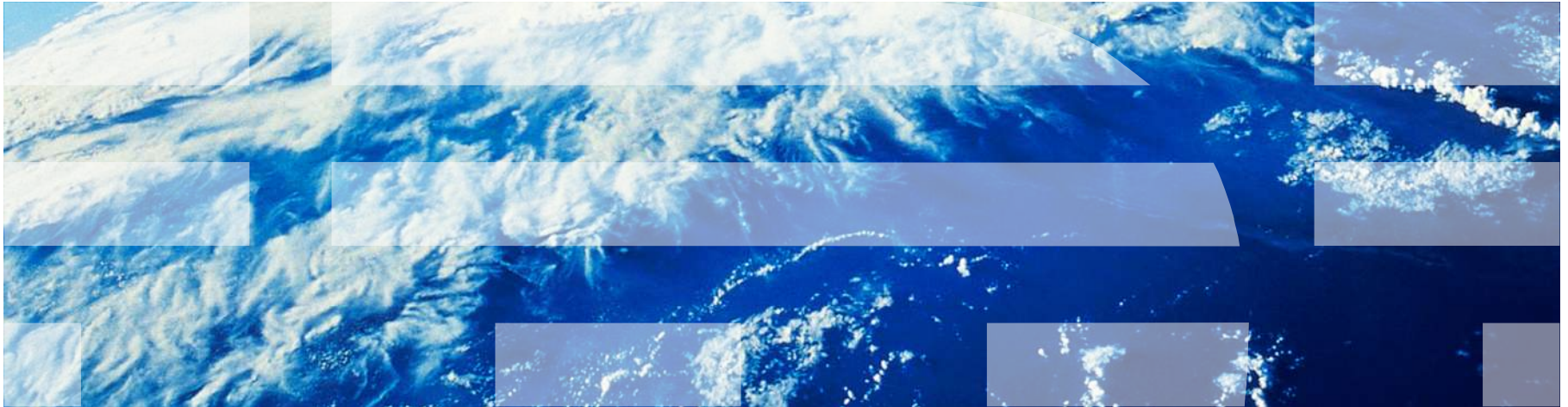


IBM Informix Change Data Capture API



IBM Informix Change Data Capture API (CDC API)

- nová vlastnost IDS 11.50 (od 11.50.xC3)
- umožňuje externím aplikacím zachytávat 'v reálném čase' změny v datech uložených v databázích IDS. Monitorované operace:
 - INSERT
 - UPDATE
 - DELETE
 - TRUNCATE
- změny jsou získávány z logických logů
- podporovány jsou všechny datové typy kromě:
 - TEXT, BYTE
 - SET, MULTISSET, LIST, ROW
 - UDTs

CDC API - Komponenty

- systémová databáze „syscdc“
 - obsahuje CDC funkce, definice CDC záznamů a chybových kódů a informace o probíhajících CDC sezeních
- CDC funkce
 - vestavěné SQL funkce pro kontrolu CDC sezení (otevření/uzavření, (de-)aktivace, ...)
- CDC záznamy
 - informace o probíhajících transakcích, strukturách tabulek i zachycená data informace jsou vrácena jako smart-blob objekty ve formě tzv. CDC záznamů
- chybové (návrátové) kódy CDC funkcí
- funkce pro čtení smart-blob objektů
 - nejsou součástí CDC API; jsou používány pro přenos CDC záznamů do aplikačního bufferu
 - `mi_lo_read()` - DB API; `ifx_lo_read()` - ESQCLC, ODBC; `IfxLoRead()` - JDBC; `IfxBlob.Read()` - .NET

CDC API – Databáze „syscdc“

- obsahuje tabulky a funkce CDC API
 - Tabulky:
 - syscdcerrcodes
 - syscdcpacketschemes
 - syscdcrectypes
 - syscdcscs
 - syscdctabs
 - syscdcvers

- slouží pro připojení aplikací, které chtějí využívat CDC API

- používá UTF-8 locale

- Před prvním použitím CDC API je nutno ji vytvořit (jako uživatel „informix“):
 - `$ dbaccess - $INFORMIXDIR/etc/syscdcv1.sql`

