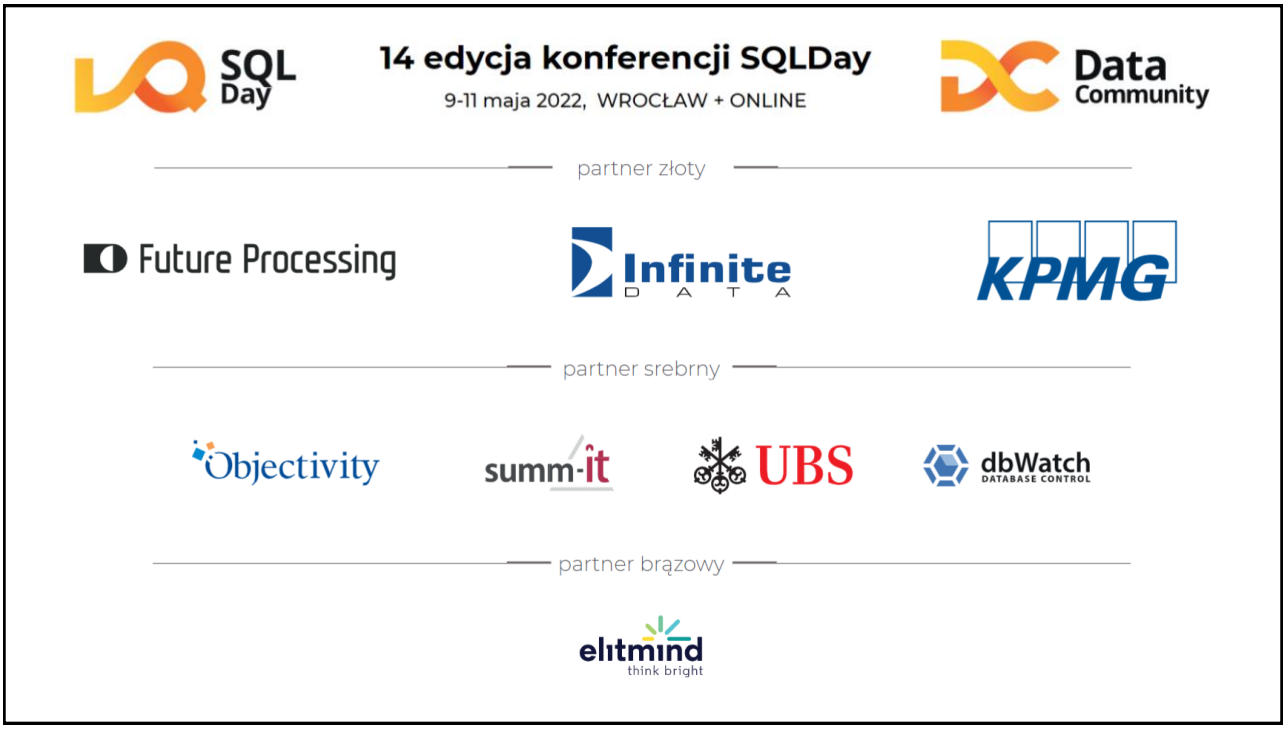
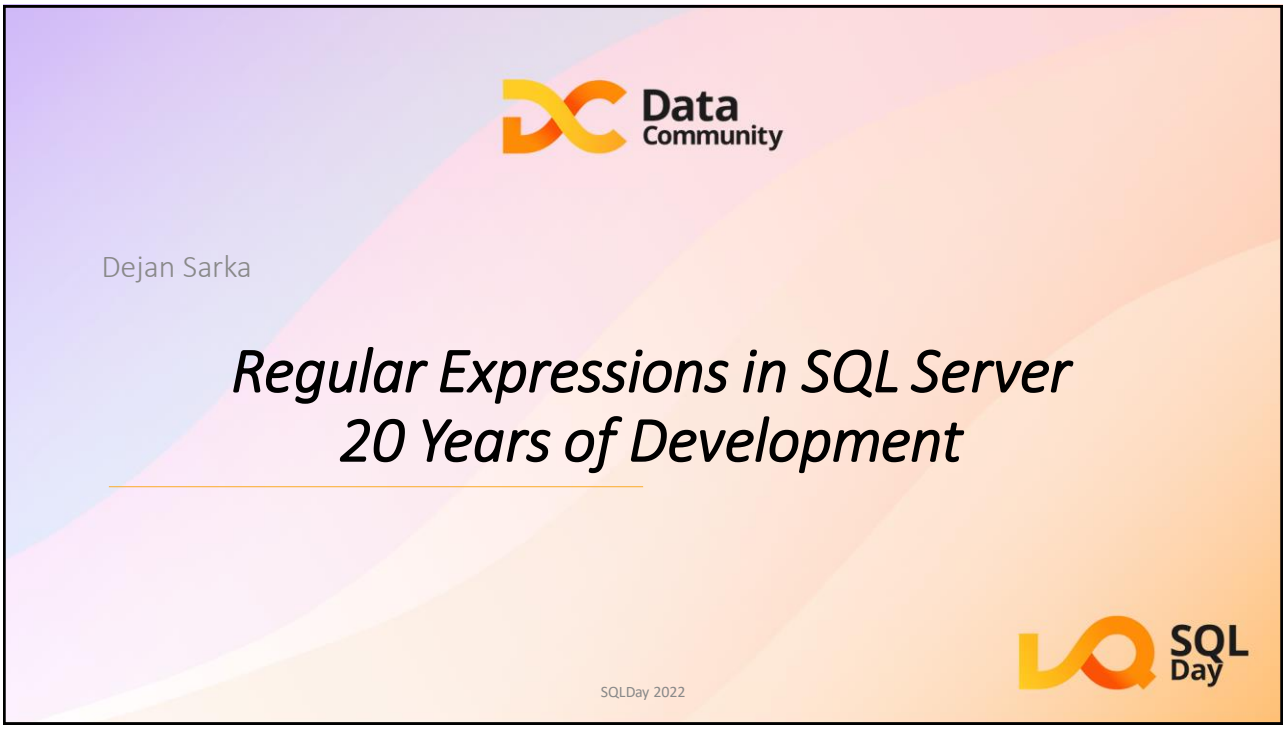

