

# Automate Accelerate & Scale IT for business

Cloud, Digital Transformation, IoT



Paris, London, Sophia Antipolis, San Jose, Dakar

- Objective
- Data Streaming Platform (**Kafka-Storm-Visdom**)
  - Architecture
  - Execution in PCA (ProActive Cloud Automation)
- Dataflow Sample
  - Overview
  - Workflow description
  - Workflow execution
  - Result visualization
- Pros of Data Streaming with (**ActiveEon**) ProActive

- We aim at demonstrating how easy and efficient to perform real-time data stream processing using ActiveEon ProActive.
- We use a data streaming platform composed of:
  - **Apache Kafka**: A distributed streaming platform (<https://kafka.apache.org/>)
  - **Apache Storm**: A distributed real-time computation system (<http://storm.apache.org>)
  - **Visdom**: A tool for creating, organizing, and sharing visualizations of live, rich data (<https://github.com/facebookresearch/visdom>)

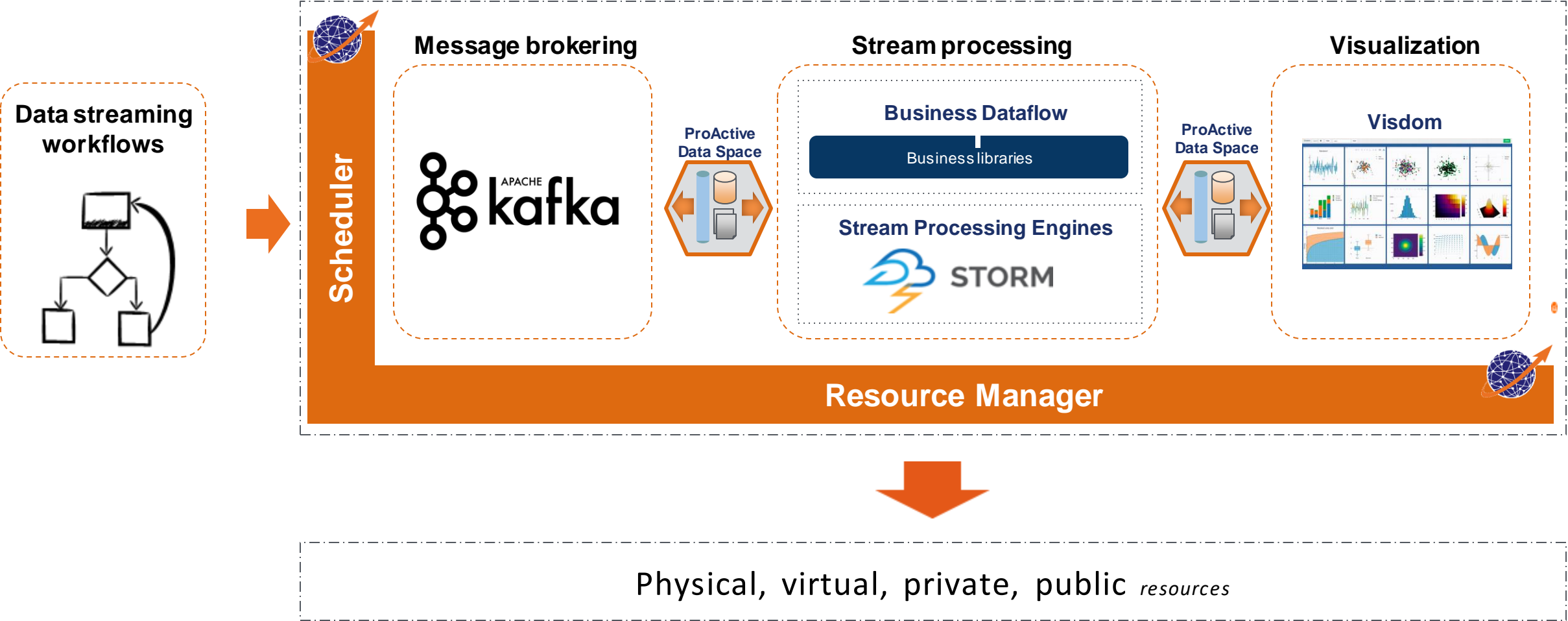


