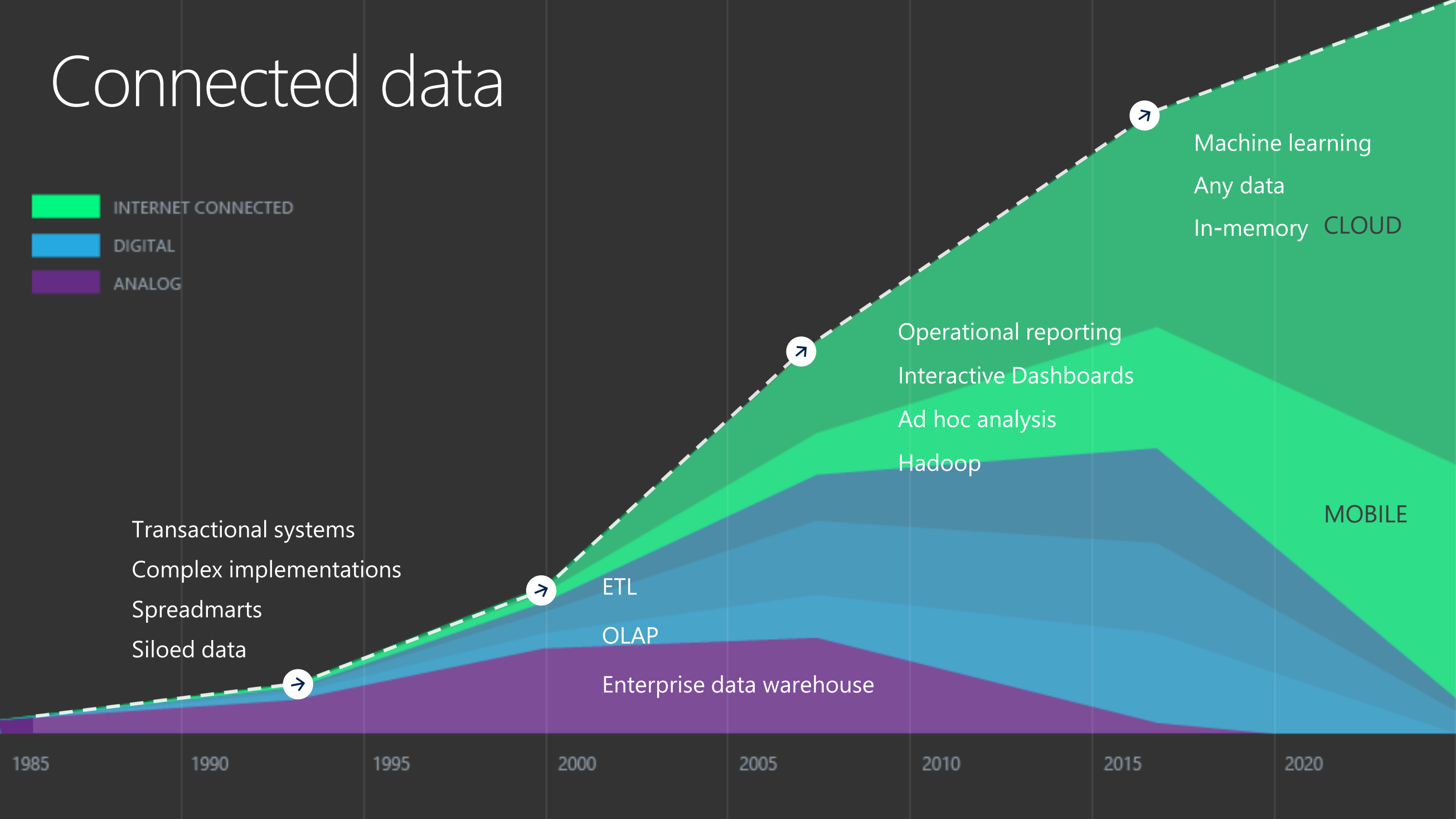


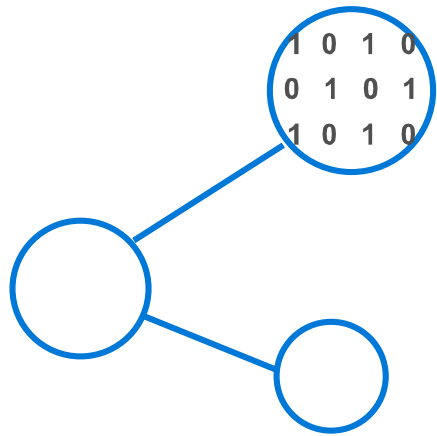
SQL Server 2016 R services

Conf.Dr. Cristian Kevorchian
Universitatea din București

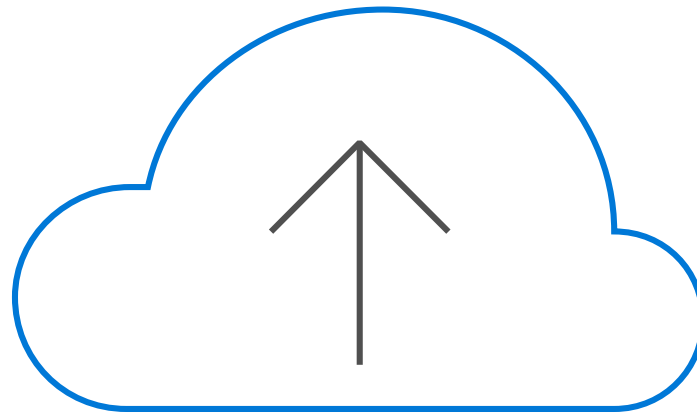
Connected data



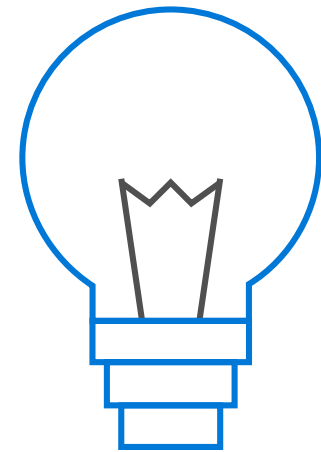
Trei tendințe convergente



Datele

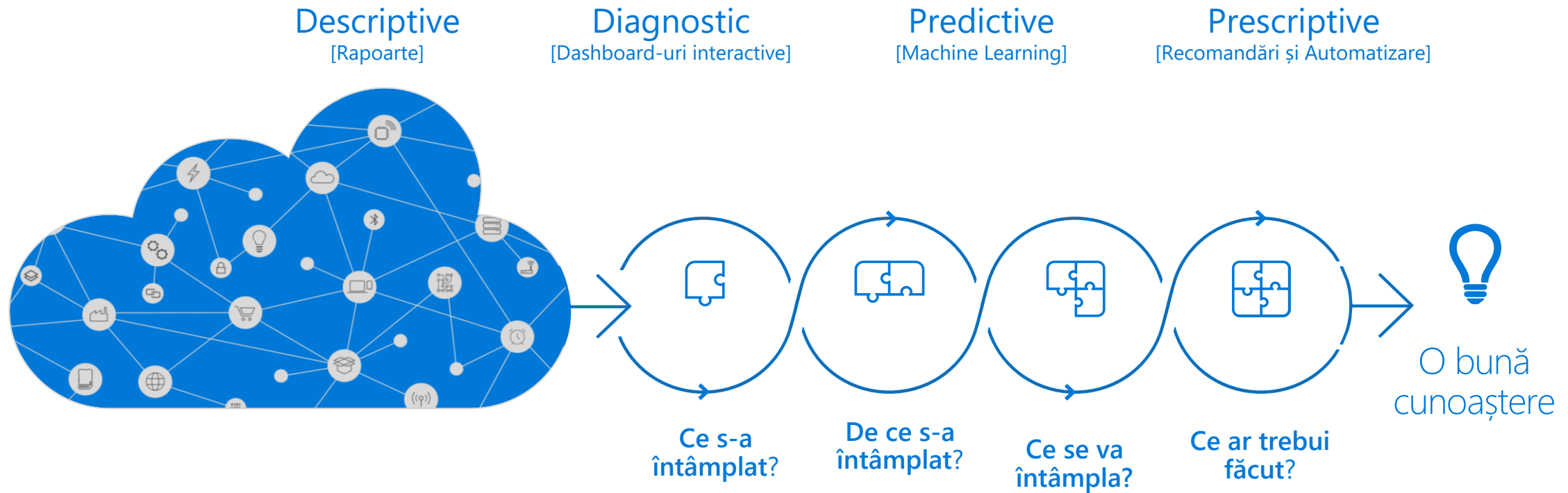


Cloud



Analiza profundă
datelor

De la date la decizie și acțiune



Aplicații industriale ale analiticelor



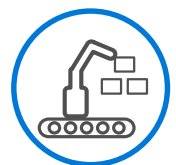
Retail & Consum



Servicii Financiare & Asigurări



Sectorul public



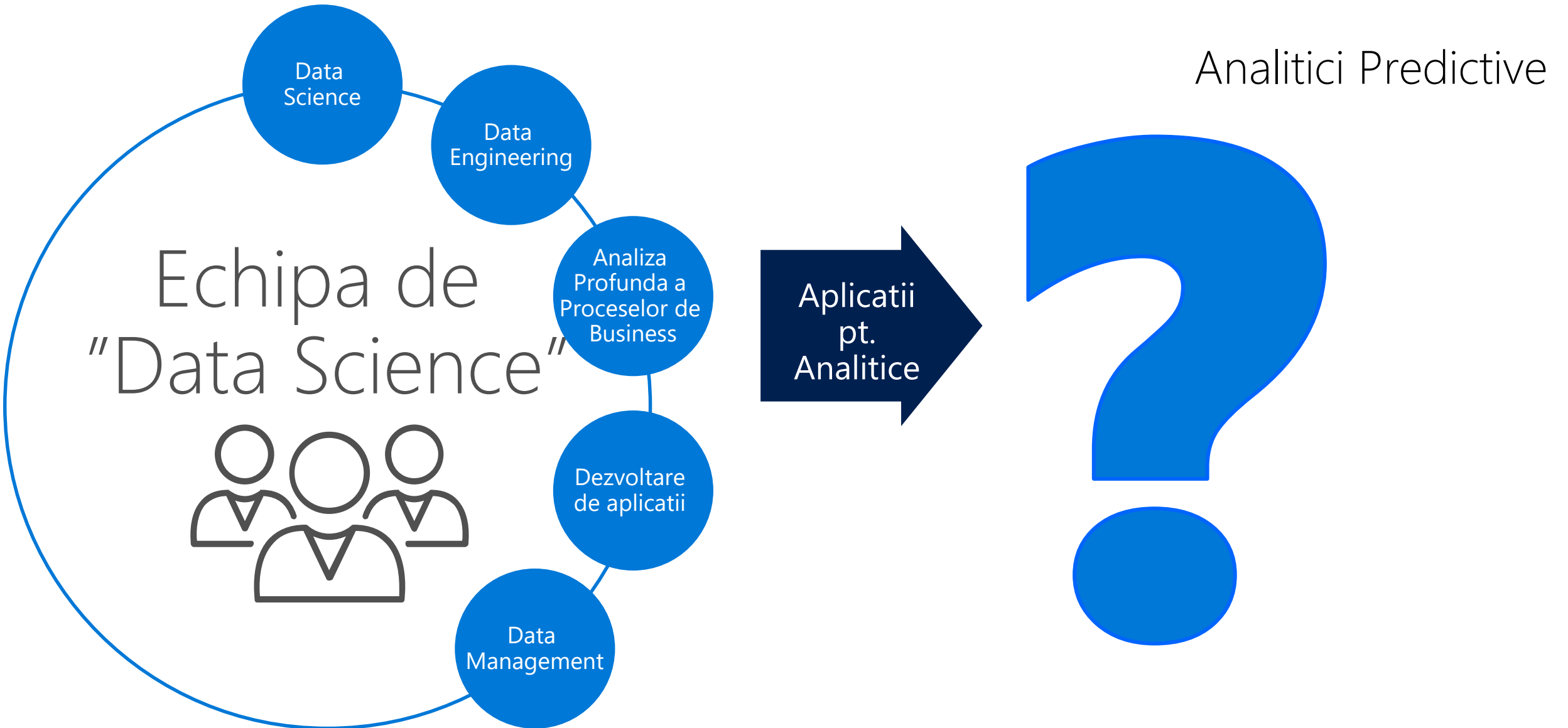
