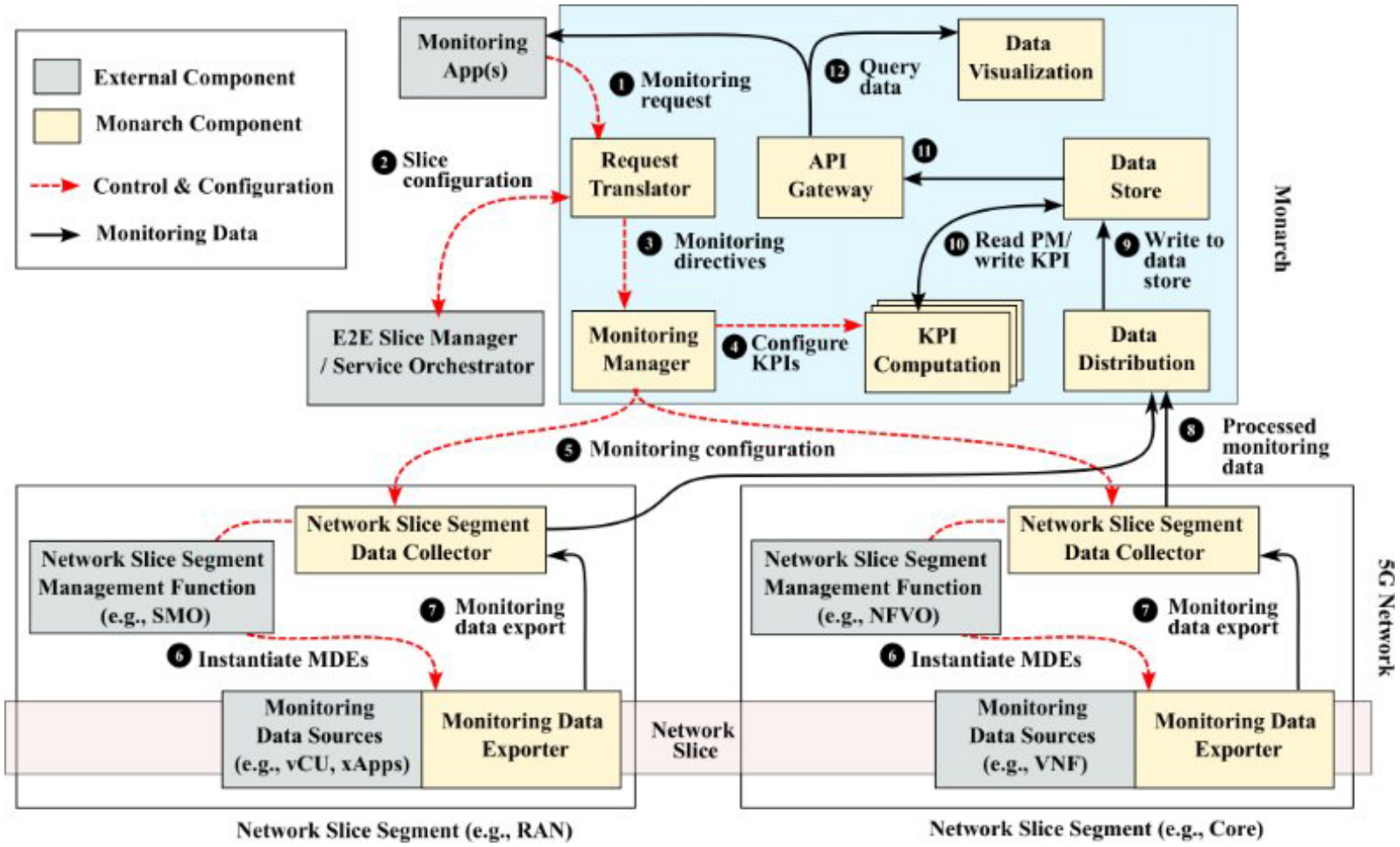
Workshop 1: 5G Core Slicing

Scalable Data Ingestion, Processing, and Visualization

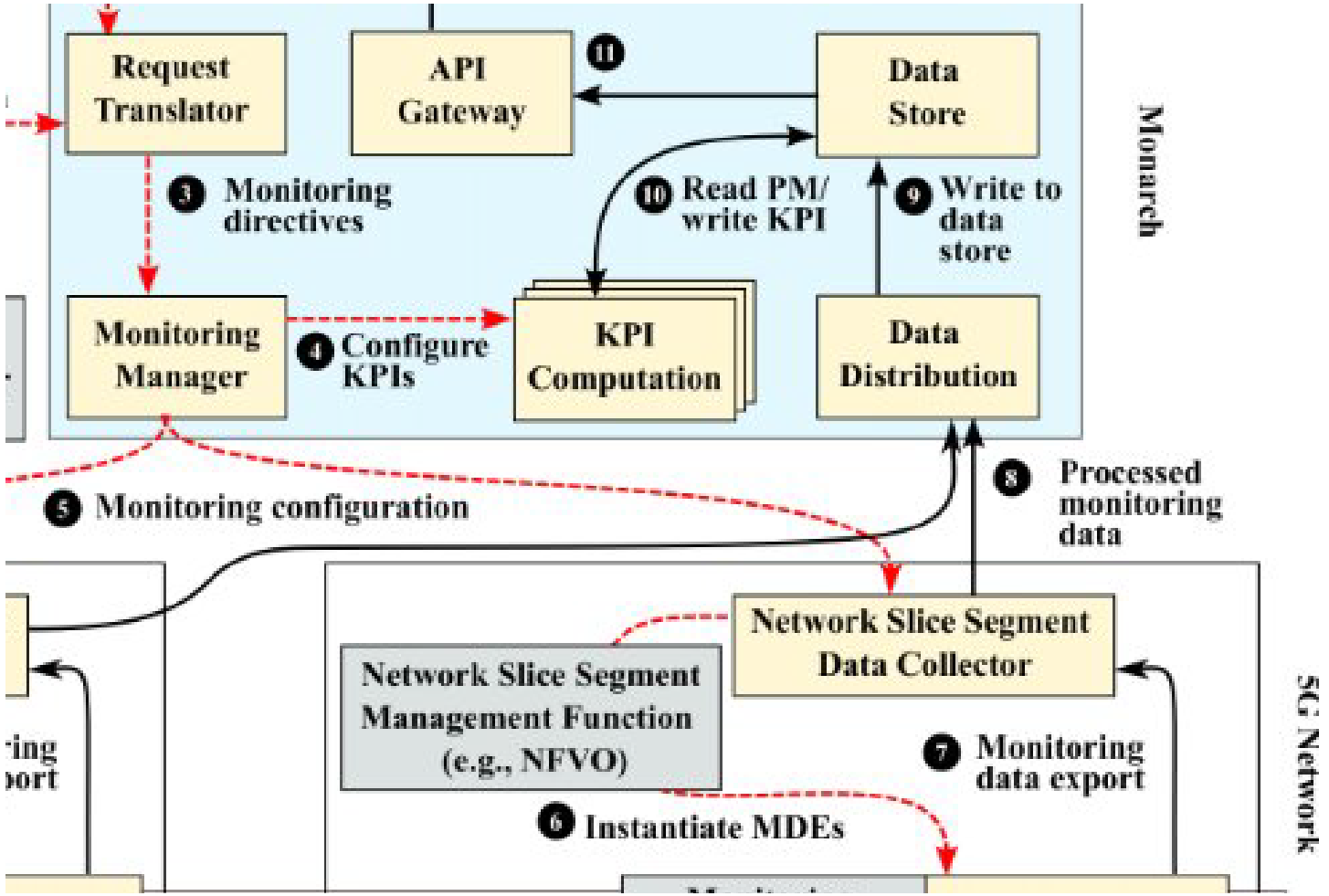
Raouf Boutaba
David R. Cheriton School of Computer Science
University of Waterloo

