



# R jako główna platforma do zaawansowanej analityki w Enterprise

Wit Jakuczun,  
WLOG Solutions

2017-09-28

**Biznes nie rozmawia o  
zaawansowanej analityce.**

**Biznes nie rozmawia o platformie  
analitycznej.**

**Biznes rozmawia o tym jak  
sprawnie zarabiać pieniądze.**

## Elastyczność

Czy mogę rozwiązać  
**ten konkretny  
problem?**

## Jakość

Na ile  
**wiarygodna/pewna** jest  
informacja?

## Wydajność

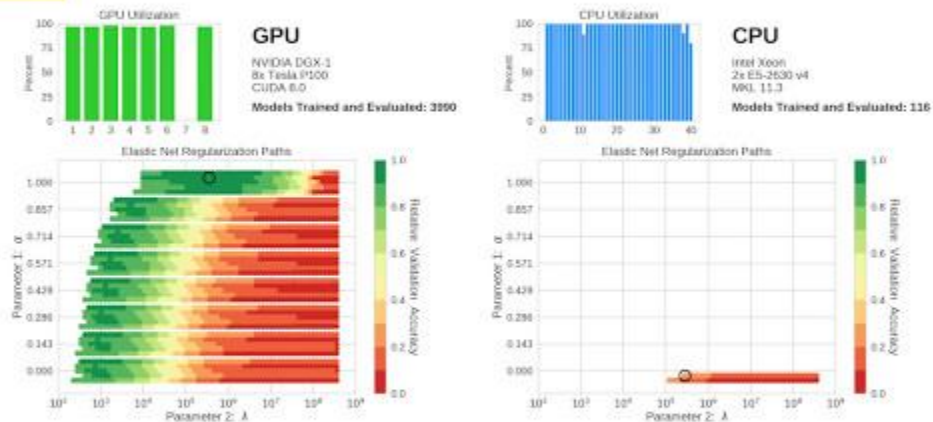
**Jak długo** muszę  
czekać na informację?

## Społeczność!

Jak **szybko** mogę dostać **wsparcie?**



## H2O.ai Machine Learning – Generalized Linear Modeling



U.S. Census dataset (predict income): 45k rows, 10k cols

Dataset split: 5-fold cross-validation,  $\alpha = 5 \times 10^{-4}$ ,  $\lambda = 0.71$ , full 1 epoch

**4000x5 modeli (CV-5) elastic-net dla danych wymiaru 45Kx10K w 1,5 minuty!**

A ile czasu Twój biznes czeka na  
wsparcie decyzji?



# R jako platforma analityczna w Enterprise

Jak to sprzedać swojemu szefowi oraz zaplanować wdrożenie?



# Obietnica dla Szefa

Będziemy w stanie **szybciej** dać  
**lepszey jakości** odpowiedź na  
**więcej pytań** biznesu  
jednocześnie obniżając tzw.  
**licence footprint.**

# Jak to osiągnąć?

# Ewolucja a nie **R**ewolucja\*

\*R i tak się obroni wynikami :)

