

Introduktion til Microsoft R

Steen Dybboe, Pragmatic BI

MsBIP 2017

Aarhus

Agenda

Introduktion til R

Data Scientists

Microsoft R implementering

R + SQL

SQL + R



Hvem er jeg

Steen Dybboe

Selvstændig konsulent

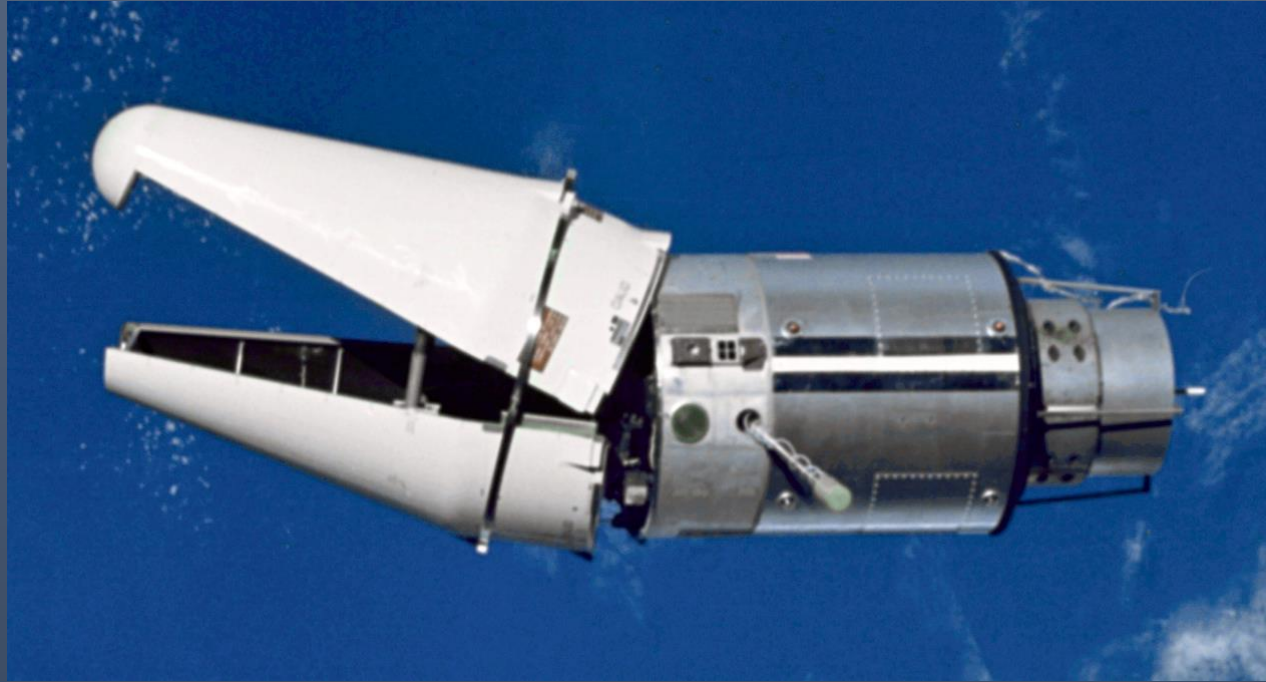
Fokus på Business Intelligence, databaser og kode

Har ingen hund, spiller lidt golf og synes "Pulp Fiction"
er en super sej film....



Pragmatic BI

Demo R



20.11.2016 KL. 06:30

I Mofibo er de fleste ansatte blevet dataanalytikere

Serieiværksætter Morten Strunge ønskede at kunne reagere lynsnart på ændringer i kundernes adfærd. Derfor ansatte han en analysechef og udstyrede medarbejderne i internetboghandlen Mofibo med **it-programmer** og **kompetencer** til selv at analysere data.