Monarch Architecture



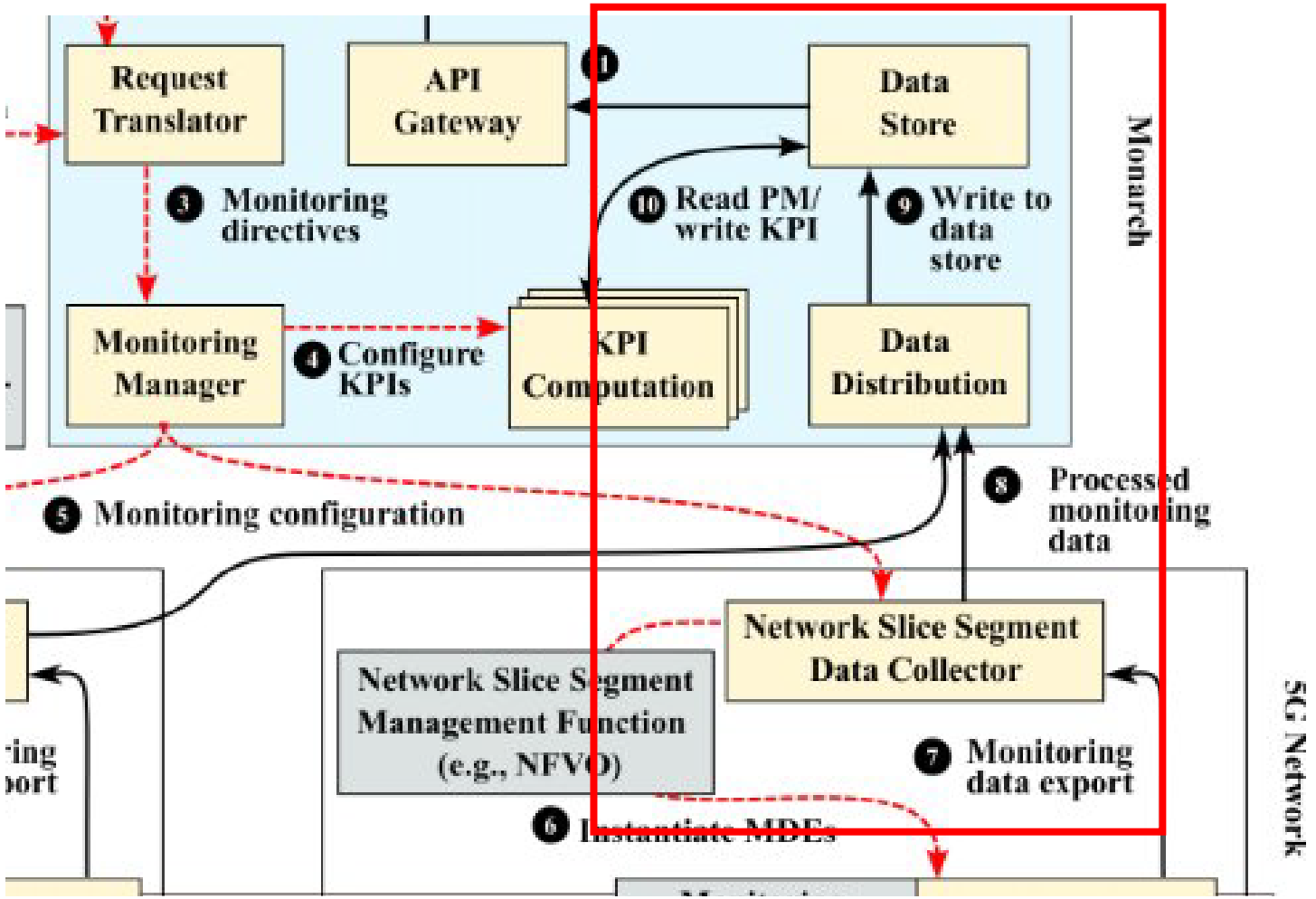
UNIVERSITY OF WATERLOO

Monarch Architecture



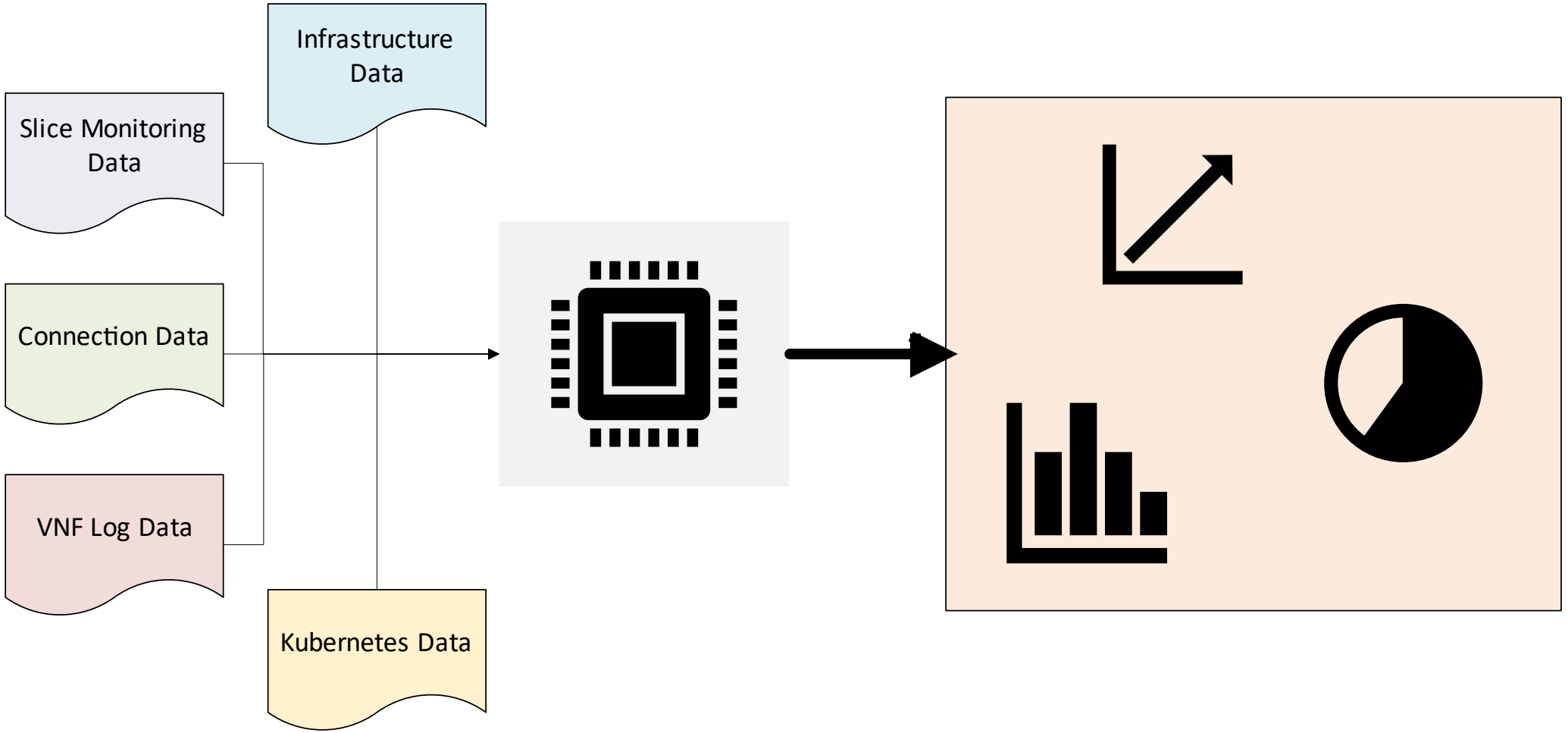
UNIVERSITY OF WATERLOO

Monarch Architecture



UNIVERSITY OF WATERLOO

Data Pipeline Architecture



Designing a Data Pipeline

- Three key questions:
  - What do I want to know?
    - Resource allocation
    - Error monitoring
    - Security
  - What data is necessary to know?
  - How do I obtain the necessary data?

Designing a Data Pipeline

- Three key questions:
  - What do I want to know?
  - What data is necessary to know?
    - Collecting resource usage
    - Server up/down status
    - Network information
  - How do I obtain the necessary data?

