A Case Study on Benchmarking IoT Cloud Services

Kevin Grünberg and Wolfram Schenck(✉)

Faculty of Engineering and Mathematics, Center for Applied Data Science Gütersloh,
Bielefeld University of Applied Sciences, Bielefeld, Germany
kevin_gruenberg@web.de, wolfram.schenck@fh-bielefeld.de

Abstract. The Internet of Things (IoT) is on the rise, forming networks of sensors, machines, cars, household appliances, and other physical items. In the industrial area, machines in assembly lines are connected to the internet for quick date exchange and coordinated operation. Cloud services are an obvious choice for data integration and processing in the (industrial) IoT. However, manufacturing machines have to exchange data in close to real-time in many use cases, requiring small round-trip times (RTT) in the communication between device and cloud. In this study, two of the major IoT cloud services, Microsoft Azure IoT Hub and Amazon Web Services IoT, are benchmarked in the area of North Rhine-Westphalia (Germany) regarding RTT, varying factors like time of day, day of week, location, inter-message interval, and additional data processing in the cloud. The results show significant performance differences between the cloud services and a considerable impact of some of the aforementioned factors. In conclusion, as soon as (soft) real-time conditions come into play, it is highly advisable to carry out benchmarking in advance to identify an IoT cloud workflow which meets these conditions.

Keywords: Cloud benchmarking · Internet of Things · IoT
Industrial IoT · Cyberphysical systems · Round-trip time
Microsoft Azure IoT Hub · Amazon Web Services IoT

1 Introduction

During the last decade, cloud computing has become a highly popular approach to provide scalable and flexible hardware and software infrastructure at various service levels (infrastructure as a service [IasS], platform as a service [PaaS], software as a service [SaaS]) [1]. With the Internet of Things (IoT) being on the rise, a new use case for cloud computing emerges. In the IoT, physical objects like sensors, machines, cars, household appliances, and other items are connected via the internet to enable interaction and cooperation of these objects. Application areas of the IoT are, among others, transportation, healthcare, smart homes and industrial environments [2]. IoT devices typically generate and transmit data at regular intervals. This data is integrated, processed, and monitored (e.g., via

M. Luo and L.-J. Zhang (Eds.): CLOUD 2018, LNCS 10967, pp. 398–406, 2018.
https://doi.org/10.1007/978-3-319-94295-7_28

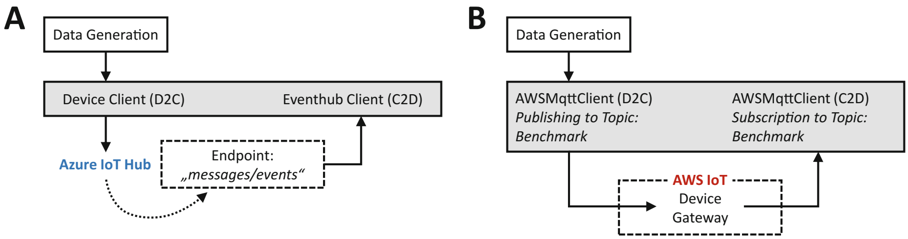

Fig. 1. Message flow between benchmarking applications (gray box incl. data generation) and cloud services. A: For MS Azure IoT Hub. B: For AWS IoT.

mobile devices or web applications), and commands for corrective actions (which are either manually or automatically generated) are sent back to the devices [3]. Furthermore, IoT devices may interact directly with each other.

It is an obvious approach to use the cloud for data integration and processing [4]. The major cloud service providers (CSP) have established meanwhile specialized services for data collection from IoT devices, supporting popular protocols like MQTT or AMQP. However, many application scenarios in the IoT—especially in the industrial area—rely on operations taking place in close to real-time: For example, when manufacturing machines are controlled via apps running in the cloud, or when these machines use the cloud as central instance for data exchange to coordinate their work in an assembly line. Therefore it is important that cloud platforms used for these purposes process and forward messages with small latency and consistent availability and performance.

Although many studies on cloud benchmarking in general exist, we are not aware of any study which addresses especially the performance of IoT cloud services. As a first step in this direction, we present benchmarking data collected from Microsoft Azure IoT Hub (MAIH) and Amazon Web Services IoT (AWSI) at three different locations in North Rhine-Westphalia (NRW; Germany) around Sept./Oct. 2017. This data collection was part of a larger research project for a local mechanical engineering company to integrate their products in the industrial IoT. Despite this local scope, we are convinced that some general conclusions can be drawn from our data.