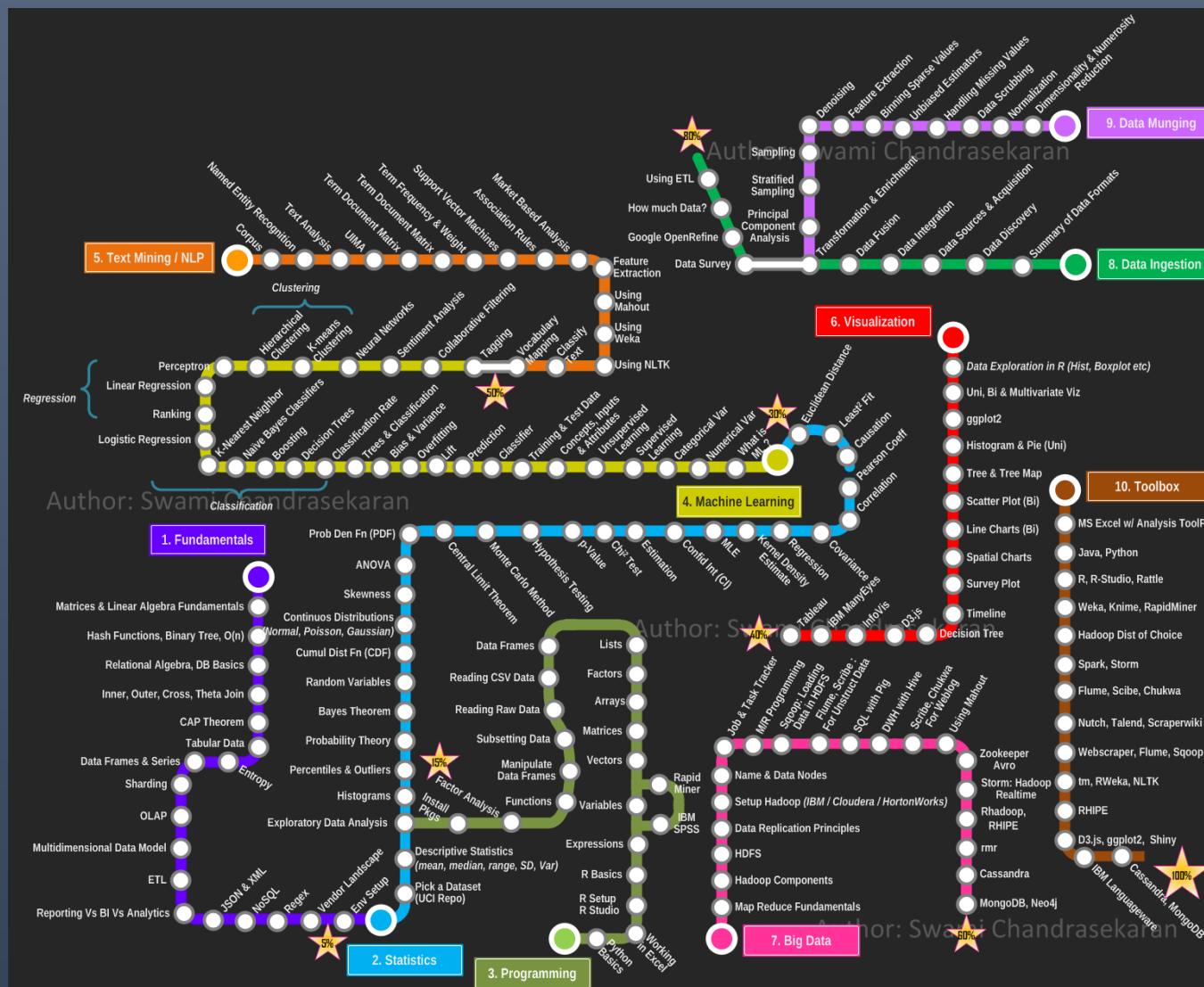


Martin Jonassen er analysechef i Mofibo og Storytel Group. Foto: Pr

Data Scientist



Data Scientist Overblik



Introduktion til R

Open Source (1995)

