
18
Win ays
Technology
Microsoft

Microsoft

SQL Server in the service of business optimization

Josip Šaban
Aldo Zelen

Agenda

- Why this lecture?
- What is NOT in this lecture?
- SQL Server and R
- Demo
- What we could have done more if we had more time?

Who is on stage today?

- Josip Šaban
  - FER Zagreb, Cotrugli MBA
  - MC* (SQL Server, Biztalk, Project Server, .NET), MCT od 2012 do 2014
  - CompTIA Project+, PRINCE2 Foundation, PMP, ITIL Foundation
  - Coursera - finished 9 courses and 3 specializations (very big fan :))
  - This is his 7th Windays lecture
  - Working in IT department in Erste Bank Croatia, also owns sole proprietorship Meridian Data
- Aldo Zelen
  - Team lead for DWH in Erste Bank Croatia, responsible for architecture and design of DWH in Croatia, Serbia and Montenegro
  - Ten years in IT, mainly on development of database and reporting system
  - Worked on relational and non-relational database systems, hosted both on-premise and in cloud
  - Huge fan of database intensive systems

Windays 18

Why this lecture?

Why this lecture?

- Industries applying advanced analytics
  - Retail and consumer products, financial services and industry, government, manufacturing, e-Commerce, healthcare, almost every on-line firm, …
- Analytic analysis is used to transform key business aspects
  - Transform your products
  - Empower your employees
  - Engage your customers
  - Optimize your operations - topic of today's lecture

Why this lecture?

- What is R and why do we use it?
  - We want to show integration of R with SQL Server
  - R is a language platform - a statistics programming language and a data visualization tool (open source) with huge community
- Why R on SQL Server and not stand-alone?
  - Leverage SQL Server resources - R on deskop has performance limitations, using SQL we gain more data analytic options
  - Data storage and analysis on the same platform - No need to install or buy something new
  - Usually huge security issue – moving data around removes security controls

What is not in this lecture?

What is not in this lecture?

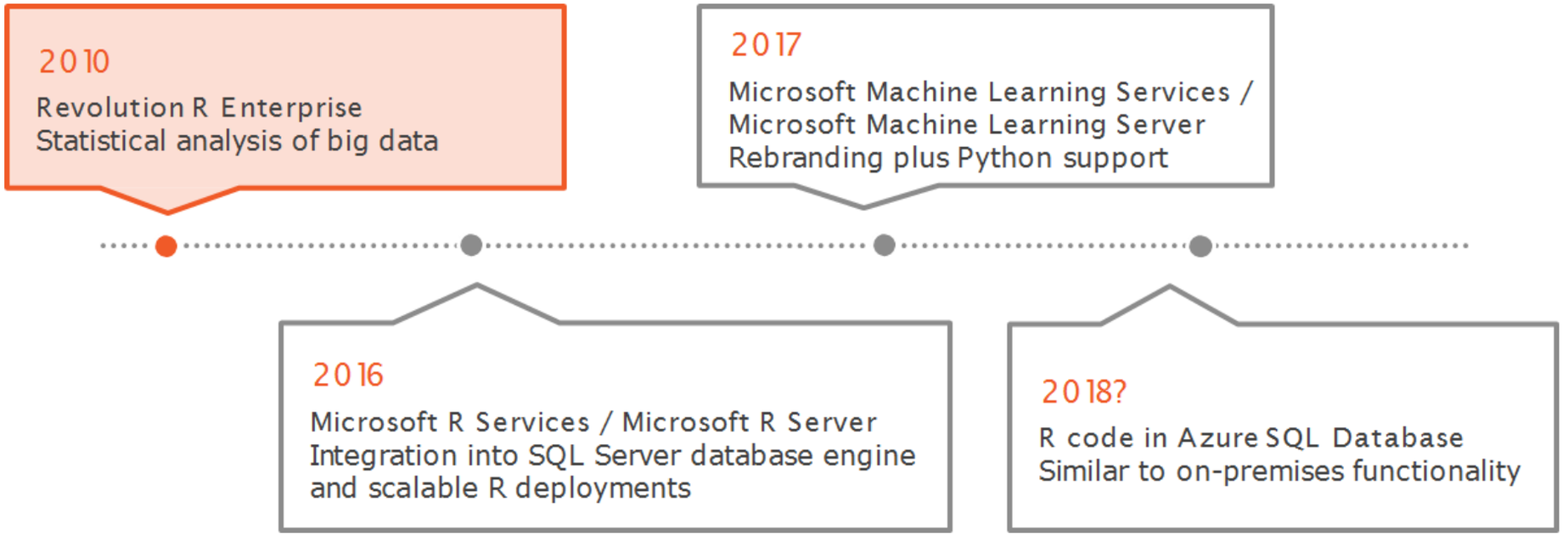
- This is NOT R, PowerBI or data statistics tutorial lecture
  - We will try to create a real-time demonstration how a SYSTEM using these technologies „could" work when we put all the black boxes together
- We will NOT discuss basic R topics which include:
  - Limited data scale issues, difficult modeling performance, lack of commercial support, complex deployment support
- We expect audience to understand…
  - Data analytics project lifecycle (Prepare – Model – Operationalize)
  - What is the role of data scientist, data engineer and business analyst
  - That creating a fancy PowerBI dashboard is the easiest part

