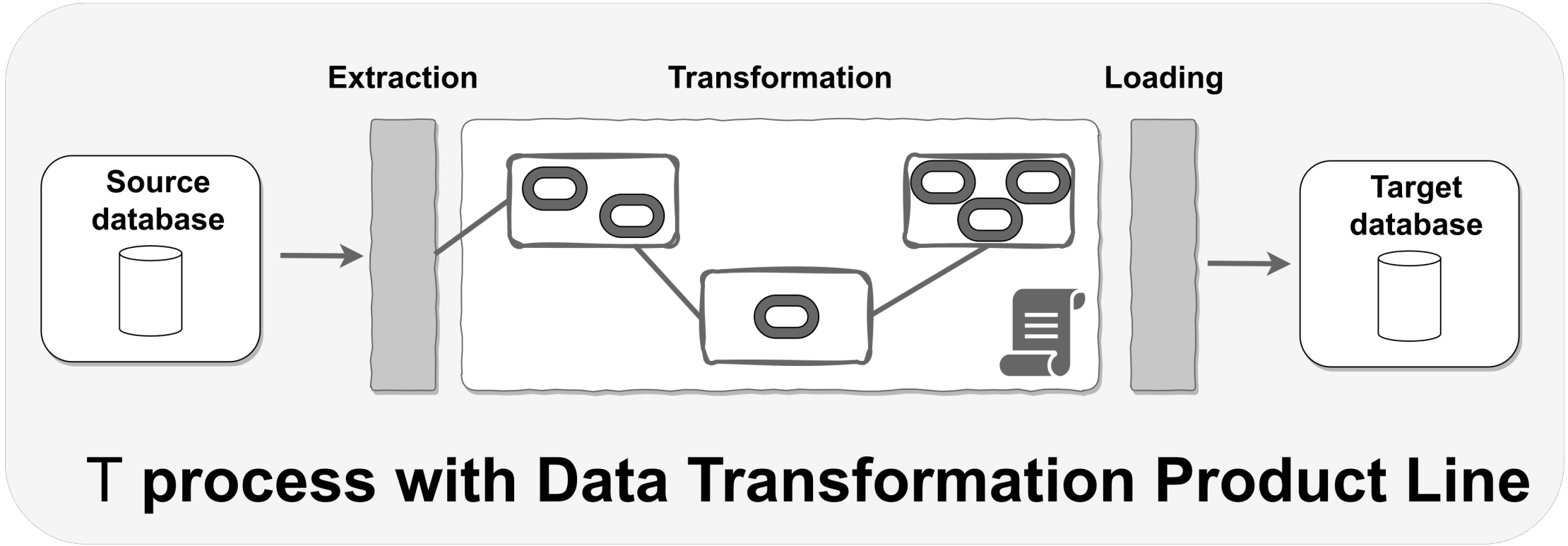
- Migrating the software step-by-step
- Each step produces a significant but controlled change

Roadmap

1. Introduction
2. Data migration background
3. **Our solution: an SPL for data transformation**
4. Transformation product line for relational databases
5. Proof of concepts
6. Implementation
7. Conclusions and future work

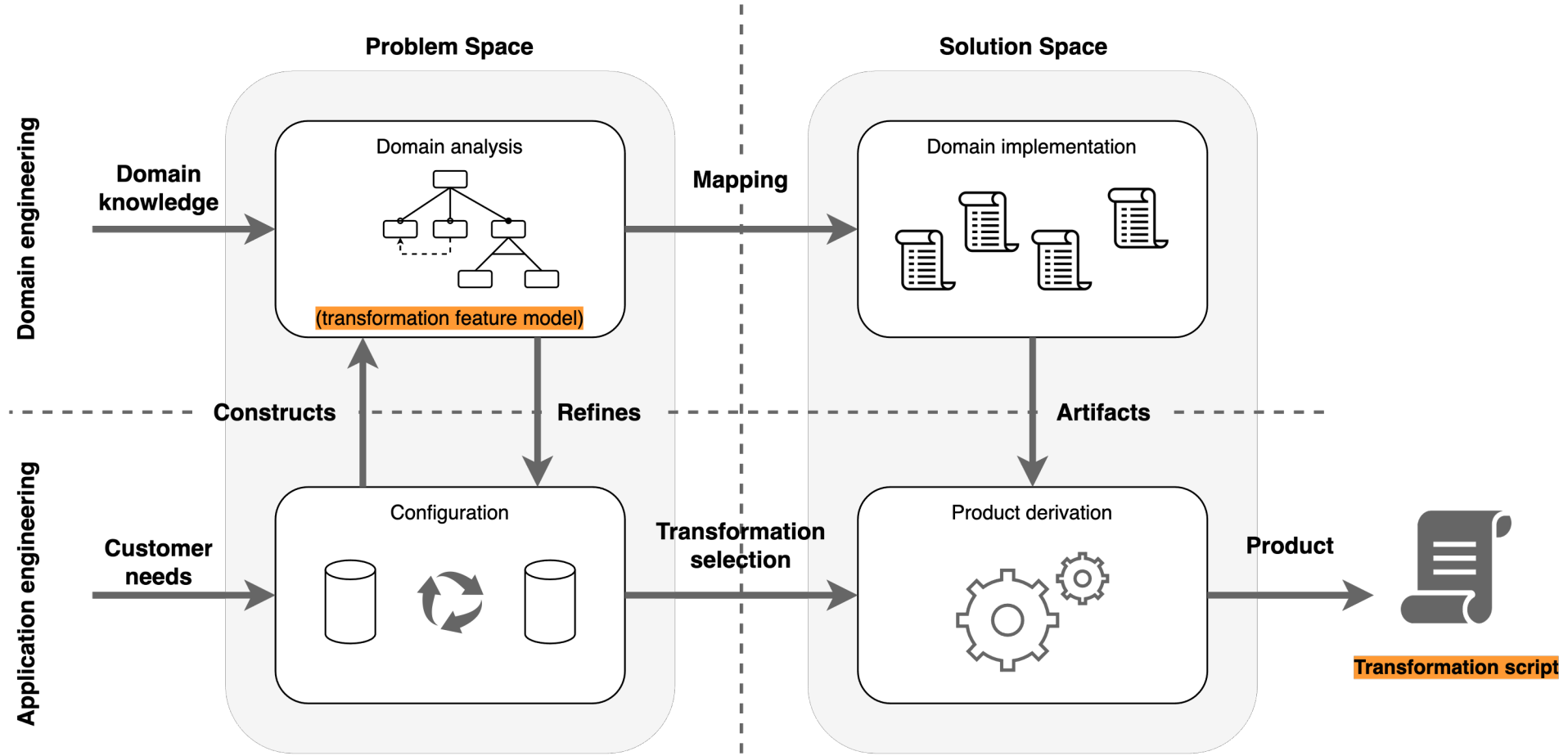
3. Our solution: an SPL for data transformation

Our approach



3. Our solution: an SPL for data transformation

Transformation product line approach



3. Our solution: an SPL for data transformation

Definitions

ACTION ($a \in A$)