“Brugervenlig” dataanalyse, statistik og grafisk modellering

Værktøj for Statistikere og Data Scientists

Cutting edge

Interaktivt

Udfordringer

Open Source – mange implementeringer af samme funktionalitet

Skalerer dårligt (fri version)

Open Source ikke altid godkendt i firmaer



R overfor DBMS

SQL Server

R

Client/Server

Interaktivt, fortolket

Database objekter

Data strukturer

DML, DDL

Funktioner

Transact SQL

Code flow

Data Control Language

Mangler

Mangler (*CLR*)

Biblioteker

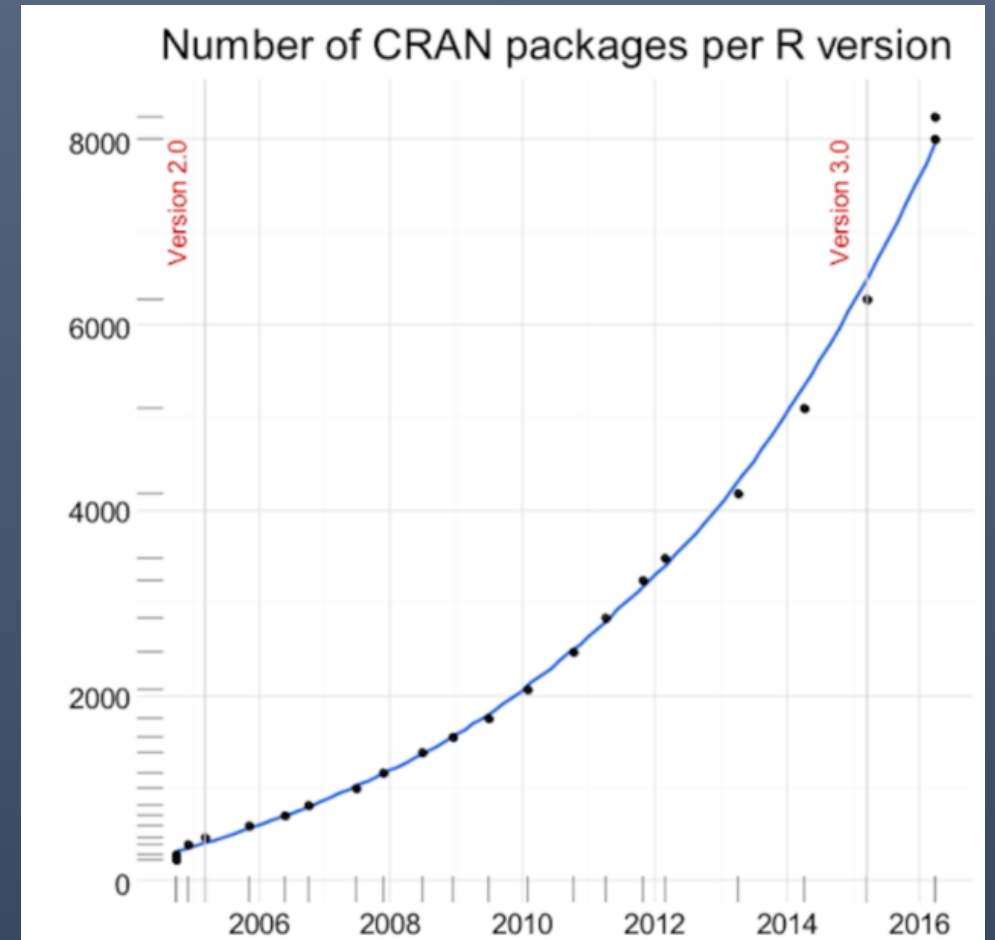
CRAN Pakker

Open Source

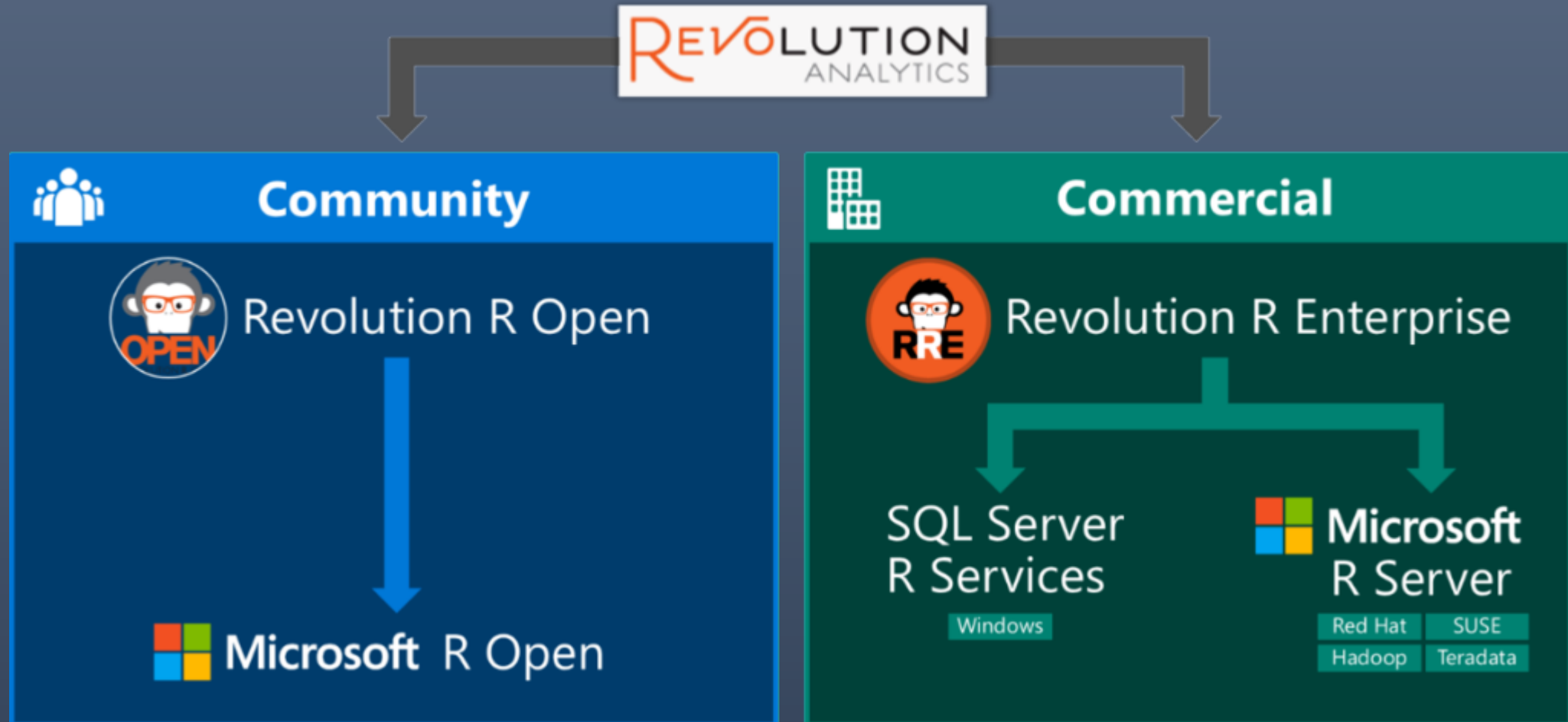
R kernen udvikles af en lukket gruppe

Kode biblioteker

Inline dokumenteret



Microsoft R



Microsoft R komponenter

CRAN-R

Open Source, In-memory

Microsoft R Open

Optimeret CRAN-R, 100 % kompatibel, fri

Microsoft R Server

Memory optimering, skalering, kontekst

SQL Server R Services

R Server på SQL Server

SQL Server R Services Install

SQL Server Installation Center

- Planning
- Installation**
- Maintenance
- Tools
- Resources
- Advanced
- Options

New SQL Server stand-alone installation or add features to an existing installation
Launch a wizard to install SQL Server 2016 in a non-clustered environment or to add features to an existing SQL Server 2016 instance.