# Zarządzanie zmianą!



**Pokaż korzyści i  
wytlumacz**

**Wdrażaj na  
realnych  
problemach**

Zafunduj  
szkolenia dla  
zespołu

Dostosuj  
projekty i  
architekturę

Idź na najbliższy  
uniwersytet

Znajdź wsparcie  
zewnętrzne

# Platforma analityczna w Enterprise

Komponenty oraz rola R

# Wymagania

## Elastyczność

Czy mogę rozwiązać  
**ten konkretny**  
**problem?**

## Jakość

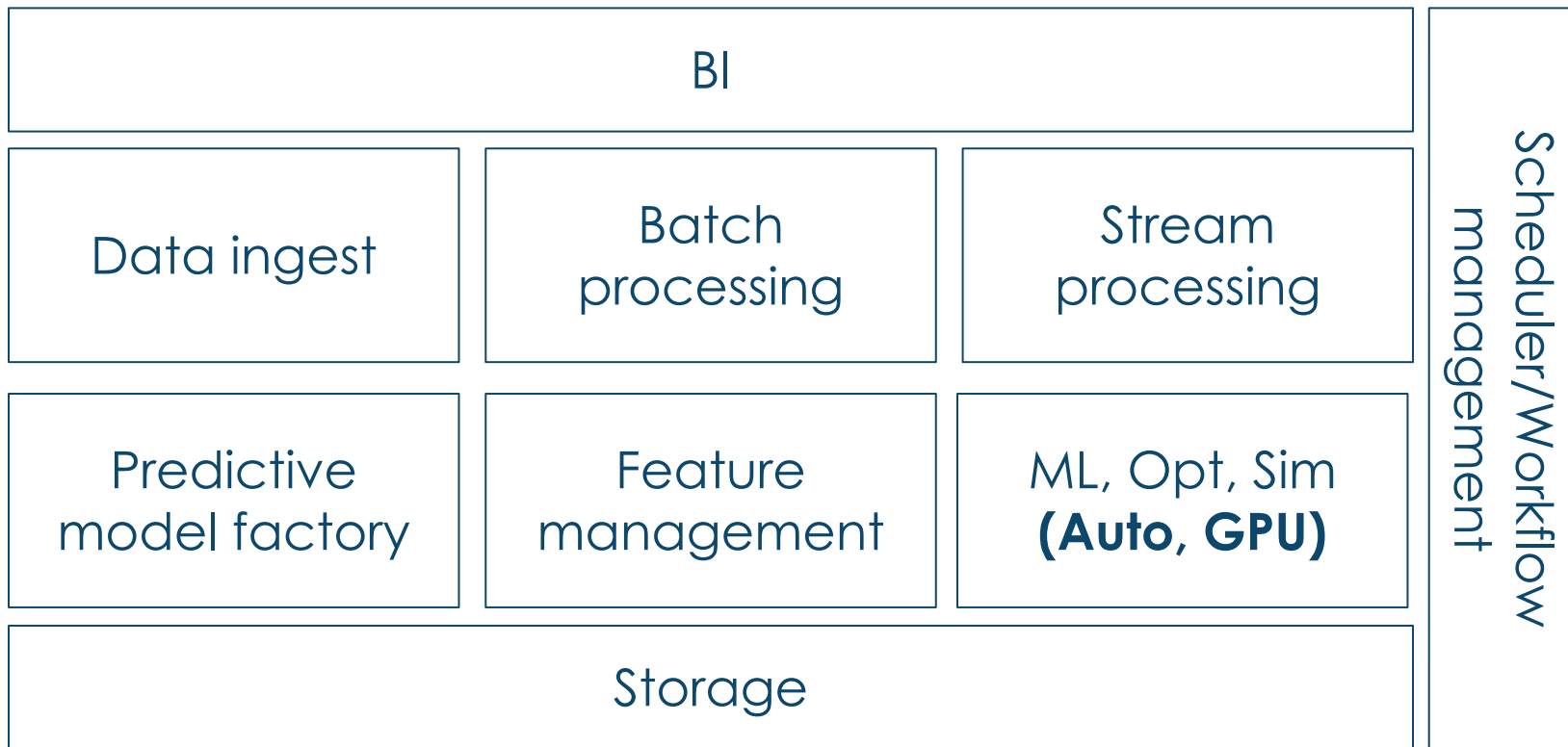
**Na ile**  
**wiarygodna/pewna** jest  
informacja?

## Wydajność

**Jak długo** muszę  
czekać na informację?



# Komponenty platformy analitycznej



# Rola i miejsce R (I)

**R jest językiem do pisania logiki  
przetwarzania w ramach  
platformy.**

# Rola i miejsce R (II)

Predictive  
model factory

Data Quality  
management

Machine  
Learning

Analizy ad-hoc

Raporty  
powtarzalne

Automatyzacja  
procesów

# Dlaczego R?

System  
dokumentacji  
pakietów

System  
zależności  
między  
pakietami

Dostęp do  
najlepszych  
technologii

“Język-klej”

