

Möglichkeiten und Grenzen von Oracle Advanced Compression



Olaf Herden

Duale Hochschule BW Campus Horb





Gliederung

- Motivation
- Kompression in Oracle
- Untersuchungen
- Zusammenfassung & Ausblick



Motivation

- Natürliches Wachstum der Geschäftsdaten
- Migration existierender Desktop-Applikationen in unternehmensweite Anwendungen
- Entwicklung neuer Applikationen
- Gesetzliche Anforderungen (z.B. Basel II oder Sarbanes Oxley)
- Zunahme unstrukturierter und multimedialer Daten
- Redundante Speicherung von Daten in Data Warehouses
- Neue Technologien wie RFID produzieren große Datenmengen



Kompressionsfaktor

- Definition

$$KF = \frac{\text{Platz unkomprimierte Speicherung}}{\text{Platz komprimierte Speicherung}}$$

- Beispiel:

- 100 TB Rohdaten
- Nach Kompression 20 TB
- Kompressionsfaktor = 5
- Oder Kompressionsrate 1:5
- Oder Gesparter Speicherplatz = 80%



Kompression in Datenbanken

- Kompression:
 - Seit langem untersucht und angewendet
 - Stand der Technik: LZ-Algorithmen-Familie
- In DB-Umfeld „nicht so einfach“:
 - Einerseits:
 - Gute Kompressionsrate
 - Andererseits:
 - Overhead für (De)kompression beschränken
 - Overhead beim Logging reduzieren
 - INSERT, UPDATE und DELETE nicht signifikant langsamer



Übersicht

- Motivation
- **Kompression in Oracle**
- Untersuchungen
- Zusammenfassung & Ausblick

