



Unicode Migration nach Oracle 12c

Johannes Ahrends
Geschäftsführer
CarajanDB GmbH



ORACLE
ACE



- **Vorstellung CarajanDB GmbH**
- **Zeichenkonvertierung**
- **Umstellung der Datenbank**
- **Migration mit Export / Import**
- **Planung einer Migration**
- **Minimal Downtime Migration**
- **... und Oracle 12c?**

- **Experten mit über 30 Jahren Oracle Erfahrung**
- **Spezialisten für**
 - Backup & Recovery
 - Hochverfügbarkeit
 - Healthchecks
 - Performance Optimierung
 - Einsatz von Oracle Standard Edition
 - Oracle in virtuellen Umgebungen und in der Cloud
 - Oracle Migrationen (HW, Unicode, Konsolidierung, Standard Edition)
 - Monitoring (Grid / Cloud Control, HLMM, Foglight, Spotlight)
- **Schulung und Workshops (Oracle, Toad)**



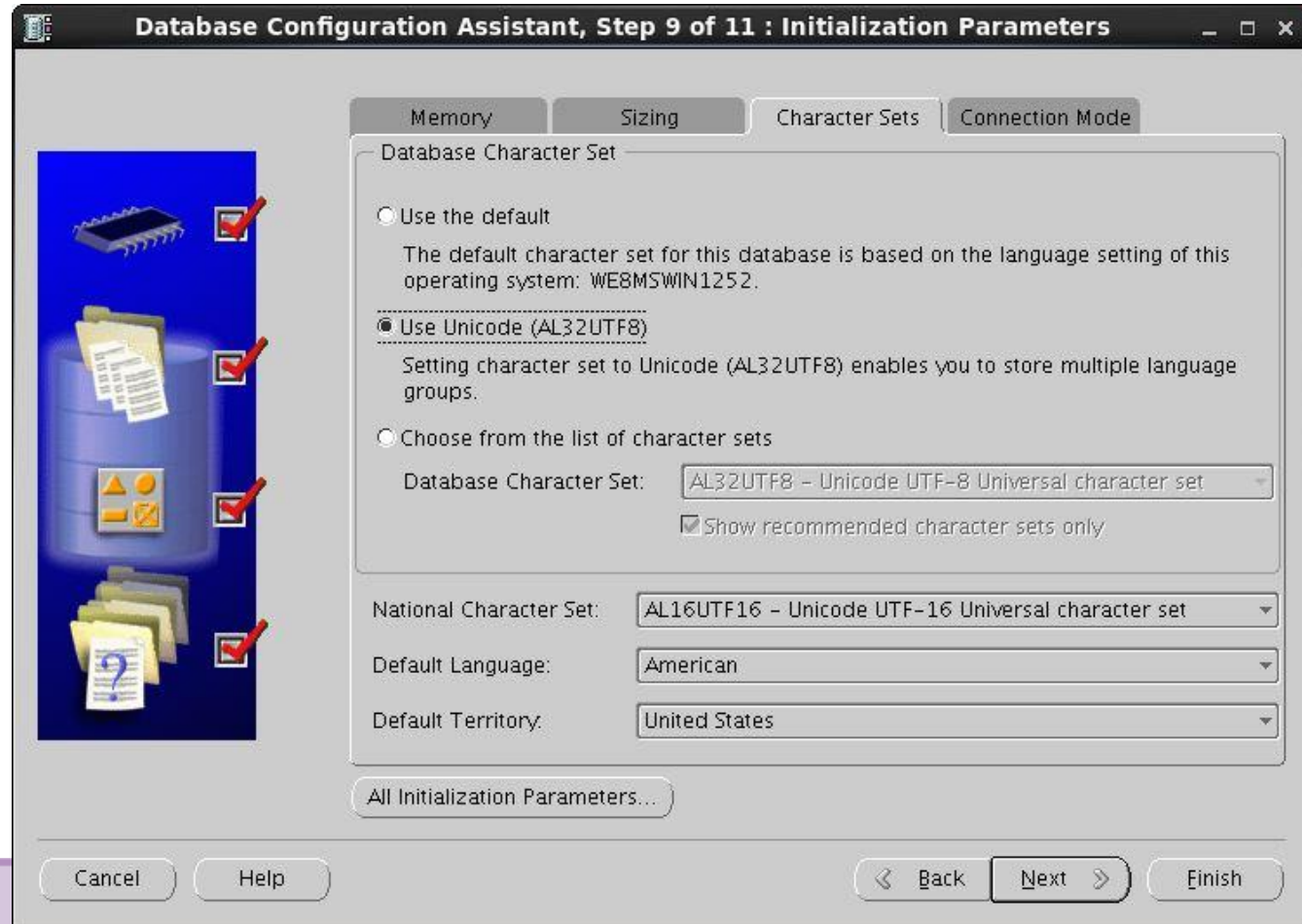


Zeichenkonvertierung

Zeichenkodierungen

Zeichensatz	A	Ö	ä	ß	€
US7ASCII	0x41	N/A	N/A	N/A	N/A
WE8PC850	0x41	0xD6	0xE4	0xDF	N/A
WE8ISO8859P1	0x41	0xD6	0xE4	0xDF	N/A
WE8ISO8859P15	0x41	0xD6	0xE4	0xDF	0xA4
WE8MSWIN1252	0x41	0xD6	0xE4	0xDF	0x80
AL32UTF8	0x41	0xC3 96	0xC3 A4	0xC3 9F	0xE2 82 AC

Auswahl des Charactersets



Zeichenkonvertierung

Linux:
WE8ISO8859P15



€ = 0xA4

Windows
WE8MSWIN1252



€ = 0x80

App-Server
UTF-8



€ = 0xE2 82 AC

€ = 0xE2 82 AC



AL32UTF8

Ein Beispiel (1)

➤ Client: DOS-Box:

```
SQL> insert into umlaute values (1, 'Test ü ß');
```

1 Zeile wurde erstellt.

```
SQL> commit;
```

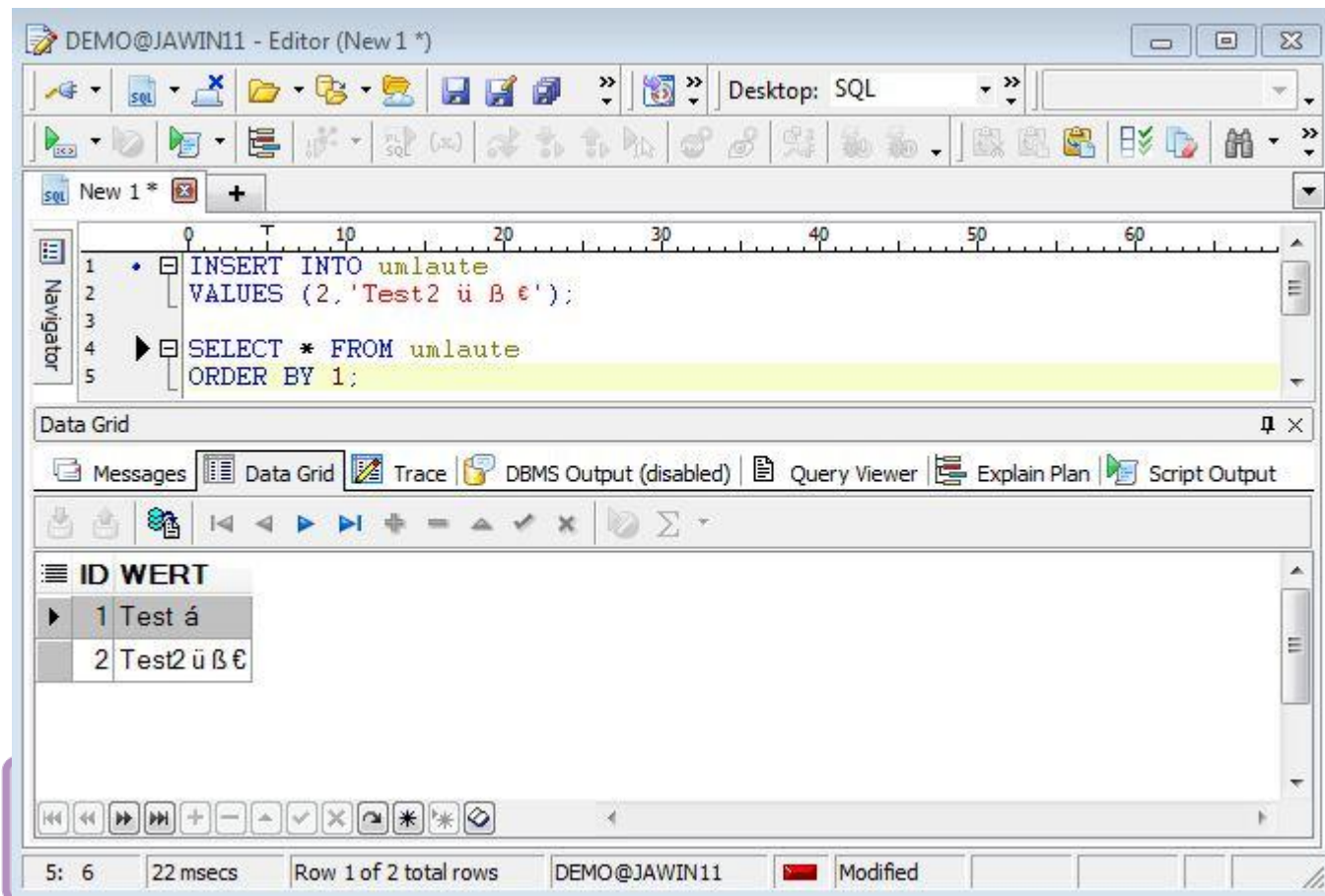
Transaktion mit COMMIT abgeschlossen.

```
SQL> select * from umlaute;
```