Szybkie  
prototypowanie  
w dużej skali

Dostęp do  
wiedzy i  
kompetencji

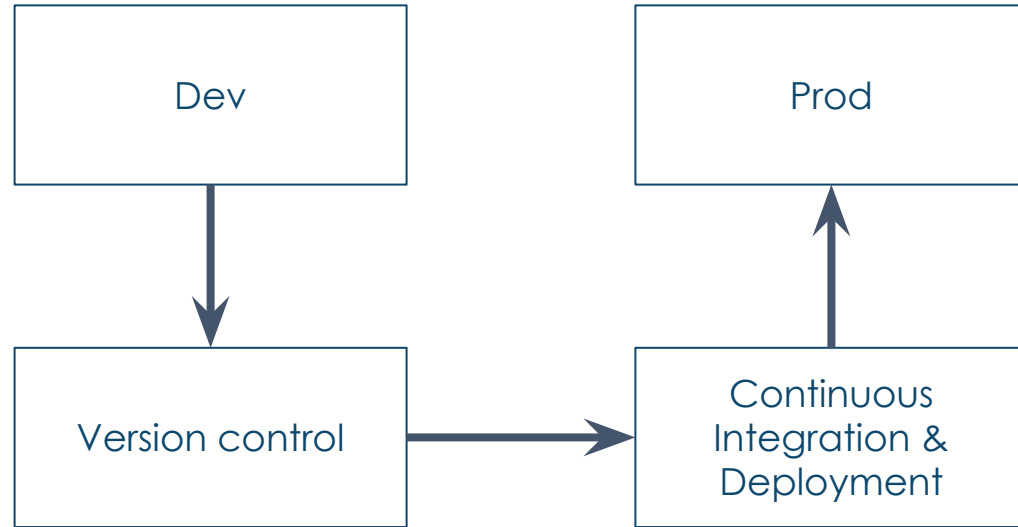
# Platforma analityczna w Enterprise

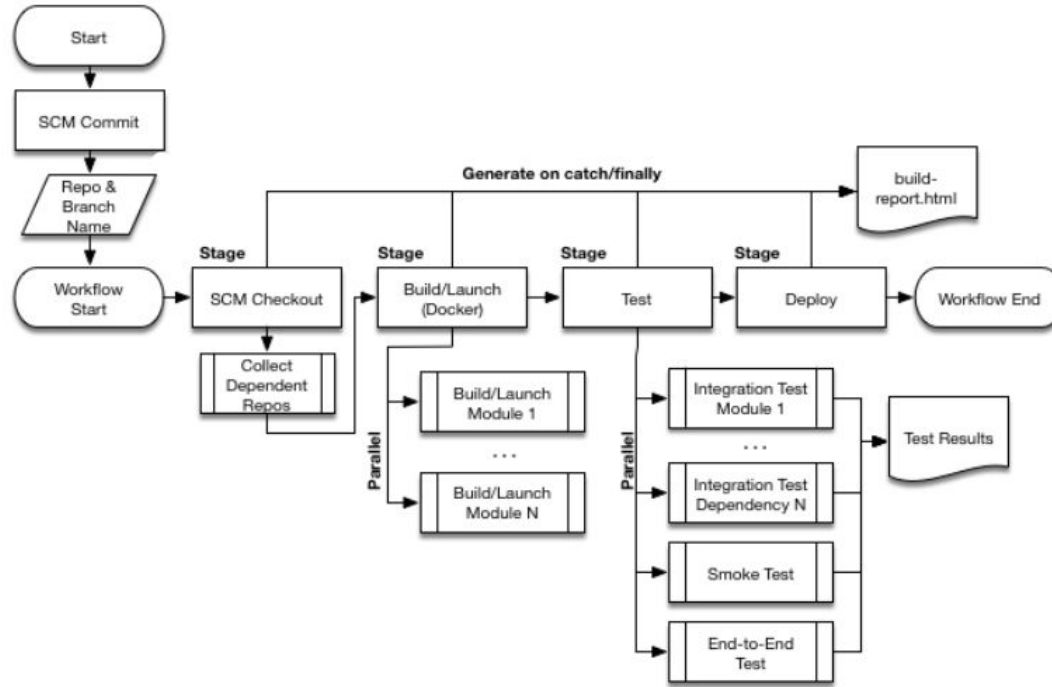
Jak zarządzać środowiskiem R?