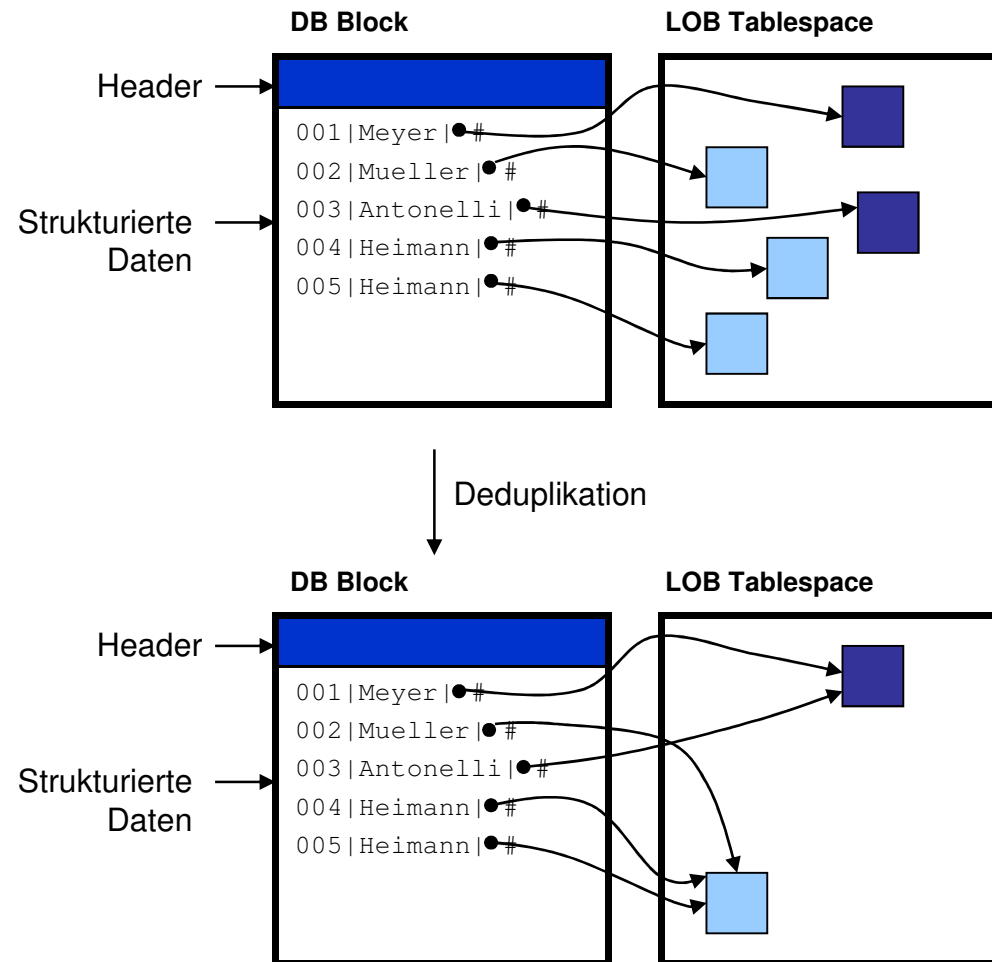


Kompression in Oracle

- Historie:
 - 8i: Kompression von Indexen
 - 9i R2: Basic Table Compression (Bulk Load)
 - 10g: LOB Kompression
 - 11g R1: OLTP Table Compression
- Advanced Compression in 11g R2:
 - Kompression von Sicherungsdateien
 - Data Guard Netzwerkkompression
 - Kompression von SecureFile-Objekten
 - Kompression strukturierter Daten
- Lizenzen:
 - Basic Table Compression in 11g EE enthalten
 - Advanced Compression zusätzlich lizenzpflichtig

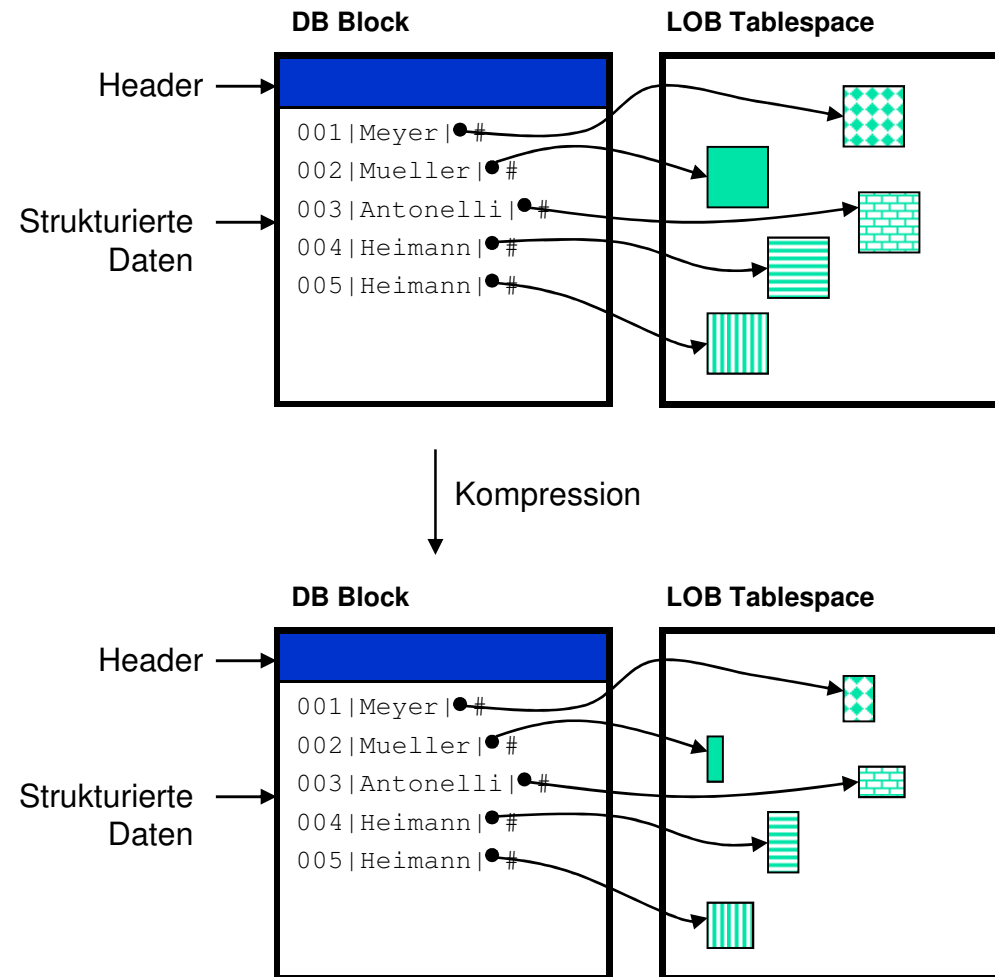
SecureFile-Objekte (I)

