



# XML-Archivierung betriebswirtschaftlicher Datenbank-Objekte\*

---

**Bernhard Zeller**<sup>1</sup>

**Axel Herbst**<sup>2</sup>

**Alfons Kemper**<sup>1</sup>

<sup>1</sup> **Universität Passau**

**94030 Passau**

**<Nachname>@db.fmi.uni-passau.de**

<sup>2</sup> **SAP AG**

**69190 Walldorf**

**axel.herbst@sap.com**

**\* Diese Arbeit wurde durch die Firma SAP im Rahmen des sog. Terabyte-Projektes gefördert.**



# Gliederung

---

- Motivation
- Archivierung bisher
- Verwandte Arbeiten
- Besonderheiten der XML-Archivierung
- Der XML-Archivierungs-Operator
- Leistungsanalyse
- Zukünftige Arbeiten
- Zusammenfassung



# Motivation

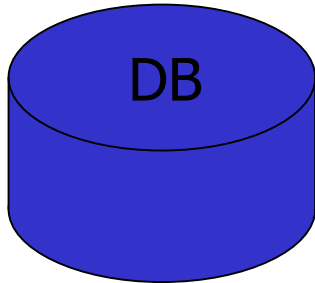
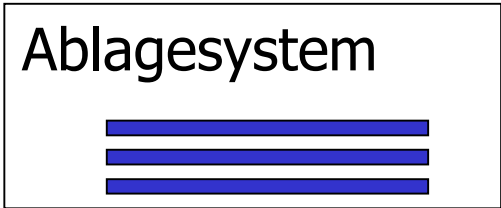
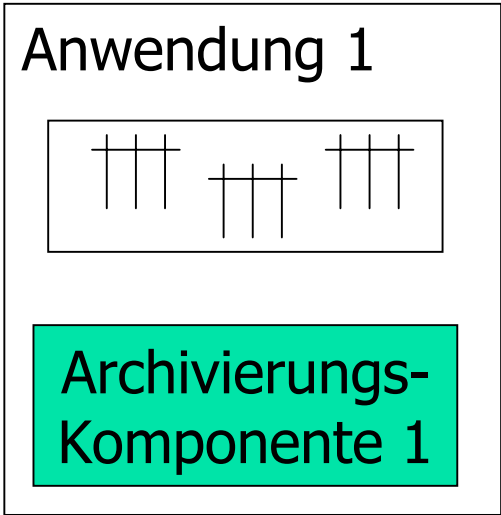
---

- Datenbanken betriebswirtschaftlicher Software wie z.B. SAP R/3 wachsen stetig (bereits mehrere Terabyte)
  - Tabellen mit sehr vielen Einträgen
  - Leistungseinbußen, schwer administrierbar
- Archivierung kann Probleme lösen / lindern durch **Verschieben der Daten von selten oder wenig benötigten betriebswirtschaftlichen Objekten aus der Datenbank heraus auf Tertiärspeichersysteme** (z.B. CDs)
  - Tabellen werden kleiner
  - Kosten werden gesenkt (CDs billiger als Platten)

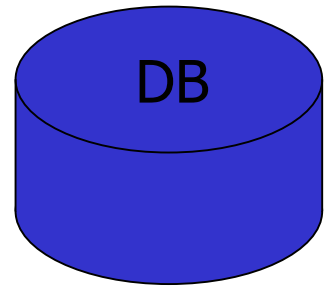
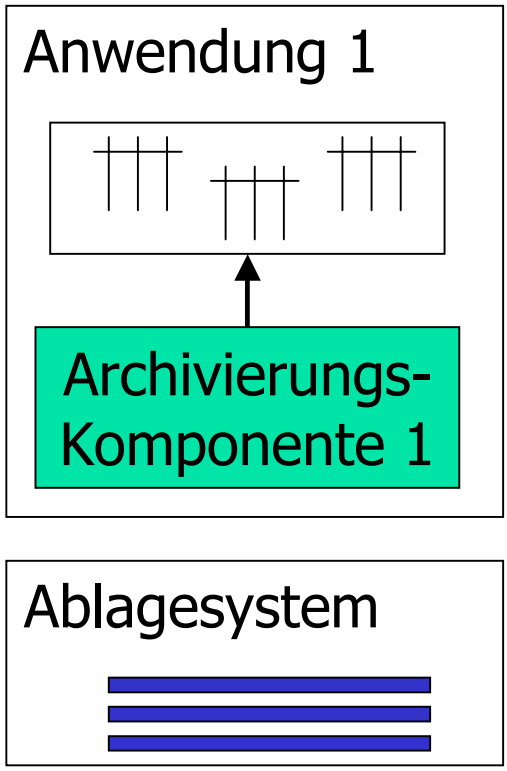


# Archivierung bisher

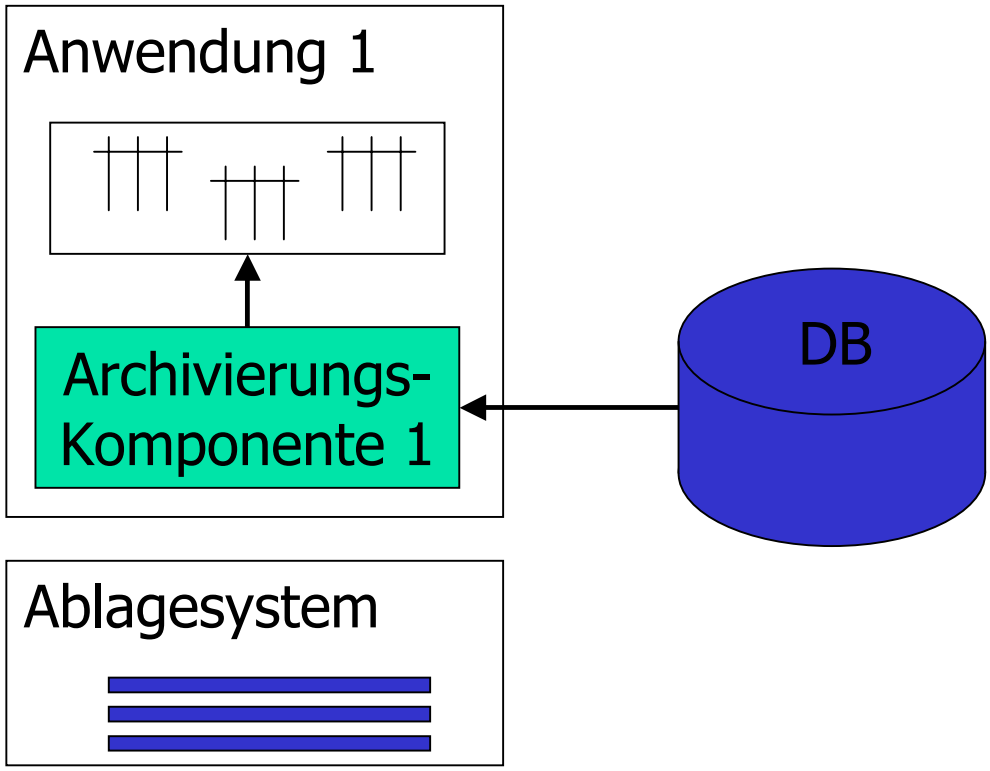
---



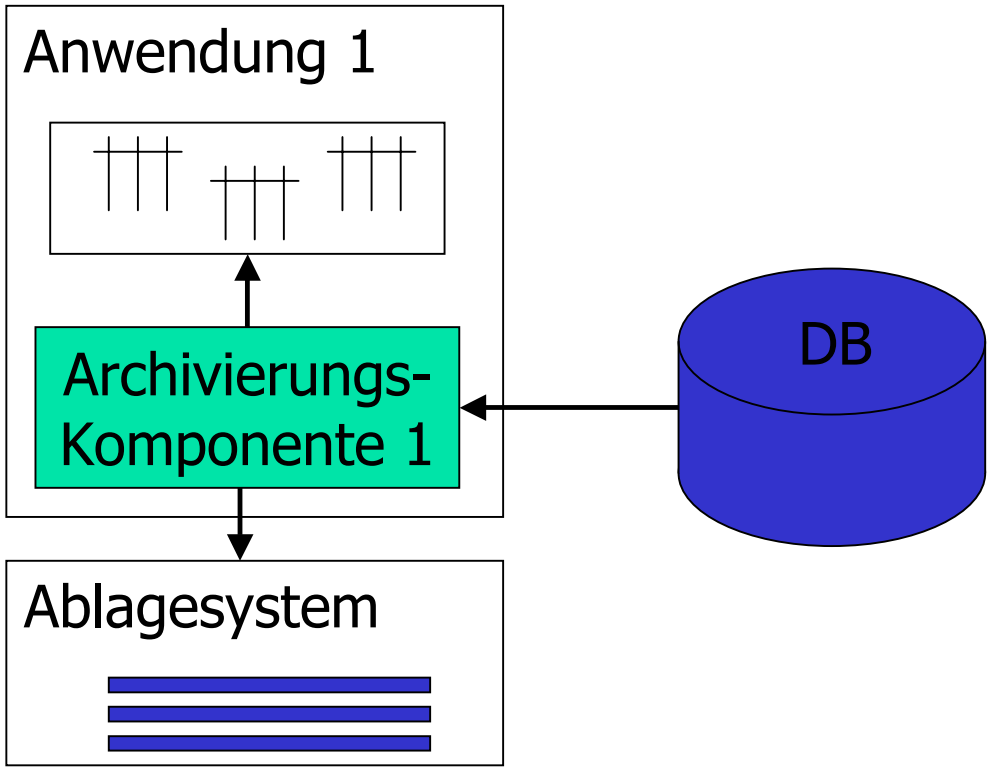
# Archivierung bisher



# Archivierung bisher

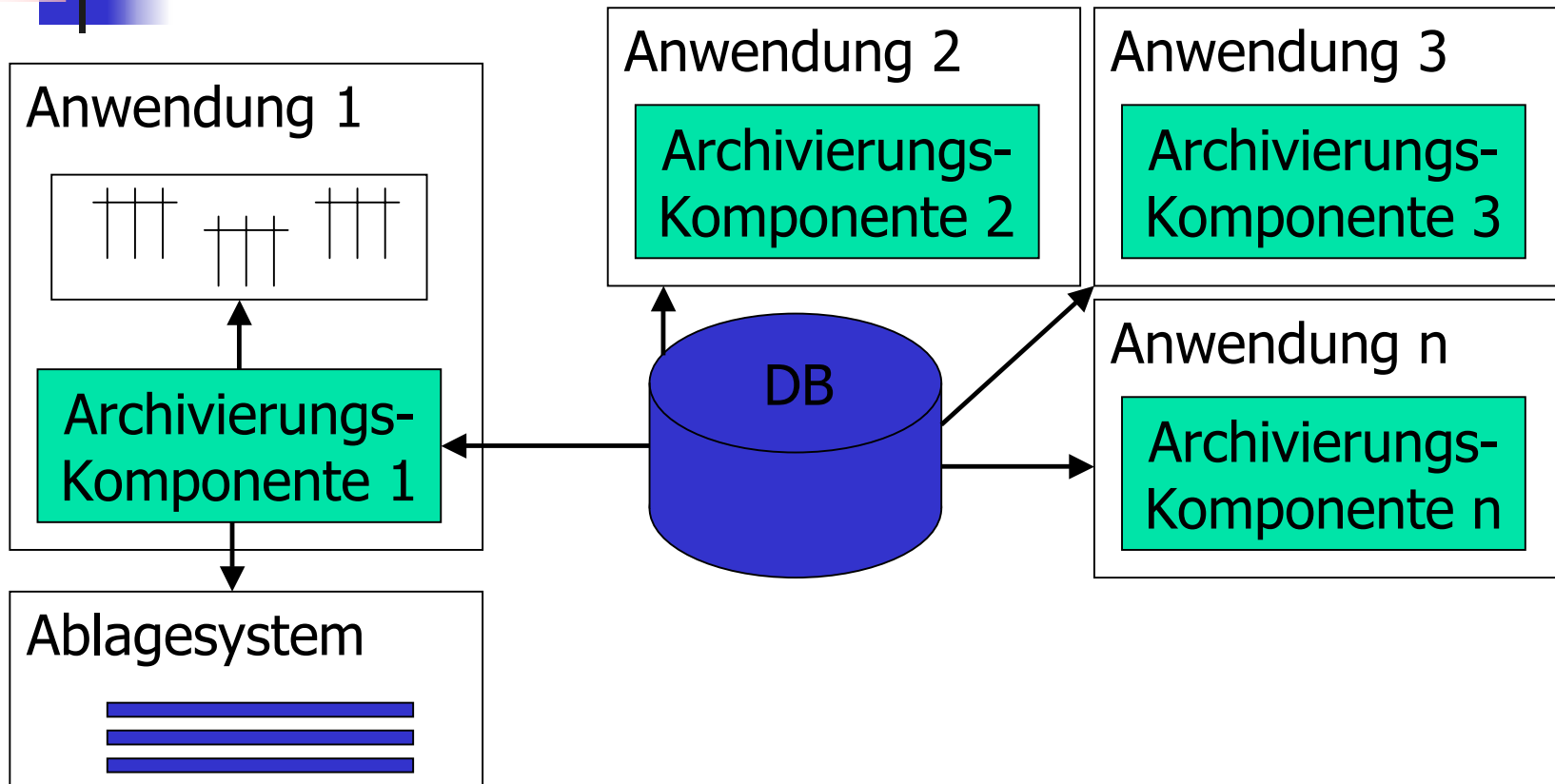


# Archivierung bisher



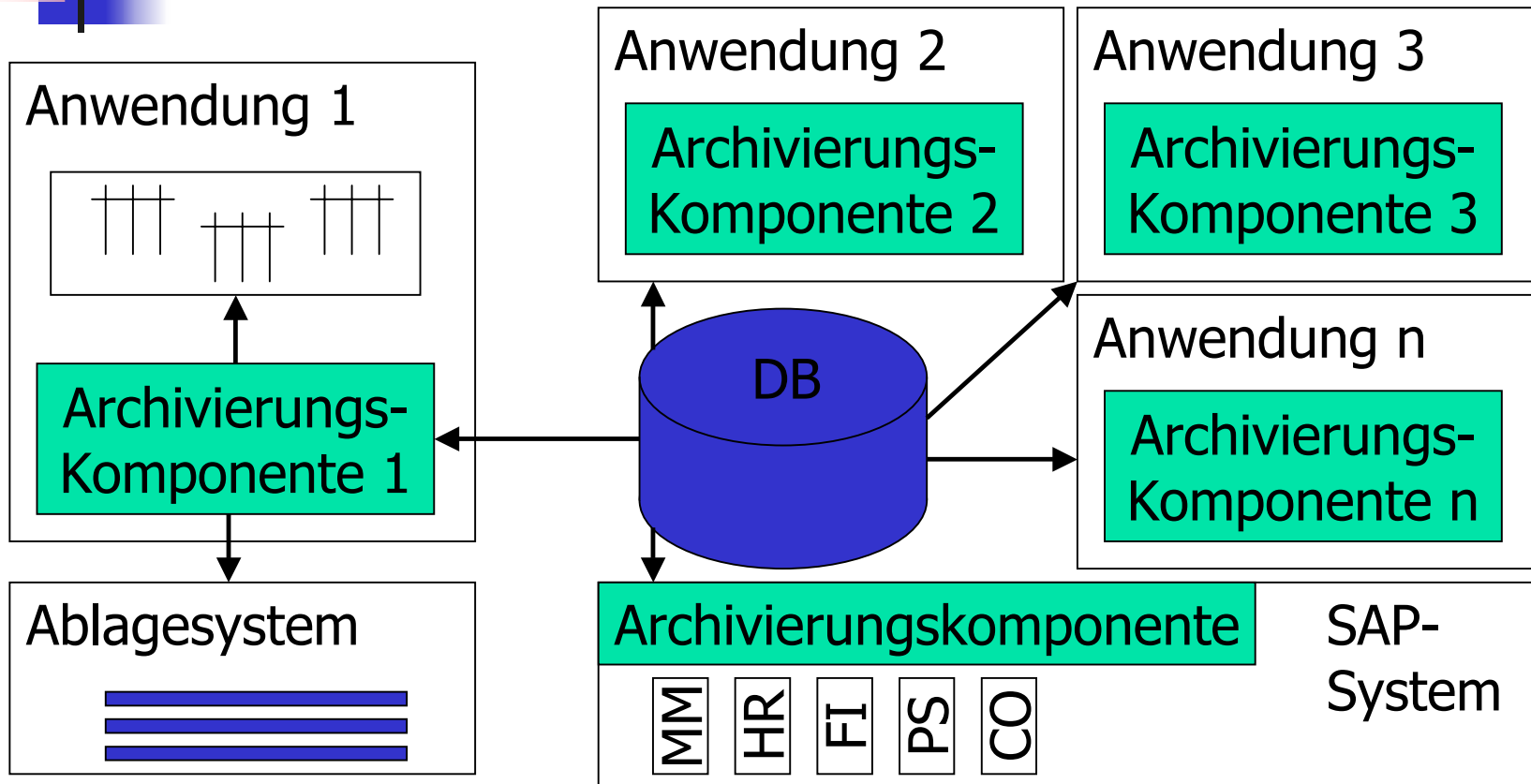
26.02.2003

# Archivierung bisher



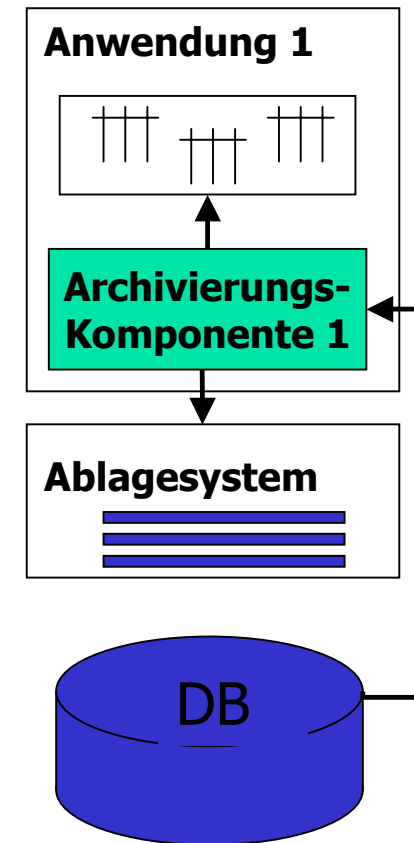


# Archivierung bisher



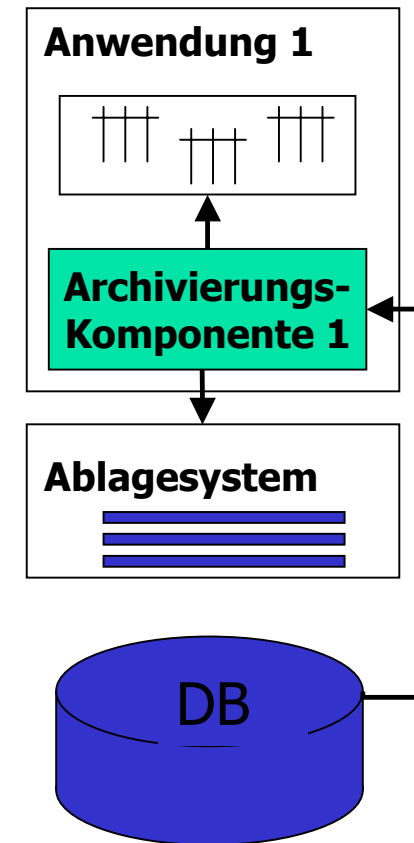
# Probleme

- Vielzahl von Archivierungskomponenten
- Daten in Ablage ohne Anwendung nicht lesbar
- Hohe Netzlast (DB → Anwendung → Ablage)
- Definition des betriebswirtschaftlichen Objektes nur in Anwendung verfügbar
- Komplexe Prüfungen für Archivierbarkeit notwendig → Nur in Anwendung



# Probleme

- **Vielzahl von Archivierungskomponenten**
- **Daten in Ablage ohne Anwendung nicht lesbar**
- Hohe Netzlast (DB → Anwendung → Ablage)
- Definition des betriebswirtschaftlichen Objektes nur in Anwendung verfügbar
- Komplexe Prüfungen für Archivierbarkeit notwendig → Nur in Anwendung





# Lösung

---

- Zentrale Archivierungskomponente,
- die Teil der Datenbank ist und
- die Daten als XML-Dateien ablegt.



