



DataBench

Evidence Based Big Data Benchmarking to Improve Business Performance

D3.3 DataBench Toolbox – Beta including end-to-end-scenario tool

Abstract

This is a supporting document of the demonstrator of the Beta version of the DataBench Toolbox. The document provides an overview of the features evolved from the Alpha version including support for defining use cases when running benchmarks as well as the definition of modules in charge of handling the results and extracting the technical metrics from the data stored in the Toolbox backend.

This document is the third deliverable related to the DataBench Toolbox after D3.1 and D3.2. It provides an update of the development presented in D3.2 in the light of the advances done in the Beta version. More updates will be provided as part of the upcoming release of the Toolbox scheduled in the DataBench WP3 lifecycle.



Deliverable D3.1	DataBench architecture
Work package	WP3
Task	3.3
Due date	31/12/2019
Submission date	23/12/2019
Deliverable lead	ATOS
Version	1.0
Authors	ATOS (Tomás Pariente, Iván Martínez and Ricardo Ruiz) GUF (Todor Ivanov)
Reviewers	JSI (Marko Grobelink), SINTEF (Arne Berre)

Keywords

Benchmarking, big data, big data technologies, architecture, business performance, performance metrics, Toolbox, use cases

Disclaimer

This document reflects the authors view only. The European Commission is not responsible for any use that may be made of the information this document contains.

Copyright Notice

Copyright belongs to the authors of this document. Use of any materials from this document should be referenced and is at the user's own risk.

Table of Contents

Executive Summary	5
1. Introduction	6
2. DataBench Toolbox Beta version essentials	7
2.1 Beta version - Toolbox processes covered	7
2.2 Beta version - Toolbox architecture update	8
2.3 Beta version - Toolbox back-end.....	9
2.4 Beta version - Toolbox front-end.....	10
2.5 Beta version – Benchmarks integrated so far	11
2.6 Beta version – Features included to describe benchmarks	11
3. DataBench Toolbox Beta version back-end.....	12
3.1 Support for adding and configuring benchmarks	12
3.2 Support for deployment of benchmarks	12
3.3 Support for retrieving results	12
3.4 Metrics spawner.....	13
4. DataBench Toolbox Beta version front-end	14
4.1 User management.....	14
4.2 Benchmark management.....	14
4.3 Results.....	17
5. Conclusions and future work.....	21
6. References.....	22

Table of Figures

Figure 1. User deletion page.....	14
Figure 2. Benchmark selection page.....	15
Figure 3. Benchmark update and delete page	16
Figure 4. Benchmark use case selector	17
Figure 5. Search by use case	17
Figure 6. Results list for a user	18
Figure 7. Detailed result with use case.....	18
Figure 8. Result charts.....	19

List of Tables

Table 1 – Processes and functions covered in the Beta version	8
Table 2 – Example of logs	9

Executive Summary

This document is a demonstrator of the second release of the DataBench Toolbox, also known as Beta version. It is therefore providing an overview of the second version of the implementation of the Toolbox. Instead of explaining the complete Toolbox, the document reports in the updates and new developments after the Alpha version released 6 months (reported in the DataBench deliverable D3.2) before outlining the status of the implementation at the time of submission (December 2019). For a complete overview of the functionalities covered it is advisable to read before deliverable D3.2.

The document is a “demonstrator” deliverable, meaning that the DataBench Toolbox is accompanied with software and a demo version of the Toolbox Dashboard accessing to the backend functionality and integrated benchmarks. The URL of the Beta version is not available yet to the public but can be accessed on request.

The final release of the Toolbox is expected by mid-2020. This document provides an overview of the expected functionalities to be covered by the Toolbox. So far, Alpha and Beta versions cover mostly aspects related to the cataloguing, deployment, execution and gathering of metrics related to existing big data benchmarks. The translation to business insights is expected to be part of the final version.

1. Introduction

This deliverable is a demonstrator of the Beta release of the DataBench Toolbox. It provides software, a web dashboard and an accompanying document (this document) to give overview of the current status of the implementation of the Toolbox.

As already stated in deliverable D3.2, “the DataBench Toolbox will include formal mechanisms to ease the process of reusing existing or new big data benchmarks into a common framework that will help stakeholders to find, select, download, execute and get a set of homogenized metrics.” The Beta version is aiming towards this particular goal, while the more ambitious objective of translating technical metrics into recommendations and business insights will be part of the final version of the Toolbox, as this particular objective involves the integration of multiple results from other work packages that are being delivered at the time of releasing this version and during the following months.

This document reports only the changes done after the Alpha version released 6 months before (reported in the DataBench deliverable D3.2). Therefore, the document does not repeat what is explained in D3.2, but refers to the and addition changes made, outlining the status of the implementation at the time of submission (December 2019).

The document follows a similar approach as Deliverable 3.2:

- Section 1 provides the introduction to the objectives of the deliverable.
- Section 2 is a summary of the essential features of the DataBench Toolbox Beta version.
- Section 3 dives into the back-end processes updated in this Beta version.
- Section 4 provides screenshots of the updated parts of the fronted in this Beta version of the Toolbox.
- Finally, Section 5 provides the conclusions of the document as well as outlining the future work on the DataBench Toolbox.