ID	WERT
1	Test ü ß

Ein Beispiel (2)

➤ Windows Tool (Toad)

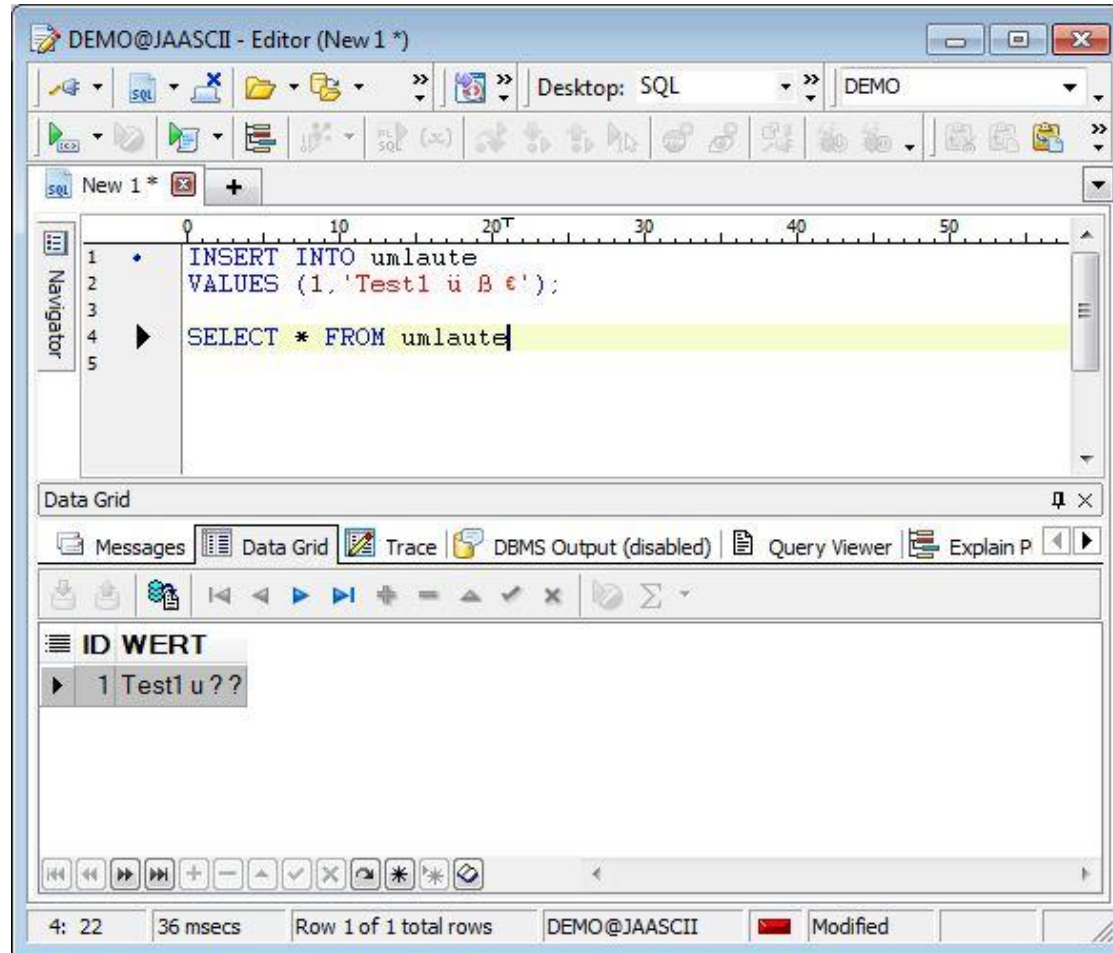


Wer hat recht?

```
SELECT s.SID, s.username, s.PROGRAM,  
       i.client_charset, i.client_version  
FROM v$session s, v$session_connect_info i  
WHERE i.SID=s.SID  
      AND s.PROGRAM IN ('Toad.exe','sqlplus.exe')  
      AND s.username='DEMO'  
      AND i.network_service_banner LIKE 'Windows NT%,;'
```

SID	USERNAME	PROGRAM	CLIENT_CHARSET	CLIENT_VERSION
331	DEMO	Toad.exe	UTF16	11.2.0.3.0
482	DEMO	Toad.exe	UTF16	11.2.0.3.0
484	DEMO	sqlplus.exe	WE8MSWIN1252	11.2.0.3.0

- Das geht doch gar nicht!
- Aber was, wenn ich der Datenbank einfach mitteile, dass keine Konvertierung notwendig ist?



... und Putty?

```
insert into umlaute  
values (2,'Test2 ü ß Ö');
```

1 row created.

```
SQL> commit;
```

Commit complete.

```
SQL> select * from umlaute  
2 ;
```

ID	WERT
1	Test1 u ? ?
2	Test2 ü ß Ö

Warum?

```
export NLS_LANG=GERMAN_GERMANY.US7ASCII
```

- **Damit geht die Datenbank (US7ASCII) davon aus, dass keine Konvertierung notwendig ist!**

Ein Extrembeispiel

Linux:
WE8ISO8859P15



€ = 0xA4

Windows
WE8MSWIN1252



€ = 0x80

App-Server
UTF-8



€ = 0xE2 82 AC

€ = ???



US7ASCII



Umstellung der Datenbank

Umstellung der Datenbank

- CSALTER
- ALTER DATABASE CHARACTER SET ...

- Anlegen eines Benutzers in der Datenbank

```
SQL> ?/rdbms/admin/csminst.sql
```

- Starten der Datenbank im restricted Mode

```
SQL> STARTUP RESTRICT
```

- Scannen der Datenbank

```
% csscan system/manager FULL=Y ARRAY=1024000  
PROCESS=2 TOCHAR=AL32UTF8 TONCHAR=AL16UTF16
```

```
scan.log
```

```
scan.txt
```

```
scan.err
```

Database Scan Individual Exception Report

[Database Scan Parameters]

Parameter	Value
-----	-----