Designing a Data Pipeline

- Three key questions:
  - What do I want to know?
  - What data is necessary to know?
  - How do I obtain the necessary data?
    - Deploying collection agents
    - Data transformations
    - Data storage

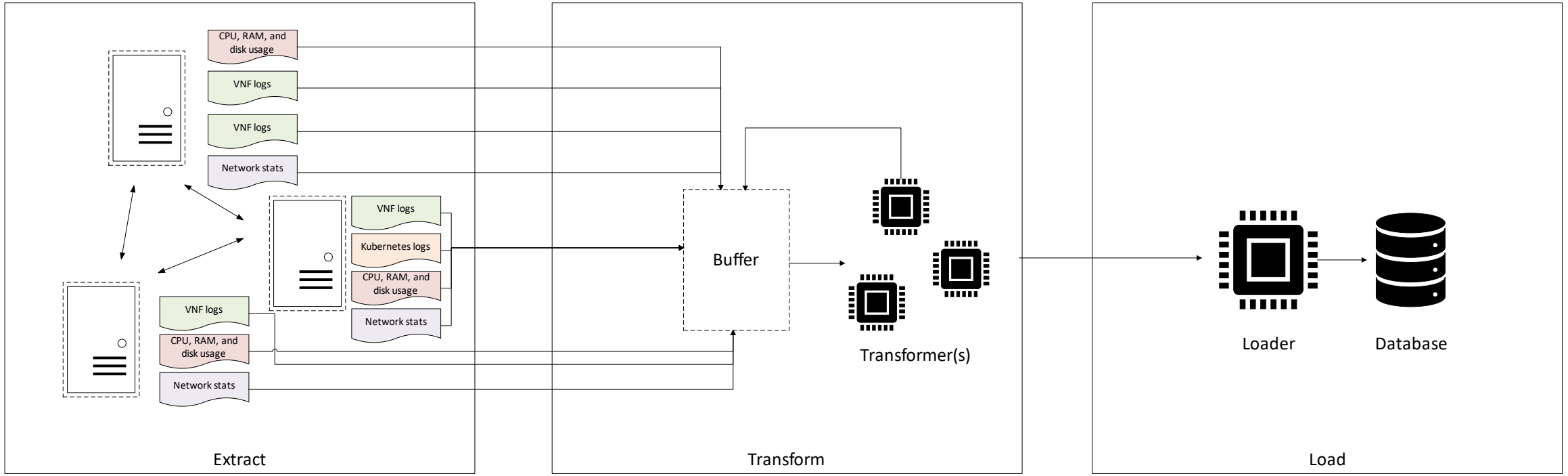
Session Outline

- Introduce the core components of a data pipeline
- Present commonly-used software in a data pipeline
- Discuss challenges of implementing and maintaining a data pipeline in practice
- Hands-on labs:
  - Show how a data pipeline in the 5G core network can be integrated
  - Utilize the data stored by the data pipeline to build dashboards

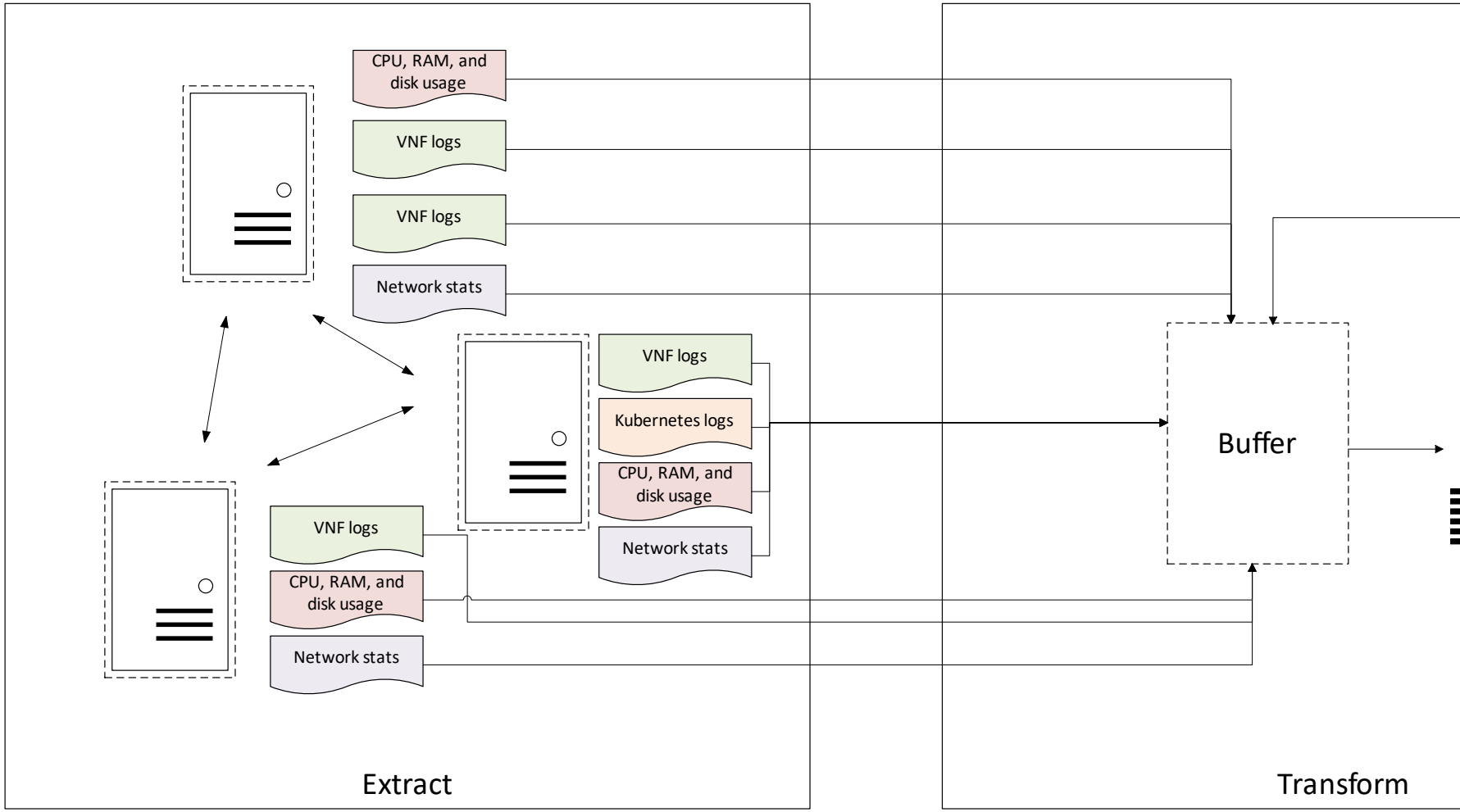
DATA PIPELINE COMPONENTS

What are the key components of a data pipeline?