**Activeeon**  
SCALE BEYOND LIMITS

# **Data Streaming Platform**

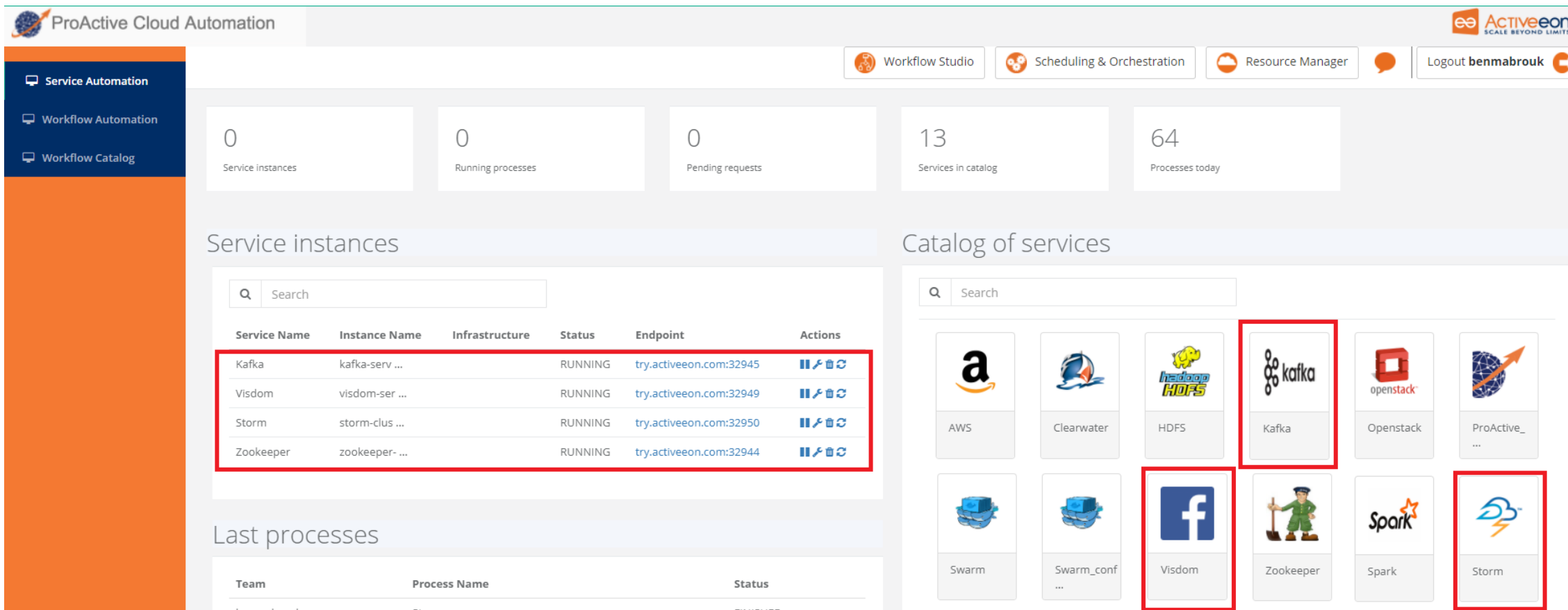
## **Kafka-Storm-Visdom**

# Data Streaming Platform - Architecture



# Data Streaming Platform in PCA

- In ProActive Cloud Automation (PCA), we created workflows for creating and removing instances of **Apache Kafka**, **Apache Storm** and **Visdom**.















The screenshot displays the ProActive Cloud Automation (PCA) dashboard. The top navigation bar includes the ProActive Cloud Automation logo, the Activeeon logo, and a user profile for 'benmabrouk'. The main dashboard features a summary section with five cards: Service instances (0), Running processes (0), Pending requests (0), Services in catalog (13), and Processes today (64). Below this, the 'Service instances' section shows a table of running services, and the 'Catalog of services' section shows a grid of available services. Red boxes highlight the Kafka, Storm, and Visdom services in both sections.