Manufactură



Sănătate



Succesul implică convergența competențelor



Bariere în adoptarea analiticelor

Deficitul de talente

- Rigoare academică
- Competiția talentelor

Productivitate

- Integrarea Complexității
- Bariere între Tool-uri, Competente & Cultură

Infrastructură Complexă

- Volume mari de date, Diversitate
- Restricții la capitolul Securitate & Guvernanță
- Evoluția rapidă a platformelor

Dinamică scăzută a inovației

- Rata scăzută a experimentării
- Operationalizare complexă

Costuri mari

- Produse vechi
- Irregular Workload

Introducere în R

Ce este R



Platforma
Limbaaj

- Un limbaj de programare pentru statistică
- Un tool pentru vizualizare date
- Open source

Comunitate

- 2.5+ milioane utilizatori
- Prezent în multe universități
- Utilizat de mulți specialiști tineri
- Numeroase grupuri de utilizatori

Ecosistem

- 8000+ algoritmi livrați gratis în CRAN
- Scalabil pentru Big Data
- "Rich application" & integrare pe platforme

R(Open Source) nu este la nivel "Enterprise"

Fluxurile de date depășesc posibilitatile R in varianta Open Source

Operații In-Memory











Paralelism limitat

Transferul Datelor și Duplicarea se realizează cu costuri mari

"Not enterprise ready"

