Model automation for policy and supervision

A cloud-based approach

DeNederlandscheBank
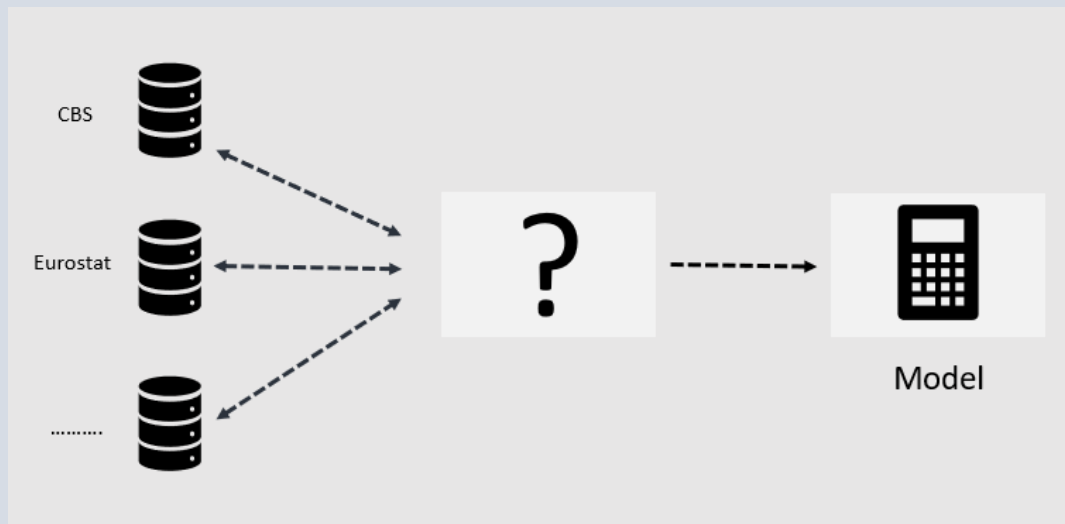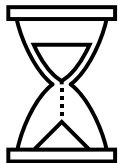
Project setting

Project setting
The scenario

Model process

DNB creates and maintains many models for prediction, stress testing and valuation using data from various sources.

However, the data are often collected and stored **manually** for each model. This results in a process that is error prone, hard to reproduce, and means that modelers must wait for someone to refresh the data every time a model has to be updated.

By **automating** the data collection in a modern way, the process becomes significantly more efficient while reducing risks.

CBS

Eurostat

....

?

Model

DeNederlandscheBank
EUROSYSTEEM

DataScience Hub

Efficiency
Time before a modeler can get started
Manual

modeler — Request data — expert — Manual download — Combine sheets — Send data back — Use in model
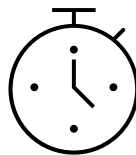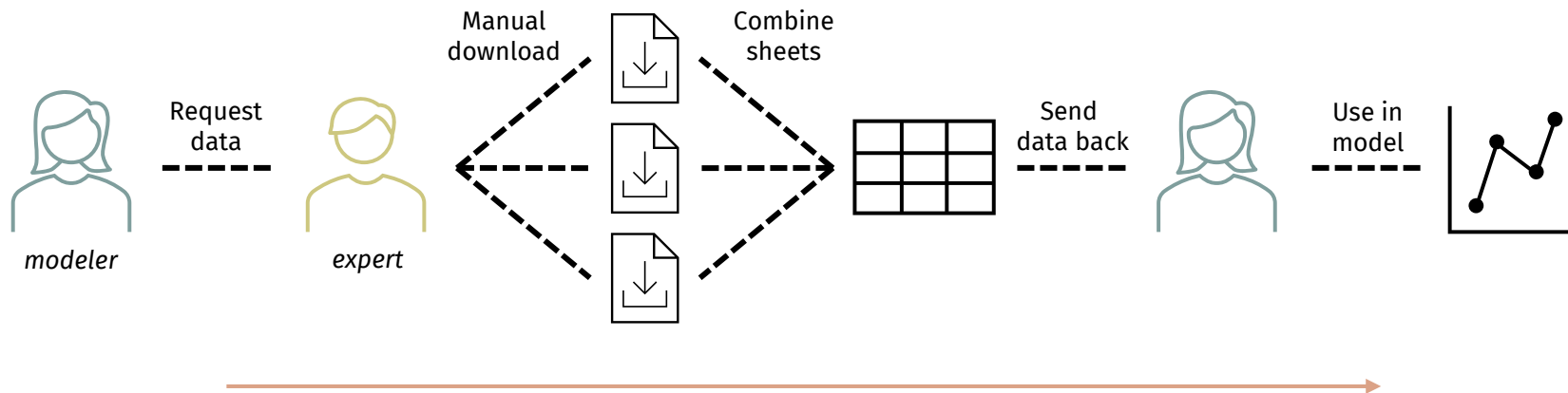
Up to multiple days
Depending on size of data
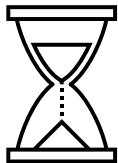
Efficiency
Time before a modeler can get started
Automated