Def: *a basic operation that is executed to transfer data from a source database to a target database*

Examples

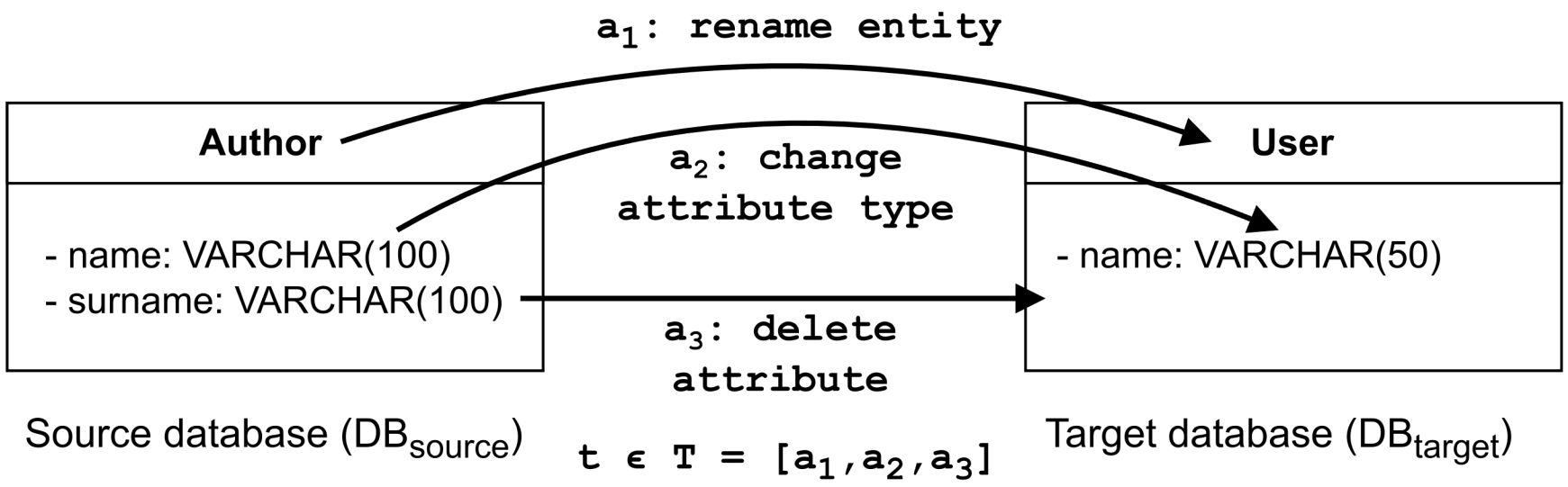
create entity, rename entity, move attribute, update value…

3. Our solution: an SPL for data transformation

Definitions

TRANSFORMATION FEATURE (TF)

Def: *a list of actions to be executed in order*



3. Our solution: an SPL for data transformation

Definitions

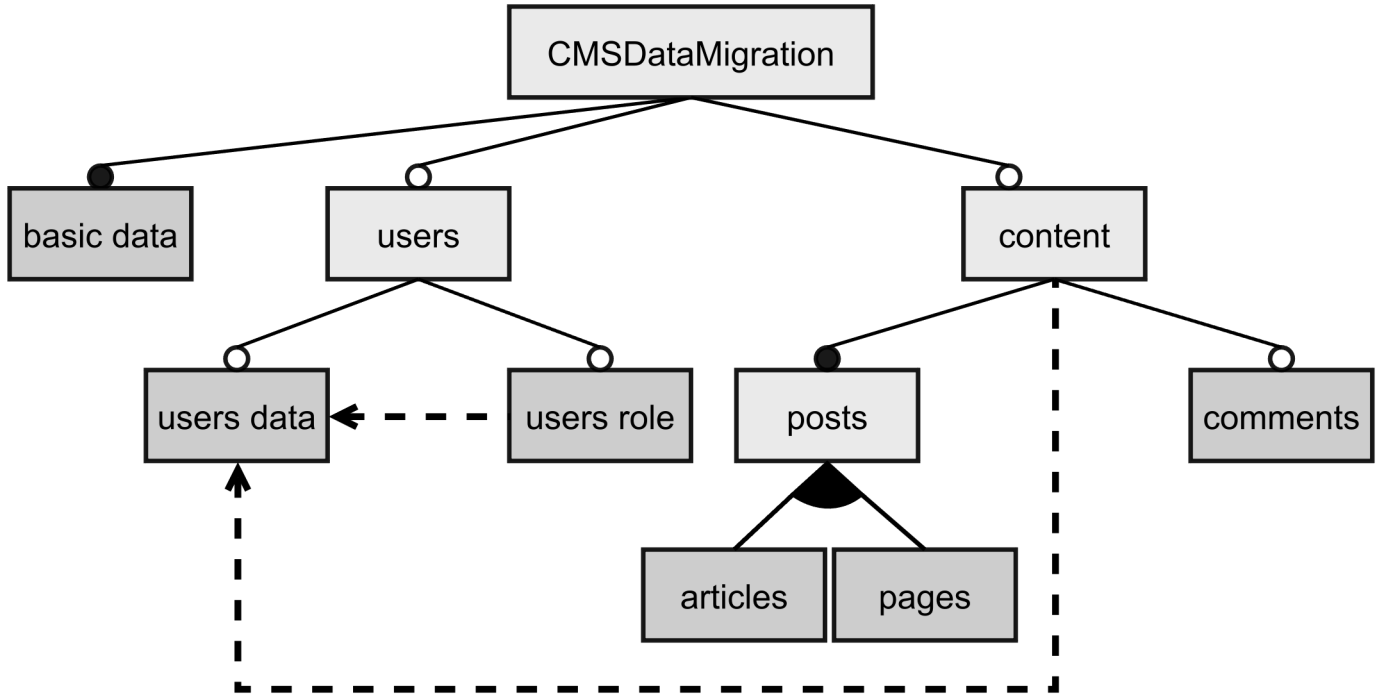
TRANSFORMATION FEATURE MODEL (FM_{TF})

Def: *a feature model* ***where the features are transformation features*** *and the relationships and constraints represent dependencies between these transformation features*

3. Our solution: an SPL for data transformation

Definitions

TRANSFORMATION FEATURE MODEL (FM_{TF})