2 Previous Work

Studies on cloud benchmarking address mostly the PaaS and IaaS levels. Concerning PaaS, Binnig et al. [5] point out that classic benchmarking approaches are not well suited for cloud APIs because resources are usually scaled up automatically with increasing load. They suggest to concentrate on scalability, cost, coping with peak load, and fault tolerance. Agarwal and Prasad [6] carried out a comprehensive benchmarking study on the storage services of the Microsoft Azure cloud platform, especially for high-performance computing. In their view, benchmarking is definitely required to understand the performance capabilities

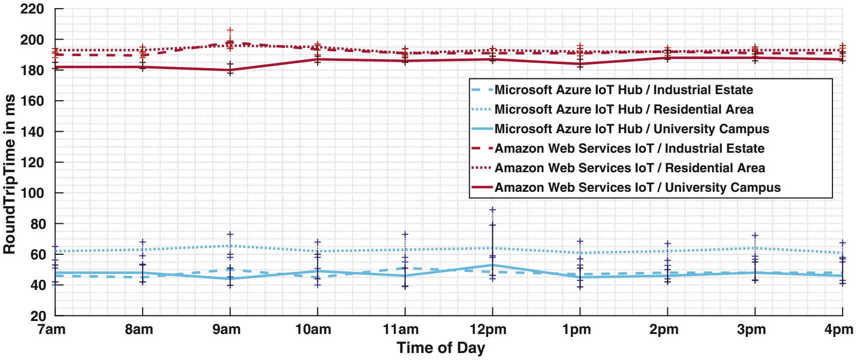

Fig. 2. RTT measurements, varying time of day and location. Each data point depicts the median value and interquartile range of around 100 measurements.

of the underlying cloud platforms because the cloud service offerings are usually only best-effort based without any performance guarantees. In a similar spirit, Kossmann et al. [7] compared the performance of major CSPs in transaction processing. Their results show that the services vary dramatically when it comes to end-to-end performance, scalability, and cost—although the price matrices of the CSPs are very similar in terms of network bandwidth, storage cost, and CPU cost (see also [8] for a holistic evaluation approach). The study by Bermbach and Wittern [9] is related to our work because the authors use response latency as performance measure for web API quality. They observed 15 different web APIs over a longer time period from different locations in the world. They results show that web API quality is very volatile over time and location, and that developers have to take this into account from an architectural and engineering point of view.

High variability is also observed at the IaaS level. Gillam et al. [10] and Zant and Gagnaire [11] carried out a large set of benchmarks on virtual machines (VM) of different specifications and from different CSPs. Even if the virtual resources are of the same specification, the observed variability is substantial (e.g., regarding CPU and memory performance). To help customers to choose the best cloud platform for their needs, benchmarking results could be made available in publicly available databases (as in [10]), or tools for easier benchmarking could be developed. Following the second approach, Cunha et al. [12] implemented "Cloud Crawler" for describing and automatically executing application performance tests in IaaS clouds for different virtual machine configurations.

3 Methods

Performance Measure. The quality of data transfer to and from cloud services can vary along two main dimensions: throughput (number and/or size of messages per time unit) and latency. Using throughput as performance measure,

Table 1. Mean values and standard deviations (in brackets) of the RTT median values (varying over time of day) shown in Fig. 2 to illustrate the systematic differences between locations. **Values given in ms.**

	Location		
	Industrial estate	Residential area	University campus
MS Azure IoT Hub	47.7 (2.0)	62.8 (1.4)	47.2 (2.7)
AWS IoT	191.8 (2.5)	199.1 (1.4)	185.4 (2.8)

however, is problematic because many cloud services automatically scale up the available communication resources with increasing load [13]. For this reason, and because we aim on real-time performance in the context of our study, which requires message transfers with low latency, we use the round-trip time (RTT) in our benchmarking. RTT is defined here as the time difference between sending a message to a communication endpoint (time of sending) and receiving the same message back from this endpoint (time of arrival).

Microsoft Azure IoT Hub (MAIH). The benchmarking application for MAIH was implemented using the .NET SDK provided by Microsoft [14]. On the client side, a so-called “DeviceClient” has to be created and initialized for sending messages to the IoT Hub (see Fig. 1A). Within the Azure cloud, messages are forwarded to an “EventHub” endpoint. Within the client, a subscription to this endpoint is created so that messages are automatically received back. (MAIH server location: Western Europe)

Amazon Web Services IoT (AWSI). The benchmarking application for AWSI has a similar structure but is based on the JAVA SDK provided by Amazon [15]. An important step is to register the device at the AWSI device gateway (see Fig. 1B) as a “thing” with a unique identifier (“topic”) under which it is published. The same topic is used for the subscription to receive the messages back from the device gateway. (AWSI server loc.: Frankfurt/Main, Germany)

