# Background

To provide a general understanding of big data and microservices, this section provides a brief overview of both concepts.

## Big Data

Since its inception (Diebold 2012), the term big data underwent a remarkable evolution. In the beginning, it was primarily used as a synonym for large data that are hard to handle for the established relational databases and technologies. Yet, nowadays it comprises a plethora of advanced data characteristics, technologies, paradigms and methods (Volk et al. 2020b). Further, even though the corresponding research community is highly active (Parlina et al. 2020), there is still no universally applied definition for the term. Instead, many rather similar ones exist that deviate in certain details (Volk et al. 2020b). However, the definition provided by the National Institute of Standards and Technology (NIST) is widely used and was therefore also chosen for this paper. It states that big data “consists of extensive datasets - primarily in the characteristics of volume, variety, velocity, and/or variability - that require a scalable architecture for efficient storage, manipulation, and analysis” (Chang and Grady 2019).

In this context, the term volume refers to the size and number of the elements that have to be processed. Variety describes the format of the data that can be structured, semi-structured or unstructured. Velocity describes the speed at which the data is captured or needs to be processed (Gandomi and Haider 2015). Finally, the variability indicates changes regarding the other characteristics that may occur over time (Chang and Grady 2019).

When it comes to the implementation and deployment of big data applications, there are still various challenges that thwart their adoption by organizations. Besides a shortage of qualified experts to build and run the systems (Gardiner et al. 2018), the design, engineering, and integration of the corresponding architectures represents a cumbersome task (Volk et al. 2019).

Consequently, there have been numerous attempts to reduce the corresponding complexity by creating aids such as reference architectures (Ataei and Litchfield 2020), decision support systems (Volk et al. 2020a), automation approaches (Fernández et al. 2020) or by applying new technologies (Freymann et al. 2020). One of those technologies, that facilitates the creation of loosely coupled and flexible applications, instead of rather static and hard to maintain monoliths, are microservices.

## Microservices

Similar to big data, there is no universally applied definition for the term microservice. However, in general, the concept can be described as an architectural approach for building software that is closely related to service oriented architectures (Jamshidi et al. 2018) and aims at the decomposition of the developed application into many small services (Nadareishvili et al. 2016, p. 6) that are rather independent of each other. Usually, they are based on business functions and focus on one specific task. As a result, the development teams are often assembled based on business capabilities instead of a specific technology layer, as it can be observed in more traditional software development. Therefore, a more comprehensive set of skills has to be present in the teams.

However, due to their independence, the distinct services can be implemented using different programming languages and technology stacks (Freymann et al. 2020), which allows the development teams greater flexibility and accommodation to their capabilities compared to a centrally dictated toolkit.

Further, each microservice runs in a separate process, the communication between the services is realized via lightweight mechanisms and they only need very little central management (Drews et al. 2017). Hence, they are perfectly geared to be deployed independently of each other, using continuous deployment tools.

In general, realizing a high degree of modularity is considered a good practice in software engineering. However, achieving it is also seen as difficult (Faitelson et al. 2018). Yet, with the microservice approach componentization is already achieved by design. Subsequently, the effort for maintenance and modification is significantly reduced and it is often sufficient to only redeploy the affected service when implementing small changes. As a result, an evolutionary design is promoted, using the services’ decomposition as a driving force to facilitate frequent and controlled changes to the system (Krylovskiy et al. 2015).

Since microservices are specialized based on the business logic, small changes or feature additions may lead to the implementation of completely new services or variations of existing ones. However, due to the overall design, those can be easily integrated into the existing application. Another advantage is that novel technologies can be at first explored on a smaller scale in just a single service before widely applying it.