Install SQL Server Management Tools
Launch a download page that provides a link to install SQL Server Management Studio, SQL Server command-line utilities (SQLCMD and BCP), SQL Server PowerShell provider, SQL Server Profiler and Database Tuning Advisor. An internet connection is required to install these tools.

Install SQL Server Data Tools
Launch a download page that provides a link to install SQL Server Data Tools (SSDT). SSDT provides Visual Studio integration including project system support for Azure SQL Database, the SQL Server Database Engine, Reporting Services, Analysis Services and Integration Services. An internet connection is required to install SSDT.

New SQL Server failover cluster installation
Launch a wizard to install a single-node SQL Server 2016 failover cluster.

Add node to a SQL Server failover cluster
Launch a wizard to add a node to an existing SQL Server 2016 failover cluster.

Upgrade from a previous version of SQL Server
Launch a wizard to upgrade a previous version of SQL Server to SQL Server 2016.

New R Server (Standalone) installation
Launch a wizard to install R Server (Standalone) on a Windows machine. This is typically used by data scientists as a standalone analysis server or as a SQL Server R Services client.

Microsoft SQL Server 2016

SQL Server 2016 Setup

Feature Selection

Select the Enterprise features to install.

Global Rules

Product Updates

Install Setup Files

Install Rules

Installation Type

Product Key

License Terms

Feature Selection

Feature Rules

Instance Configuration

Server Configuration

Database Engine Configuration

Consent to install Microsoft R ...

Feature Configuration Rules

Ready to Install

Installation Progress

Complete

Features:

- Database Engine Services
- SQL Server Replication
- R Services (In-Database)
- Full-Text and Semantic Extractions for Search
- Data Quality Services
- PolyBase Query Service for External Data
- Analysis Services
- Reporting Services - Native

Shared Features

- R Server (Standalone)
- Reporting Services - SharePoint
- Reporting Services Add-in for SharePoint Prod
- Data Quality Client

Feature description:

The configuration and operation of each instance feature of a SQL Server instance is isolated from other SQL Server instances. SQL Server instances can operate side-by-side on

Prerequisites for selected features:

Already installed:

- Windows PowerShell 3.0 or higher
- Microsoft Visual Studio 2010 Redistributable

Disk Space Requirements

Drive C: 2320 MB required, 105401 MB available

Select All Unselect All

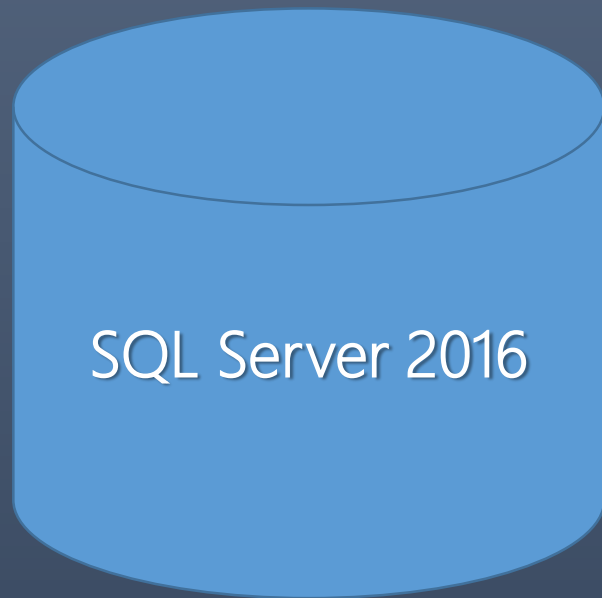
Instance root directory: C:\Program Files\Microsoft SQL Server\

