## 实验一: 分布式数据库数据定义和数据分片

学号: <u>202103151422</u>			姓名:温	家伟	班级: <u>大数据分析 2101</u>	
1、	实验	҈环境安装情况(用☑替	换口即可)			
	a)	□自安装环境	□高校邦系统	(如果选择后	后者,以下三个不用回答)	
	b)	□单机	☑多机			
	c)	□单数据库实例	☑两个数据库剪	它例	□多个数据库实例	
	d)	☑同类型数据库实例	□两个类型数据	居库实例	□多类型数据库实例	
	e)	□Mycat	☑ShardingSphe	ere		
	f)	☑MySQL	□OpenGauss		□其他	
2、	实验	內容				
	a)	针对实验运行环境(用	☑替换□即可)			
		☑命令行下进行实验	□Java 应用程序	序进行实验	□Python 应用程序进行实验	
		□其他				
	b)	创建学生表 Stu、课程	表 Cou 和学生选	课表 SC		
		推荐采用 Create Table	创建请将 Create	Table 完整语	5	
	若别	<b>采用其他方式创建表,</b> 同	样可以把相关作	代码或者图片	计截图放在下表的右侧空白格中。	
	表		创建表的过程	1		
	wer	njw_Stu(	CREATE TABLE wenjw_Stu0 (			
		Sno//学号	Sno INT	PRIMARY I	KEY,	
		Sname//姓名	Sname V	ARCHAR(10	00),	
		Sdept//学院	Sdept VA	RCHAR(100	0),	
		Stel//联系方式	Stel VAR	CHAR(20),		
		Sex//性别	Sex VAR	CHAR(10),		
		Sage//年龄	Sage INT	,		
		Saddr//家庭住址	Saddr VA	RCHAR(20	0),	
	Spost//邮编		Spost VA	RCHAR(10)		
		);	);			
			CREATE TAE	BLE wenjw_S	Stu1 (	
			Sno INT	PRIMARY I	KEY,	
			Sname V	ARCHAR(10	00),	
			Sdept VA	RCHAR(100	0),	
			Stel VAR	CHAR(20),		
			Sex VAR	CHAR(10),		
			Sage INT	,		
			Saddr VA	RCHAR(20	0),	
			Spost VA	RCHAR(10)		

);

wenjw_Cou(	CREATE TABLE wenjw_Cou (		
Cno//课号	Cno INT PRIMARY KEY,		
Cname//课名	Cname VARCHAR(100),		
Cdept//开课学院	Cdept VARCHAR(100),		
Ctea//开课教师	Ctea VARCHAR(100),		
Cdate//开课学期	Cdate VARCHAR(20),		
CScore//学分	CScore INT,		
Caddr//开课教室	Caddr VARCHAR(200)		
);	);		
	// 这里 wenjw_Cou 配置的是广播表,所以在两个数据		
	库实例中都相同。		
wenjw_SC(	CREATE TABLE wenjw_SC0 (		
Sno//学号	Sno INT,		
Cno//课号	Cno INT,		
Grade//成绩	Grade DECIMAL(4, 1) NOT NULL CHECK (Grade		
);	BETWEEN 0 AND 100),		
	PRIMARY KEY (Sno, Cno),		
	FOREIGN KEY (Sno) REFERENCES		
	wenjw_Stu0(Sno),		
	FOREIGN KEY (Cno) REFERENCES		
	wenjw_Cou(Cno)		
	);		
	CREATE TABLE wenjw_SC1 (		
	Sno INT,		
	Cno INT,		
	Grade DECIMAL(4, 1) NOT NULL CHECK (Grade		
	BETWEEN 0 AND 100),		
	PRIMARY KEY (Sno, Cno),		
	FOREIGN KEY (Sno) REFERENCES		
	wenjw_Stu1(Sno),		
	FOREIGN KEY (Cno) REFERENCES		
	wenjw_Cou(Cno)		
	);		

在两个数据库中查看三张表的模式,查看到的表模式截图插入下表中(如果是多数据库的请自行扩展列; Navicat 可以查看数据库,即可以借助其他软件查看数据库表模式;还可以用命令行工具查看表模式,下表需要放截图)。

