UNIVERSITY OF CAMBRIDGE

Data-oriented Architectures and Real-World ML Systems

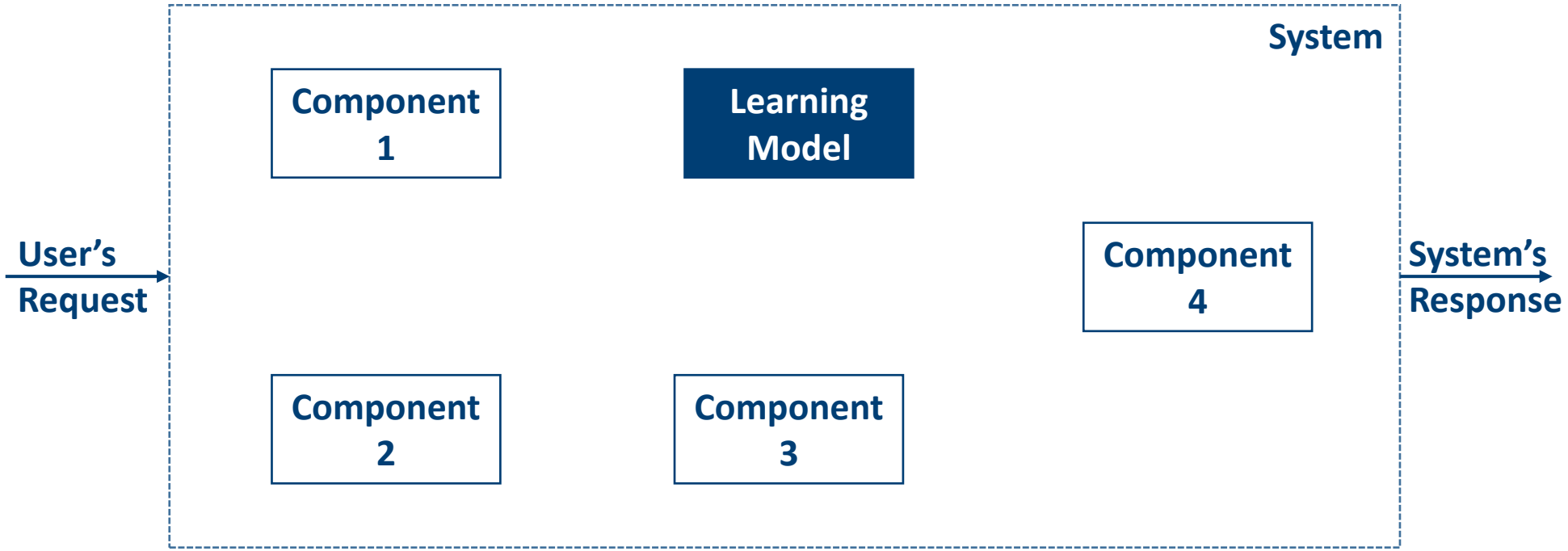
Christian Cabrera Jojoa
Research Associate
chc.79@cam.ac.uk

Department of Computer Science and Technology

ML Models

ML Models

ML Models at deployment

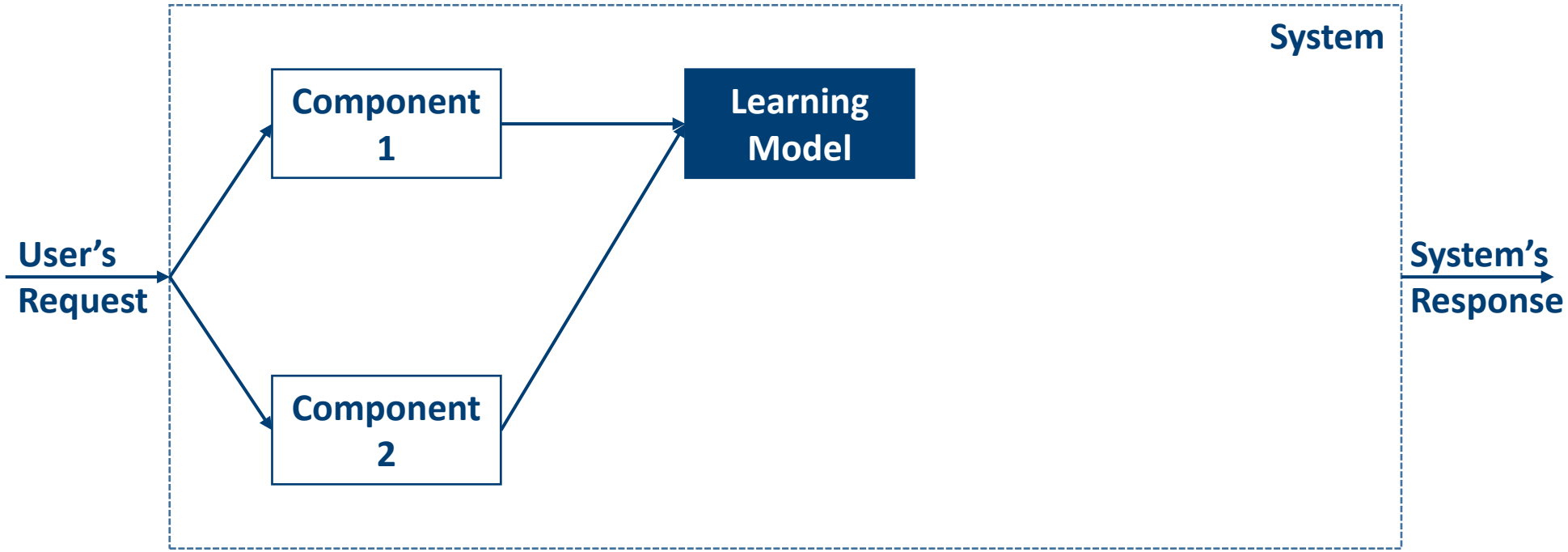


Learning models are deployed as part of larger systems.

Systems address end users' requirements.

Users' requirements must drive the design of the systems.

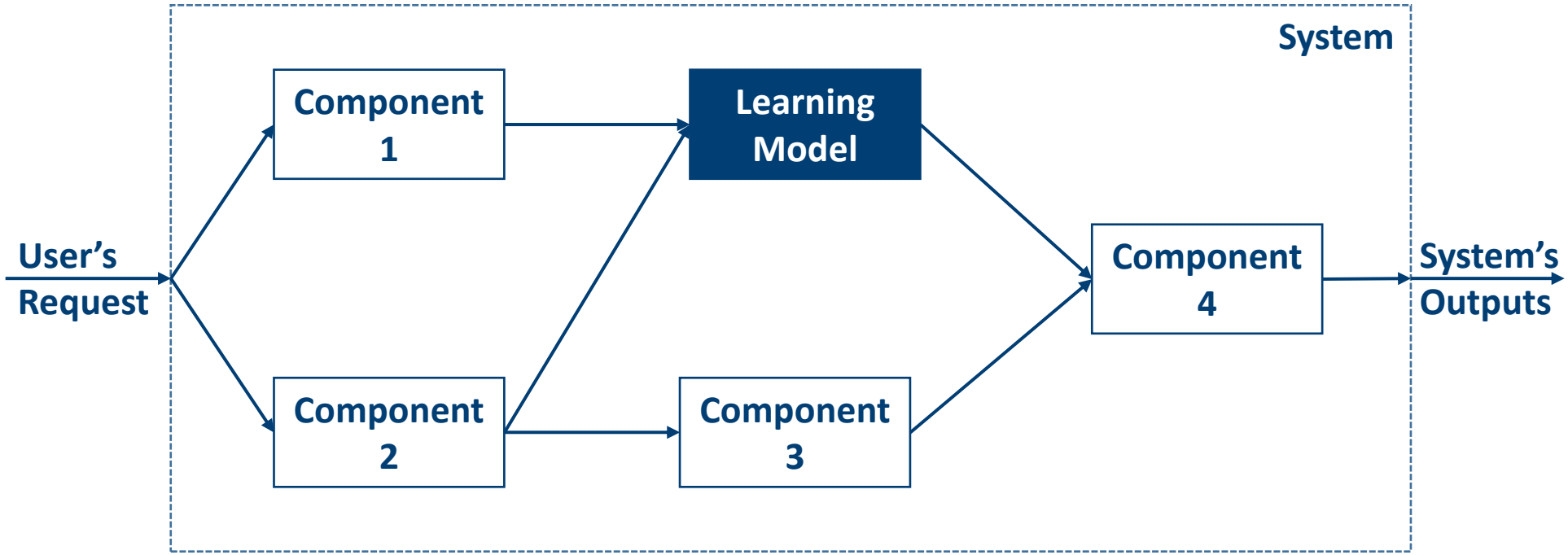
ML Models at deployment



Systems are composed of multiple components

These components can impact the behavior of the learning models

ML Models at deployment



Learning models outputs can be inputs for other components

Systems' outputs are aggregations of the data processed by different components.

Systems Design

Users' requirements drive the design of the systems…

Systems Design

Software Design Decisions:

You have three learning models that predict Y.

Systems Design

Software Design Decisions:

You have three models that predict X.

Learning Model	Accuracy
Model1	88%
Model2	90%
Model3	98%

Systems Design

Software Design Decisions:

You have three models that predict X.

Learning Model	Accuracy	Latency
Model1	88%	3 secs
Model2	90%	4 secs
Model3	98%	10 secs

Low latency requirement!

< 5secs

Systems Design

Software Design Decisions:

You have three models that predict X.

Learning Model	Accuracy	Latency	Resources Demand
Model1	88%	3 secs	Low
Model2	90%	4 secs	Medium
Model3	98%	10 secs	High

Constraint resources!

Systems Design

Software Design Decisions:

You have three models that predict X.

Learning Model	Accuracy	Latency	Resources Demand
Model1	88%	3 secs	Low
Model2	90%	4 secs	Medium
Model3	98%	10 secs	High

Which one should I select?

Systems Design

Software Design Decisions:

You have three models that predict X.

Learning Model	Accuracy	Latency	Resources Demand
Model1	88%	3 secs	Low
Model2	90%	4 secs	Medium
Model3	98%	10 secs	High

Which one should I select?

System design decisions…

Systems Design

Systems' design decisions change from system to system…

Systems Design

Systems' design decisions change from system to system…

But we can identify common requirements between current ML-based systems.