SQL Server and R

SQL Server and R

- SQL Server and R – key points
  - Fully integrated in SQL Server installation package
    - Gives „enteprise" speed and scale, memory and disk scalability
    - Can use SQL Server parallel threading and processing, no R memory limits
  - End-to-end in SQL Server (Prepare – Model – Operationalize)
    - Run R inside SQL Server using IDE or embed R in T-SQL or run stored procedure
    - Reduce security exposure
    - Faster alghorithm execution due to parallelization
    - R on a server pulling data using SQL
  - Tools we use
    - R Tools for Visual Studio
    - RStudio – „everybody" in R world is using IT
    - „Real men" use Notepad++ and run it from command line ☺

SQL Server and R



- Common R use cases
  - Exploratory data analysis
  - Data visualization
  - Predictive modeling
- New in Machine Learning Server 2017
  - Added Python support
  - Expanding R capabilities

SQL Server and R

- Required Components - Developer
  - Integrated Development Environment (IDE) + Microsoft R client
- Required Components - Server
  - In Database - Standard Edition (limited performance), Enterprise Edition (full resource governance)
    - In Standard - data has to fit within memory + no Resource Governor capability
- Microsoft R Services (2016) / Microsoft Machine Learning Services (2017)
  - Standalone - Parallelism, Chunked processing of data (Microsoft R Server)
  - Microsoft R Server (2016)/Microsoft Machine Learning Server (2017)

SQL Server and R

- Unhandled SQL data types
  - Cursor, geography, geometry, hierarchyid, image, sql_variant, table, timestamp, uniqueid, xml, datetime2
  - Consult documentation how to convert (some) of these data types to R data types

Demo

Demo 1 - intro

- Required server configuration after default installation
  - Enabling external scripting
  - Granting permissions to run R scripts on database (if not “dbowner”)
  - OPTIONAL - Enabling implied authentication (optional step to support remote compute context), new group “SQLRUserGroup”
  - Option to enable users to run R scripts from their dev environment on SQL Server
  - Of course, remote connections have to also be enabled
  - Using key RevoScaleR function “RxSetComputeContect(sqlCompute)”
    - Rx - Data sources (RxSqlServerData) and Compute contexts functions (RxComputeContext)
    - rx - Get/Set data or medata (rxGetVarInfo) and Transformations functions (rxDataStep)

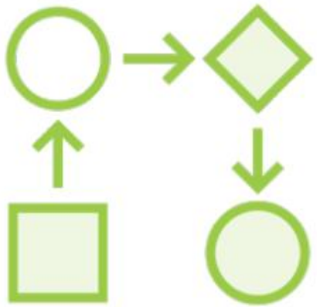
Demo 1 – data loading and data exploration

- We load data from CSV file and, using RevoScaleR function, load it to SQL Server
- After the load is complete, we use data exploration functions to get a feeling of what is inside this dataset
- Why do we show this demo?
  - To show how R and RevoScaleR functions work in Visual Studio
  - This is basis of all data analysis projects
  - It is a usual requirement in software development
    - Sometimes not so easy to do using „standard" SQL or OOP languages

Demo 2 – business optimization problem

- Basic case study description:
  - Every large company employs outside help (in form of consultants) that work on some project
    - Each consultant is avaliable at a specific moment in time because of other obligations
    - Work is divided by months
  - Optimal consultant usage, considering their cost and avaliability, is visualized in PowerBI, and calculated using R

Demo 2 – business optimization problem



- n people, 12 months
- Each person has list of months and price
- Each month requires m people
- Minimize total cost, given that…
  - Cost of person „i" working in month „j" is given by „$c_{i,j}$"

Demo 2 – business optimization problem

Assigment problem

Objective Function

Minimize cost

Constraints

Each month requires only certain number of resources

Decision Variables

Binary variables matching people to months

Demo 2 – business optimization demo

Objective Function

Cost of person „j" working in month „i" is given by „$c_{i,j}$"

$$X_{i,j} = \begin{cases} \text{1 if person j is working in month i} \\ \text{0 if not} \end{cases}$$