表	数据库 1	数据库 2	

Stu	mysql> describe wenjw_Stu0;	<pre>MySQL [wenjw_db1]&gt; describe wenjw_Stul; +</pre>
Sta	Field   Type   Null   Key   Default   Extra	Field   Type   Null   Key   Default   Extra
	Sno   int(11)   NO   PRI   NULL   NO   NO   NULL   NULL   NO   NULL   NULL	Sno
	mysql> describe wenjw Cou;	
Cou	Field   Type   Null   Key   Default   Extra	MySQL [wenjw_db1]> describe wenjw_Cou;
	÷+	Field   Type   Null   Key   Default   Extra
	Cno   int(11)   NO   PRI   NULL	Cno   int(11)   NO   PRI   NULL
	7 rows in set (0.00 sec)	7 rows in set (0.00 sec)
SC	mysql> describe wenjw_SC0;	MySQL [wenjw_db1]> describe wenjw_SC1;
	Field   Type   Null   Key   Default   Extra	++   Field   Type
	Sno   int(11)   NO   PRI   NULL	The state of the
	3 rows in set (0.00 sec)	

主要目的是对比数据库1和数据库2上的表模式结构。

c) 针对自安装环境的数据分片所涉及的配置文件内容(以下以 Mycat 为例)

创建分库分表的过程中可能会涉及到分库分表相关配置文件的修改,且需要按照需求设计分库分表策略,所以配置文件信息是重要的一部分内容,所以请把相关配置文件的修改部分粘贴到下表右侧空白格中。

配置文件名	配置内容 <mark>修改部分</mark> (加注释,注释用以阐明分库分表策略)
Server.yaml	# 访问认证的配置部分:
	rules:
	- !AUTHORITY
	users:
	- root@%:root
	- sharding@:sharding
	provider:
	type: ALL_PERMITTED
	- !TRANSACTION
	defaultType: XA
	providerType: Atomikos
	- !SQL_PARSER
	sqlCommentParseEnabled: true
	sqlStatementCache:
	initialCapacity: 2000
	maximumSize: 65535
	parseTreeCache:
	initialCapacity: 128
	maximumSize: 1024

# 其他的一些参数: props: max-connections-size-per-query: 1 kernel-executor-size: 16 # Infinite by default. proxy-frontend-flush-threshold: 128 # The default value is 128. proxy-hint-enabled: true sql-show: true check-table-metadata-enabled: false # Proxy backend query fetch size. A larger value may increase the memory usage of ShardingSphere Proxy. # The default value is -1, which means set the minimum value for different JDBC drivers. proxy-backend-query-fetch-size: -1 proxy-frontend-executor-size: 0 # Proxy frontend executor size. The default value is 0, which means let Netty decide. # Available options of proxy backend executor suitable: OLAP(default), OLTP. The OLTP option may reduce time cost of writing packets to client, but it may increase the latency of SQL execution # and block other clients if client connections are more than 'proxy-frontend-executor-size', especially executing slow SQL. proxy-backend-executor-suitable: OLAP proxy-frontend-max-connections: 0 # Less than or equal to 0 means no limitation. # Available sql federation type: NONE (default), ORIGINAL, **ADVANCED** sql-federation-type: NONE # Available proxy backend driver type: JDBC (default), ExperimentalVertx proxy-backend-driver-type: JDBC proxy-mysql-default-version: 5.7.22 # In the absence of schema name, the default version will be used. proxy-default-port: 3307 # Proxy default port. proxy-netty-backlog: 1024 # Proxy netty backlog. # 逻辑数据库 Config-Sharding.yaml databaseName: wenjw db # 数据源

```
dataSources:
 wenjw_db0:
   # 实例
   url:
jdbc:mysql://127.0.0.1:3306/wenjw db0?serverTimezone=UTC
&useSSL=false&characterEncoding=utf8
   username: root
   password: dees
   connectionTimeoutMilliseconds: 30000
   idleTimeoutMilliseconds: 60000
   maxLifetimeMilliseconds: 1800000
   maxPoolSize: 50
   minPoolSize: 1
 wenjw db1:
   # 实例
   url:
jdbc:mysql://120.26.168.159:3306/wenjw_db1?serverTimezone=
UTC&useSSL=false&characterEncoding=utf8
   username: root
   password:
   connectionTimeoutMilliseconds: 30000
   idleTimeoutMilliseconds: 60000
   maxLifetimeMilliseconds: 1800000
   maxPoolSize: 50
   minPoolSize: 1
# 配置分库分表规则
rules:
-!SHARDING
 tables:
   wenjw Stu:
   # 逻辑表 wenjw_Stu 划分为以下两个实际表
     actualDataNodes: wenjw_db0.wenjw_Stu0,
wenjw db1.wenjw Stu1
     tableStrategy:
       standard:
         shardingColumn: Sno
         shardingAlgorithmName: wenjw_Stu_inline
```

```
keyGenerateStrategy:
       column: Sno
       keyGeneratorName: snowflake
   wenjw SC:
   # 逻辑表 wenjw SC 划分为以下两个实际表
     actualDataNodes: wenjw db0.wenjw SC0,
wenjw_db1.wenjw_SC1
     tableStrategy:
        standard:
          shardingColumn: Sno
          shardingAlgorithmName: wenjw_SC_inline
     keyGenerateStrategy:
       column: Sno
       keyGeneratorName: snowflake
 # 配置绑定表
 bindingTables:
   - wenjw Stu, wenjw SC
 # 配置广播表
 broadcastTables:
   - wenjw Cou
 # 默认的分库策略: database_inline (见后面)
 defaultDatabaseStrategy:
  standard:
     shardingColumn: Sno
     sharding Algorithm Name: \ database\_in line
 defaultTableStrategy:
   none:
 # 配置分库分表算法
 shardingAlgorithms:
   database inline:
   # 分库策略: 按照 Sno % 2 的方式进行划分
     type: INLINE
     props:
       algorithm-expression: wenjw_db${Sno % 2}
       allow-range-query-with-inline-sharding: true
   wenjw Stu inline:
   #wenjw_Stu 分表策略:按照 Sno % 2 的方式划分
```