# Verwandte Arbeiten

---

- K. Küspert and R. Schaarschmidt. ***Archivierung in Datenbanksystemen***, Informatik Spektrum, 1998
- R. Schaarschmidt, ***Archivierung in Datenbanksystemen: Konzept und Sprache***, Verlag B.G. Teubner, 2001.
- M. J. Carey, D. Florescu, Z. G. Ives, Y. Lu, J. Shanmugasundaram, E. J. Shekita, and S. N. Subramanian. ***XPERANTO: Publishing Object-Relational Data as XML***, In *Proc. of the Third International Workshop on the Web and Databases*, 2000.
- M.F. Fernandez, W.-C. Tan, and D. Suciu. ***SilkRoute: Trading between Relations and XML***. In Int'l World Wide Web Conf. (WWW), 2000.



# Probleme

---

- Vielzahl von Archivierungskomponenten ✓ nur eine
- Daten in Ablage ohne Anwendung nicht lesbar ✓ XML
- Hohe Netzlast (DB→Anwendung→Ablage) ✓ innerhalb DB
- Definition des betriebswirtschaftlichen Objektes nur in Anwendung verfügbar
- Komplexe Prüfungen für Archivierbarkeit notwendig → Nur in Anwendung



# Probleme

---

- Vielzahl von Archivierungskomponenten ✓ nur eine
- Daten in Ablage ohne Anwendung nicht lesbar ✓ XML
- Hohe Netzlast (DB→Anwendung→Ablage) ✓ innerhalb DB
- Definition des betriebswirtschaftlichen Objektes nur in Anwendung verfügbar ✓ XML Schema
- Komplexe Prüfungen für Archivierbarkeit notwendig → Nur in Anwendung



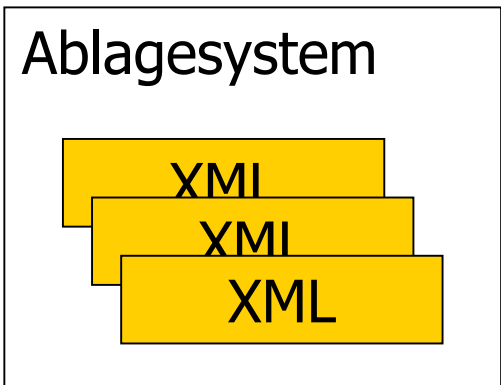
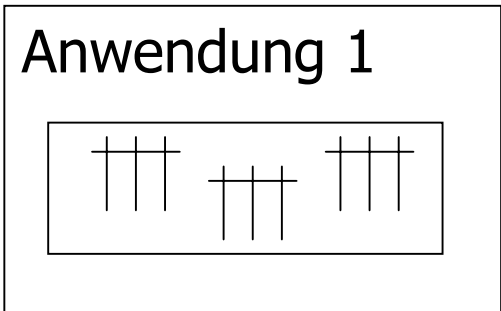
# Probleme

---

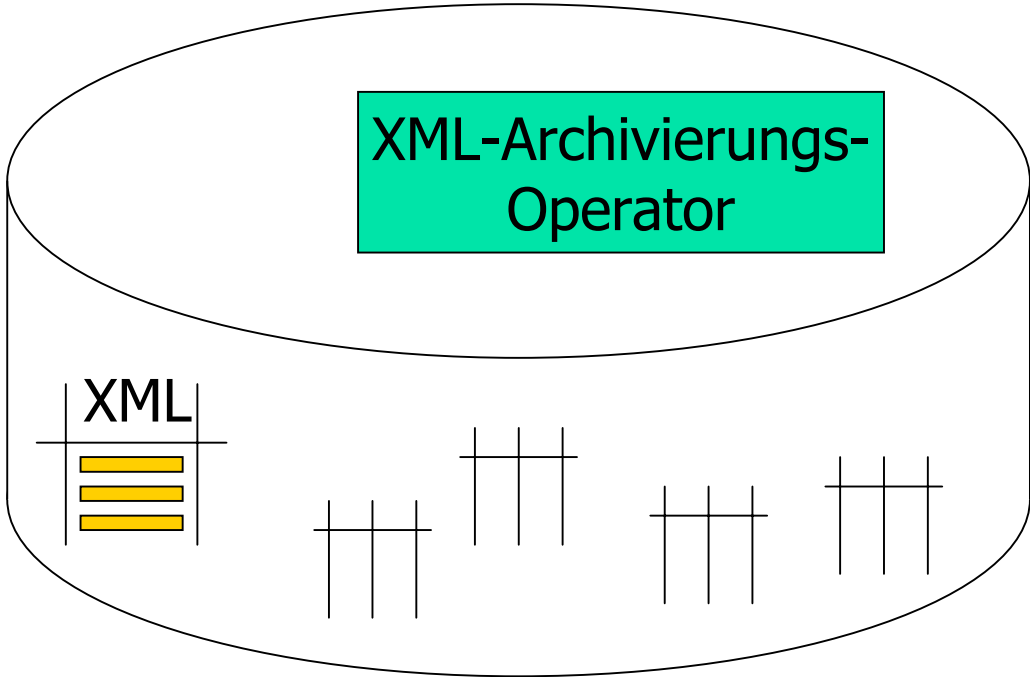
- Vielzahl von Archivierungskomponenten ✓ nur eine
- Daten in Ablage ohne Anwendung nicht lesbar ✓ XML
- Hohe Netzlast (DB→Anwendung→Ablage) ✓ innerhalb DB
- Definition des betriebswirtschaftlichen Objektes nur in Anwendung verfügbar ✓ XML Schema
- Komplexe Prüfungen für Archivierbarkeit notwendig → Nur in Anwendung ✓ temporäre Tabellen



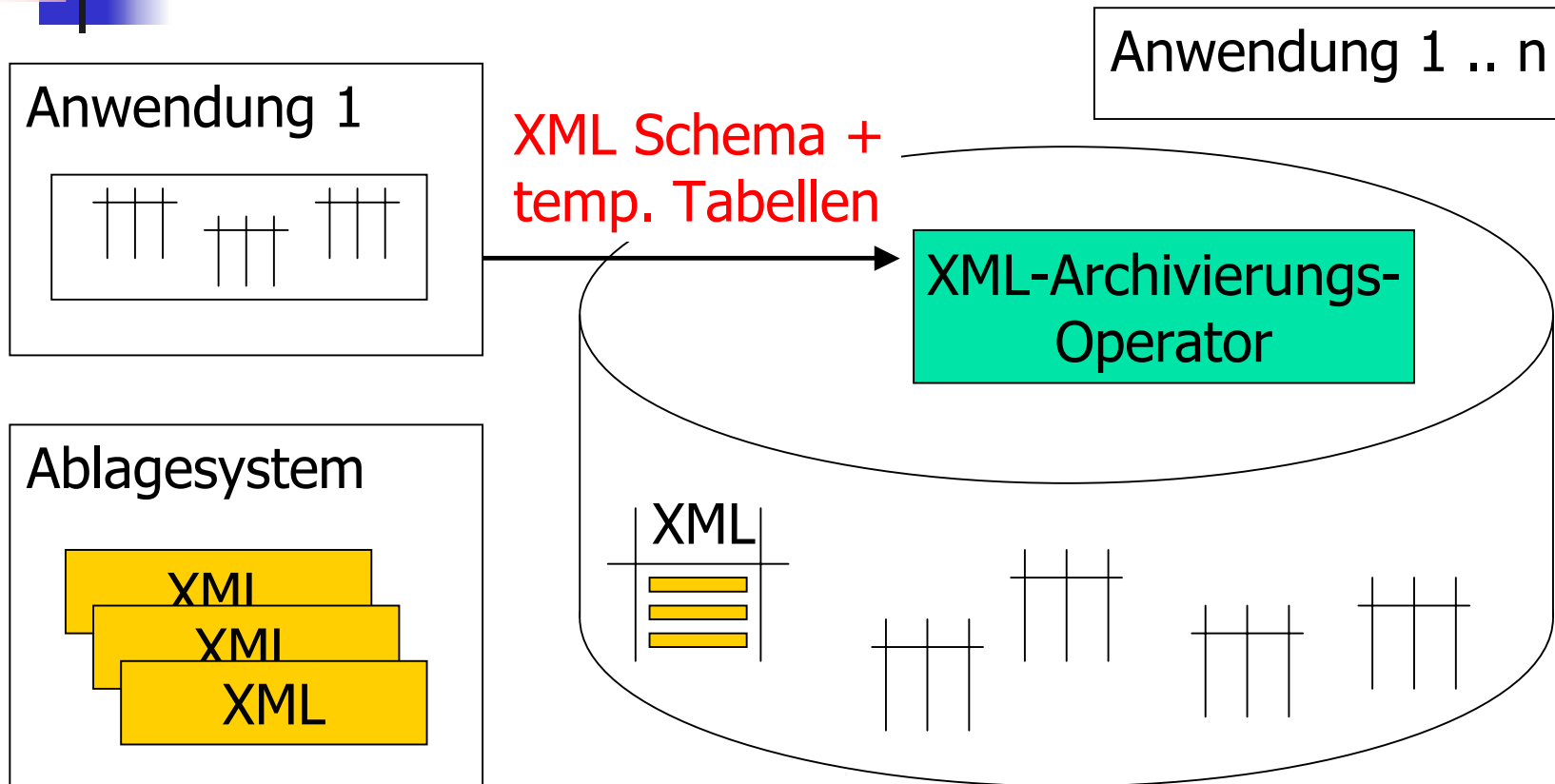
# XML Archivierung



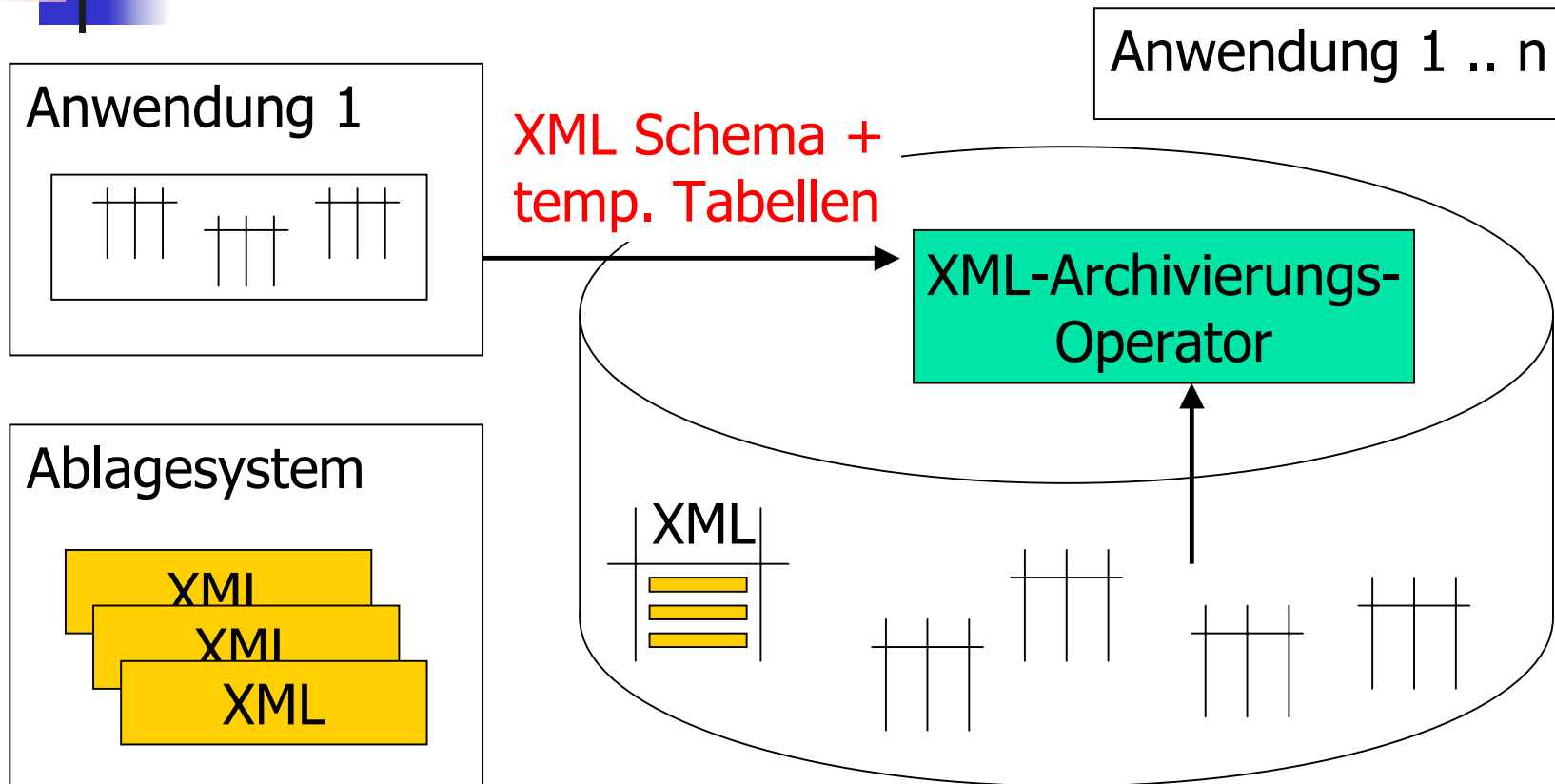
Anwendung 1 .. n



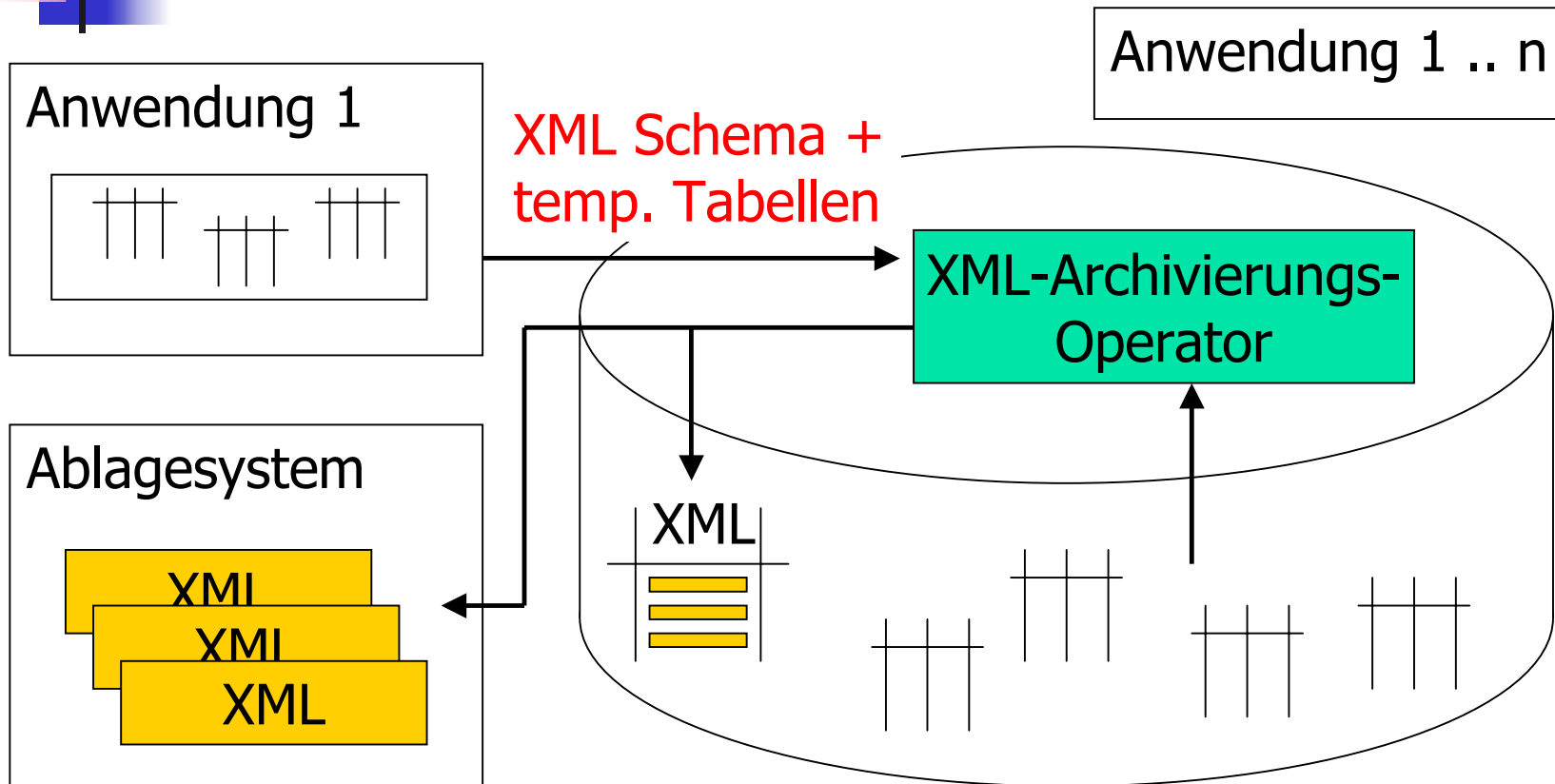
# XML Archivierung



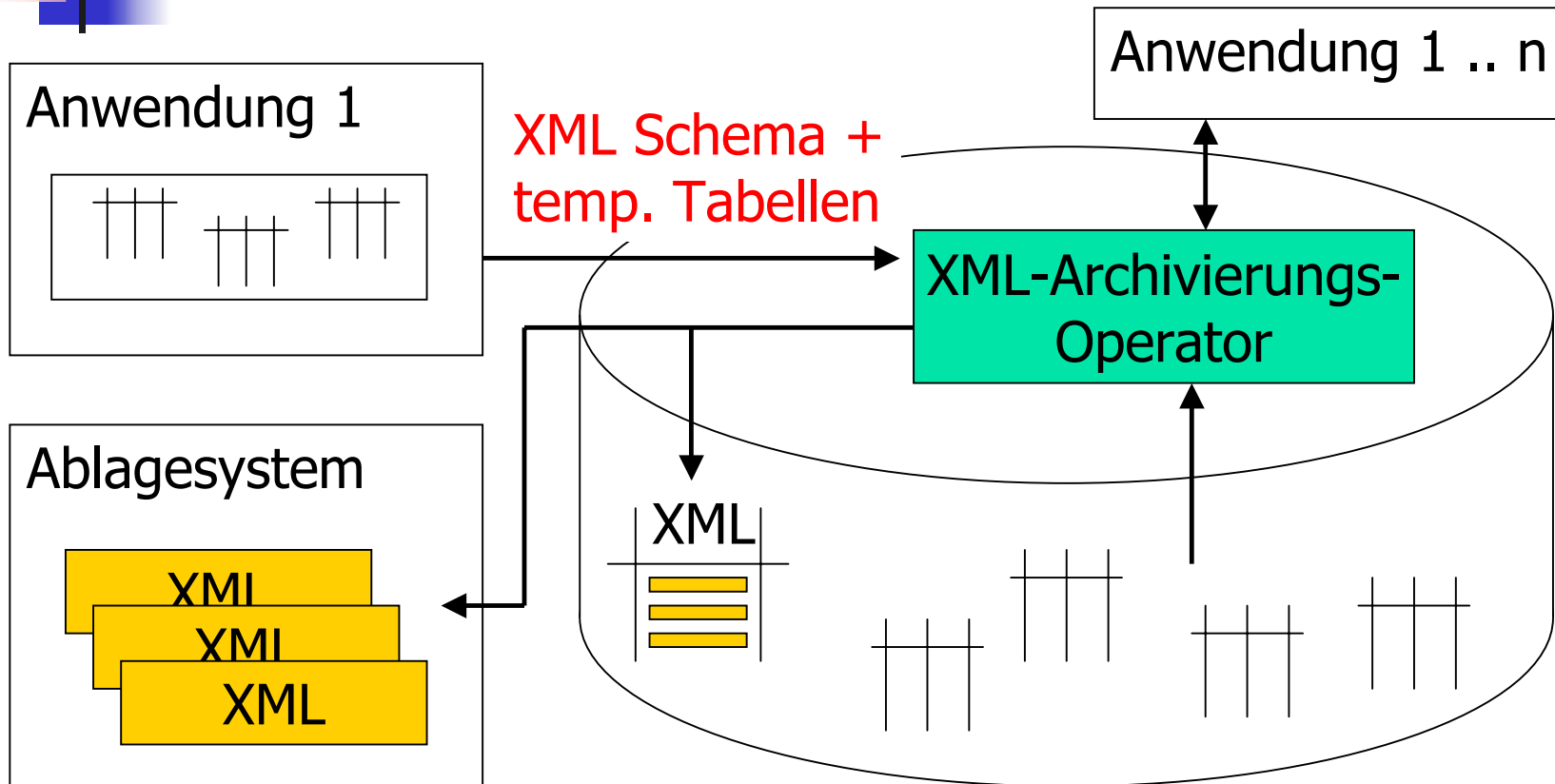
# XML Archivierung



# XML Archivierung



# XML Archivierung





# XML Schema Parameter

---

- Definition des betriebswirtschaftlichen Objektes, d.h. auf welche Tabellen sind die Daten eines Objektes verteilt
- bestimmt das Aussehen der erzeugten XML Dateien
- Einbinden der Daten mittels Annotationen:

```
<xsd:annotation>  
  <xsd:appinfo>  
    sql:[Schema].[Tabelle].[Spalte]  
  </xsd:appinfo>  
</xsd:annotation>
```



