目录

[目录 1](#_Toc483661385)

[1. 引言 2](#_Toc483661386)

[1.1 目标测试数据库 2](#_Toc483661387)

[1.2 参考文献 2](#_Toc483661388)

[1.3 本文读者 2](#_Toc483661389)

[1.3.1 需求提出者 2](#_Toc483661390)

[1.3.2 系统开发者 2](#_Toc483661391)

[2. 测试场景模式定义 3](#_Toc483661392)

[2.1 测试过程 3](#_Toc483661393)

[2.2 定义 3](#_Toc483661394)

[2.3 测试场景模式分类 3](#_Toc483661395)

[2.3.1 前提 3](#_Toc483661396)

[2.3.2 各列定义解释 3](#_Toc483661397)

[2.3.3 测试场景模式明细 3](#_Toc483661398)

[3. 测试项的测试过程 6](#_Toc483661399)

[3.1 核心测试项(生产中用的较多的功能，读写测试) 6](#_Toc483661400)

[3.1.1 数据导入测试(预装载) 6](#_Toc483661401)

[3.1.2 数据压缩比测试 6](#_Toc483661402)

[3.1.3 写入性能测试 6](#_Toc483661403)

[3.1.4 简单查询性能测试 7](#_Toc483661404)

[3.1.5 聚合查询性能操作 7](#_Toc483661405)

[3.2 数据管理的其他辅助功能(生产中用的较少的功能） 8](#_Toc483661406)

[3.2.1 数据update测试 8](#_Toc483661407)

[3.2.2 数据random insert 8](#_Toc483661408)

[3.2.3 数据delete 8](#_Toc483661409)

[3.2.4 数据导出 8](#_Toc483661410)

[首先导入 8](#_Toc483661411)

[4. 配置文件说明 9](#_Toc483661412)

[4.1 基本函数设置（base-function.conf） 9](#_Toc483661413)

[4.2 传感器值配置(sensor.conf) 9](#_Toc483661414)

[4.3 设备配置(device.conf) 9](#_Toc483661415)

[4.4 offline 数据生成(generate-offline.conf) 9](#_Toc483661416)

[4.5 online 数据负载(generate-online.conf) 10](#_Toc483661417)

[4.6 数据库连接配置(database.conf) 10](#_Toc483661418)

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，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 测试过程

先配置好配置文件，然后根据配置文件进行历史数据(一般为一周)生成，生成历史数据后，将历史数据导入目标数据库中，使数据处于准备状态，然后对数据进行混合负载测试，包含数据导入测试，数据压缩比测试，写入性能测试，查询性能测试，聚合查询性能测试，数据更新操作测试，数据insert操作测试，数据删除测试，数据导出测试。

