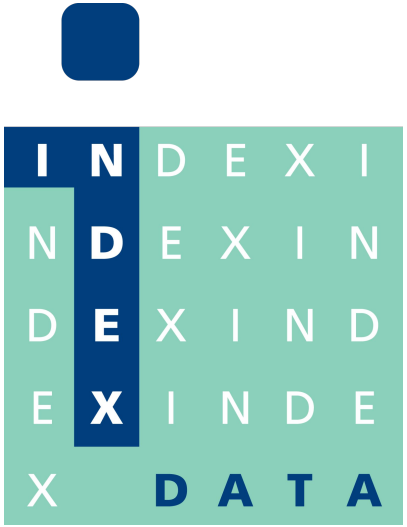
Setting up and using a Metadb reporting database to provide your institution with real-time data reports from FOLIO

Mike Taylor
Software Guy
Index Data
mike@indexdata.com

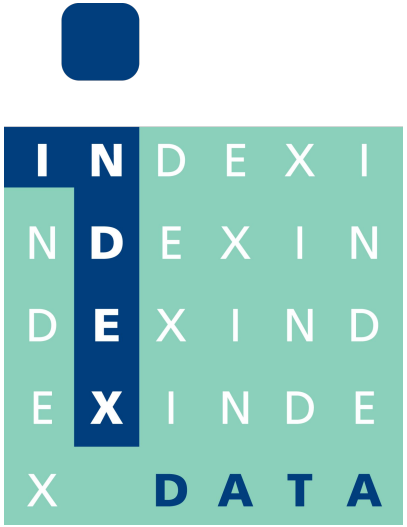


Setting up and using a Metadb reporting database to provide your institution with real-time data reports from FOLIO

Mike Taylor
Software Guy
Index Data
mike@indexdata.com

A gentle introduction to Metadb

Mike Taylor
Software Guy
Index Data
mike@indexdata.com



Introduction

If your library runs FOLIO, you need reports.

Introduction

Why not just run SQL queries against the FOLIO database?

Introduction

Why not just run SQL queries against the FOLIO database?

- No record is kept of old data.

Introduction

Why not just run SQL queries against the FOLIO database?

- No record is kept of old data.
- All the data is in JSON so queries would have to transform it.

Introduction

Why not just run SQL queries against the FOLIO database?

- No record is kept of old data.
- All the data is in JSON so queries would have to transform it.
- The records are highly denormalized with data fragments split across tables.

Introduction

Why not just run SQL queries against the FOLIO database?

- No record is kept of old data.
- All the data is in JSON so queries would have to transform it.
- The records are highly denormalized with data fragments split across tables.
- Running expensive reports has performance implications for regular users.

Introduction

What's needed is

- a separately provisioned database.

Introduction

What's needed is

- a separately provisioned database,
- containing old data as well as the present data.

Introduction

What's needed is

- a separately provisioned database,
- containing old data as well as the present data,
- with a structure optimised for report-type queries.

Introduction

What's needed is

- a separately provisioned database,
- containing old data as well as the present data,
- with a structure optimised for report-type queries,
- that stays up to date as the FOLIO database changes.

Introduction

What's needed is

- a separately provisioned database,
- containing old data as well as the present data,
- with a structure optimised for report-type queries,
- that stays up to date as the FOLIO database changes.

That's Metadb!

How Metadb works

FOLIO stores its data in a Postgres database.

How Metadb works

FOLIO stores its data in a Postgres database.

When changes are made in this database by the day-to-day operation of the FOLIO ILS, change records are written to a Postgres *replication slot*.

