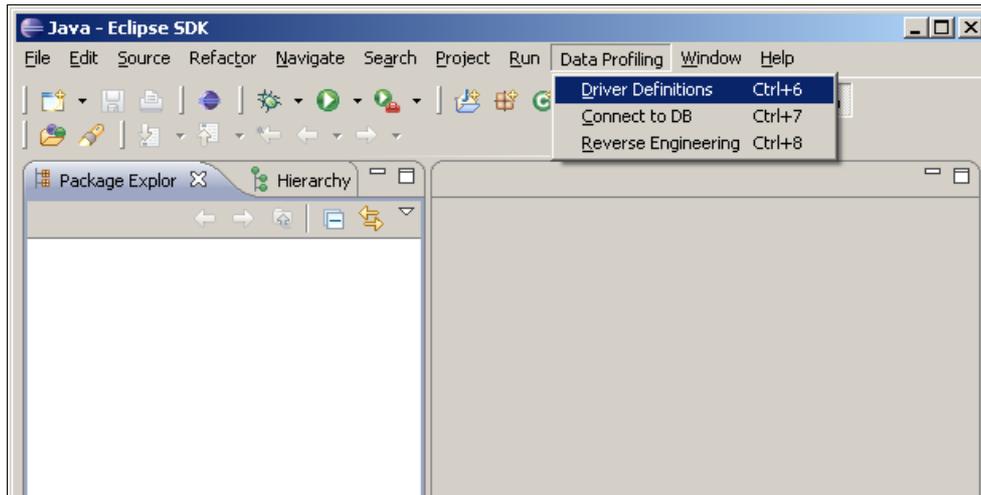


Data-Profiling Anwendungsleitfaden (Erste Schritte)

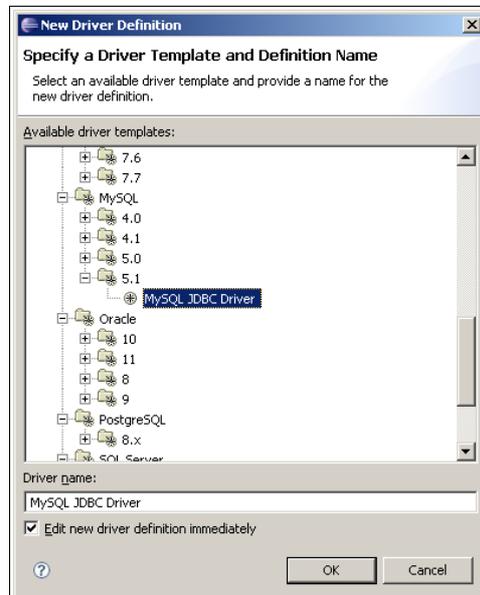
Zum **Installieren** des Data-Profiling-Werkzeugs bitte **install.txt** lesen!

Dieses Dokument soll einen kurzen Überblick über die prototypische Data-Profiling-Anwendung geben. Dafür wird ein typisches Anwendungsszenario von der Datenbanktreiber-Installation bis zur spezifischen Data-Profiling-Analyse dargestellt.