2. DataBench Toolbox Beta version essentials

The Beta version of the Toolbox is an iterative update over the Alpha version. It is focused on the end-to-end functionality and tries to cover a bit more the business user point of view. This last aspect is still work in progress and some functionalities are left for the final version due in M30 of the project.

This section aims to explain briefly the main aspects evolved in the Beta version of the DataBench Toolbox.

2.1 Beta version - Toolbox processes covered

Deliverable D3.1 [1] provides a detailed list of the processes that take part in the life-cycle of the DataBench Toolbox. This section describes the updates and improvements done to the Toolbox from the Alpha version grouped in the same way as described in D3.1 and a set of expected functionalities related to the main elements expected for the final version. For more details in the processes, please refer to D3.1. section 3.

Process	Covered in Beta version	Remaining functionality
Accessing	Improved user management. The Toolbox now allows deleting them from the web.	Better management of access control for guest users Improve the GUI for sign-in, log-in and administrative management Improve and complete the definition for Business users and Benchmark providers, as well as the user grants by the administrator.
User Intentions	Guided search by features Guided search by use case	Improved visual search
Setup and Runtime	Improved the module that monitors Kafka[11] and stores the results into the database to handle any new benchmark that will be added. Improved the result injection to handle any future benchmark with little to no changes in the code. Implemented the metrics spawner module to extract technical metrics from the benchmark results	Integrate the configuration of more benchmarks (so far only 4 have been integrated) for deployment and execution Implement functionality for Technical users to download the Ansible Playbook for “offline” execution of benchmarks Improve the configuration process Improve software to homogenize technical metrics as much as possible

Process	Covered in Beta version	Remaining functionality
Analytics and metrics management	<p>Some simple analytics are done over the results to link use cases and benchmarks.</p> <p>Results are parsed into metrics and can be shown to the user for better understanding of their results</p>	<p>Further analytics for the results still pending to be implemented</p> <p>Analysis of the usage of the Toolbox</p>
Visualization and Reporting	<p>Inventory of benchmarks</p> <p>Metadata-based Google-like search box</p> <p>Filtered advanced search by an initial set of metadata</p>	<p>The visualization functionality is so far minimal, but enough for showcasing the Beta version.</p> <p>Most improvements will be dealing with search functionality (advanced and guided search), visualization of technical metrics and visualization of business-related aspects and knowledge</p>
Benchmark Management	<p>Full CRUD operations over the benchmarks: update and deletion now possible.</p> <p>Added a logging system in cooperation with WP5 to analyse the platform usage.</p> <p>Developed a module to extract metrics from the benchmark results</p>	<p>No changes or just small improvements</p>

Table 1 – Processes and functions covered in the Beta version

Table 1 shows the main elements covered and what is missing for the final version, but there is more functionality to be integrated, especially related to the identification of business insights. This has not been covered yet in the Alpha or Beta versions.

As it was highlighted in the report of the Alpha version of the Toolbox, the Graphical User Interface (GUI) was not planned until the final version, but it was delivered on purpose to help in the dissemination a potential early testing of the Toolbox. By having an easy to use GUI the Toolbox can be demoed in different conferences and with end users. Moreover, it is available for the Beta projects to start using it speeding up the development and bug fixing of the project.

2.2 Beta version - Toolbox architecture update

The architecture of the DataBench Toolbox is described in detail in Deliverable D3.1 and it remains mostly unchanged in this Beta version.

The main change made to the original architecture is the addition of the logging framework needed for the metrics dashboard defined in Deliverable D5.3 [3] which is explained in more detail in the following section.

2.3 Beta version - Toolbox back-end

The Beta version of the Toolbox introduced some modifications to support the implementation of the metrics dashboard defined in Deliverable D5.3 [3].

Specifically, a new log framework was developed ad-hoc so that we can extract from the Toolbox logs the information needed by the Metrics dashboard defined in D5.3.

The Metrics dashboard and how will it extract the information from the different Toolbox data sources is further explained in D5.3. This section explains how the Beta version of the Toolbox implements the logging.

The Toolbox is conformed of several modules in charge of different tasks. The main module is the web user interface from where users can select and run catalogued benchmarks. There are also modules to get the benchmark execution results from Apache Kafka storing the results into the backend database or extracting those results from the database to be parsed into metrics that can be further analyzed and shown to the users. All these modules create in the Beta version logs in different ways due to the different functionalities that they cover.