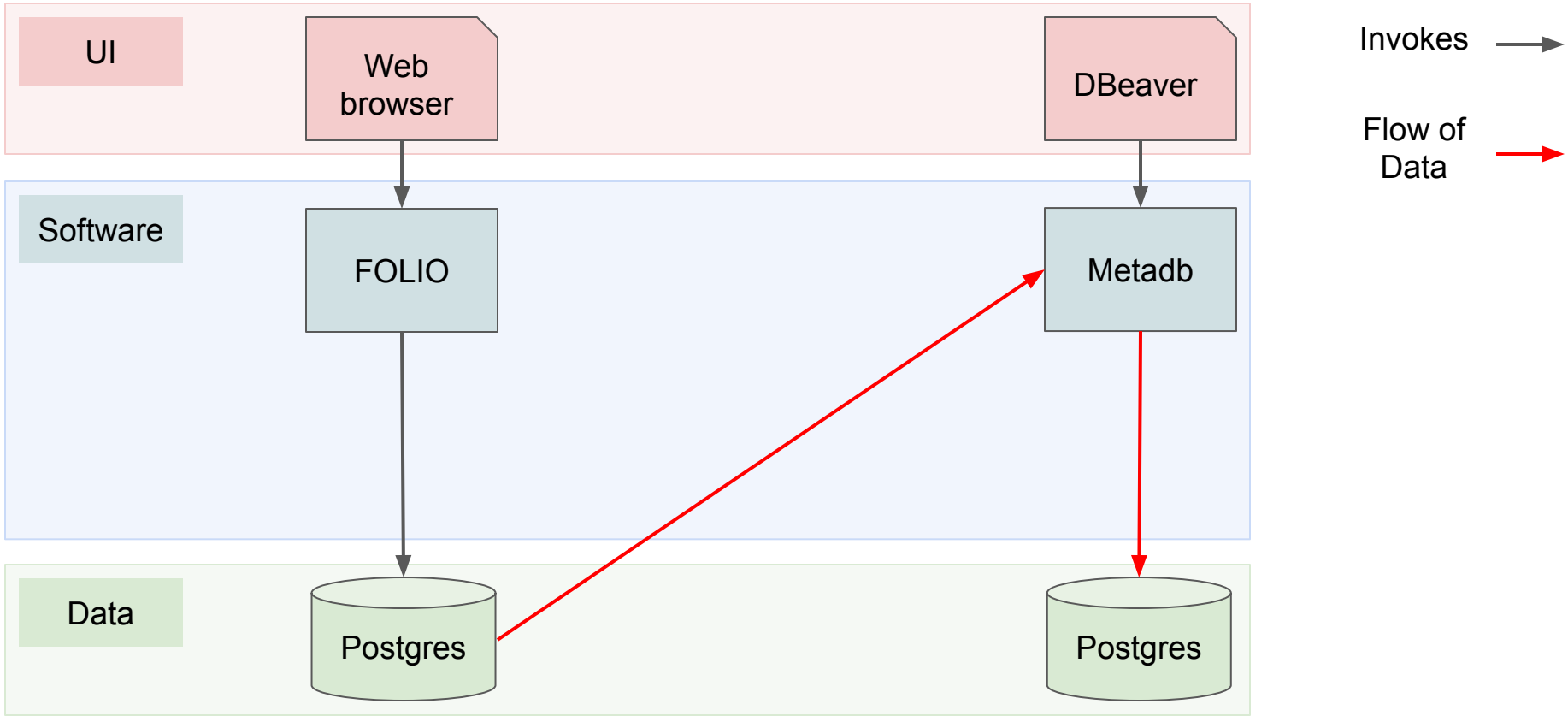
How Metadb works

FOLIO stores its data in a Postgres database.

When changes are made in this database by the day-to-day operation of the FOLIO ILS, change records are written to a Postgres *replication slot*.

Metadb harvests these change records to keep its own database up to date.

How Metadb works



How Metadb works

(Actually, it's more complicated than that — stay tuned.)

How Metadb works

Old versions of changed records are retained and accessible via historical tables that record all previous versions of each record with the start and end of their validity period.

How Metadb works

Old versions of changed records are retained and accessible via historical tables that record all previous versions of each record with the start and end of their validity period.

Each record coming into Metadb has its JSON blob extracted and transformed into SQL-level fields.

How Metadb works

Old versions of changed records are retained and accessible via historical tables that record all previous versions of each record with the start and end of their validity period.

Each record coming into Metadb has its JSON blob extracted and transformed into SQL-level fields.

MARC records, which are stored as opaque blobs in FOLIO, are expanded into queryable tables.

How Metadb works

Periodic queries, run automatically, generate derived tables that contain denormalized versions of FOLIO records which include data from related records.