ML-based Systems and SOA

Maintainability

High availability

Scalability

Low latency

Data Ownership

Resource constraint environments

Transparency and monitoring

ML-based Systems and SOA

Maintainability

High availability

Scalability

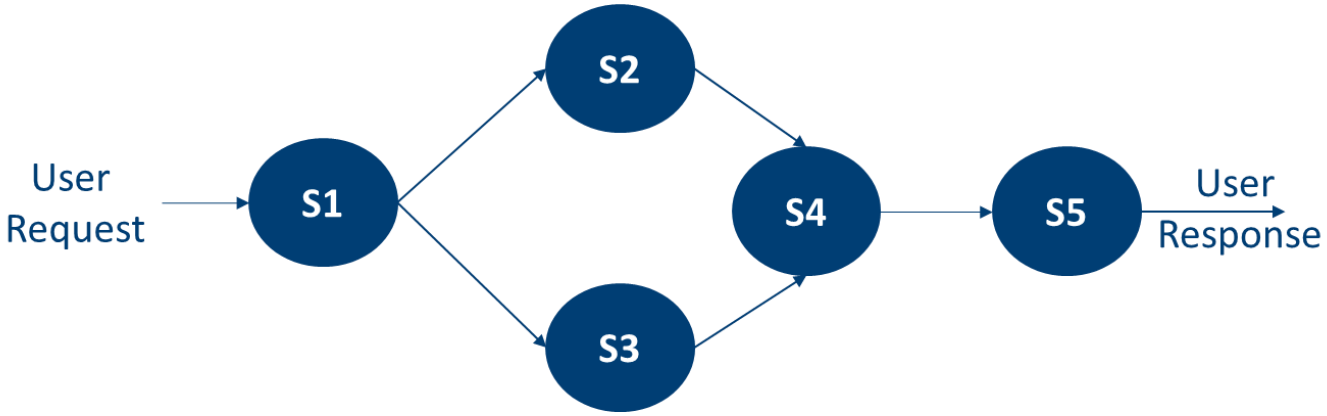
Low latency

Data Ownership

Resource constraint environments

Transparency and monitoring

Service-oriented Architectures (SOAs)



ML-based Systems and SOA

Maintainability

High availability

Scalability

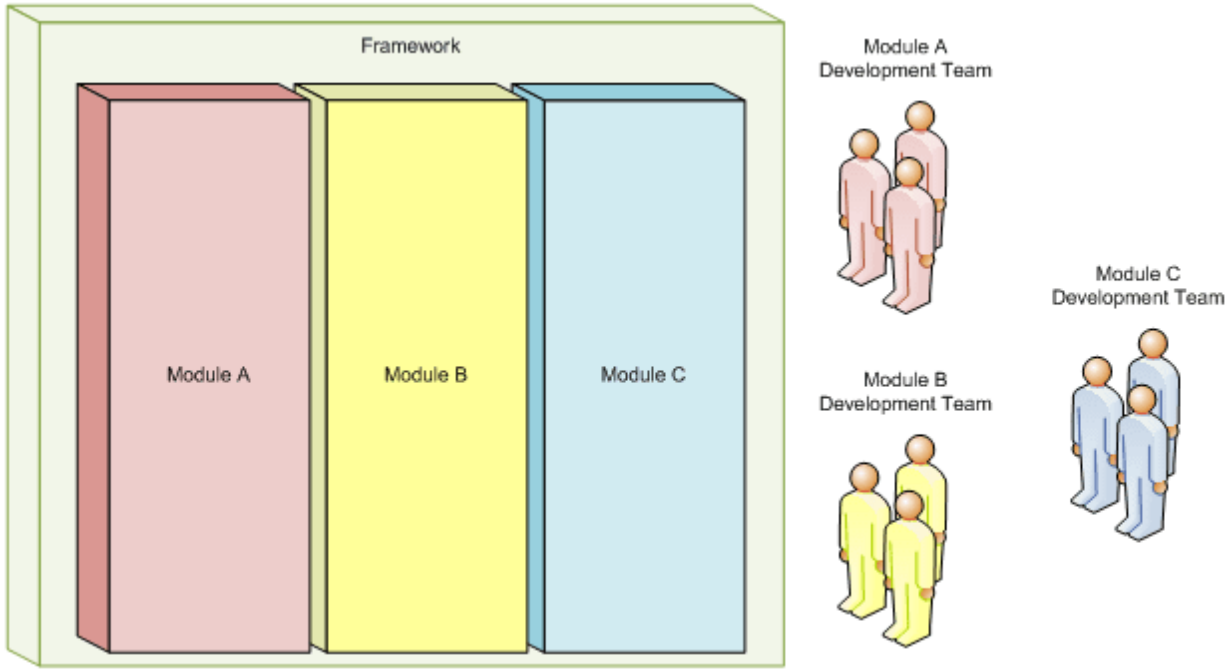
Low latency

Data Ownership

Resource constraint environments

Transparency and monitoring

Service-oriented Architectures (SOAs)



Source: https://dev.to/suspir0n/soc-separation-of-concerns-5ak7

ML-based Systems and SOA

Maintainability

High availability

Scalability

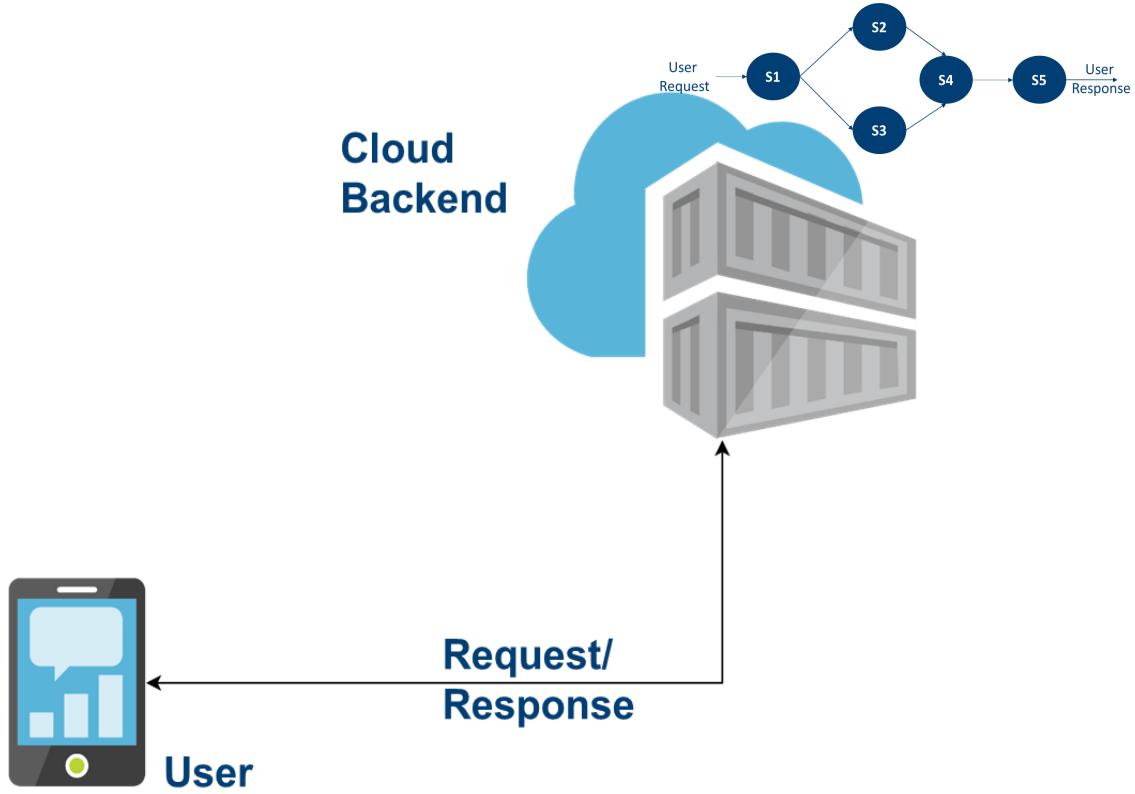
Low latency

Data Ownership

Resource constraint environments

Transparency and monitoring

Service-oriented Architectures (SOAs)



ML-based Systems and SOA

Maintainability

High availability

Scalability

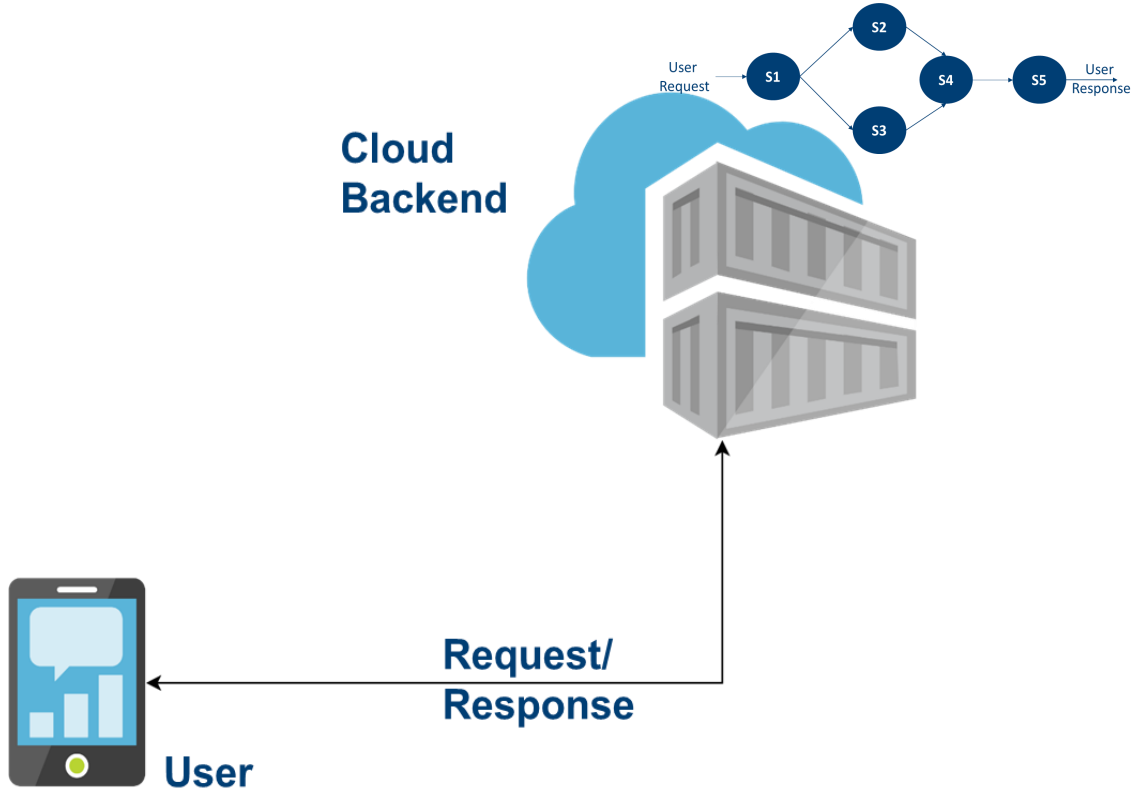
Low latency

Data Ownership

Resource constraint environments

Transparency and monitoring

Service-oriented Architectures (SOAs)