To ease the work of extracting those metrics to show them in a dashboard WP3 in collaboration with WP5 have defined a common structure of logs to be created by the different tools used in the Toolbox. The logs of all the modules now share the structure shown in Table 2:

Logs example:
<code>[info] c.Ansible_Tower_Controller - [User: rruizsaiz][Action: List Benchmarks][Extra messages][Status: SUCCESS]</code>
<code>[info] c.Ansible_Tower_Controller - [User: rruizsaiz][Action: List benchmark for update][Integrated: 3][Not_integrated:12][Status: SUCCESS]</code>
<code>[info] c.UserController - [User: rruizsaiz][Action: About me][Status: SUCCESS]</code>
<code>[info] c.Ansible_Tower_Controller - [User: rruizsaiz][Action: Search][Integrated: 0][Not_integrated:1][Search param:hadoop][Status: SUCCESS]</code>

Table 2 – Example of logs

The meaning of the tags used in Table 2, that follows the pattern: `[info] [User: <USER>][Action: <Action>][Extra messages][Status: <Status>]`, is the following:

- **[info]:** Standard Apache log level. Can be any of the following in severity order: debug, info, warning, error
- **<User>:** Defines the user that performs the action
- **<Action>:** Describes the action performed. E.g.: Index page render, results page render, update benchmark...
- **<Status>:** SUCCESS or ERROR, indicates whether the action described was fulfilled completely or if there was any error in the Toolbox while performing it.
- **[Extra messages]:** This is an undefined number of extra messages that can be important for the metrics collection point of view. The current options include the number of benchmarks shown in the page, the benchmarks integrated, the ones still not integrated, the search parameter used in the search page...

2.4 Beta version - Toolbox front-end

Although the main effort has been centred in the results processing and in the business user point of view, this Beta version delivers some extra changes done to the GUI that complement the previously commented changes.

These are not big impact changes, but additional pages and small tweaks introduced to slightly improve the user feeling when using the platform. The technologies behind the front-end remains unchanged from the previous deliverables.

A summary of the main updates made in the Beta version with respect to the Alpha version are the following:

- User management:
 - Improved the look and feel of the users' list.
 - Added the possibility to delete a user from the platform.
- Definition and searching of benchmarks:
 - Improved the look and feel of the benchmarks list.
 - Added logic to update metadata for an already added benchmark.
 - Added logic to delete/disable benchmarks.
 - Added the use case selection to the configuration and launch page.
 - Improved the guided search to allow searching by use case.
- Gathering and showing results from the execution of the benchmarks:
 - Improved the inject results to allow extra benchmarks.
 - Added some graphical visualization to the results.

More detailed information as well as some screenshots with the new changes is described in Section 4

2.5 Beta version – Benchmarks integrated so far

As stated in the previous sections, this Beta version of the Toolbox has been more focused on improving the modules working with the results of the benchmarks and trying to highlight the link with the business part defined in the other WPs.

As a result of that, no extra benchmark has been integrated in the Toolbox, only some minor fixes and updates to already integrated benchmarks have occurred during this Beta version.

On the other hand, as a result of the work performed by partners in the project (especially the work done in WP1 in the scope of the revision of the deliverable D1.2 [5]), there has been an extension of the list of benchmarks categorized and complemented with extra information. These benchmarks have been added to the Toolbox to be searched and its metadata has been properly filled with the information given in that list.

2.6 Beta version – Features included to describe benchmarks

In this Beta version of the Toolbox no further features have been included to describe the benchmarks. The only update on this topic is the inclusion of new benchmarks to be searched with the already existing features.

3. DataBench Toolbox Beta version back-end

3.1 Support for adding and configuring benchmarks

This functionality detailed in Deliverable D3.2 has been improved in the Beta version.

The user can now specify the use case for what they are running this benchmark when configuring the run. The platform provides a combo box with the list of use cases extracted from Deliverable D1.1 [4], so the user can select one of them to tag their run.

These results and use cases are later on analysed by the platform to provide the rest of the users with some insights of it. The search functionality makes use of this analysis to calculate the benchmarks that are most used within the platform to test a use case. This information can help present and future users that come to DataBench to look for benchmarks for a similar use case.

3.2 Support for deployment of benchmarks

The changes introduced in this version of the Toolbox have been minor fixes and improvements done to the ansible recipes of the already integrated benchmarks.

Moreover, some extra information has been introduced to help the user identify and understand some of the configuration variables needed for some benchmarks.

Other than that, the support for the deployment of benchmarks is the same as it was in the Alpha version.

3.3 Support for retrieving results

Handling the results of the benchmark runs and extracting information for future analysis have been the objectives of this Beta version. Hence the main modifications have been done under the hood of the Toolbox by improving the modules related to handling the results and extracting the metrics from those results.

To adopt those updates, some extra changes needed to be done also in the data model handling them. Specifically, the list of changes inserted in the database are the following:

- Added a field for each benchmark to set up the Apache Kafka topic related to it where it will send its results.
- Added a field to define the use case for each benchmark run.
- Redefined the tables related to the metrics to accommodate the new modules.

Some of these modifications in the data model have led us to introduce new updates in the Toolbox.

For example, since the database now stores a list of Kafka topics that are available for the benchmarks, some updates were introduced in the logic to inject results from the web tool to the platform. It now allows to upload any result related to any of the benchmarks stored

in the database (even if they are not integrated to be run from the Toolbox) and the results will follow the process and be stored in the database ready to be shown to the user.

On top of that, the module in charge of monitoring the Kafka queues to look for new results have also been updated so that now it gets the list of Kafka topics from the database and updates the monitored topics in real time rather than needing to be restarted every time a new topic needed to be monitored.

3.4 Metrics spawner

One of the biggest parts of this new iteration of the Toolbox is the one related to the results parsing to extract the metrics.

Since this is a key component of the Toolbox it has involved tuning or updating several parts of the platform, from the data model to the frontend to show the results of this parsing. This module is the one in charge of extracting technical metrics from the text results of the benchmarks run as soon as they are available in the database. This is key for the project and especially problematic due to the heterogeneous nature of the benchmark results because they usually don't calculate the same metrics and the near real time interaction needed to show the users the calculated metrics as soon as the results are sent to the platform.

To overcome this situation, we have created a modular implementation of a metric spawner focusing the efforts on making the implementation of new benchmark parser as easy as possible. For this we have implemented a set of shared functionalities hidden from the developers, so they only must worry about parsing the text result into specific metrics. In particular we have implemented the data model to store the metrics in a standardized way as well as the logic to save those metrics in the database.

For a benchmark provider or anyone interested in providing a parser for any benchmark the only task is to extend one of the Java classes and implement the parsing method. This abstract class has some utilities implemented that are described below:

- Data structure of a Metric already defined and ready to be used.
- List of metrics ready to add new metrics
- Logic to automatically save the metrics from that list into the database
- Additional methods to retrieve the information of the benchmark run and the benchmark metadata associated to the results that is parsed.

This module has been implemented for YCSB[13] and Hibench [14] benchmarks already integrated in the platform so far and the results of that extraction is visualized as will be further explained in Section 4

4. DataBench Toolbox Beta version front-end

As stated before, this GUI was not planned for this Deliverable, but we found it useful to showcase the Toolbox in the different conferences and meetings.

In this section we will present the evolution of the GUI developed for the Toolbox. This document will not go again through the pages shown in Deliverable D3.2 but will only showcase the pages that have been improved or newly created.

We can categorize the modifications in three groups.

4.1 User management

User management from the Toolbox is an important need for the Beta version of the Toolbox since one of the objectives is to use it to start onboarding projects. Due to that, some work has been done in the user management functionality. In the Alpha version the Toolbox allowed to create new users and categorize them in technical, business or provider. Now in the Beta version of the Toolbox we have extended the use of that categories throughout all the platform, restricting the sections that each user can make use of.

Moreover, the functionality to delete users from the web Toolbox have been added as is showcased in Figure 1 (users masked for privacy sake).

Users list:

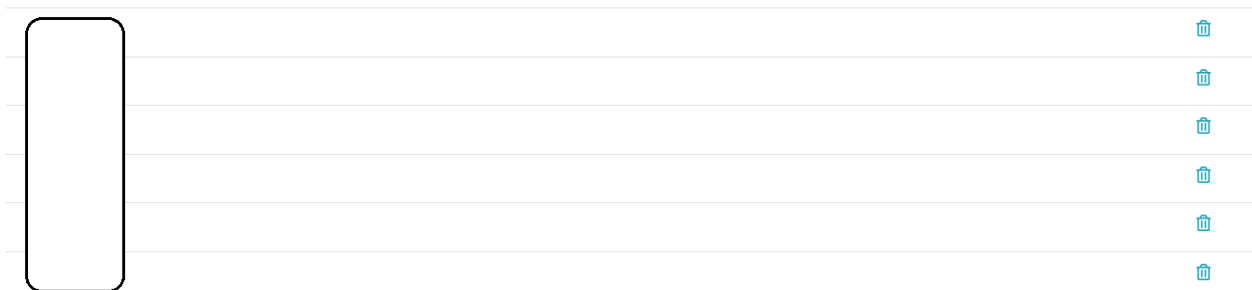


Figure 1. User deletion page

4.2 Benchmark management

Another key pillar of the Toolbox is the functionality to manage the benchmarks information as well as the possibility to integrate them and run them from the Toolbox. This is another key functionality for demoing the Beta version of the Toolbox and there have been some evolutions since the Alpha Version.

Specifically, regarding the actual benchmark information management we have worked to provide full CRUD functionalities for managing the benchmarks.

The UI for the list of the benchmarks and the visualization of which ones are integrated to be run from the Toolbox has been improved as can be seen in Figure 2. It now offers a cleaner list of benchmarks and shows with an icon which ones are integrated.

The same visualization is used to present the user with the list of benchmarks to update.