How Metadb works

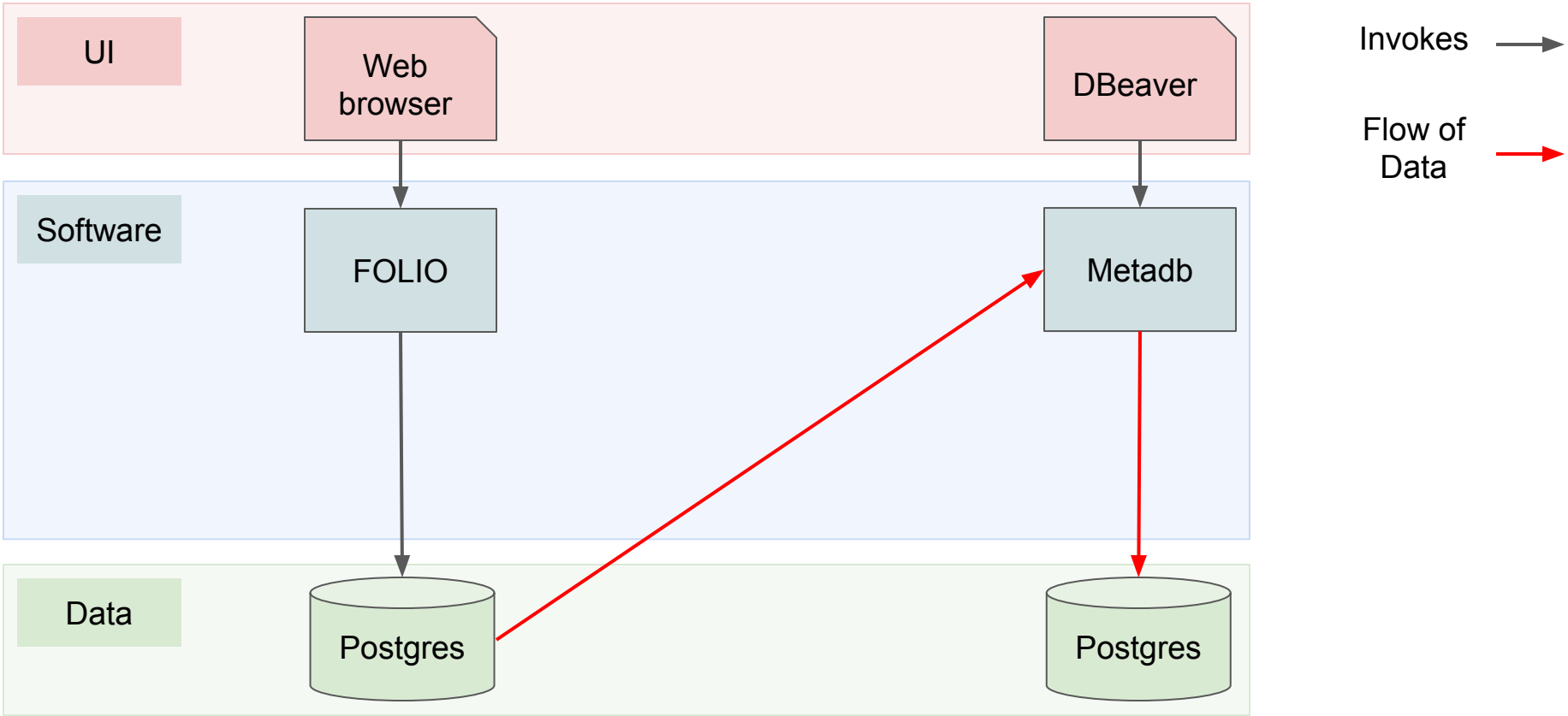
Periodic queries, run automatically, generate derived tables that contain denormalized versions of FOLIO records which include data from related records.

For example, user records in Metadb's **users_groups** table contain not just a **patron_group** ID column but also a **group** column containing the name of the group that the user belongs to.

How data gets from FOLIO into Metadb

It would be possible for Metadb to read directly from the replication slot of the FOLIO system's Postgres database.

How data gets from FOLIO into Metadb



How data gets from FOLIO into Metadb

It would be possible for Metadb to read directly from the replication slot of the FOLIO system's Postgres database.

That would not be ideal:

- Change events could back up in the Postgres replication slot

How data gets from FOLIO into Metadb

It would be possible for Metadb to read directly from the replication slot of the FOLIO system's Postgres database.

That would not be ideal:

- Change events could back up in the Postgres replication slot
- State could get confused if FOLIO or Metadb goes down

How data gets from FOLIO into Metadb

It would be possible for Metadb to read directly from the replication slot of the FOLIO system's Postgres database.

