Project Caliper Proposal

# **HIP Identifier**

# Hyperledger Caliper

# **Sponsors**

Haojun ZHOU ([zhouhaojun@huawei.com](mailto:zhouhaojun@huawei.com))

Victor HU ([huruifeng@huawei.com](mailto:huruifeng@huawei.com))

# **Introduction**

# Caliper is a blockchain benchmark framework which allows users to measure the performance of a specific blockchain implementation with a set of predefined use cases. Caliper will produce reports containing a number of performance indicators, such as TPS (Transactions Per Second), transaction latency, resource utilisation etc. The intent is for Caliper results to be used as a reference in supporting the choice of a blockchain implementation suitable for the users specific use-cases. Given the variety of blockchain configurations, network setup, as well as the specific use-cases in mind, it is not intended to be an authoritative performance assessment, nor to be used for simple comparative purposes (e.g. blockchain A does 5 TPS and blockchain B does 10 TPS, therefore B is better).The Caliper project references the definitions, [metrics and terminology](https://docs.google.com/document/d/1DQ6PqoeIH0pCNJSEYiw7JVbExDvWh_ZRVhWkuioG4k0/edit?usp=sharing) as defined by the [Performance & Scalability Working Group](https://wiki.hyperledger.org/groups/pswg/performance-and-scale-wg) (PSWG).

# **Context**

Currently, there are some reports on the performance of different blockchain framework/platform in various scenarios, however there is not a commonly accepted blockchain benchmark framework. Some existing testing tools are listed below:

* PTE in fabric (https://github.com/hyperledger/fabric/tree/release/test/tools/PTE), which only supports hyperledger fabric
* Blockbench (https://github.com/ooibc88/blockbench), developed by Natil University of Singapore and Zhejiang University, which provides a benchmark framework for multiple blockchain platform including Fabric0.6, Ethereum and Parity. A good framework for reference, but it does not support Fabric1.0 and other hyperledger blockchain platforms like Sawtooth. Neither does it provide pluggable capability for supporting multiple test cases.

# **Dependent Projects**

Caliper is a benchmark framework for blockchain framework/platform and relies on a functioning blockchain framework/platform as the benchmarking target. However, tools that can quickly bring up a blockchain network are ideal to cooperate with Caliper.

# **Motivation**

The performance of a blockchain solution is one of the features of most concern for blockchain users. However, currently there is no general framework that serves to provide performance evaluations for different blockchain solutions based on a set of neutral & commonly accepted rules. Clearly, every blockchain framework/platform proposed nowadays has it’s strengths and abilities. In the light of such variety, we consider the absence of a consistentbenchmark framework an obstaclel in promoting blockchain across multiple industry verticals. Here are 3 reasons why we think it is necessary to have a general purpose benchmark platform:

1. There are some performance reports for several projects but if they do not provide the source code, it is hard to re-perform the evaluation and impossible to replicate evaluation on different projects.
2. There is no common definition of performance indicators (e.g. TPS, latency, resource utilization, etc.) and we think PSWG is the ideal place to define these. Caliper would be a good starting point and place to hold the paper work. A good example of having a well-defined performance indicator is TPS (Transaction per Second), as in any blockchain solution that supports smart contracts, it is a common agreement that the complexity of a smart contract is a key factor of TPS. Benchmarks that use different trial smart contracts, no matter how simple these smart contracts could be, is biasing the results.
3. There is no common accepted benchmarking use cases for benchmarking. Users are always curious about use cases, while proper use cases, they help the users to understand the blockchain itself as well as the performance indicators. The use cases for benchmarking are also open to discussion and welcome contribution.

The Caliper project is dedicated to provide a general purpose blockchain benchmark framework, starting from the following 3 aspects:

1. A unified blockchain benchmark framework. We provide a common layer to integrate with major existing blockchain framework/platforms, so that the same benchmarks can be run for different blockchain systems Some benchmark test environment will be provided to help different people run tests under the same environment, blockchain management tools like Hyperledger Cello could be integrate later to deploy and operate the environment. Also users can use their existing environment and configure Caliper to run the test under the environment.
2. A commonly accepted definition of performance indicators. You cannot compare an apple and a pear directly unless some common criteria is set. We will work closely to PSWG to provide common definition of performance indicators that users care about, such as TPS, latency, resource utilization, etc.
3. A set of commonly accepted benchmark cases. The goal of Caliper includes providing a set of easy-understandable benchmark cases so that each blockchain solution can be compared in various scenarios. This calls for much collaboration from PSWG, Requirement WG and other WG in Hyperledger community as well as blockchain practitioners to cover as many use cases that are of user’s interest as possible.

Caliper is not intended to make judgments and will not publish benchmark results, but provide benchmark tools for users. Users should not claim the result is tested by Caliper until the test environment is revealed. Such specification should be defined later to guide people to publish the result.

# **Status**

Caliper started in May, 2017 and is being actively designed and developed within Huawei. The code is available here - <https://github.com/Huawei-OSG/caliper>. We look forward to more contributors from the industry for system design, benchmark case design, as well as code contribution.

# Progress Report March 2018

1. Ongoing work with PSWG to define the metrics document, some consensus has already been reached and we hope that the document will be finalized in near future.
2. We discussed Caliper’s scope with PSWG and agree that Caliper should be a pure test tool to provide metrics of various performance aspects instead of providing any ‘score’ or direct comparison result for DLT systems.
3. We modified our metrics and measurement methods to keep consistent with the current output of PSWG, and will continue to work on this according to the progress of PSWG
4. Three hyperleger project are supported now, including:
   * Fabric v1.0.5
   * Sawtooth Lake v0.8, and we are working on v1.0 now
   * Iroha Jan version (develop version)
5. Currently supported performance indicators:
   * Success rate
   * Throughput (TPS)
   * Transaction confirmation latency
   * Resource consumption (CPU, Memory, Network IO,...)
6. Some new features:
   * Support launching/removing local dockerized SUT automat in ically before/after test or interacting with an existing SUT by configuration
   * Support running a test with multiple local or remote clients, caliper clients can register themselves and receive test jobs from caliper server
   * Support monitoring and outputting resource consumptions of local or remote dockerized SUT
   * An HTML format test report will now be generated (by Mustache) automatically after the test
   * A simple GUI has been provided to launch the test and monitor the test progress
   * Other updates to improve test flow, workload definition and so on

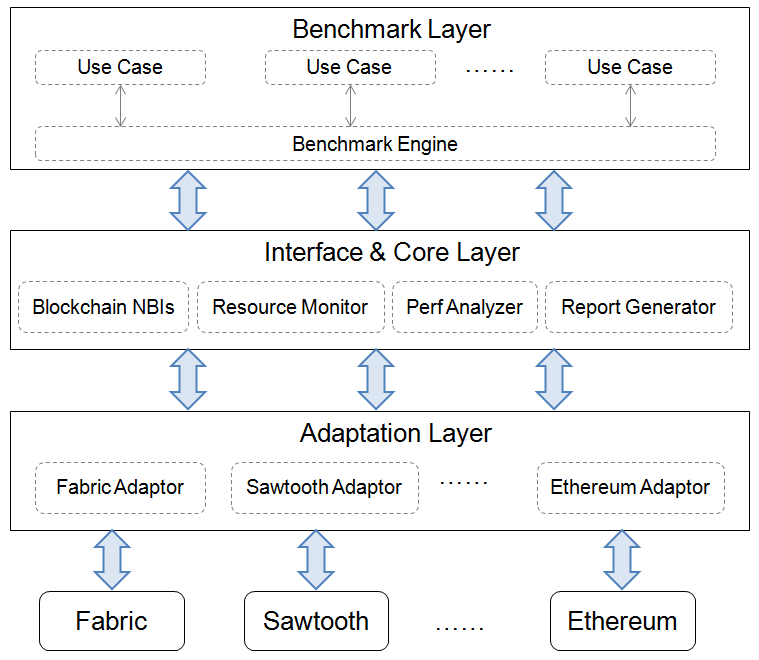
# Roadmap

* June 2018
  + Update with the latest version of fabric/sawtooth lake/iroha
  + Update to align with PSWG definitions
  + Investigate to integrate more hyperledger DLTs
* Dec 2018
  + Stable version
  + Integrate more hyperledger DLTs
  + Support large scale test
  + More test cases for typical scenarios (with PSWG coorperation)
  + New GUI
  + Investigate to integrate DLT management tools like Cello
  + Investigate to integrate non-hyperlerledger DLTs

# **Solution**

## **Framework Introduction**

The key component in Caliper is the Adaptation Layer which is introduced to integrate multiple blockchain solutions into Caliper framework. An adaptor is implemented for each blockchain system under test (SUT), the adaptor is responsible for translation of Caliper NBIs into corresponding blockchain protocol. Caliper NBI is a set of common blockchain interfaces which contains operations to interact with backend blockchain system, for example, to install smart contract, invoke contract, query state from the ledger, etc. The NBIs can be used for upstream applications to write tests for multiple blockchain systems. For more information, please see the documentation of [Caliper](https://github.com/Huawei-OSG/caliper).



For now, Fabric, Sawtooth Lake and Iroha are in scope and we sincerely welcome contributions for integrations to other blockchain solutions.

# **Effort and resources**

Cooperation with PSWG (Performance & Scalability Work Group) in Hyperledger community is extremely important to the success of Caliper and we propose to have regular discussion of Caliper on PSWG meeting.

Caliper is not going to limit the benchmark environment, but will leave it to user’s choice depending on their own use cases. Collaborating with Requirement WG and using Cello to provide user defined environment will also require close discussion among the WGs and projects.

Huawei, Hyperchain, Oracle, Bitwise, Soramitsu, IBM and the Budapest University of Technology and

Economics are all contributing to the project. Others are welcome.

Haojun ZHOU (zhouhaojun@huawei.com)

Baohua YANG (yangbaohua@gmail.com)

Iushkevich Nikolai (nikolai@soramitsu.co.jp)

Victor Drobny (drobny@soramitsu.co.jp)

Vyacheslav Bikbaev (viacheslav@soramitsu.co.jp)

Lizhong Kuang (lizhong.kuang@hyperchain.cn)

Jame Mitchell (mitchell@bitwise.io)

Dongming Hwang (dming@us.ibm.com)

Victor HU (huruifeng@huawei.com)

Imre Kocsis (ikocsis@mit.bme.hu)

# **How to**

Documentation of Caliper will be provided along with the code publication and will be updated regularly.

# **Closure**

The first goal of the project is to provide the community a functioning benchmark framework capable of running against both Fabric and Sawtooth Lake. Meanwhile we will be putting continuous effort with the community into new definition of performance indicators and new benchmark use cases. The success of the project can be that it attracts many users within or out of the community to use it as the benchmark framework. If other project shows up and the community is more interested in putting more resource on it over Caliper, a merge should occur and development will be moved to the new project