

MARKET NEED

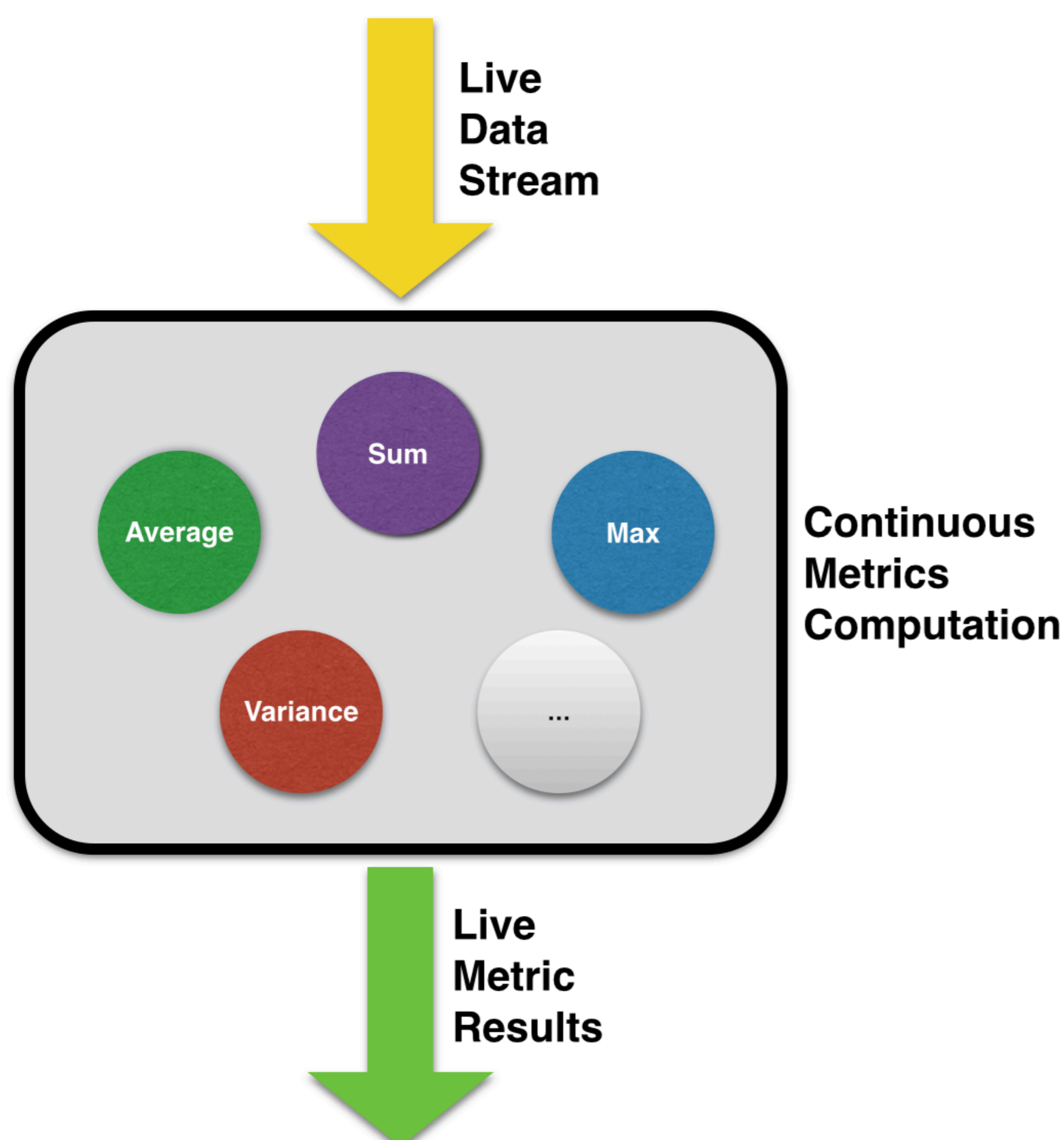
Computing metrics in real-time over live streaming data sources is required in many industries.

- Financial services companies compute metrics within a sliding window for analysing stock prices, trading volumes and fraud detection.
- Energy suppliers analyse live power usage and grid statistics. Telecommunications companies analyse real-time network usage and health statistics.
- Transportation and logistics agencies provide real-time traffic information to drivers based on live metrics computed over traffic flows at key road intersections.