Messages. The content of the sent messages was a single boolean value[1] which was embedded into a JSON container (size: 250 Bytes). This container also contained the time of sending as timestamp. In most benchmarking tests, messages were sent every two seconds. To measure a single data point, about 100 messages were sent; the median and the interquartile range of this sample are shown in most of the figures in the results section. The benchmarking applications were run on a notebook with a quadcore CPU (Intel i7-4712MQ). Messages retrieved under a CPU load of more than 20% were discarded to eliminate CPU load as a cause of variability [13]. The percentage of discarded messages was in the low single-digit range. Messages were transferred via MQTT.

[1] Pretests did not show any significant RTT differences between different basic data types like `int`, `float`, etc.

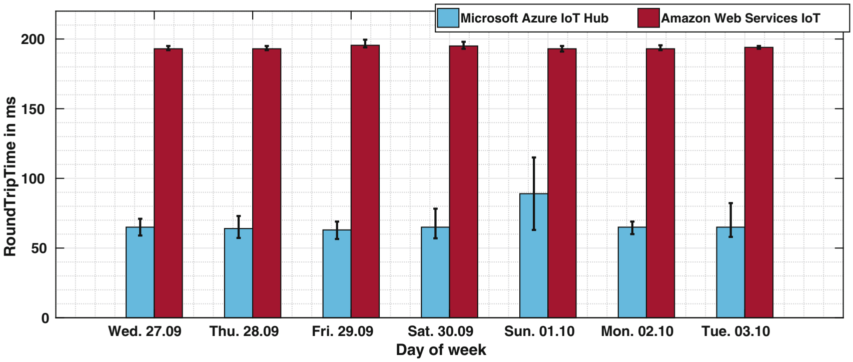

Fig. 3. RTT measurements, varying day of week. Each data point depicts the median value and interquartile range of around 100 measurements.

Locations. Benchmarking took place at three different locations in NRW (Germany): (1) At a mechanical engineering company within an industrial park in Rheda-Wiedenbrück (denoted as "industrial estate" in the following; connection to internet via optic fiber); (2) at a residential building in Herzebrock-Clarholz ("residential area"; connection to internet via DSL over regular copper phone lines); (3) at the main campus of Bielefeld University of Applied Sciences in Bielefeld ("university campus"; connection to internet via the German National Research and Education Network [DFN] [optic fiber]).

4 Results

Time of Day and Location. In the first set of benchmarks, time of day and location were varied systematically. Figure 2 shows considerable RTT differences between MAIH (mean median value: 52.5 ms) and AWSI (mean median value: 192.1 ms) while performance stays nearly constant during the day. Interquartile ranges vary during the day, and sometimes strong outliers with large RTT values were observed (not shown). Location has systematic impact too, with "university campus" exhibiting the smallest RTT values for both CSPs, followed by "industrial estate" and "residential area" (in this order; see Table 1). Measurements were carried out around midweek.

Day of Week. RTT measurements were repeated between Sept. 27th and Oct. 3rd at the residential area to check the impact of the day of the week. The results in Fig. 3 show small variability between the days of the week regarding median values with one notable exception: MAIH on Sunday with an RTT increase of about 30%. Because this effect does not appear in the AWSI measurements carried out at nearly the same time, it can be most likely attributed to the Azure platform or to the connecting network route.

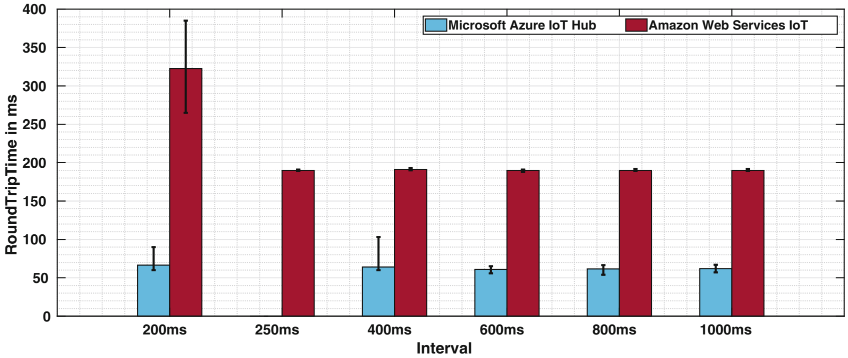

Fig. 4. RTT measurements, varying the time interval between sending messages. Each data point depicts the median value and interquartile range of around 100 measurements. The data point at 250 ms was only measured for AWSI.