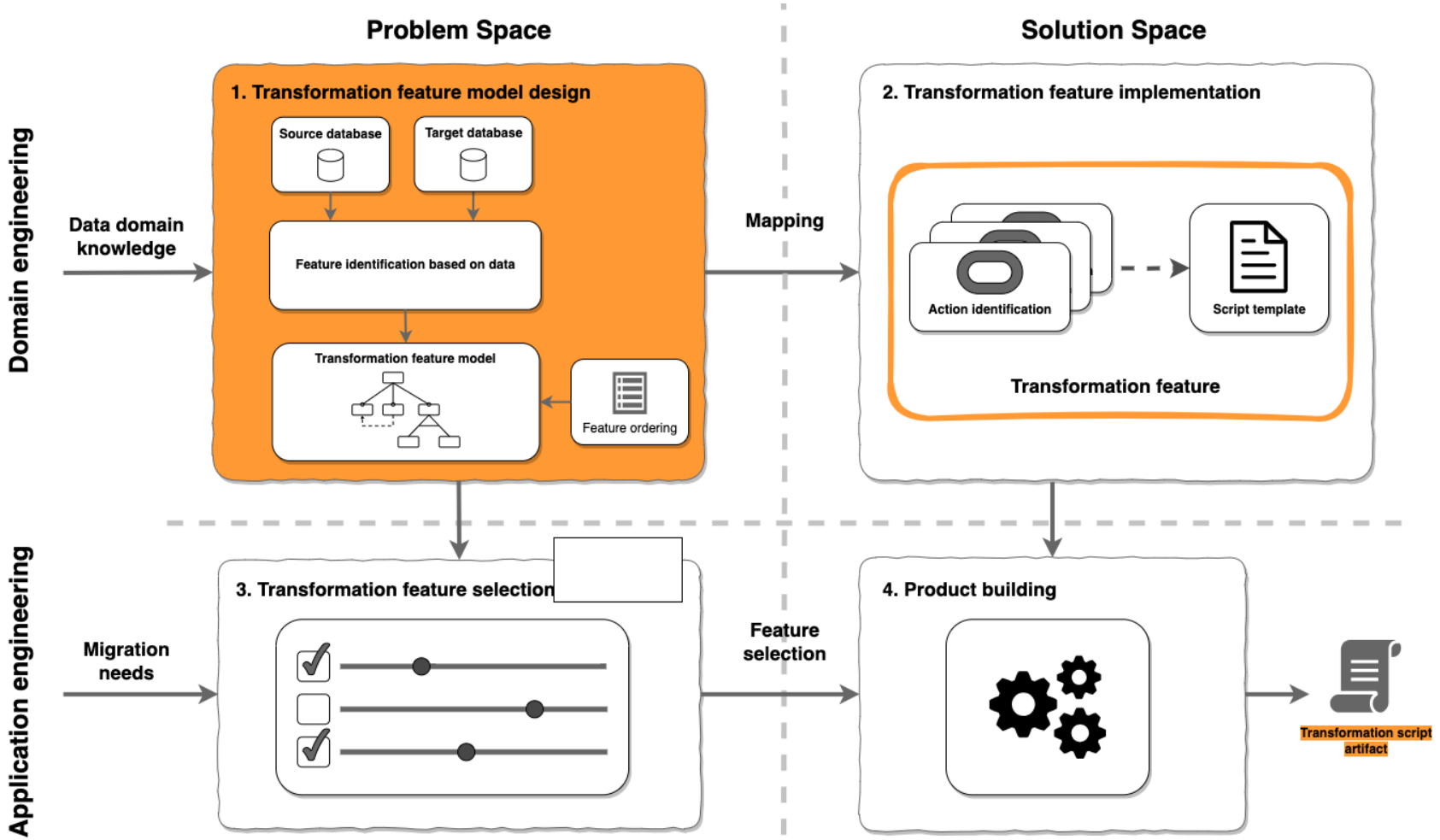


Roadmap

1. Introduction
2. Data migration background
3. Our solution: an SPL for data transformation
4. **Transformation product line for relational databases**
5. Proof of concepts
6. Implementation
7. Conclusions and future work

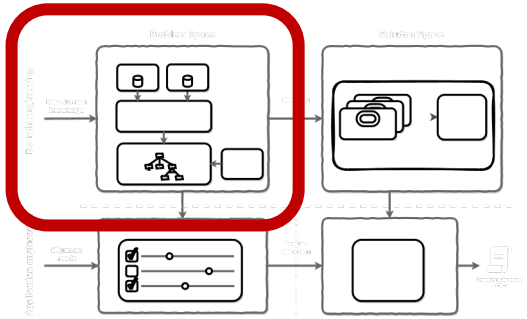
4. Transformational product line for relational databases