## **R “z pudełka”**

nie ma bezpośredniego wsparcia  
dla rozwiązań wielkiej skali  
w środowisku Enterprise.

# Proces developerski





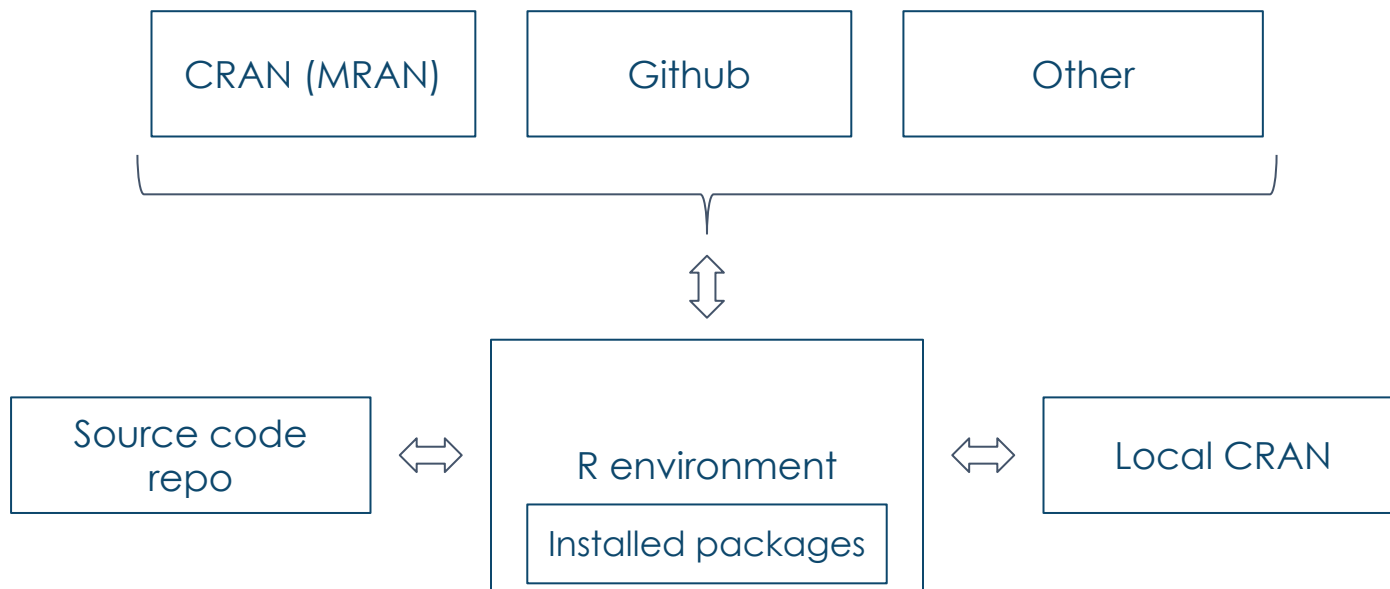
[Jenkins](#) exemplary pipeline