**Service instances**

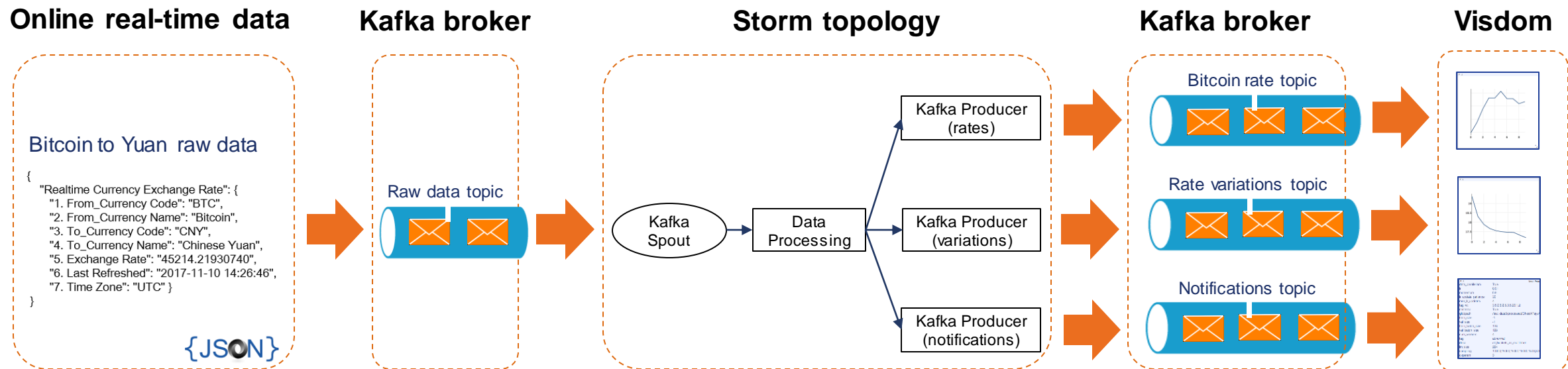
Service Name	Instance Name	Infrastructure	Status	Endpoint	Actions
Kafka	kafka-serv ...		RUNNING	try.activeeon.com:32945	⏏ ⏏ ⏏ ⏏
Visdom	visdom-ser ...		RUNNING	try.activeeon.com:32949	⏏ ⏏ ⏏ ⏏
Storm	storm-clus ...		RUNNING	try.activeeon.com:32950	⏏ ⏏ ⏏ ⏏
Zookeeper	zookeeper- ...		RUNNING	try.activeeon.com:32944	⏏ ⏏ ⏏ ⏏

**Catalog of services**

Service Name	Icon
AWS	
Clearwater	
HDFS	
Kafka	
Openstack	
ProActive_...	
Swarm	
Swarm_conf ...	
Visdom	
Zookeeper	
Spark	
Storm	

# Dataflow Sample

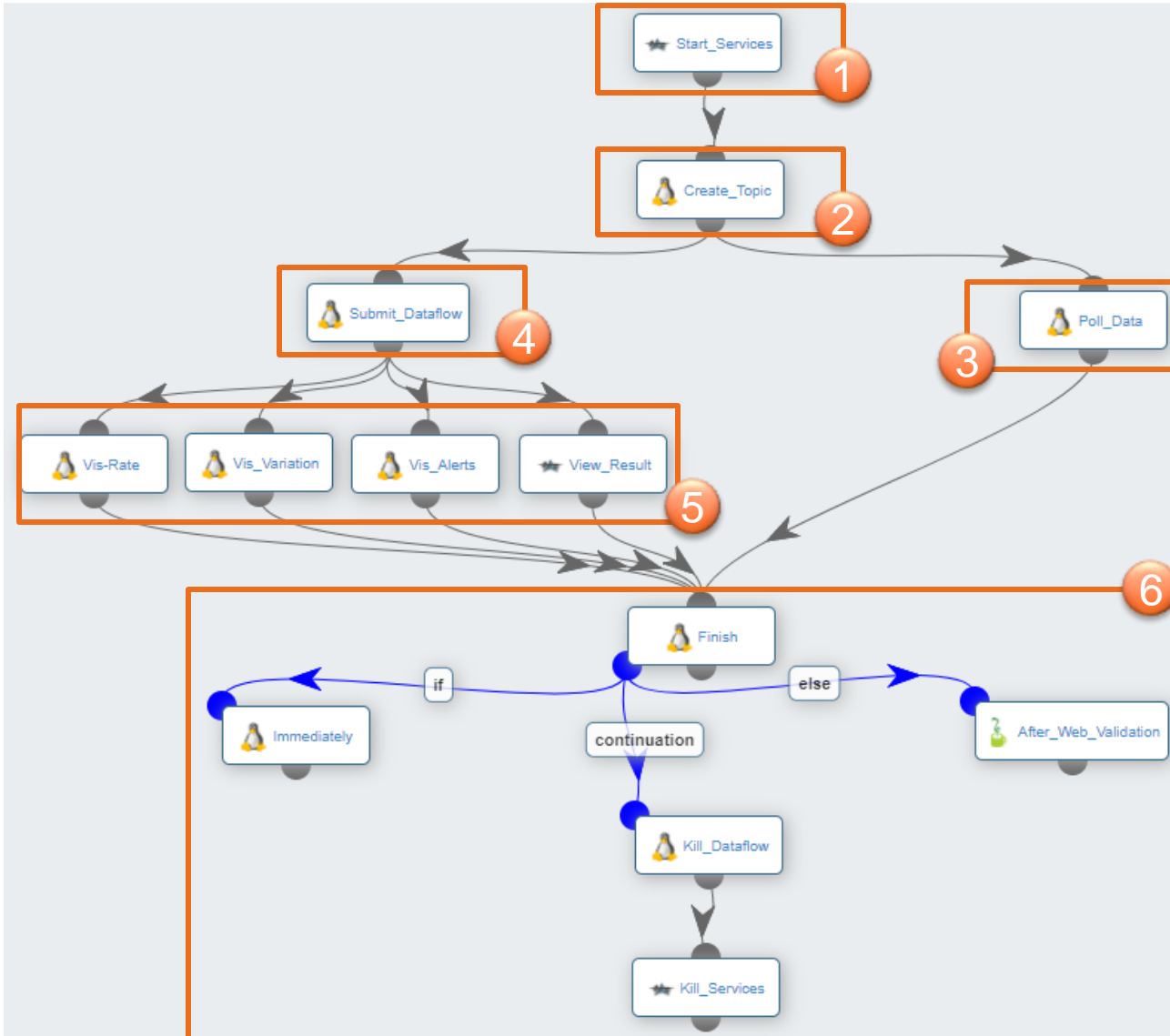
# Dataflow Sample - Bitcoin to Yuan rate processing