- Kompression durch Deduplikation:



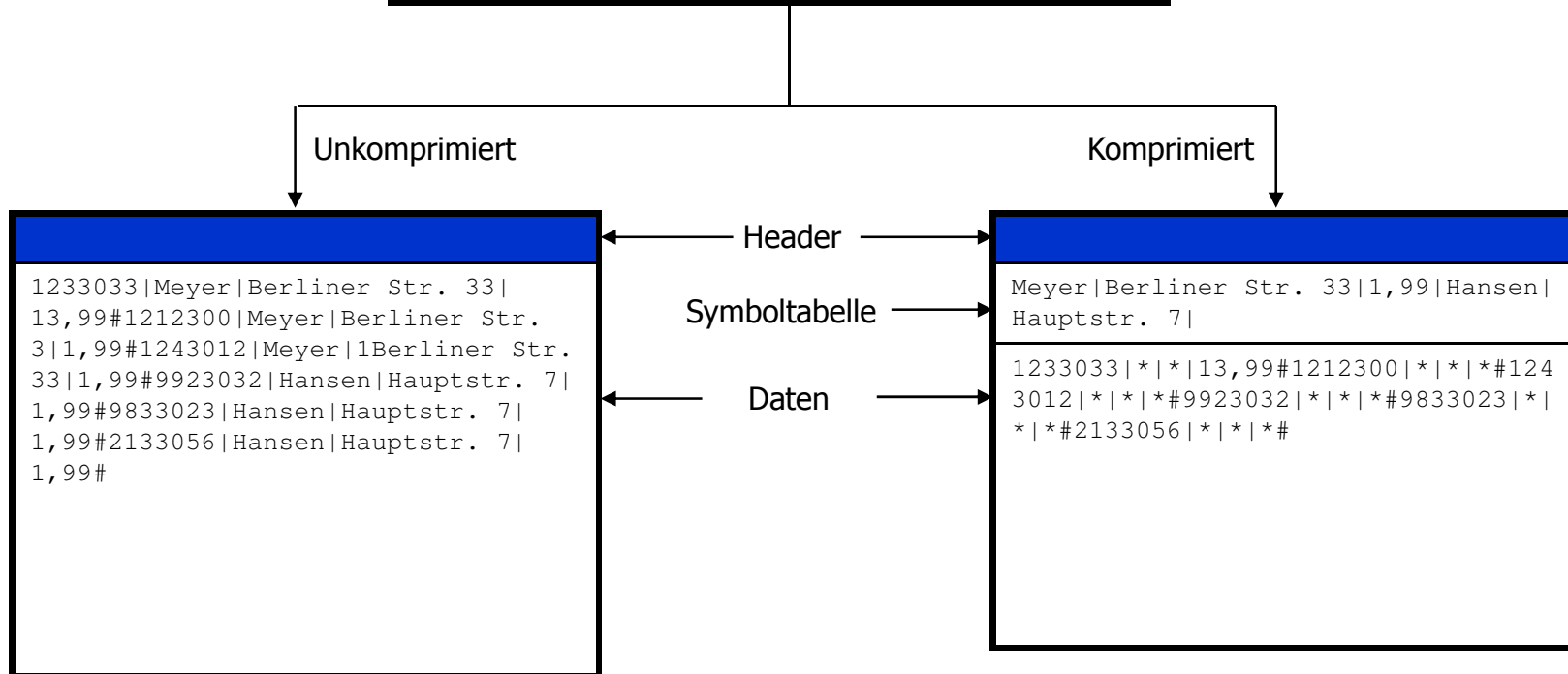
SecureFile-Objekte (II)