CSSCAN Version	v2.1
Instance Name	JAWE8
Database Version	11.2.0.3.0
Scan type	Full database
Scan CHAR data?	YES
Database character set	WE8ISO8859P15
FROMCHAR	WE8ISO8859P15
TOCHAR	AL32UTF8
Scan NCHAR data?	NO
Array fetch buffer size	1024000
Number of processes	2
Capture convertible data?	NO
-----	-----

...

...

[Data Dictionary individual exceptions]

[Application data individual exceptions]

User : DEMO

Table : UMLAUTE2

Column: WERTGROSS

Type : VARCHAR2(4000)

Number of Exceptions : 1

Max Post Conversion Data Size: 4096

ROWID	Exception Type	Size	Cell Data(first 30 bytes)
AAADZdAAEAAAACOOAA	exceed column size	4096	ÖÜsafda sf saffsa sfsaf ü ö ä

Wenn kein Fehler...

- **Umstellung der Datenbank mit dem Package csalter:**
`SQL> @?/rdbms/admin/csalter.plb`
- ... in den meisten Fällen gibt es aber Fehler
- Erst Fehler beheben, dann Scanner noch einmal laufen lassen

```
SQL> @?/rdbms/admin/csalter.plb
```

```
0 rows created.  
Function created.  
Function created.  
Procedure created.
```

```
This script will update the content of the Oracle Data Dictionary.  
Please ensure you have a full backup before initiating this procedure.  
Would you like to proceed (Y/N)?Y
```

```
old 6:      if (UPPER('&conf') <> 'Y') then
```

```
new 6:      if (UPPER('Y') <> 'Y') then
```

```
Checking data validity...
```

```
Unrecognized convertible data found in scanner result
```

```
PL/SQL procedure successfully completed.
```

```
Checking or Converting phase did not finish successfully  
No database (national) character set will be altered  
CSALTER finished unsuccessfully.
```

```
PL/SQL procedure successfully completed.
```

```
0 rows deleted.  
Function dropped.  
Function dropped.  
Procedure dropped.
```

Wenn kein Fehler...

