

Bekannte Probleme bei der Umstellung auf Oracle 12c R2 (UTF-8)

Inhalt

1	Einleitung	2
1.1	Technischer Hintergrund Oracle VARCHAR2	2
2	Das Test- und Korrektur-Paket	3
3	Korrekturen in der Anwendung	6
3.1	Textbausteine bis 1.24.13.29417 nicht lesbar	6

1 Einleitung

Dieses Dokument ist ein Nachtrag zum Release 1.24.13, das aufgrund von Rückmeldungen aus Bayern und Baden-Württemberg die aktuellen Probleme bei der Umstellung der Anwendung auf UTF-8 (Oracle Character set AL32UTF8) erläutert und Hilfen zur Behebung dieser Probleme bietet.

1.1 Technischer Hintergrund Oracle VARCHAR2

Grundsätzlich gibt es bei Oracle das Problem, dass Textfelder vom Typ „Varchar2“ im Modus „Anzahl Zeichen in BYTE“ und „Anzahl Zeichen in CHAR“ angelegt werden können. Bis Release 1.24.5 wurden alle von BALVI angelegten Tabellen im Modus „VARCHAR2([Anzahl])“ ohne die Angabe BYTE oder CHAR angelegt, um Oracle die Entscheidung selbstständig treffen zu lassen. Der Oracle-Standard ist BYTE, was bei ISO-Zeichensätzen kein Problem darstellt, da ein Zeichen auch immer mit einem BYTE abgebildet und gespeichert werden kann.

Mit der Umstellung auf „Multi-Byte-Zeichensätze“ wie UTF-8 ändert sich jedoch der Bedarf an Platz in Bytes für diverse Zeichen, z.B. die erweiterten Lateinischen Zeichen für die deutsche Sprache wie Umlaute, „ß“, „€“ und „Office-Spezifische Sonderzeichen“, z.B. der Aufzählungspunkt benötigen nun mehr als ein Byte.

Wenn man also davon ausgeht, ein Textfeld im Modus VARCHAR2(4000 CHAR) anzulegen, sollte nach dem technischen Verständnis egal sein, wie viele BYTE wirklich benötigt werden. Leider ist das bei Oracle nicht korrekt.

Die Dokumentation von Oracle 12g R2¹ sagt dazu immer noch: Die max. Größe ist 4000 BYTE, auch wenn CHAR angegeben wurde.

VARCHAR2	Maximum size: 4000 bytes, or 32767 bytes if the MAX_STRING_SIZE initialization parameter is set to EXTENDED
----------	---

Daher kommt es beim Import der vorhandenen Daten aus Oracle 11g R2 oder 12c R1 (mit Zeichensatz ISO 8859 P15 oder MSWIN CP 1252) in der 12c R2 mit Zeichensatz AL32UTF8 ggf. zu folgendem Fehler:

```
ORA-02374: Konvertierungsfehler beim Laden von Tabelle "SCHEMA"."VIS_D0301"
ORA-12899: Wert zu groß für Spalte BEM (aktuell: 4000, maximal: 4000)
ORA-02374: Konvertierungsfehler beim Laden von Tabelle "SCHEMA"."VIS_D90"
ORA-12899: Wert zu groß für Spalte WERT_ALT (aktuell: 3906, maximal: 4000)
```

Der Fehler kann mehrere Tabellen betreffen und tritt nur dann ein, wenn der Text nach der Umwandlung in UTF-8 das Limit von 4000 BYTE überschreitet.

Um heraus zu finden, ob Ihre Daten betroffen sind, haben wir eine Kontrollabfrage zur Verfügung gestellt, die diesem Paket beiliegt.