Shared feature directory: C:\Program Files\Microsoft SQL Server\

Shared feature directory (x86): C:\Program Files (x86)\Microsoft SQL Server\

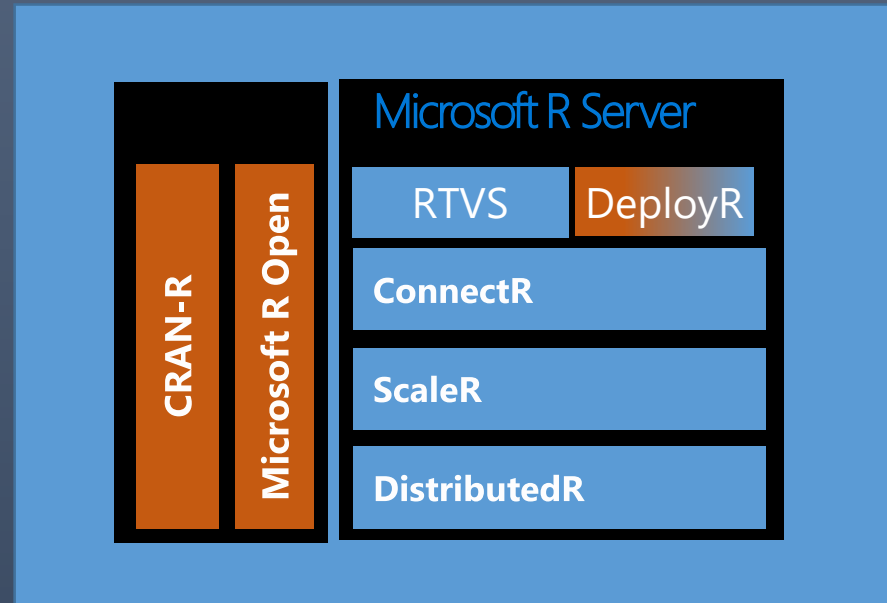
< Back Next > Cancel

SQL Server R Services Communication



BXL Server
SQL Satellite

SQL Server R Services



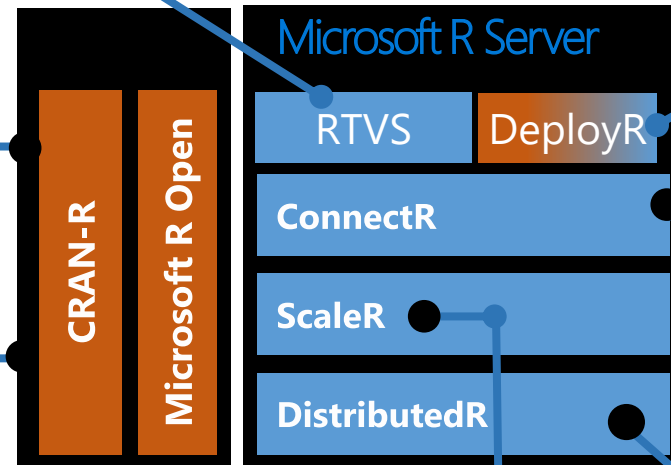
Microsoft R Server

CRAN-R

- Open source R interpreter
- Freely-available huge range of R algorithms
- Algorithms callable by RevoR
- Embeddable in R scripts
- 100% Compatible with existing R scripts, functions and packages

R Tools for Visual Studio

- State of the art, R Tools for Visual Studio IDE



ScaleR

- Ready-to-Use high-performance big data big analytics
- Fully-parallelized analytics
- Descriptive statistics & statistical tests
- Range of predictive functions
- User tools for distributing customized R algorithms across nodes
- Wide data sets supported – thousands of variables

DeployR

- RESTful APIs for easy integration from Java, JavaScript, .NET
- Enterprise authentication & security
- Horizontal scaling

ConnectR

- High-speed & direct connectors

Available for:

- High-performance XDF
- SAS, SPSS, delimited & fixed format text data files
- Hadoop HDFS (text & XDF)
- Teradata Database & Aster
- EDWs and ADWs
- ODBC