# Temporäre Tabellen

- enthalten die Schlüsselwerte der zu archivierenden Tabelleneinträge
- Kennzeichnen, welche Tabelleneinträge zu welchem Objekt gehören:

Bestellung (**BID**, Kunde, Datum, Summe)

Bestellposition (**BID**, **Zeile**, Ware, Anz, Preis)

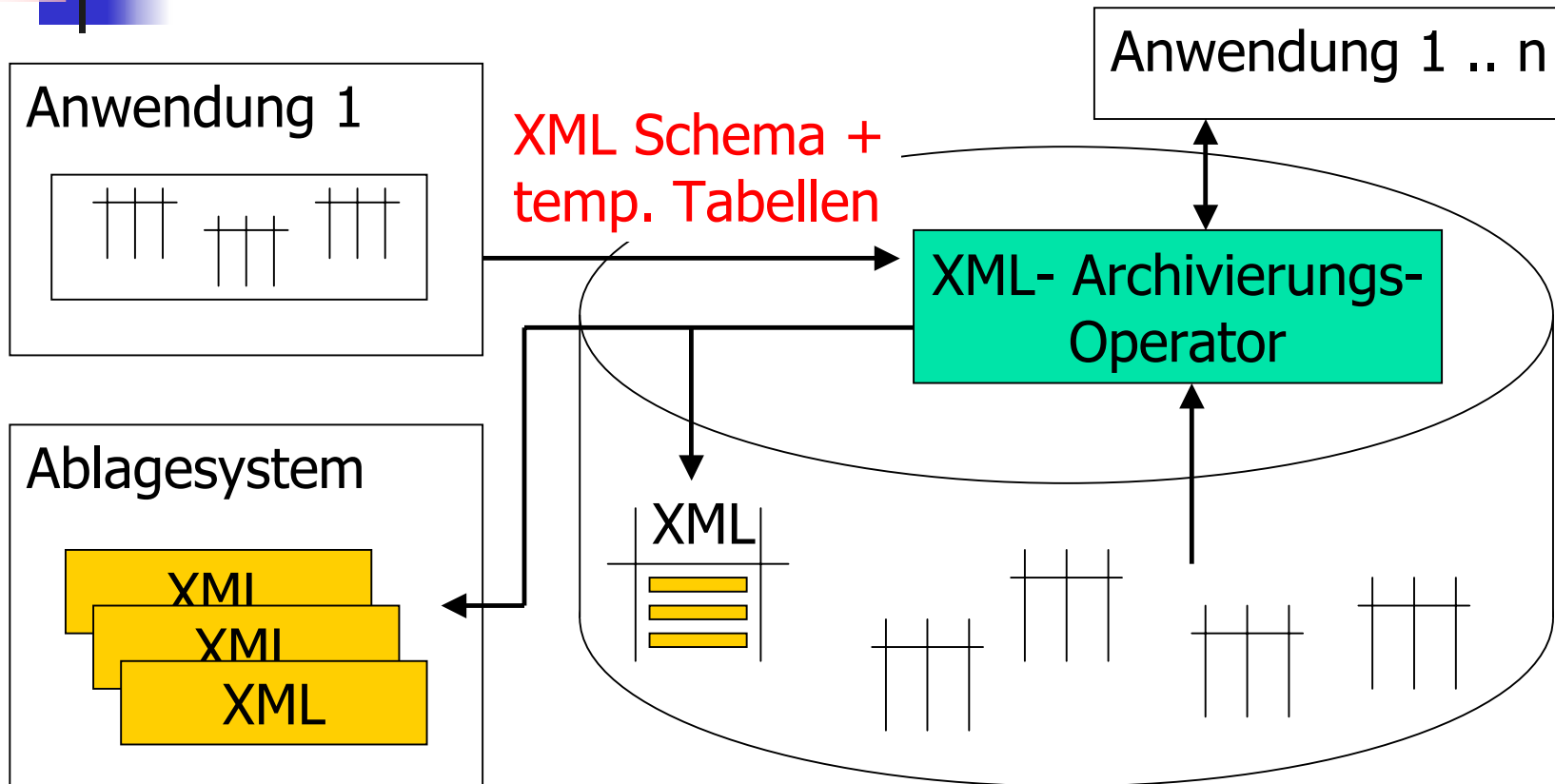
Arch\_Bestellung

ArchKey	BID
Obj_1	1
Obj_2	3

Arch\_Bestellposition

ArchKey	BID	Zeile
Obj_1	1	1
Obj_1	1	2
Obj_2	3	1

# XML Archivierung







# Beispiel

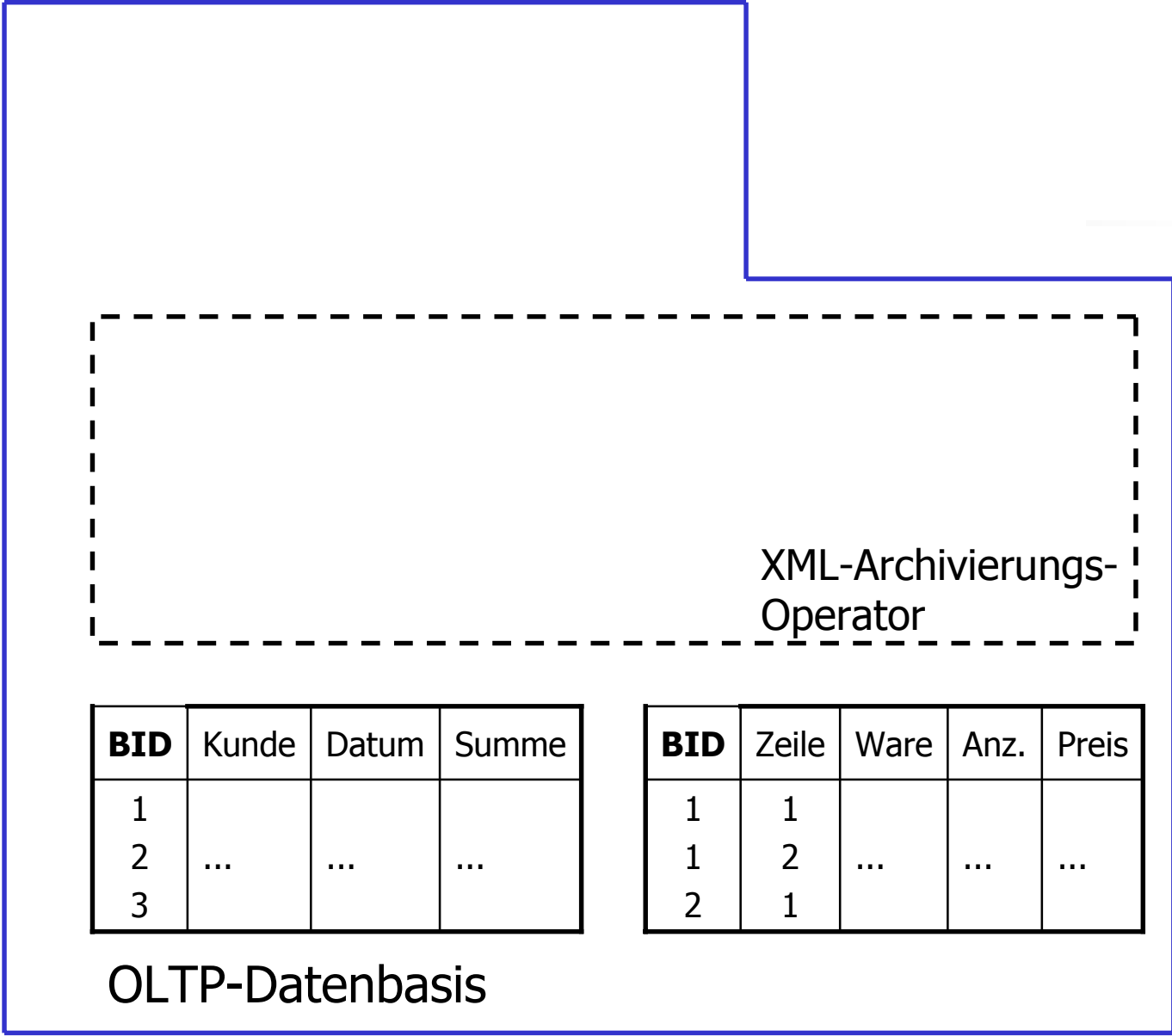
## Bestellung

<b>BID</b>	Kunde	Datum	Summe
1	4711	3.5.02	1200
2	5678	3.5.02	57
3	3456	4.5.02	107
4	5678	4.5.02	10
...	...	...	...

## Bestellpositionen

<b>BID</b>	<b>Zeile</b>	Ware	Anz.	Preis
1	1	Fernseher	1	1000
1	2	DVD-Player	1	200
2	1	Radio	1	57
3	1	Batterien	10	0,10
3	2	Lautsprecher	2	23
3	3	CDs	2	30
...				

Anwendung



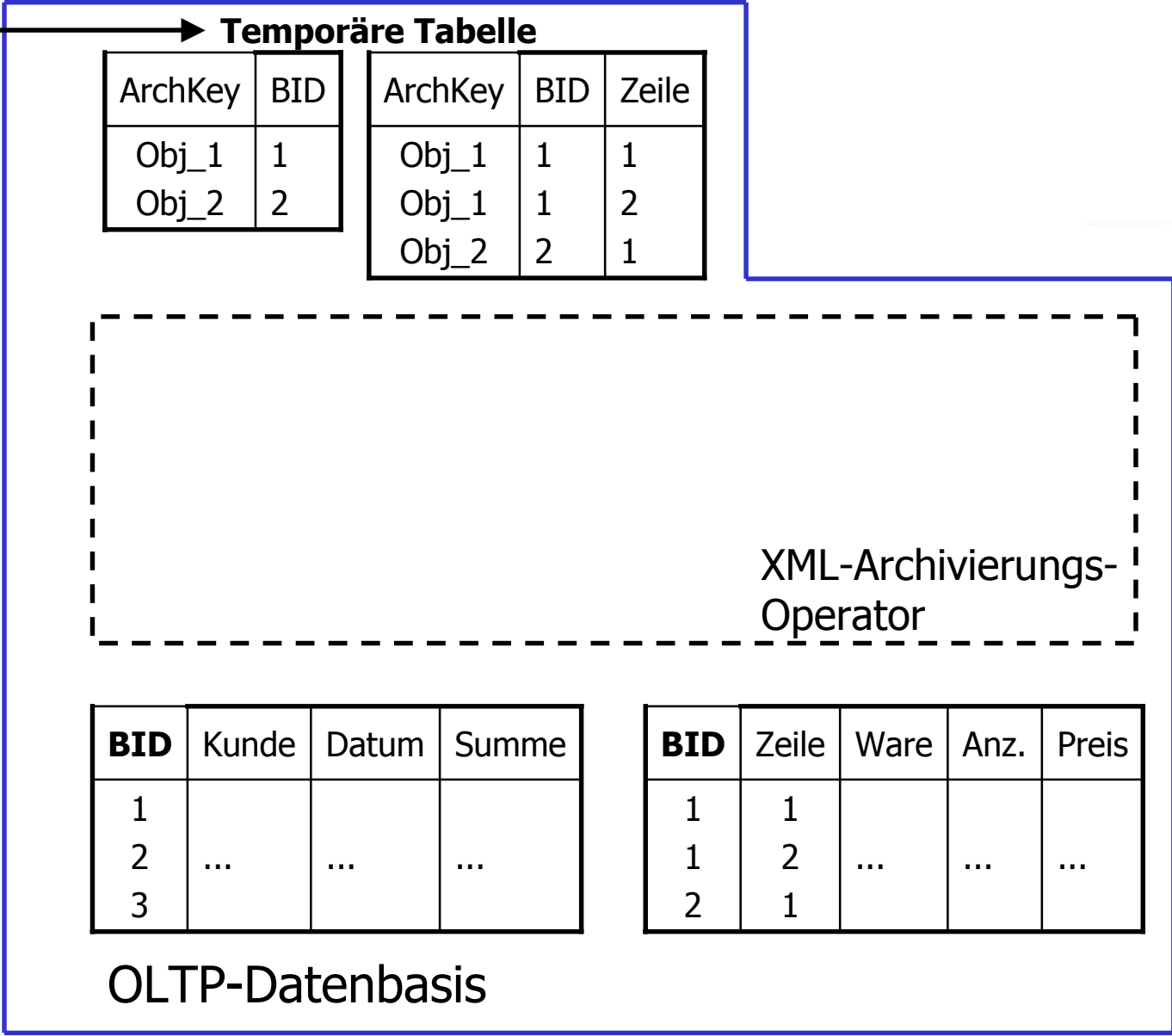
<b>BID</b>	Kunde	Datum	Summe
1			
2	...	...	...
3			

OLTP-Datenbasis

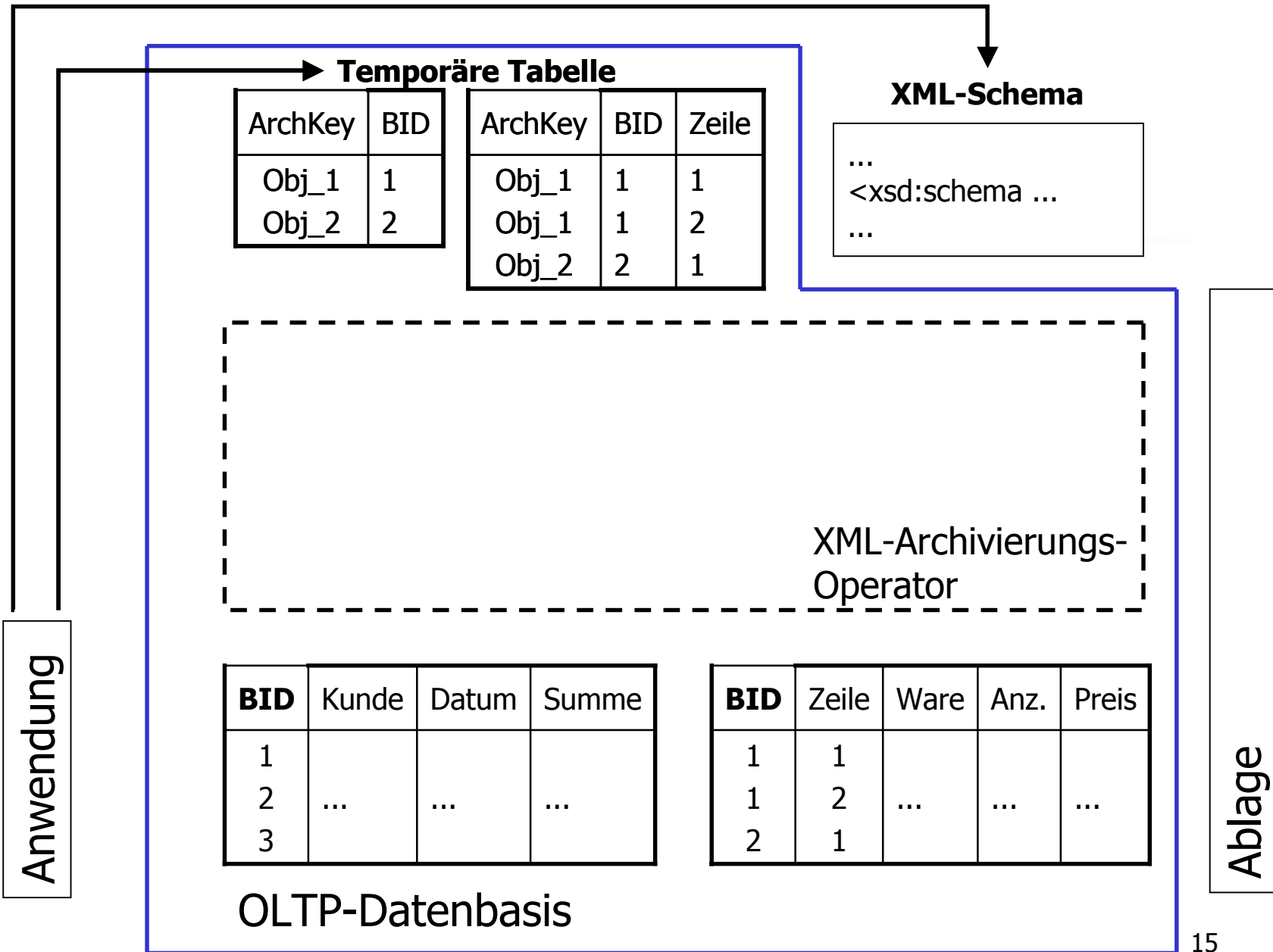
<b>BID</b>	Zeile	Ware	Anz.	Preis
1	1			
1	2	...	...	...
2	1			