Suport pentru comunitati încă neadecvat

Nu are SLAs sau Modele de suport

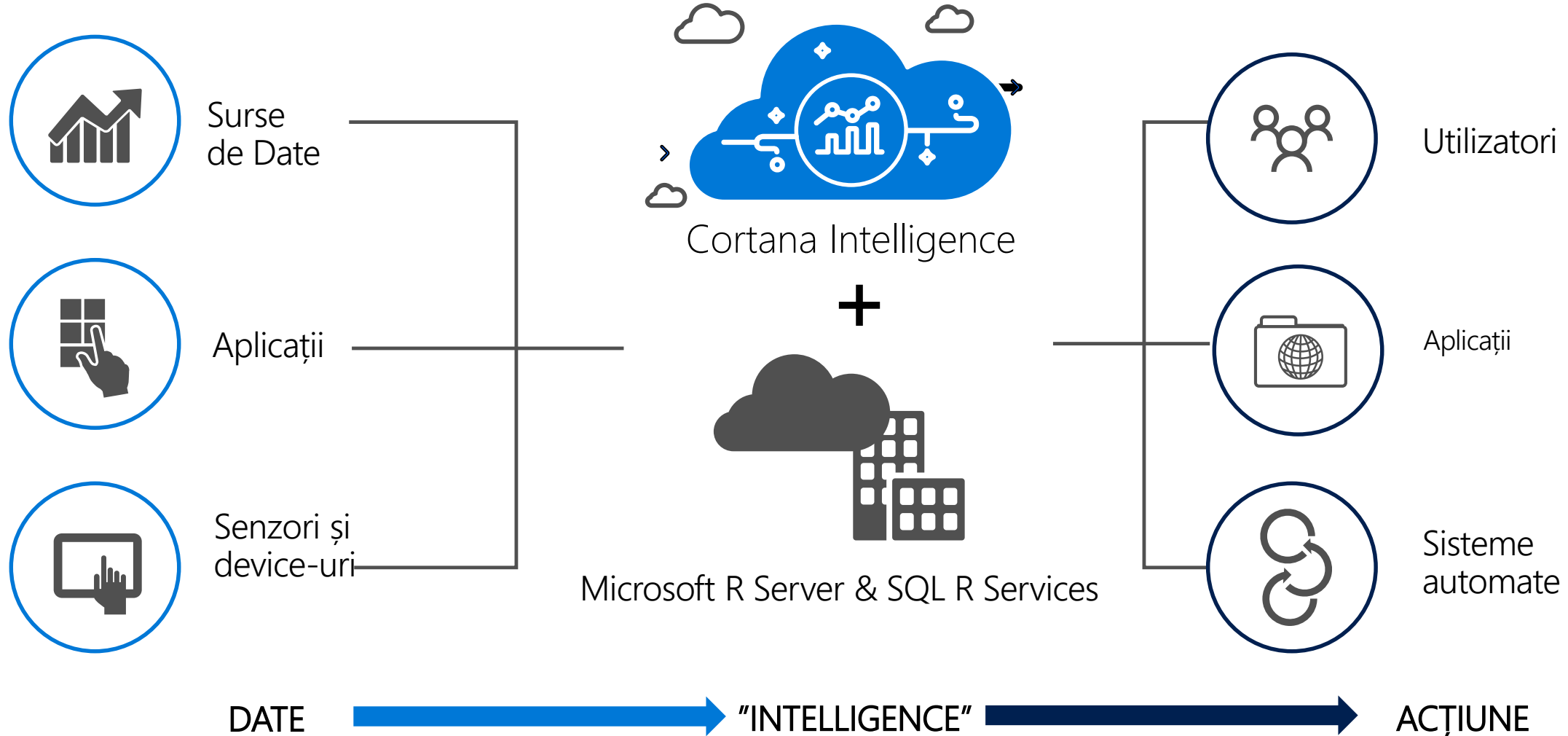
Language Rank	Types	2016 Spectrum Ranking	2015 Spectrum Ranking	2014 Spectrum Ranking
1. C		100.0	100.0	100.0
2. Java		98.1	99.9	99.3
3. Python		98.0	99.4	95.5
4. C++		95.9	96.5	93.5
5. R		87.9 <i>2016</i>	81.3	92.4
6. C#		86.7	84.8 <i>2015</i>	84.8
7. PHP		82.8	84.5	84.5
8. JavaScript		82.2	83.0	78.9
9. Ruby		74.5	76.2	74.3 <i>2014</i>
10. Go		71.9	72.4	72.8

