

Masterarbeitsthema

Erstellung und Wiederherstellung von Testdaten für die Simulation von Produktionssystemen



Hintergrund:

Produktionssysteme von Unternehmen enthalten meist sehr große Mengen an Daten, die über die Jahre gewachsen sind. Oft ist ein Last-Test mit Kopien der Daten des Produktionssystems aus organisatorischen oder rechtlichen Gründen leider nicht möglich. Dann müssen geeignete Testdaten generiert werden. Damit ein Last-Test aussagekräftig ist, sollte die Größe und Verteilung der generierten Datenmenge auf die Seiten der Datenbank dem Produktionssystem weitgehend entsprechen. Aufgrund der großen Datenmenge muss die Erzeugung der Testdaten möglichst effizient sein.

Auslöser dieses Thema ist ein Projekt (Merger zweier deutscher Netzbetreiber im Mobilfunk), bei dem sehr große Datenmigrationen durchgeführt werden müssen, welche durch geeignete Last-Tests mit generierten produktions-nahen Daten ausführlich getestet worden sind.



Aufgabenstellung:

Ziel dieser Arbeit ist der strukturierte Vergleich von Techniken sowohl zum Erstellen als auch zum Wiederherstellen von Testdaten nach einem Testlauf.

Weiter soll ein generisches Werkzeug zum Erstellen und Wiederherstellen solcher Testdaten implementiert werden. Dieses soll verschiedene Strategien für das Erstellen und Wiederherstellen der Testdaten unterstützen (s.u.). Bei der Entwicklung der Software soll eine sinnvolle, begründete Architektur sowie Erweiterbarkeit gemäß Software-Engineering-Grundprinzipien sichergestellt werden.

Als zugrunde liegendes Datenbanksystem soll die Oracle Datenbank verwendet werden.

Die Performanz folgender Techniken zum Einfügen von soll mindestens untersucht werden:

- Datei-basiert (via Import von Datei, die von Java-Programm erzeugt wurde):
 - Import mit imp
 - Import mit impdp (via Data Pipe)
 - SQL+ Loader
- Direkt (via Stored Procedure in PL/SQL):
 - Direct Path Insert (einfach)
 - Direct Path Insert (block-weise)
 - Merge (block-weise)
 - Forall (mit Arrays / Bulks)

Darüber hinaus soll beim Generieren der Daten darauf geachtet werden, dass die Daten „produktions-nah“ sind. Die Verteilung der Tupel auf die Seiten der Datenbank soll maximal ungünstig sein. D.h. z.B. 100 Tupel mit aufsteigenden Primärschlüsseln sollen auf 100 verschiedene Seiten verteilt werden.

Weiter soll untersucht werden, wie sich die Test-Daten, welche durch einen Testlauf verändert wurden, am besten wieder in den ursprünglichen Zustand versetzen lassen.

Folgende Strategien sollen mindestens betrachtet werden:

- Datei-basiert (via Export und anschließendem Import):
 - Export mit exp und Import mit imp
 - Export mit expdp und Import mit impdp (via Data Pipe)
 - Export und Import mit SQL+ Loader
- Direkt (via SQL/ SQL+):
 - Kopie erstellen durch:
 - CTAS (Create Table As Select)
 - SQL+ copy
 - Daten zurückspielen mit:
 - Drop und Rename
 - Drop und CTAS
 - Truncate und Direct Path Insert
 - SQL+ copy



Voraussetzungen:

- Gute Kenntnisse in der Vorlesung „Datenbanken und Informationssysteme“
- Gute Kenntnisse in der Vorlesung „Transaktionssysteme“

Organisation & Prozedere



Thema:

Interessantes, aktuelles Thema, welches zwischen Lehrstuhl für Informationsmanagement und ConSol bereits abgesprochen ist



Betreuung:

Sowohl einen Betreuer am Lehrstuhl für Informationsmanagement als auch einen Betreuer bei ConSol:

- Regelmäßiges Betreuungsgespräch mit Betreuer am Lehrstuhl
- Betreuung bei ConSol durch



*Dr. Fabian Stäber,
Head of Research
& Development*



*Dr. Christoph Ehlers,
Alumnus des Lehrstuhl für
Informationsmanagement*



Zwei Varianten:

- Betreuungsgespräch alle ein bis zwei Wochen bei ConSol in München (weitere Absprachen via E-Mail, Telefon oder Skype)
- Abschlussarbeit als Praktikum vor Ort bei ConSol in München



Ablauf:

1. Bewerbung
2. Einarbeitung in das Thema der Abschlussarbeit
3. Antrittsvortrag im Oberseminar des Lehrstuhls in Passau
4. Zwischenvortrag im Oberseminar des Lehrstuhls in Passau
5. Abgabe der schriftlichen Ausarbeitung
6. Abschlussvortrag im Oberseminar des Lehrstuhls in Passau
7. Abschlussvortrag in der ConSol-Akademie in München



Fragen?

Fragen zum Thema oder zum Ablauf der Arbeit gern via E-Mail an:

abschlussarbeiten@consol.de

Interesse?

Sende Deine aussagekräftige Bewerbung (mit aktuellem HISQIS-Notenauszug) für diese Abschlussarbeit am Lehrstuhl für Informationsmanagement und bei ConSol bitte an:

abschlussarbeiten@consol.de