That would not be ideal:

- Change events could back up in the Postgres replication slot
- State could get confused if FOLIO or Metadb goes down
- We would be limited to harvesting from Postgres, and Metadb is more general than that — e.g. could harvest from a VuFind database.

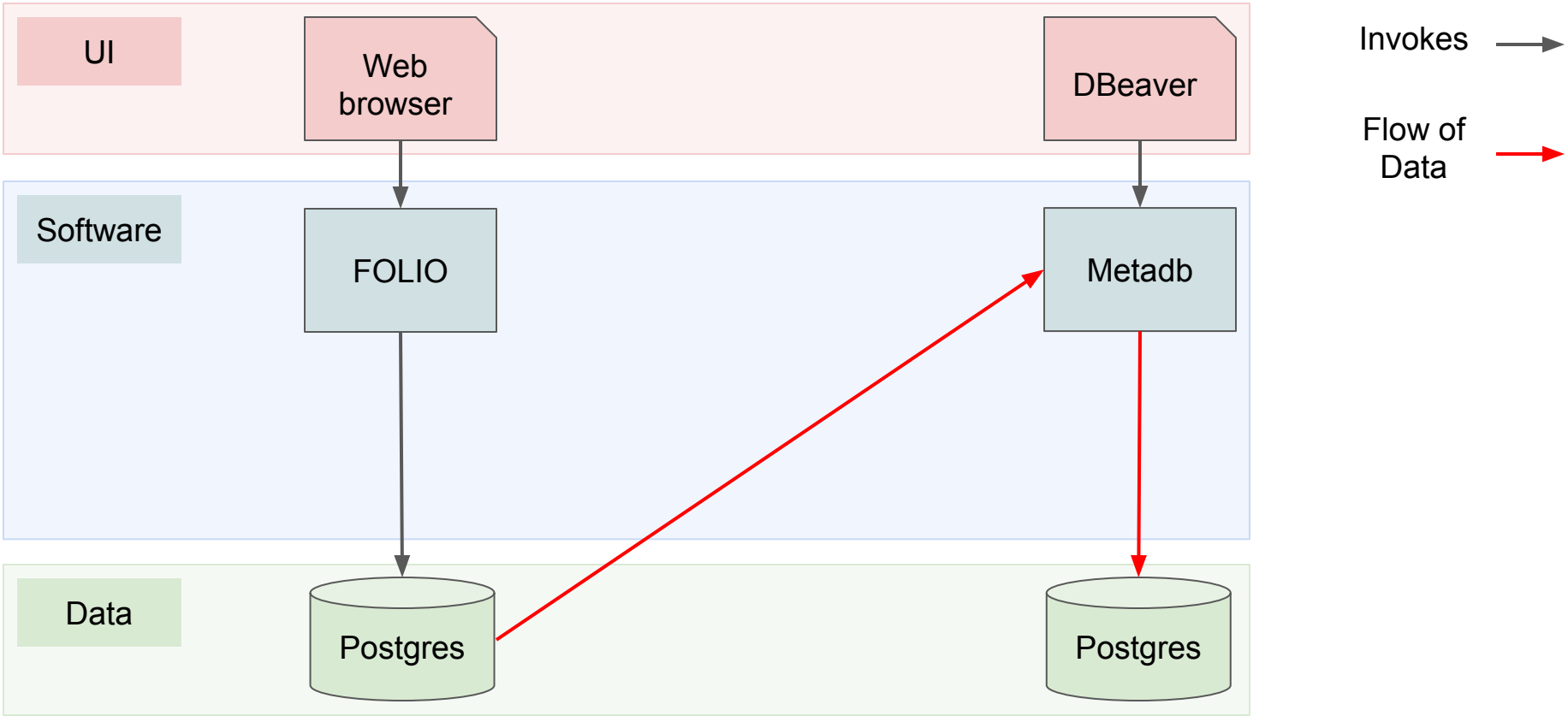
How data gets from FOLIO into Metadb

Instead we use a more flexible but complex system:

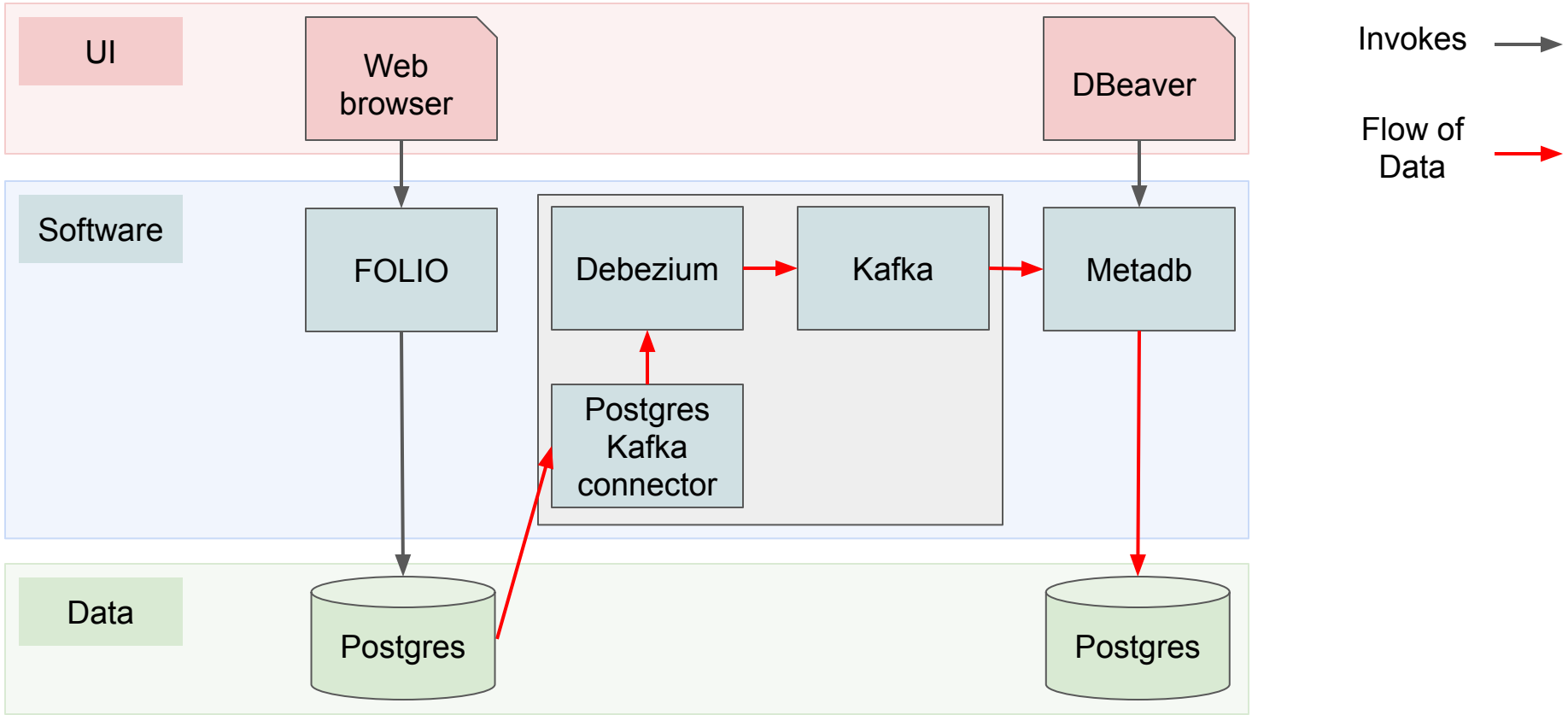
- Debezium uses the Postgres connector from Kafka Connect to read from the FOLIO database's replication slot
- It feeds those events into Kafka
- Metadb reads events from Kafka

This provides buffering, robustness and database-independence.