- Kompression:

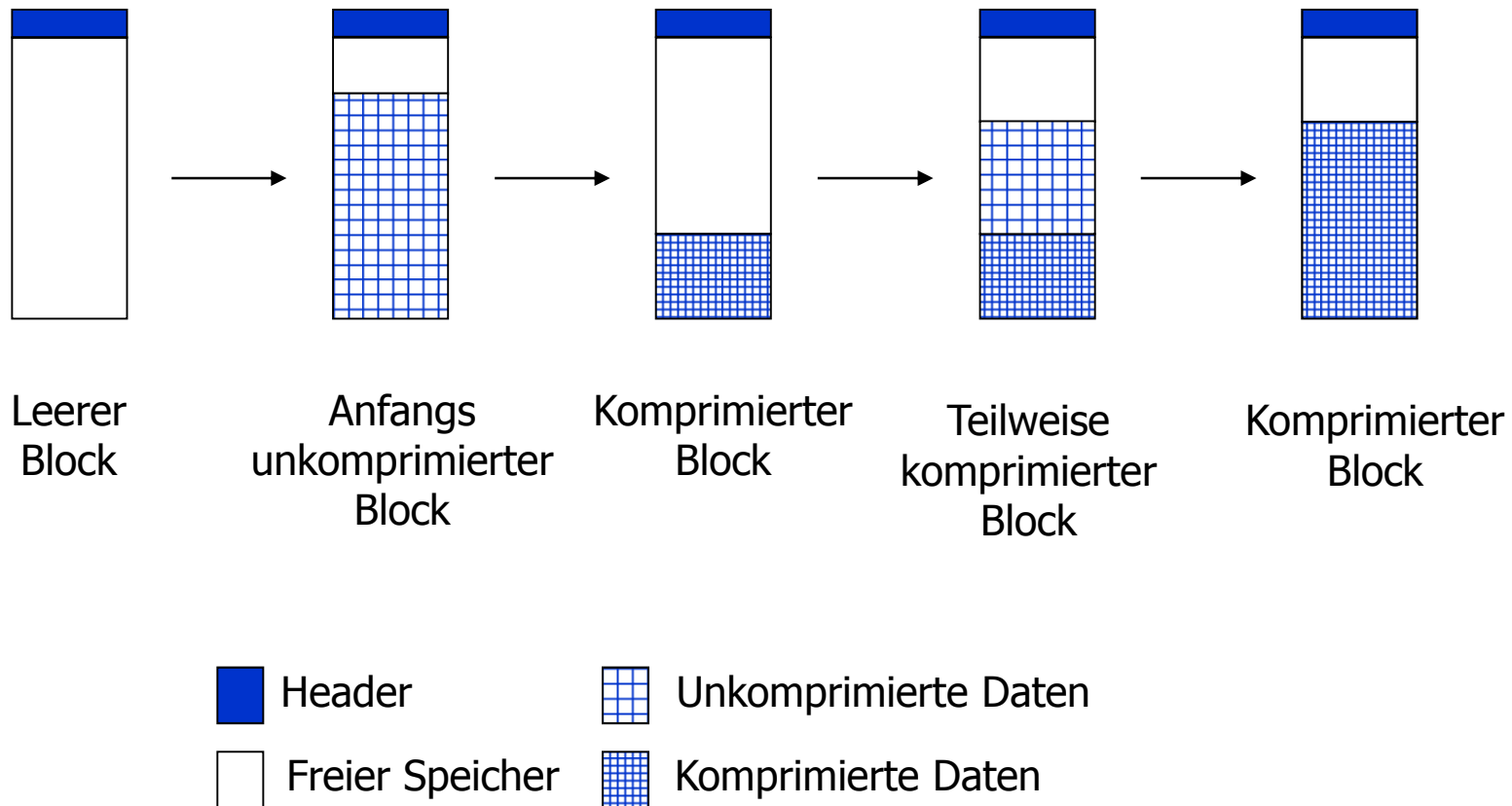


Strukturierte Daten

Re_Nr	Name	Adresse	Betrag
1233033	Meyer	Berliner Str. 33	13,99
1212300	Meyer	Berliner Str. 33	1,99
1243012	Meyer	Berliner Str. 33	1,99
9923032	Hansen	Hauptstr. 7	1,99
9833023	Hansen	Hauptstr. 7	1,99
2133056	Hansen	Hauptstr. 7	1,99



Kompressionsprozess





Übersicht

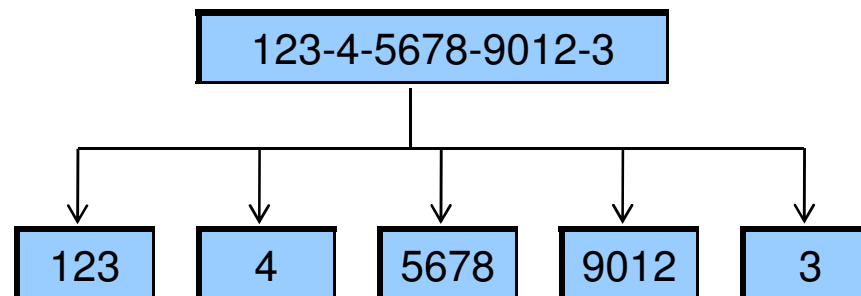
- Motivation
- Kompression in Oracle
- Untersuchungen
- Zusammenfassung & Ausblick

Ähnliche, nicht-gleiche Daten

- Beispiel:
 - Fortlaufend generierte ISBN-Werte
 - Keine Kompression

ISBN
000-0-0000-0000-0
000-0-0000-0000-1
000-0-0000-0000-2
000-0-0000-0000-3
...

- Feld-Separierung:



- Kompressionsfaktor 1,8



Transaktionsdaten

- Beispiel:

Tabelle Event	
Attribut	Typ
ID	NUMBER(10)
Event_Time	DATE
Terminal_ID	NUMBER(10)
Value	NUMBER(10,2)

Tabelle Event_Separiert	