- Extract, (buffer), transform, load (ETL)

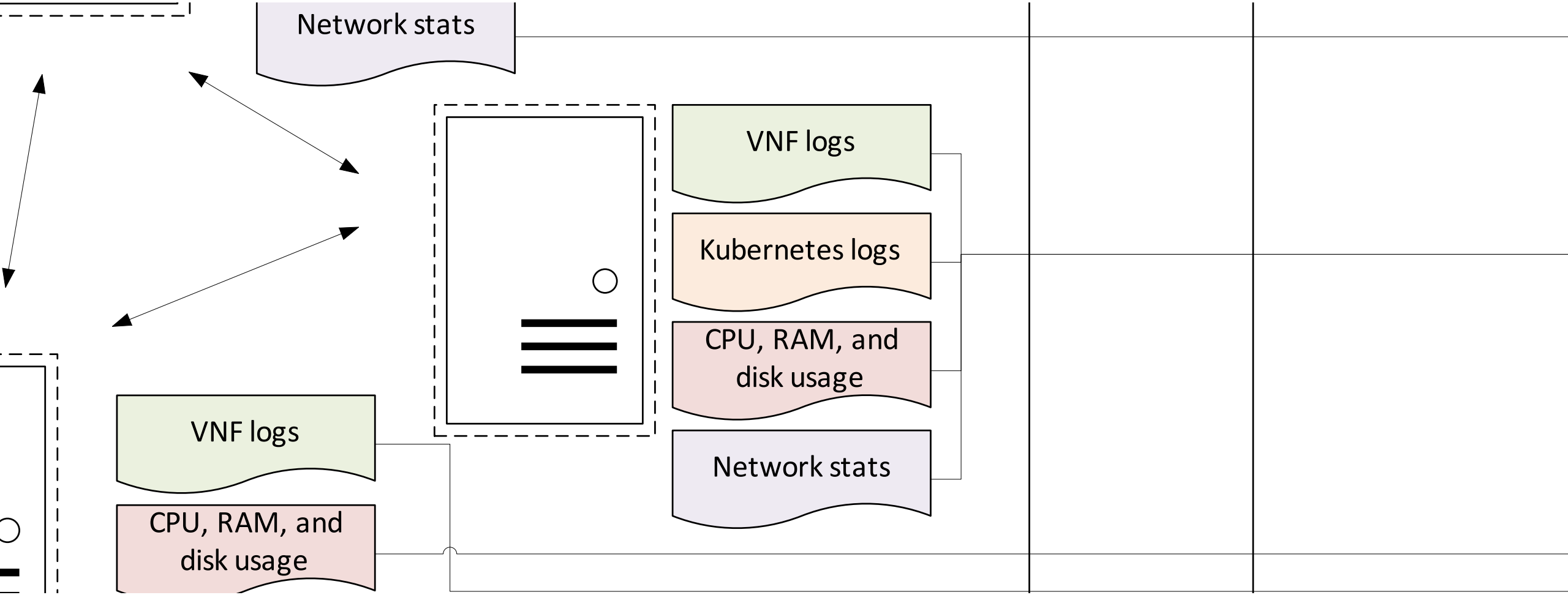


What are the key components of a data pipeline?