Zitat (www.orafaq.com):

> **Unrecognized convertible date found in scanner result**

It appears you have "dirty" data.

If you can't find/fix the bad data, your choices are limited.

exp/imp is THE approved method to change character set.



3. Migration mit Export / Import

Export der alten Datenbank

- Muss es unbedingt ein Full Export sein?
- ➔ Besser Schemaexport!

```
DUMPFIL="export.dmp"  
LOGFILE="exp_export.log"  
DIRECTORY=DATA_PUMP_DIR  
FLASHBACK_TIME="TO_TIMESTAMP('2012-10-13 09:00:00','YYYY-MM-DD HH24:MI:SS')"  
COMPRESSION=NONE  
CONTENT=ALL  
SCHEMAS= ('BASIS', 'DEMO')
```

- Dadurch Bereinigung (z.B. HR Schema)

- **Kopieren der Dump Datei auf den Zielrechner**
- **Alternative: Datenbank Link verwenden**

```
DUMPFIL="export.dmp"  
LOGFILE="imp_export.log"  
DIRECTORY=DATA_PUMP_DIR  
STREAMS_CONFIGURATION=n  
TABLE_EXISTS_ACTION=SKIP  
SKIP_UNUSABLE_INDEXES=y  
CONTENT=ALL  
PARTITION_OPTIONS=none
```

Erfolgreich?

```
Processing object type SCHEMA_EXPORT/SEQUENCE/SEQUENCE
Processing object type SCHEMA_EXPORT/TABLE/TABLE
ORA-39083: Object type TABLE:"DEMO"."AUFTRAEGE" failed to create with error:
ORA-00439: feature not enabled: Partitioning
Failing sql is:
CREATE TABLE "DEMO"."AUFTRAEGE" ("AUFID" NUMBER(10,0) NOT NULL ENABLE, "PERSID" NUMBER(10,0) NOT
NULL ENABLE, "AUFDATUM" DATE, "LIEFERDATUM" DATE, "AUFSTATUS" CHAR(1 BYTE) NOT NULL ENABLE) PCTFREE
10 PCTUSED 0 INITRANS 1 MAXTRANS 255 NOCOMPRESS LOGGING STORAGE(INITIAL 65536 NEXT 1048576
MINEXTENTS 1 MAXEXTENTS 2147483645 PCTINCREASE 0 BUFFER_POOL DEFAULT FLASH_CACHE DEF
Processing object type SCHEMA_EXPORT/TABLE/TABLE_DATA
. . imported "DEMO"."PRODUKTE"                172.5 MB      374 rows
. . imported "DEMO"."POSITIONEN"              714.9 KB     29780 rows
. . imported "BASIS"."NACHNAMEN"               177.2 KB     11552 rows
. . imported "BASIS"."ORTE"                   190.7 KB      6104 rows
. . imported "BASIS"."VORNAMEN"               178.2 KB      8856 rows
. . imported "DEMO"."ADRESSEN"                106.7 KB      2000 rows
. . imported "DEMO"."PERSONEN"                39.68 KB      1000 rows
. . imported "DEMO"."TELEFONE"               108.1 KB      4000 rows
. . imported "BASIS"."STRASSEN"               20.24 KB       640 rows
. . imported "DEMO"."BUNDESLAENDER"           5.531 KB        16 rows
. . imported "DEMO"."PRODUKTGRUPPEN"          10.61 KB       156 rows
ORA-02374: conversion error loading table "DEMO"."STATUS"
ORA-12899: value too large for column STATUSID (actual: 2, maximum: 1)

ORA-02372: data for row: STATUSID : 0X'C4'

. . imported "DEMO"."STATUS"                  5.835 KB      5 out of 6 rows
```



Planung einer Migration

➤ Tabellendefinition

- `CREATE TABLE status (
 statusid CHAR(1),
 beschreibung VARCHAR2(50));`

➤ Was bedeutet die Längenangabe?

- `CHAR(1)` = 1 BYTE oder 1 Zeichen?
- `VARCHAR2(50)` = 50 BYTE 0 der 50 Zeichen?