ML-based Systems and SOA

Maintainability

High availability

Scalability

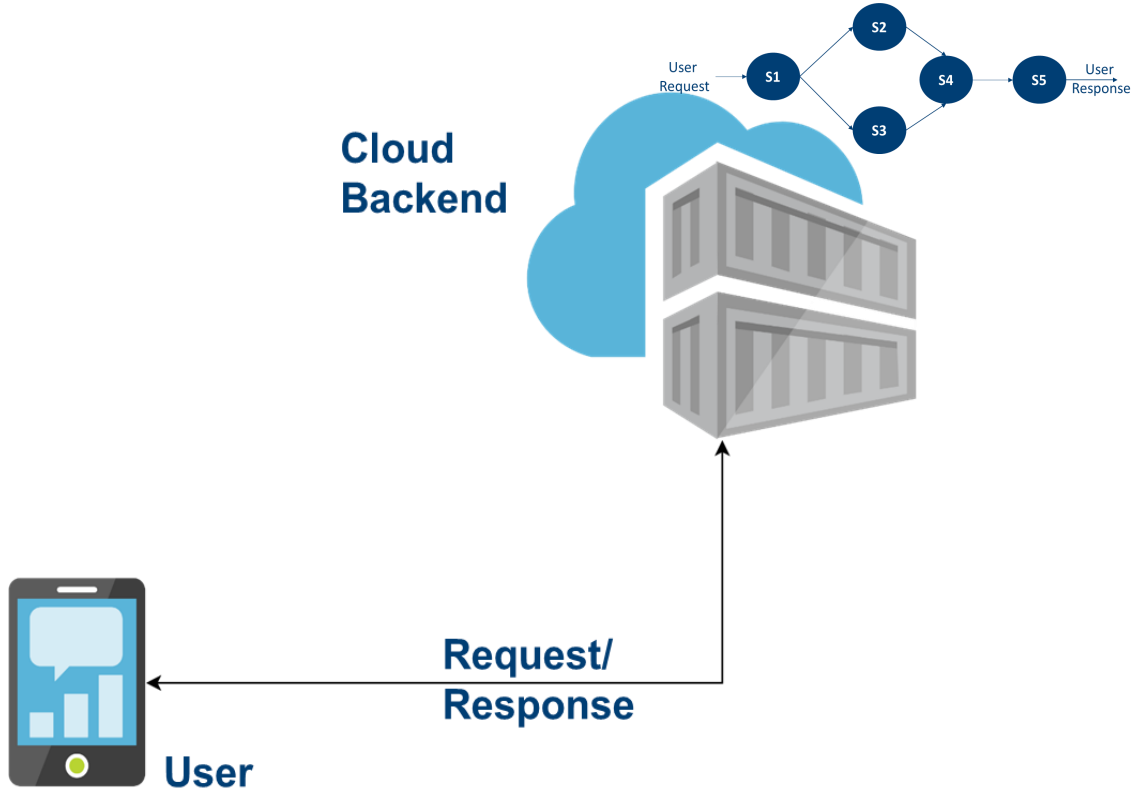
Low latency

Data Ownership

Resource constraint environments

Transparency and monitoring

Service-oriented Architectures (SOAs)



ML-based Systems and SOA

Maintainability

High availability

Scalability

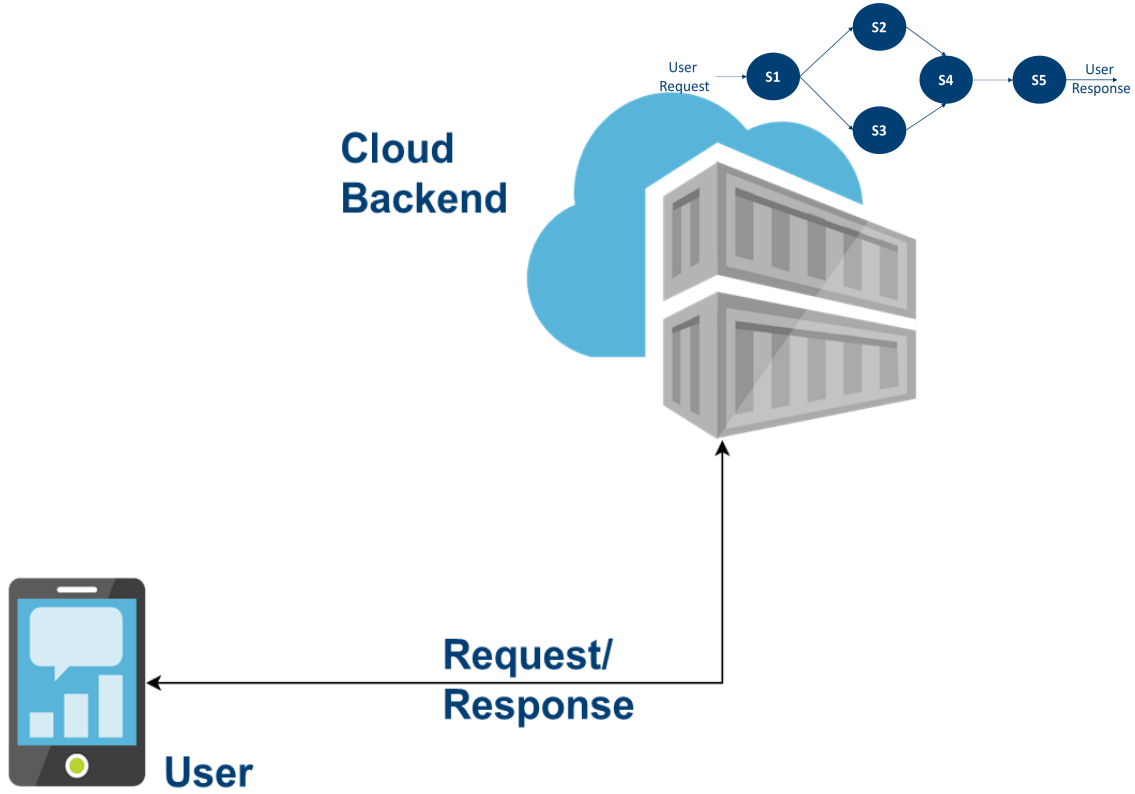
Low latency

Data Ownership

Resource constraint environments

Transparency and monitoring

Service-oriented Architectures (SOAs)



ML-based Systems and SOA

Maintainability

High availability

Scalability

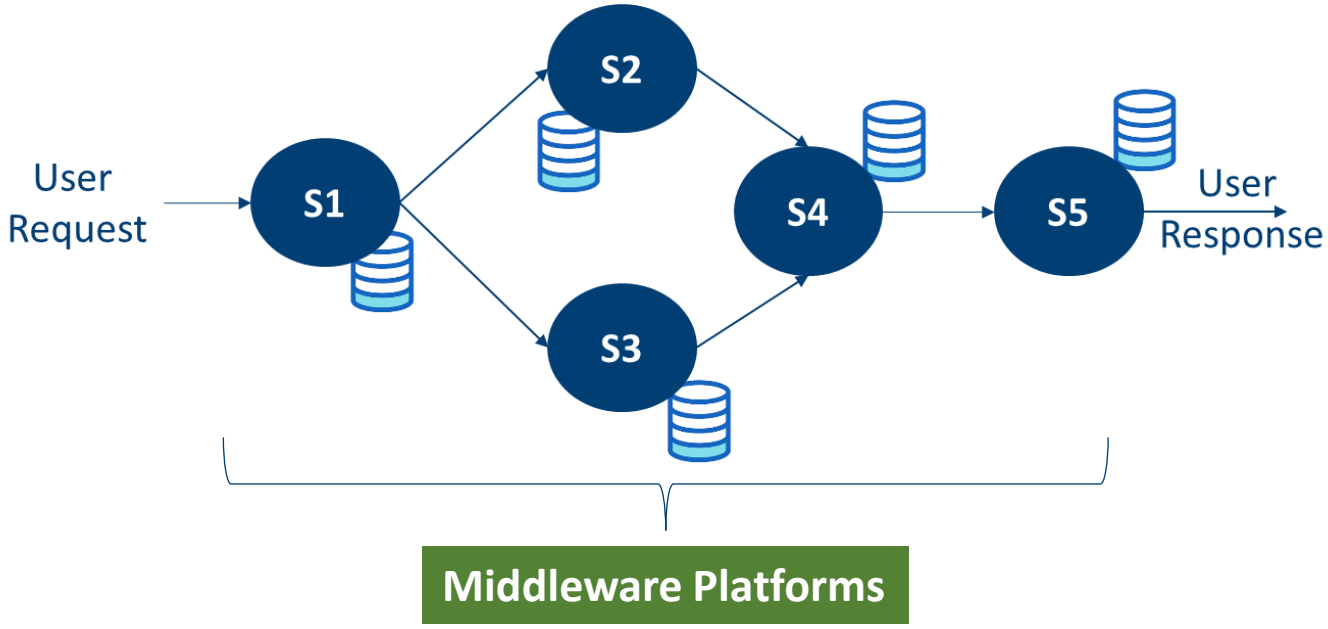
Low latency

Data Ownership

Resource constraint environments

Transparency and monitoring

Service-oriented Architectures (SOAs)



ML-based Systems and SOA

Maintainability

High availability

Scalability

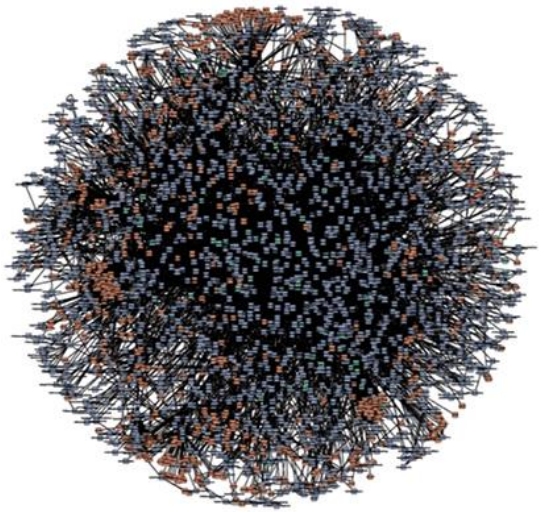
Low latency

Data Ownership

Resource constraint environments

Transparency and monitoring

Service-oriented Architectures (SOAs)



amazon.com

NETFLIX

Source: https://www.divante.com/blog/10-companies-that-implemented-the-microservice-architecture-and-paved-the-way-for-others

ML-based Systems and DOA

Data-oriented Architectures (DOAs)

Maintainability

High availability

Scalability

Low latency

Data Ownership

Resource constraint environments

Transparency and monitoring

ML-based Systems and DOA

Data-oriented Architectures (DOAs)