Ablage

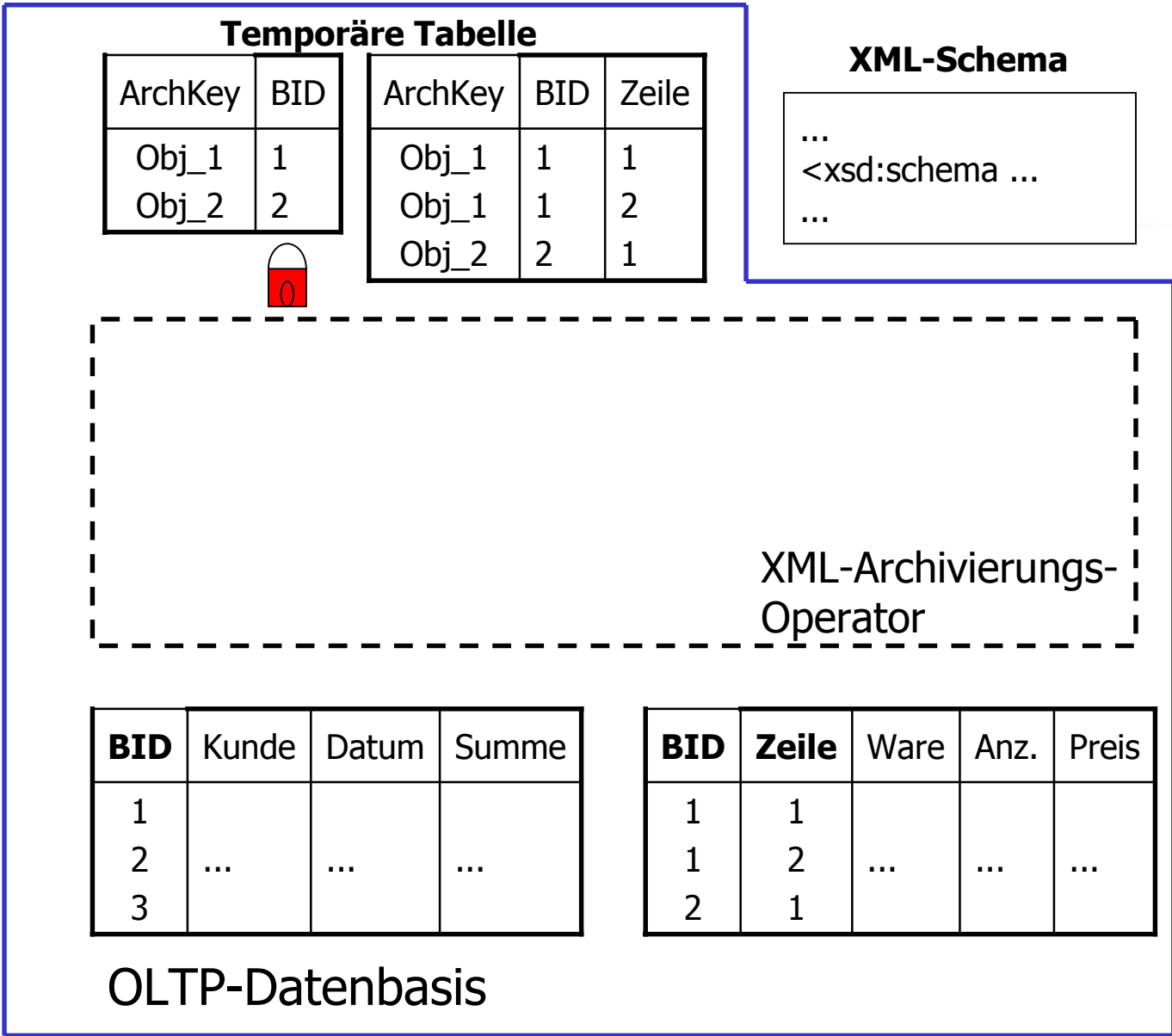
Anwendung



Ablage

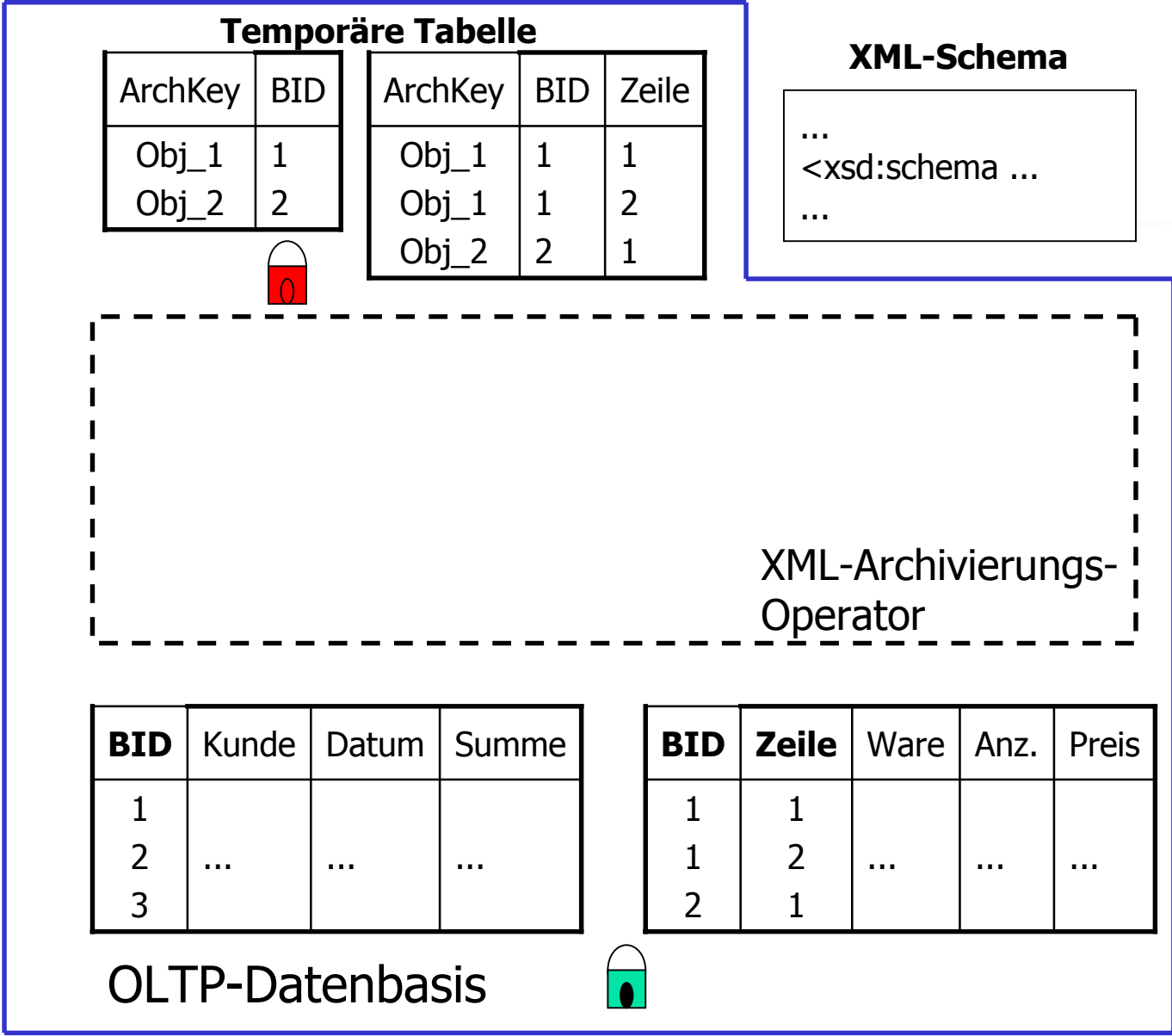


Anwendung



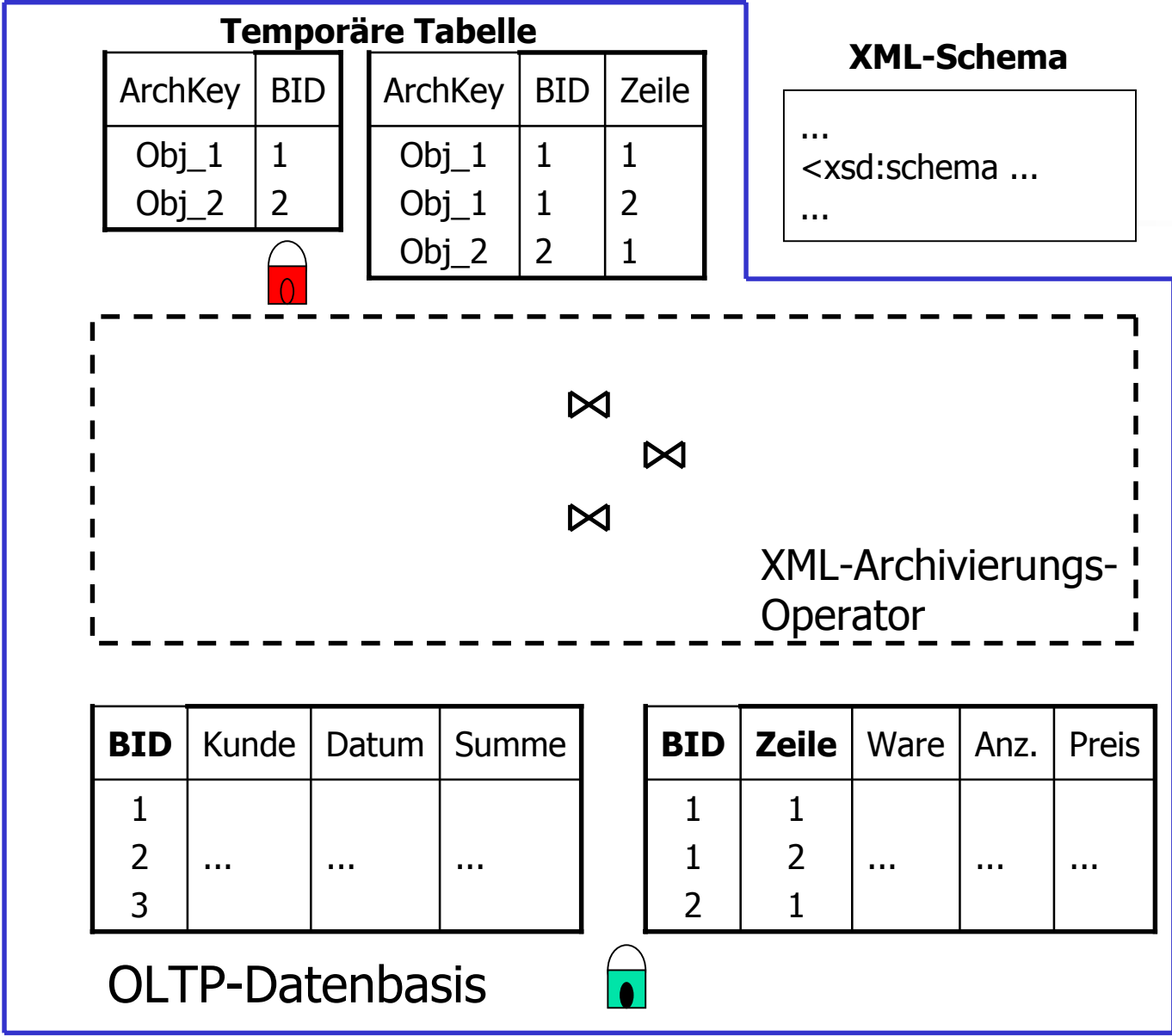
Ablage

Anwendung



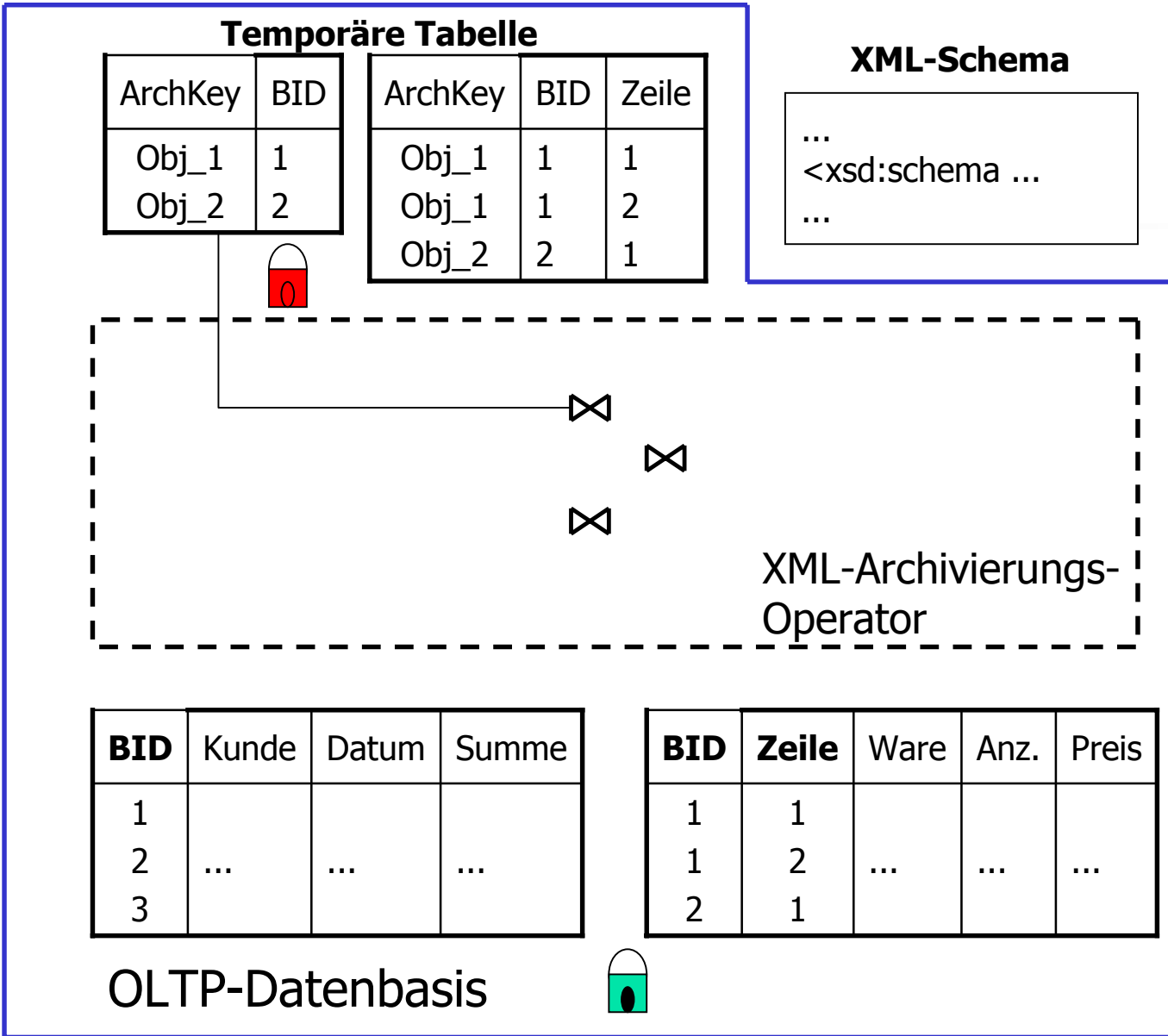
Ablage

Anwendung



Ablage

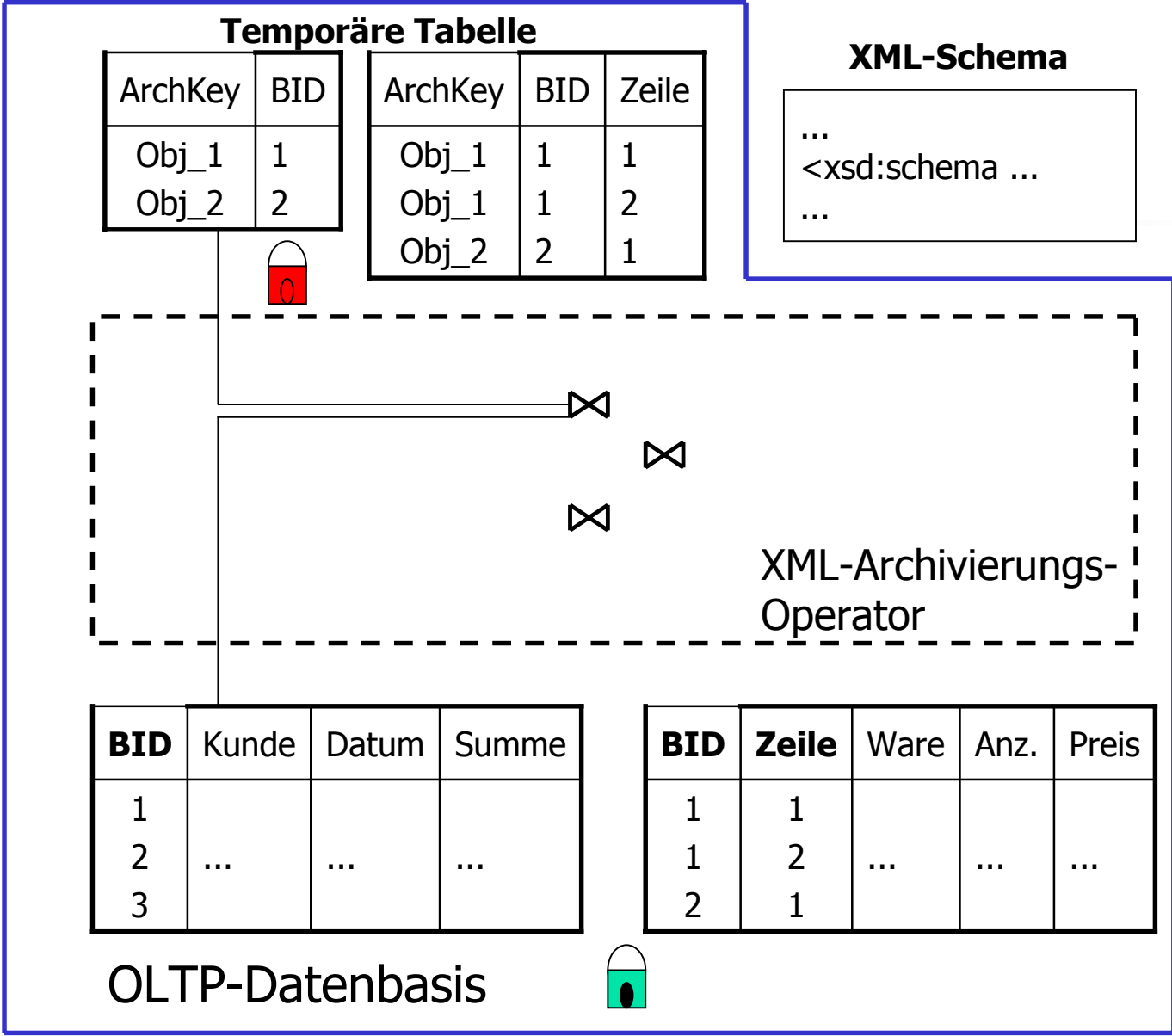
Anwendung



Ablage

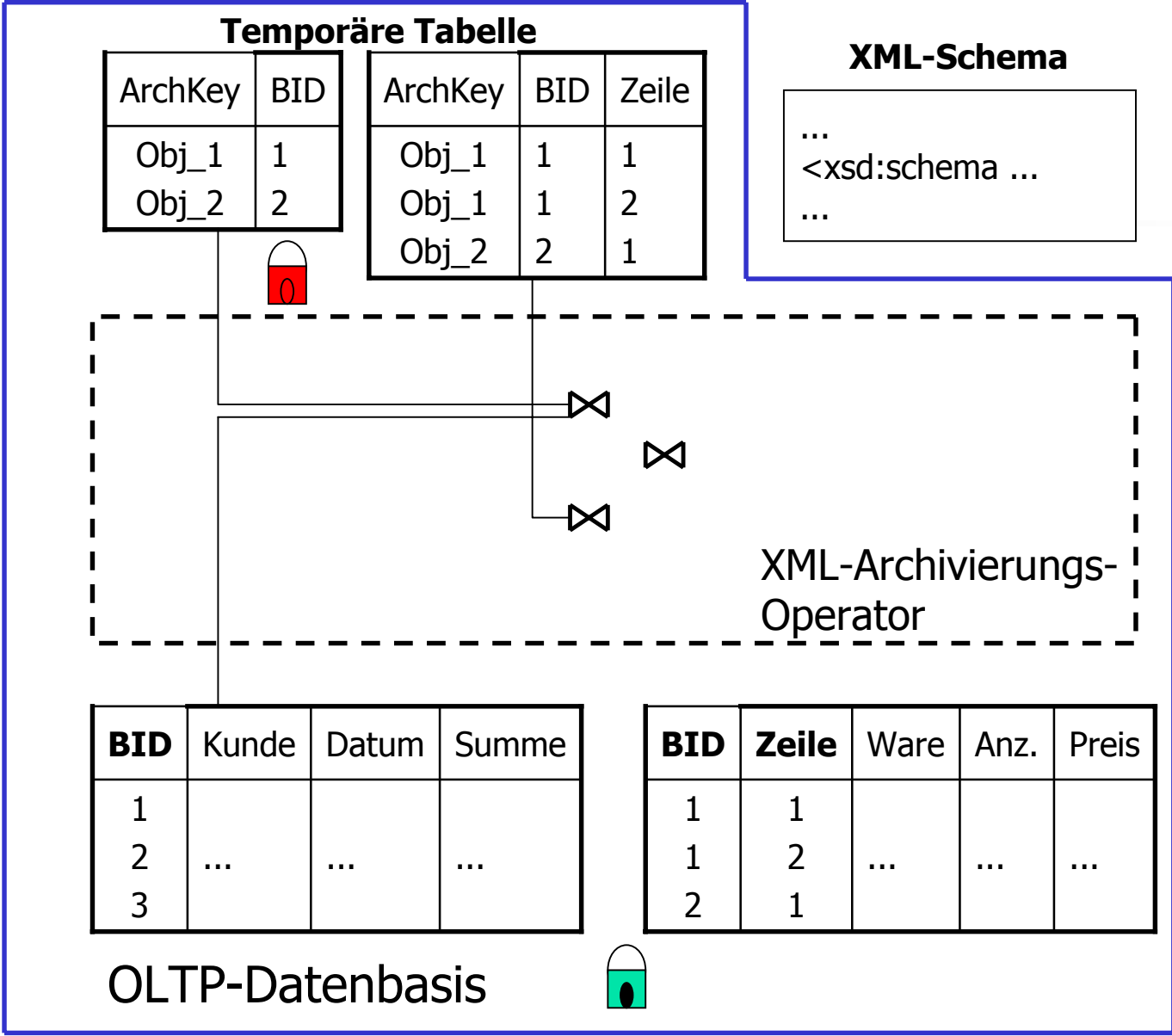


Anwendung



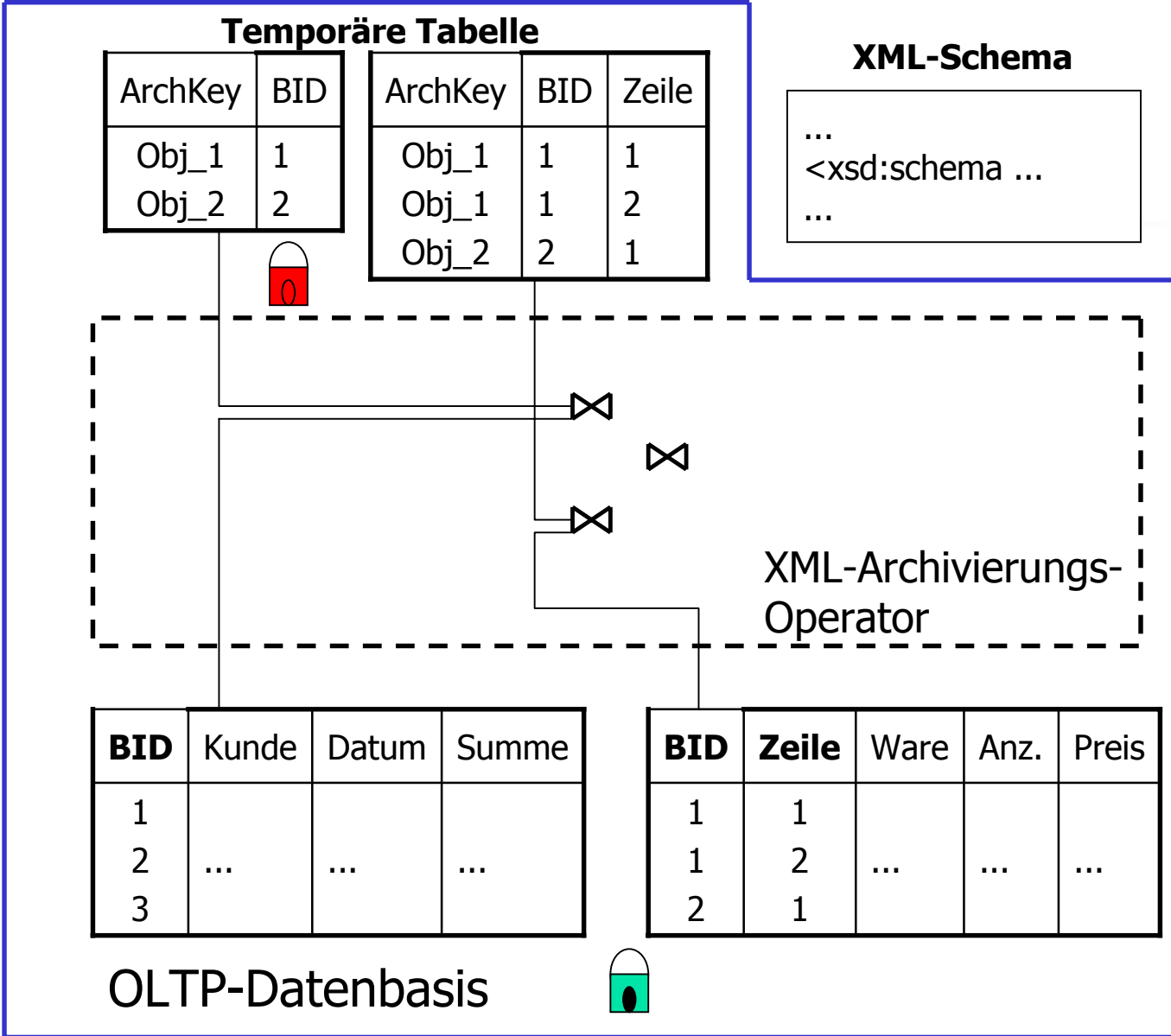
Ablage

Anwendung



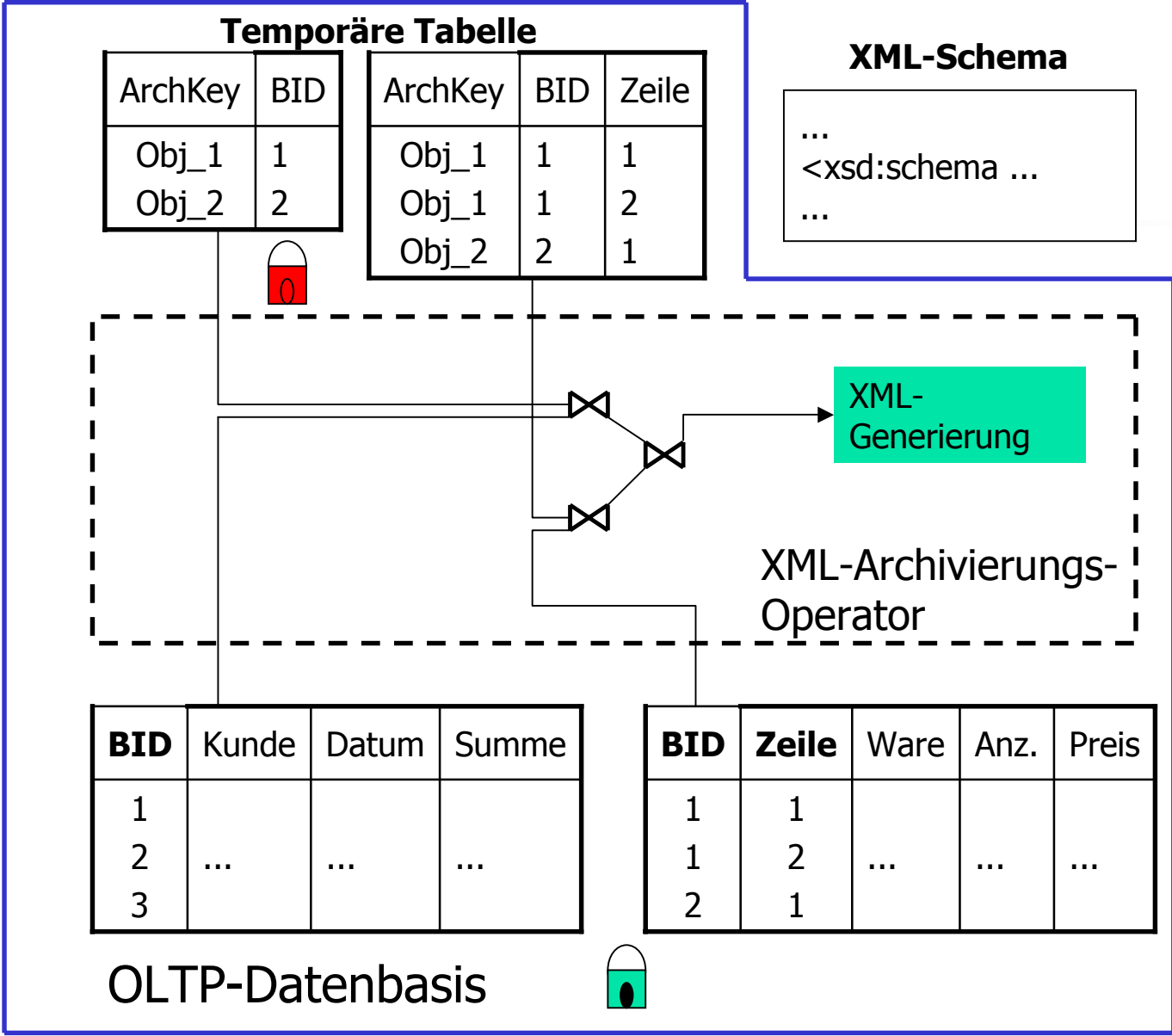
Ablage

Anwendung



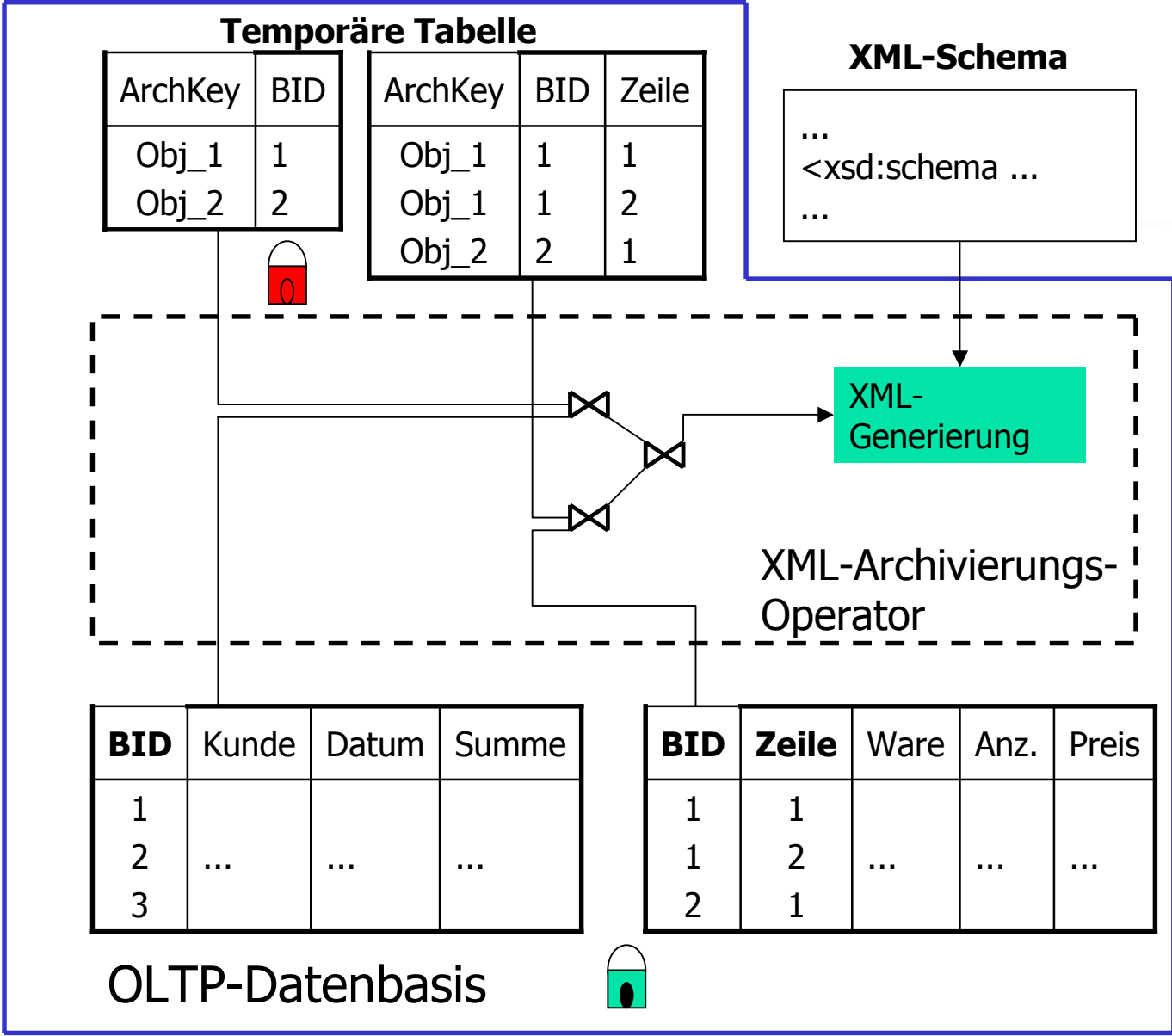
Ablage

Anwendung



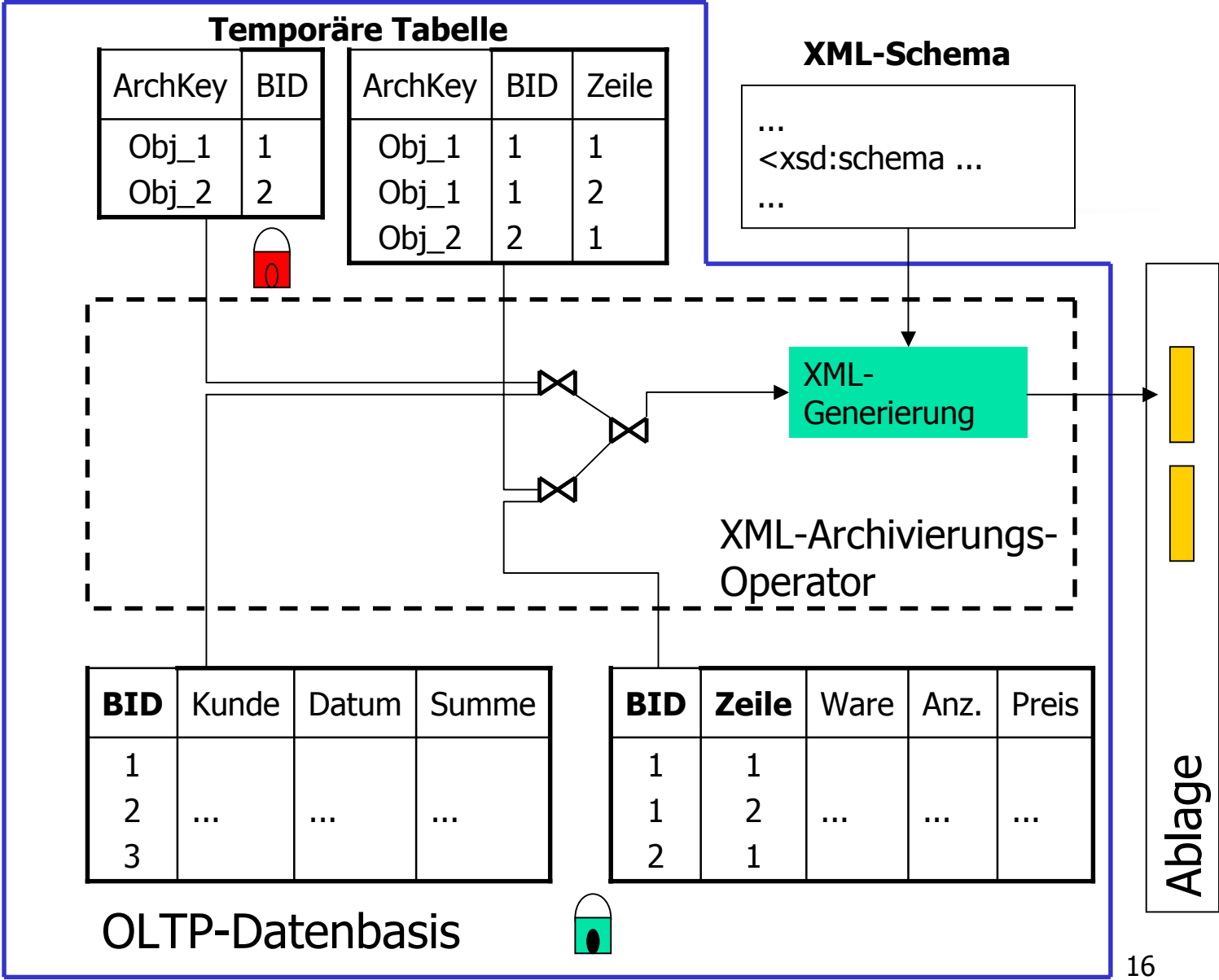
Ablage

Anwendung

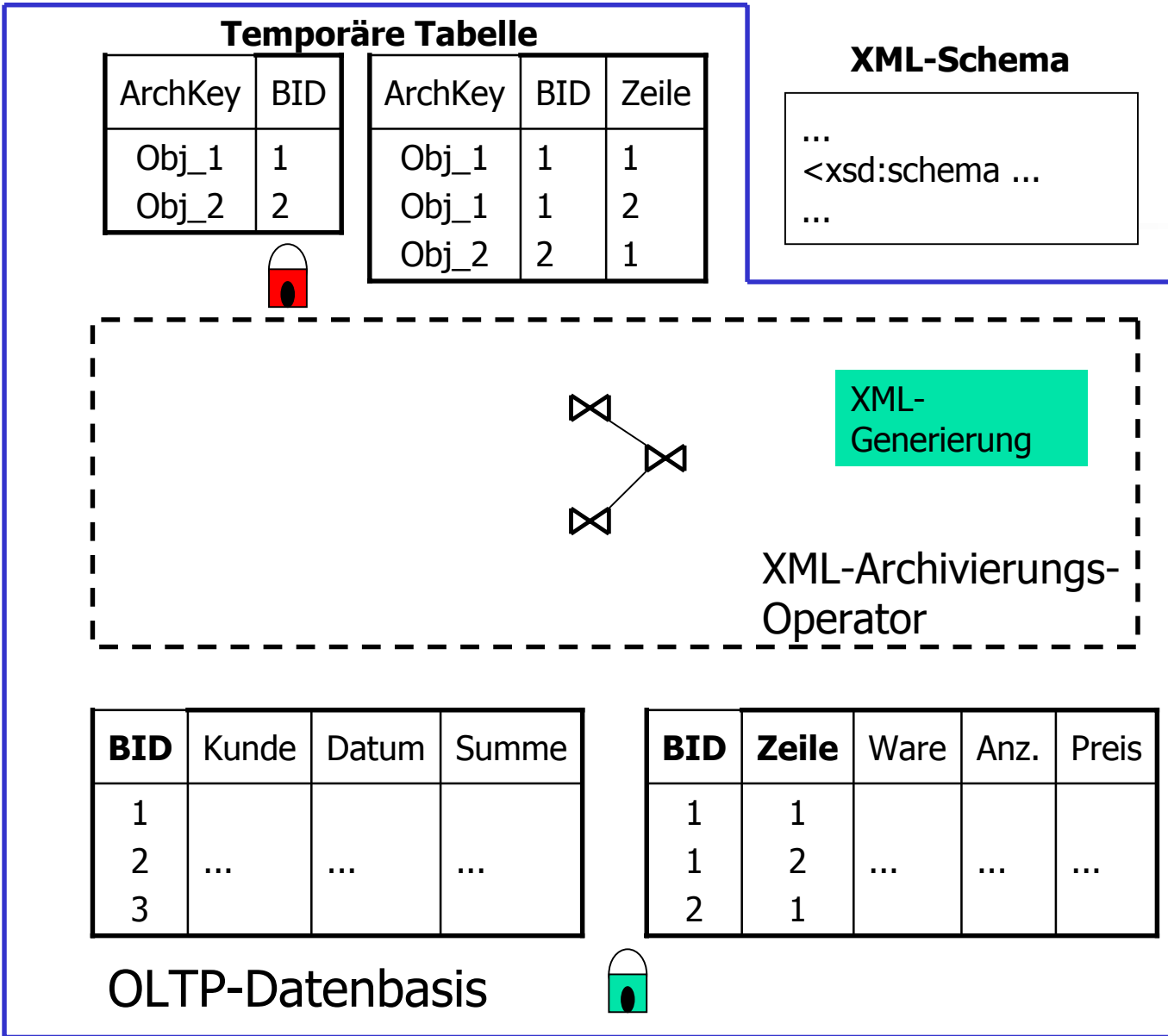


Ablage

Anwendung



Anwendung



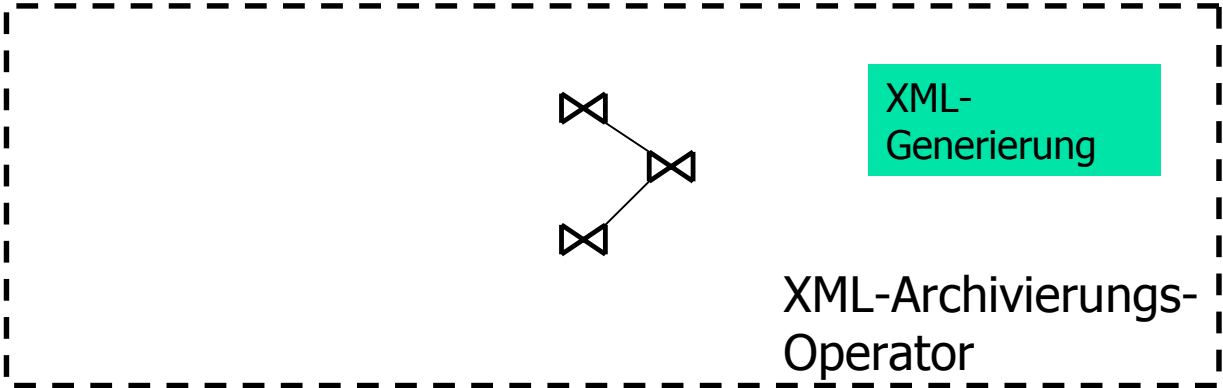
**Temporäre Tabelle**

ArchKey	BID
Obj_1	1
Obj_2	2

ArchKey	BID	Zeile
Obj_1	1	1
Obj_1	1	2
Obj_2	2	1

XML-Schema

```
...  
<xsd:schema ...  
...
```

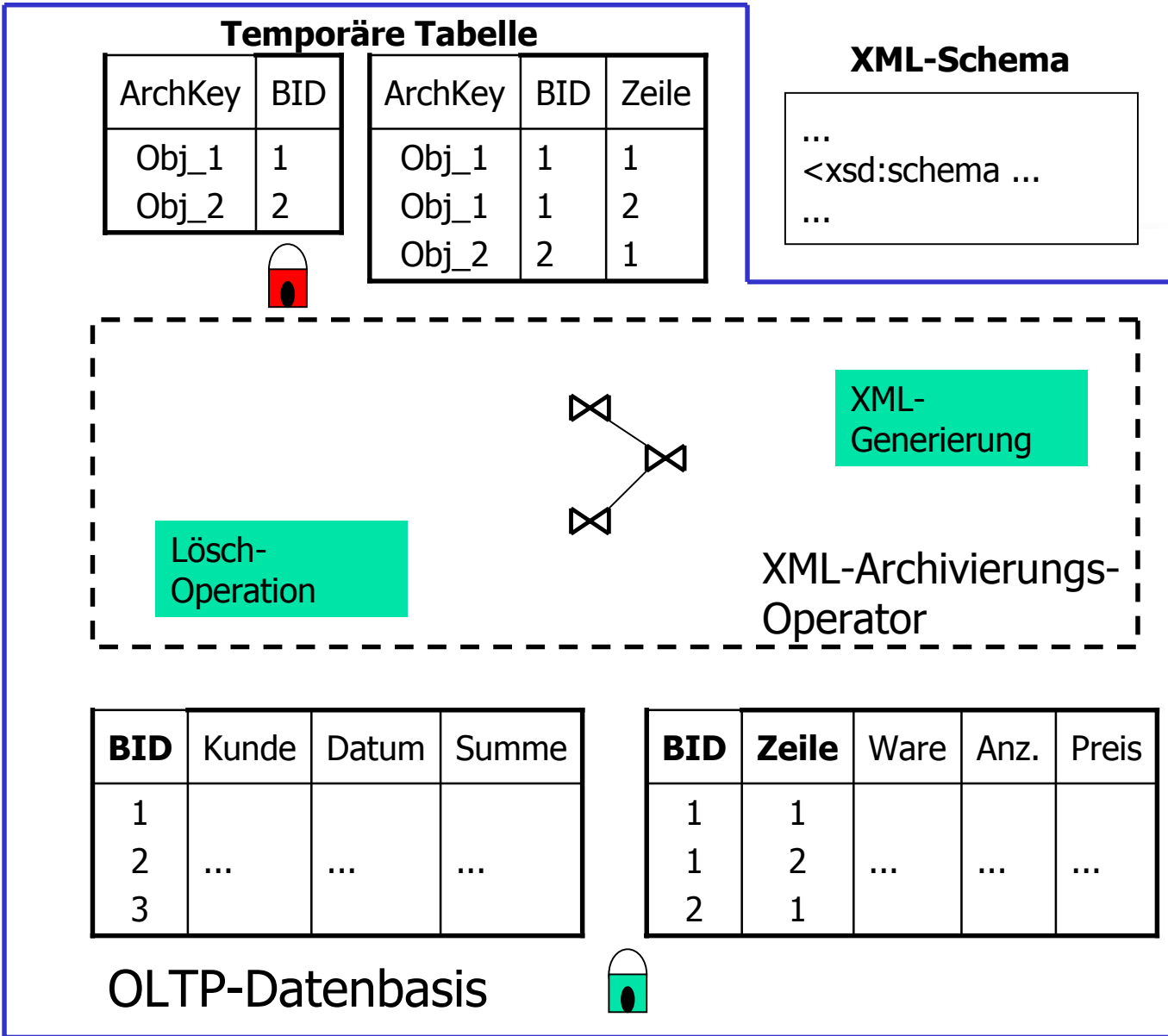


<b>BID</b>	Kunde	Datum	Summe
1			
2	...	...	...
3			

<b>BID</b>	<b>Zeile</b>	Ware	Anz.	Preis
1	1			
1	2	...	...	...
2	1			

Ablage

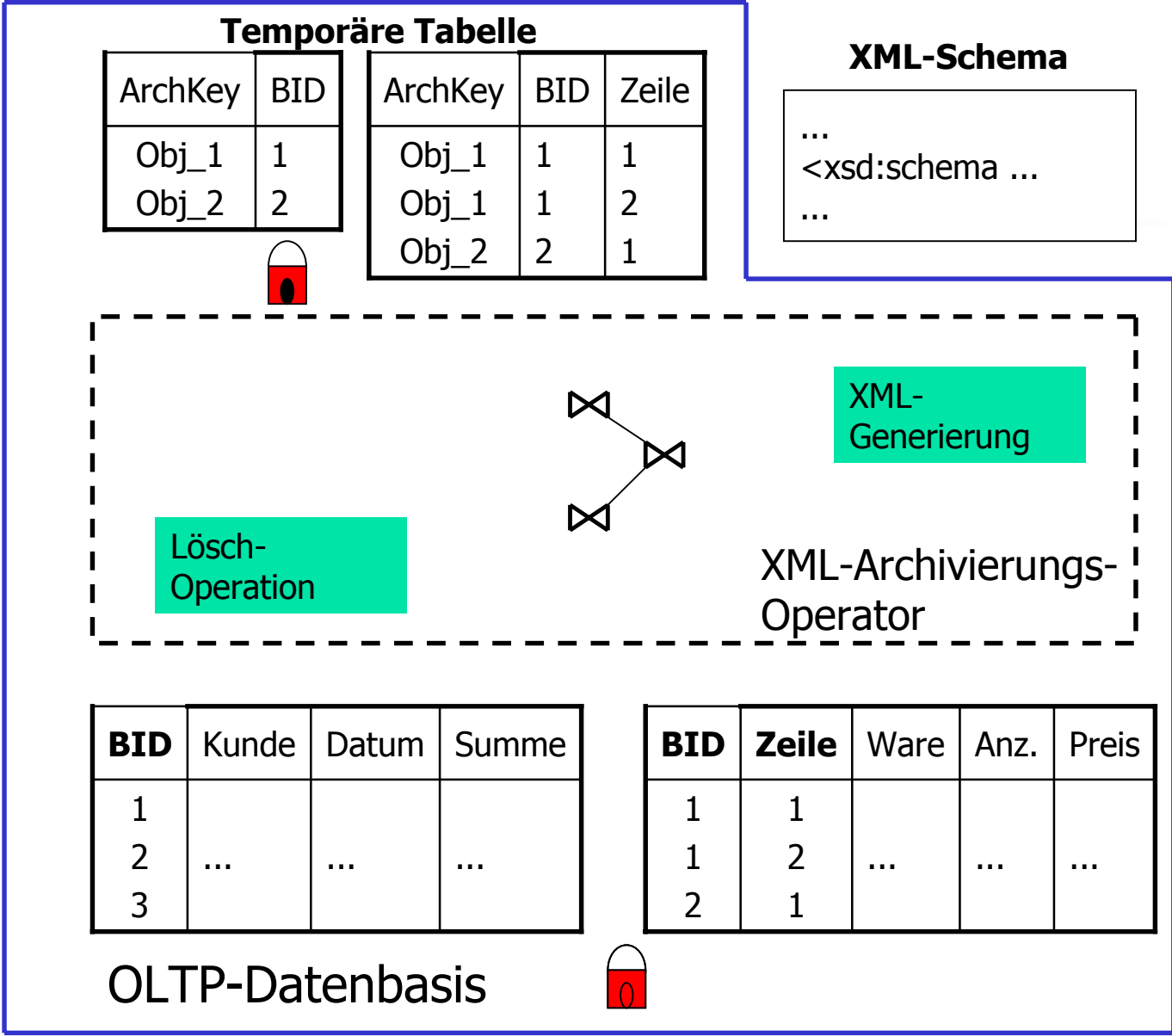
