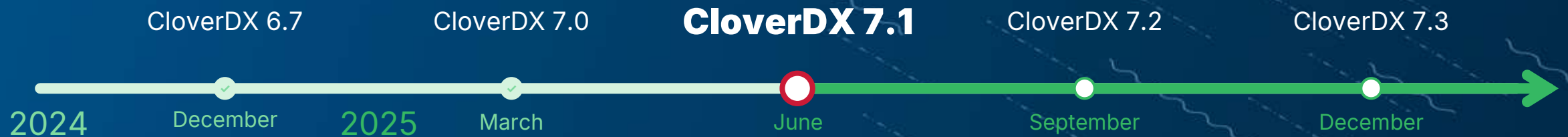


# CloverDX 7.1

Branislav Repcek, VP, Product  
8<sup>th</sup> July 2025

Four major releases every year

# CloverDX releases



## Highlights

# CloverDX 7.1

- **AI features in CloverDX 7.1**
  - ✓ OpenAIClient
  - ✓ New ML data classification and anonymization components
  - ✓ Machine learning model deployments and CloverDX Marketplace
- **Centralized Data Manager**
- **Reference data sets in Data Catalog and Wrangler**
- **Descriptions for Event Listeners and Schedulers**
- **Other changes**



# AI in data integration

- **How do we help different kinds of users?**
  - ✓ Technical as well as non-technical data experts
  - ✓ Data engineers, data stewards, business users
- **How do we help users in different stages of their project lifecycle?**
  - ✓ Increase productivity during design-time
  - ✓ Use AI or machine learning to transform your data





# AI in data integration



## Usage

**Transformation**  
*Cost per record*


**Productivity**  
*Capabilities*

 Data classification and anonymization Using local models	 Clover Assistant for suggestions in Data Manager
 OpenAIClient transformer	 Clover Assistant for self-service Wrangler

# AI in CloverDX 7.1: incubation components

- **Local execution: machine learning data classification**

- ✓ Machine learning models built by 3<sup>rd</sup> parties and wrapped in CloverDX libraries
  - Get models from HuggingFace, GitHub, etc. or use your own models
  - PyTorch model format
- ✓ Strong data security and privacy – data never leaves your CloverDX Server
- ✓ Requires powerful hardware to run with good performance

	Transformation	Productivity
Local		
Cloud LLM		

# AI in CloverDX 7.1: incubation components

- **Local execution: machine learning data classification**

- ✓ Machine learning models built by 3<sup>rd</sup> parties and wrapped in CloverDX libraries
  - Get models from HuggingFace, GitHub, etc. or use your own models
  - PyTorch model format
- ✓ Strong data security and privacy – data never leaves your CloverDX Server
- ✓ Requires powerful hardware to run with good performance

- **Using LLMs in the cloud: work with LLMs provided by OpenAI**

- ✓ All LLM processing in the cloud – data must be sent to OpenAI
- ✓ Use any LLM available in the API to control cost and performance

	Transformation	Productivity
Local	✓	
Cloud LLM		

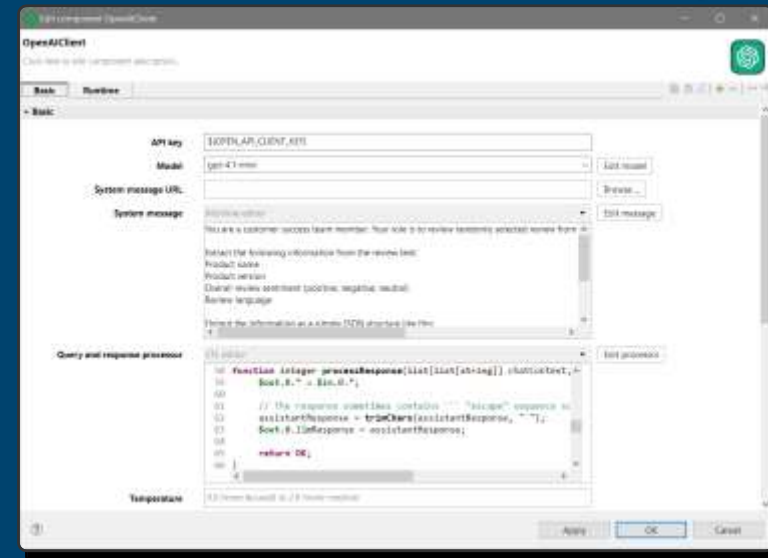
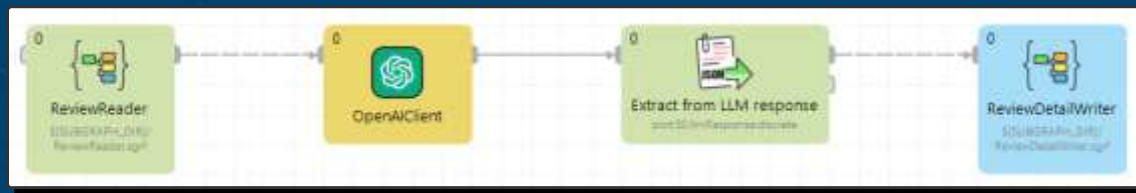
## Incubation