Maintainability

High availability

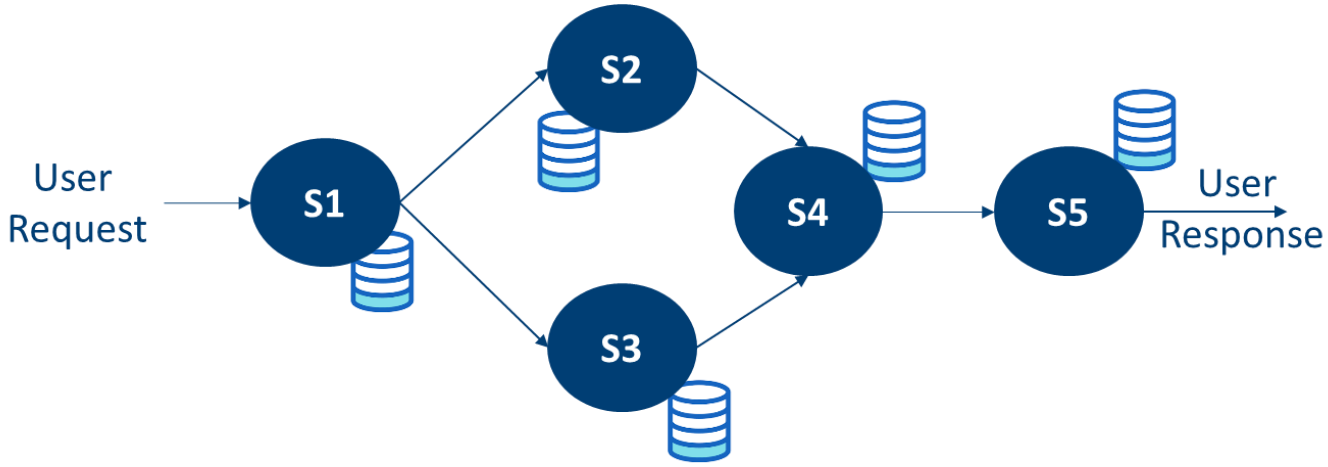
Scalability

Low latency

Data Ownership

Resource constraint environments

Transparency and monitoring



ML-based Systems and DOA

Maintainability

High availability

Scalability

Low latency

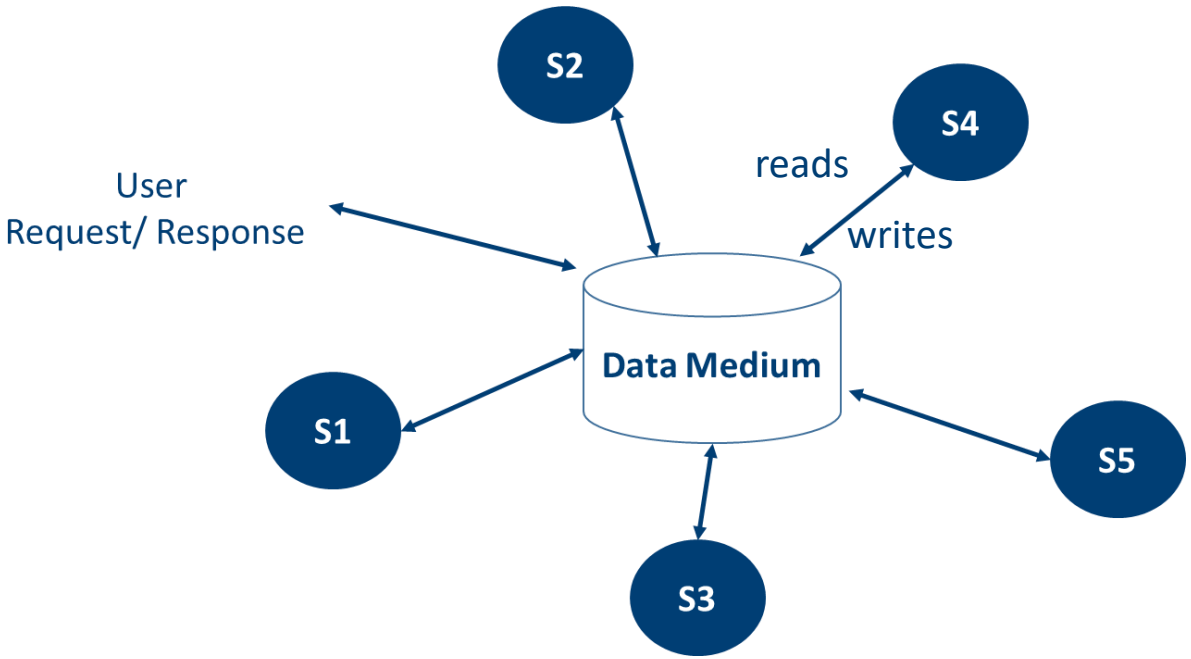
Data Ownership

Resource constraint environments

Transparency and monitoring

Data-oriented Architectures (DOAs)

Data First



ML-based Systems and DOA

Maintainability

High availability

Scalability

Low latency

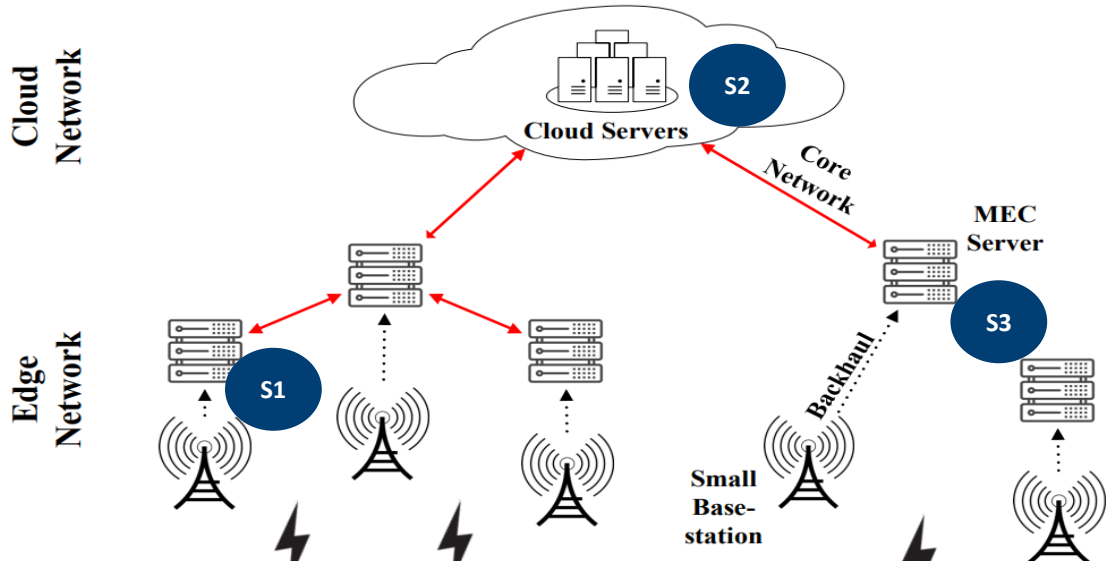
Data Ownership

Resource constraint environments

Transparency and monitoring

Data-oriented Architectures (DOAs)

Prioritise decentralisation



ML-based Systems and DOA

Maintainability

High availability

Scalability

Low latency

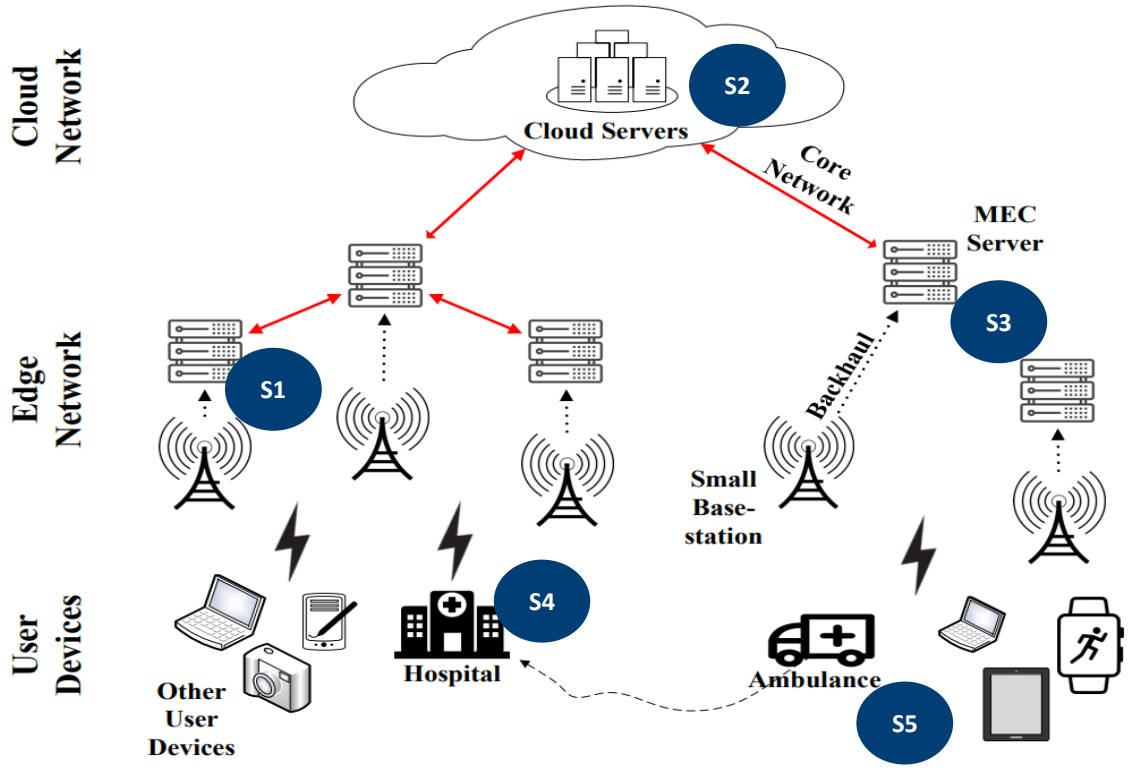
Data Ownership

Resource constraint environments

Transparency and monitoring

Data-oriented Architectures (DOAs)

Openness



ML-based Systems and DOA

Ewaso Nyiro River

Source: https://en.wikipedia.org/wiki/Ewaso_Ng%27iro

Water level monitoring project at DeKUT (DSA 2022 - Arusha)