# Bibtex

% This file was created with Citavi 6.11.0.0

@inproceedings{Ataei.2020,

author = {Ataei, Pouya and Litchfield, Alan},

title = {Big Data Reference Architectures, a systematic literature review},

publisher = {AIS},

booktitle = {Australasian Conference on Information Systems (ACIS) 2020},

year = {2020}

}

@misc{Chang.2019,

author = {Chang, Wo L. and Grady, Nancy},

year = {2019},

title = {NIST Big Data Interoperability Framework: Volume 1, Definitions},

url = {https://doi.org/10.6028/NIST.SP.1500-1r2},

file = {Chang, Grady - NIST Big Data Interoperability Framework:Attachments/Chang, Grady - NIST Big Data Interoperability Framework.pdf:application/pdf}

}

@article{Diebold.2012,

author = {Diebold, Francis X.},

year = {2012},

title = {On the Origin(s) and Development of the Term 'Big Data'},

journal = {SSRN Electronic Journal},

doi = {10.2139/ssrn.2152421}

}

@inproceedings{Drews.2017,

author = {Drews, Paul and Schirmer, Ingrid and Horlach, Bettina and Tekaat, Carsten},

title = {Bimodal Enterprise Architecture Management: The Emergence of a New EAM Function for a BizDevOps-Based Fast IT},

pages = {57--64},

publisher = {IEEE},

isbn = {978-1-5386-1568-3},

booktitle = {2017 IEEE 21st International Enterprise Distributed Object Computing Workshop (EDOCW)},

year = {2017},

doi = {10.1109/EDOCW.2017.18}

}

@incollection{Faitelson.2018,

author = {Faitelson, David and Heinrich, Robert and Tyszberowicz, Shmuel},

title = {Functional Decomposition for Software Architecture Evolution},

pages = {377--400},

volume = {880},

publisher = {{Springer International Publishing}},

isbn = {978-3-319-94763-1},

series = {Communications in computer and information science},

editor = {Pires, Lu{\'i}s Ferreira and Hammoudi, Slimane and Selic, Bran},

booktitle = {Model-Driven Engineering and Software Development},

year = {2018},

address = {Cham},

doi = {10.1007/978-3-319-94764-8{\textunderscore }16}

}

@article{Fernandez.2020,

author = {Fern{\'a}ndez, A. M. and Guti{\'e}rrez-Avil{\'e}s, D. and Troncoso, A. and Mart{\'i}nez--{\'A}lvarez, F.},

year = {2020},

title = {Automated Deployment of a Spark Cluster with Machine Learning Algorithm Integration},

url = {http://www.sciencedirect.com/science/article/pii/S2214579620300034},

pages = {100135},

volume = {19-20},

issn = {22145796},

journal = {Big Data Research},

doi = {10.1016/j.bdr.2020.100135}

}

@inproceedings{Freymann.2020,

author = {Freymann, Andreas and Maier, Florian and Schaefer, Kristian and B{\"o}hnel, Tom},

title = {Tackling the Six Fundamental Challenges of Big Data in Research Projects by Utilizing a Scalable and Modular Architecture},

pages = {249--256},

publisher = {{SCITEPRESS - Science and Technology Publications}},

isbn = {978-989-758-426-8},

booktitle = {Proceedings of the 5th International Conference on Internet of Things, Big Data and Security},

year = {2020},

doi = {10.5220/0009388602490256}

}

@article{Gandomi.2015,

author = {Gandomi, Amir and Haider, Murtaza},

year = {2015},

title = {Beyond the hype: Big data concepts, methods, and analytics},

pages = {137--144},

volume = {35},

number = {2},

issn = {02684012},

journal = {International Journal of Information Management},

doi = {10.1016/j.ijinfomgt.2014.10.007}

}

@article{Gardiner.2018,

author = {Gardiner, Adrian and Aasheim, Cheryl and Rutner, Paige and Williams, Susan},

year = {2018},

title = {Skill Requirements in Big Data: A Content Analysis of Job Advertisements},

pages = {374--384},

volume = {58},

number = {4},

issn = {0887-4417},

journal = {Journal of Computer Information Systems},

doi = {10.1080/08874417.2017.1289354}

}

@article{Jamshidi.2018,

author = {Jamshidi, Pooyan and Pahl, Claus and Mendonca, Nabor C. and Lewis, James and Tilkov, Stefan},

year = {2018},

title = {Microservices: The Journey So Far and Challenges Ahead},

pages = {24--35},

volume = {35},

number = {3},

issn = {0740-7459},

journal = {IEEE Software},

doi = {10.1109/MS.2018.2141039}

}

@inproceedings{Krylovskiy.2015,

author = {Krylovskiy, Alexandr and Jahn, Marco and Patti, Edoardo},

title = {Designing a Smart City Internet of Things Platform with Microservice Architecture},