# Using OpenAI LLMs



- **OpenAIClient component**

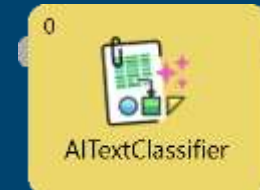
- ✓ Submit data to OpenAI LLM and process the result.
- ✓ Requires OpenAI API key, pricing depends on model selected.
- ✓ Customizable response processing



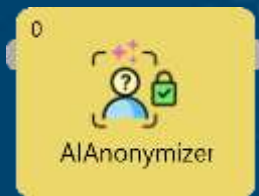
Incubation

# Data classification components

- **Run data classification models locally in your CloverDX**
  - ✓ Token classification, text classification
  - ✓ Components pass data to ML model to classify and return full classification response as variant
- **Models shared via CloverDX Server**
  - ✓ Also available in local Designer-only projects if needed

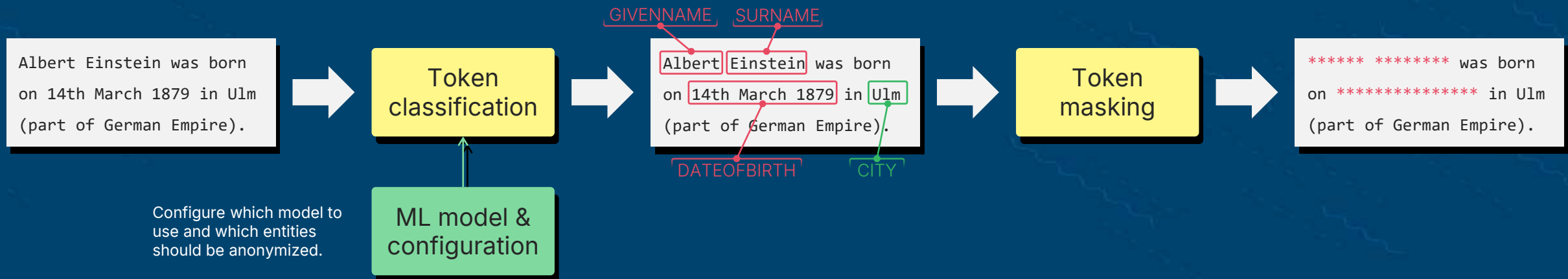


# AI-based data anonymization



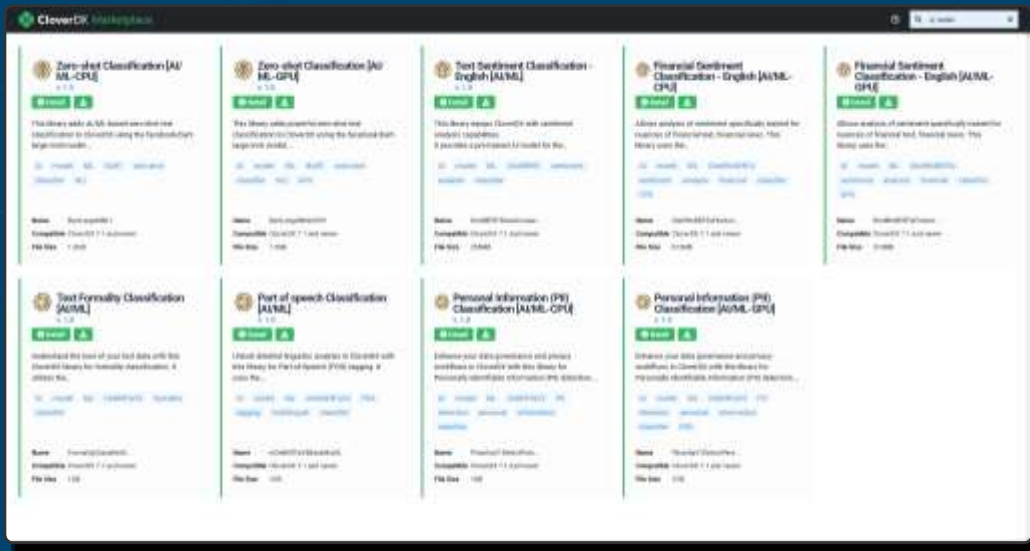
## ● AIAnonymizer component

- ✓ Token classifier + automated transformer to mask the data
- ✓ Identify classes with ML model and then apply data masking to data classes that are selected in the component configuration
- ✓ Use the same models as for AITokenClassifier