¹ <https://docs.oracle.com/en/database/oracle/oracle-database/12.2/refrn/datatype-limits.html>

2 Das Test- und Korrektur-Paket

Dem Paket liegt eine ZIP-Datei [UTF8-Kontroll_Korrektur-Paket_trp.zip] bei, die folgende Erweiterungen in BALVI iP ergänzt:

- 1.) Kontrollabfrage: Feld-Analyse vor der UTF-8-Umstellung (27559001)
- 2.) Korrektur-Skript: Daten Kürzen für Tabellen VIS_D0301, VIS_D90 (27560001)

Diese beiden Vorlagen sind nur für den SYSTEM/ADMIN** im Objekt „Korrektur-Skripte (systemweit) unter dem jeweiligen Menüpunkt zu erreichen.

Führen Sie die Kontrollabfrage aus, erhalten Sie im Anschluss im modalen Fenster ein Ergebnis ausgegeben, das folgendermaßen aussehen könnte:

betroffene Tabelle : log-Protokoll (VIS_D90)			
WERT_NEU	4000	4002	
WERT_NEU	4000	4032	Messwert = 5.500 mg/kg
WERT_NEU	4000	4032	Messwert = 5.500 mg/kg
WERT_NEU	4000	4007	::no_commit::
WERT_NEU	4000	4002	::no_commit::
WERT_NEU	4000	4002	
WERT_NEU	4000	4007	::no_commit::
WERT_NEU	4000	4021	Zusammenfassung
WERT_NEU	4000	4002	Drop Table VS:ssid::_5300_Betriebe
WERT_NEU	4000	4002	Drop Table VS:ssid::_5180_Betriebe
WERT_NEU	4000	4002	Drop Table VS:ssid::_5190_Betriebe
WERT_NEU	4000	4002	Drop Table VS:ssid::_5280_Betriebe
WERT_NEU	4000	4002	Drop Table VS:ssid::_5290_Betriebe
WERT_NEU	4000	4002	Drop Table VS:ssid::_LE_FIELDS;
WERT_NEU	4000	4002	Drop Table VS:ssid::_2761_RegNrt

Sie sehen die betroffene Tabelle (Gelber Pfeil), die betroffene Spalten (im Beispiel WERT_NEU), die Spalte „Länge Zeichen“ und die Spalte „Länge Byte“ und den berechneten Ersatz nach Kürzung der Daten, so dass es kein Feld mehr gibt, was das Limit 4000 BYTE von Oracle beim Import in UTF-8 überschreitet.

Es ist unvermeidbar, dass der Kürzung logische Zeichen im betroffenen Feld verloren gehen. Im oben beschriebenen Fall „ORA-12899: Wert zu groß für Spalte“ wird beim Import jedoch der komplette Datensatz nicht gespeichert, was zu höheren Datenverlusten führen würde.

Nach Sichtung der aufgetretenen Fehler wurden jedoch nicht automatisch alle Felder gekürzt, da BALVI die Entscheidung nicht automatisch für alle Kunden treffen kann.

Daher wurden im Korrektur-Skript bislang genau zwei Fälle behoben:

- 1.) VIS_D90 – Die Protokolltabelle für die Änderungshistorie
Hier werden alte und neue Zustände bei einer Änderung an den Originaldaten (über Datenbank-Trigger) gespeichert. Dieses Funktion ist nur in wenigen Bundesländern aktiv und da es sich um ein Protokoll handelt, auch nicht als Datenverlust bei Kürzung anzusehen. Im Rahmen der Löschrotation sind diese Daten sowieso irgendwann veraltet und werden automatisch gelöscht.
- 2.) VIS_D0301 (BEM) – Bemerkung in den Terminen.
Termine sind flüchtige Daten, mit Erledigung eines Termins kann dieser auch gelöscht werden. Die Bemerkung ist daher nach dem Ermessen von BALVI keine kritische Information, die unbedingt gerettet werden muss.

Beide Fälle werden mit dem „Korrektur-Skript: Daten Kürzen für Tabellen VIS_D0301, VIS_D90 (27560001)“ automatisch behoben.

Berichtsvorlagen		Nummer
#kürz		
Berichtsvorlagen		Nummer
..... Korrektur-Skript: Daten Kürzen für Tabellen VIS_D0301, VIS_D90		27560001

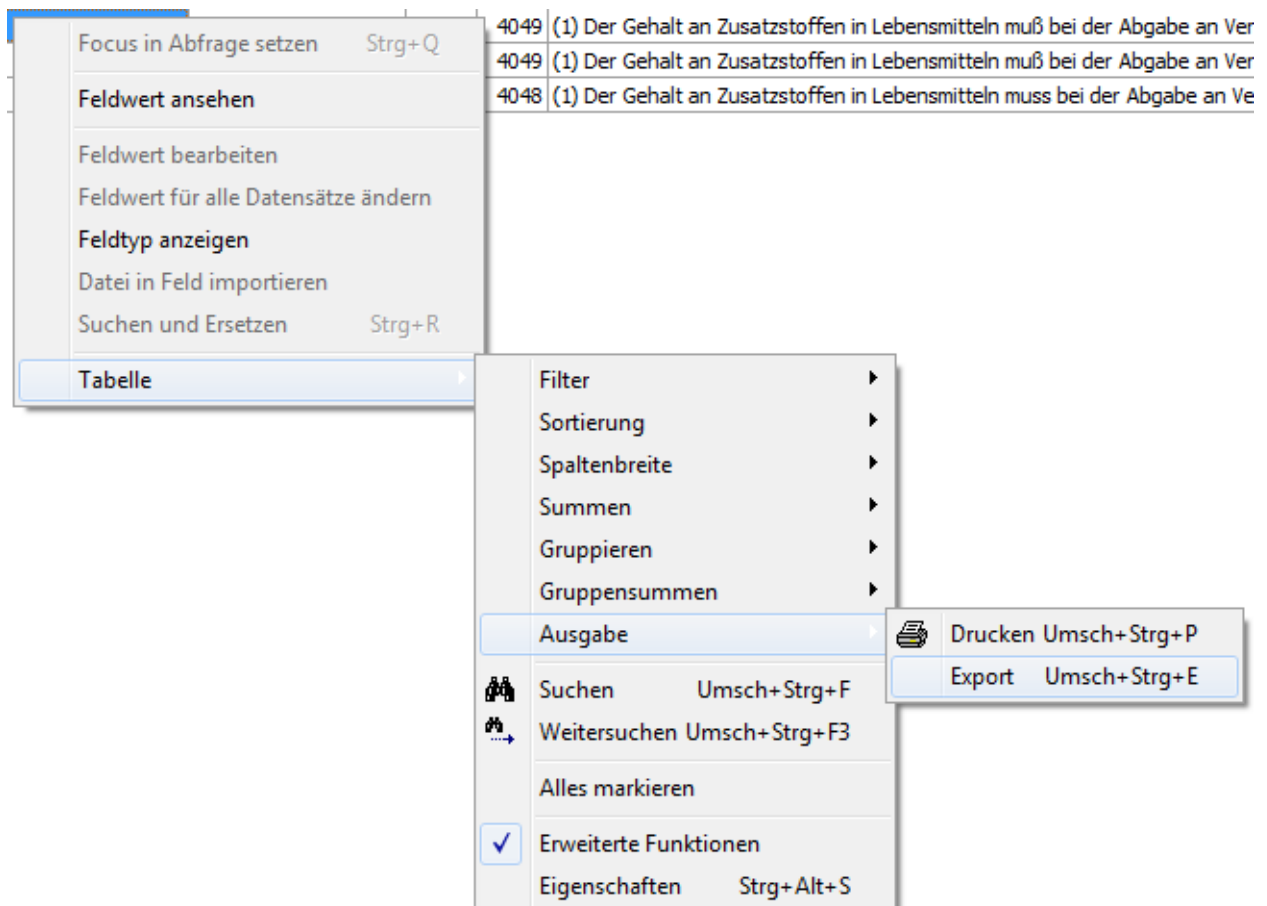
Das Korrektur-Skript muss immer nach der Kontrollabfrage ausgeführt werden. Ansonsten werden keine weiteren Daten geändert. Nach Ausführung wird die Anzeige erneut aufgerufen und alle Fälle ausgegeben, die nach der Korrektur immer noch nicht behoben sind.