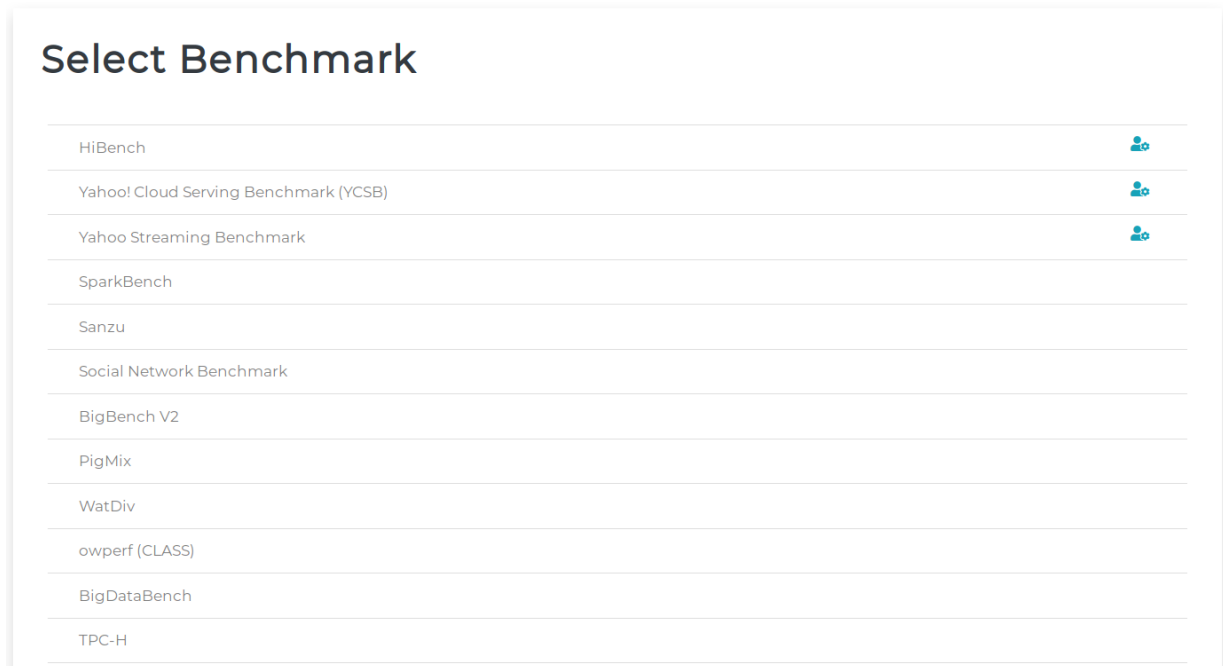


Figure 2. Benchmark selection page

Moreover, the UI for deleting any of the stored benchmarks is also available and presents the list with an icon to the right most column to delete any listed benchmark as can be seen in Figure 3. From the same page the user can access the functionality to update any information regarding any benchmark, this page will not be shown here since it looks the same as the one to create a new benchmark showed in Deliverable 3.2 but already populated with the stored data.

Select Benchmark
















HiBench	 
Yahoo! Cloud Serving Benchmark (YCSB)	 
Yahoo Streaming Benchmark	 
SparkBench	
Sanzu	
Social Network Benchmark	
BigBench V2	
PigMix	
WatDiv	
owperf (CLASS)	
BigDataBench	
TPC-H	

Figure 3. Benchmark update and delete page

The work done in the benchmarks section is not only limited to managing the benchmarks information. As commented in the previous sections, an extra effort has been made to try to link the benchmarks to the business metrics in the Toolbox. One of the most visible ways is that now the Toolbox, when trying to run a benchmark from the Toolbox, offers the possibility to the user to specify the use case that they are running the benchmark for. In the GUI this is shown as a combo box in the configuration page of the benchmark with the list of use cases provided by WP1 in Deliverable D1.1[4]. This update can be seen in Figure 4

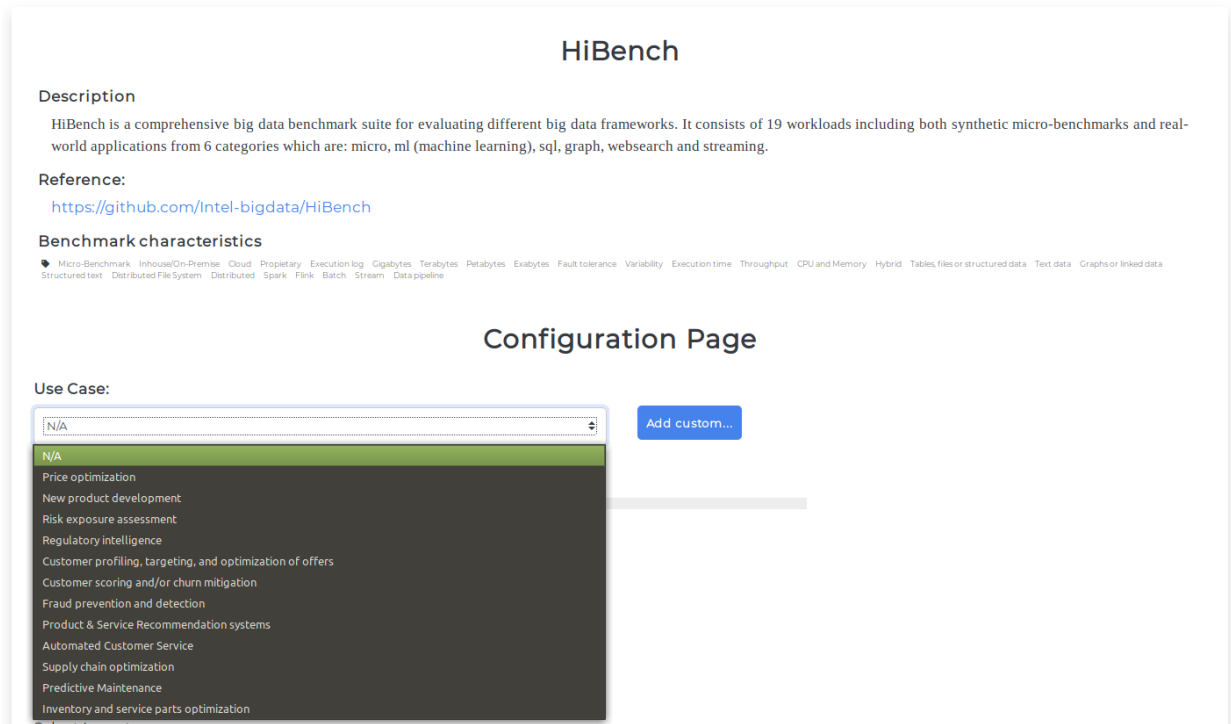


Figure 4. Benchmark use case selector

Since we have now the possibility to specify the use case for the benchmarks, a new filter based in this parameter has been added to the search functionality. A new section was created in the guided search to expose that functionality as can be seen in Figure 5.

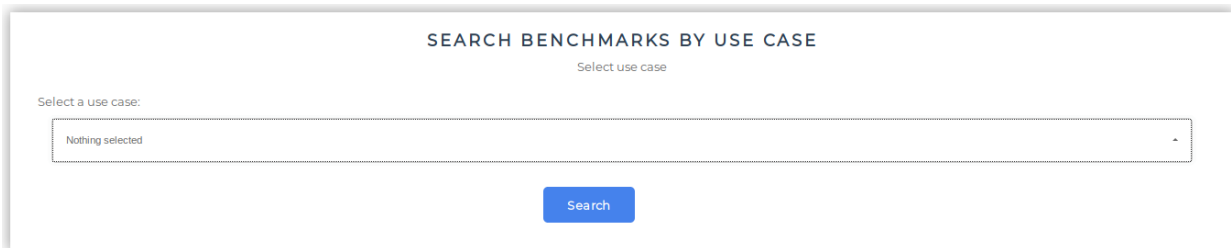


Figure 5. Search by use case

4.3 Results

The Beta version of the Toolbox focused on the end-to-end scenario so, on top of the Alpha version, an update was made not only to the previously stated sections but also a large overhaul was made to the results section.

As described in Section 3 some large improvements have been performed in the results part of the backend of the Toolbox that reflects in what information the frontend is able to show.

One of these examples is the result list for the user, its look and feel has been modified to look like the list of benchmarks (Figure 6), moving the functionality of seeing the detailed

results to a new page that can be opened by clicking any of the results from the list. This improves the user experience by showing a consolidated UI for every aspect of the Toolbox.

Results
Yahoo! Cloud Serving Benchmark (YCSB) on 2019-11-14T10:47:23.828
HiBench on 2019-11-04T12:09:31.171
Yahoo! Cloud Serving Benchmark (YCSB) on 2019-09-03T10:51:18.371
Yahoo! Cloud Serving Benchmark (YCSB) on 2019-09-03T09:12:56.765
Yahoo! Cloud Serving Benchmark (YCSB) on 2019-09-02T14:30:18.963
Yahoo! Cloud Serving Benchmark (YCSB) on 2019-09-02T12:42:41.726
Yahoo! Cloud Serving Benchmark (YCSB) on 2019-09-02T12:42:41.329
Yahoo! Cloud Serving Benchmark (YCSB) on 2019-09-02T10:43:09.749
Yahoo Streaming Benchmark on 2019-09-02T10:40:59.724
Yahoo Streaming Benchmark on 2019-09-02T10:40:59.400

Figure 6. Results list for a user

On top of that new visualization, since the Toolbox now is able to store the use case for each run, we can now show it in the detailed page for the results as can be seen in the example in Figure 7.

93 : Yahoo! Cloud Serving Benchmark (YCSB)

Run timestamp: 2019-09-03T10:51:18.371

Use case: Risk exposure assessment

```

2019-09-03 10:50:53.327 [Thread-3] ERROR c.y.ycsb.db.arangodb.ArangoDBClient - Failed to create database: ycsb with ex: com.arangodb.ArangoDBException: Response: 409, Error: 1207 - duplicate name
2019-09-03 10:50:53.339 [Thread-3] ERROR c.y.ycsb.db.arangodb.ArangoDBClient - Failed to create collection: usertable with ex: com.arangodb.ArangoDBException: Response: 409, Error: 1207 - duplicate name
2019-09-03 10:50:53.339 [Thread-3] INFO c.y.ycsb.db.arangodb.ArangoDBClient - ArangoDB client connection created to localhost:8529
2019-09-03 10:50:53.340 [Thread-3] INFO c.y.ycsb.db.arangodb.ArangoDBClient - Arango Configuration: dropDBBeforeRun: false, address: localhost:8529, databaseName: ycsb, waitForSync: false, transactionUpdate: false;
2019-09-03 10:50:54.256 [Thread-3] INFO c.y.ycsb.db.arangodb.ArangoDBClient - Local cleaned up.
[OVERALL], Runtime(ms), 1087
[OVERALL], Throughput(ops/sec), 919.9632014719411
[TOTAL_GC_TIME_PS_Scavenge], Count, 1
[TOTAL_GC_TIME_PS_Scavenge], Time(ms), 11
[TOTAL_GC_TIME_%_PS_Scavenge], Time(%), 1.011959521619135
[TOTAL_GC_TIME_PS_MarkSweep], Count, 0
[TOTAL_GC_TIME_PS_MarkSweep], Time(ms), 0
[TOTAL_GC_TIME_%_PS_MarkSweep], Time(%), 0.0
[TOTAL_GC_S], Count, 1
[TOTAL_GC_TIME], Time(ms), 11
[TOTAL_GC_TIME_%], Time(%), 1.011959521619135
[READ], Operations, 479
    
```

Figure 7. Detailed result with use case

In addition to these aesthetic modifications, since the Toolbox has a module to extract metrics from the results that was explained in Section 3.4, it has now the capability to show the user some graphics about some of the calculated metrics for their benchmark runs.

This will be further extended in future versions of the Toolbox but, as of now, it already allows showing some informational charts to the user. As can be seen in the example screenshot in Figure 8, the Toolbox now has the possibility of showing some of the numerical benchmarks calculated in a chart allowing comparison of values between different runs of the benchmark as well as showing if the metrics are related between each other.