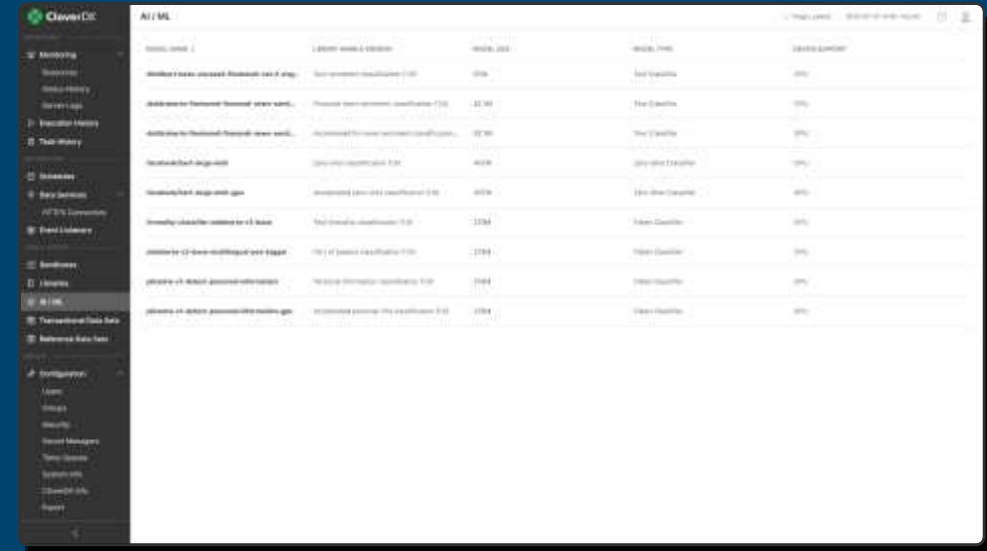


# Managing local machine learning models

- **Models deployed on Server via CloverDX Marketplace**
- **Models are converted 3<sup>rd</sup> party models from HuggingFace**
- **Models shared with all users on the Server**
  - ✓ Each job / component can use its own settings for model execution



Deploy

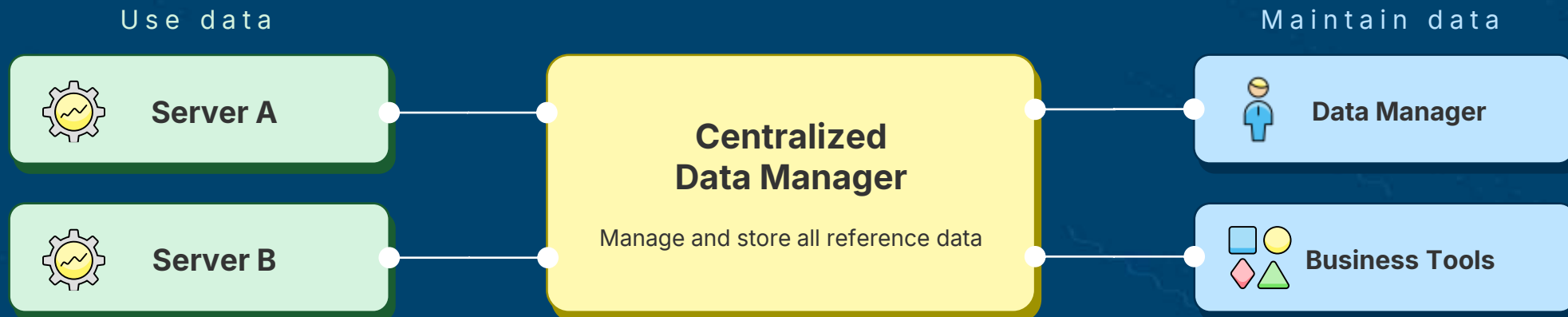


# Hardware-acceleration for CloverDX deployments

- **Use NVIDIA GPU to get maximum performance when running models locally**
  - ✓ CPU is the default, always works, but slow
- **Use our new Docker image on DockerHub**
  - ✓ <https://hub.docker.com/r/cloverdx/cloverdx-server-ai>
  - ✓ Preconfigured with all software needed to use GPU-acceleration
- **Easily deploy to AWS EC2 G5 and EC2 G4 or similar instances**
  - ✓ Instances already come with pre-installed software requirements that will allow you to deploy the AI Docker image easily