type: INLINE props: algorithm-expression: wenjw Stu\${Sno % 2} allow-range-query-with-inline-sharding: true wenjw SC inline: # wenjw SC 分表策略: 按照 Sno % 2 的方式划分 type: INLINE props: algorithm-expression: wenjw SC\${Sno % 2} allow-range-query-with-inline-sharding: true keyGenerators: snowflake: type: SNOWFLAKE auditors: sharding key required auditor: type: DML SHARDING CONDITIONS

配置文件视情况而定,上述是以 Mycat 为例

- ➤ 采用 Mycat 可能涉及多个配置文件;
- ➤ 采用 ShardingSphere-JDBC 可能只涉及一个应用项目相关的配置文件;
- ▶ 采用 ShardingSphere-Proxy 可能涉及多个配置文件。

针对配置文件回答如下问题(用□替换□即可):

- ☑ 配置文件在 Mycat/ShardingSphere 内部
- □ 配置文件在 MySQL/OpenGauss 内部
- □ 配置文件在 Java 应用项目内部
- □ 配置文件在 Python 应用项目内部
- □ 未做任何配置文件的修改

## 注意:

- ▶ 首先,如果是采用高校邦系统,此部分可能也不需要;
- ▶ 其次,命令行/Java 应用程序/Python 应用程序等均可;
- ▶ 如果完全是默认,未做任何配置,上面题目请选最后一个。
- d) 向三张表中插入数据(确保能看到分库分表的效果)

建议采用 Insert 语句插入数据,请务必插入足够数据以确保能看到数据库分库分表的效果。将一系列的 Insert 语句(建议不要超过 50 条,只要能够看到分库分表效果即可)粘贴到下表的右侧空白格中。若采用其他方式导入数据,也请详细阐述相关过程(加注释)。

表	插入数据的过程
Stu	INSERT INTO wenjw_Stu (Sno, Sname, Sdept, Stel, Sex, Sage, Saddr, Spost)

```
VALUES
       (2101, '温家伟', '理学院', '19858180826', '男', 19, '山西省', '037000'),
       (2102, '金娇薄', '计院', '19157934071', '女', 20, '浙江省', '315000'),
       (2103, '毛假发', '法学院', '15678901234', '女', 21, '美国', '510000'),
       (2104, '曾辽远', '理学院', '13609876543', '男', 20, '山东市', '518000'),
       (2005, '秋晓荷', '理学院', '13789012345', '女', 21, '浙江省', '610000'),
       (2006, '朱甜甜', '计院', '15908765432', '女', 20, '吉林省', '311100');
       INSERT INTO wenjw Cou (Cno, Cname, Cdept, Ctea, Cdate, CScore, Caddr)
Cou
       VALUES
       (12011, '分布式数据库', '计院', '范玉雷', '大三上', 3, '健行楼'),
       (12012, '人工智能', '计院', '高楠', '大三上', 3, '健行楼'),
       (12013, '数字图像处理', '计院', '刘盛', '大三上', 2, '广知楼'),
       (12014, '数学建模', '理学院', '周凯', '大二下', 3, '健行楼'),
       (12015, '运筹学', '理学院', '丁晓东', '大二下', 4, '广知楼'),
       (12016, '数据结构', '计院', '蒋莉', '大二上', 4, '郁文楼'),
       (12017, '大学物理', '理学院', '施建青', '大一下', 3, '新教楼'),
       (12018, '程序设计基础 C', '计院', '蒋莉', '大一上', 4, '子良楼');
SC
       INSERT INTO wenjw SC (Sno, Cno, Grade) VALUES
       (2101, 12011, 100),
       (2101, 12012, 100),
       (2101, 12013, 100),
       (2101, 12014, 100),
       (2101, 12018, 100),
       (2102, 12011, 65),
       (2102, 12013, 55),
       (2102, 12015, 59),
       (2102, 12016, 58),
       (2102, 12017, 60),
       (2103, 12012, 60),
       (2103, 12013, 15),
       (2103, 12015, 55),
       (2103, 12017, 66),
       (2103, 12018, 55),
       (2104, 12014, 79),
       (2104, 12015, 80),
```

```
(2104, 12016, 73),

(2104, 12017, 66),

(2104, 12018, 85),

(2005, 12012, 70),

(2005, 12013, 75),

(2005, 12014, 81),

(2005, 12016, 68),

(2005, 12018, 58),

(2006, 12013, 88),

(2006, 12014, 96),

(2006, 12015, 87),

(2006, 12016, 92),

(2006, 12018, 100);
```