Anwendung



Ablage

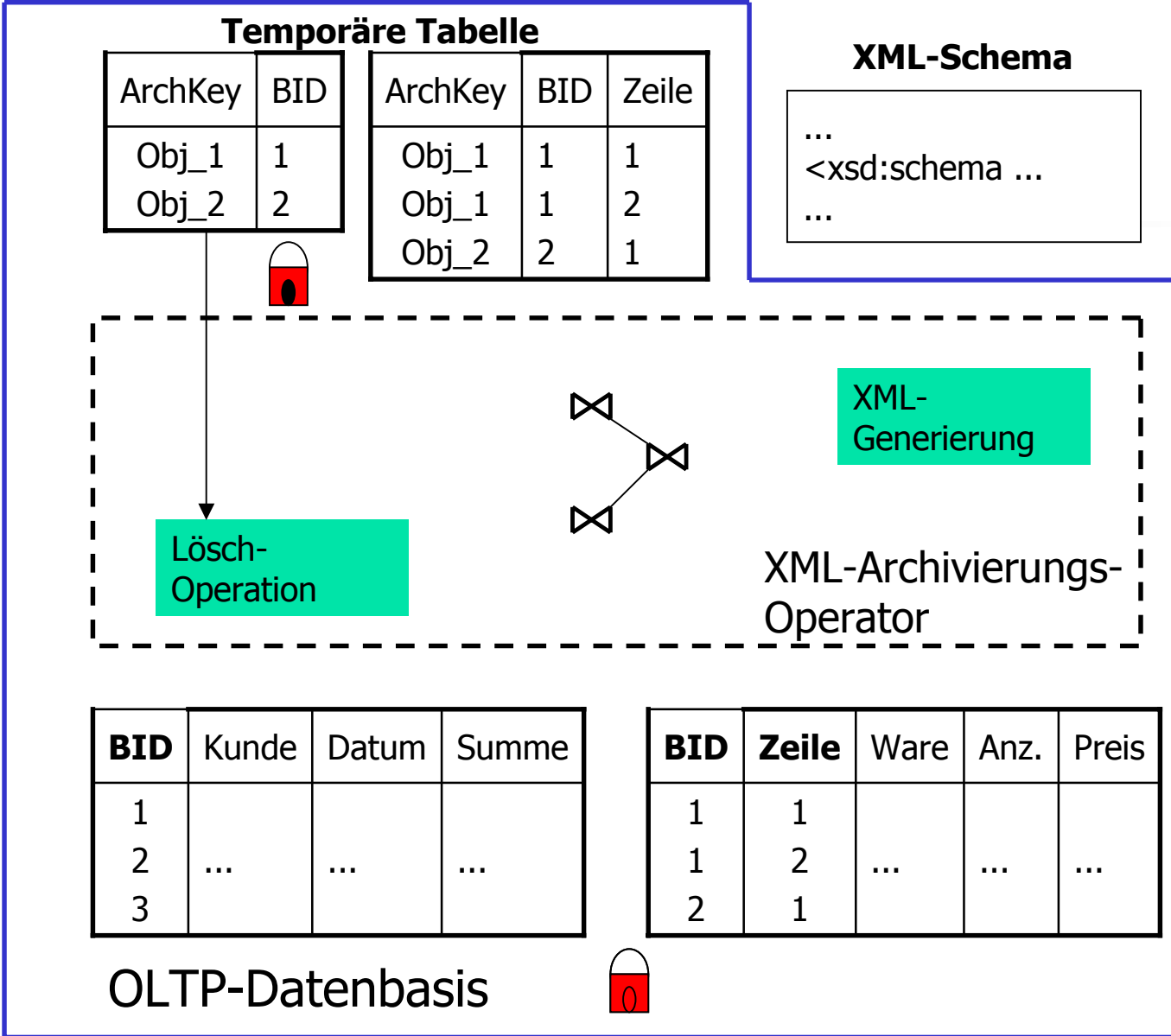


Anwendung



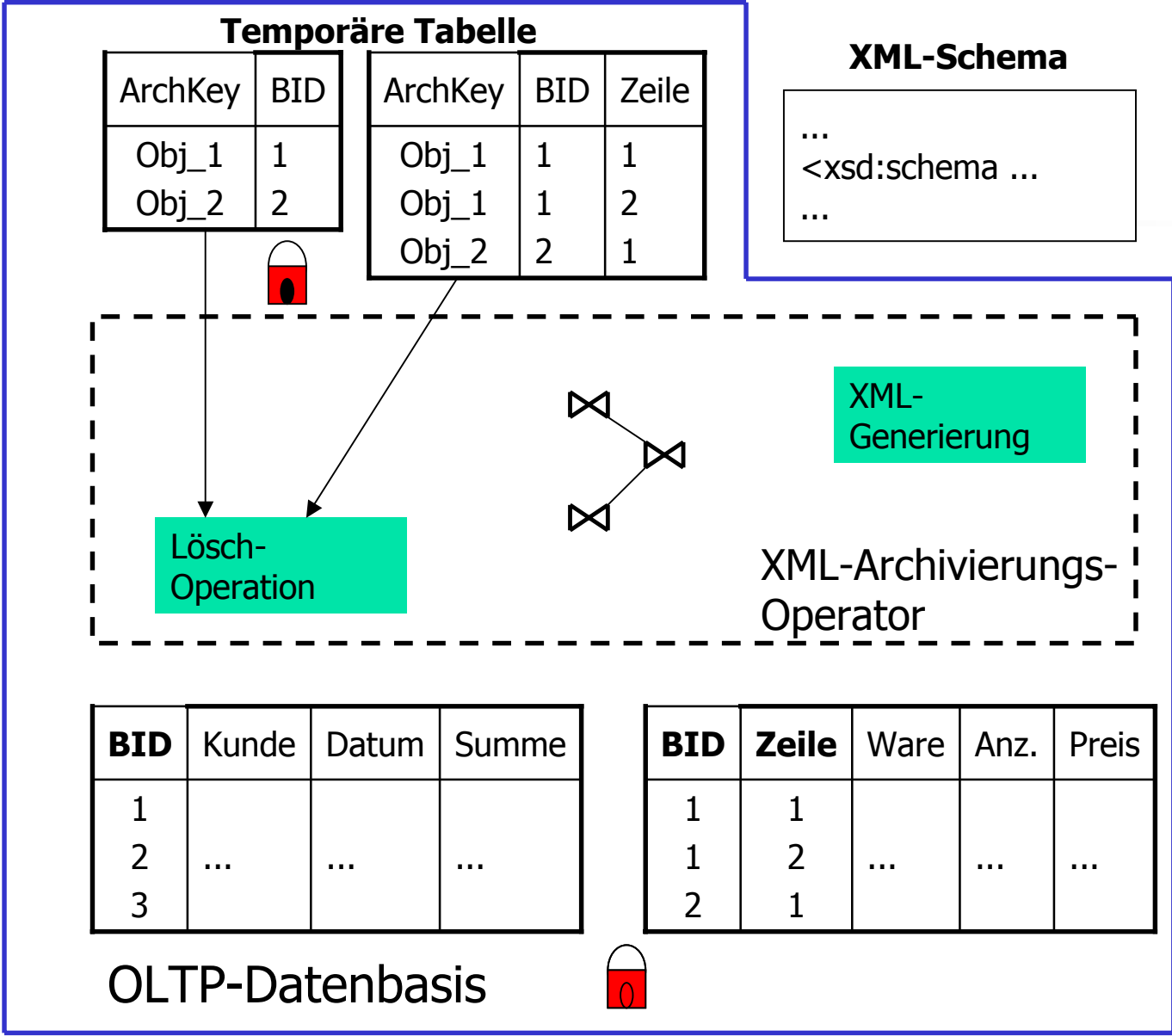
Ablage

Anwendung



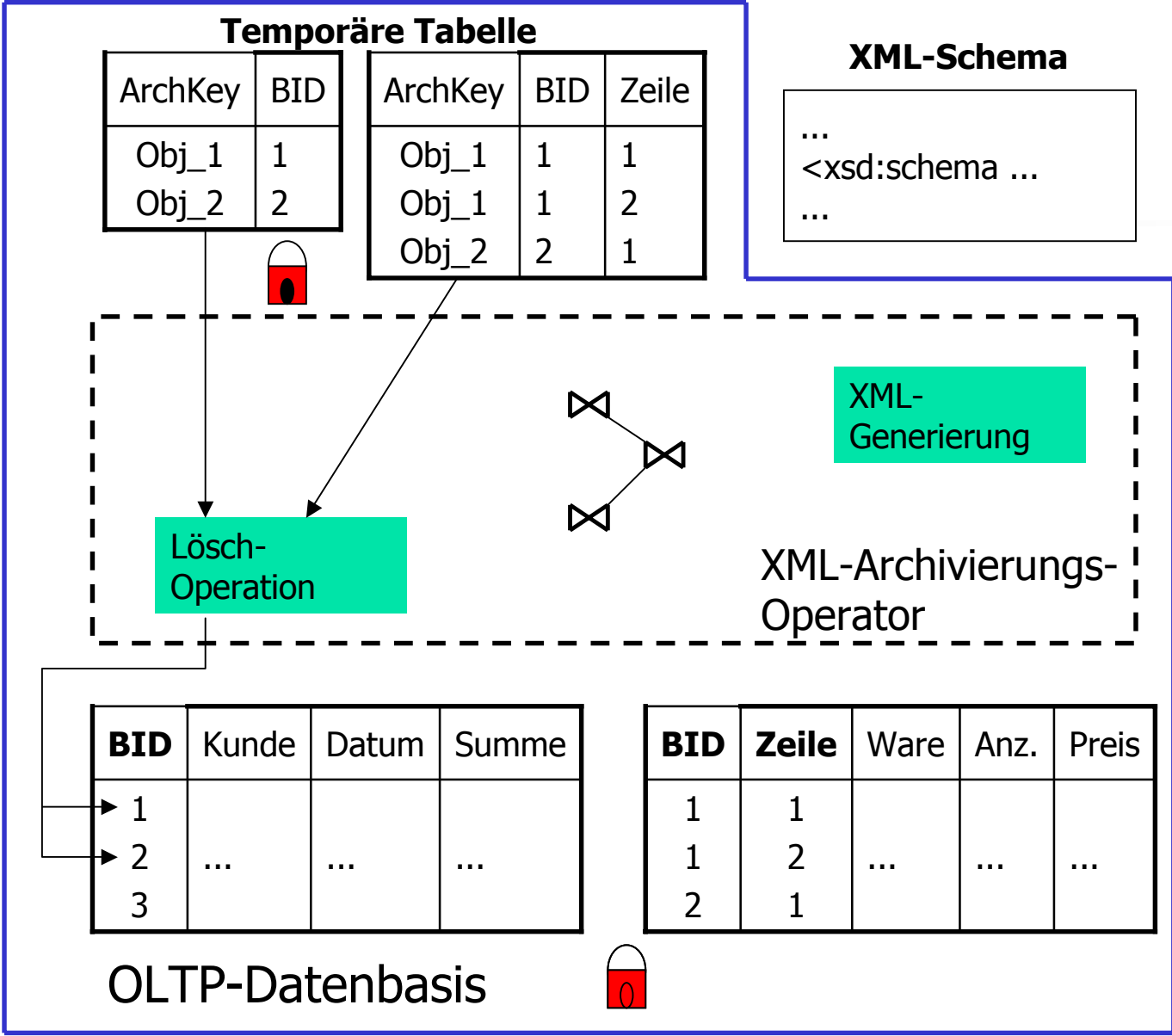
Ablage

Anwendung



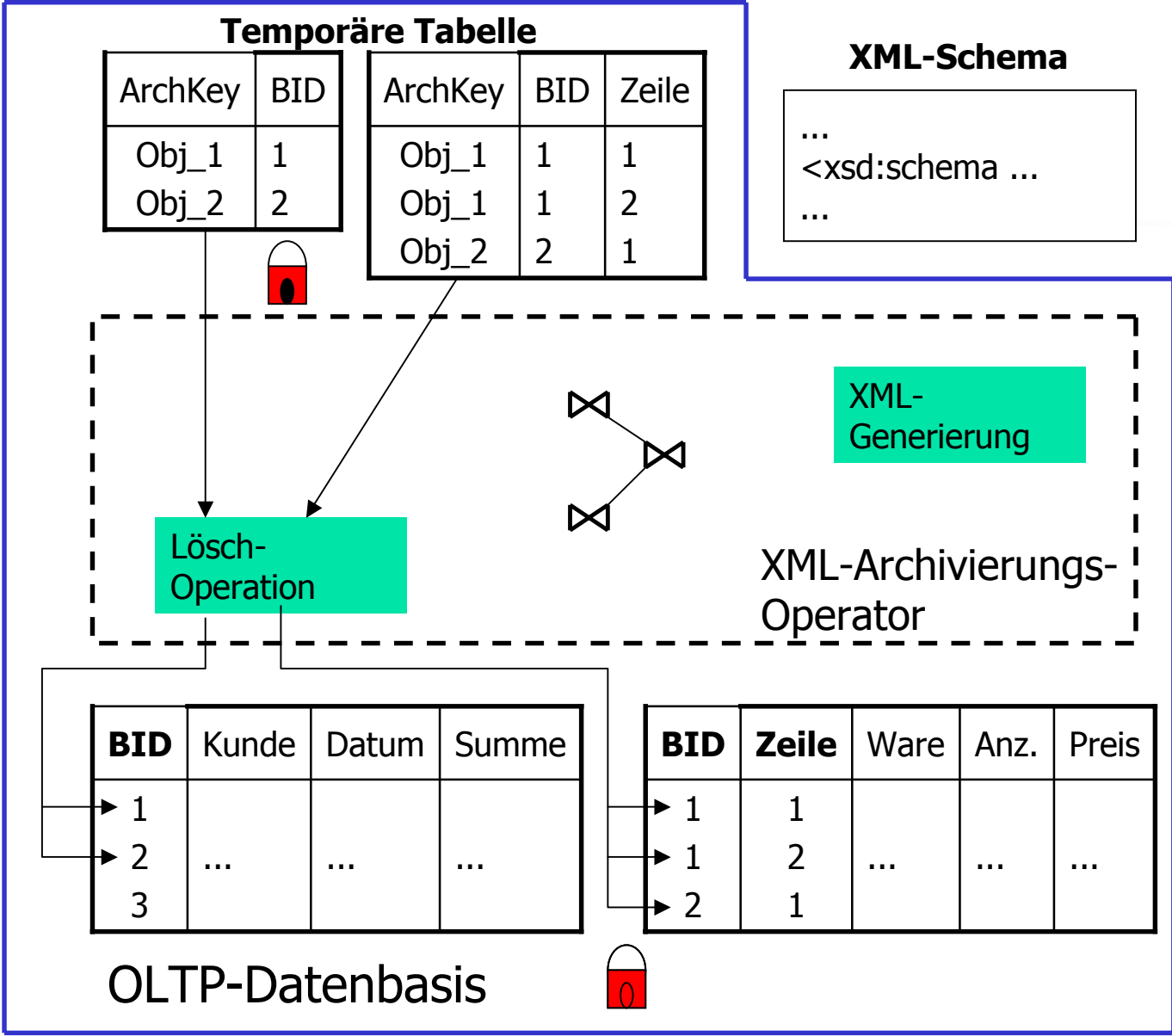
Ablage

Anwendung



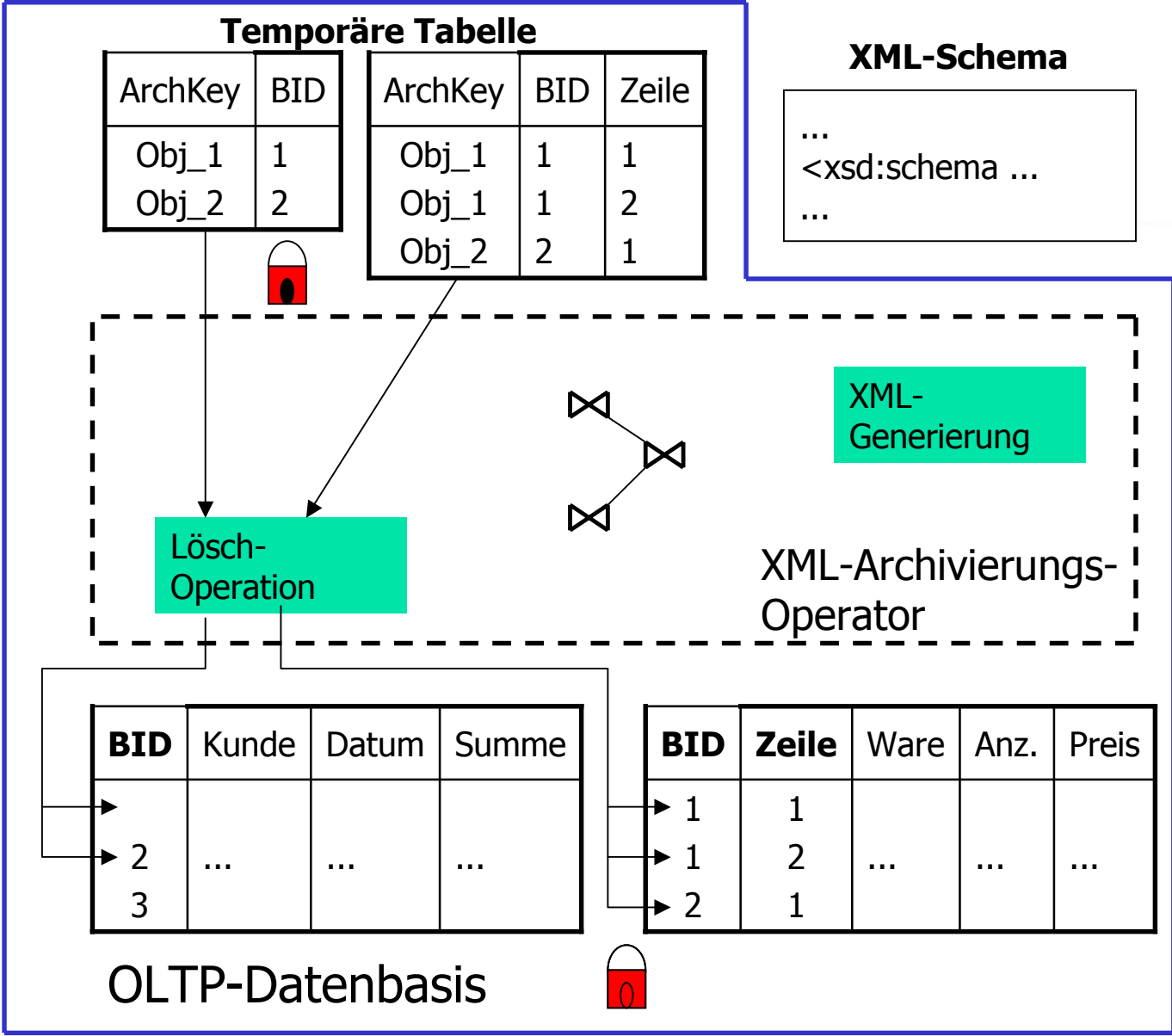
Ablage

Anwendung



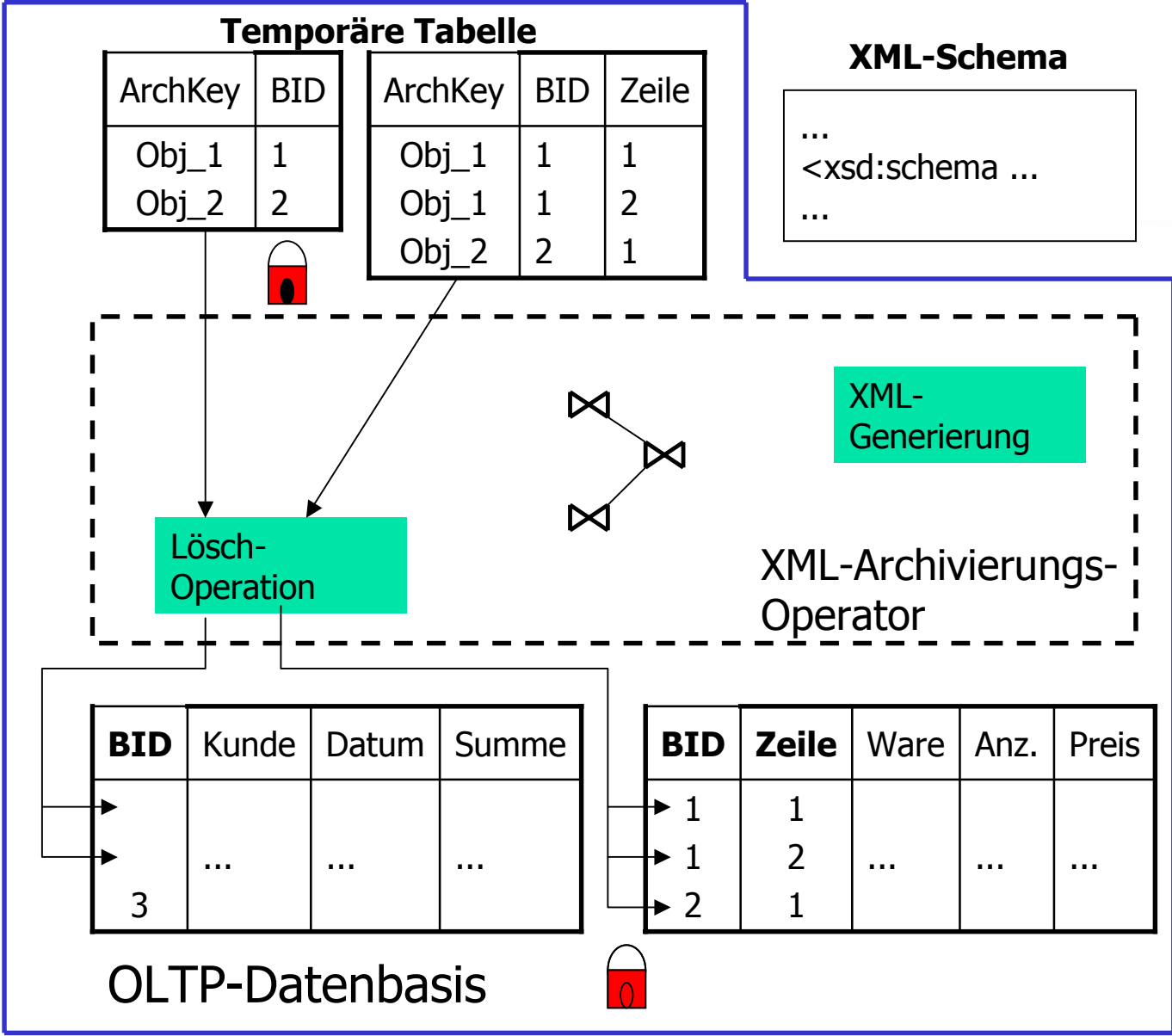
Ablage

Anwendung

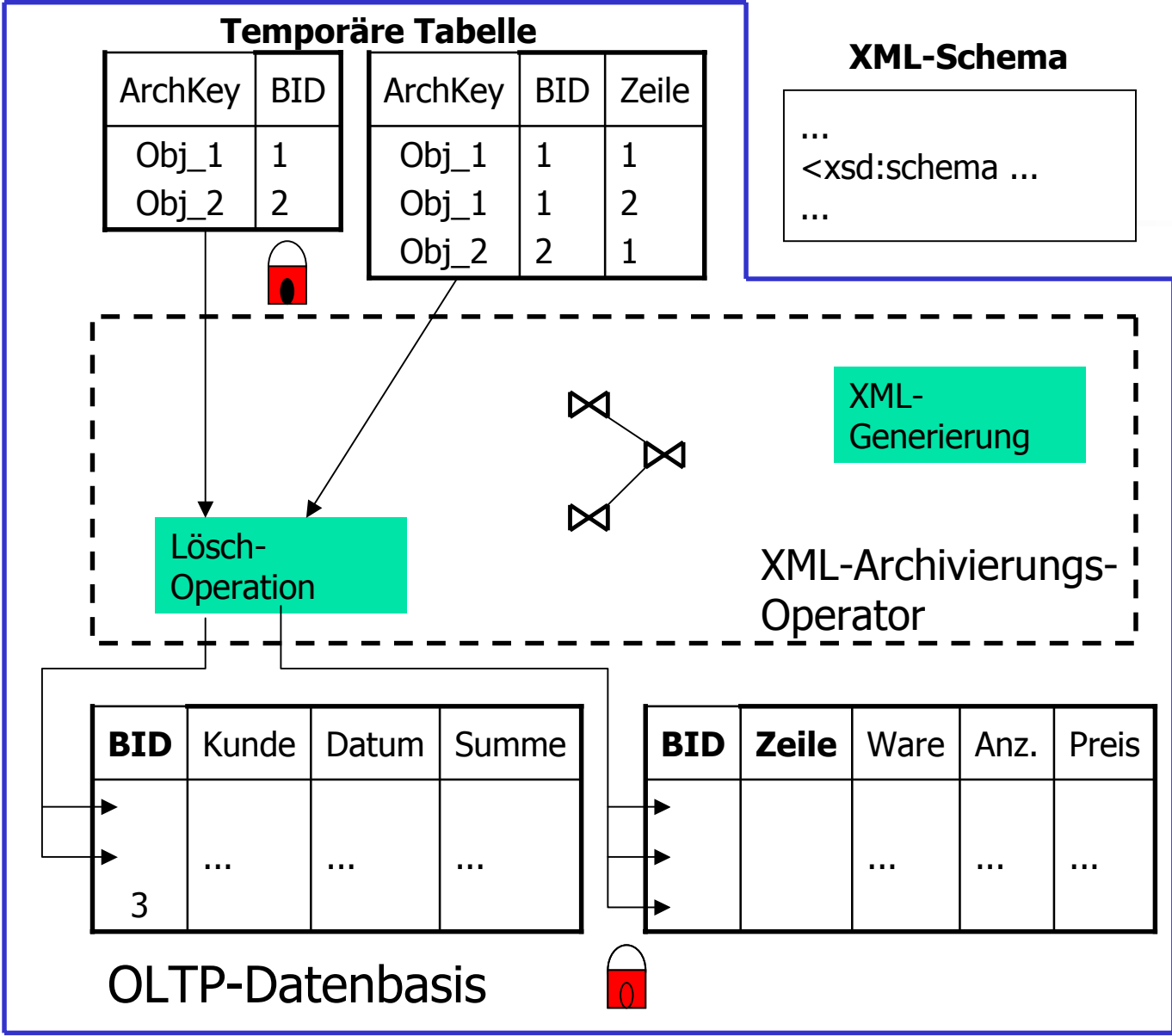


Ablage

Anwendung

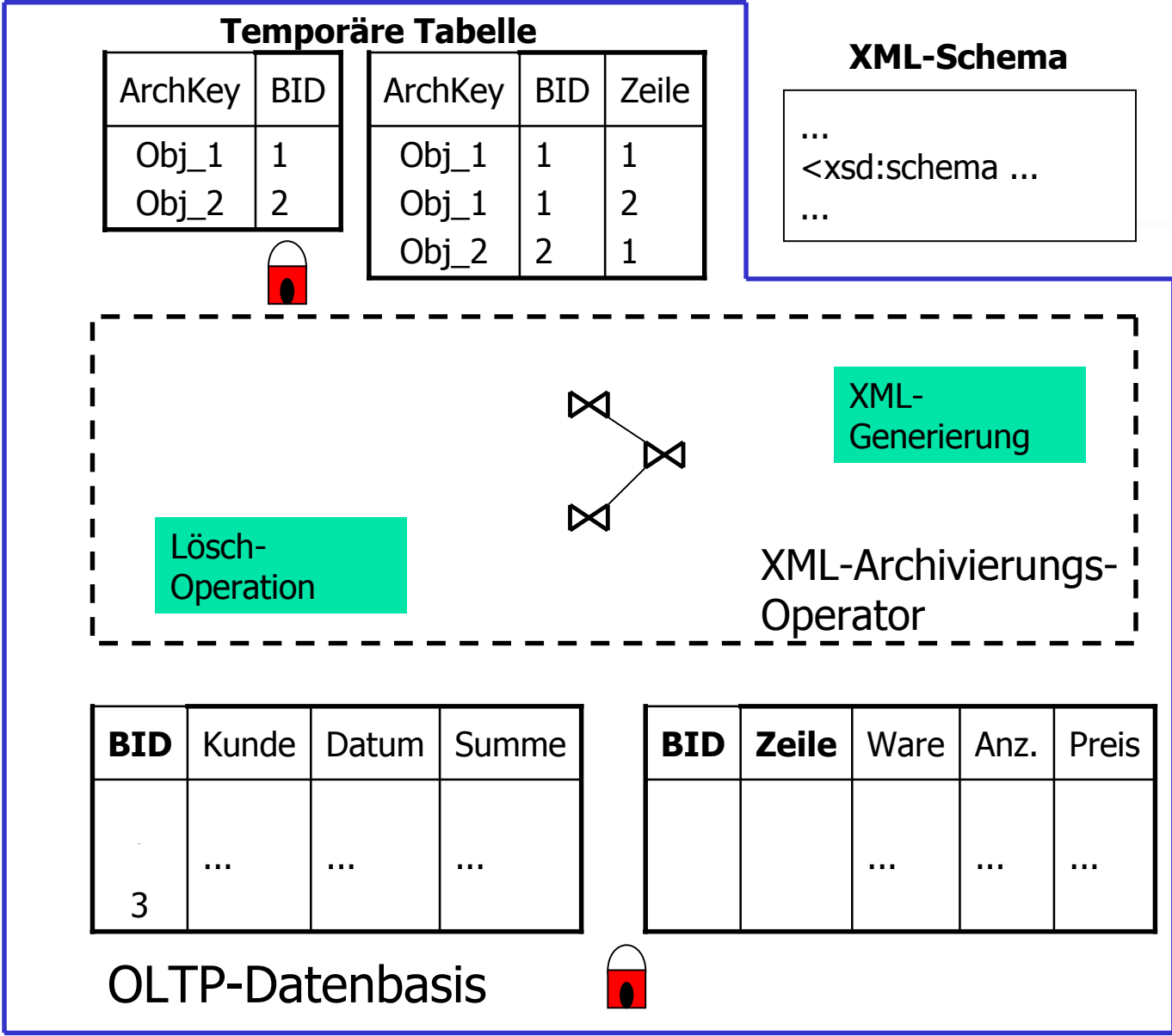


Anwendung





Anwendung



Ablage

Anwendung

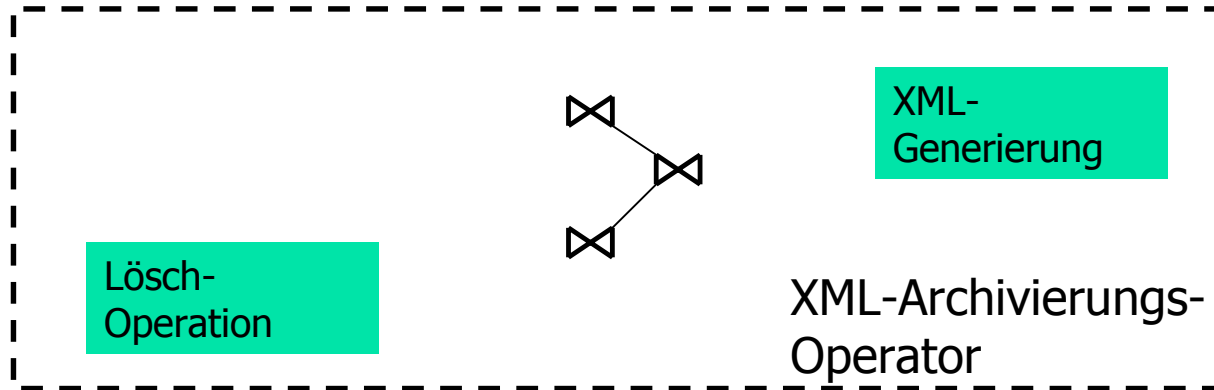
**Temporäre Tabelle**

ArchKey	BID
Obj_1	1
Obj_2	2

ArchKey	BID	Zeile
Obj_1	1	1
Obj_1	1	2
Obj_2	2	1

**XML-Schema**

```
...  
<xsd:schema ...  
...
```



BID	Kunde	Datum	Summe
3	...	...	...

OLTP-Datenbasis

BID	Zeile	Ware	Anz.	Preis
		...	...	...

Ablage

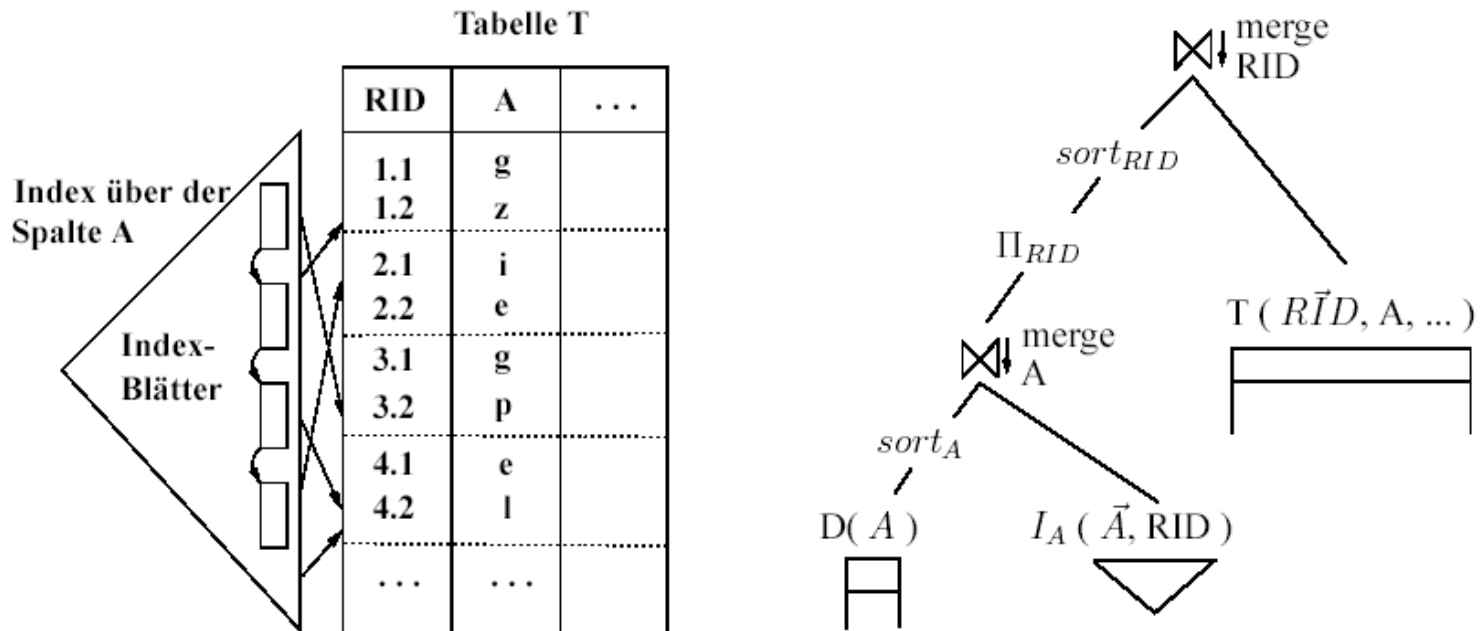


# Ablauf einer XML-Archivierung

---

1. Temporäre Tabellen erzeugen
2. Temporäre Tabellen mit Schlüsselwerten füllen
3. Der XML-Archivierungs-Operator wird aufgerufen.
  - (a) Temporäre Tabellen sperren