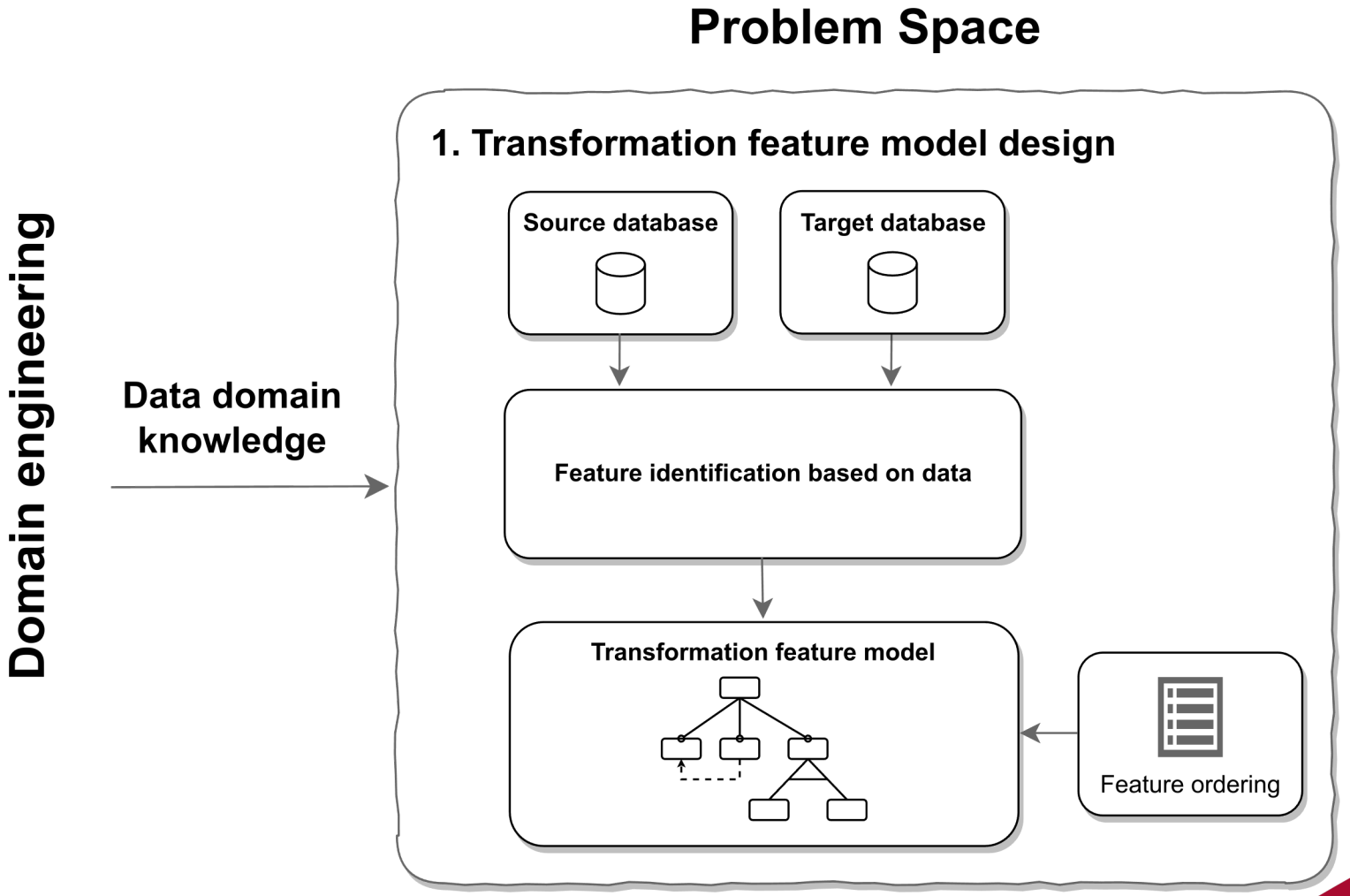
Overview



4. Transformational product line for relational databases

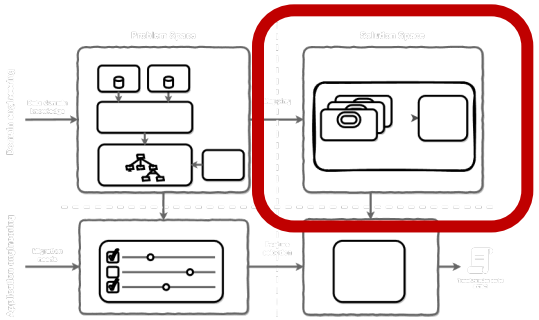
4.1 Transformation feature model design

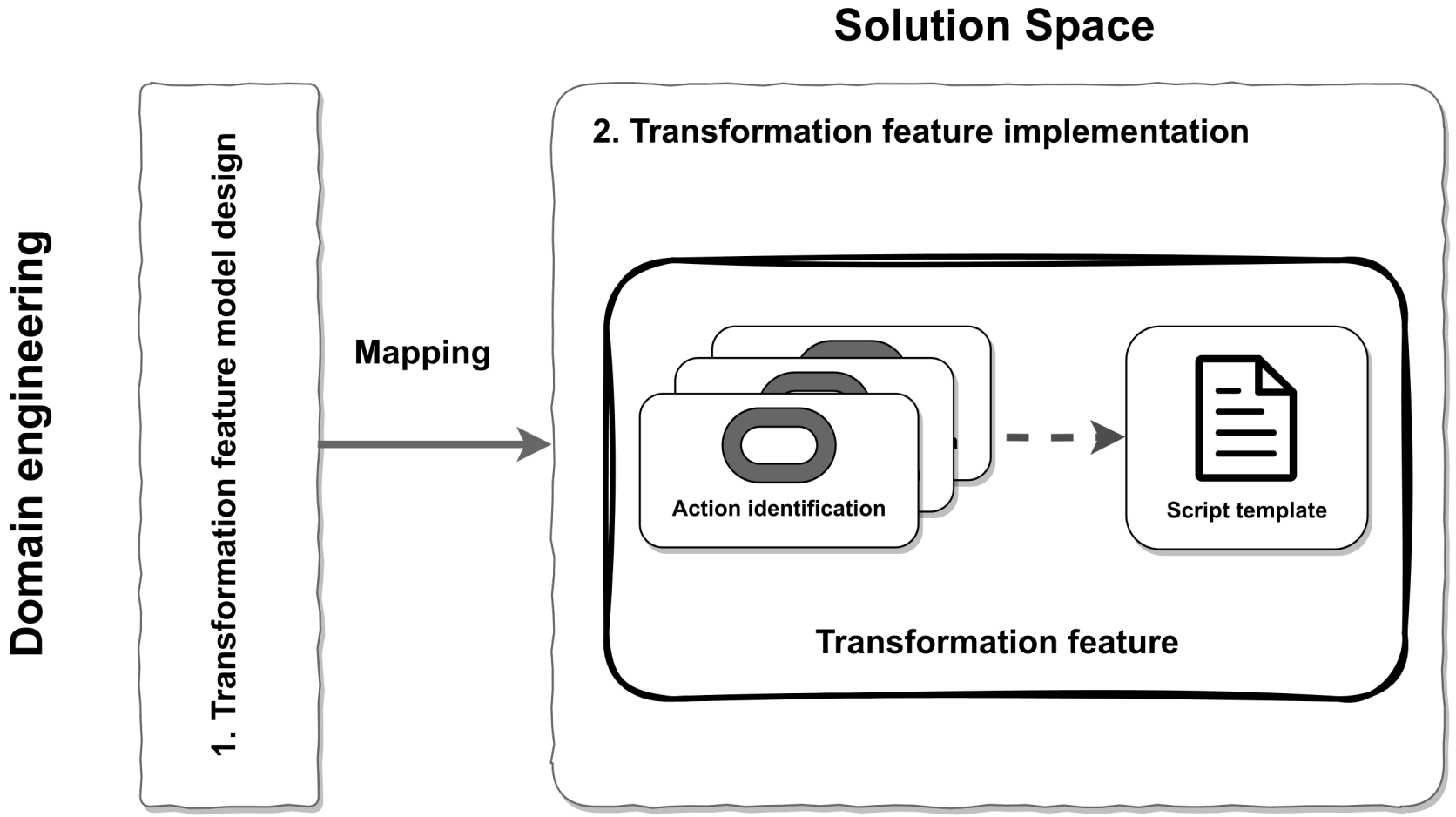




4. Transformational product line for relational databases

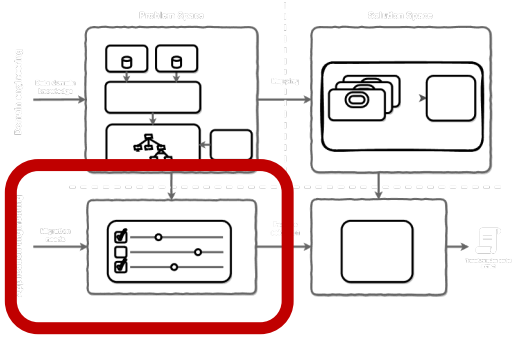
4.2 Transformation feature implementation