Total cost is sum-product of the decision variables and the costs

$$\sum_{i=1}^{12} \sum_{j=1}^{n} C_{i,j} X_{i,j}$$

Demo 2 – business optimization demo

Constraints

Each month requires a certain number of resources

$$\sum_{i=1}^{12} X_i = y_i$$

where y_i is required recourses in month „i”

$X_{i,j}$ are binary

Demo 2 – business optimization demo

Decision Variables

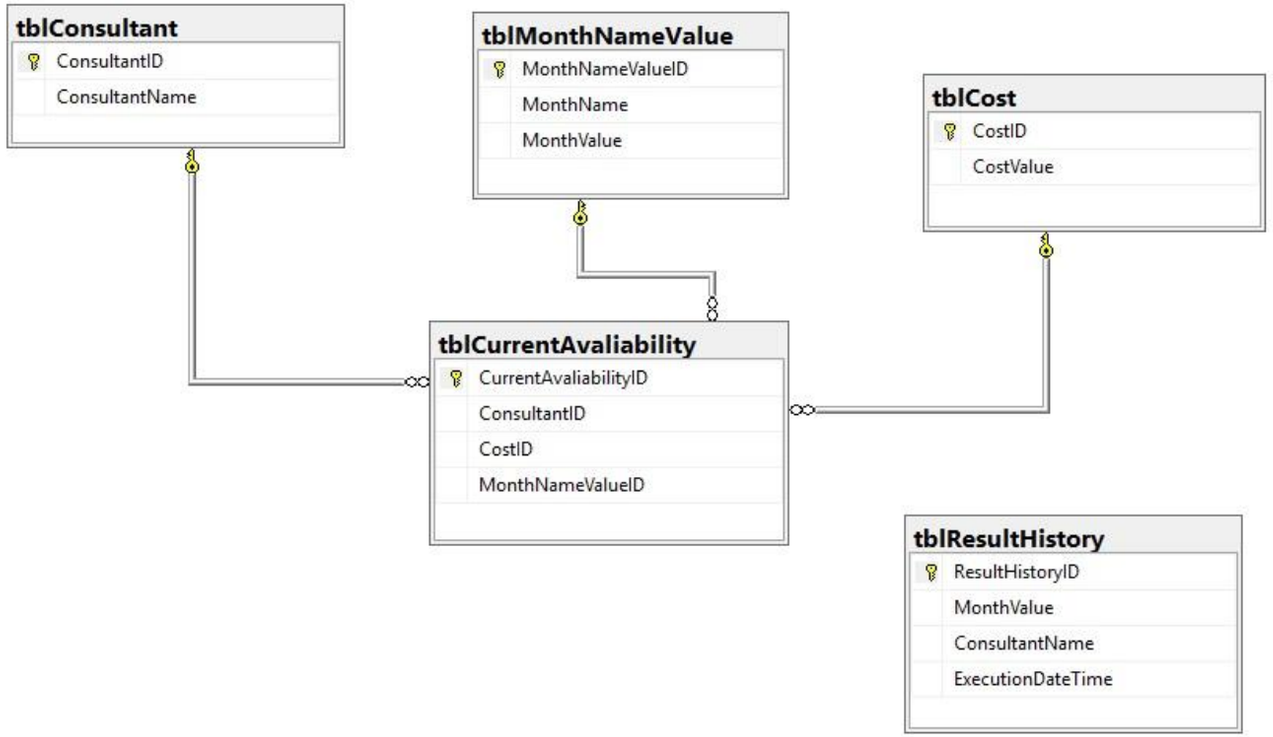
$$X_{i,j} = \begin{cases} \text{1 if pearson j is working in month i} \\ \text{0 if not} \end{cases}$$

Total cost is sum-product of the decision variables and the costs

Each „j" varies from 1 to n and „i" from 1 to 12
n*12 decision variables in total

Demo 2 – business optimization problem

- Database model



ConsultantName
Arya
Bran
Cersei
Daenerys
Jaime
Joh
Melisandre
Theon
Tyrion

CostValue
10
15
20
25
30
35

MonthName	MonthValue
January	1
February	2
March	3
April	4
May	5
June	6
July	7
August	8
September	9
October	10
November	11
December	12

Name	Cost	Availability
Arya	20	1
Arya	20	2
Arya	20	5
Bran	15	2
Bran	15	3
Bran	15	4
Bran	15	5
Cersei	35	3
Cersei	35	4
Daenerys	35	4
Daenerys	35	5
Jaime	20	2
Jaime	20	3

@jan = 1,@feb = 4, @mar = 3, @apr = 5,@may = 2, @jun = 0,
@jul = 0, @aug = 0, @sept = 0, @oct = 0, @nov = 0, @dec =0;

@jan = 1,@feb = 4, @mar = 3, @apr = 5,@may = 2, @jun = 0,
@jul = 1, @aug = 3, @sept = 2, @oct = 4, @nov = 1, @dec =0;

Windays 18

What we could have done more if we had more time?