  - (b) Einträge in prod. Tabellen sperren (Lesesperre)
  - (c) Daten und XML-Schema verknüpfen und XML-Dokumente generieren
  - (d) XML-Dokumente ablegen
  - (e) Lesesperren in Schreibsperren umwandeln und Daten in prod. Tabellen löschen
  - (f) Der Operator gibt die Sperren frei.
4. Temporäre Tabellen löschen

# Bulkdelete (ICDE 2001)





# Light-Variante des Bulkdelete

---

- Wenn RID vorhanden (und order by erlaubt):

```
delete from <produktive Tabelle P> where <P.Schlüssel> in
(select <P.Schlüssel> from <produktive Tabelle P>,
                             <temporärer Tabelle T>
 where <P.Schlüssel> = <T.Schlüssel>
 order by <P.RID>)
```

- Sortierung bzgl Primärschlüssel und auf gute Clusterung hoffen



# Gliederung

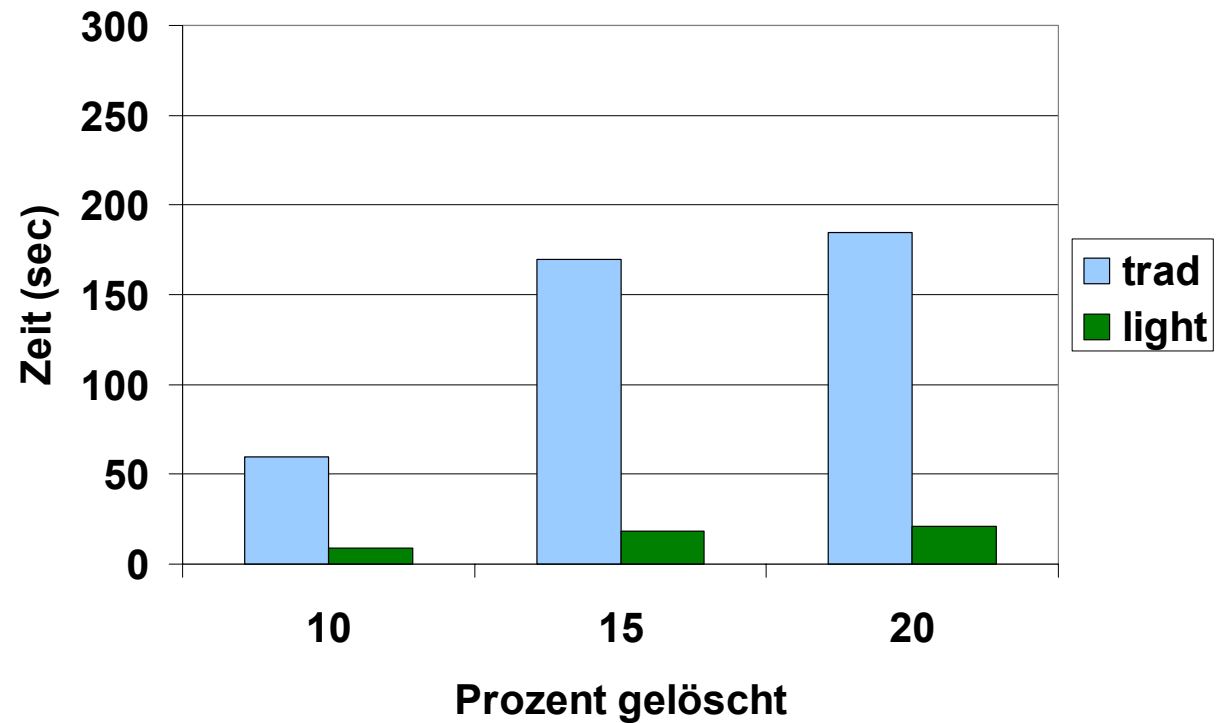
---

- Motivation
- Archivierung bisher
- Verwandte Arbeiten
- Besonderheiten der XML-Archivierung
- Der XML-Archivierungs-Operator
- **Leistungsanalyse**
- **Zukünftige Arbeiten**
- **Zusammenfassung**



# Leistungsanalyse

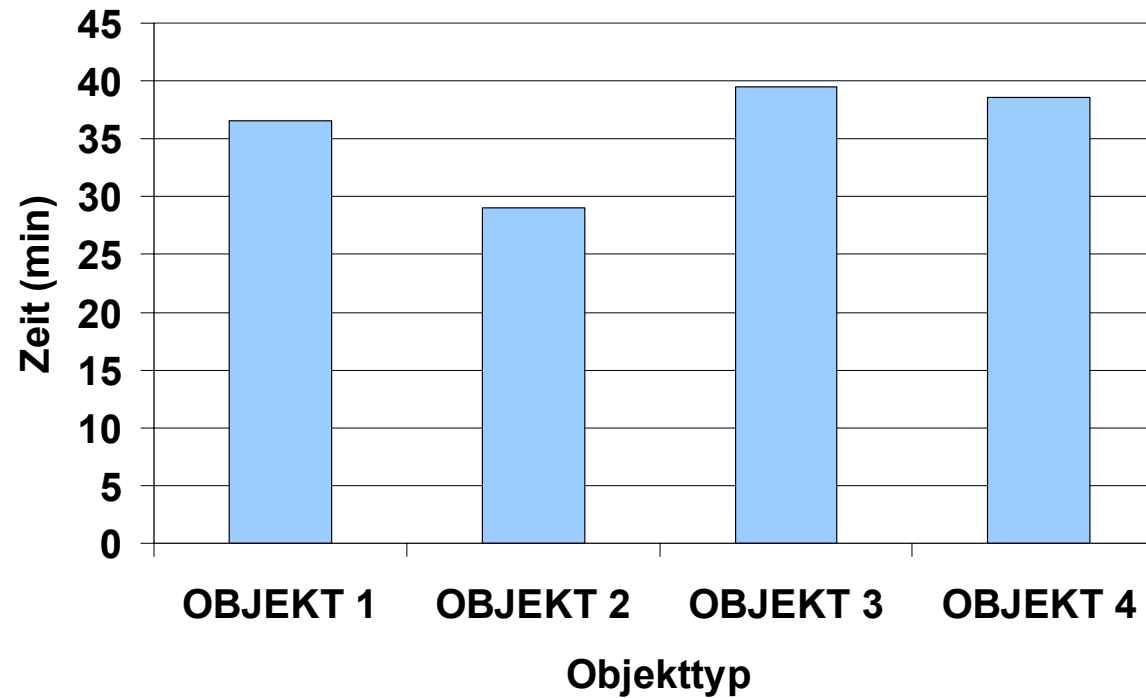
Objektyp 3, 10.000 Objekte