- We acquire real-time data about the exchange rate of Bitcoin to the Chinese Yuan.
- This data is available online in a json format. Each json message is stored in a Kafka publish-subscribe topic.
- We further deploy a topology (i.e., a dataflow application) in storm, which:
  - (i) reads json messages from the kafka topic,
  - (ii) parses the json structure,
  - (iii) extracts the exchange rate,
  - (iv) computes the rate variations,
  - (v) sends a notification if the variation is greater than 50 Yuan.
- The extracted rate, variations and notifications are sent to Kafka topics, and further visualized via Visdom.

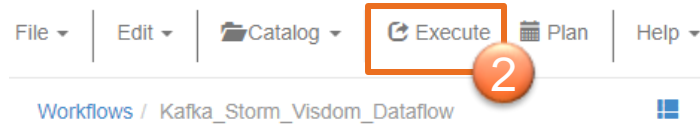


# Dataflow Sample – Workflow overview



- 1 Start PCA services (Kafka, Storm and Visdom)
- 2 Create a topic in Kafka to receive Bitcoin raw data
- 3 Acquire Bitcoin raw data available online and send it to the created topic in Kafka
- 4 Submit the dataflow (jar) code to Storm. The processed data will be sent to some Kafka topics
- 5 Acquire processed data from Kafka topics and send it to Visdom for visualization
- 6 Remove the dataflow and PCA platforms, either immediately, or after a web validation

# Dataflow Sample – Workflow execution



General Parameters ⓘ

Workflow Variables ⓘ

Variables ⓘ

zookeeper_instance_name	zookeeper-server-1	x
kafka_instance_name	kafka-server-1	x
storm_instance_name	storm-cluster-1	x
visdom_instance_name	visdom-server-1	x
visdom_service_model	http://models.activeeon.cc	x
storm_service_model	http://models.activeeon.cc	x
kafka_service_model	http://models.activeeon.cc	x
zookeeper_service_model	http://models.activeeon.cc	x
message_topic	BitCoinExchangeTopic	x
dataflow_name	BitcoinExchangeDataflow	x
execution_duration	120	x
automatic_kill	true	x