pages = {25--30},

publisher = {IEEE},

isbn = {978-1-4673-8103-1},

editor = {Awan, Irfan},

booktitle = {Proceedings of the 3rd International Conference on Future Internet of Things and Cloud (FiCloud 2015)},

year = {2015},

address = {Piscataway, NJ},

doi = {10.1109/FiCloud.2015.55}

}

@book{Nadareishvili.2016,

author = {Nadareishvili, Irakli and Mitra, Ronnie and McLarty, Matt and Amundsen, Michael},

year = {2016},

title = {Microservice architecture: Aligning principles, practices, and culture},

address = {Beijing and Boston and Farnham and Sebastopol and Tokyo},

edition = {First edition},

publisher = {O´Reilly},

isbn = {978-1-491-95625-0}

}

@article{Parlina.2020,

author = {Parlina, Anne and Ramli, Kalamullah and Murfi, Hendri},

year = {2020},

title = {Theme Mapping and Bibliometrics Analysis of One Decade of Big Data Research in the Scopus Database},

pages = {69--94},

volume = {11},

number = {2},

journal = {Information},

doi = {10.3390/info11020069}

}

@incollection{Volk.2020,

author = {Volk, Matthias and Staegemann, Daniel and Bosse, Sascha and Nahhas, Abdulrahman and Turowski, Klaus},

title = {Towards a Decision Support System for Big Data Projects},

pages = {357--368},

publisher = {{GITO Verlag}},

isbn = {9783955453350},

editor = {Gronau, Norbert and Heine, Moreen and Poustcchi, K. and Krasnova, H.},

booktitle = {WI2020 Zentrale Tracks},

year = {2020},

doi = {10.30844/wi{\textunderscore }2020{\textunderscore }c11-volk}

}

@inproceedings{Volk.2019,

author = {Volk, Matthias and Staegemann, Daniel and Pohl, Matthias and Turowski, Klaus},

title = {Challenging Big Data Engineering: Positioning of Current and Future Development},

pages = {351--358},

publisher = {{SCITEPRESS - Science and Technology Publications}},

isbn = {978-989-758-369-8},

booktitle = {Proceedings of the 4th International Conference on Internet of Things, Big Data and Security},

year = {2019},

doi = {10.5220/0007748803510358}

}

@incollection{Volk.2020b,

author = {Volk, Matthias and Staegemann, Daniel and Turowski, Klaus},

title = {Big Data},

pages = {1--18},

publisher = {{Springer Fachmedien Wiesbaden}},

isbn = {978-3-658-17345-6},

series = {Springer Reference Wirtschaft},

editor = {Kollmann, Tobias},

booktitle = {Handbuch Digitale Wirtschaft},

year = {2020},

address = {Wiesbaden},

doi = {10.1007/978-3-658-17345-6{\textunderscore }71-1}

}

Publication bibliography

Ataei, Pouya; Litchfield, Alan (2020): Big Data Reference Architectures, a systematic literature review. In : Australasian Conference on Information Systems (ACIS) 2020. Wellington, New Zealand: AIS.

Chang, Wo L.; Grady, Nancy (2019): NIST Big Data Interoperability Framework: Volume 1, Definitions. Available online at https://doi.org/10.6028/NIST.SP.1500-1r2.

Diebold, Francis X. (2012): On the Origin(s) and Development of the Term 'Big Data'. In *SSRN Journal*. DOI: 10.2139/ssrn.2152421.

Drews, Paul; Schirmer, Ingrid; Horlach, Bettina; Tekaat, Carsten (2017): Bimodal Enterprise Architecture Management: The Emergence of a New EAM Function for a BizDevOps-Based Fast IT. In : 2017 IEEE 21st International Enterprise Distributed Object Computing Workshop (EDOCW). 2017 IEEE 21st International Enterprise Distributed Object Computing Workshop (EDOCW). Quebec City, QC, 10.10.2017 - 13.10.2017: IEEE, pp. 57–64.