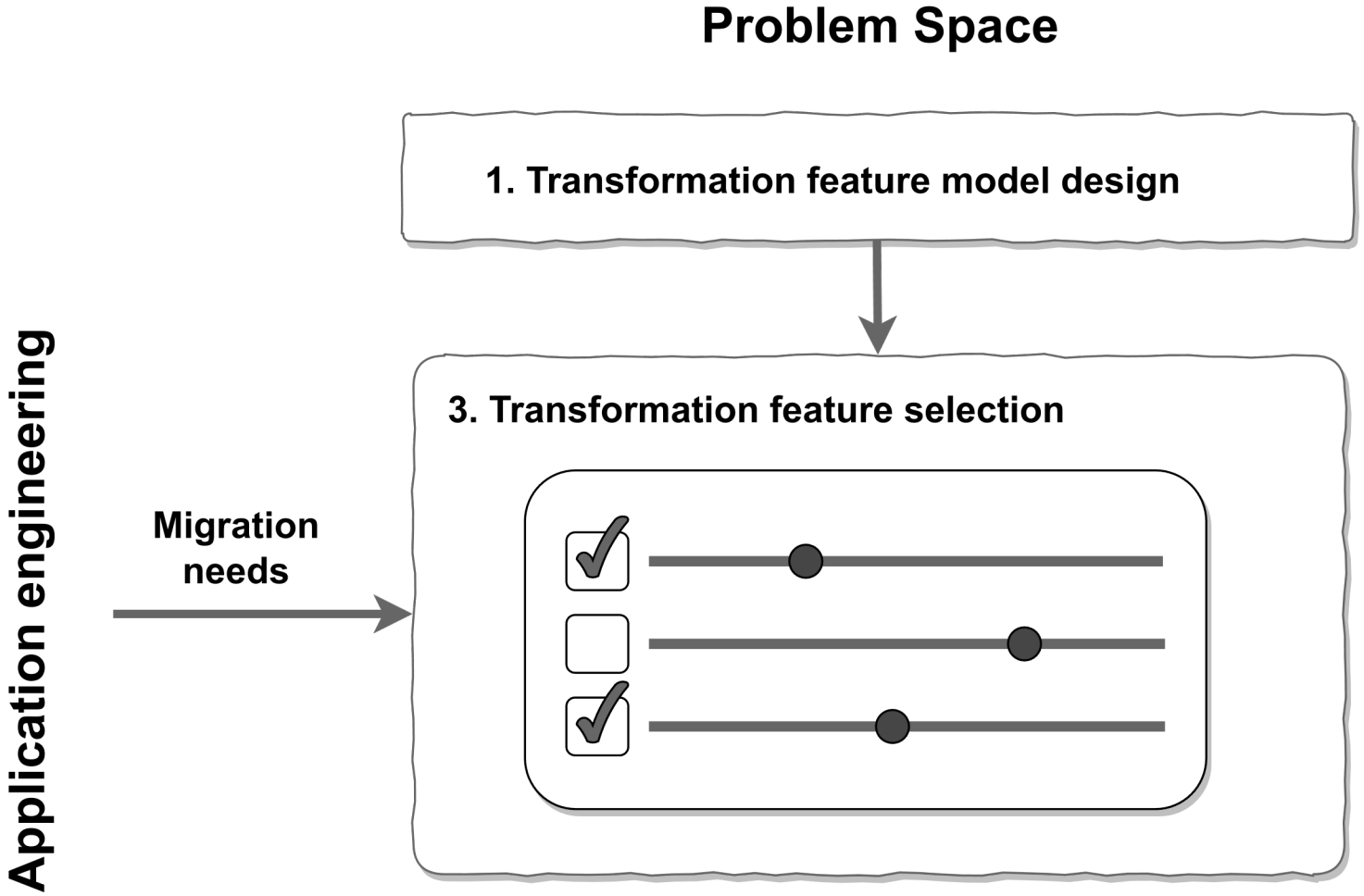




4. Transformational product line for relational databases

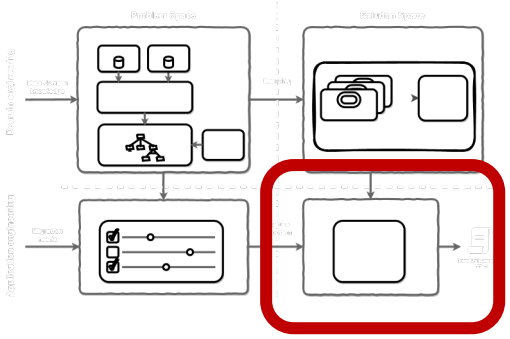
4.3 Transformation feature selection

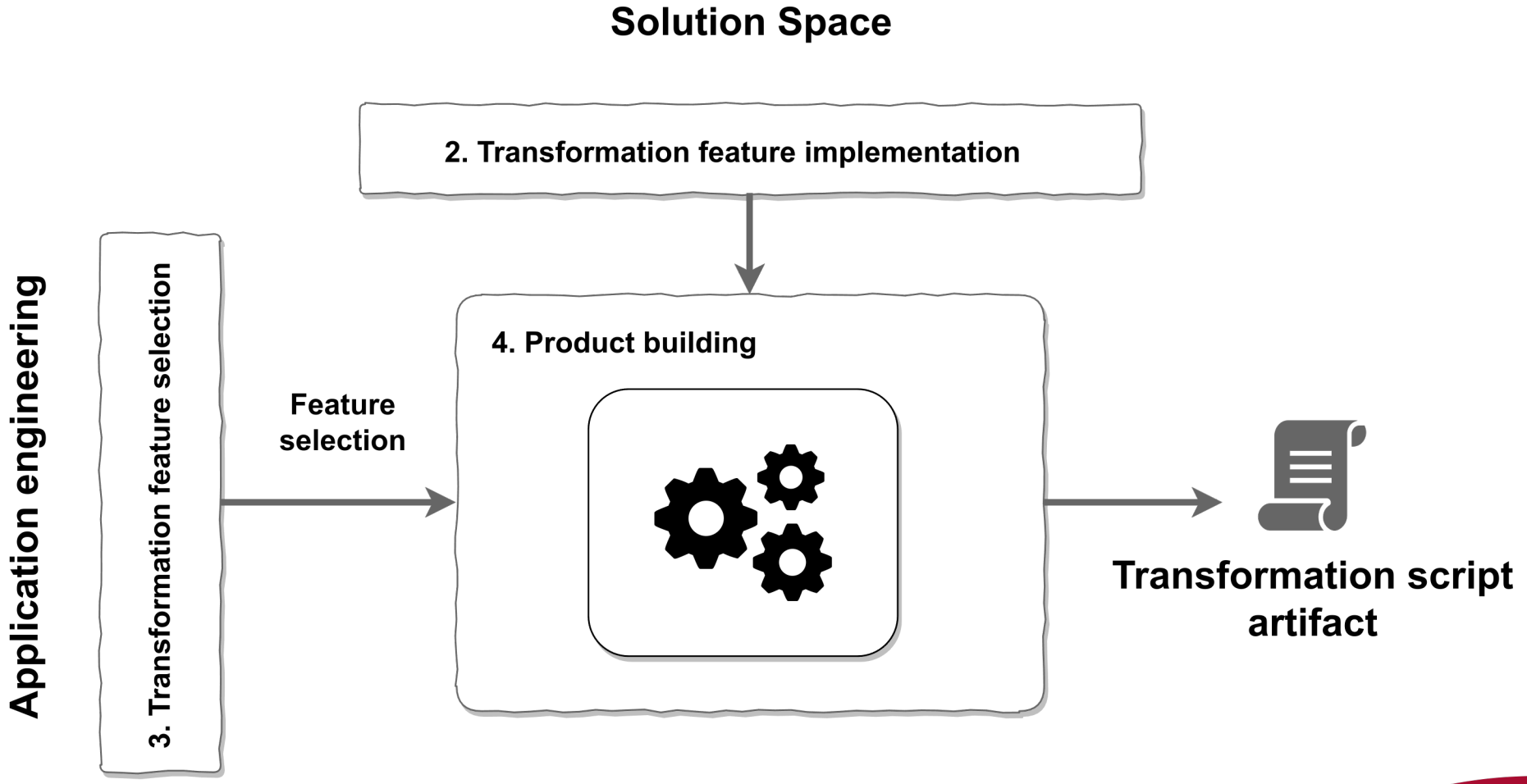




4. Transformational product line for relational databases