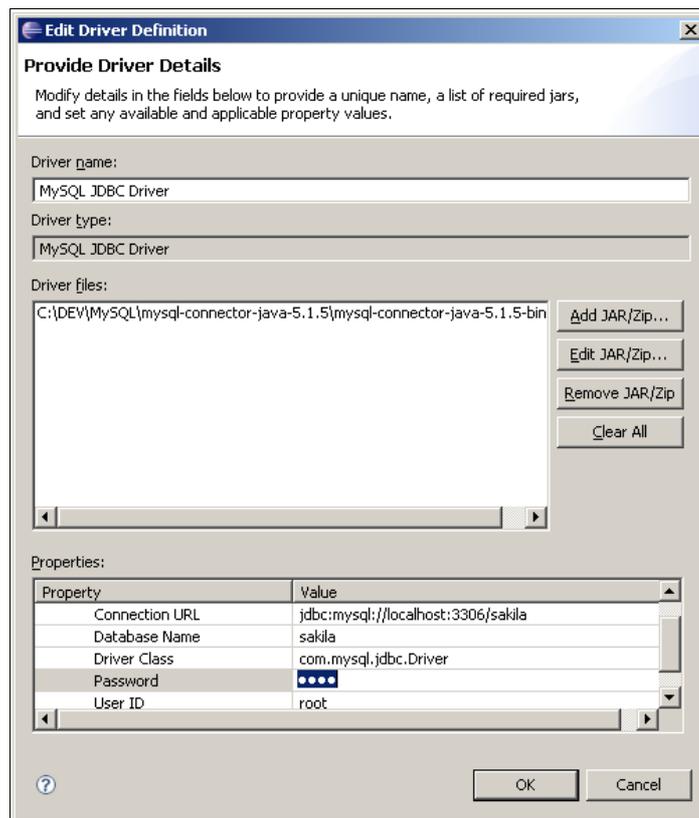
- Wenn Eclipse mit allen erforderlichen Plugins erfolgreich geladen wurde, sollte im Hauptmenü ein **'Data Profiling'-Menü** erscheinen.



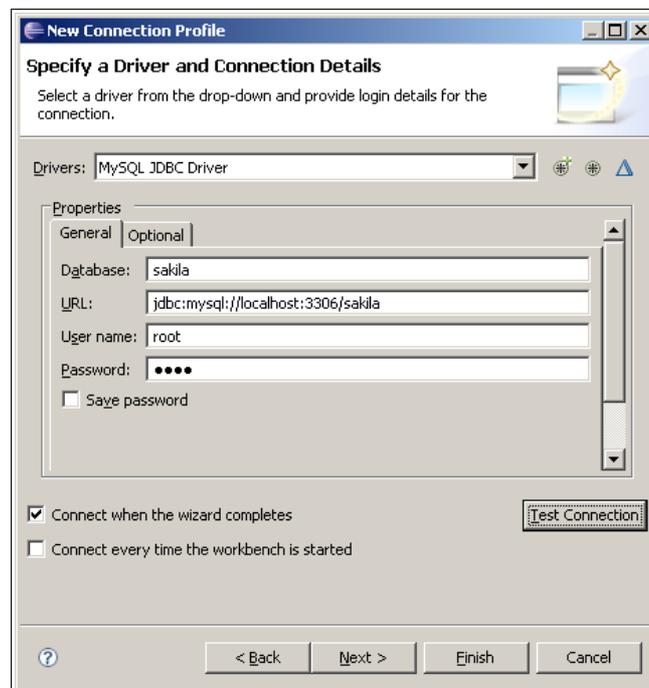
- '*Data Profiling* → *Driver Definitions* → *Add...*' ausführen, um einen neuen **Datenbanktreiber** zu definieren. Im Beispiel wird ein **'MySQL JDBC Driver'** ausgewählt.



- Anschließend müssen die entsprechenden **Treibereigenschaften**, wie Datenbankname, User und Passwort angegeben werden. Die Treiberdatei muss die entsprechende JDBC-jar-Datei des Datenbankherstellers referenzieren.