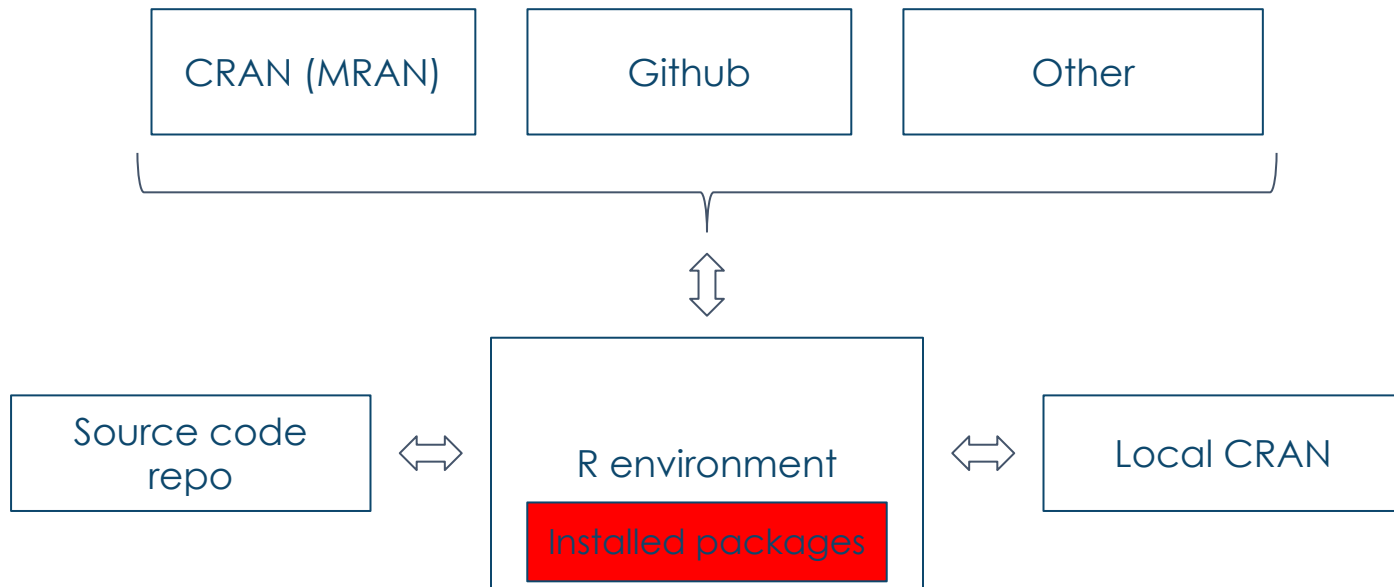
# Best practices by WLOG



# Co jest blokerem?



# Co jest blokerem?



# data.table

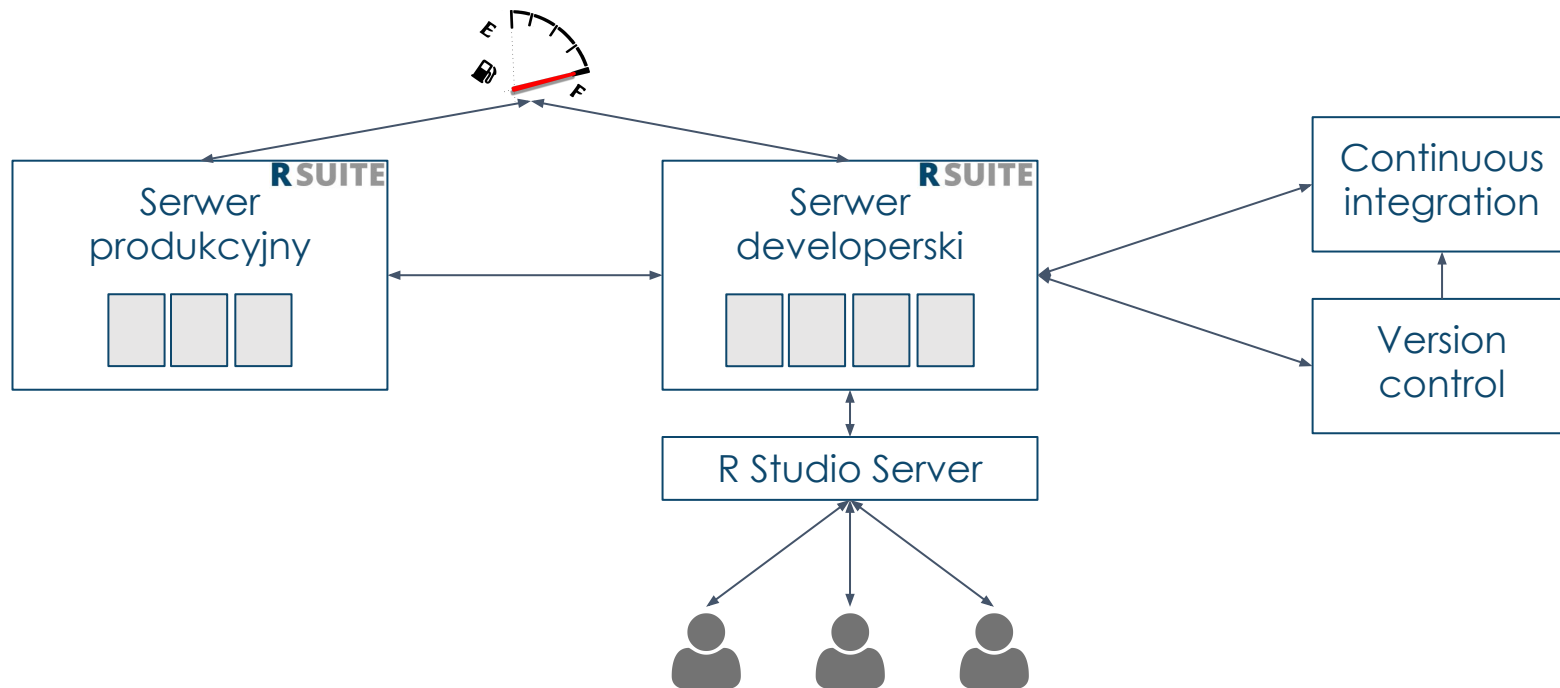
## 🔗 Changes in v1.9.8 (on CRAN 25 Nov 2016)