DistributedR

- Distributed computing framework
- Delivers cross-platform portability

Microsoft R Open

- Based on open source R
- High-performance math library to speed up linear algebra functions
- Checkpoint package to easily share R code and replicate results using specific R package versions

rproject1 - Microsoft Visual Studio

File Edit View Project Build Debug Team Tools Architecture Test R Tools Analyze Window Help

Debug Any CPU Source startup file

demolrisComplete.R* x demolris.R script.R

```
rm(list = ls())

library(e1071);
data(iris)
irismodel <- naiveBayes(iris[, 1:4], iris[, 5])

pred <- predict(irismodel, iris)

data.frame(iris, pred)

rm(test)
test <- data.frame(Sepal.Length = 5.9, Sepal.Width = 3.0, Petal.Length = 5.1, Petal.Width = 0.2)

pred <- predict(irismodel, test)
data.frame(test, pred)

#Hvordan performer modellen
table(predict(irismodel, iris[, 1:4]), iris[, 5])
```

100 %

R Interactive

Attach Debugger C:/Users/Steen/Dropbox/Undervisning

```
+ 
+ irismodel <- naiveBayes(iris[, 1:4], iris[, 5])
+ 
+ pred <- predict(irismodel, iris)
+ 
+ data.frame(iris, pred)
+ 
+ Sepal.Length Sepal.Width Petal.Length Petal.Width Species pred
+ 1 5.1 3.5 1.4 0.2 setosa setosa
+ 2 4.9 3.0 1.4 0.2 setosa setosa
+ 3 4.7 3.2 1.3 0.2 setosa setosa
+ 4 4.6 3.1 1.5 0.2 setosa setosa
+ 5 5.0 3.6 1.4 0.2 setosa setosa
+ 6 5.4 3.9 1.7 0.4 setosa setosa
+ 7 4.6 3.4 1.4 0.3 setosa setosa
+ 8 5.0 3.4 1.5 0.2 setosa setosa
+ 9 4.4 2.9 1.4 0.2 setosa setosa
+ 10 4.9 3.1 1.5 0.1 setosa setosa
+ 11 5.4 3.7 1.5 0.2 setosa setosa
+ 12 4.8 3.4 1.6 0.2 setosa setosa
+ 13 4.8 3.0 1.4 0.1 setosa setosa
+ 14 4.3 3.0 1.1 0.1 setosa setosa
```

100 %

Data Tools Operations Error List Output

Solution Explorer

Search Solution Explorer (Ctrl+)

Solution 'rproject1' (1 project)

- rproject1
 - .Rhistory
 - rproject1.rproj
 - script.R

Tabular Model Explorer Solution Explorer Team Explorer Class View

R Plot - Device 3 - Active

Histogram of iris\$Petal.Length

iris\$Petal.Length	Frequency
1.0 - 1.5	32
1.5 - 2.0	13
2.0 - 2.5	1
2.5 - 3.0	4
3.0 - 3.5	11
3.5 - 4.0	21
4.0 - 4.5	21
4.5 - 5.0	17
5.0 - 5.5	16
5.5 - 6.0	5
6.0 - 6.5	4

R Plot R Help Properties

Ready Ln 184 Col 3 Ch 3 INS ↑ Publish

R + SQL

Klassik data analyse fra R

Klassisk

Hente data I SQL Server og anvende dem
RODBC

```
install.packages("RODBC")  
library(RODBC)
```

```
cn <- odbcDriverConnect(connection="Driver={SQL Server Native Client 11.0};  
                        server=localhost;database=PegasusPark;trusted_connection=yes;")  
  
dataFetchCity <- sqlFetch(cn, 'vwCityList', colnames=FALSE, rows_at_time=1000)  
  
View(dataFetchCity)
```