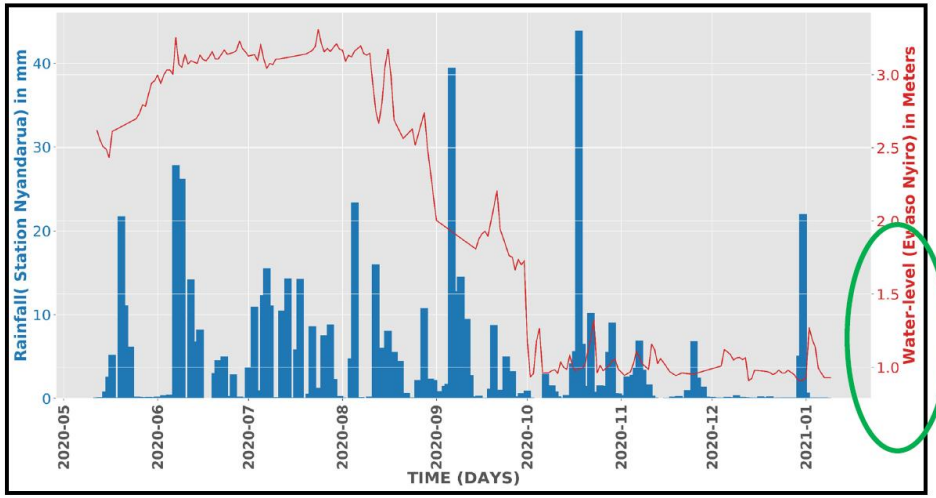


Figure 7: Rainfall profile comparison to the water level profile

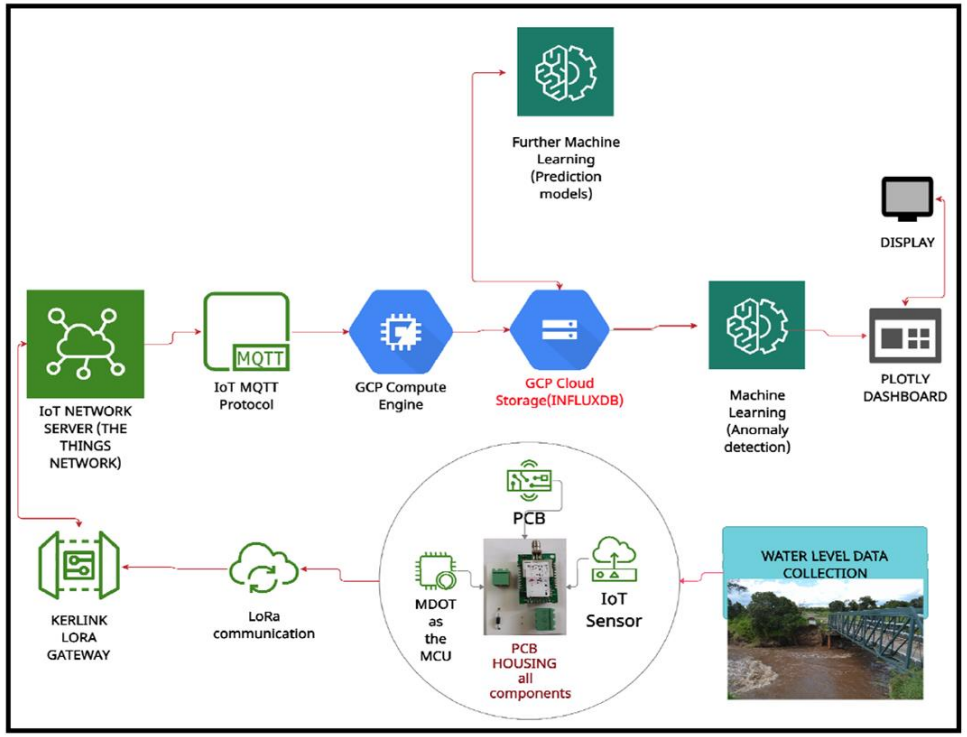
Catchment test

ML-based Systems and DOA

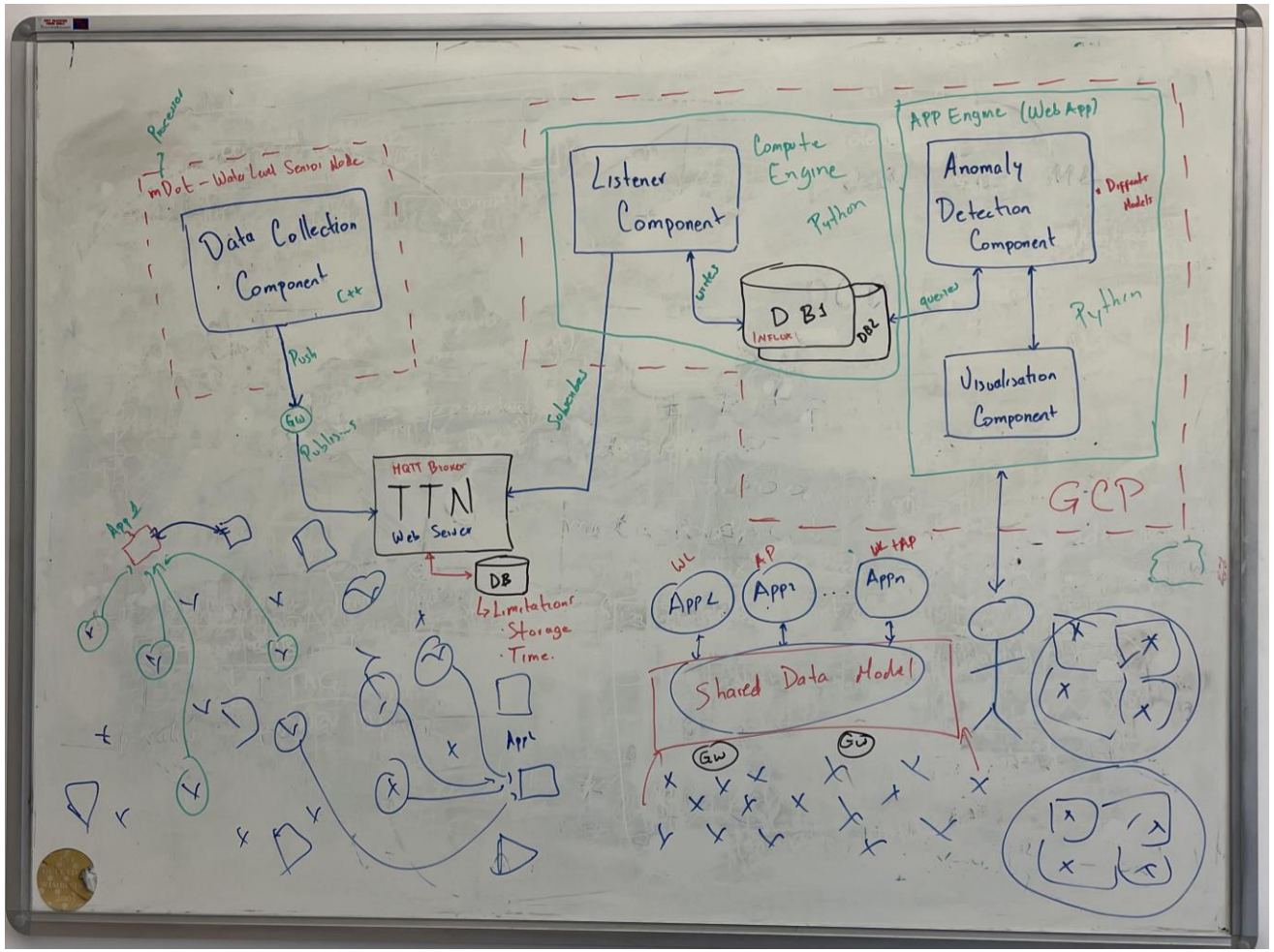
Ewaso Nyiro River

Source: https://en.wikipedia.org/wiki/Ewaso_Ng%27iro

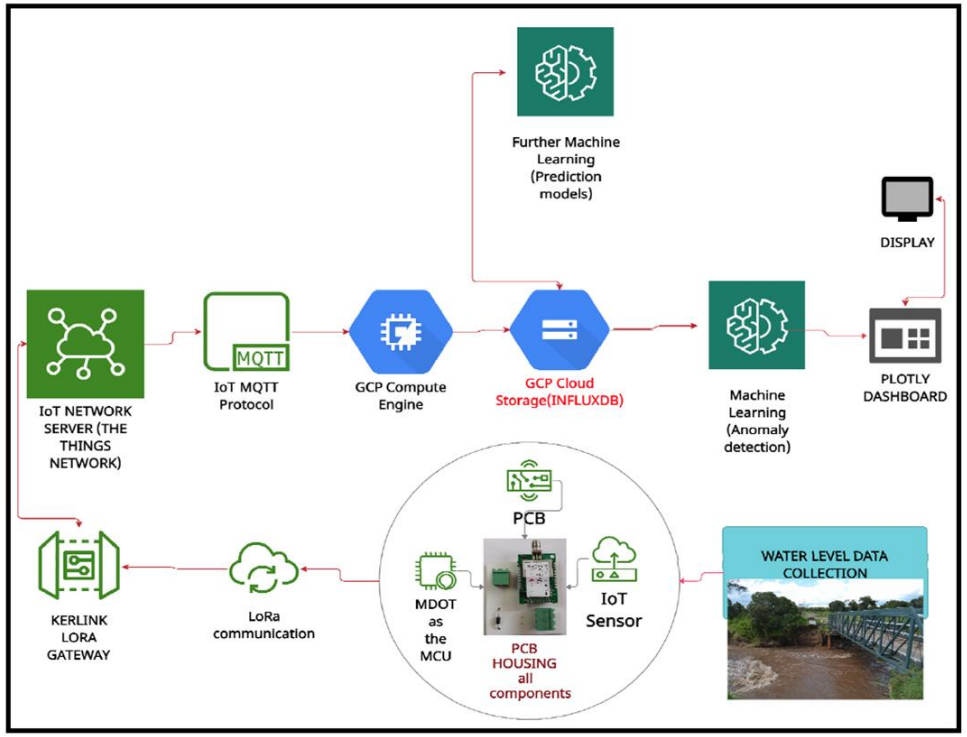
Water level monitoring project at DeKUT (DSA 2022 - Arusha)