Source: IEEE Spectrum July 2014, 2015 & 2016

Introducere în Microsoft R

De la date și analiza datelor(intelligence) la acțiune

- "On-Prem" sau în the Cloud



Introducere în Microsoft R Server

Linux, Windows, Hadoop & Teradata, and SQL Server 2016, 2017, 2019

Performanță ridicată, R Scalabil

R 100% open source

Compatibil cu CRAN, MRAN, GitHub

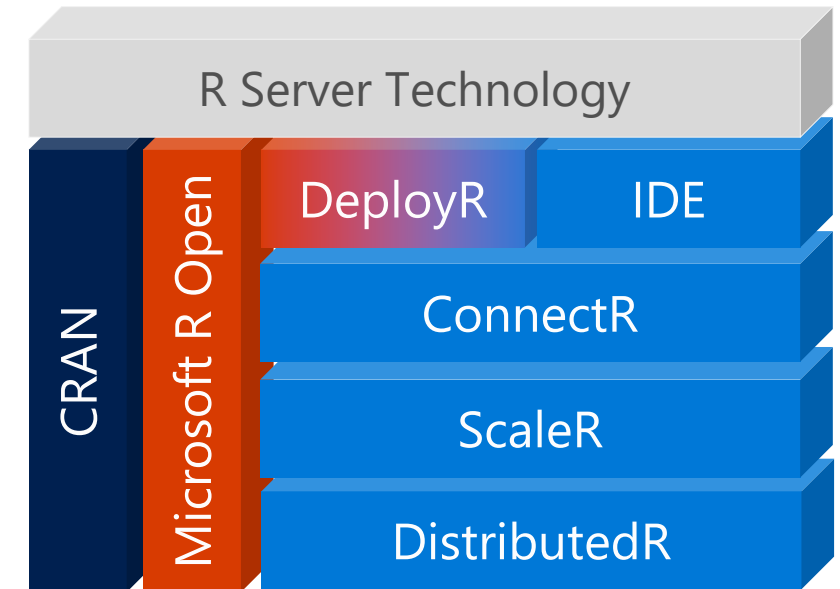
Masiv Scalabil

Multi-platformă

Conectivitate Big Data

Arhitectură Hibridă

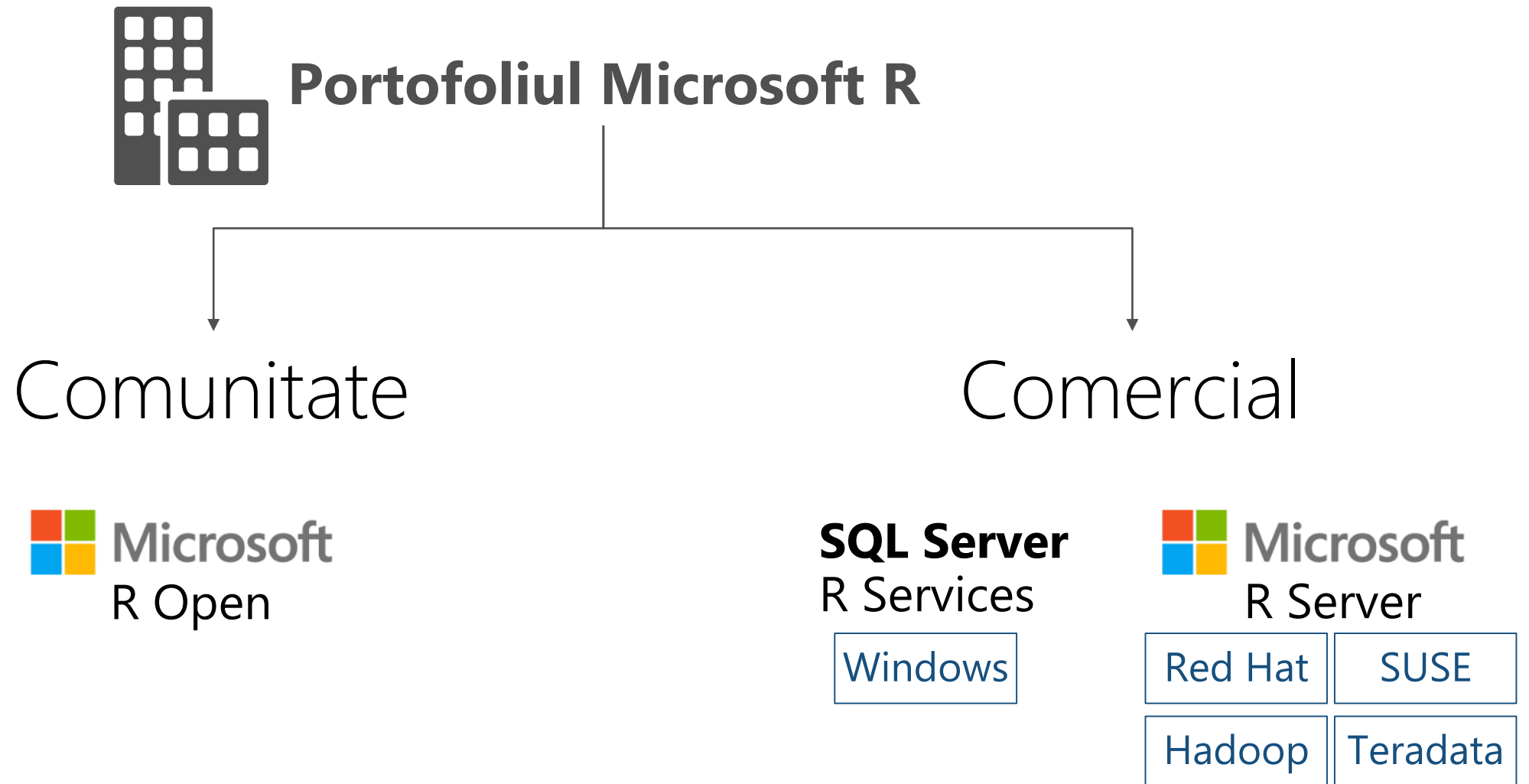
Alegerea IDE-ului



Componente
Open Source

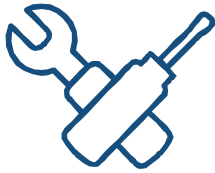
Componente licențiate

Portofoliul Microsoft R



Introducere SQL Server R Services

Introducere SQL Server R Services



Simplitate

Viteză de execuție și
scalare -enterprise

Analitice DB intedrate

Procesare și threading
paralel

Reutilizare competente
SQL pentru "data
engineering"



Scalabilitate

In-database
deployment-ul BD

Scalabilitate
memorie+disk

Fără limite de memorie
pentru R

"Write once, deploy
anywhere"