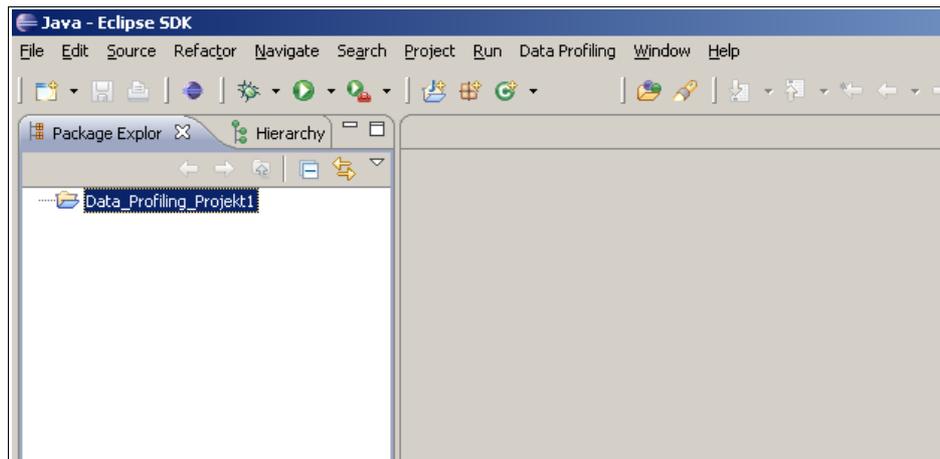


- Im nächsten Schritt muss ein **Verbindungsprofil** angelegt und aktiviert werden. 'Data Profiling → Connect to DB → New...' ausführen. Im folgenden Fenster den entsprechenden Treibertyp doppelt anklicken.

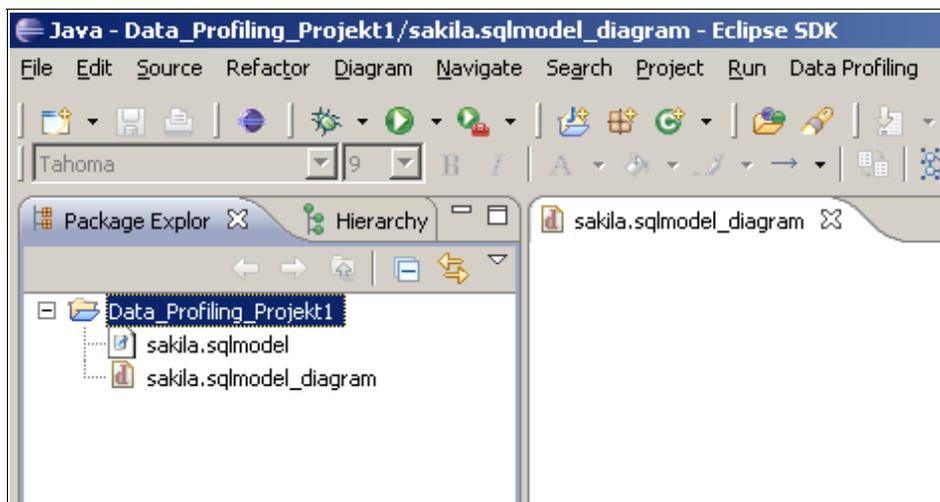


Das zuvor angelegte Treiberprofil wird standardmäßig angezeigt und sollte auch einen Verbindungstest erfolgreich bestehen. **Passwort speichern** und mit **Finish** abschließen.

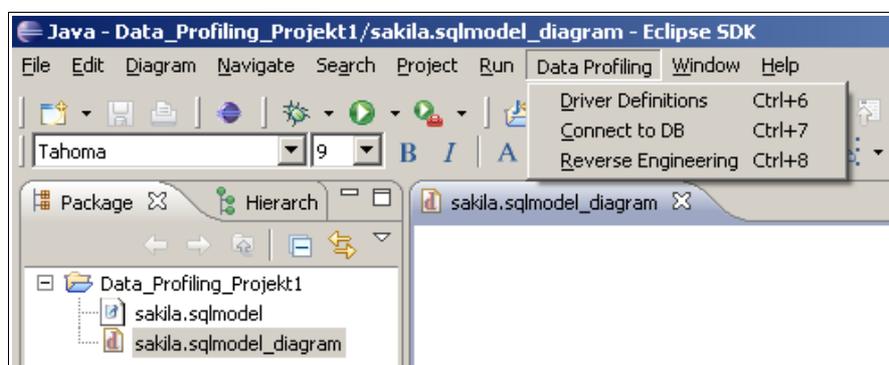
- Im Nächsten Schritt wird ein **neues Data-Profiling-Projekt** angelegt. Über '*File → New → Project...*' eine allgemeines Projekt mit einem beliebigen Namen anlegen.