CDC API – CDC funkce

- jsou (musí být) volány z aplikace (není možné je volat z procedur v IDS)
 - `cdc_opensess()`
 - otevře CDC sezení a vrátí jeho ID
 - současně nastavuje prodlevu, po jaké je sezení uzavřeno v případě, že nejsou zachycena žádná data
 - `cdc_set_fullrowlogging()`
 - zapíná/vypíná full-row logovací mód pro tabulku
 - pro jakýkoli update budou do log. logu uloženy věty HUPDBEF + HUPDAFT namísto jednoduché HUPDATE
 - `cdc_startcapture()`
 - specifikuje tabulku (a její sloupce), pro kterou se má zachytávání provádět
 - `cdc_activatesess()`
 - pro dané sezení spustí zachytávání dat od dané pozice v logickém logu (def. current)

CDC API – CDC funkce - pokr.

- `cdc_errortext()`
 - vrací text chyby odpovídající zadanému symbolickému jménu chyby

- `cdc_recboundary()`
 - restartuje zachytávání dat od začátku aktuálního CDC záznamu

- `cdc_endcapture()`
 - ukončí zachytávání dat pro danou tabulku

- `cdc_deactivatesess()`
 - deaktivuje dané aktivní sezení

- `cdc_closesess()`
 - uzavře dané sezení

CDC API - CDC záznamy

- popisují:
 - změny dat zachycené nad tabulkou
 - insert, update (before a after image), delete, truncate
 - transakce, které tyto změny provedly
 - begin, commit, rollback
 - další „režijní“ informace
 - schéma tabulky, chyba, timeout, vyřazení (discard)
- všechny začínají **jednotnou hlavičkou**, která popisuje délku a typ záznamu
- dále obsahují **hlavičku specifickou pro daný typ záznamu**
- + samotná data

Header size	Payload size	Packet scheme	Record type	Record specific header	Data
4 bytes	4 bytes	4 bytes	4 bytes	8 – 24 bytes	>= 0 bytes

CDC API – LSN - sledování pozice v log. lozích

- většina CDC záznamů obsahuje tzv. 'logical sequence number'
- jedná se o pozici v logických lozích, odkud byl daný záznam přečten
- datový typ BIGINT (8 bytů)
 - horní 4 byty udávají číslo log. logu
 - dolní 4 byty udávají pozici v log. logu ve tvaru 0xLLLLLLBBB
 - LLLLLL = logická adresa stránky v daném logu
 - BBB = byte offset na stránce
 - př. 17191440536 = 0x0000000400B09098
 - log. log č. 4
 - log. stránka 0x00B09 = 2825
 - offset 0x098 = 152
- LSN lze použít k restartu zachytávání od určitého místa v log. lozích (`cdc_activatesess()`)

CDC API – Použití

- Aplikace by měla obsahovat
 - strukturu pro popis schématu tabulky(-ek)
 - funkce pro naplnění této struktury
 - funkce pro zpracování jednotlivých typů záznamu (především hodnot datových sloupců)
 - funkce pro ošetření chybových stavů
 - mechanismus pro sledování/uložení aktuálního LSN

CDC API – Použití - pokr.

▪ Postup:

- musí být spuštěna pod uživatelem 'informix'
- nastavení locales dle zdrojové (zachytávané) databáze
- připojení k databázi syscdc1
- otevření zachytávací session – `cdc_opensess()`
- zapnutí full-row logování pro všechny monitorované tabulky – `cdc_set_fullrowlogging()`
- definice dat (sloupců), která se budou monitorovat – `cdc_startcapture()`
- start zachytávání – `cdc_activatesess()`
- čtení CDC záznamů pomocí funkcí pro čtení smart-blob objektů
- zpracování zachycených dat (zápis do souboru, modifikace non-IDS databáze, ...)
- ukončení zachytávání – buď po tabulkách (`cdc_stopcapture()`), nebo pro celou session (`cdc_deactivatesess()`)
- vypnutí full-row logování pro tabulky
- uzavření zachytávací session - `cdc_closesess()`