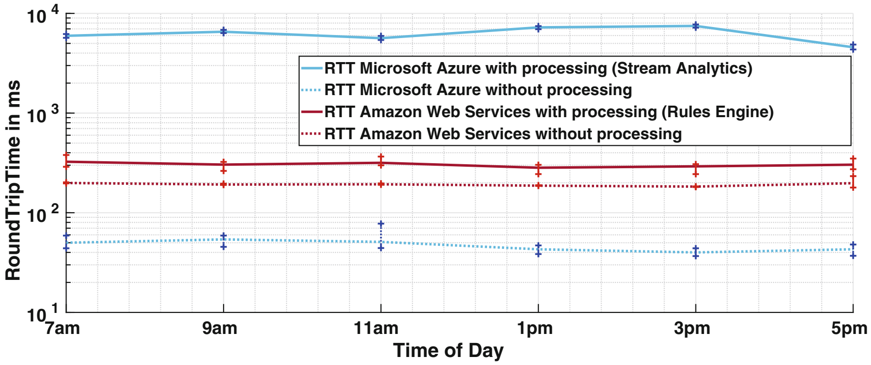

Fig. 5. RTT measurements, with/without processing in the cloud; in addition the time of day was varied. Each data point depicts the median value and interquartile range of around 100 measurements. Note the logarithmic scaling of the y-axis.

Message Interval. In this benchmark, the interval between sending messages was varied between 200 ms and 1000 ms. Measurement location was the residential area. On the whole, performance does not vary significantly depending on the message interval (see Fig. 4). However, RTT values are much larger (and have higher variability) for AWSI when an interval of 200 ms is used; they amount to ca. 320 ms in this case. Since AWSI RTT values amount otherwise to around 190 ms, this shows most likely a detrimental effect of sending and receiving messages at around the same time. We think that the causes of this effect reside mostly on the cloud server side, because the client performs sending and receiving of messages in a concurrent way on a quadcore CPU which is under low load. In such a setting, it is very unlikely that the rather large time difference of 130 ms (320 ms–190 ms) is caused by a resource conflict on the client side alone.

Additional Processing in the Cloud. For the (industrial) IoT, it is a common scenario that device data is processed in the cloud before it is sent back to another endpoint (e.g., data preprocessing in the cloud before data visualization on a mobile device). As a benchmark for this use case, we sent in each message a single float value to the cloud where it was multiplied by two and sent back. For this purpose, the workflows in the cloud had to be modified: Within Microsoft Azure, messages are forwarded from the IoT Hub to the service "Stream Analytics" where processing takes place. Afterwards, the "Service Bus" sends the modified messages back to a so-called "Queue Client" implemented as part of the benchmarking application. In the AWS cloud, received messages are forwarded from the Device Gateway to the "Rules Engine" for processing. The processed messages are made available under a new topic to which the MQTT client in the benchmarking application is subscribed.

Benchmarks were carried out at the industrial estate. The results are shown in Fig. 5. Performance does not vary significantly during the course of the day, but the additional processing in the cloud has a clear impact on the RTT values. For AWSI, the mean value of the median values increases from 192.8 ms (without processing) to 304.3 ms (with processing). For MAIH, an increase from 49.2 ms (without processing) to 6336 ms (with processing) was observed. The latter is a really significant slowdown by a factor of 129.

5 Discussion and Outlook

Results of cloud benchmarking usually have to be taken with a grain of salt because there exist many uncontrollable factors along the route between client and cloud servers [13]. Furthermore, CSPs may upgrade and modify their systems without notice. Therefore, we are well aware that our results only represent a local snap-shot, and that similar performance measurements may be different in other parts of the world and in the future. For this reason, it is definitely not our goal to attribute good or bad performance to MAIH or AWSI in general.

However, our results clearly show that one has to be careful in choosing the right cloud platform for IoT operations in one's local area, especially if we consider the use case of cyber-manufacturing systems (industrial IoT) which are intended to interact in real-time—not in the very strict sense of hard real-time conditions, but in the sense of seamless coordination between machines (and maybe machines and humans). Our data shows a clear performance advantage for one of the two benchmarked CSPs, and how this pattern completely reverses in an extreme fashion as soon as additional processing of the data in the cloud is required. Mean performance and variability is rather stable when varying the day of week, the time of day, or the location. However, the latter factor has a clearly visible and systematic impact on the performance of both CSPs. A considerable performance drop could be observed for AWSI when decreasing the sending interval between messages to the usual RTT. This is a clear warning that performance is subject to factors in an unexpected way (one message every 200 ms does not look like heavy load, does it?).

Although the number of clients in a manufacturing line, which are independently connected to an IoT cloud service, is usually not that large compared to e.g. the millions of smart home devices, it is definitely an important step for further benchmarking to increase the number of clients which send data in parallel to the cloud. Furthermore, we plan to repeat our measurements over a longer time period and, in addition, to vary message size. Message size is important because it is common that programmable logic controllers in manufacturing collect sets of sensory and control variables and transmit them at once, e.g. to a local OPC-UA client which in turn forwards the data in the form of larger packages to the cloud.

Given our already existing benchmarking results, we are convinced that manufacturing companies, when they move forward to cyberphysical production systems, have to carefully plan the data flow to the cloud and to carry out corresponding benchmarks because it may depend strongly on the CSP if specific real-time conditions for IoT data processing are met in the end.

Acknowledgements. We appreciate the general support by Dominik Osthues (Rheda-Wiedenbrück) for our benchmarking work. Furthermore, we thank Martin Kohlhase (Bielefeld Univ. of Applied Sciences) for many helpful discussions.

References

1. Vaquero, L.M., Rodero-Merino, L., Caceres, J., Lindner, M.: A break in the clouds: towards a cloud definition. SIGCOMM Comput. Commun. Rev. **39**(1), 50–55 (2008)
2. Jeschke, S., Brecher, C., Meisen, T., Özdemir, D., Eschert, T.: Industrial internet of things and cyber manufacturing systems. In: Jeschke, S., Brecher, C., Song, H., Rawat, D.B. (eds.) Industrial Internet of Things – Cybermanufacturing Systems. SSWT, pp. 3–19. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-42559-7_1
3. Lee, G.M., Crespi, N., Choi, J.K., Boussard, M.: Internet of things. In: Bertin, E., Crespi, N., Magedanz, T. (eds.) Evolution of Telecommunication Services: The Convergence of Telecom and Internet: Technologies and Ecosystems. LNCS, vol. 7768, pp. 257–282. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41569-2_13
4. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Future Gener. Comput. Syst. **29**(7), 1645–1660 (2013)
5. Binnig, C., Kossmann, D., Kraska, T., Loesing, S.: How is the weather tomorrow?: towards a benchmark for the cloud. In: Proceedings of the 2nd International Workshop on Testing Database Systems, DBTest 2009, pp. 9:1–9:6. ACM, New York (2009)
6. Agarwal, D., Prasad, S.K.: AzureBench: benchmarking the storage services of the Azure cloud platform. In: 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops PhD Forum, pp. 1048–1057 (2012)

7. Kossmann, D., Kraska, T., Loesing, S.: An evaluation of alternative architectures for transaction processing in the cloud. In: Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, SIGMOD 2010, pp. 579–590. ACM, New York (2010)
8. Li, Z., O'Brien, L., Zhang, H.: Boosting metrics: measuring cloud services from the holistic perspective. Int. J. Cloud Comput. (IJCC) **2**(4), 12–22 (2014)
9. Bermbach, D., Wittern, E.: Benchmarking web API quality. In: Bozzon, A., Cudre-Maroux, P., Pautasso, C. (eds.) ICWE 2016. LNCS, vol. 9671, pp. 188–206. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-38791-8_11
10. Gillam, L., Li, B., O'Loughlin, J., Tomar, A.P.S.: Fair benchmarking for cloud computing systems. J. Cloud Comput.: Adv. Syst. Appl. **2**(1), 6 (2013)
11. Zant, B.E., Gagnaire, M.: Performance and price analysis for cloud service providers. In: 2015 Science and Information Conference (SAI), pp. 816–822 (2015)
12. Cunha, M., Mendonca, N., Sampaio, A.: A declarative environment for automatic performance evaluation in IaaS clouds. In: 2013 IEEE Sixth International Conference on Cloud Computing, pp. 285–292 (2013)
13. Bermbach, D., Wittern, E., Tai, S.: Cloud Service Benchmarking – Measuring Quality of Cloud Services from a Client Perspective. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55483-9
14. Microsoft Corporation: Get started with Azure IoT Hub (.NET) – Microsoft Docs (2017). https://docs.microsoft.com/en-us/azure/iot-hub/iot-hub-csharp-csharp-getstarted. Accessed 08 Aug 2017
15. AWS Inc.: AWS IoT – Developer Guide (2017). http://docs.aws.amazon.com/iot/latest/developerguide/iot-dg.pdf. Accessed 02 Sept 2017