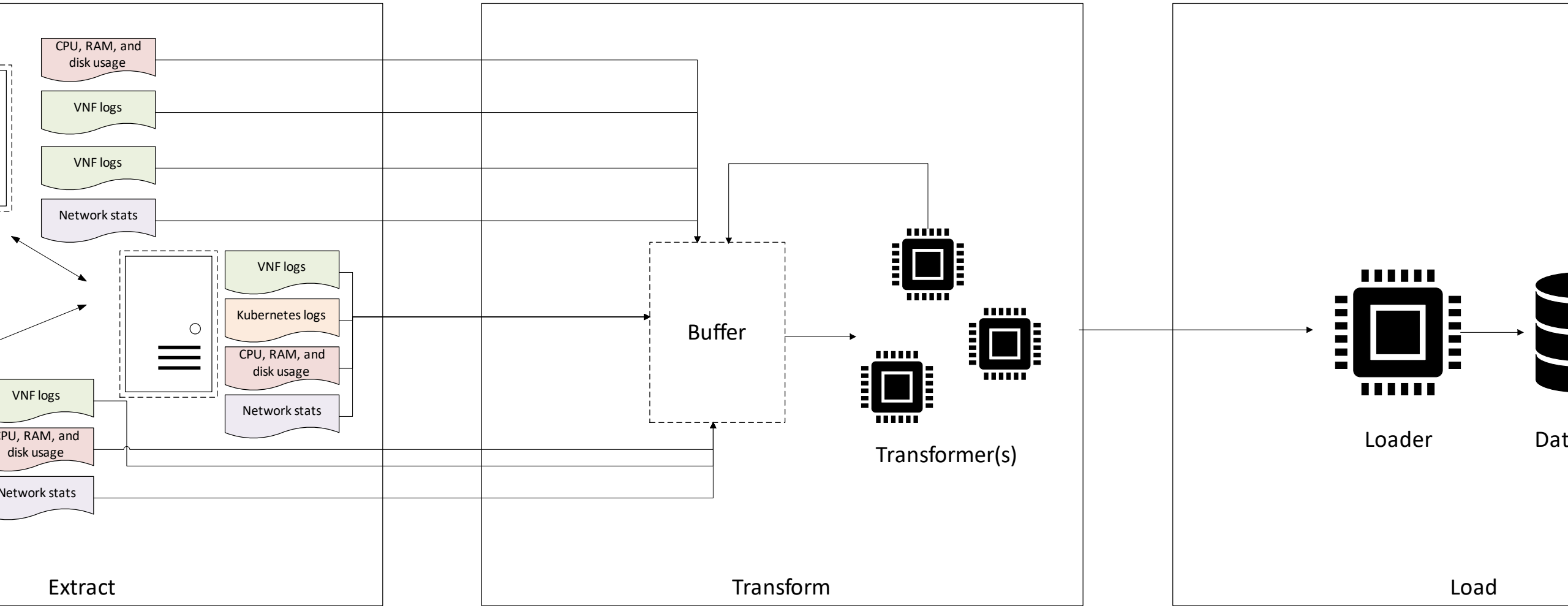


UNIVERSITY OF WATERLOO

What are the key components of a data pipeline?

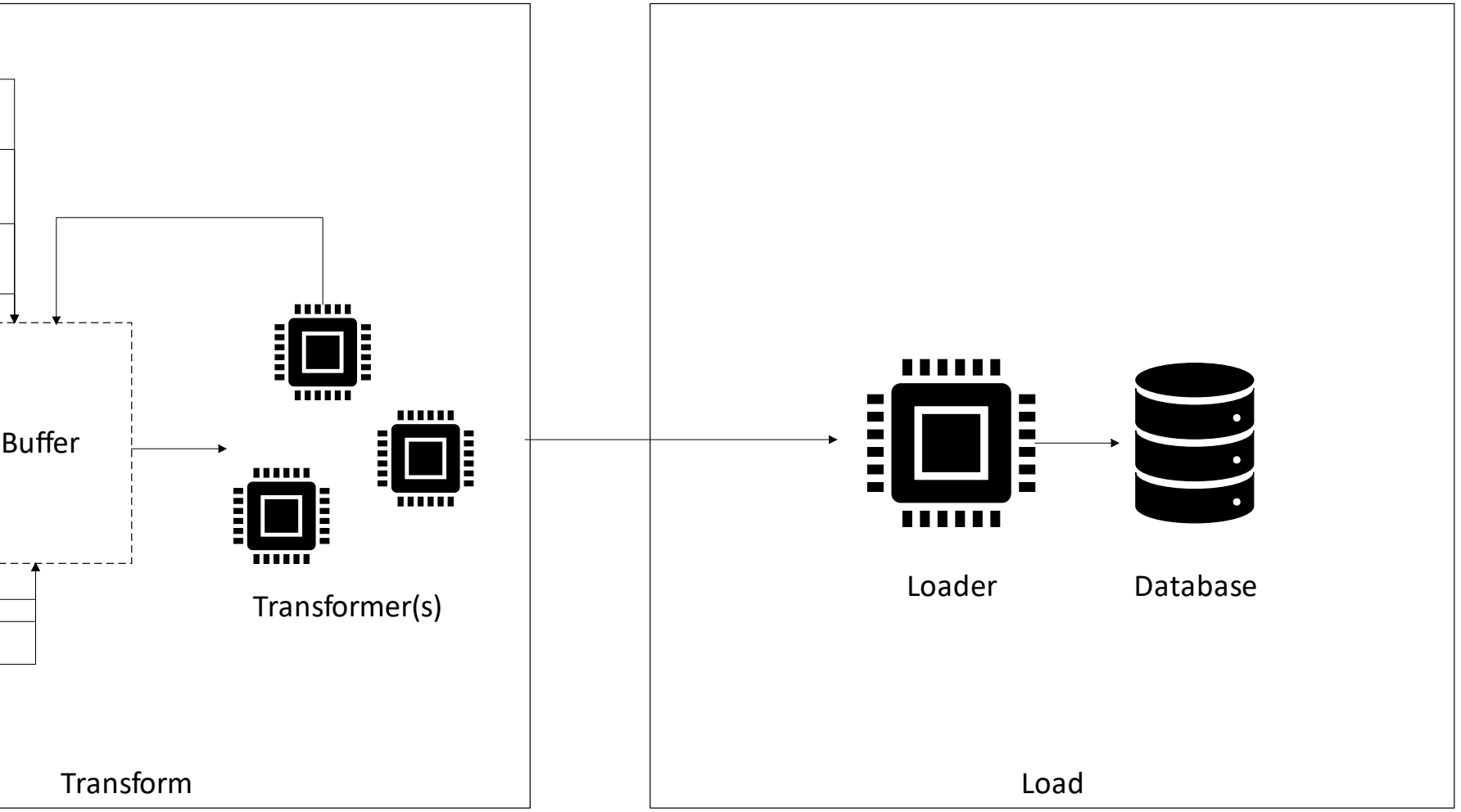


What are the key components of a data pipeline?



UNIVERSITY OF WATERLOO

What are the key components of a data pipeline?



COMMON SOFTWARE

Extraction

- Generally, the type of extraction software depends on:
  - Data you want to ingest,
  - Output format,
  - Resource requirements
- Common extraction-type utilities:
  - Syslog
  - Fluent bit
  - Beats

Buffer

- Depending on the use-case, buffers may not be necessary
- Buffer-type utilities:
  - Apache Kafka
  - RabbitMQ

Transformation

- Transformation-type utilities:
  - Logstash
  - Fluentd
  - Apache NiFi
  - Data computation programs, e.g., Apache Spark, Ray

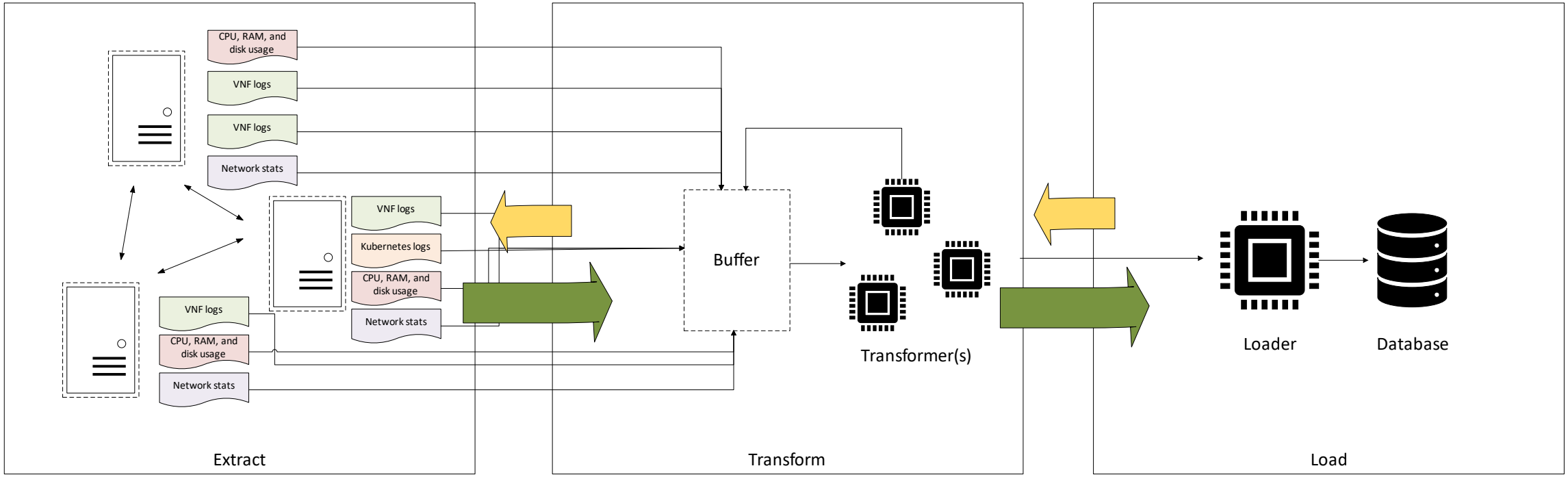
Loading

- Loading-type utilities:
  - Usually integrated with the transformation software
  - Examples include:
    - Logstash
    - Fluentd
    - Apache NiFi

DATA PIPELINE CHALLENGES

What are the challenges of a data pipeline?

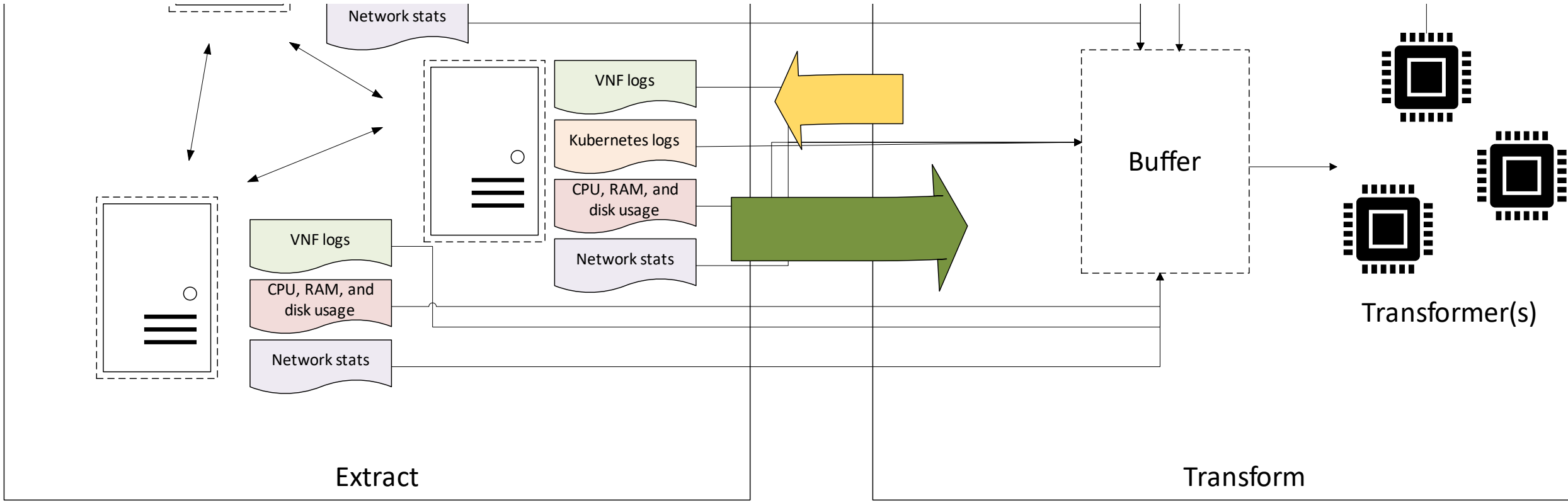
- Backpressure!



UNIVERSITY OF WATERLOO

What are the challenges of a data pipeline?