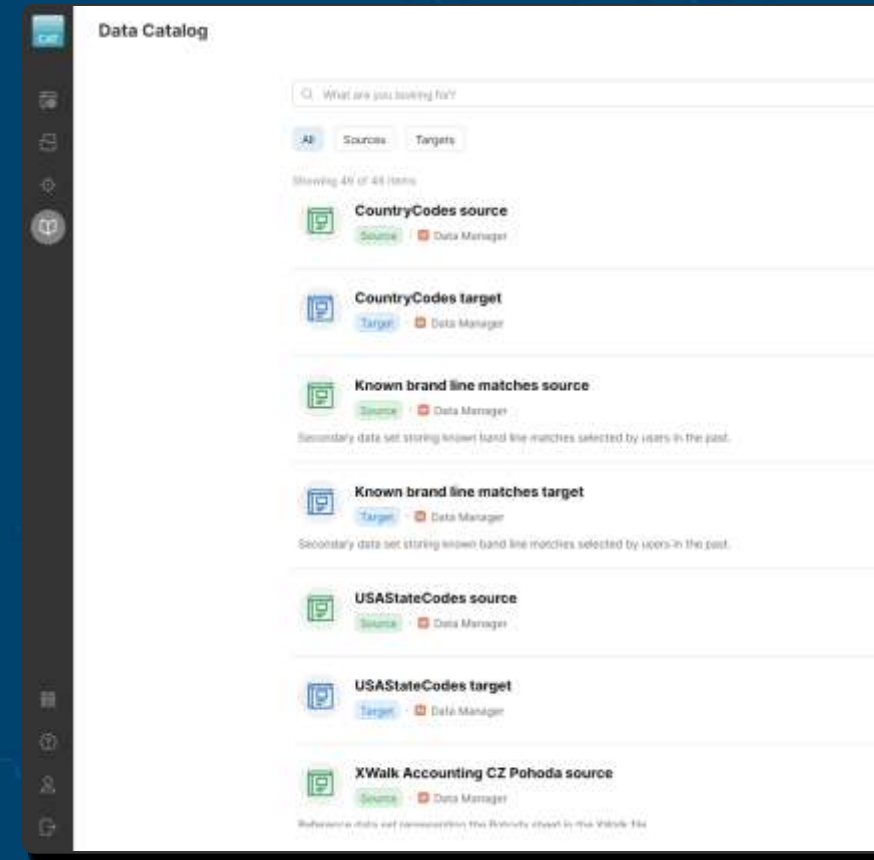
# Centralized Data Manager

- **Use single shared instance of Data Manager for centralized data governance**
- **Manage reference data on one CloverDX instance**
  - ✓ Dedicate the instance to Business Tools users
- **Share across any number of other Server instances**



# Reference data sets in Wrangler

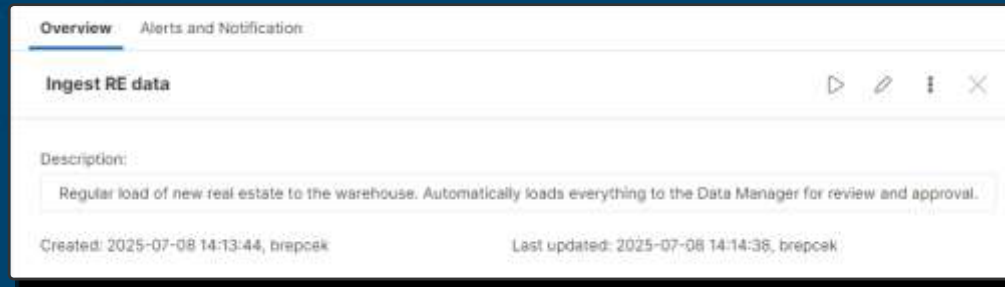
- **Use data sets directly as lookups in Wrangler**
  - ✓ All reference data sets available as sources in the Data Catalog
- **Update reference data in Wrangler jobs**
  - ✓ Use targets for data sets from the Data Catalog
- **Manage permissions on the data set level for maximum granularity**



# Other changes

- **Descriptions in Schedulers and Event Listeners**

- ✓ Additional properties also available: last edit time and user



- **Using gRPC for node to node communication in cluster**

- ✓ May require changes in your cluster configuration

- **Additional configuration for embedded Tomcat**

- ✓ Improve security by allowing you to hide details in error messages

# Demo

# Q & A

## CloverDX 7.1

Release notes: <https://support.cloverdx.com/releases/7.1>

Documentation: <https://doc.cloverdx.com/latest/>