modeler → Press refresh button ----- ↺ ----- Data is fetched ----- Stored in database ----- Use in model
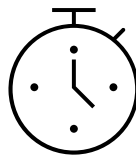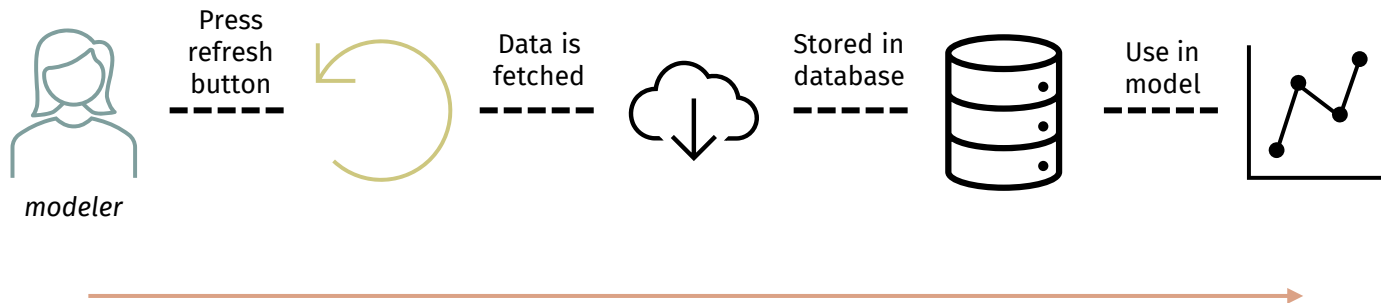
~ 10 minutes
Until new data appears in database
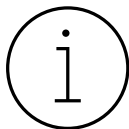
The problem

Internal model outcomes are difficult to reproduce, audit, and can become unreliable.

Manual operations cannot be avoided completely. However, when incorporated, two aspects are key:

- Clear procedure & documentation
- Discipline

Without these, **reproducibility** is at stake.

In many cases, manual operations are not (or incompletely) logged. Therefore, **audits** are difficult.

Additionally, changes in model data, configuration files, and other important model components **are not tracked.**

These risks lead to a less **reliable** process, that can potentially decrease the trust in the model outcomes.

Eventually, DNB's reputation could be affected.

Reliability & reproducibility
Why it is important

What can go wrong?

Model outcomes could be based on incorrect data.

This might stay undetected but errors could also become known to the public.

Obviously, this could impact DNB's reputation.

NOS Nieuwe rekenfout RIVM in lijst met grote uitstoters, minister 'baalt enorm'

Analysis **The Guardian**

Covid: how Excel may have caused loss of 16,000 test results in England

● ● ●

Reliability and reproducibility are important!

DeNederlandscheBank
EUROSYSTEEM

DataScience Hub

7

Model automation

Goal & Scope

The goal of the NIPE automation project is to automate the NIPE model process in a modern way.

Scope of the project:

- Automating the data collection part of the model, such that model developers can easily update model data, as well as creating an easy way to communicate with the data on Azure: users can connect their model directly to the database to always have access to the most recent data
- It is clear who has access to the data, and who has used it
- New data sources can be easily added to the database

Components of the solution:

- The dnbdatafetcher package: a python package that can automatically fetch all required model data from sources like CBS and Eurostat
- Resources in the Azure Cloud
- Auditability and Reproducibility ensured using Azure DevOps Version Control
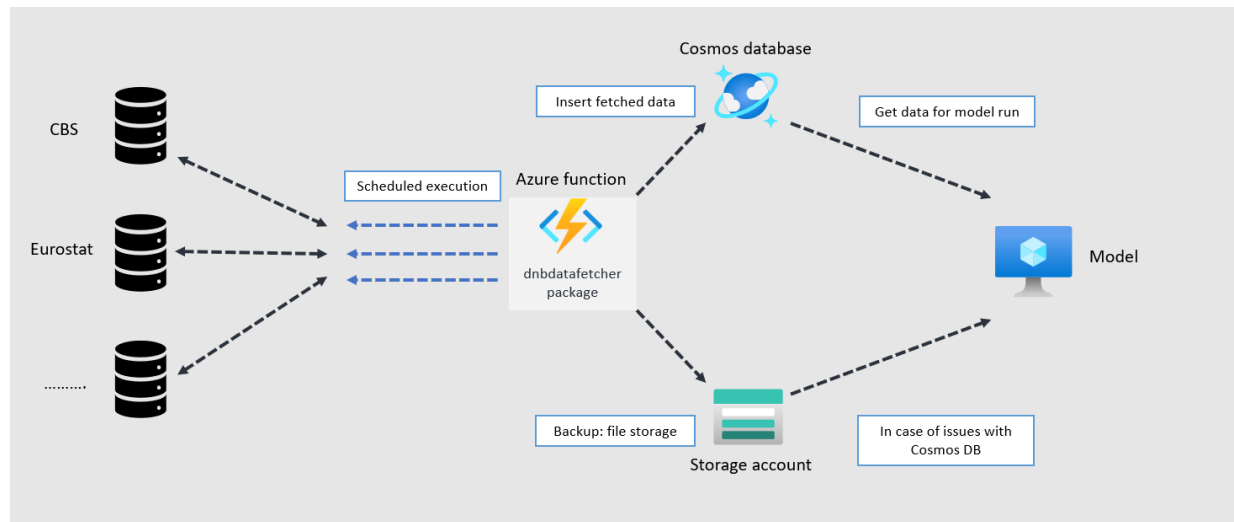
Application

The diagram to the right illustrates how the application works

The function fetches data on a predefined schedule and stores it in a database and in a storage account
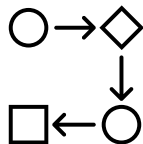
Always up-to-date

Because data is fetched on a schedule, modelers always have access to up-to-date data for their model(s)

Getting data is easy and can be done via e.g. DSW. For extra certainty, data are also stored as a file in a storage account

CBS

Eurostat

....

Scheduled execution

Azure function

dnbdatafetcher package

Cosmos database

Insert fetched data

Get data for model run

Model

Backup: file storage

Storage account

In case of issues with Cosmos DB

Workflow

Workflow
Ensure reproducibility

Code reviews & tests

Ensuring that the code does what was intended is important to ensure the integrity of the application.

Our workflow encompasses this important aspect by:

- *Requiring code reviews for each line of code that is changed;*
- *Running automated tests for each change to ensure the behavior is unchanged;*
- *Requiring acceptance tests from the business before a new deployment to production is released.*

Correctness
&
Review
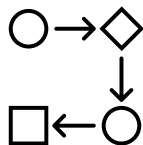
Workflow
Ensure reproducibility

Reproducibility & Auditability

Version history & logging

The whole application is under version control. Meaning that every change is tracked. This ensures that everything is reproducible. Actions performed in the application are logged such that activities and events can be audited.

On the modeling side, this has benefits as well:

- *At any point in time, it is clear which data was fetched;*
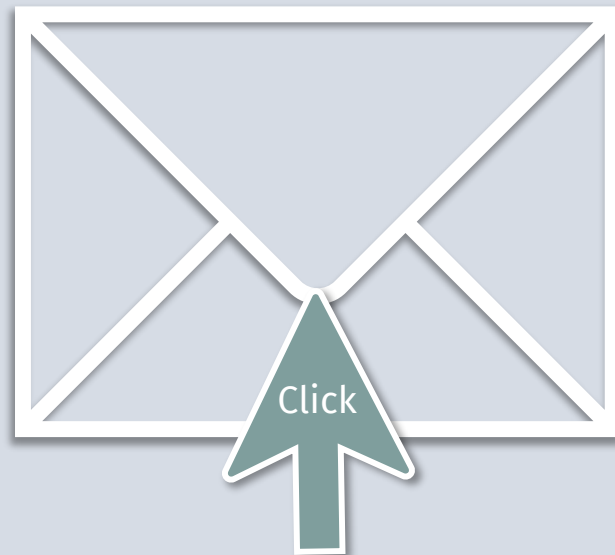- *The database has versioning, meaning that each model run can be related to a unique dataset in the database.*

Questions?
Do not hesitate to contact us.

More info

At the Data Science Hub, we are excited to improve the way we all work and are happy to share knowledge whenever possible.

If you want to know more or have any questions, feel free to reach out to us!

Click

Send us a mail at
data_science@dnb.nl

DeNederlandscheBank
EUROSYSTEEM

DataScience
Hub