4.4 Product building



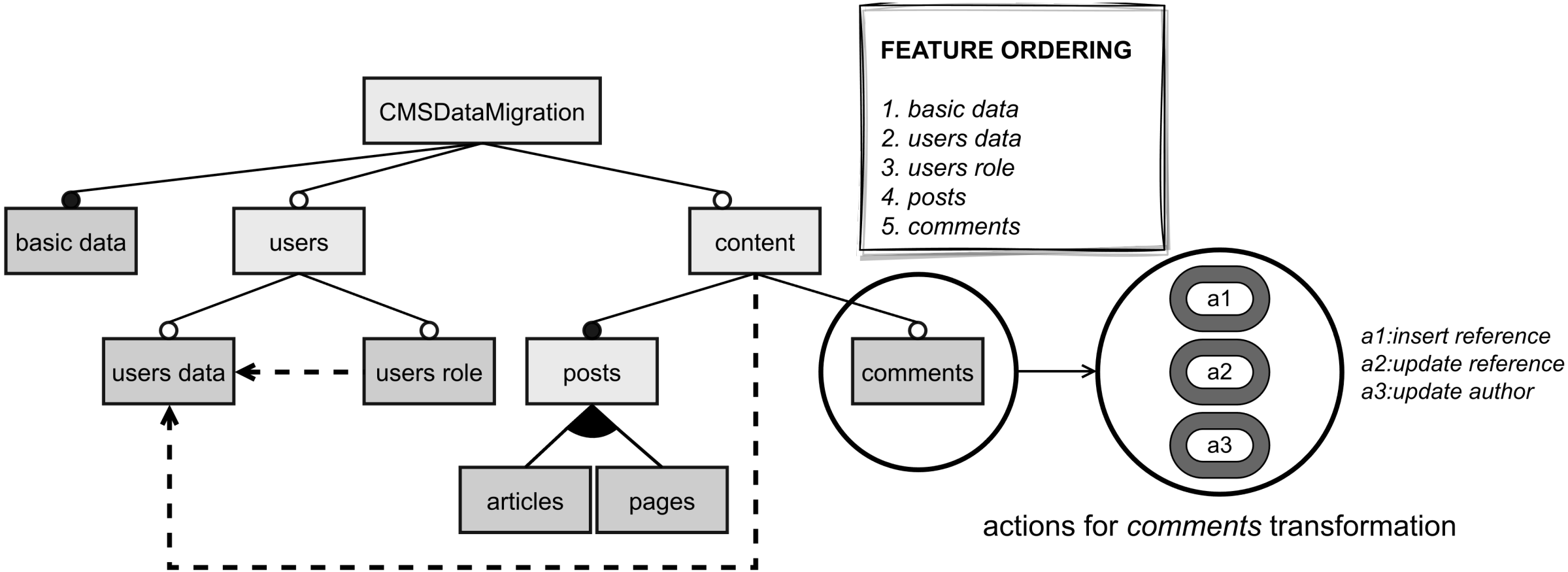


Roadmap

1. Introduction
2. Data migration background
3. Our solution: an SPL for data transformation
4. Transformation product line for relational databases
5. **Proof of concepts**
6. Implementation
7. Conclusions and future work