SQL Day
14 edycja konferencji SQLDay
9-11 maja 2022, WROCŁAW + ONLINE
Data Community
partner złoty
Future Processing
Infinite DATA
KPMG
partner srebrny
Objectivity
summ-it
UBS
dbWatch DATABASE CONTROL
partner brązowy
elitmind think bright


Data Community
Dejan Sarka
Regular Expressions in SQL Server
20 Years of Development
SQL Day
SQLDay 2022

Instructor Bio

- Dejan Sarka
  - 36 years of experience
  - Data Platform Server MVP, MCT
  - 19 books
  - 20+ courses
  - Focus:
    - Data science
    - Data quality
    - Data modeling

Agenda

- Introduction
- Comparing:
  - SQL Server 2019: Java
  - SQL Server 2000: TSQL LIKE
  - SQL Server 2000: FTS CONTAINS
  - SQL Server 2005: C# (and VB)
  - SQL Server 2005: XQuery
  - SQL Server 2016: R
  - SQL Server 2017: Python
- Conclusion

Introduction (1)

- SQL Server 2019 supports **Java**
  - Tutorial: Search for a string using regular expressions (regex) in Java
  - A lot of Java code for finding two out of three rows with word „Java“
  - Also not nice code (no schema.object, no semicolons)

```
CREATE TABLE testdata (
    id int NOT NULL,
    "text" nvarchar(100) NOT NULL
)
GO

-- Insert data into test table
INSERT INTO testdata(id, "text") VALUES (1, 'This sentence contains java')
INSERT INTO testdata(id, "text") VALUES (2, 'This sentence does not')
INSERT INTO testdata(id, "text") VALUES (3, 'I love Java!')
GO
```

Other Possibilities (1)

- In 2005, the first CLR examples (**C#** and **VB**) were about regular expressions as well
  - String Utility Functions Sample from the documentation
  - Stricter security in latest versions
- SQL Server 2005 also brought the **XML** data type with **XQuery** language
  - Can use the **contains()** function
  - SQL Server does not support the **matches()** function
  - Can use XML indexes

Other Possibilities (2)

- Of course, one could simple use the **TSQL LIKE** operator
  - With us from... Huh... At least 1992
  - Can benefit from nonclustered indexes
- The Full-Text Search (**FTS**) **CONTAINS** operator works as well
  - FTS available from Sql Server 2000
  - Uses special FTS indexes

Other Possibilities (3)