Attribut	Typ
ID	NUMBER(10)
Event_Year	NUMBER(4)
Event_Month	NUMBER(12)
Event_Day	NUMBER(2)
Event_Hour	NUMBER(2)
Event_Minute	NUMBER(2)
Event_Second	NUMBER(2)
Terminal_ID	NUMBER(10)
Value	NUMBER(10,2)

- In beiden Fällen keine Kompression

Fakttabellen (I)

Tabelle F1		
Attribut	Typ	Wertebereich
Dim1_ID	NUMBER(10)	Zusammengesetzter Primärschlüssel
Dim2_ID	NUMBER(10)	
Dim3_ID	NUMBER(10)	
IntAttribut_1	NUMBER(10)	0-2
IntAttribut_2	NUMBER(10)	1-3
IntAttribut_3	NUMBER(10)	1-10
EnumAttribut_1	VARCHAR2	red, yellow, green
EnumAttribut_2	VARCHAR2	true, false
EnumAttribut_3	VARCHAR2	cheese, cloth, toys, sale, foto, cosmetics, furniture, fruit, meat, frozen
RealAttribut_1	NUMBER(10,2)	0,0 – 2,0
RealAttribut_2	NUMBER(10,2)	0,0 – 2,0
RealAttribut_3	NUMBER(10,2)	0,0 – 2,0
RealAttribut_4	NUMBER(10,2)	1,0 – 10,0
RealAttribut_5	NUMBER(10,2)	1,0 – 10,0