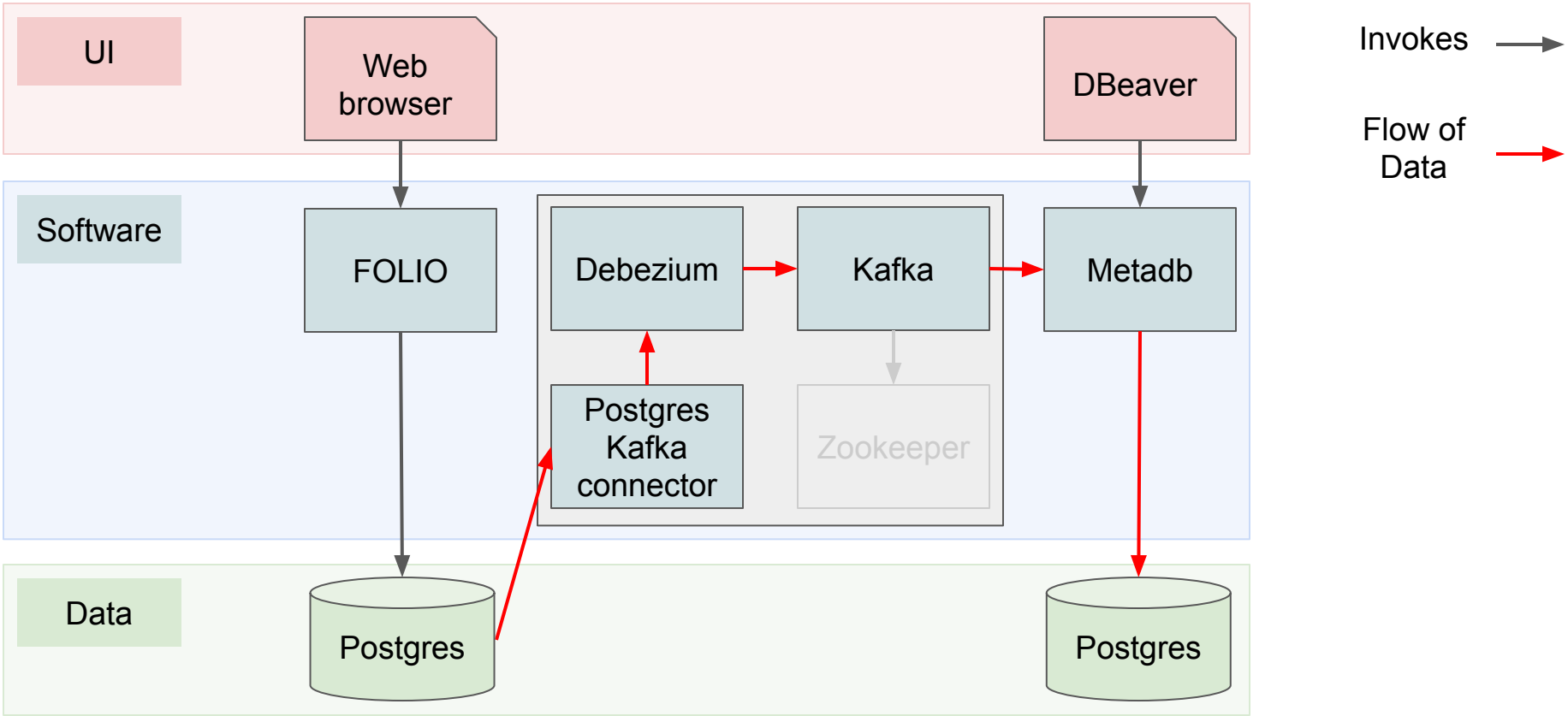
How data gets from FOLIO into Metadb



How data gets from FOLIO into Metadb



How data gets from FOLIO into Metadb



How Metadb data can be interrogated

The Metadb database can be queried using SQL:

- On the command-line with **psql**

How Metadb data can be interrogated

The Metadb database can be queried using SQL:

- On the command-line with **psql**
- Using a GUI like DBeaver or Tableau (Suzette will tell us more about this)

How Metadb data can be interrogated

The Metadb database can be queried using SQL:

- On the command-line with **psql**
- Using a GUI like DBeaver or Tableau (Suzette will tell us more about this)
- Directly within FOLIO (Scott will tell us more about this)

How Metadb data can be interrogated

The Metadb database can be queried using SQL:

- On the command-line with **psql**
- Using a GUI like DBeaver or Tableau (Suzette will tell us more about this)
- Directly within FOLIO (Scott will tell us more about this)
- Programatically by client software

How Metadb data can be interrogated

The Metadb database can be queried using SQL:

- On the command-line with **psql**
- Using a GUI like DBeaver or Tableau (Suzette will tell us more about this)
- Directly within FOLIO (Scott will tell us more about this)
- Programatically by client software

And what kinds of queries might you run? Sharon will tell us more about this!

Setting up and using a Metadb reporting database to provide your institution with real-time data reports from FOLIO

Mike Taylor
Software Guy
Index Data
mike@indexdata.com