ML-based Systems and DOA

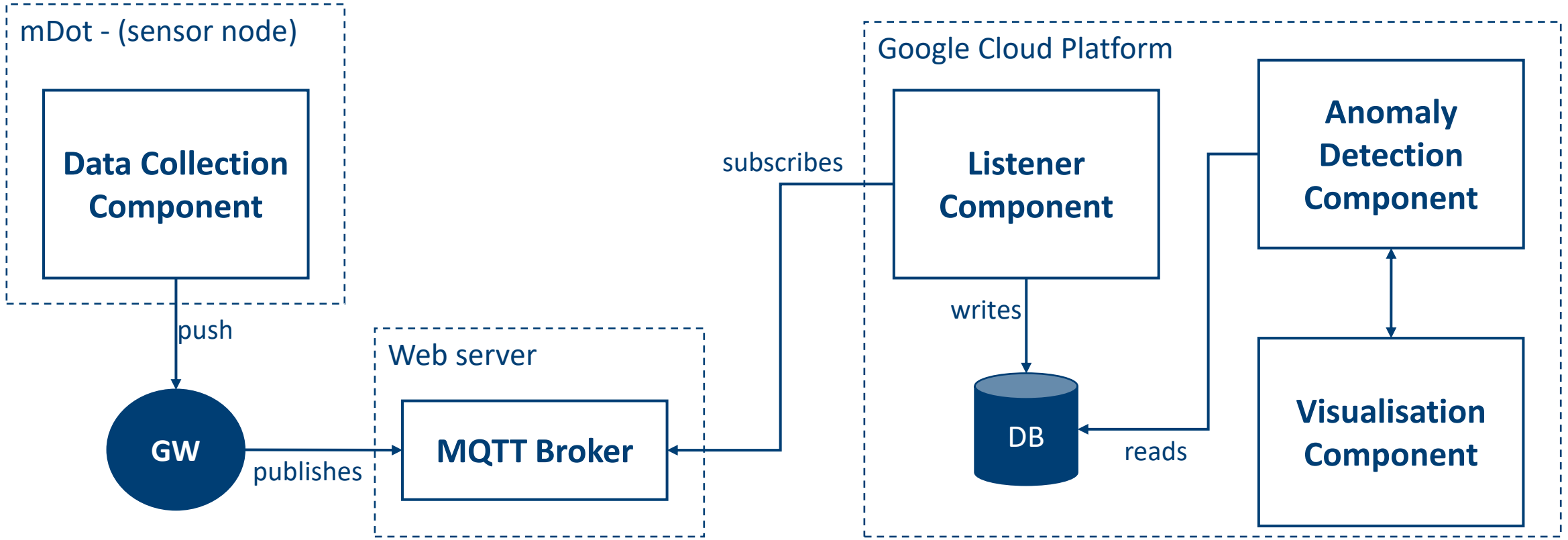


Water level monitoring project at DeKUT (DSA 2022 - Arusha)



