Real-Time Analytics for students

The 8th Workshop Advanced Analytics and Data Science. Business Analytics & Academia

Sebastian Zając

Warsaw School of Economics

14.09.2021

About Us

In 2020/21 **the Statistical Methods & Business Analytics Unit** is celebrating 15 years of activity.

As the Unit we are responsible for:

- International Workshop Advanced Analytics & Data Science
- Realization of Certificate Program: „Data Scientist with SAS" for Advanced Analytics-Big Data students.
- Two post-diploma studies: „Analytical Academy with SAS, R & Python" and „Statistical Analysis & Data Mining in Business"

Education

Education process at the Advanced Analytics & Big Data program with implementation of the following items for last 6 years.

- Basic & Advanced Programming with SAS with Statistics
- Data Mining
- Credit Scoring
- Logistic Regression
- Event History Analysis
- Advanced Business Analytics. Power of Predictive Models.
- **Real Time Analytics**

Next ...

- Quantum Machine Learning
- MLOps for Business

Why Real Time Analytics

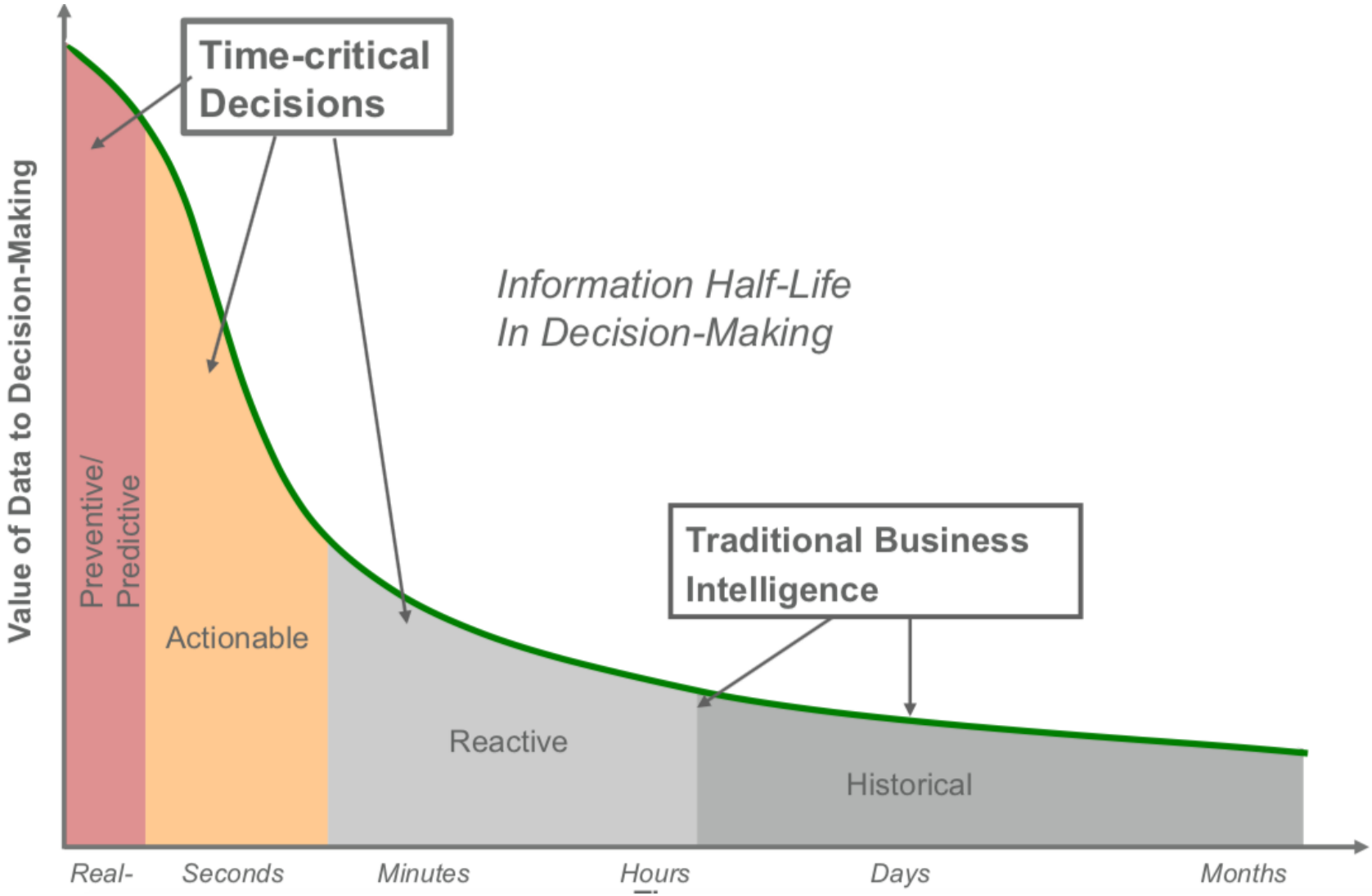
Real time processing deals with streams of data that are captured in real-time and processed with minimal latency to generate real-time (or near-real-time) reports or automated responses.
Streaming analytics, also known as event stream processing, is the analysis of current and "in-motion" data (infinite datasets) through the use of continuous queries, called event streams. These streams are triggered by a specific event that happens as a direct **result of an action** (or set of actions), like:

- a financial transaction,
- equipment failure,
- a social post or a website click,
- other measurable activity.

Data are always produced in some stream process!!

Real Time for Business

Time-critical decisions - Business point of view.



Tools for RTA

All lectures and labs with python codes, env instalation instructions and others (books, links) are available on my website
`https://sebkaz-teaching.github.io/RealTime/`

Environment You will know

- GIT - save work and colaborate with others

Tools for RTA

All lectures and labs with python codes, env instalation instructions and others (books, links) are available on my website
`https://sebkaz-teaching.github.io/RealTime/`

Environment You will know

- GIT - save work and colaborate with others
- Docker - Run (prepare) environment

Tools for RTA

All lectures and labs with python codes, env instalation instructions and others (books, links) are available on my website
`https://sebkaz-teaching.github.io/RealTime/`

Environment You will know

- GIT - save work and colaborate with others
- Docker - Run (prepare) environment
- OOP Python and Jupyter Notebook (Colab) with docker image
  docker run -d -p 8888:8888 sebkaz/docker-data-science

Tools for RTA

All lectures and labs with python codes, env instalation instructions and others (books, links) are available on my website
`https://sebkaz-teaching.github.io/RealTime/`

Environment You will know

- GIT - save work and colaborate with others
- Docker - Run (prepare) environment
- OOP Python and Jupyter Notebook (Colab) with docker image
  docker run -d -p 8888:8888 sebkaz/docker-data-science
- Apache Spark for stream
  docker run -d -p 8888:8888 sebkaz/docker-spark-jupyter

Tools for RTA

All lectures and labs with python codes, env instalation instructions and others (books, links) are available on my website
`https://sebkaz-teaching.github.io/RealTime/`

Environment You will know

- GIT - save work and colaborate with others
- Docker - Run (prepare) environment
- OOP Python and Jupyter Notebook (Colab) with docker image
  docker run -d -p 8888:8888 sebkaz/docker-data-science
- Apache Spark for stream
  docker run -d -p 8888:8888 sebkaz/docker-spark-jupyter
- Apache Kafka with docker compose image
  `https://github.com/sebkaz/docker-kafka-rta`

Tools for RTA

All lectures and labs with python codes, env instalation instructions and others (books, links) are available on my website
`https://sebkaz-teaching.github.io/RealTime/`

Environment You will know

- GIT - save work and colaborate with others
- Docker - Run (prepare) environment
- OOP Python and Jupyter Notebook (Colab) with docker image
  docker run -d -p 8888:8888 sebkaz/docker-data-science
- Apache Spark for stream
  docker run -d -p 8888:8888 sebkaz/docker-spark-jupyter
- Apache Kafka with docker compose image
  `https://github.com/sebkaz/docker-kafka-rta`
- Databricks - free cloud env for Spark

Tools for RTA

All lectures and labs with python codes, env instalation instructions and others (books, links) are available on my website
`https://sebkaz-teaching.github.io/RealTime/`

Environment You will know

- GIT - save work and colaborate with others
- Docker - Run (prepare) environment
- OOP Python and Jupyter Notebook (Colab) with docker image
  docker run -d -p 8888:8888 sebkaz/docker-data-science
- Apache Spark for stream
  docker run -d -p 8888:8888 sebkaz/docker-spark-jupyter
- Apache Kafka with docker compose image
  `https://github.com/sebkaz/docker-kafka-rta`
- Databricks - free cloud env for Spark
- SAS Event Stream Processing