插入数据后观察两个数据库中数据的分布情况(截图插入到下表右侧空白格中)。

表	数据库 1			数据库 2		
Stu	mysql> select * from wenjw_Stu0;		u0;	<pre>MySQL [wenjw_db1]&gt; select * from wenjw_Stu1;</pre>		
	Sno   Sname	Sdept	Stel	Sno   Sname   Sdept   Stel		
	2006   朱甜甜   2102   金娇薄   2104   曾辽远	计院   计院   理学院	15908765432   19157934071   13609876543	2005   秋晓荷   理学院   13789012345   2101   温家伟   理学院   19858180826   2103   毛假发   法学院   15678901234		
	3 rows in set (6	0.00 sec)		3 rows in set (0.00 sec)		
Cou	mysql> select * from wenjw_Cou;			<pre>MySQL [wenjw_db1]&gt; select * from wenjw_Cou;</pre>		
	Cno   Cname	Cd	ept Ctea	Cno   Cname   Cdept   Ctea		
	12011   分布式数   12012   人工智能   12013   数字包像模   12014   数字包像模   12015   运筹学   12016   数据结构   12017   大学物理   12018   程序设计	计 处理   计   理   计   理 基础C   计	院   高楠 院   刘周晓 学院   周 明	12011   分布式数据库		
SC	mysql> s	select * from	wenjw_SC0;	<pre>MySQL [wenjw_db1]&gt; select * from wenjw_SC1;</pre>		
	Sno	Cno Grad	e	Sno		
	2006   2006   2006   2006   2006   2102   2102   2102   2102   2104   2104   2104   2104	12013   88. 12014   96. 12015   87. 12016   92. 12018   100. 12011   65. 12013   66. 12015   70. 12016   69. 12017   60. 12014   79. 12015   80. 12016   73. 12017   66. 12018   85.	0   0   0   0   0   0   8   6   0   0   0   0	2005   12012   70.0		

## 3、实验总结:

- (1) 多主机相连:最开始我是尝试用服务器和虚拟机相连搭建分布式分库分表,但是失败了,后来尝试和同学一起搭建,但是无一例外都失败了。后来我发现是开放端口号和防火墙的问题,在服务器的控制台开放了3306的端口号,用 tcping 工具测试,发现就能 ping 通了。
- (2) 查询语句的报错: ERROR 13000 (44000): SQL check failed,error message:Not allow DML operation without sharding conditions.这可能是分片审计算法做的限制,即避免一些不合理的 sql 对数据库做破坏,这也可以解释为什么有些公司在写 sql 语句的时候会有"where 1=1"的条件: 因为通过审计对 sql 做了限制,不允许写不带 where 条件的 sql 语句(以免修改时破坏整个数据库)。后来我把配置文件中的相关代码注释掉,就不报错了。
- (3)配置每个表的分片时,要配置分库算法,否则按默认的来。这次实验我对学生表和选课记录表做了分片,都是通过 Sno 来分,课程表则是在两个数据库实例中各自保存相同的一份。最开始我是想三个表都分片的,但是,如果学生表按照学号奇偶来分,课程表按照课号奇偶来分,那么在两个数据库实例中就只有学号和课号同奇同偶的记录。由于选课记录表定义了外键,这样一来就只能插入学号和课号同奇同偶的记录了,显然不合逻辑。
- (4)由于(3)的限制,所以我把wjw\_Cou 配置成了广播表,即每个数据库实例中都预先创建好一个同名的表,这样在逻辑数据库中会根据广播表的配置把他们合并为一个逻辑表。本实验的库表逻辑如下:

