

# BenchX

BaseX performance platform comparison

Andy Bunce  
Feb 2015  
BaseX Usergroup  
Prague

# About me

- BaseX user-group-talk XML Prague 2013  
[Adventures with BaseX and Web applications](#)
- A node.js client for the BaseX server.  
<https://github.com/apb2006/basex-node>
- A quick start to run BaseX on OpenShift  
[openshift-basex-quick-start](#)

Twitter: [@apb1704](#)

Github: [apb2006](#)

# What is BenchX?

BenchX is a Web application for comparing BaseX performance across platforms and configurations.

In this talk:

- Why create it?
- How does it work?
- What are the results?

# Why?

BaseX is easy to install!

Generally as simple as...

1. Ensure Java is installed
2. download and unzip `basex.zip`
3. Run `basexgui` or `basexhttp`

... and you have great tools to query and transform data.

# Out of curiosity...

I have tried installing BaseX on various machines around the house...

- PC s Windows and Unix
- Raspberry PI
- NAS (Netgear ReadyNAS Pro)
- Android Tablet and Phone
- How does performance compare?

# Raspberry Pi - installation notes

Runs Debian, costs £25

1. `sudo apt-get install oracle-java8-jdk`
2. `wget http://files.basex.org/releases/7.9/BaseX79.zip`
3. `etc...`



# NAS - NETGEAR ReadyNAS Pro 2

- runs customized, old, version of Debian.
- music+photo store=plenty of metadata for XML query.



1.8 GHz Intel Atom Dual Core  
Processor 1 GB DDR2 SODIMM

# Android - installation notes

Add a range of linux distros to your android device through a method known as 'chroot'



1. Install linux.  
Linux Deploy, on Play store worked for me
2. ssh in
3. Install Java <http://www.webupd8.org/2014/03/how-to-install-oracle-java-8-in-debian.html>
4. wget <http://files.basex.org/releases/7.9/BaseX79.zip>
5. etc
6. setup startup script for basexhttp





# Cloud - installation notes

## OpenShift

- A quick start to run BaseX on OpenShift  
[Quodatum/openshift-basex-quick-start](#)
- Use BaseX on OpenShift  
[dirkk/openshift-basex](#)

## Amazon Web Services

- Amazon Web Services EC2 setup (Paul Swennenhuis)  
[http://www.swennenhuis.nl/basexfordummies/BaseX\\_for\\_dummies\\_part\\_2.pdf](http://www.swennenhuis.nl/basexfordummies/BaseX_for_dummies_part_2.pdf)

## Docker

- Docker image for BaseX 8.0 (Andreas Jung)  
[docker-basex](#)

# Why build BenchX?

- To get some numbers (and charts) to answer “How does the performance of these platforms compare?”
- ...and because I like building web applications

# How?

Run the same queries on each platform and compare times.

Which queries?

- Start with the XMark queries

## XMark-An XML Benchmark Project

- The classic XQuery benchmark
- 20 Queries
- test data models auction site
- C executable to generate test data
  - factor to vary generated data size

# Sample XMark query (q05.xq)

```
let $auction := collection("benchx-db") return
count(
  for $i in $auction/site/closed_auctions/closed_auction
  where $i/price/text() >= 40
  return $i/price
)
```

# XMark data

```
7      <quantity>1</quantity>
8      <name>but nine eighteen </name>
9      <type>teard</payment>
10     <description>
11       <parlist>
12         <listitem>
13           <text> page rous lady idle authority capt professes stabs monster
14             petition heave humbly removes rescue runs shady peace most piteous
15             worser oak assembly holes patience but malice whoreson mirrors
16             master tenants smocks yielded <keyword> officer embrace such fears
17             distinction attires </keyword>
18           </text>
19         </listitem>
20         <listitem>
21           <text> shepherd noble supposed dotage humble servilius bitch theirs
22             venus dismal wounds gum merely raise red breaks earth god folds
23             closet captain dying reek </text>
24         </listitem>
25       </parlist>
26     </description>
27     <shipping>Will ship internationally, See description for charges</shipping>
28     <incategory category="category0"/>
29     <incategory category="category0"/>
30     <incategory category="category0"/>
31     <incategory category="category0"/>
32     <incategory category="category0"/>
33     <mailbox>
34       <mail>
35         <from>Dominic Takano mailto:Takano@yahoo.com</from>
36         <to>Mechthild Renear mailto:Renear@acm.org</to>
37         <date>10/12/1999</date>
38         <text> asses scruple learned crowns preventions half whisper logotype
39           weapons doors factious already pestilent sacks dram atwain girdles
40           deserts flood park lest graves discomfort sinful conceiv therewithal
41           motion stained preventions greatly suit observe sinews enforcement
42           <emph> armed </emph> gold gazing set almost catesby turned servilius
43           cook doublet preventions shrunk </text>
44       </mail>
```

Factor	Size
0	27kb
.5	58Mb
1	116Mb
10	1.4Gb
100	11Gb

# JacobSnyder/XQuery-Benchmark

- Shell script
- Runs all XMark queries
  - for range of data sizes
- Has query timeout
  - kill run after 10mins
- tab-separated text output

<https://github.com/JacobSnyder/XQuery-Benchmark>

# But...

Unix only.

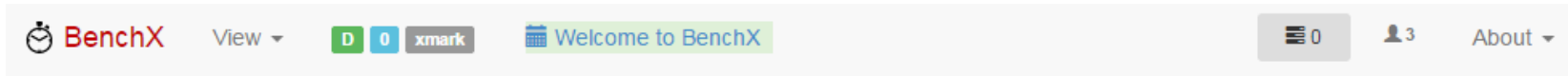
No support after the numbers are generated.

For viewing, managing, comparing

So BenchX...



# BenchX : A Web Application

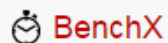


## BenchX

This application times the execution of [▶ suites](#) of XQuery queries. It also manages a [📖 library](#) of results comparing the relative performance across different [🌐 environments](#) for the [BaseX](#) XML database.

- This application requires internet access (for CDN and Google chart).
- This application is not designed for multiple session use.
- BenchX source on [Github](#)
- Stopwatch icon from [SimpleIcon](#)

# Suites



View ▾

F 0 xmark

Welcome to BenchX

☰ 0

👤 1

About ▾

🏠 / suites

## sample1

queries: 2 , results: 2

Sample suite mainly as an example of suite functionality. Contains the xmark q20 and rewritten version using XQuery 3 group by features

## xmark-polo

queries: 18 , results: 3

The xmark-polo suite is the standard xmark suite with queries 11 and 12 removed for performance reasons <http://www.ins.cwi.nl/projects/xmark/>

## xmark

queries: 20 , results: 0

The aim of the XMark project is to provide a benchmark suite that allows users and developers to gain insights into the characteristics of their XML repositories. <http://www.ins.cwi.nl/projects/xmark/>

# Library

[Home](#) / [library](#) Action: [Schema \(xsd\)](#)

Title	Suite	BaseX	Java	OS	Hostname	OS Arch	Created	Runs
<a href="#">short</a>	<a href="#">sample1</a>	8.0 beta 2f72800	1.7.0_60	Windows 8.1	redfish	amd64	3 days ago	16
<a href="#">12</a>	<a href="#">sample1</a>	8.0 beta 2f72800	1.7.0_60	Windows 8.1	redfish	amd64	4 days ago	12
<a href="#">DF .1 .2</a>	<a href="#">xmark</a>	8.0 beta 2f72800	1.7.0_60	Windows 8.1	(unknown)	amd64	4 days ago	8
<a href="#">f0.3</a>	<a href="#">sample1</a>	8.0 beta 2f72800	1.7.0_60	Windows 8.1	(unknown)	amd64	4 days ago	2
<a href="#">one</a>	<a href="#">sample1</a>	8.0 beta 2f72800	1.7.0_60	Windows 8.1	(unknown)	amd64	4 days ago	2
<a href="#">save</a>	<a href="#">sample1</a>	8.0 beta 2f72800	1.7.0_60	Windows 8.1	(unknown)	amd64	4 days ago	96
<a href="#">test3</a>	<a href="#">sample1</a>	8.0 beta 2f72800	1.7.0_60	Windows 8.1	(unknown)	amd64	4 days ago	72
<a href="#">another</a>	<a href="#">sample1</a>	8.0 beta 2f72800	1.7.0_60	Windows 8.1	redfish	amd64	4 days ago	48

Filter (@TODO)

Name Value

# Environments

[Home](#) / Environments

Environment display needs more work

<b>Windows 8.1</b> Basex: 8.0, Java: 1.7.0_60	env: 0
<b>Linux</b> Basex: 8.0 beta 89eb29e, Java: 1.7.0_21	env: 1
<b>Windows 7</b> Basex: 8.0 beta 3a4b860, Java: 1.8.0_20	env: 2
<b>Linux</b> Basex: 8.0, Java: 1.8.0_31	env: 3
<b>Linux</b> Basex: 8.0 beta 2f72800, Java: 1.7.0_71	env: 4
<b>Linux</b> Basex: 8.0 beta 2f72800, Java: 1.7.0_75	env: 5
<b>Linux</b> Basex: 8.0 beta 49700b6, Java: 1.7.0_40	env: 6
<b>Windows 8.1</b>	env: 7

# results - table

Results: Suite: **sample1**, Host: **(unknown)**, Created: 2 days ago, Runs:12

file vs db

Query	F0	D0	F0.5	D0.5	F1	D1
q20.xq	Runs: 1 0.218	Runs: 1 0.179	Runs: 1 2.687	Runs: 1 0.19	Runs: 1 5.34	Runs: 1 0.235
xq20.xq	Runs: 1 0.11	Runs: 1 0.106	Runs: 1 2.626	Runs: 1 0.13	Runs: 1 5.054	Runs: 1 0.154



# Data sources

- Want to run against file or database
- So queries use `collection("benchx-db")`  
will use database if it exists otherwise the folder

The datasource is defined by two parameters:

mode: "D" or "F" meaning database or file

factor: data size, floating point number used by `xmlgen`

```
xmlgen /f 100 /o benchx-db/s /s 400
```

# BenchX Databases

Two databases are used:

1. *benchx*: results library and application data.
    - Created if not found using data folder
    - results are documents within the db
  2. *benchx-db*: query data generated by *xmlgen*.  
created and dropped by the application.
- (This enables application bootstrap)



# BenchX Database in DBA

## Database Administration

[Databases](#) | [Queries](#) | [Logs](#) | [Users](#) | [Settings](#) | [Logout](#)

### Databases » benchx

#### 13 RESOURCES:

- [library/118f256e-035c-44b6-a618-c0d09c924e33.xml](#)
- [library/16715e33-5a08-46a3-9c5b-7800ecb3face.xml](#)
- [library/30a2a185-d0f6-4de4-b08e-66393c887c15.xml](#)
- [library/5c3dadd7-197e-48f1-9111-3f30670267d9.xml](#)
- [library/6ae0f484-7bf5-4643-9e05-07e868793937.xml](#)
- [library/6e40f025-0054-4ca1-be99-9ba1930680fa.xml](#)
- [library/7edd38b9-bc3b-45ec-848f-788b2291a02b.xml](#)
- [library/914d16c7-02aa-42a8-9c38-a394db9c9181.xml](#)
- [library/998fef1c-53b9-4582-96f8-ffe59b226f4.xml](#)
- [library/b2ed7ad7-3b34-4346-a676-0f507cf5d295.xml](#)
- [library/e004d7da-2268-4253-ae1e-1debb48048e0.xml](#)
- [benchmark.xml](#)
- [state.xml](#)

#### CONTENT TYPE RAW

application/xml	-
application/xml	-
application/xml	-
application/xml	-
application/xml	-
application/xml	-
application/xml	-
application/xml	-
application/xml	-
application/xml	-
application/xml	-
application/xml	-
application/xml	-

#### DATABASEPROPERTIES

NAME	benchx
SIZE	922 KB
NODES	43407
DOCUMENTS	13
BINARIES	0
TIMESTAMP	2015-02-08-15-14-41

#### RESOURCEPROPERTIES

TIMESTAMP	2015-02-08-15-14-36
ENCODING	UTF-8
CHOP	✓

#### INDEXES

UPTODATE	✓
TEXTINDEX	✓
ATTRINDEX	✓

# Library XML format

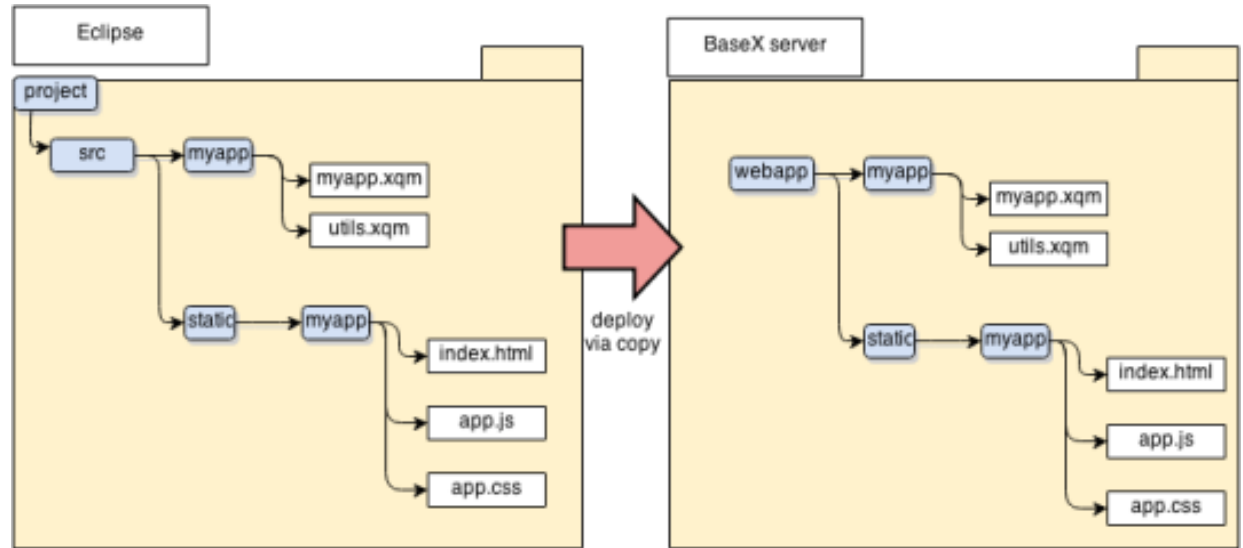
- Unique ID (GUID)
- Metadata
- Environment
- Runs
  - query name
  - mode + factor
  - time

```
<benchmark>
  <id>5c3dadd7-197e-48f1-9111-3f30670267d9</id>
  <suite>xmark-polo</suite>
  <created>2015-01-13T21:34:44.707Z</created>
  <description>full on laptop</description>
</meta>
<server>
<environment>
  <basex.version>8.0 beta 2f72800</basex.version>
  <java.vendor>Oracle Corporation</java.vendor>
  <java.version>1.7.0_60</java.version>
  <java.vm.version>24.60-b09</java.vm.version>
  <os.arch>amd64</os.arch>
  <os.name>Windows 8.1</os.name>
  <os.version>6.3</os.version>
  <runtime.availableProcessors>4</runtime.availableProcessors>
  <runtime.freeMemory>436 MB</runtime.freeMemory>
  <runtime.maxMemory>455 MB</runtime.maxMemory>
  <runtime.totalMemory>455 MB</runtime.totalMemory>
</environment>
<runs type="array">
  > > <run>
  > > <run>
    <runtime type="number">0.441</runtime>
    <status/>
    <name>q01.xq</name>
    <mode>D</mode>
```

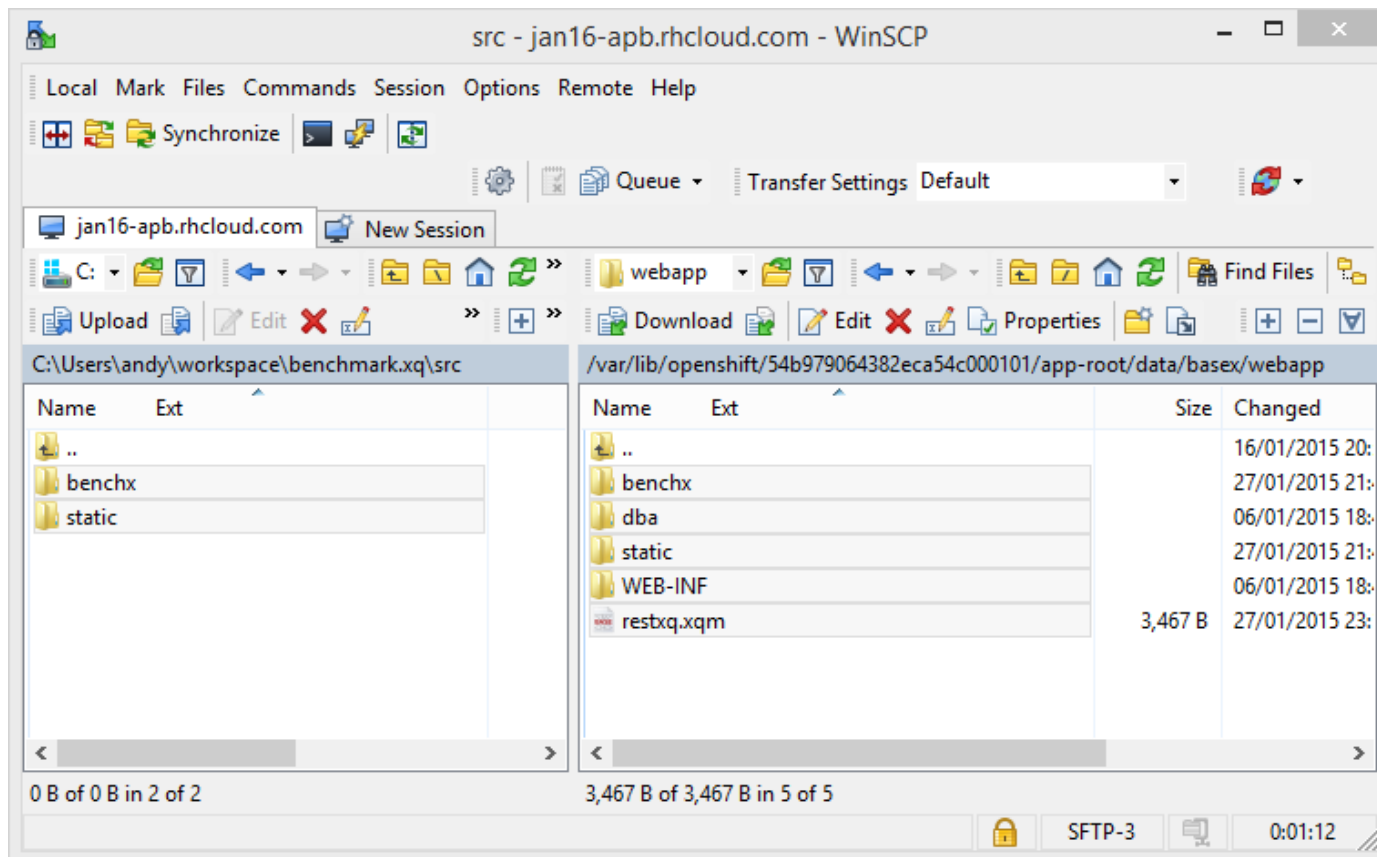
benchmark.xsd

# Project structure

- `{$app-name}/`
- `static/{$app-name}`



# Deploy via copy...



# Client Architecture (static)

- Angular SPA talks to BaseX RESTXQ using json

• benchx	GET
• benchx/api/environment	GET
• benchx/api/execute	POST
• benchx/api/library	GET
• benchx/api/library/{\$id}	DELETE
• benchx/api/library/{\$id}	GET
• benchx/api/library/{\$id}/compare	GET
• benchx/api/session	DELETE
• benchx/api/session	POST
• benchx/api/state	GET
• benchx/api/state	POST
• benchx/api/suite	GET
• benchx/api/suite/{suite}	GET
• benchx/api/thisenv	GET
• benchx/api/~testbed	GET