Im Beispiel wurde eine Kopie „VTMP_D90“ der Tabelle VIS_D90 angelegt, um das Problem zu erläutern:

BETROFFENE_TABELLE	SPALTENNAME	LAENG	LAENG	ABGESCHNITTENER_WERT
▶ Tabelle: VTMP_D90	WERT_ALT	4000	4048	(1) Der Gehalt an Zusatzstoffen in Lebensmitteln muss bei der Abgabe an Verbraucher wie folgt na...
Tabelle: VTMP_D90	WERT_ALT	4000	4049	(1) Der Gehalt an Zusatzstoffen in Lebensmitteln muß bei der Abgabe an Verbraucher wie folgt nac...
Tabelle: VTMP_D90	WERT_ALT	4000	4049	(1) Der Gehalt an Zusatzstoffen in Lebensmitteln muß bei der Abgabe an Verbraucher wie folgt nac...
Tabelle: VTMP_D90	WERT_ALT	4000	4049	(1) Der Gehalt an Zusatzstoffen in Lebensmitteln muß bei der Abgabe an Verbraucher wie folgt nac...
Tabelle: VTMP_D90	WERT_ALT	4000	4049	(1) Der Gehalt an Zusatzstoffen in Lebensmitteln muß bei der Abgabe an Verbraucher wie folgt nac...

Die nun immer noch aufgeführten Daten in den betroffenen Tabellen müssen geprüft werden, bevor die Übernahme zu Oracle 12c R2 (AL32UTF8) fehlerfrei durchgeführt werden kann.

Um die Prüfung vornehmen zu können, können Sie die Excel-Liste exportieren



BETROFFENE_TABELLE	SPALTENNAME	LAENG	LAENG	ABGESCHNITTENER_WERT
▶ Tabelle: VTMP_D90	WERT_ALT	4000	4048	(1) Der Gehalt an Zusatzstoffen in Lebensmitteln muss bei der Abgabe an Verbraucher wie folgt na...
Tabelle: VTMP_D90	WERT_ALT	4000	4049	(1) Der Gehalt an Zusatzstoffen in Lebensmitteln muß bei der Abgabe an Verbraucher wie folgt nac...
Tabelle: VTMP_D90	WERT_ALT	4000	4049	(1) Der Gehalt an Zusatzstoffen in Lebensmitteln muß bei der Abgabe an Verbraucher wie folgt nac...
Tabelle: VTMP_D90	WERT_ALT	4000	4049	(1) Der Gehalt an Zusatzstoffen in Lebensmitteln muß bei der Abgabe an Verbraucher wie folgt nac...
Tabelle: VTMP_D90	WERT_ALT	4000	4049	(1) Der Gehalt an Zusatzstoffen in Lebensmitteln muß bei der Abgabe an Verbraucher wie folgt nac...