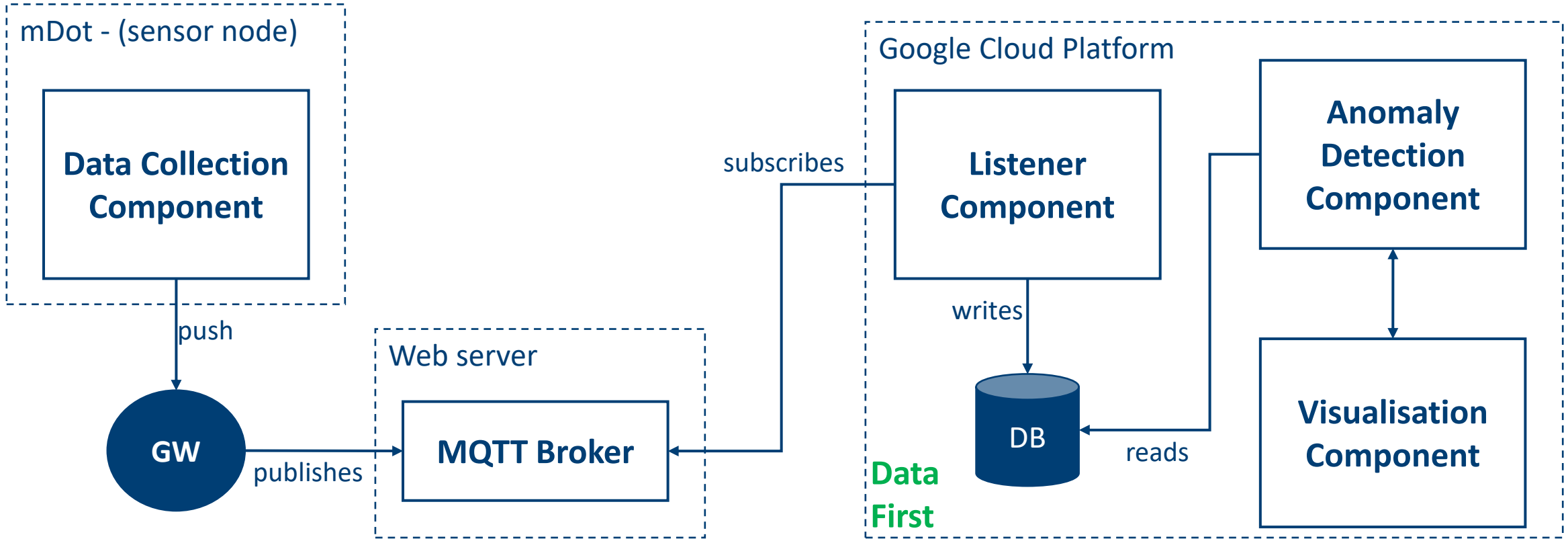
ML-based Systems and DOA

Water level monitoring project at DeKUT



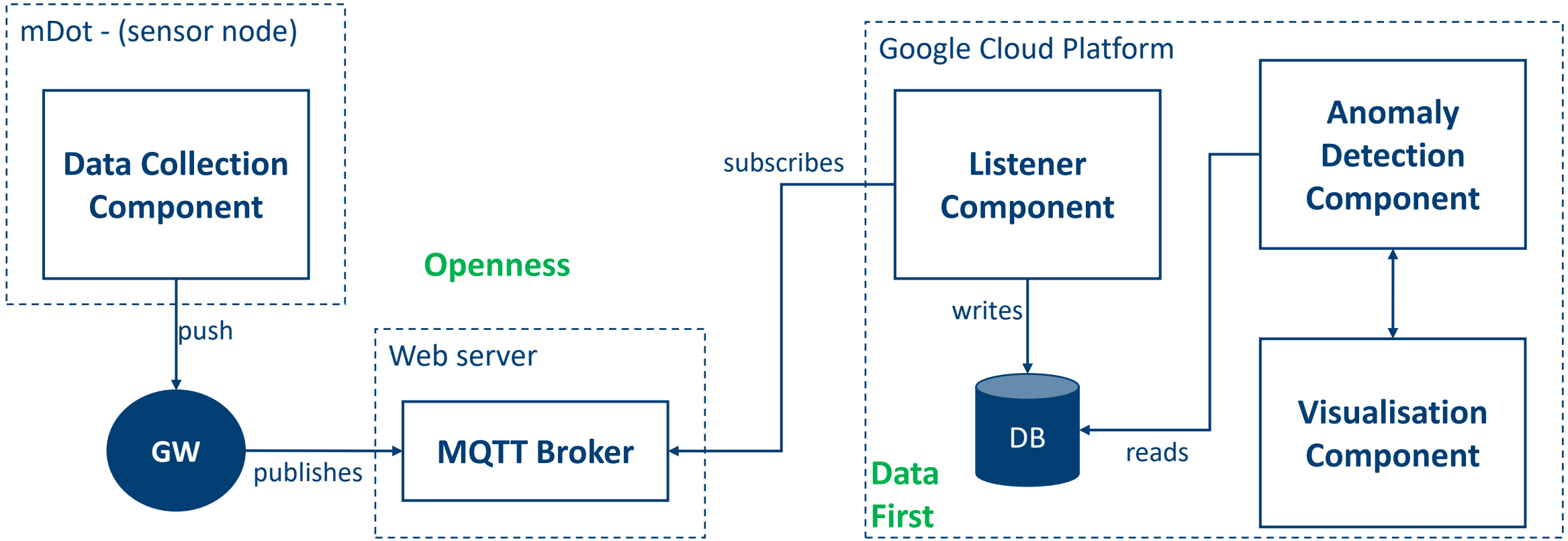
ML-based Systems and DOA

Water level monitoring project at DeKUT



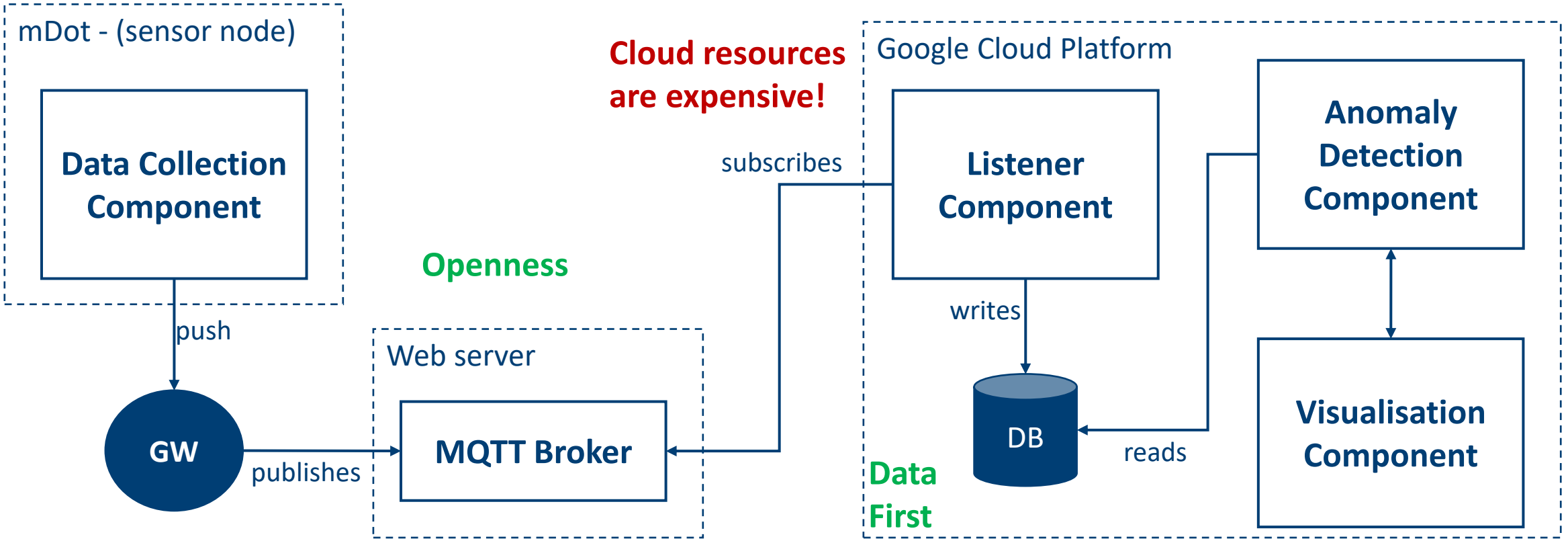
ML-based Systems and DOA

Water level monitoring project at DeKUT



ML-based Systems and DOA

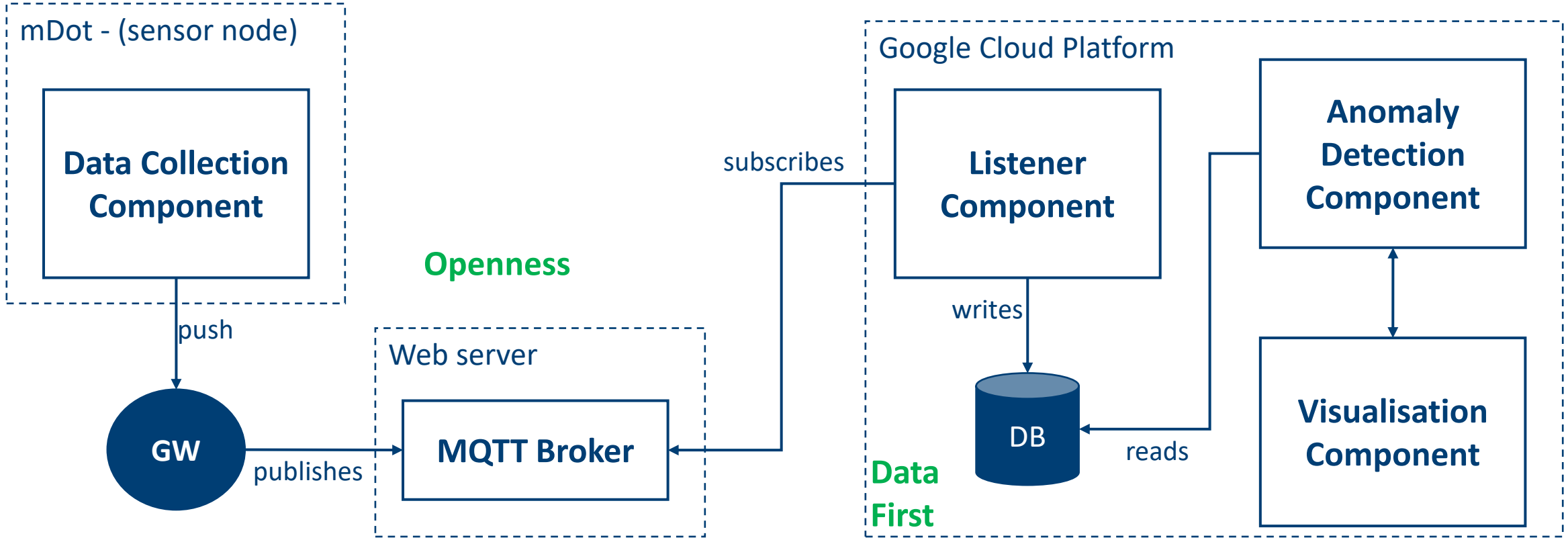
Water level monitoring project at DeKUT



ML-based Systems and DOA

Decentralisation

Water level monitoring project at DeKUT

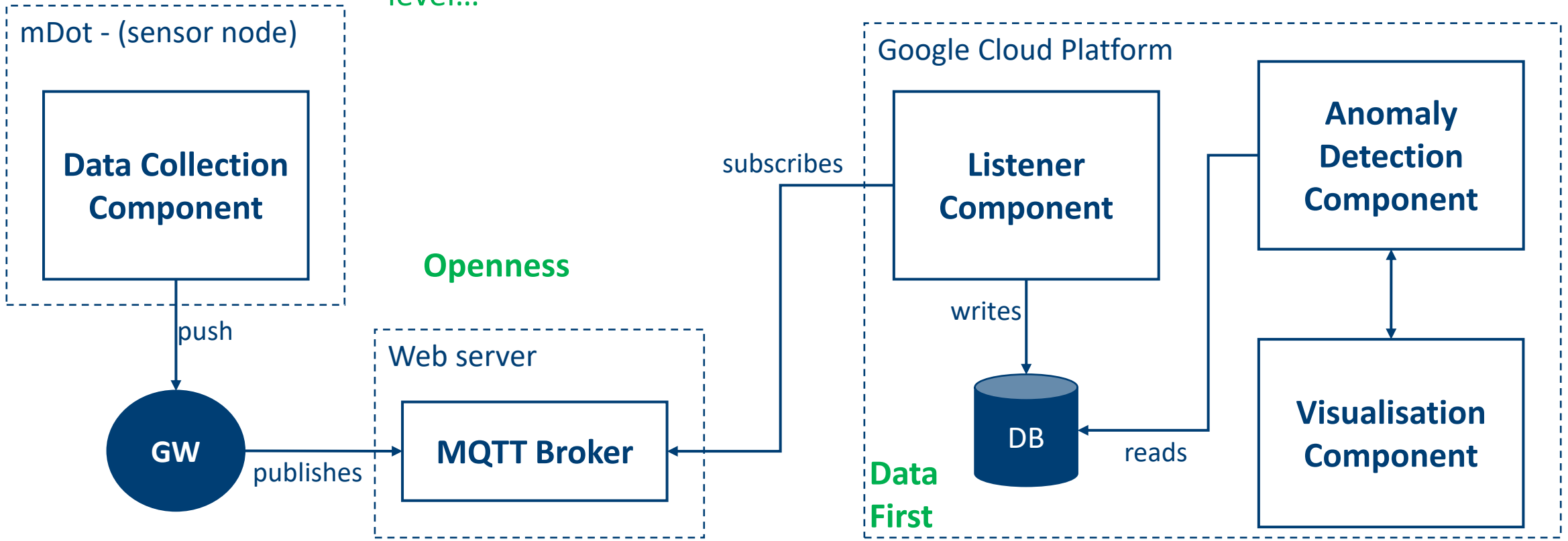


ML-based Systems and DOA

Decentralisation

Preprocess data at the gateway level…

Water level monitoring project at DeKUT



ML-based Systems and DOA

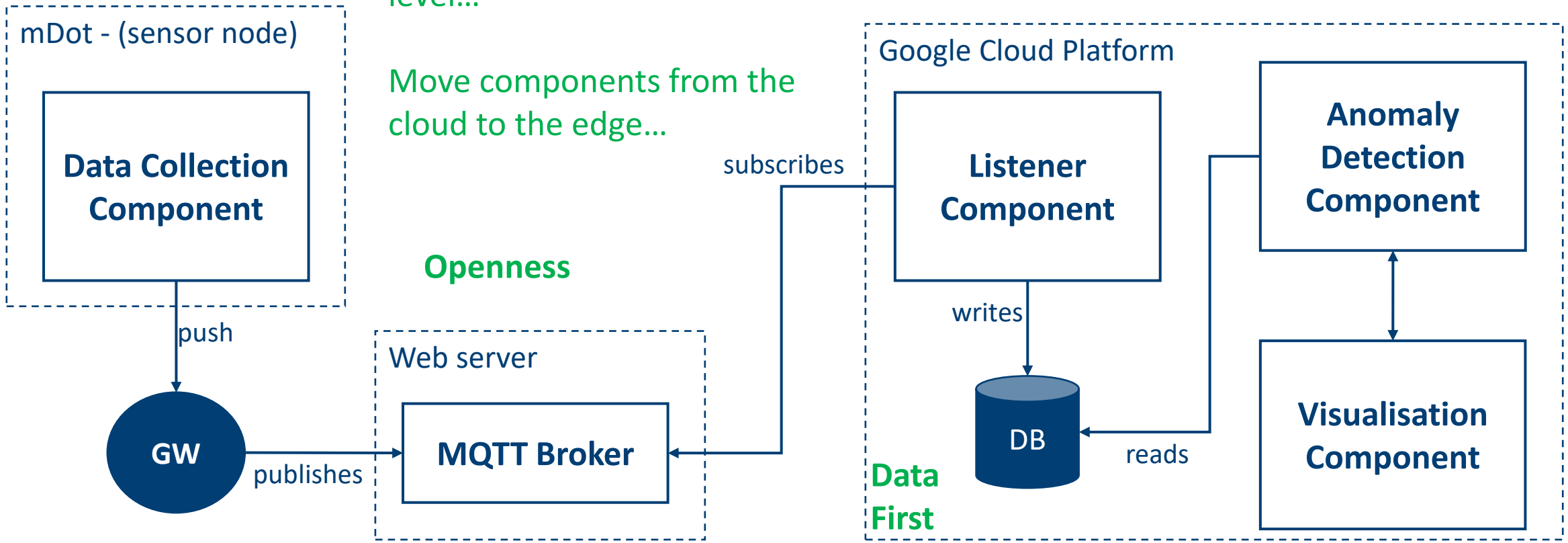
Decentralisation

Preprocess data at the gateway level…

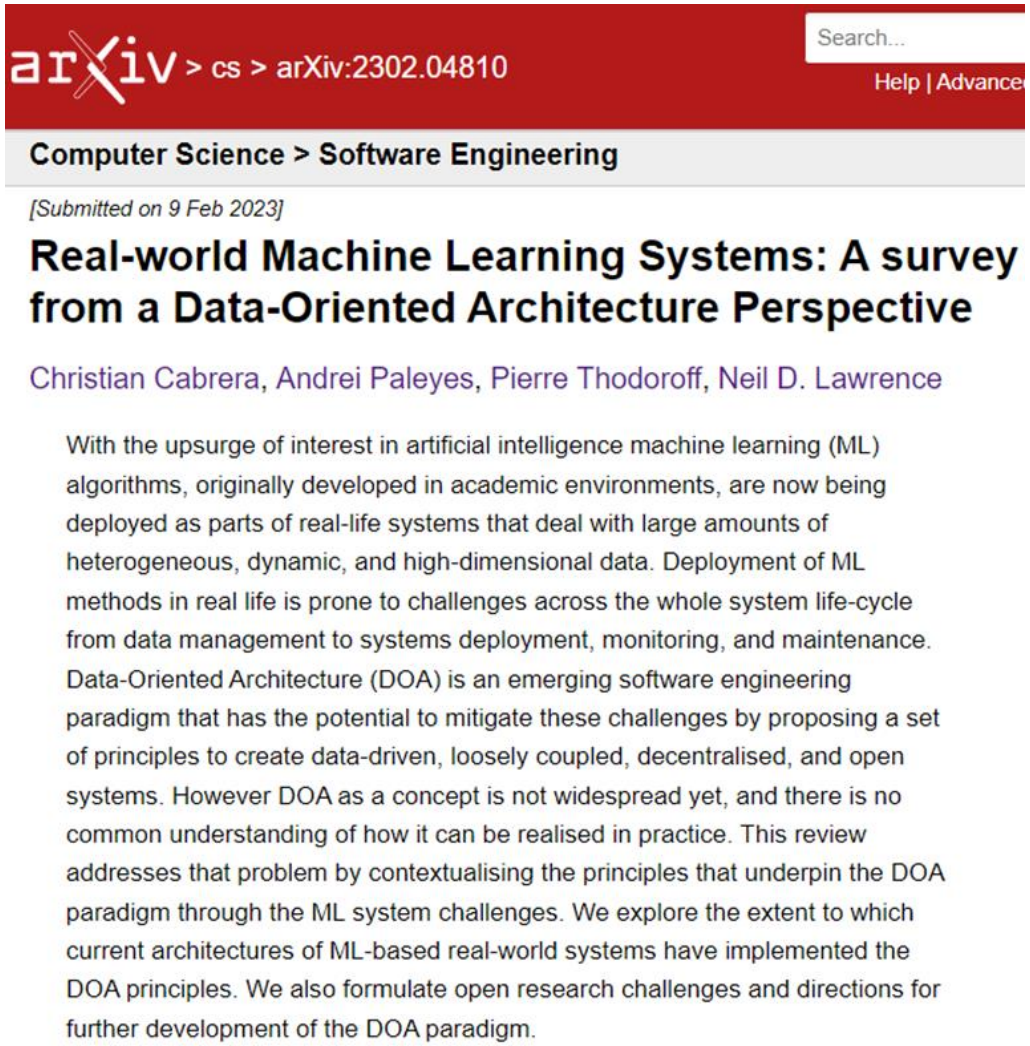
Move components from the cloud to the edge…

Openness

Water level monitoring project at DeKUT



ML-based Systems and DOA



arXiv > cs > arXiv:2302.04810

Search...