➤ Besser:

- `CREATE TABLE status (
 statusid CHAR(1 CHAR),
 beschreibung VARCHAR2(50 CHAR));`

➤ Alternativ:

- `ALTER SESSION SET NLS_LENGTH_SEMANTICS='CHAR';`

➤ Längensemantik eingeschaltet

```
CREATE TABLE STATUS (  
  STATUSID          CHAR(1 BYTE)    NOT NULL,  
  KURZBESCHREIBUNG VARCHAR2(10 BYTE),  
  BESCHREIBUNG      VARCHAR2(255 BYTE)  
) TABLESPACE USERS;
```

➤ Längensemantik ausgeschaltet

```
CREATE TABLE STATUS (  
  STATUSID          CHAR(1)    NOT NULL,  
  KURZBESCHREIBUNG VARCHAR2(10),  
  BESCHREIBUNG      VARCHAR2(255)  
) TABLESPACE USERS;
```

- Maximale Länge der Datentypen:
 - CHAR = 2000 **BYTE**
 - VARCHAR2 = 4000 **BYTE**
- Was ist, wenn der Datentyp nicht ausreicht?
 - VARCHAR2(4000) → CLOB
- Maximale Länge von Objektnamen = 30 **BYTE**
- CLOB Datentyp beansprucht mehr Platz
- Indexlänge maximal ein Oracle Block

„Oracle recommends using Unicode for all new system deployments.”

“Data in CLOB columns is stored in a format that is compatible with UCS-2 when the database character set is multibyte, such as UTF8 or AL32UTF8. This means that the storage space required for an English document doubles when the data is converted.”

- **Objektdefinitionen ohne Längensemantik exportieren**
 - Z.B. Toad → Generate Schema Script
- **Anpassen der Definitionen**
 - Ändern von VARCHAR2 → CLOB, wenn 4000 Zeichen
 - Ändern von NVARCHAR2 → VARCHAR2
 - Ändern von Objektnamen (hoffentlich nicht!)
- **Erstellen der Schemata auf dem Zielsystem mit Längensemantik 'CHAR'**
- **Ausschalten von Foreign Key Constraints**
- **Exportieren der Daten (CONTENT=DATA_ONLY)**
- **Importieren der Daten (TABLE_EXISTS_ACTION=APPEND)**
- **Foreign Key Constraints einschalten (NOVALIDATE)**

... und was war mit der US7ASCII DB?

- Export der Datenbank führt dazu, dass alle Umlaute weg sind!
- Umgehung:
 - Starten der Datenbank im Restricted Mode:
`SQL> STARTUP RESTRICT`
 - Switch des Datenbank Zeichensatzes:
`SQL> ALTER DATABASE
CHARACTERSET INTERNAL_USE WE8ISO8859P15;`
 - Export der Schemata, die diesen Zeichensatz benutzen
 - Switch des Datenbank Zeichensatzes:
`SQL> ALTER DATABASE
CHARACTERSET INTERNAL_USE AL32UTF8;`
 - Export der Schemata, die diesen Zeichensatz benutzen