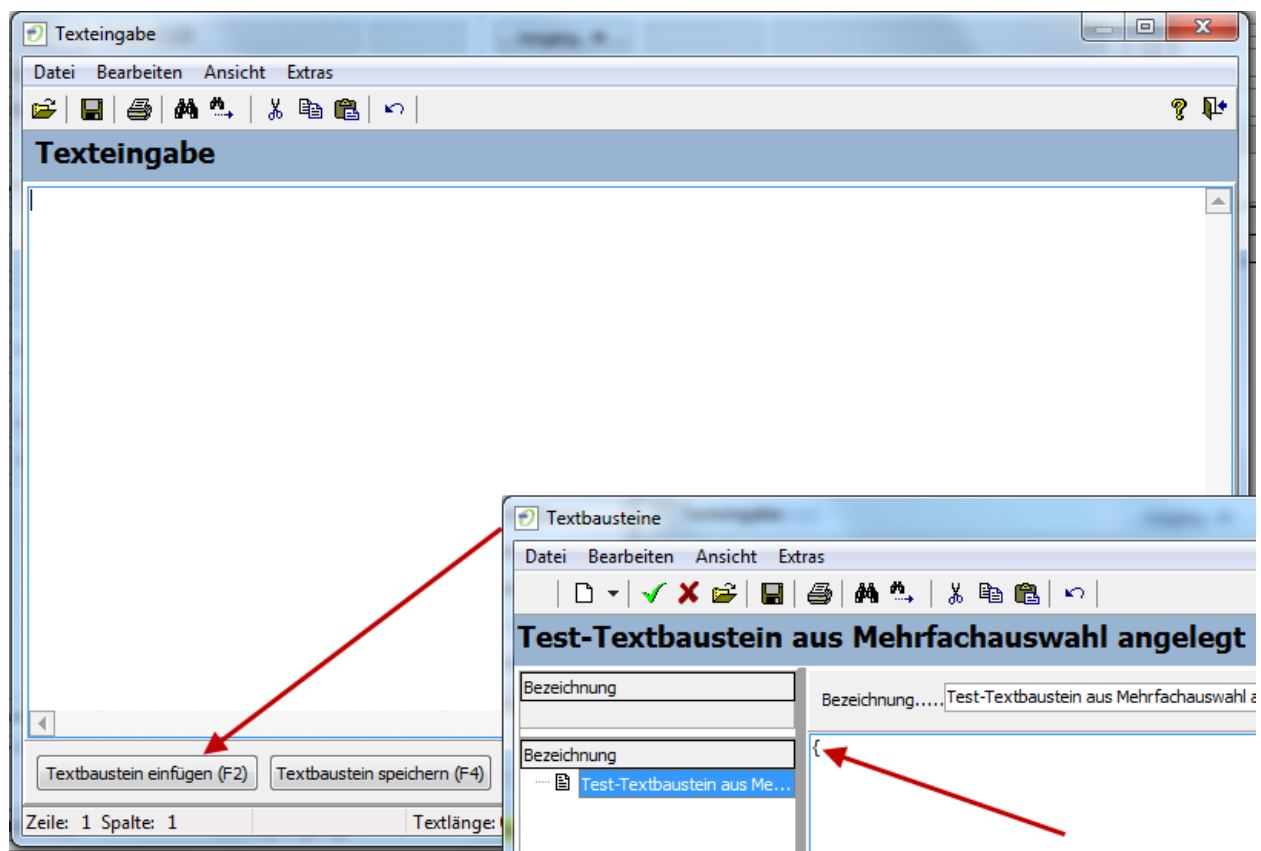
und BALVI zur Prüfung zukommen lassen. Wir werden dann Ihnen dann die möglichen Korrektur-Vorschläge zukommen lassen.

3 Korrekturen in der Anwendung

Die Anwendung BALVI iP wurde zu Release 1.24.5 komplett auf UTF-8-Tauglichkeit untersucht und umgestellt. Leider ist dabei eine Stelle nicht gefunden worden.

3.1 Textbausteine bis 1.24.13.29417 nicht lesbar

Diese Stelle „Textbausteine“ benutzt intern ein RTF-kompatibles Datenformat, um die Absatzformatierungen beibehalten zu können, die dann in Word eingefügt werden.



Defekt geöffneter Textbaustein.

Dieser Fehler wurde mit der Anwendung 1.24.13.29417 behoben, so dass der Inhalt auch nach Umstellung auf UTF-8 wieder korrekt gelesen werden kann.

Die Datei liegt der Auslieferung bei. Tauschen Sie bitte die vorhandene BALVI iP Anwendung gegen die neue Version der balviip.exe (1.24.13.29417) aus, sofern Sie noch keine neuere Version vorliegen haben.