Datensätze	Blöcke Unkomprimiert	Blöcke Komprimiert	Faktor
100000	249	123	2,0
200000	469	217	2,2
500000	1131	532	2,1
1000000	2268	1129	2,0



Fakttabellen (II)

- Faktttabelle mit
 - Zusammengesetztem, dreiattributigem Primärschlüssel
 - Und 10 Integer-Attributen
 - F2: 10 Attribute aus [1,10]
 - F3: 10 Attribute aus [1001,1010]
 - F4: 10 Attribute aus [100001,100010]
- Kompressionsraten:

Datensätze	Tabelle F2			Tabelle F3			Tabelle F4		
	Blöcke Unkomprimiert	Blöcke Komprimiert	Faktor	Blöcke Unkomprimiert	Blöcke Komprimiert	Faktor	Blöcke Unkomprimiert	Blöcke Komprimiert	Faktor
100000	172	91	1,9	217	91	2,4	280	91	3,1
200000	343	154	2,2	406	154	2,6	532	154	3,5
500000	879	375	2,3	1036	375	2,8	1320	375	3,5
1000000	1729	753	2,3	2013	753	2,7	2778	753	3,7



Verwandte Ansätze

- [Orac09]:
 - „Normale OLTP-Daten“
 - Kompressionsfaktor 2-3
- [Nand08]:
 - Synthetisch erzeugte Stammdatentabelle
 - Kompressionsfaktor 2,75
- [PoPo03]:
 - „Reale Daten“ Kompressionsfaktor 3-4
 - Lineitem-Fakttabelle des TPC-H-Benchmark
Kompressionsfaktor 1,6



Übersicht

- Motivation
- Kompression in Oracle
- Untersuchungen
- Zusammenfassung & Ausblick



Zusammenfassung: Möglichkeiten

- Kompressionsfaktoren 1-4
- Nicht das Optimum, aber ausgewogen bzgl. Overhead
- Vorteile der Kompression:
 - Einsparung Plattenplatz
 - Schnellere Antwortzeiten
 - Reduzierung Energieverbrauch
 - Kleinere Backups, schnellere Sicherung und Wiederherstellung



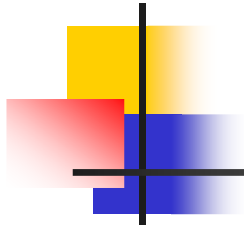
Zusammenfassung: Grenzen

- (Kaum/keine) Kompression bei „normalen“ OLTP-Daten
- Keine allgemeinen Aussagen der Form „Wenn Daten soundso, dann wird Kompression Dasunddas bringen“
- Also: Immer Test mit realen Daten notwendig
- Lizenzierung kritisch zu hinterfragen
- Not open, not complete, integrated?



Ausblick

- **Ausblick:**
 - Weitere Untersuchungen
 - Bezgl. Kompressionsfaktor
 - Bezgl. Verarbeitungsgeschwindigkeit
- **Bessere Kompressionsraten:**
 - Berücksichtigung Teilworte (MS SQL Server, IBM DB2)
 - Tabellen- statt blockbezogener Kompression (IBM DB2)
 - Spaltenorientierte Speicherung (Oracle Exadata, Sybase IQ)
- **Wünsche für die Zukunft:**
 - Weitere Kompressionsmöglichkeiten
 - Offenheit, Erweiterbarkeit
 - Calculator/Advisor



Vielen Dank für Ihre
Aufmerksamkeit!

Haben Sie Fragen?