- The **R** language came in SQL Server 2016
  - Scalable **RevoScaleR** library
  - Use the **rxDataStep()** function with the **grepl()** function
  - Or use the **data.table** object
- SQL Server 2017 added support for the **Python** language
  - The **revoscalepy** library also supports the **rxDataStep()** function
  - However, the regular **pandas** function **str.contains()** also supports regular expressions

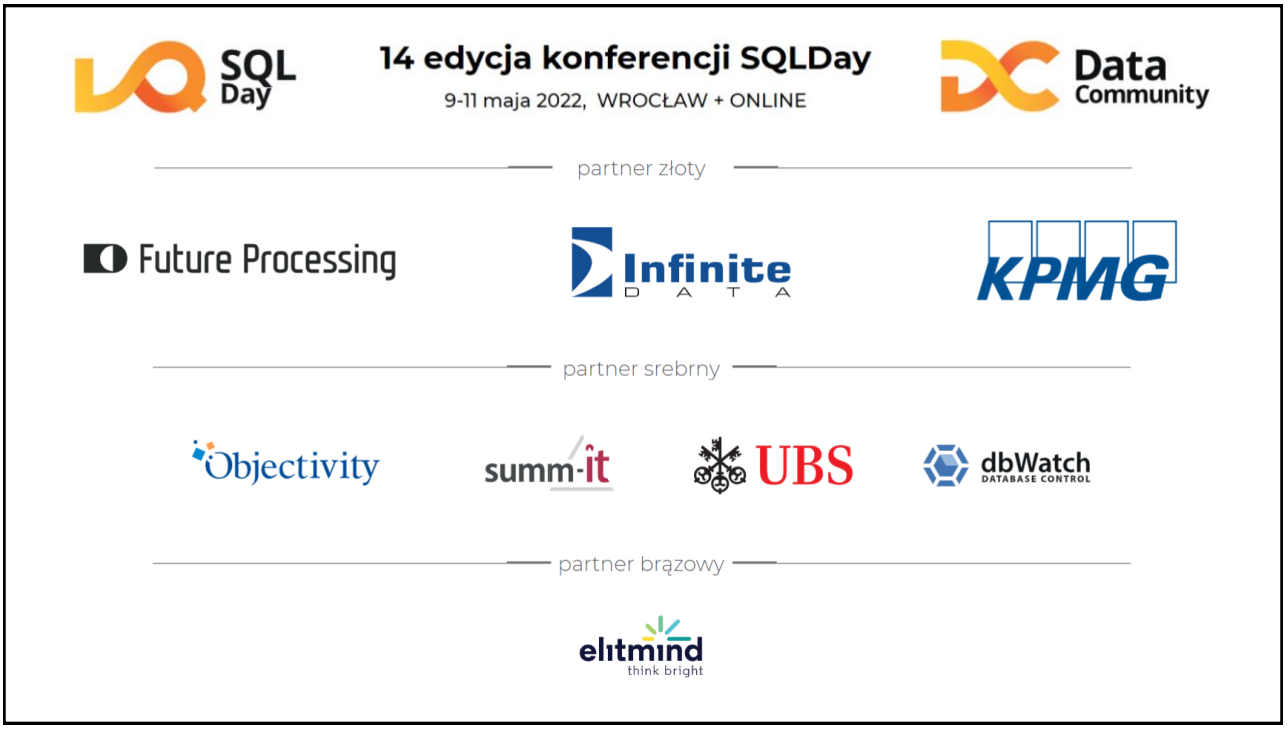
The Results

- Demo table with 100,000 rows, where 20,000 rows contain „Java“ in one nvarchar and one xml column

languageExpression	indexType	cYea	linesOfCode	timeMs
TSQL FTS CONTAINS	FTS table function	2000	3	110
TSQL LIKE	NCL covering scan	2000	3	453
TSQL LIKE	CL scan	2000	3	470
XQuery	XML Primary	2005	3	390
XQuery	CL scan	2005	3	593
CS	CL scan	2005	49	907
R data table	NCL covering scan	2016	20	1350
R rxDataStep()	NCL covering scan	2016	25	4737
Python	NCL covering scan	2017	17	953
Java	CL scan	2019	126	14646

Q & A

Questions?


SQL Day
14 edycja konferencji SQLDay
9-11 maja 2022, WROCŁAW + ONLINE
Data Community
partner złoty
Future Processing
Infinite DATA
KPMG
partner srebrny
Objectivity
summ-it
UBS
dbWatch DATABASE CONTROL
partner brązowy
elitmind think bright