# Leistungsanalyse

**XML Generierung: 10.000 Objekte, 10% archiviert**

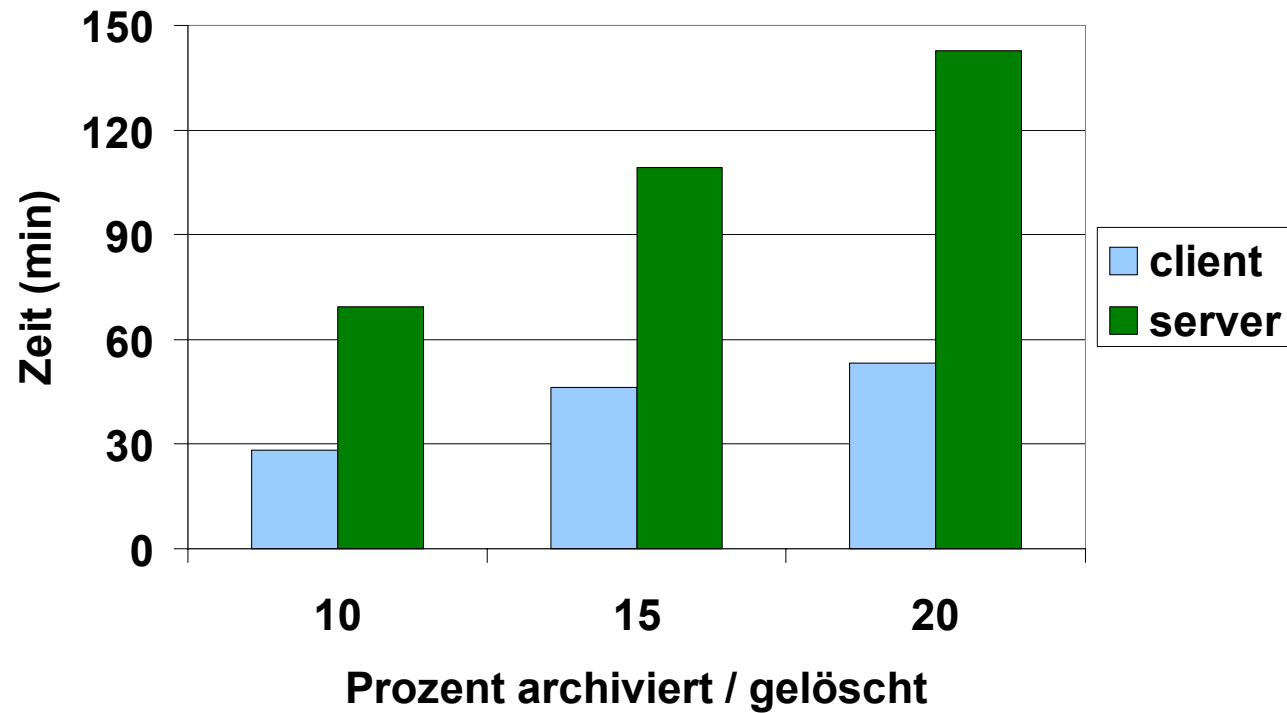






# Leistungsanalyse

10.000 Objekte, Objekttyp 2





# Zukünftige Arbeiten

---

- Implementierung verschiedener XML-Generierungstechniken
- Vergleich mit den XML-Generierungswerkzeugen der Datenbankhersteller
- Entwicklung eines dafür geeigneten Benchmarks
- Leistungsanalyse

**Archive Browser** FILE SYSTEM - Archive:  
 Current Path → /Diplomarbeit/archiv/DAservice/

**Collections**  
 Actual Directory  
 axml  
 axml2  
 fi\_dec  
 matbel  
 mm\_matbel

**Metadata**  
 Path: /Diplomarbeit/archiv/DAservice/  
 File: 1005.xml  
 Documents: 1005.xml

**Referenced Tables:**  
 YSIMP\_1  
 YSIMP\_2  
 YSIMP\_3  
 YSIMP\_4  
 YSIMP\_5  
 YSIMP\_6  
 YSIMP\_KOPF

**Xml**  
 <ItemList xsi:schemaLocation="http://www.sap.com/archive/axml/document SCHEMA.XSD">  
 <MKPF>  
 <MANDT>001</MANDT>  
 <MBLNR>4900008001</MBLNR>  
 <MJAHR>2001</MJAHR>  
 <VGART>WA</VGART>  
 <BLART>WA</BLART>  
 <BLAUM>PR</BLAUM>  
 <BLDAT>2001-01-02</BLDAT>  
 <BUDAT>2001-01-02</BUDAT>  
 <CPUDT>2001-01-02</CPUDT>  
 <CPUTM>17:13:40</CPUTM>  