Add

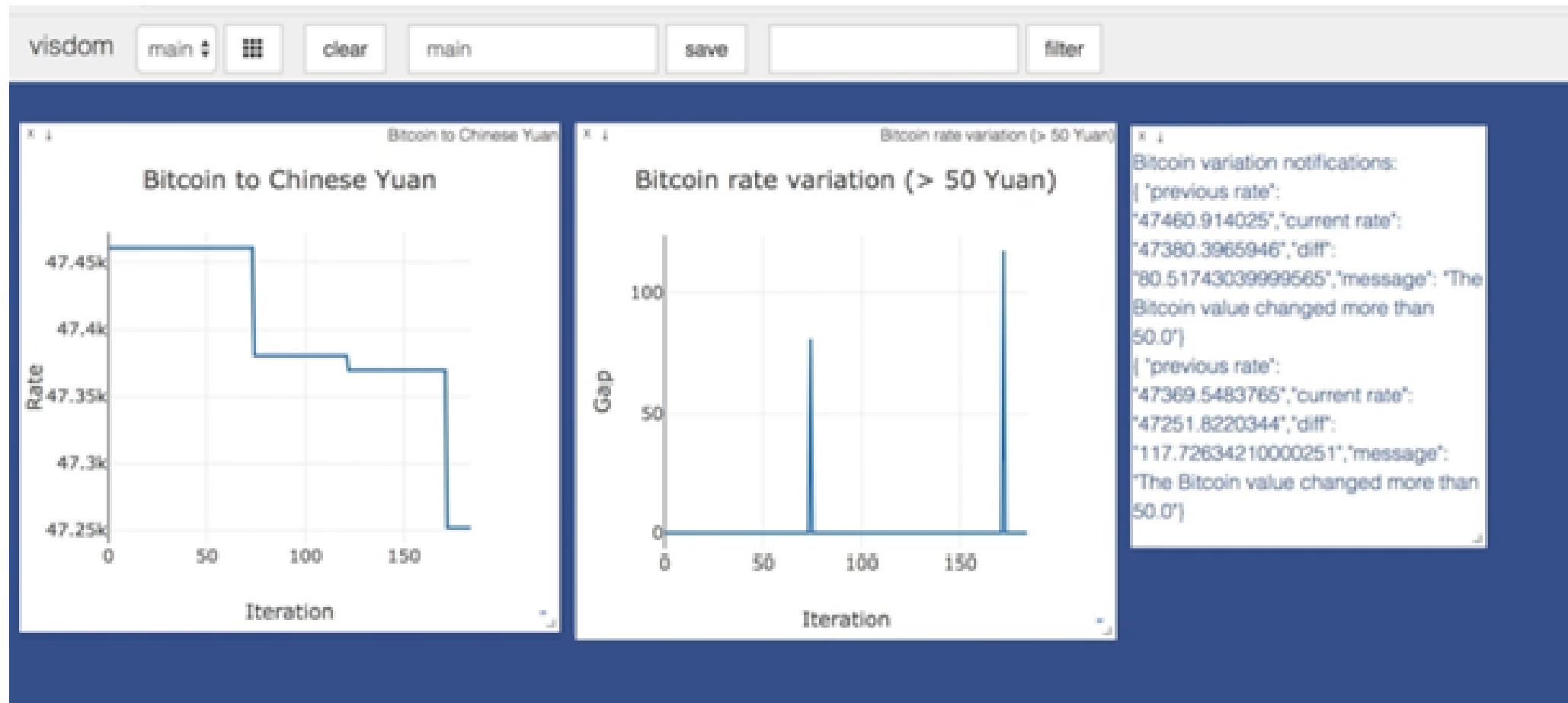
Generic Info ⓘ

Data Management ⓘ

Error Handling ⓘ

- 1 Configure key workflow parameters
  - **execution\_duration:** The execution duration of the dataflow  
Default 120 seconds
  - **automatic\_kill:** Automatic kill of the dataflow and PCA platforms.  
if set to true, the dataflow and PCA platforms are immediately removed after elapsing the execution duration. Otherwise, a web notification is sent to the user, and an acknowledgement is required to remove the running dataflow and PCA platforms.
- 2 Launch the dataflow

# Dataflow Sample – *Result visualization*



# Summary

- Deploy/undeploy complex big data platforms in **One Click**
- Loose coupling and easy integration between big data platforms using ProActive

## **Dataspaces** and **Endpoints**

- Schedule big data workloads and manage resources **efficiently**
- Fault tolerance and recovery for business dataflows

# Automate Accelerate & Scale IT for business



**Activeeon SAS**  
Paris, Sophia Antipolis & San  
Jose, London, Dakar with  
Azure, AWS, and Google cloud



+33 988 777 660



contact@activeeon.com



@activeeon