

Austausch von Daten zwischen dem MS SQL Server und Hadoop

Andreas Jordan
PASS Rhein-Main 10.09.2019



- Installation der SQL Server Instanz
- Konfiguration der SQL Server Instanz
- Konfiguration der SQL Server Datenbank
- Demonstration

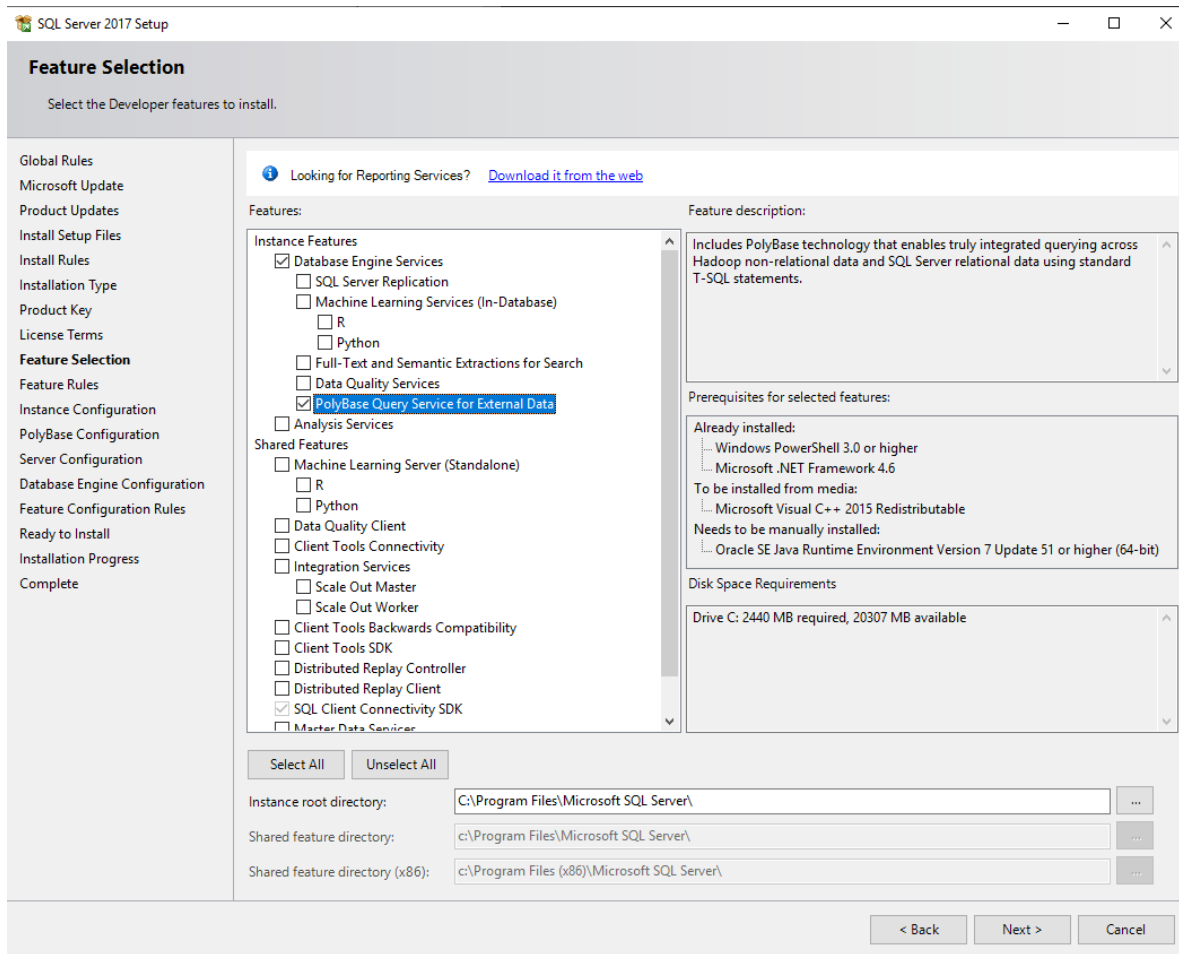
Installation der SQL Server Instanz

- Voraussetzung
- Installation per GUI
- Installation per Kommandozeile

Voraussetzungen

- Java (jre-8u221-windows-x64.exe)