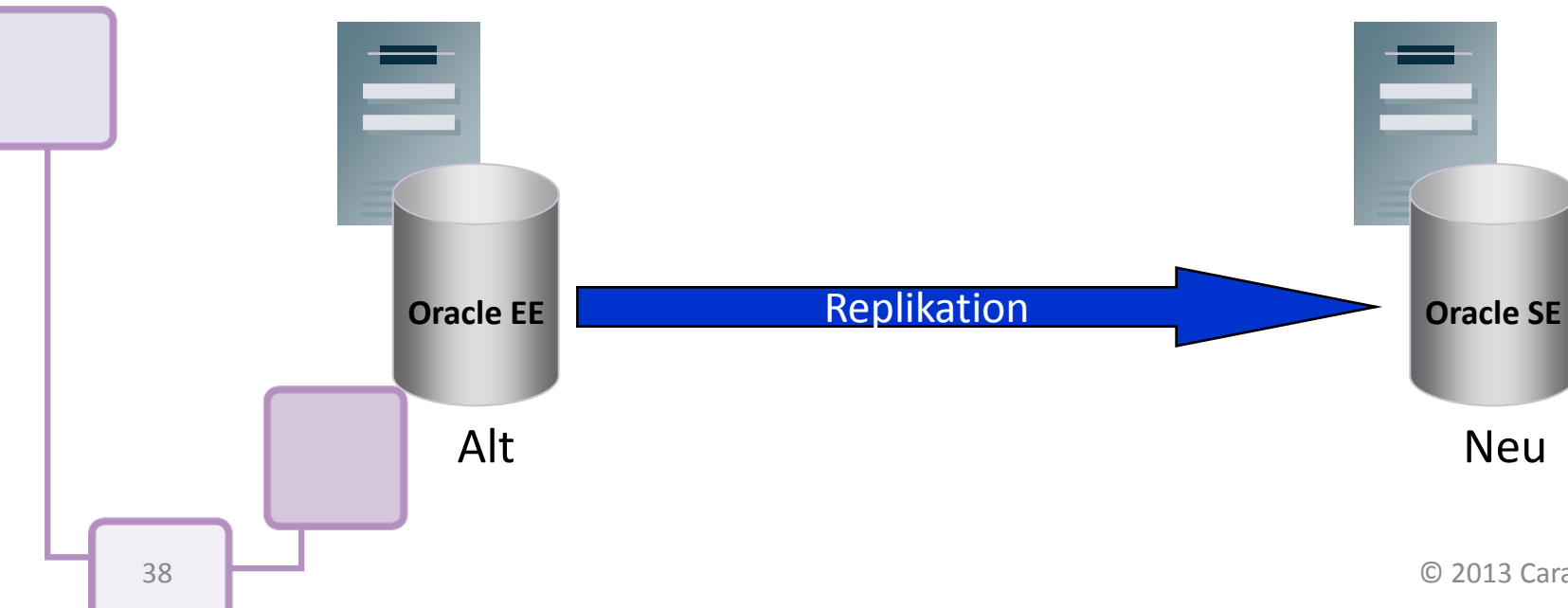
Minimal Downtime Migration

- **Definition einer „NO DDL“ Periode**
 - Es dürfen keine weiteren Änderungen an den Datenbankstrukturen vorgenommen werden
- **Exportieren und importieren der Definitionen wie bei „normaler“ Migration.**
- **Aufbau einer Replikationsumgebung mit**
 - Quest / DELL SharePlex for Oracle
 - Oracle Goldengate
 - Dbvisit Replicate

- Capture / Mine aller Änderungen mit der Replikationssoftware
- Kein Apply / Post auf dem Zielsystem!!!
- Data Pump Export / Import mit Flashback SCN
- Reconcile des Apply / Post Prozesses auf die SCN
- Starten des Apply / Post Prozesses
- Warten ...
- Datenbanken in Sync ➔ Umdrehen der Replication (Fallback Lösung)

Minimal Downtime Migration

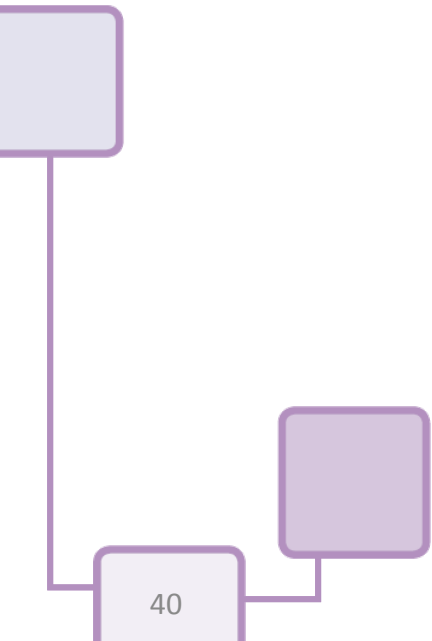
- Migration von Zeichensätzen
- Plattformwechsel
- Umstellung auf Oracle Standard Edition
- Migration nach Oracle 12c





... und Oracle 12c

???



- **Kritisch für die Umstellung auf Unicode sind:**
 - Ausreizen der maximalen Datentyp Längen (4000 Byte)
 - „Schmutzzeichen“ (keine Zeichenkonvertierung)
 - CLOB Länge
- **Migration am Besten mit Data Pump Export / Import**
 - Schemata vorher anlegen mit Längensemantic = „CHAR“
 - Einzelne Schemata exportieren und Importieren
- **Replikationslösungen für Minimal Downtime Migration**
 - Migration im laufenden Betrieb
 - Fallback Möglichkeit



Fragen?

Johannes Ahrends
www.carajandb.com

Johannes.ahrends@carajandb.com



ORACLE
ACE