## RestXQ API: /benchx

wadl.xml 

benchx	text/html	GET	Benchmark html application entry point. Will create db if required	run	ang
benchx/api/environment	application/json	GET	distinct environments from the library	run	ang
benchx/api/execute	application/json	POST	Execute one test and store to session	run	ang
benchx/api/library	application/json	GET	list of library files	run	ang
benchx/api/library/{\$id}	application/json	DELETE			

# Javascript libraries

angular-hotkeys

angular-moment

angular-ui-bootstrap

angularjs

AngularLogExtender

async

FileSaver

lodash

moment

ng-google-chart

ngStorage

stacktrace

twitter-bootstrap

Components used 13

angular-hotkeys 1.2.0

Configuration-centric keyboard shortcuts for your angular apps

Runat: browser

Used by:

Depends on: angularjs

Versions

- 1.2.0

# Server architecture

- RESTXQ
- run queries with `xquery:eval` (allows timeout)
- `proc:execute` out to run `xmlgen`

# Server Architecture

No good way to start long running queries from http. Request is timed out after some minutes.

So Client

- creates a queue of tasks to run on server.
- Runs as sequence REST calls



# Run options

[Home](#) / [suites](#) / [xmark](#) / [run](#) /  Grid  Graph

Action:

## 1. Initial state

Mode:

Factor:  (27.65 kB)

## 2. then

Run other mode:

use increments:

Factor incr:

Factor max:

## 3. Repeat (1) and (2)

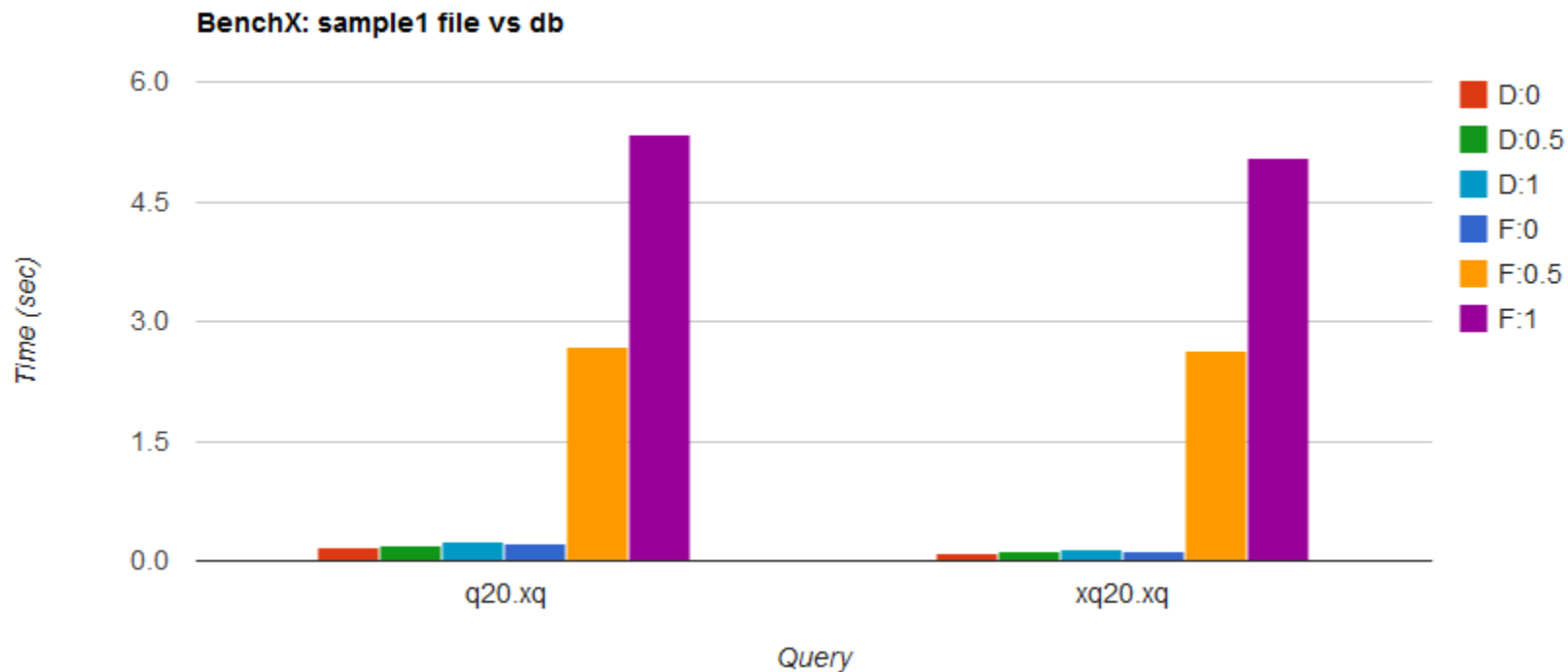
Times:

Queue and execute a set of BenchX tasks

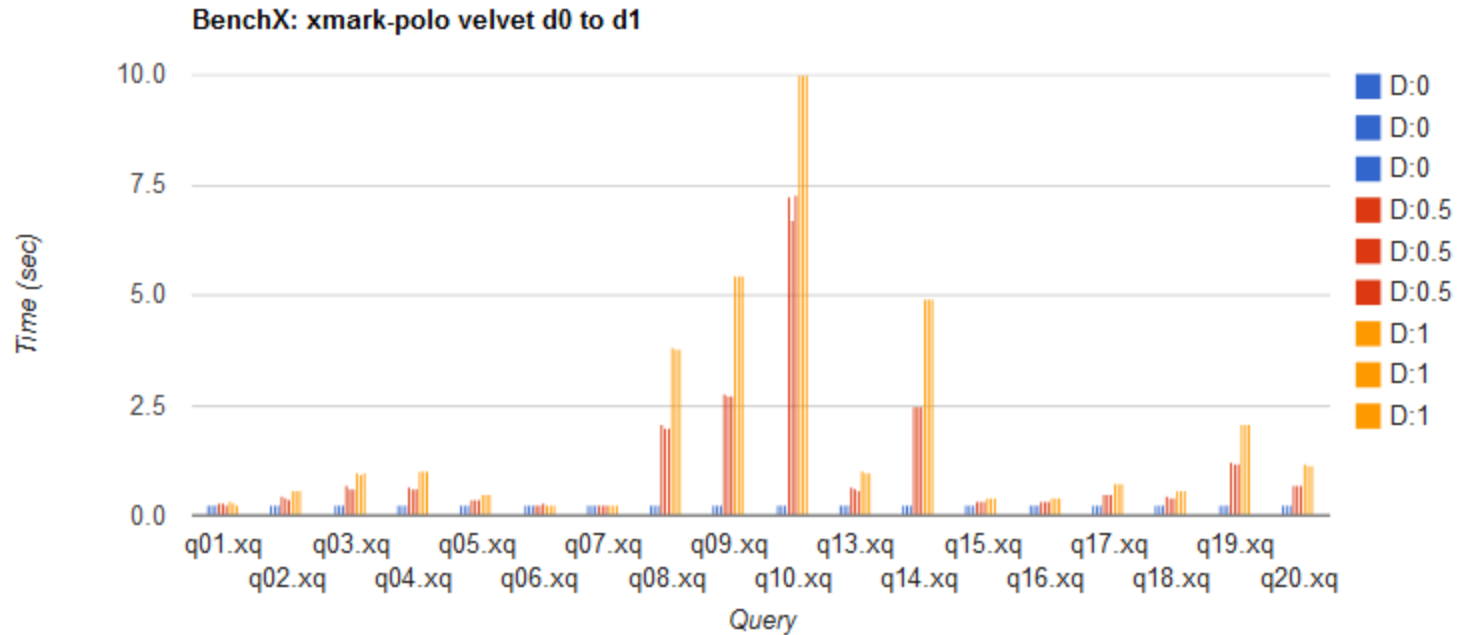
# Demo

- run sample f+d 0 and 0.5
- save
- compare

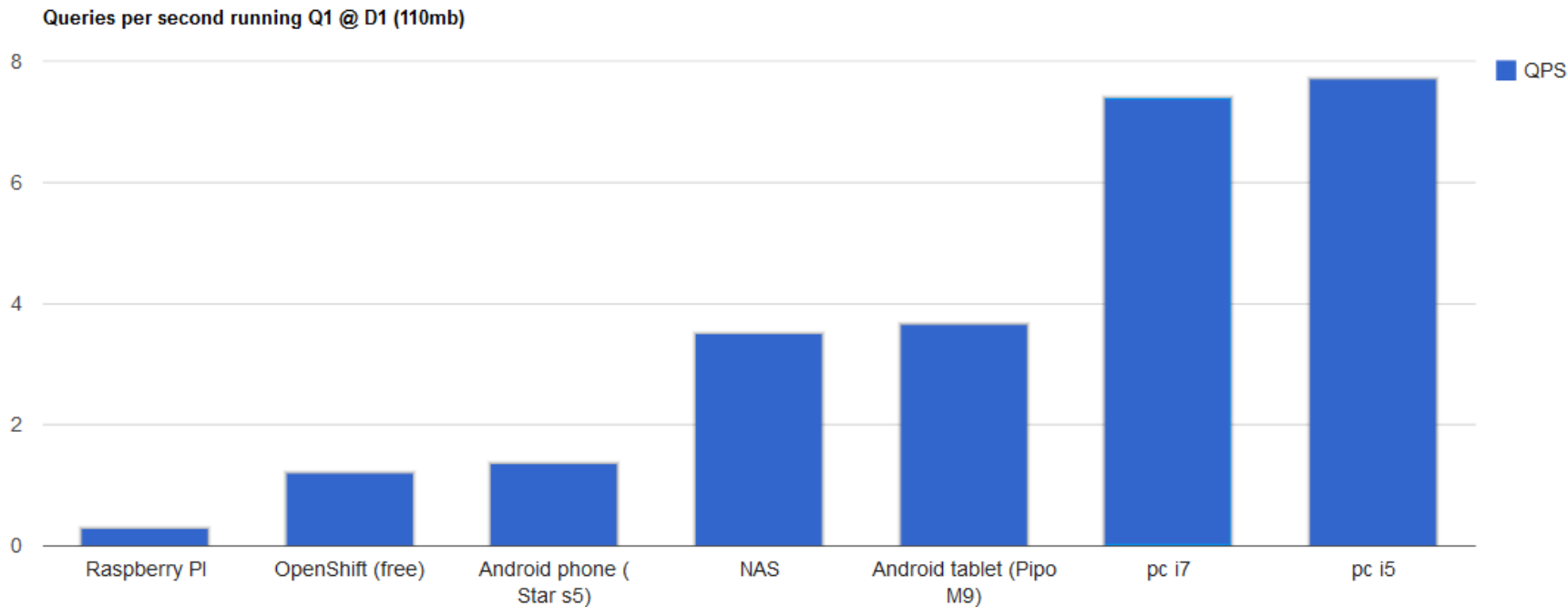
# Results - File vs db



# Results - varying the factor



# Results - platforms



# Related work

- [An Analysis of the Current XQuery Benchmarks](#)  
Loredana Afanasiev, Maarten Marx
- [MemBeR: A Micro-benchmark Repository for XQuery](#)  
Loredana Afanasie, Ioana Manolesc, Philippe Michiels
- [Benchmarking XSLT Performance](#)  
Michael Kay, Debbie Lockett
- [MonetDb XMark Benchmark](#)  
MonetDb

# The BenchX source

<https://github.com/Quodatum/BenchX>

## Contributions welcome

- Results for the library
- New suites
- Bug fixes/improvements

# Future work

- clean up code
- xmlgen use option to split into smaller files
- create pluggable data sources per suite

@TODO

- \* If you have real-life data, store this data in the systems of your choice x
- \* Formulate real-life queries that give you similar results ✓
- \* Compare these measurements to get an impression on the overall performance ✓

[Christian Grün Tue, 20 Jan 2015 10:43:13 -0800](#)



# The End

- Thanks
- Questions