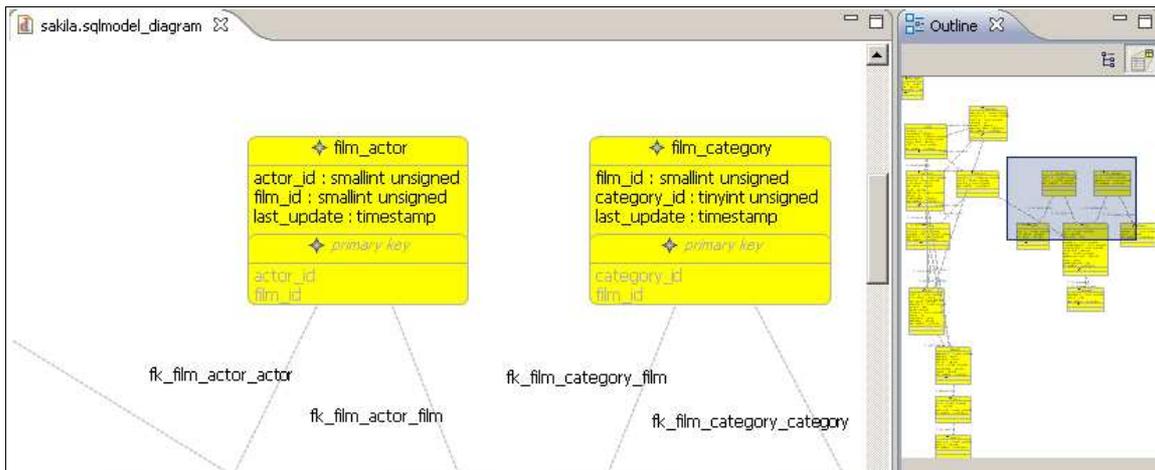
- Danach über '*File → New → Example... → Sqlmodel Diagram*' die **Diagramm-Dateien erzeugen**. Beide Dateien müssen den **selben Dateinamen** haben und die vorgeschlagene Dateiendung beibehalten.



- Um nun das **Reverse Engineering** durchzuführen, muss die Datei mit der Endung ***.sqlmodel_diagram** im Editorfenster **geladen** sein (s.oben). Der eigentliche Aufruf erfolgt dann über '*Data Profiling → Reverse Engineering*'.



- Nachdem das Reverse Engineering erfolgreich durchgelaufen ist, erscheint das **grafische Datenmodell** im Editorfenster.



- Die **Table Profile View** kann über 'Window → Show View → Other...' und die Auswahl von 'Data Profiling/Table Profile View' eingeblendet werden. Wird nun ein Tabellenelement ausgewählt, zeigt das eingeblendete Fenster die entsprechenden Data-Profiling-Kennzahlen an.

The screenshot shows the 'Table Profile View' window for the 'actor' table. The table lists various data profiling metrics for each column.

Column name	Total	Distinct	% Distinct	Null	% Null	Min Value	Max Value
actor_id	200	200	100,0	0	0,0	1	200
first_name	200	128	64,0	0	0,0	ADAM	ZERO
last_name	200	121	60,5	0	0,0	AKROYD	ZELLWEGER
last_update	200	1	0,5	0	0,0	2006-02-15 04:34:33.0	2006-02-15 04:34

- Die **Mustermanalyse** wird über das Kontextmenü einer Datenbankspalte angestoßen:

The screenshot shows a context menu for a column in a table. The 'Pattern Analysis' option is selected. The resulting 'Pattern Analysis' dialog box displays a table of patterns and their counts.

Pattern	Count	%
AAAAAAA	160	26,7
AAAAAA	114	19,0
AAAAAAA	76	12,7
AAAA	70	11,7
AAAAAAA	32	5,3
AAA	17	2,8
AAAAAAA	9	1,5
AAAAAAA...	6	1,0
AA	1	0,2