## 2.2 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(包含数据实时写入压力，实时聚合查询压力)，以有效测试出各系统的性能。

## 2.3 测试场景模式分类

2.3.1 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）。

### 2.3.2 各列定义解释

模式名称：测试场景模式的名称

Append写入压力：这种模式下，Append写入压力状态

Read 查询压力：这种模式下，Read 查询压力状态

描述：场景描述

性能指标：这种模式下的性能指标

### 2.3.3 测试场景模式明细

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模式名称 | Append写入压力 | Read 查询压力 | 描述 | 性能指标 |
| 模式一 | 从0不断增加写入压力 | 0 | 1. Read查询压力为0； 2. 通过增加设备数不断增加压力，直到写入性能到达瓶颈 | 吞吐量throughput：单位为每秒所写入的点数 |
| 模式二 | 0 | 从0不断增加读压力 | 1. Read查询加压分两类，第一类简单查询为主力，分析聚合查询为辅，比例可配置；第二类为分析聚合查询为主，简单查询为辅，比例可配置； 2. Append写入压力为0，   3，通过增加每秒的查询数，不断增加查询压力，直到到达查询瓶颈 | 1，延迟：单位为平均每查一个点的延迟时间（毫秒/point），  2，吞吐量throughput：单位为每秒所查询的点数 |
| 模式三 | 从0不断增加写入压力 | 固定 | 1. Read查询压力固定（分别为最大压力的0.1，0.4，0.7）； 2. 通过增加设备数不断增加压力，直到写入性能到达瓶颈 | Read:  1，延迟：单位为平均每查一个点的延迟时间（毫秒/point），  2，吞吐量throughput：单位为每秒所查询的点数  Append:  吞吐量throughput：单位为每秒所写入的点数 |
| 模式四 | 固定 | 从0不断增加读压力 | 1. Read查询加压分两类，第一类简单查询为主力，分析聚合查询为辅，比例可配置；第二类为分析聚合查询为主，简单查询为辅，比例可配置； 2. Append写入压力为固定（分别为最大压力的0.1，0.4，0.7）；   3，通过增加每秒的查询数，不断增加查询压力，直到到达查询瓶颈 | Read:  1，延迟：单位为平均每查一个点的延迟时间（毫秒/point），  2，吞吐量throughput：单位为每秒所查询的点数  Append:  吞吐量throughput：单位为每秒所写入的点数 |
| 模式五 | 固定 | 固定 | 1. Read查询压力固定，Append写入压力固定，可配置； 2. 测试update/random insert/delete/导出 | 延迟：单位为平均每查一个点的延迟时间（毫秒/point） |

# 3. 测试项的测试过程

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。通过数据生成器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。

性能指标：吞吐量throughput，单位为每秒所写入的点数

### 3.1.2 数据压缩比测试

本测试主要测试目标数据库的存储性能。

通过3.1.1导入数据后，查看导入前数据包和导入数据库后数据包的大小，计算出前后压缩比。

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

首先用测试模式一，测试模式二测出最大的写入性能MAP（Max Append Perform），最大的简单读取性能MSRP（Max Simple Read Perform），最大的聚合读取性能MARP（Max Aggregate Read Perform），然后在测试模式三下测试在不同读取压力下的写入性能。

进行连续加压，加压到每秒钟写入1000的数据点后，然后通过加设备进行加压。

写入加压过程描述:

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: 吞吐量throughput，单位为每秒所写入的点数

3.1.4 简单查询性能测试

首先用测试模式一，测试模式二测出最大的写入性能MAP（Max Append Perform），最大的简单读取性能MSRP（Max Simple Read Perform），最大的聚合读取性能MARP（Max Aggregate Read Perform），然后在测试模式四下测试在不同写入压力下的简单读取性能。

读取加压过程描述:

从每秒中读取0个数据点，不断增加每秒简单查询的数据点，直到增加简单查询的数据点，而真正的每秒查询到数据点不再增加。

性能指标：

1，延迟：单位为平均每查一个点的延迟时间（毫秒/point），

2，吞吐量throughput：单位为每秒所查询的点数。

3.1.5 聚合查询性能操作

首先用测试模式一，测试模式二测出最大的写入性能MAP（Max Append Perform），最大的简单读取性能MSRP（Max Simple Read Perform），最大的聚合读取性能MARP（Max Aggregate Read Perform），然后在测试模式四下测试在不同写入压力下的简单读取性能。

读取加压过程描述:

从每秒中读取0个数据点，不断增加每秒简单查询的数据点，直到增加简单查询的数据点，而真正的每秒查询到数据点不再增加。

性能指标：

1，延迟：单位为平均每查一个点的延迟时间（毫秒/point），

2，吞吐量throughput：单位为每秒所查询的点数。

## 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

在测试模式五下，在线插入数据，0.001几率错掉一个数据并且记录， 1小时后进行更新操作，计算update延迟时间。

性能指标：延迟时间，单位为每个数据点的延迟时间。

3.2.2 数据random insert

在测试模式五下，在线插入数据，0.001几率漏掉数据并且记录，在后面进行批量random insert 操作，计算insert延迟时间。

性能指标：延迟时间，单位为每个数据点的延迟时间。

3.2.3 数据delete

首先插入2周内的数据

删除昨天同时间之前的数据，然后计算所用的延迟时间，单位为每删除一个数据点所用的毫秒

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. 配置文件设计与说明

* 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
     3. 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
  2. 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(device.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 device.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，向数据库中导入数据

4.5 online 数据负载(generate-online.conf)

(多线程负载)

配置五种模式中，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