Demo

Database adgang fra R

SQL + R

Udnyt styrken fra R i SQL

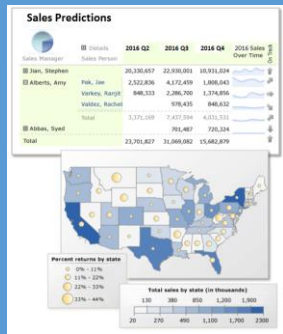
sp_execute_external_script

Kalder R fra SQL

Flytter data frem, og tilbage

```
EXEC sp_execute_external_script
  @language = N'R'
, @input_data_1 = N'SELECT 1 as Installed'
, @script = N'OutputDataSet <- data.frame(path = .libPaths())'
WITH RESULT SETS (([path] VARCHAR(250) NOT NULL));
```

Advanced Analytics – Med T-SQL



Application

1 Call Stored Procedure

3 Results: Data, plots



Advanced Analytics Extensions

Microsoft R Open

Microsoft R Server

2 The stored procedure contains R code and executes in-database

```
exec sp_execute_external_script  
  @ language = 'R'  
  , @script =  
  -- R code --
```



Iris Virginica



Iris Setosa



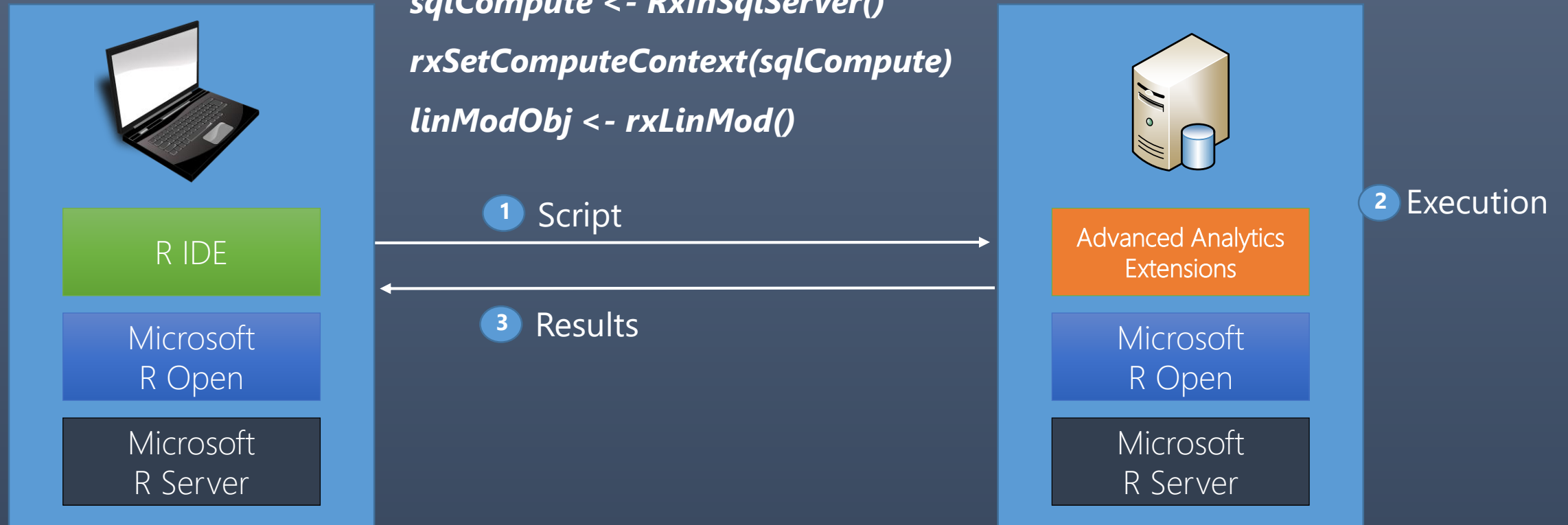
Iris Versicolor

Demo

Introduktion af Iris

Anvendelse af R fra SQL

Advanced Analytics – R Client



Opsamling

Microsoft er gået "All-in" på R

Spændende muligheder

Forventninger

Næste skridt

Installer R klient og prøv R
Afprøv Microsoft R Services

Kursus "Introduktion til Microsoft R"
Marts 2017

ORANGEMAN



Pragmatic BI