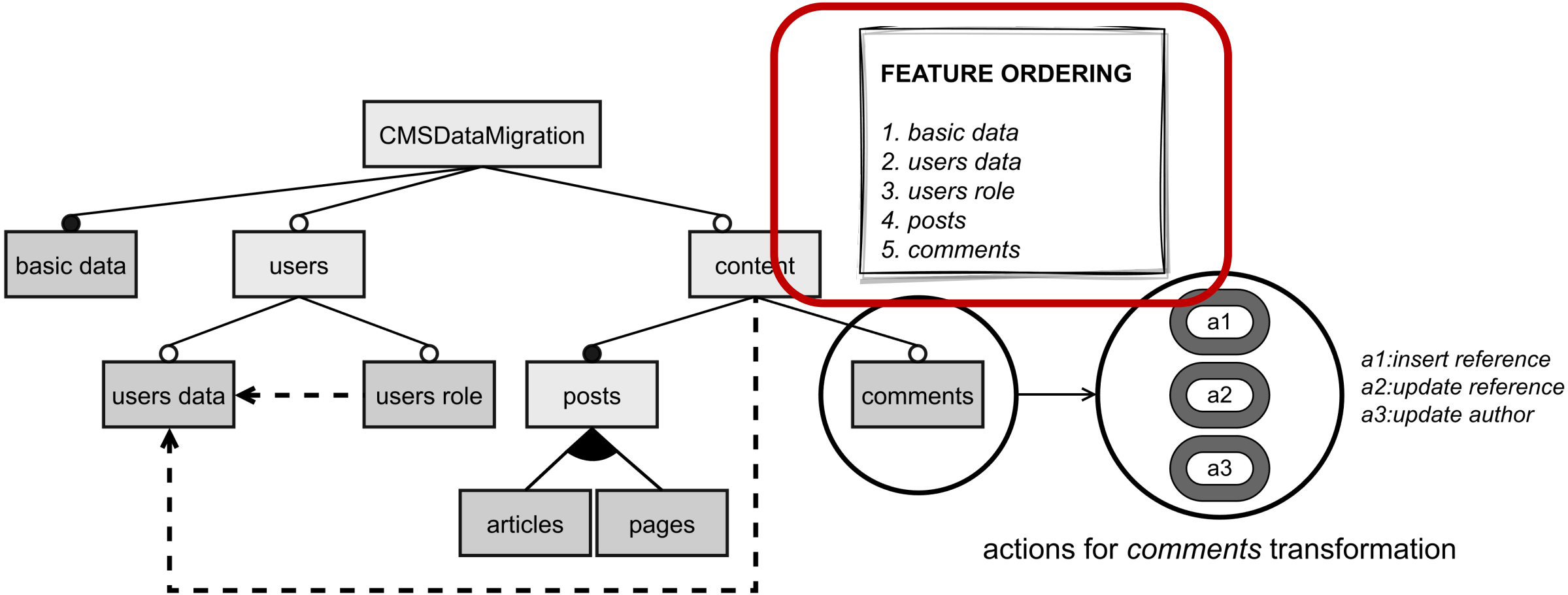
5. Proof of concepts

Validation: Drupal to WordPress



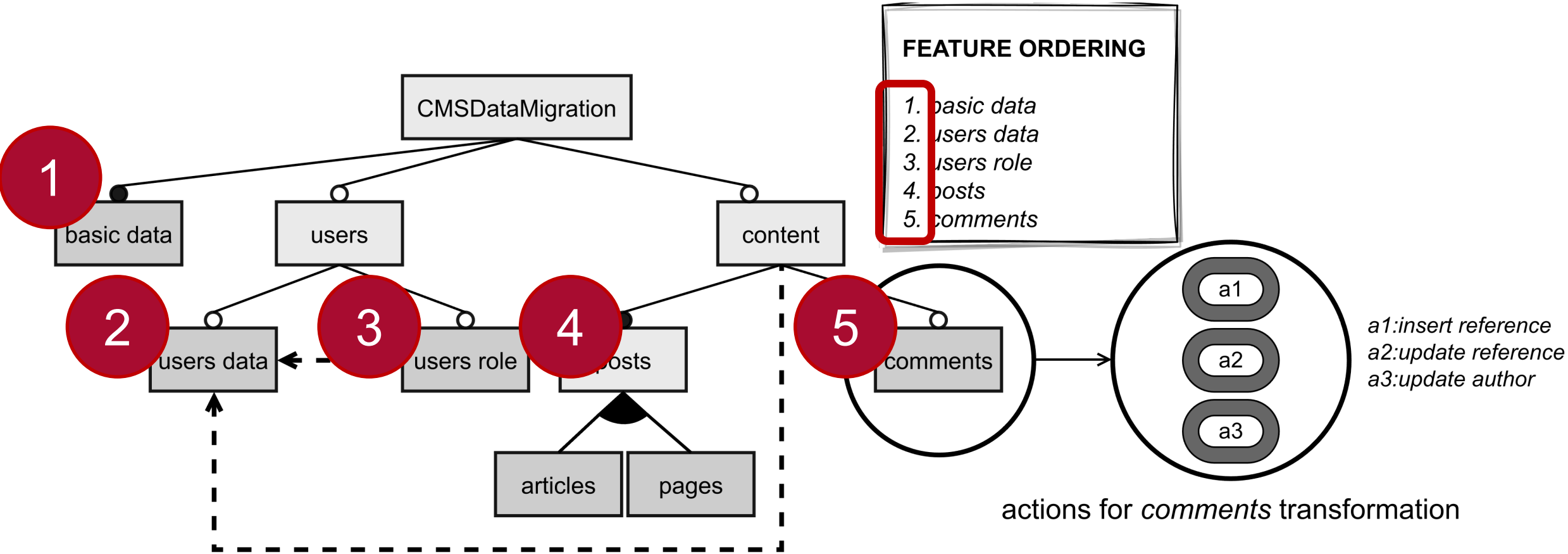
5. Proof of concepts

Validation: Drupal to WordPress



5. Proof of concepts

Validation: Drupal to WordPress



5. Proof of concepts

Validation: Drupal to WordPress

For example…

[basic, user data, user roles, pages, comments]

5. Proof of concepts

Validation: Drupal to WordPress

For example…

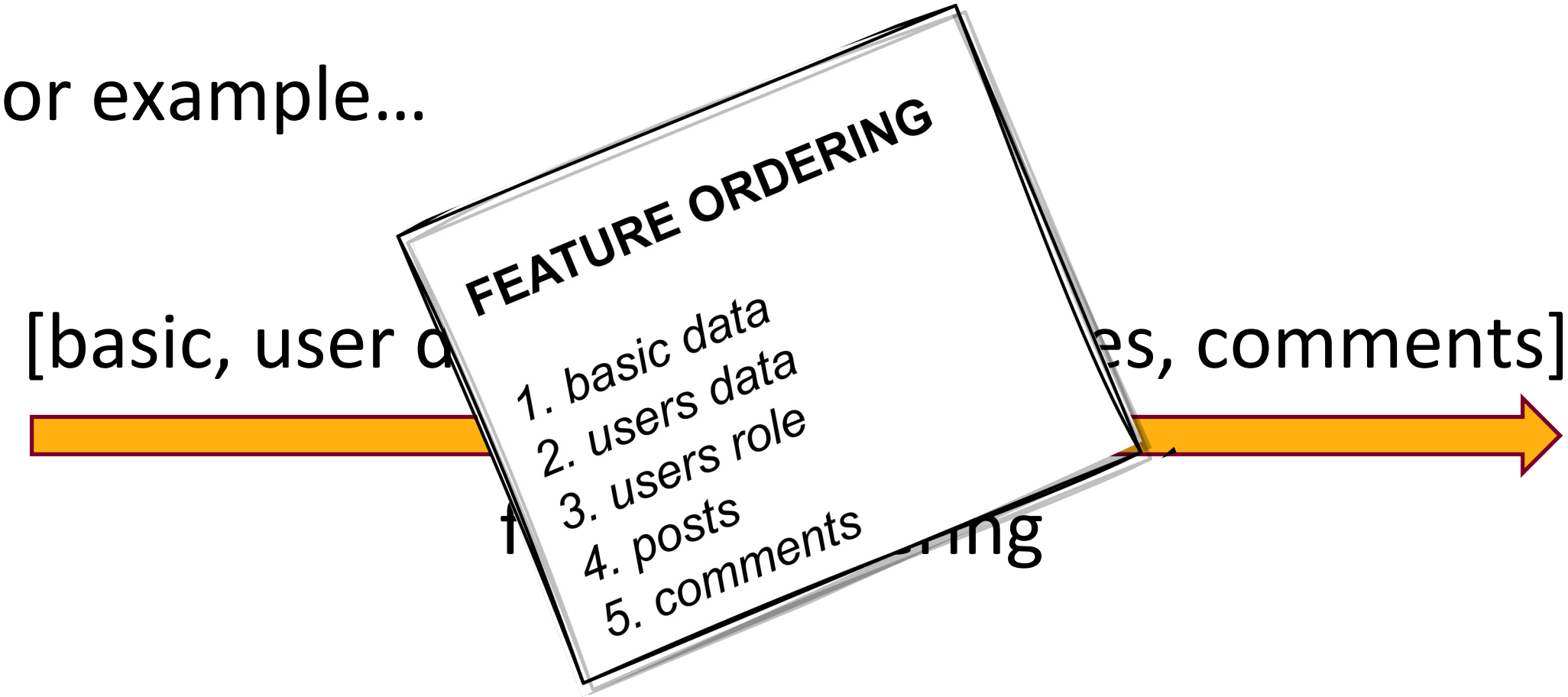
[basic, user data, user roles, pages, comments]

feature ordering

5. Proof of concepts

Validation: Drupal to WordPress

For example…



5. Proof of concepts

Validation: Drupal to WordPress

(1) [basic]
(2) [basic, user data]
(3) [basic, user data, articles]
(4) [basic, user data, articles, pages]
(5) [basic, user data, pages]
(6) [basic, user data, articles, comments]
(7) [basic, user data, articles, pages, comments]
(8) [basic, user data, pages, comments]
(9) [basic, user data, user roles, pages, comments]
(10) [basic, user data, user roles, pages]
(11) [basic, user data, user roles, articles, pages]
(12) [basic, user data, user roles, articles, pages, comments]
(13) [basic, user data, user roles, articles, comments]
(14) [basic, user data, user roles]
(15) [basic, user data, user roles, articles]