Faitelson, David; Heinrich, Robert; Tyszberowicz, Shmuel (2018): Functional Decomposition for Software Architecture Evolution. In Luís Ferreira Pires, Slimane Hammoudi, Bran Selic (Eds.): Model-Driven Engineering and Software Development, vol. 880. Cham: Springer International Publishing (Communications in computer and information science), pp. 377–400.

Fernández, A. M.; Gutiérrez-Avilés, D.; Troncoso, A.; Martínez–Álvarez, F. (2020): Automated Deployment of a Spark Cluster with Machine Learning Algorithm Integration. In *Big Data Research* 19-20, p. 100135. DOI: 10.1016/j.bdr.2020.100135.

Freymann, Andreas; Maier, Florian; Schaefer, Kristian; Böhnel, Tom (2020): Tackling the Six Fundamental Challenges of Big Data in Research Projects by Utilizing a Scalable and Modular Architecture. In : Proceedings of the 5th International Conference on Internet of Things, Big Data and Security. 5th International Conference on Internet of Things, Big Data and Security. Prague, Czech Republic, 07.05.2020 - 09.05.2020: SCITEPRESS - Science and Technology Publications, pp. 249–256.

Gandomi, Amir; Haider, Murtaza (2015): Beyond the hype: Big data concepts, methods, and analytics. In *International Journal of Information Management* 35 (2), pp. 137–144. DOI: 10.1016/j.ijinfomgt.2014.10.007.

Gardiner, Adrian; Aasheim, Cheryl; Rutner, Paige; Williams, Susan (2018): Skill Requirements in Big Data: A Content Analysis of Job Advertisements. In *Journal of Computer Information Systems* 58 (4), pp. 374–384. DOI: 10.1080/08874417.2017.1289354.

Jamshidi, Pooyan; Pahl, Claus; Mendonca, Nabor C.; Lewis, James; Tilkov, Stefan (2018): Microservices: The Journey So Far and Challenges Ahead. In *IEEE Softw.* 35 (3), pp. 24–35. DOI: 10.1109/MS.2018.2141039.

Krylovskiy, Alexandr; Jahn, Marco; Patti, Edoardo (2015): Designing a Smart City Internet of Things Platform with Microservice Architecture. In Irfan Awan (Ed.): Proceedings of the 3rd International Conference on Future Internet of Things and Cloud (FiCloud 2015). Rome, Italy, 24.08.2015 - 26.08.2015. Piscataway, NJ: IEEE, pp. 25–30.

Nadareishvili, Irakli; Mitra, Ronnie; McLarty, Matt; Amundsen, Michael (2016): Microservice architecture. Aligning principles, practices, and culture. First edition. Beijing, Boston, Farnham, Sebastopol, Tokyo: O´Reilly.

Parlina, Anne; Ramli, Kalamullah; Murfi, Hendri (2020): Theme Mapping and Bibliometrics Analysis of One Decade of Big Data Research in the Scopus Database. In *Information* 11 (2), pp. 69–94. DOI: 10.3390/info11020069.

Volk, Matthias; Staegemann, Daniel; Bosse, Sascha; Nahhas, Abdulrahman; Turowski, Klaus (2020a): Towards a Decision Support System for Big Data Projects. In Norbert Gronau, Moreen Heine, K. Poustcchi, H. Krasnova (Eds.): WI2020 Zentrale Tracks: GITO Verlag, pp. 357–368.

Volk, Matthias; Staegemann, Daniel; Pohl, Matthias; Turowski, Klaus (2019): Challenging Big Data Engineering: Positioning of Current and Future Development. In : Proceedings of the 4th International Conference on Internet of Things, Big Data and Security. 4th International Conference on Internet of Things, Big Data and Security. Heraklion, Crete, Greece, 02.05.2019 - 04.05.2019: SCITEPRESS - Science and Technology Publications, pp. 351–358.

Volk, Matthias; Staegemann, Daniel; Turowski, Klaus (2020b): Big Data. In Tobias Kollmann (Ed.): Handbuch Digitale Wirtschaft. Wiesbaden: Springer Fachmedien Wiesbaden (Springer Reference Wirtschaft), pp. 1–18.