Installation per GUI



Installation per GUI

The screenshot displays a software installation window with two main panes. The left pane, titled 'Features:', is divided into 'Instance Features' and 'Shared Features'. Under 'Instance Features', several options are listed with checkboxes: 'Database Engine Services' (checked), 'SQL Server Replication' (unchecked), 'Machine Learning Services (In-Database)' (unchecked), which includes sub-options for 'R' and 'Python' (both unchecked), 'Full-Text and Semantic Extractions for Search' (unchecked), 'Data Quality Services' (unchecked), 'PolyBase Query Service for External Data' (checked and highlighted with a blue dashed border), and 'Analysis Services' (unchecked). Under 'Shared Features', there are options for 'Machine Learning Server (Standalone)' (unchecked), 'Data Quality Client' (unchecked), 'Client Tools Connectivity' (unchecked), and 'Integration Services' (unchecked). The right pane, titled 'Feature description:', contains a text box describing the selected feature: 'Includes PolyBase technology that enables truly integrated querying across Hadoop non-relational data and SQL Server relational data using standard T-SQL statements.' Below this, a section titled 'Prerequisites for selected features:' lists requirements categorized into 'Already installed:' (Windows PowerShell 3.0 or higher, Microsoft .NET Framework 4.6) and 'Needs to be manually installed:' (Microsoft Visual C++ 2015 Redistributable, Oracle SE Java Runtime Environment Version 7 Update 51 or higher (64-bit)).

Installation per GUI

The screenshot displays the 'Features' selection window in SQL Server Enterprise Manager. It is divided into two main sections: 'Instance Features' and 'Shared Features'. Under 'Instance Features', the 'PolyBase Query Service for External Data' checkbox is checked and highlighted with a blue background. Other options like 'Database Engine Services', 'SQL Server Replication', and 'Machine Learning Services' are also visible. The 'Feature description' pane on the right provides details for the selected feature, including its description and a list of prerequisites. The prerequisites are categorized into 'Already installed' (Windows PowerShell 3.0 or higher, Oracle SE Java Runtime Environment, Microsoft .NET Framework 4.6) and 'To be installed from media' (Microsoft Visual C++ 2015 Redistributable).

Features:	Feature description:
<p>Instance Features</p> <ul style="list-style-type: none"><input checked="" type="checkbox"/> Database Engine Services<ul style="list-style-type: none"><input type="checkbox"/> SQL Server Replication<input type="checkbox"/> Machine Learning Services (In-Database)<ul style="list-style-type: none"><input type="checkbox"/> R<input type="checkbox"/> Python<input type="checkbox"/> Full-Text and Semantic Extractions for Search<input type="checkbox"/> Data Quality Services<input checked="" type="checkbox"/> PolyBase Query Service for External Data<input type="checkbox"/> Analysis Services <p>Shared Features</p> <ul style="list-style-type: none"><input type="checkbox"/> Machine Learning Server (Standalone)<ul style="list-style-type: none"><input type="checkbox"/> R<input type="checkbox"/> Python<input type="checkbox"/> Data Quality Client	<p>Includes PolyBase technology that enables truly integrated querying across both non-relational data and SQL Server relational data using standard T-SQL statements.</p> <p>Prerequisites for selected features:</p> <p>Already installed:</p> <ul style="list-style-type: none">... Windows PowerShell 3.0 or higher... Oracle SE Java Runtime Environment Version 7 Update 51 or higher (64-bit)... Microsoft .NET Framework 4.6 <p>To be installed from media:</p> <ul style="list-style-type: none">... Microsoft Visual C++ 2015 Redistributable

Installation per Kommandozeile

- Wichtige Zeilen der ConfigurationFile.ini:
 - FEATURES=SQLEngine,PolyBase
 - TCPENABLED=1

Agenda

- Installation der SQL Server Instanz
- Konfiguration der SQL Server Instanz
- Konfiguration der SQL Server Datenbank
- Demonstration