CDC API - Omezení

- pokud nad tabulkou existují in-place altery, není možné ji zachytávat

- pokud je tabulka zachytávána, není možné provádět následující operace:
 - alter table modify/add/drop column
 - drop table
 - alter fragment
 - create/alter cluster index
 - rename table/column

CDC API - Monitoring

- `onstat -g cdc [session_ID] [bufm | config | table]`

```
$ onstat -g cdc [long]
```

```
CDC subsystem structure at 0x44148d20
```

```
  CDC session structure at 0x4d2716a8
```

```
    CDC session id: 15728679 (0xf00027)
```

```
    Associated SQL session id: 30
```

```
    Number of tables captured: 1
```

```
[long] State: ACTIVATED (0x50534555)  
Create time: 1252570632 (Thu Sep 10 10:17:12 2009)  
Open time: 1252570632 (Thu Sep 10 10:17:12 2009)  
Activate time: 1252570632 (Thu Sep 10 10:17:12 2009)  
Activate Sequence Number: 0x0  
Total client read calls: 25  
Last client read time: 1252570690 (Thu Sep 10 10:18:10 2009)  
Last Sequence Number returned to client: 0x400b092ac  
Total number records examined: 38  
Total number records kept (approximate): 27  
Total number I/U/D records examined: 14  
Total number I/U/D records kept (approximate): 12  
Client required to close: NO  
Read exit error code: 0
```

CDC API – Monitoring - pokr.

```
$ onstat -g cdc 23592999  bufm [long]
```

```
IBM Informix Dynamic Server Version 11.50.UC5      -- On-Line -- Up 00:48:07 -- 195884 Kbytes
CDC subsystem structure at 0x44148d20
  CDC session structure at 0x4cf8b6a8
    CDC session id: 23592999 (0x1680027)
```

```
  Buffer Manager  at 0x4cf8b898
    Number of allocated buffers high watermark: 25
    Number of currently allocated buffers: 2
    Minimum prepend for allocated buffers: 172
```

```
[long]          Buffer at 0x4d4563d0
                Buffer at 0x4d458930
```

CDC API – Monitoring - pokr.

```
$ onstat -g cdc table cdctest:informix.tab1 [long]
```

```
IBM Informix Dynamic Server Version 11.50.UC5      -- On-Line -- Up 00:46:51 -- 195884 Kbytes
```

```
CDC subsystem structure at 0x44148d20
```

```
  CDC session structure at 0x4d2716a8
```

```
    CDC session id: 15728679 (0xf00027)
```

```
    Captured Table Manager found at 0x4d27e018
```

```
    Number of tables captured: 1
```

```
      Captured Table structure at 0x4d271960
```

```
        Full Table Name: cdctest:informix.tab1
```

```
        Version Sequence Number: 0x400af807c
```

```
        Time capture started: 1252570632 (Thu Sep 10 10:17:12 2009)
```

```
[long]
```

```
  Captured Columns:
```

```
    Number having fixed length: 2
```

```
    Number having variable length: 0
```

```
    Total bytes in columns with fixed length : 9
```

```
    a integer, b char(5)
```

CDC API – Monitoring - pokr.

```
$ onstat -g cdc config
```

```
IBM Informix Dynamic Server Version 11.50.UC5      -- On-Line -- Up 00:47:20 -- 195884 Kbytes
CDC subsystem structure at 0x44148d20
  CDC session structure at 0x4d2716a8
    CDC session id: 15728679 (0xf00027)
    Read Timeout (seconds): 500
    Maximum buffers per read call: 1
    Survive DATALOST errors: NO
```

CDC API – Vzorový program

- součástí CDC API Programmer's Guide
- ukazuje základní použití CDC API (čte CDC záznamy a vypisuje změněná data na standardní výstup)
- http://publib.boulder.ibm.com/infocenter/idshelp/v115/topic/com.ibm.cdc.doc/ids_cdc_060.htm

CDC API - Otázky