Tools for RTA

Lectures

- Structured and unstructured Small and Big Data
- Classical (OLTP and OLAP) and new data processing models.
- Batch vs Streaming processing
- Lambda and Kappa architecture
- Time in stream data processing

Labs Batch Python

- Structured and unstructured data in python
- OOP Perceptron and Adeline neural networks
- Production env with flask and docker

Tools for RTA

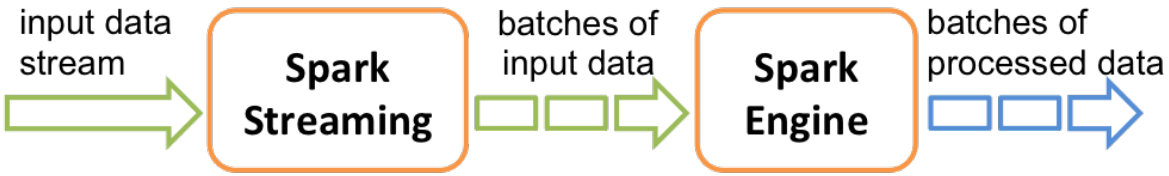
Labs Batch Spark

- Spark RDD - Transformations (map, reduce...) and Actions for Lazy evaluation
- Spark DataFrame - basic transformations with Spark SQL
- Spark ML - short example of Binary Classification

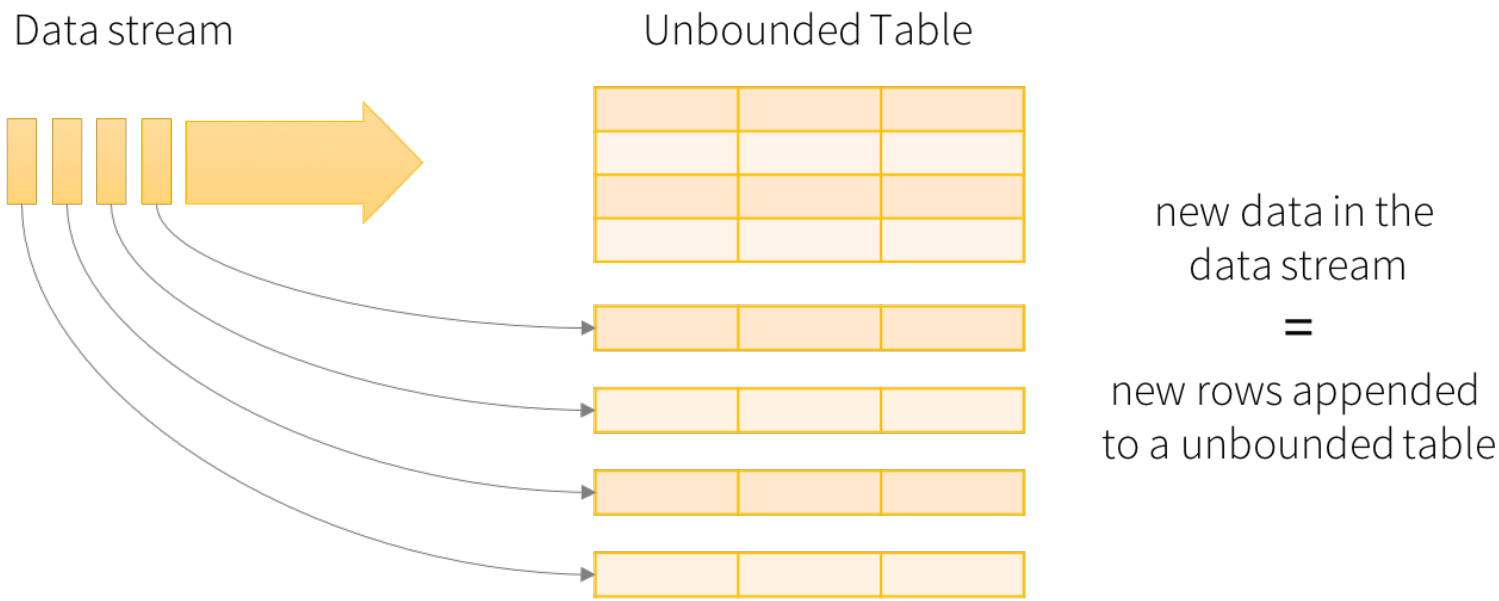
Labs Stream Spark

- Spark DStreams are built on Spark RDDs (SparkContext to StreamingContext)
- Stateless and statefull operations
- Spark Streaming with Spark DataFrame

