目录

[1. 引言 2](#_Toc483525670)

[2. 测试场景模式定义 3](#_Toc483525671)

[2.1 定义 3](#_Toc483525672)

[2.2 测试场景模式分类 3](#_Toc483525673)

[2.2.1 前提 3](#_Toc483525674)

[2.2.2 各列定义解释 3](#_Toc483525675)

[2.2.3 压力模式明细 4](#_Toc483525676)

[3. 测试项的需求分析 6](#_Toc483525677)

[3.1 核心测试项(生产中用的较多的功能，读写测试) 6](#_Toc483525678)

[3.1.1 数据导入测试(预装载) 6](#_Toc483525679)

[3.1.2 数据压缩比测试 6](#_Toc483525680)

[3.1.3 写入性能测试 6](#_Toc483525681)

[3.1.4 查询性能测试 7](#_Toc483525682)

[3.1.5 聚合查询性能操作 7](#_Toc483525683)

[3.2 数据管理的其他辅助功能(生产中用的较少的功能） 7](#_Toc483525684)

[3.2.1 数据update测试 7](#_Toc483525685)

[3.2.2 数据random insert 8](#_Toc483525686)

[3.2.3 删除某个时间之前的数据 8](#_Toc483525687)

[3.2.4 数据导出 8](#_Toc483525688)

[首先导入 8](#_Toc483525689)

[4. 配置文件说明 9](#_Toc483525690)

[4.1 基本函数设置（base-function.conf） 9](#_Toc483525691)

[4.2 传感器值配置(sensor.conf) 9](#_Toc483525692)

[4.3 设备配置(mertric.conf) 9](#_Toc483525693)

[4.4 offline 数据生成(generate-offline.conf) 9](#_Toc483525694)

[4.5 online 数据负载(generate-online.conf) 10](#_Toc483525695)

[4.6 数据库连接配置(database.conf) 10](#_Toc483525696)

# 1. 引言

本文档目的旨在是根据工业时序大数据评测基准《需求规格说明书》进行分析说明。

* 1. 目标测试数据库

TsFile，openTSDB，InfluxDB，Graphite，mySQL。

* 1. 参考文献

2017-05-07工业时序大数据评测基准《需求规格说明书》.docx

* 1. 本文读者
     1. 需求提出者

需求提出者阅读1,2,3章，以便评审，保证后续的设计和实现，不至于偏离原先的需求。

* + 1. 系统开发者

阅读1,2,3,4章，明白具体所要实现的需求，并根据文档进行程序的代码实现。

# 测试场景模式定义

## 2.1 定义

测试场景模式：在进行数据导入性能和查询性能测试时，要模拟真实的测试场景，增加测试压力，以有效测试出各系统的性能。

## 2.2 测试场景模式分类

2.2.1 前提

有一周的历史数据。单设备默认配置，有500个数据采集点，每个采集点数据波形与清华金凤的生产实例数据类似，平均每秒生成一次，采集过去一周，共302,400,000个数据点（所有的参数都是可以在配置文件中配置）。

### 2.2.2 各列定义解释

模式名称：测试场景模式的名称

查询负载压力这种模式下，单位为每秒的聚合查询次数(times/s)

写入负载压力：这种模式下，单位为每秒的单个设备写入数据点数(points/s)

应用测试项：这种模式的应用场景，所用于的测试项

### 2.2.3 压力模式明细

|  |  |  |  |
| --- | --- | --- | --- |
| 模式名称 | 查询压力 | 写入压力 | 应用测试项 |
| 模式一 | 0 | 0 | 3.1.3，3.1.4，3.1.5，3.2.1，3.2.2 |
| 模式二 | 0 | 10 | 3.1.4，3.1.5，3.2.1 |
| 模式三 | 0 | 100 | 3.1.4，3.1.5，3.2.1 |
| 模式四 | 0 | 1000 | 3.1.4，3.1.5，3.2.1 |
| 模式五 | 10 | 0 | 3.1.3，3.1.4，3.1.5，3.2.1，3.2.2 |
| 模式六 | 10 | 10 | 3.1.4，3.1.5，3.2.1 |
| 模式七 | 10 | 100 | 3.1.4，3.1.5，3.2.1 |
| 模式八 | 10 | 1000 | 3.1.4，3.1.5，3.2.1 |
| 模式九 | 100 | 0 | 3.1.3，3.1.4，3.1.5，3.2.1，3.2.2 |
| 模式十 | 100 | 10 | 3.1.4，3.1.5，3.2.1 |
| 模式十一 | 100 | 100 | 3.1.4，3.1.5，3.2.1 |
| 模式十二 | 100 | 1000 | 3.1.4，3.1.5，3.2.1 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **A设备数** | **Read client** |  | **备注** |
| **1** | **A 增加** | **没有用户** | **加压设备数** | **只插入 测写入性能 througput** |
| **2** | **没有设备** | **R增加** | **Client加压** | **只读 测读的性能，Responsedtime、R1、R2** |
| **3** | **A增加** | **R固定** | **加压设备数** | **读写混合 A+Rtime** |
| **4** | **A固定** | **R增加** | **加压clien** | **读写混合 Athrougput + R1+R2** |
| **5** | **A固定** | **R固定** | **分别加压modified负载**  **（update、Rinsert、delete）** | **关于辅助功能的测试**  **ResonseTime** |

**Metric指标**

**:**

# 3. 测试项的需求分析

测试性能指标: 测试项如无特殊说明则性能指标为每秒的数据点数 (sensor,有意义的value值)