Konfiguration der SQL Server Instanz

- `sp_configure @configname = 'hadoop connectivity', @configvalue = 6;`
- `sp_configure @configname = 'allow polybase export', @configvalue = 1;`
- Dienste neu starten, dabei auch an die PolyBase-Dienste denken.
- Links:
<https://docs.microsoft.com/de-de/sql/relational-databases/polybase/polybase-configure-hadoop>
<https://docs.microsoft.com/de-de/sql/database-engine/configure-windows/polybase-connectivity-configuration-transact-sql>

Konfiguration der SQL Server Instanz

```
$Instance = Connect-DbInstance -SqlInstance "SQL2\POLYBASE"
```

```
Set-DbaspConfigure -SqlInstance $Instance -Name 'HadoopConnectivity' -Value 6
```

```
Set-DbaspConfigure -SqlInstance $Instance -Name 'AllowPolybaseExport' -Value 1
```

```
Get-Dbaservice | ? ServiceType -EQ PolyBase | Stop-Dbaservice
```

```
Stop-Dbaservice -InstanceName 'POLYBASE'
```

```
Start-Dbaservice -InstanceName 'POLYBASE'
```

```
Get-Dbaservice | ? ServiceType -EQ PolyBase | Start-Dbaservice
```

Agenda

- Installation der SQL Server Instanz
- Konfiguration der SQL Server Instanz
- Konfiguration der SQL Server Datenbank
- Demonstration

Konfiguration der SQL Server Datenbank

- CREATE MASTER KEY
- CREATE DATABASE SCOPED CREDENTIAL
- CREATE EXTERNAL DATA SOURCE
- CREATE EXTERNAL FILE FORMAT
- CREATE EXTERNAL TABLE

CREATE MASTER KEY

```
CREATE MASTER KEY ENCRYPTION BY PASSWORD = 'P@ssw0rd';
```

CREATE DATABASE SCOPED CREDENTIAL

```
CREATE DATABASE SCOPED CREDENTIAL HadoopUser  
WITH IDENTITY = 'ordix'  
    , SECRET = 'ordix';
```

CREATE EXTERNAL DATA SOURCE

```
CREATE EXTERNAL DATA SOURCE HadoopCluster
WITH ( TYPE = HADOOP
      , LOCATION = 'hdfs://192.168.130.134:8022'
      , RESOURCE_MANAGER_LOCATION = '192.168.130.134:8032'
      , CREDENTIAL = HadoopUser
      );
```


CREATE EXTERNAL FILE FORMAT

```
CREATE EXTERNAL FILE FORMAT ParquetFileFormat
WITH ( FORMAT_TYPE = PARQUET
      , DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
      );
```

CREATE EXTERNAL TABLE

```
CREATE EXTERNAL TABLE Hadoop.Posts
( Id          INT
, CreationDate DATETIME
, OwnerUserId INT
, ParentId    INT
...
)
WITH ( LOCATION='/user/pdw_user/posts/'
      , DATA_SOURCE = HadoopCluster
      , FILE_FORMAT = ParquetFileFormat
) ;
```

Herausforderung: Nicht alle Datentypen sind erlaubt

```
CREATE EXTERNAL TABLE Hadoop.Posts
( Id          INT
, CreationDate DATETIME
, OwnerUserId INT
, ParentId    INT
...
, Body        NVARCHAR(MAX)
)
WITH ( LOCATION='/user/pdw_user/posts/'
      , DATA_SOURCE = HadoopCluster
      , FILE_FORMAT = ParquetFileFormat
      );
```

Meldung 46518, Ebene 16, Status 12, Zeile 1

The type 'nvarchar(max)' is not supported with external tables.

**Vielen Dank für
Ihre Aufmerksamkeit**

ORDIX[®] best practice
einfach. gut. beraten.

ORDIX AG
Aktiengesellschaft für
Softwareentwicklung, Schulung,
Beratung
und Systemintegration

Zentrale Paderborn
Karl-Schurz-Straße 19a
33100 Paderborn
Tel.: 05251 1063-0
Fax: 0180 1 67349 0

Seminarzentrum Wiesbaden
Kreuzberger Ring 13
65205 Wiesbaden
Tel.: 0611 77840-00

info@ordix.de
www.ordix.de