 <AEDAT/>  
 <USNAM>HUETT</USNAM>  
 <TCODE/>  
 <XBLNR>I0000015t000001</XBLNR>  
 <BKTXT>POS/R310</BKTXT>  
 <FRATH>6.16161616161612E10</FRATH>  
 <FRBNR> 2</FRBNR>  
 <WEVER/>  
 <XABLN> H</XABLN>  
 <AWSYS>P1CLNT999</AWSYS>  
 <BLA2D> M</BLA2D>

**Commercial View**

**Stylesheets**  
 - matbel0.xml  
 - matbel1.xml  
 - matbel2.xml

MKPF	MJAHR	MBLNR	VGART	BUDAT	USNAM
	2001	4900008001	WA	2001-01-02	HUETT

MKPF	MJAHR	MBLNR	VGART	BUDAT	USNAM
	2001	4900008002	WL	2001-01-03	HUETT

MSEG	ZEILE	BWART	MATNR	WERKS	LGORT	CHARG	SOBKZ	MENGE
	0001	601	R100000	R311	0001			6.153542061616E9

MKPF	MJAHR	MBLNR	VGART	BUDAT	USNAM
	2001	4900008003	WL	2001-01-03	HUETT



# Zusammenfassung

---

- In dieser Arbeit wurde ein **XML-Archivierungs-Operator**, der
- Daten betriebswirtschaftlicher Objekte als **XML-Dokumente** ablegt und
- in die Datenbank **integriert** ist
- Durch Archivierung **verbessert** sich die **Leistung** des DBMS,
- es werden **Kosten gespart** und
- durch Ablage als XML-Dokumente sind die **Daten** für andere Anwendungen **lesbar**



# Fragen

---

Danke für Ihre Aufmerksamkeit.

Fragen ?