## 3.1 核心测试项(生产中用的较多的功能，读写测试)

### 3.1.1 数据导入测试(预装载)

单设备，有500个数据采集点，每个采集点数据波形根据清华金凤的生产实例数据类似，平均每秒采集一次，采集过去一周，共302,400,000个数据点。通过offline模式生成数据，计划手动导入数据或者利用python脚本进行文件各个数据库的文件导入，判断各个数据库的导入性能。

用java执行python脚本进行代码实现

导入性能不同于写入性能

### 3.1.2 数据压缩比测试

根据配置文件，读取数据导入后的文件路径，判断前后数据导入的大小差别，进行压缩比计算，导入前的文件大小为数据生成器的数据文件大小，导入后的文件问数据库文件大小。

性能指标:导入前与导入后的数据压缩比。

3.1.3 写入性能测试

在测试压力模式一，模式五，模式九3种模式下进行测试；

进行连续加压，加压到每秒钟写入1000的点后，然后通过加设备进行加压。

加压过程:

首先使用一个设备，每秒写入1个数据点，计算每秒写入的数据点数，然后再每秒写入2个数据点，计算每秒的数据点数，计算完后再每秒写入3个数据点，计算每秒的数据点数，以此类推，直到单设备每秒写入1000个数据点，计算每秒写入的数据点数，然后增加一个设备，原设备每秒写入1000个数据点不变，新设备从每秒写入1个数据点不断增加到每秒写入1000个数据点，以此类推，不断增加设备，直到全被设备每秒写入的点数增加，而入库的数据点数不再增加，停止加压。

性能指标:连续加压系统每秒写入的数据点数

3.1.4 查询性能测试

进行读取测试(对某个数据点进行查询)

查询限制为单设备

十二个模式下分别对1min,1Hour,1day,1week数据全字段进行查询100次，计算出 min,max,avg,95%,99%;

3.1.5 聚合查询性能操作

查询限制为单设备，同一数据点测试聚合条件，聚合查询完就一条记录，分别是十二个模式下进行计算。

1）1分钟内对设备进行max,min,avg,sum聚合查询各100次，聚合字段为float值字段，分组为sec，设备，设备属性。

2）1小时内对设备进行max,min,avg,sum聚合查询各100次，聚合字段为float类型字段，分组为minute，设备，设备属性。

3）1天内对设备进行max,min,avg,sum聚合查询各100次，聚合字段为float类型字段，分组为hour，设备，设备属性。

3）1周内对设备进行max,min,avg,sum聚合查询各100次，聚合字段为float类型字段，分组为day，设备，设备属性。  
测试指标为延迟时间: ms (对应的时间段，聚合条件,where条件)。

## 3.2 数据管理的其他辅助功能(生产中用的较少的功能）

3.2.1 数据update测试

十二种模式下，在线插入数据，0.001几率错掉一个字段的数据，后面update计算延迟时间, 1小时后进行更新操作。

3.2.2 数据random insert

插入数据时候，随机漏掉数据并且记录，在后面进行批量random insert 操作，计算insert延迟时间，每秒的insert数目。

模式一，模式五，模式九3种模式下，实时插入数据，0.001几率漏掉数据，后面再插入漏掉的数据，1H进行insert。

3.2.3 删除某个时间之前的数据

首先插入2周内的数据

删除昨天同时间之前的数据所用的延迟ms, 平均每sec的所删除的数目

删除所有数据

进行100次操作，计算出min,max,avg,95%,99%

3.2.4 数据导出

首先导入单设备，有500个数据采集点，每个采集点数据波形根据清华金凤的生产实例数据类似，平均每秒采集一次，采集过去一周，共302,400,000个数据点。首先判断目标数据库是否支持导出功能，然后测试导出性能

导出1周数据，3次，计算延迟时间

# 4. 配置文件说明

**4.1 基本函数设置（base-function.conf）**

* 1. 配置各个基本函数参数，基本函数为系统内置，这里可以配置各个函数的参数，各个函数的类型可以重复，但是各个函数的id不可以重复，这个配置文件中的函数id，用于其他配置文件中的引用
  2. 数据类型为数字型(整型，浮点型)的基本函数(均为周期函数)包含正弦，单调斜率唯一函数(包含波和锯齿波)，抛物线，指数，对数

数字类型为非数字型(枚举类型，字符串类型，bool类型) 均显示配置在文件中

* 1. 每个函数均有自己的id，后面所有用到的函数，均需要在这里配置，需要时引用这里的id

应用场景:用于offline数据生成，和online数据负载的函数引用

**4.2 传感器值配置(sensor.conf)**

配置各个传感器的基本函数，以及名称

传感器波形可引用base-function.conf中的各个函数id进行函数组合，设置好连续点可自由组合函数生成数据(必须是相同数据类型，如果所配置的连续点不存在，则按照该不连续的点之前的函数生成数据)

应用场景:用于设备配置引用

**4.3 设备配置(mertric.conf)**

配置各个设备的参数 包含设备的元数据信息(位置，各个传感器的位置及引用)，对应各个传感器的有效时间，设备中对应各个数据采集点（传感器）的 生成频率，数据发送的缓存时间（batchSize）

Ps:配置时候，数据结构要满足各个数据库

**4.4 offline 数据生成(generate-offline.conf)**

(多线程生成，一个设备最多支持一个线程，一个设备对应一个目标文件)

配置目标数据库类型，可配置多个设备，引用4.3 mertirc.conf中的设备，配置每个设备的采集时间段，最后根据设备及目标数据库类型生成相应数据库的目标导入文件

应用场景:Offline数据导入，主要针对相应的数据库import，向数据库种导入数据

**4.5 online 数据负载(generate-online.conf)**

(多线程负载)

配置十二种模式中，各个模式所对应的设备，设备的各个参数(步长，缓存时间等)

**4.6 数据库连接配置(database.conf)**

数据库类型，url，端口，账号，密码信息