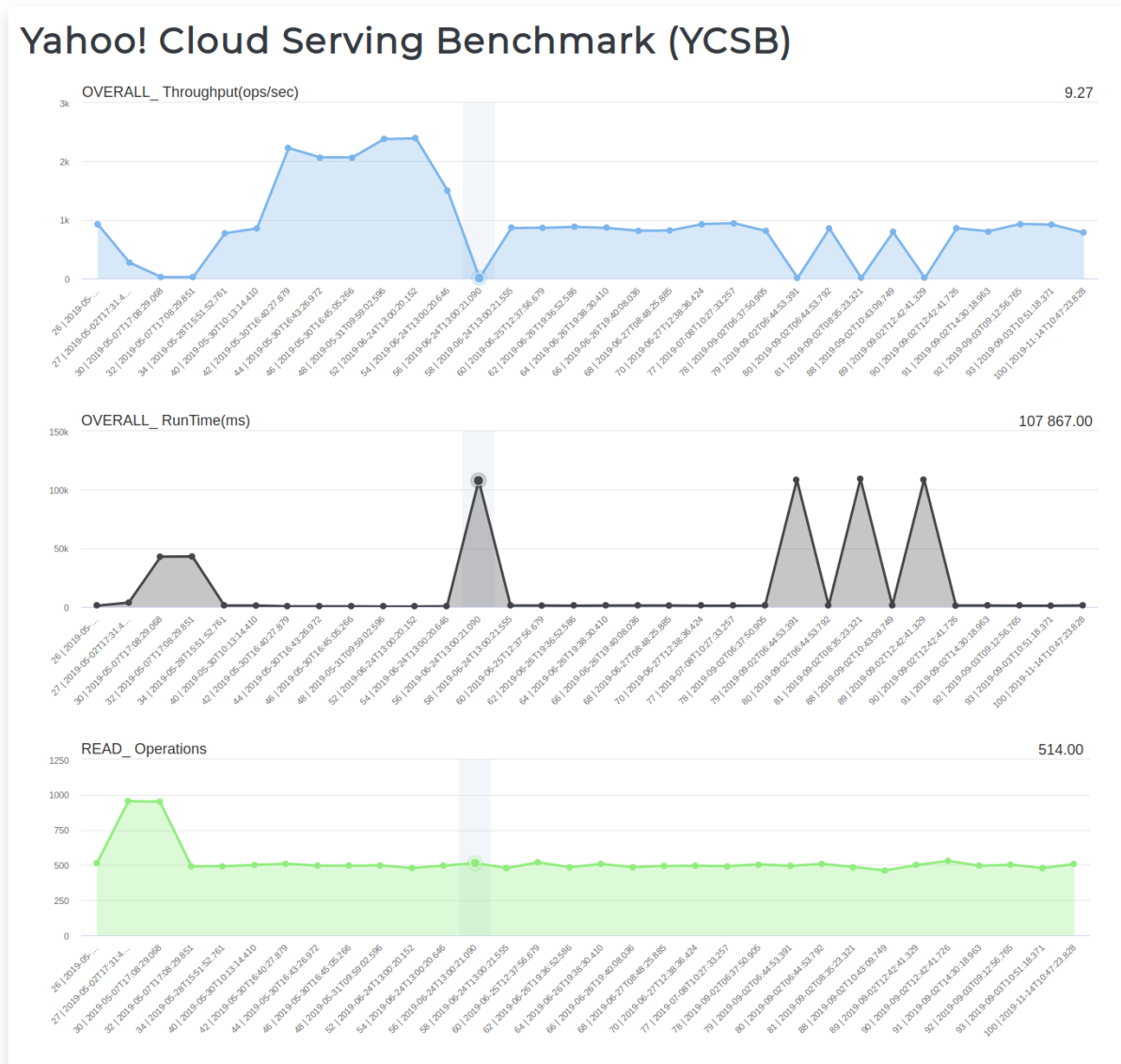


Figure 8. Result charts

In this section we have presented the updates of the DataBench Toolbox dashboard front-end as released for the Beta version by December 2019. The usage of this version is intended only for a selected set of users, mainly early adopters from sister EU-funded projects, but also for other organizations that would like to test its functionality and use some of the integrated benchmarks. The URL of the dashboard is public and therefore accessible to anybody to browse and search for the existing benchmarks. However, only registered users will be able to access to the whole set of functionalities explained in this section and in deliverable D3.2. We encourage interested users to get in touch via the DataBench contact web page (<https://www.databench.eu/contact/>) in case you would like to use and test the tool at its fullest.

The front-end of the Beta version dashboard can be accessed through the following URL:

<http://databench.ijs.si:9000/>

Note that this is a temporary URL intended for testing purposes of the Beta version, to be replaced at the time of making available the final release.

5. Conclusions and future work

This document provides a summary and an update of the DataBench Toolbox as it has been delivered for its Beta version. The Beta version is an incremental version based on the same structure and architecture presented for the Alpha version in deliverable D3.2.

Therefore, the document presents the new functions covered in the Beta version highlighting the main changes with respect to the previous release. In this sense, to understand this Beta version it is advisable to check the content of D3.2 as well.

This document will be followed by another official release: D3.4 (release version) in M30 (June 2020).

In the final version (future deliverable D3.4) we expect to make operational the work performed in other work packages. In particular, the plan is the following:

- Integrate and make searchable the catalogue of all benchmarks listed in WP1 deliverables (D1.2, D1.3 and D1.4), as well as using the categories presented in the deliverables to facilitate search, filtering and recommendations of technical benchmarks.
- Showcase a tighter integration of the most promising benchmarks in terms of uptake by the community in the Toolbox. These benchmarks will not be only searchable, but also deployable, executable and their results integrated in the platform. The Beta version already showcases three benchmarks integrated.
- Use the business indicators and benchmarks developed in WP2 (mostly presented in deliverable D2.4) drawing on European industry requirements to be able to give business insights.
- Similarly, WP4 will provide business evaluation indicators mainly based on the notion of use cases interviews and desk research. This bottom-up approach will serve as basis for recommendations of big data technologies and similar use cases to be made available via the Toolbox.
- Last, but not least WP5 will provide access to encoded knowledge and ontologies that will provide further insights and relations between technical and business aspects, making the results of previous work packages more easily available to the Toolbox. The plan is to enhance the searching and delivering of recommendations to the users by leveraging the power of the semantics developed in WP5.

The URL to access the Beta version of the DataBench Toolbox is public, and anyone could use some specific functionality related to search and browse metadata about the benchmarks integrated so far. However, only a few registered users, mainly early adopters from sister EU-funded projects or selected organizations, will be able to access to the advanced functionalities explained in the deliverable D3.2 and further extended in this document.

The Final version of the DataBench Toolbox is expected by mid-2020, although improvements will be carried out until the end of the project based on the feedback from users.

6. References

- [1]. D3.1. DataBench Deliverable D3.1. DataBench Architecture.
- [2]. D3.2. DataBench Toolbox - Alpha including support for reusing of existing benchmarks
- [3]. D5.3 Assessment of Technical Usability, Relevance, Scale and Complexity
- [4]. D1.1. DataBench Deliverable D1.1. Industry Requirement with benchmark metrics and KPIs.
- [5]. D1.2. DataBench Deliverable D1.2. DataBench Framework – with Vertical Big Data Type benchmarks
- [6]. D1.3 DataBench Deliverable D1.3. Horizontal Benchmarks – Analytics and Processing
- [7]. D1.4 DataBench Deliverable D1.4. Horizontal Benchmarks – Data Management
- [8]. D2.4 DataBench Deliverable D2.4. Benchmarks of European and Industrial Significance
- [9]. D5.1 DataBench Deliverable D5.1. Initial Evaluation of DataBench Metrics
- [10]. Ansible - <https://www.ansible.com/>
- [11]. Apache Kafka - <https://kafka.apache.org/>
- [12]. CLASS EU project - <https://class-project.eu/>
- [13]. YCSB, <https://github.com/brianfrankcooper/YCSB>, 2018
- [14]. HiBench Suite - <https://github.com/intel-hadoop/HiBench>, 2018