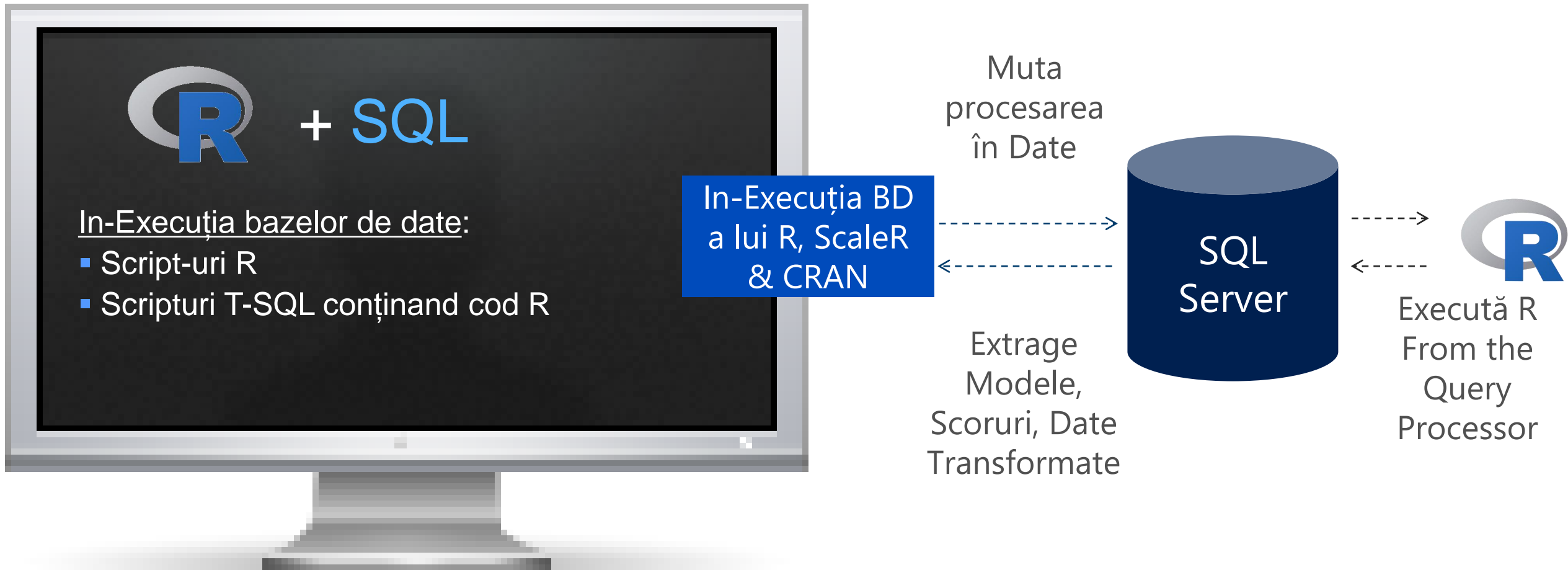
Cost redus

Included in SQL Server
2016

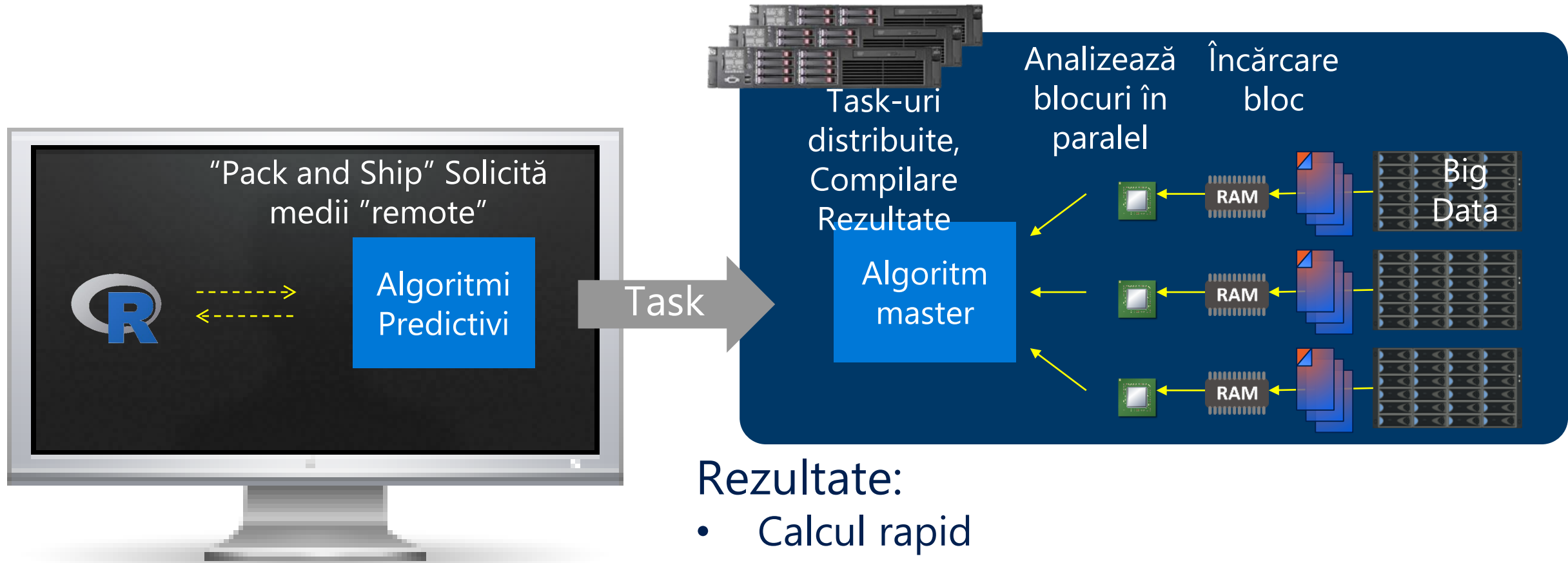
Reutilizează codul R

Elimină mutarea datelor

Noi funcționalități SQL Server :



Execuția la distanță



Rezultate:

- Calcul rapid
- Capacitate de procesare a datelor mărită
- Preocupare minima asupra securității
- Fără mutarea datelor, Fără Copii

Microsoft R Server delivers

