**What is the response time with and without caching? How much does caching help?**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Without cache** | **Cache miss** | **Cache hit** |
| **Info** | 0.026 s | 0.021 s | 0.0 |
| **Search** | 0.025 s | 0.023 s | 0.0 |

We achieve the low latency performance by adding an **In-Memory** **cache**, as we see from the above result, caching speed up the operation, because it reduces the need to access server so, the response time we need to return data to the client is always 0.

**What is the overhead of cache consistency operations?**

To implement consistency this required more response time as shown in the table, because purchase operation with cache consistency required visiting more servers, from order to catalog and then to frontend server to implement the consistency and delete this item from cache.

**Purchase - consistency**

|  |  |
| --- | --- |
| **Without consistency** | **With Consistency** |
| 0.09611940383911133 | 1.5585124492645264 |

**What is the latency of a subsequent request that sees a cache miss?**

In cache miss, the response time is very close to using server without cache as shown above in the plot, because we need to visit server to get data from it and store it to cache.