Help | Advance

Computer Science > Software Engineering

[Submitted on 9 Feb 2023]

Real-world Machine Learning Systems: A survey from a Data-Oriented Architecture Perspective

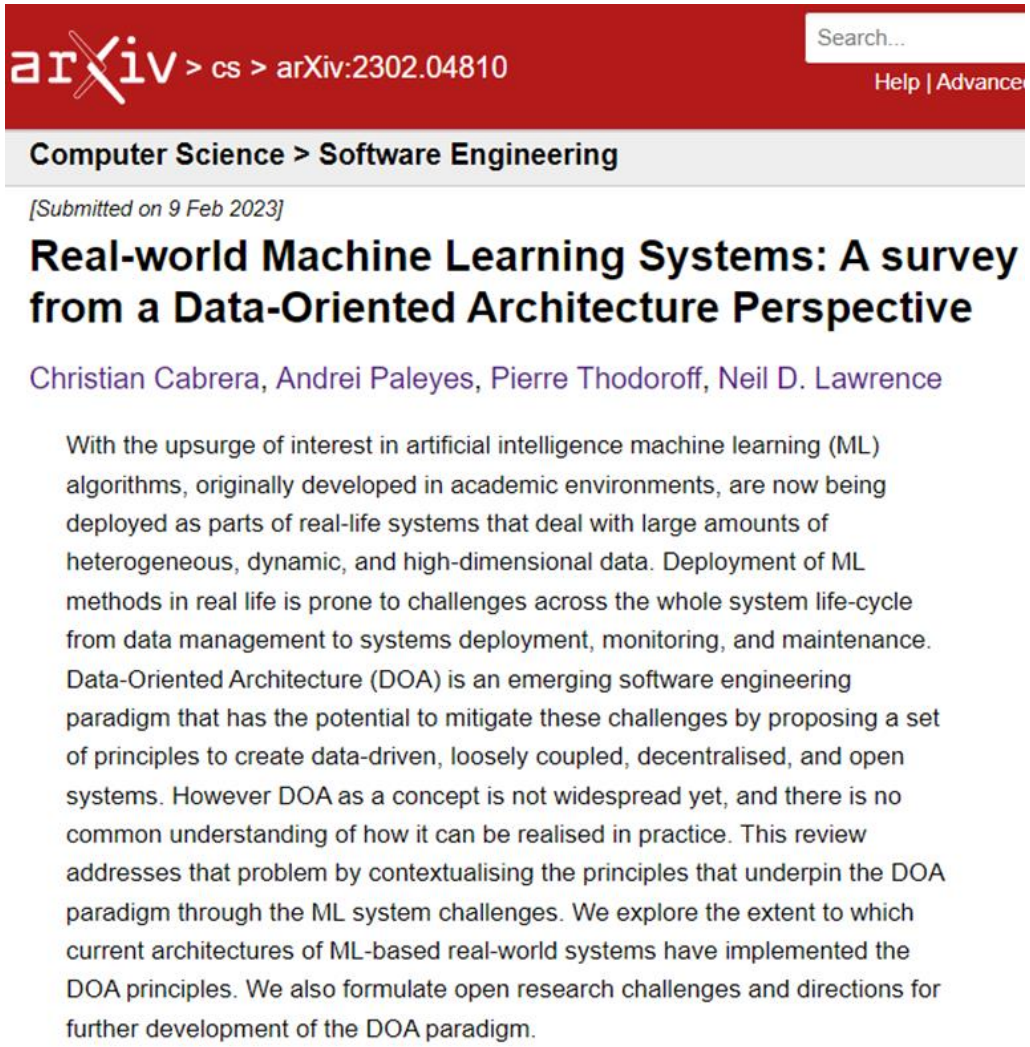
Christian Cabrera, Andrei Paleyes, Pierre Thodoroff, Neil D. Lawrence

With the upsurge of interest in artificial intelligence machine learning (ML) algorithms, originally developed in academic environments, are now being deployed as parts of real-life systems that deal with large amounts of heterogeneous, dynamic, and high-dimensional data. Deployment of ML methods in real life is prone to challenges across the whole system life-cycle from data management to systems deployment, monitoring, and maintenance. Data-Oriented Architecture (DOA) is an emerging software engineering paradigm that has the potential to mitigate these challenges by proposing a set of principles to create data-driven, loosely coupled, decentralised, and open systems. However DOA as a concept is not widespread yet, and there is no common understanding of how it can be realised in practice. This review addresses that problem by contextualising the principles that underpin the DOA paradigm through the ML system challenges. We explore the extent to which current architectures of ML-based real-world systems have implemented the DOA principles. We also formulate open research challenges and directions for further development of the DOA paradigm.

	Data as a First Class Citizen			Prioritise Decentralisation			Openness		
Research work	Data driven	Invariant and shared data mode	Data coupling	Local data chunks	Local first	Peer-to-peer first	Autonomous entities	Asynchronous entities	Message exchange protocol
Junchen et al. [60]	✓	✓	✓	✓	✓	✓	✓	✓	✓
Lebofsky et al. [72]	✓	✓	✓	✓	✓	✓	✓	✓	✓
Herrero et al. [59]	✓	✓	✓	✓	✓	✓	✓	✓	✓
Zhang et al. [125]	✓	✓	✓	✓	✓	✓	✓	✓	✓
Karageorgou et al. [66]	✓	–	✓	✓	✓		✓	✓	✓
Sultana et al. [116]	✓	✓	✓	–			✓	✓	✓
Calancea et al. [29]	✓	✓		–	–		–	–	–
Schumann et al. [106]	✓	✓	✓				✓	✓	✓
Alves et al. [9]	✓	✓	✓				✓	✓	✓
De Caro et al. [36]	✓	✓	✓				✓	✓	✓
Nguyen et al. [83]	✓	✓	✓				–	✓	✓
Xu et al. [123]	✓	✓		✓	✓	✓	✓		
Alonso et al. [8]	✓	–		✓	–		✓		✓
Sarabia-Jácome et al. [103]	✓	–		✓	✓		–		–
Santana et al. [102]	✓	✓		–	–		–		✓
Shih et al. [112]	✓	✓	–				–	–	–
Lu et al. [75]	✓	✓		–		–	–		–
Brumbaugh et al. [22]	✓	–	–	–	–	–			
Shan et al. [109]	✓	✓		✓	✓	✓			
Schubert et al. [105]	✓	✓	✓					✓	✓
Dai et al. [33]	✓	✓	✓	✓	✓				
Zhang et al. [126]	✓	✓		✓	–		✓		
Quintero et al. [99]	✓			✓	✓		–		–
Habibi et al. [56]	✓	✓	–					–	–
Gorkin et al. [55]	✓			–			–	–	–
Shi et al. [111]	✓			✓	✓			✓	
Franklin et al. [47]	✓						✓	✓	✓
Bayerl et al. [15]	✓	–		✓	✓				
Bellocchio et al. [16]	✓						–	✓	✓
Johny et al. [62]	✓				✓		–		–
Barachi et al. [14]	✓	✓		–			–		
Salhaoui et al. [101]	✓				–		–		–
Hegemier et al. [58]	✓				–		–		–
Cabanes et al. [23]	✓	✓	✓						
Agarwal et al. [2]	✓	✓	✓						
Müller et al. [81]	✓	✓	–						
Gao et al. [49]	✓	✓	–						
Amrollahi et al. [10]	✓	–	–						
Niu et al. [86]	✓			–			–		
Gallagher et al. [48]	✓			–			–		
Conroy et al. [31]	✓	✓					–		
Falcao et al. [43]	✓	✓							
Hawes et al. [57]	✓	✓							
Kemsaram et al. [67]	✓				✓				
Qiu et al. [98]	✓						–		
Ali et al. [7]	✓						–		

✓= Adopted, – = Partially adopted, = Not adopted

ML-based Systems and DOA



arXiv > cs > arXiv:2302.04810

Search...

Help | Advance

Computer Science > Software Engineering

[Submitted on 9 Feb 2023]

Real-world Machine Learning Systems: A survey from a Data-Oriented Architecture Perspective

Christian Cabrera, Andrei Paleyes, Pierre Thodoroff, Neil D. Lawrence

With the upsurge of interest in artificial intelligence machine learning (ML) algorithms, originally developed in academic environments, are now being deployed as parts of real-life systems that deal with large amounts of heterogeneous, dynamic, and high-dimensional data. Deployment of ML methods in real life is prone to challenges across the whole system life-cycle from data management to systems deployment, monitoring, and maintenance. Data-Oriented Architecture (DOA) is an emerging software engineering paradigm that has the potential to mitigate these challenges by proposing a set of principles to create data-driven, loosely coupled, decentralised, and open systems. However DOA as a concept is not widespread yet, and there is no common understanding of how it can be realised in practice. This review addresses that problem by contextualising the principles that underpin the DOA paradigm through the ML system challenges. We explore the extent to which current architectures of ML-based real-world systems have implemented the DOA principles. We also formulate open research challenges and directions for further development of the DOA paradigm.

Few projects fully follow DOA principles.

Most of the solutions are centralized and cloud based.

Databases, streams and message queues enable the data first principle.

Distributed storage and computing technologies for decentralisation.

Asynchronous communication for openness.

Thank you!

chc79@cam.ac.uk