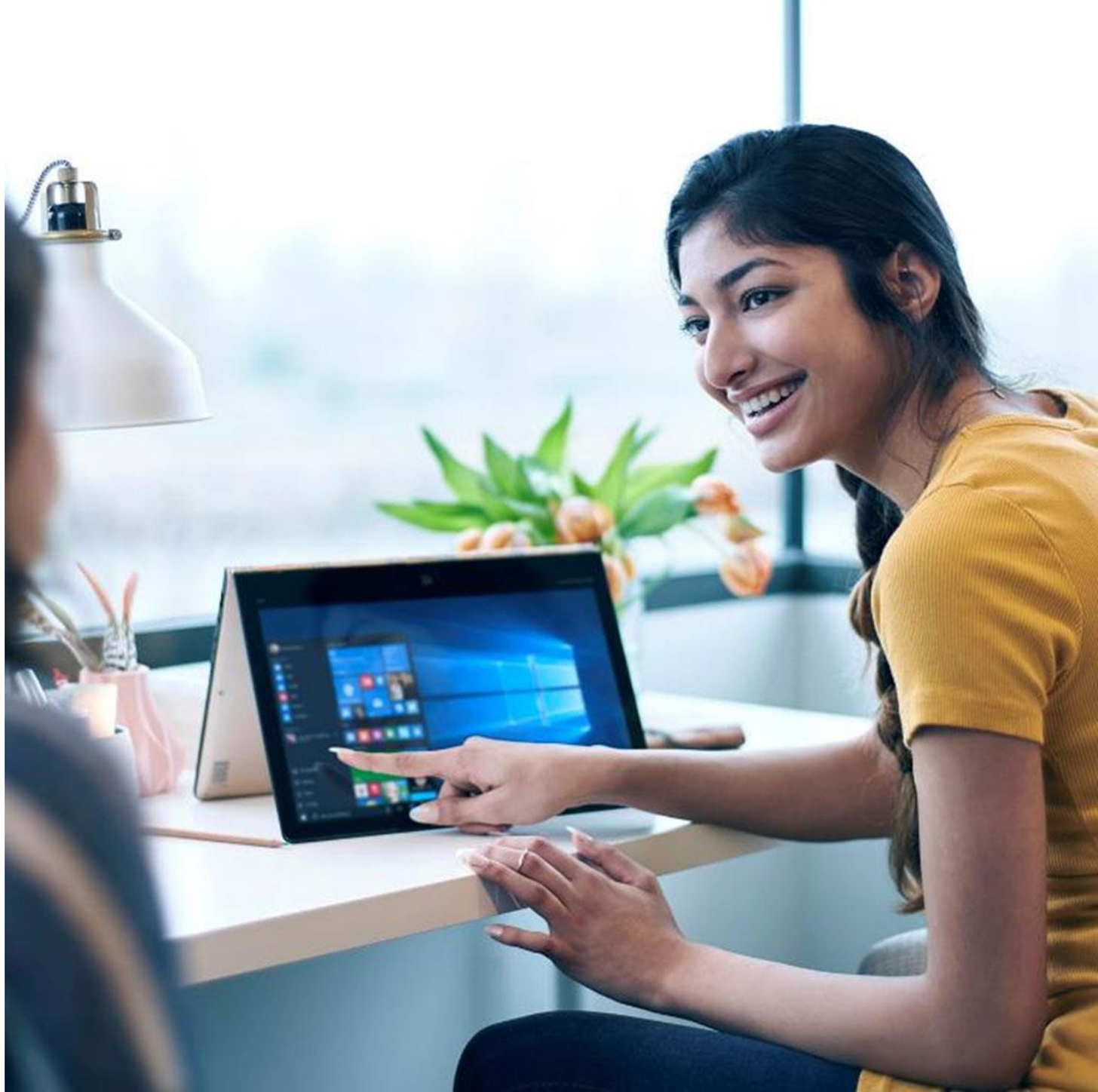
With more time, we could …

- Solve a more complex business problem
  - Add personal „quality" metrics
  - Add company inter-relationship
  - Add option for people to work with breaks in time periods
  - …
- Build predictive models and show how SQL Server handles model loading and predictive model processing
- Show how to integrate R directly into…
  - …Reporting Services
  - …SSIS (using either Process Task, R script or Data Flow task for scoring)
  - …PowerBI

Thank you ☺

https://hr.linkedin.com/in/josipsaban
jsaban@erstebank.com
meridiandatasoftware@gmail.com

https://www.linkedin.com/in/aldo-zelen-88646414/
azelen@erstebank.com
aldo.zelen@databreeze.co



Windays 18