- Scalare enterprise în spark, hadoop, RDBMSs & EDWs
- Nicio limitare privind memoria
- Variante de IDE-uri Windows si Linux
- Deployment stabil
- Portabilitate "write-once-deploy-anywhere"
- Evoluție în direcția cloud hibrid

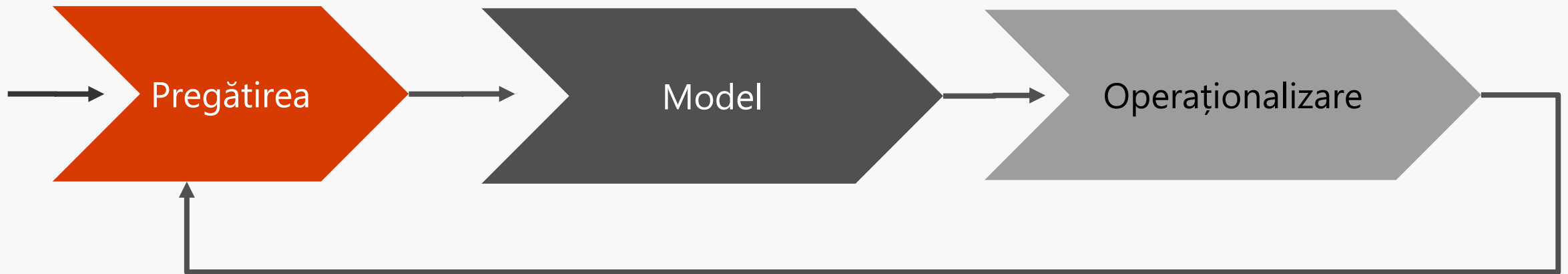


Ciclul de viață al analiticelor

Pregătirea: Colectarea, "curățarea", profilarea și transformarea diverselor date relevante pentru subiect.