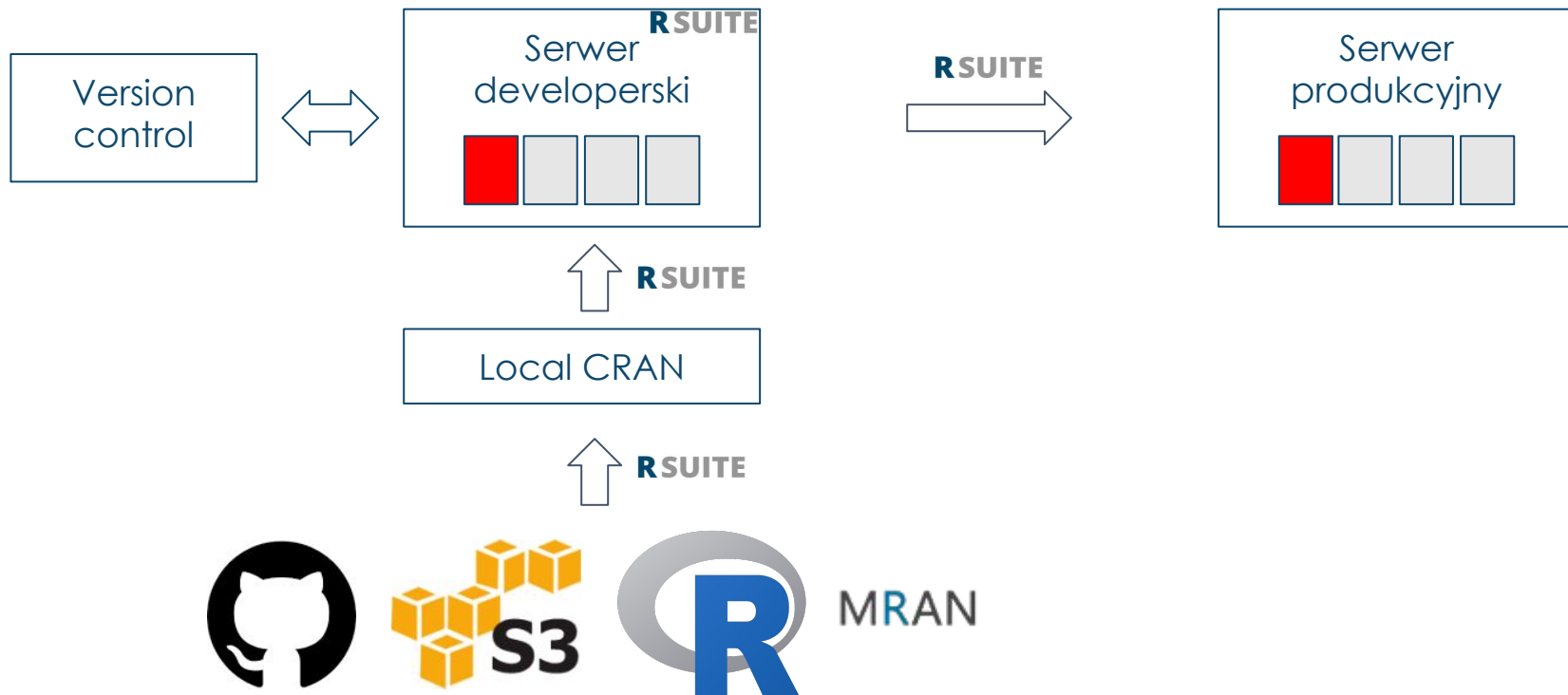
### POTENTIALLY BREAKING CHANGES

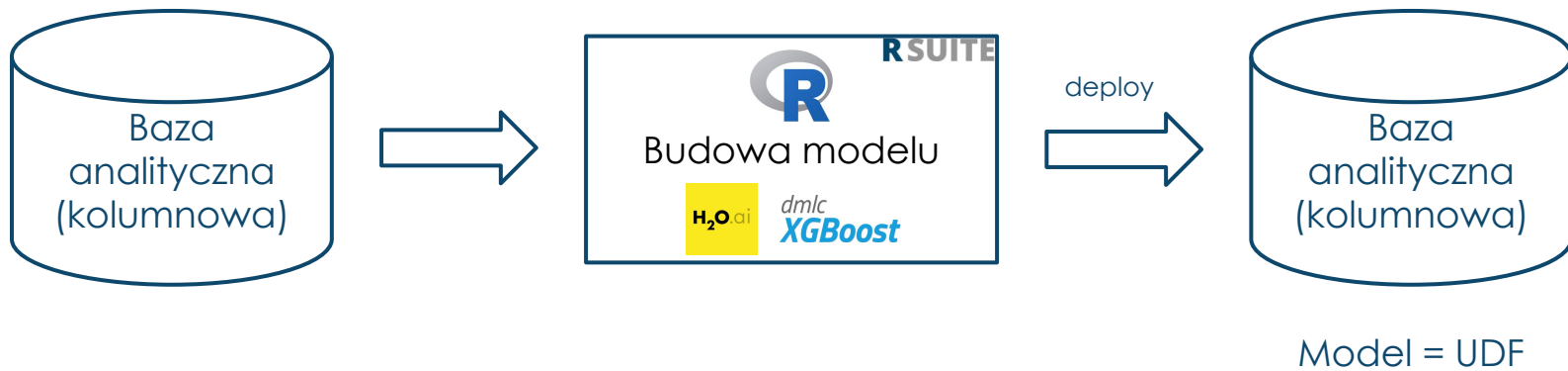
1. By default all columns are now used by `unique()`, `duplicated()` and `uniqueN()` data.table methods, [#1284](#) and [#1841](#).  
To restore old behaviour: `options(datatable.old.unique.by.key=TRUE)`. In 1 year this option to restore the old default will be deprecated with warning. In 2 years the option will be removed. Please explicitly pass `by=key(DT)` for clarity. Only code that relies on the default is affected. 266 CRAN and Bioconductor packages using data.table were checked before release. 9 needed to change and were notified. Any lines of code without test coverage will have been missed by these checks. Any packages not on CRAN or Bioconductor were not checked.

# Platforma analityczna w Enterprise

Przykłady z życia









# Platforma analityczna w Enterprise

Podsumowanie

# Platforma R w środowisku Enterprise



**Środowisko R** pozwala stworzyć  
platformę analityczną, która  
**wspiera biznes** sprawnie  
**zarabiać pieniądze.**



# RSUITE

Field tested R ecosystem for Enterprise

<http://rsuite.io>

**Wit Jakuczun, PhD**

[wit.jakuczun@wlogsolutions.com](mailto:wit.jakuczun@wlogsolutions.com)