Apache Spark RDD Streaming



Apache Spark DataFrame Streaming



Data stream as an unbounded table

Tools for RTA

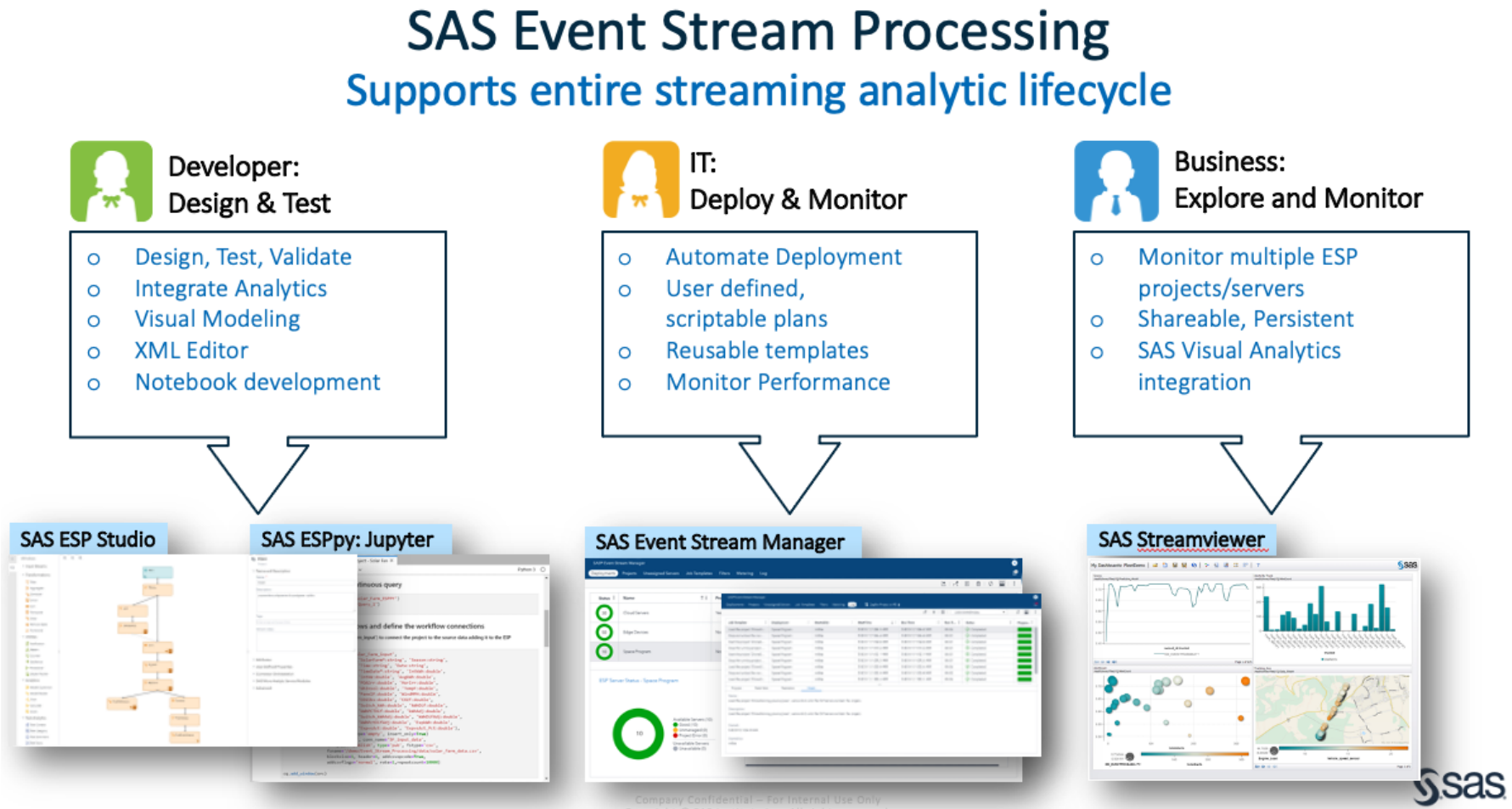
Labs Apache Kafka with Spark

- Apache Kafka Stream with Apache Spark

Clodu env for Stream

- Databricks
- IBM Cloud
- Microsoft Azure Stream
- Google Cloud Stream
- AWS Stream

Tools for RTA



Summary

Thanks for Your Attention!
sebastian.zajac@sgh.waw.pl