Model: Utilizarea statisticii și algoritmicii ML pentru a construi clasificatori și modele de regresie

Operaționalizare: Efectuarea de predicții și vizualizări pentru susținerea app. de business



Aplicațiile înainte și după

Înainte



Online Shopping

Your Cart



Black and White Diamond Heart



Diamond Pave Earrings



Diamond Tennis Bracelet

Add Items



Online Shopping

Shopping Cart

Your Cart



Black and White Diamond Heart

\$130

x



Diamond Pave Earrings

\$569

x



Diamond Tennis Bracelet

\$360

x

Add Items

Account ID

131395

Total Purchase \$1,059

Purchase

Reset

There is a problem
with this order.
Please call [800-555-
2222](tel:800-555-2222) for more
information

Din Perspectiva BD

Crearea cardului de
cumpărături

Primirea_Solicitării

Linie_de_Produse_Solicitare

MODELE

	"Intelligence"	

Submitted orders
O_ORDERSTATUS = 'S'

Solicitări

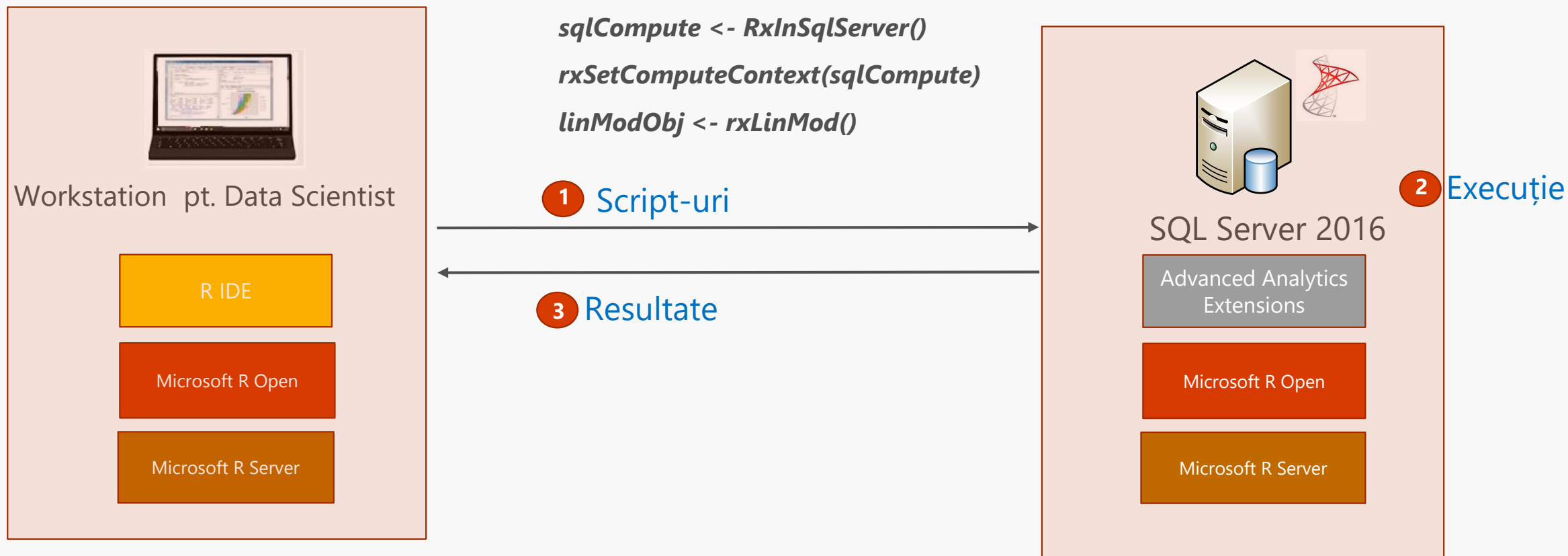
Linie_de_produse

Card de cumpărături

Efectuarea solicitării

Exploatarea Datelor și Modelare Predictivă ('Data Scientist')

Working from my R IDE on my workstation, I can execute an R script that runs in-database, and get the results back.



Operaționalizarea codului R prin T-SQL

Pot aplea Procedură Stocată de Sistem în T-SQL din propria aplicație and pot executa scriptul R in-baza de date. Rezultatele sunt intoarse in aplicatie (predicții, grafice, etc).