15 **products** =
15 **transformation scripts**

5. Proof of concepts

Validation: Drupal to WordPress

```
-- -----------------------------------------------
-- Transformation  InsertReferenceAction
-- -----------------------------------------------

INSERT INTO `wordpress`.`wp_users` (`ID`)
  SELECT MIN(`uid`) FROM `drupal`.`users_field_data`
  GROUP BY `uid`
  ORDER BY `uid`;

-- -----------------------------------------------
---
-- Transformation  UpdateFromFieldAction
-- -----------------------------------------------
---

UPDATE `wordpress`.`wp_users` table_target
       INNER JOIN `drupal`.`users_field_data`
table_source
       ON table_source.`uid` = table_target.`ID`
SET table_target.`user_nicename` =
table_source.`name`
WHERE table_source.`uid` = table_target.`ID`;
```

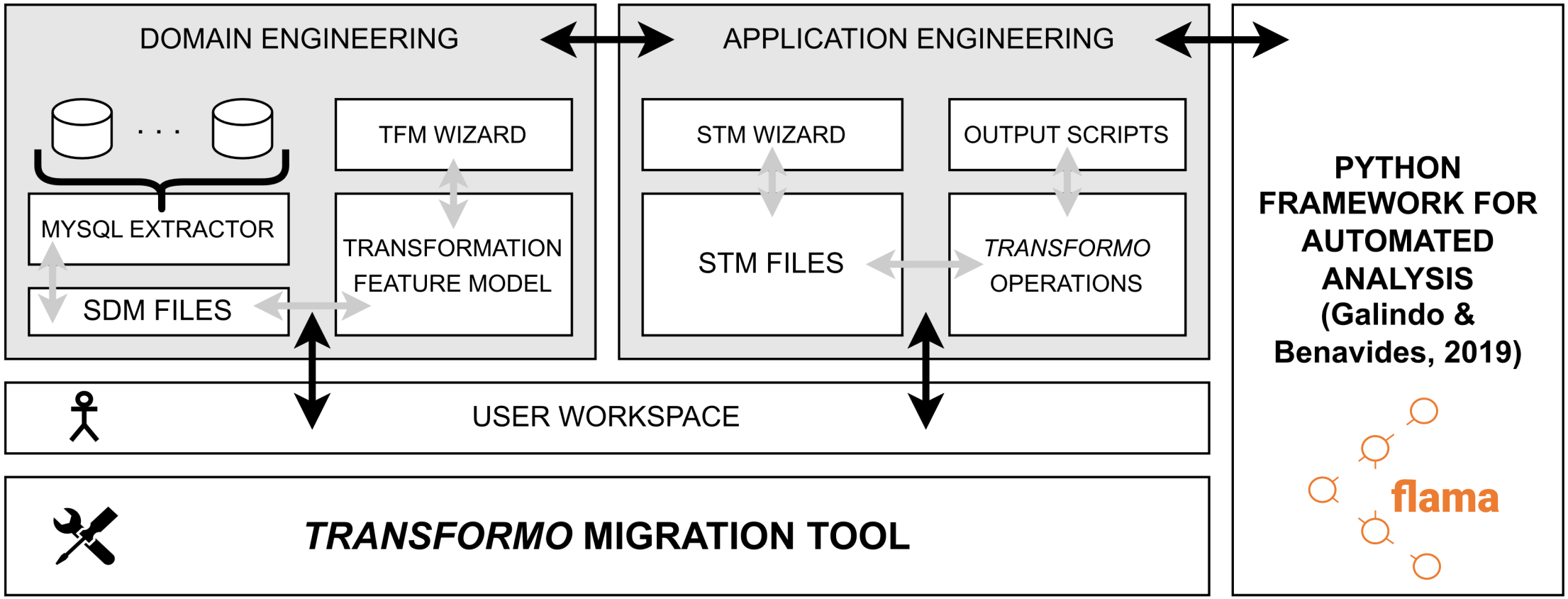
```
-- -----------------------------------------------
-- Transformation  UpdateFromFieldAction
-- -----------------------------------------------

UPDATE `wordpress`.`wp_users` table_target
       INNER JOIN `drupal`.`users_field_data`
table_source
       ON table_source.`uid` = table_target.`ID`
SET table_target.`user_login` = table_source.`name`
WHERE table_source.`uid` = table_target.`ID`;

-- -----------------------------------------------
-- Transformation  UpdateFromFieldAction
-- -----------------------------------------------

UPDATE `wordpress`.`wp_users` table_target
       INNER JOIN `drupal`.`users_field_data`
table_source
       ON table_source.`uid` = table_target.`ID`
    SET table_target.`display_name` =
table_source.`name`
WHERE table_source.`uid` = table_target.`ID`;
```

Roadmap

1. Introduction
2. Data migration background
3. Our solution: an SPL for data transformation
4. Transformation product line for relational databases
5. Proof of concepts
6. **Implementation**
7. Conclusions and future work

6. Implementation

Transformo migration tool



Roadmap

1. Introduction
2. Data migration background
3. Our solution: an SPL for data transformation
4. Transformation product line for relational databases
5. Proof of concepts
6. Implementation
7. **Conclusions and future work**

7. Conclusions and future work

Using a transformation feature model can eliminate the redundancy of code in transformation scripts

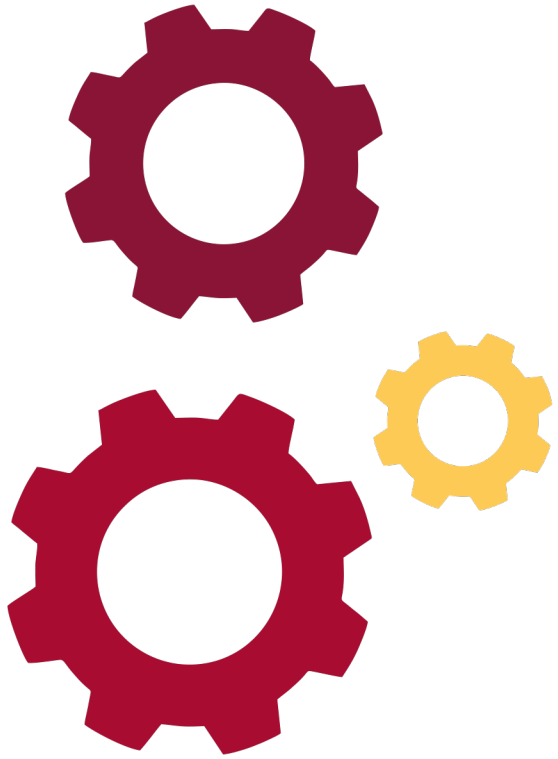
7. Conclusions and future work

Our approach addresses the challenge of migrating data between different databases by providing a flexible and customize solution.

7. Conclusions and future work

We plan to extend our approach to analyze the variability in the E and L stages of the data migration process

7. Conclusions and future work



We plan to provide inference engines to automatically determine the TFM (or semi-automatic way)

7. Conclusions and future work

We plan to extend our proposal to NoSQL databases (e.g. MongoDB)

vamos'24 | u^b

Feb 7-9, Bern, Switzerland

Variability in data transformation: towards data migration product lines

David Romero-Organvidez, David Benavides, Jose-Miguel Horcas, María Teresa Gómez-López