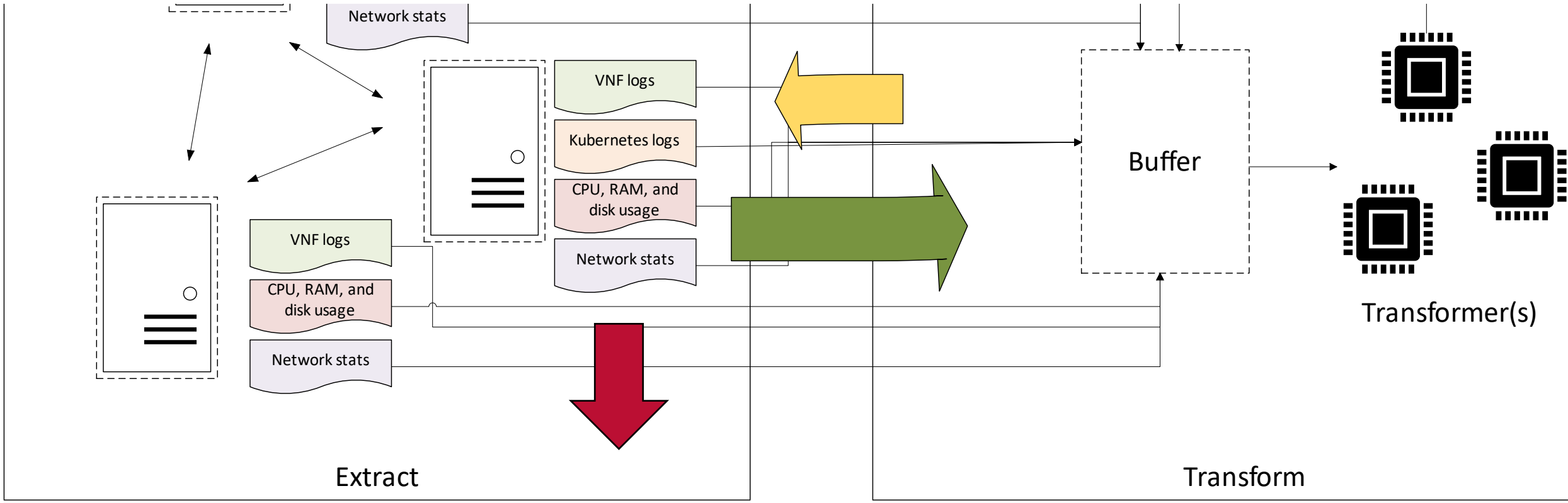
- Backpressure!



UNIVERSITY OF WATERLOO

What are the challenges of a data pipeline?

- Backpressure!

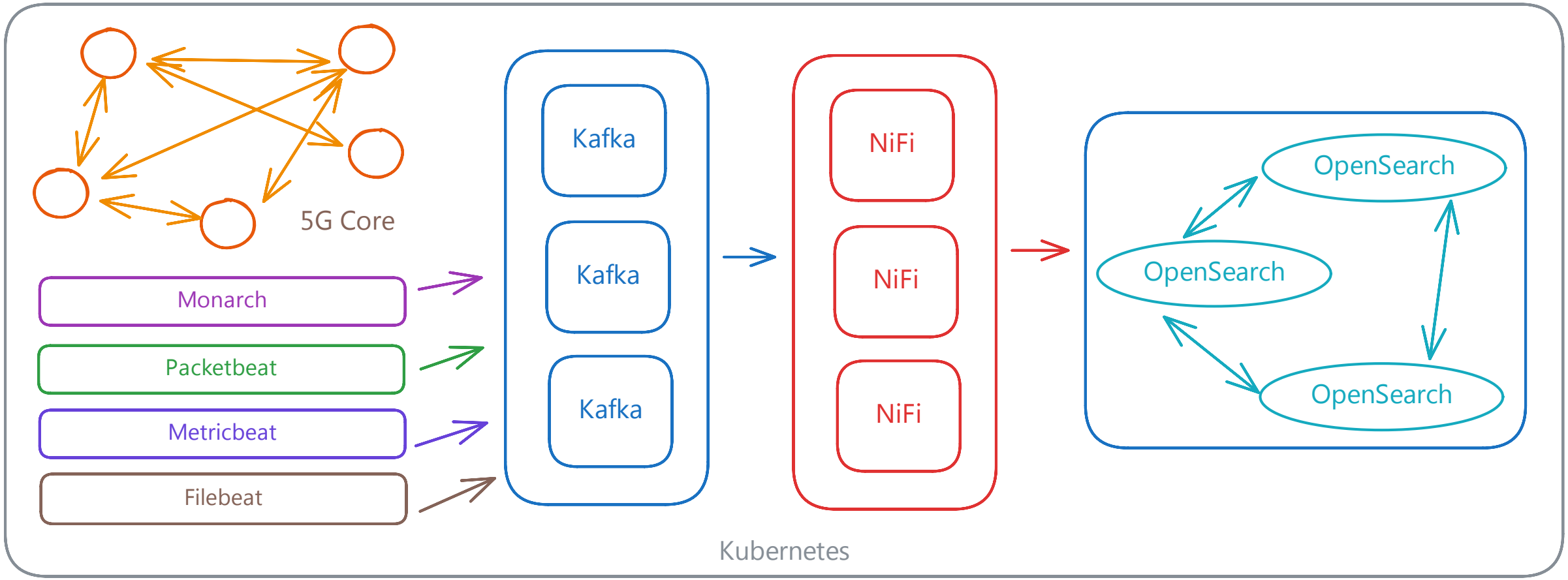


UNIVERSITY OF WATERLOO

What are the challenges of a data pipeline?

- Changing data formats
  - Examples:
    - New types of data
    - Software updates
    - Configuration file changes
  - These cause subtle incompatibilities that may not be noticed until it is too late to fix
  - Checking data schema in the pipeline itself may be a good idea
    - Dead letter queues

Hands-on Lab



UNIVERSITY OF WATERLOO