In many scenarios these metrics must often be computed as fast as possible (low latency) over high volume, high velocity data streams (high throughput).

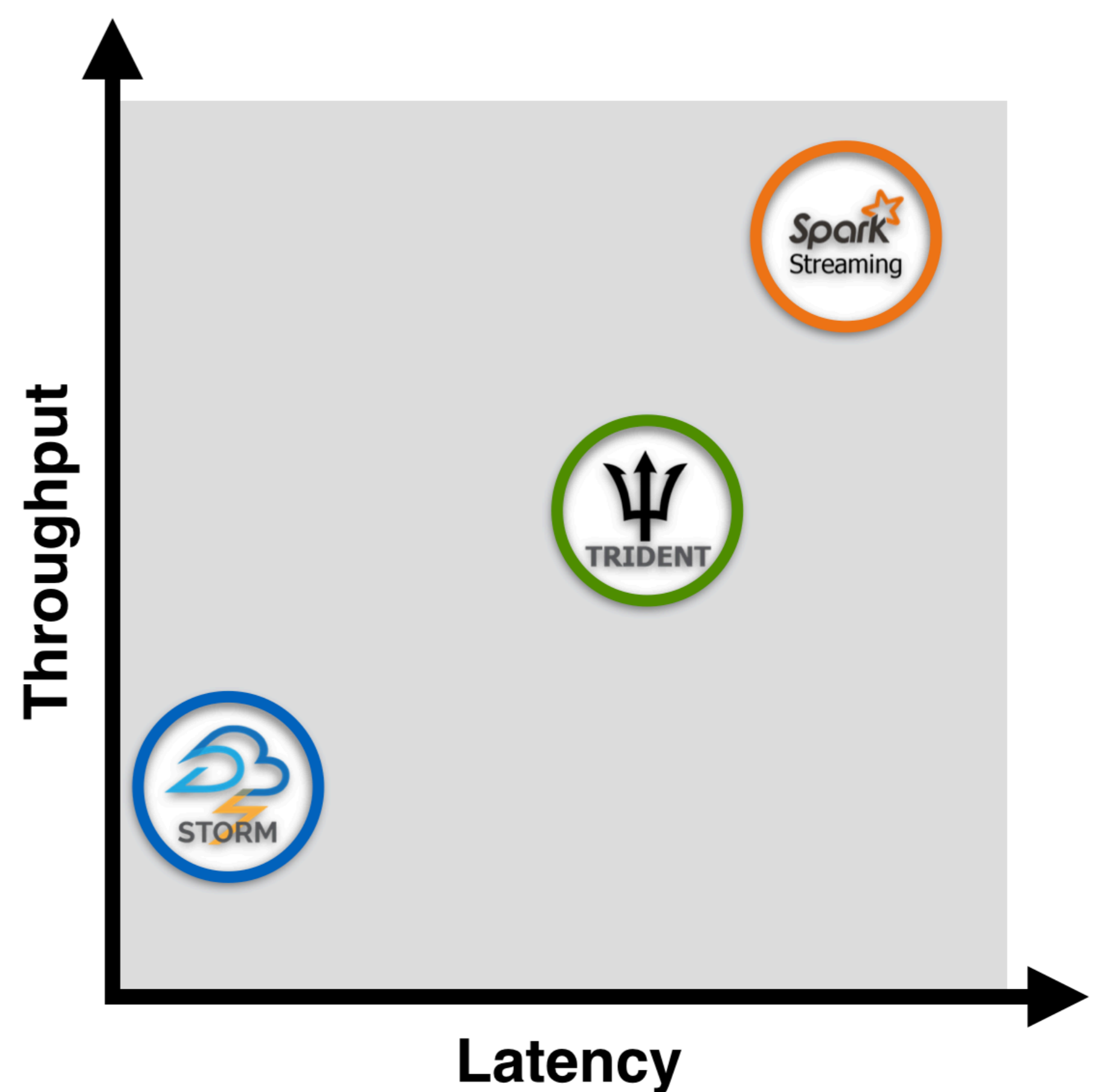
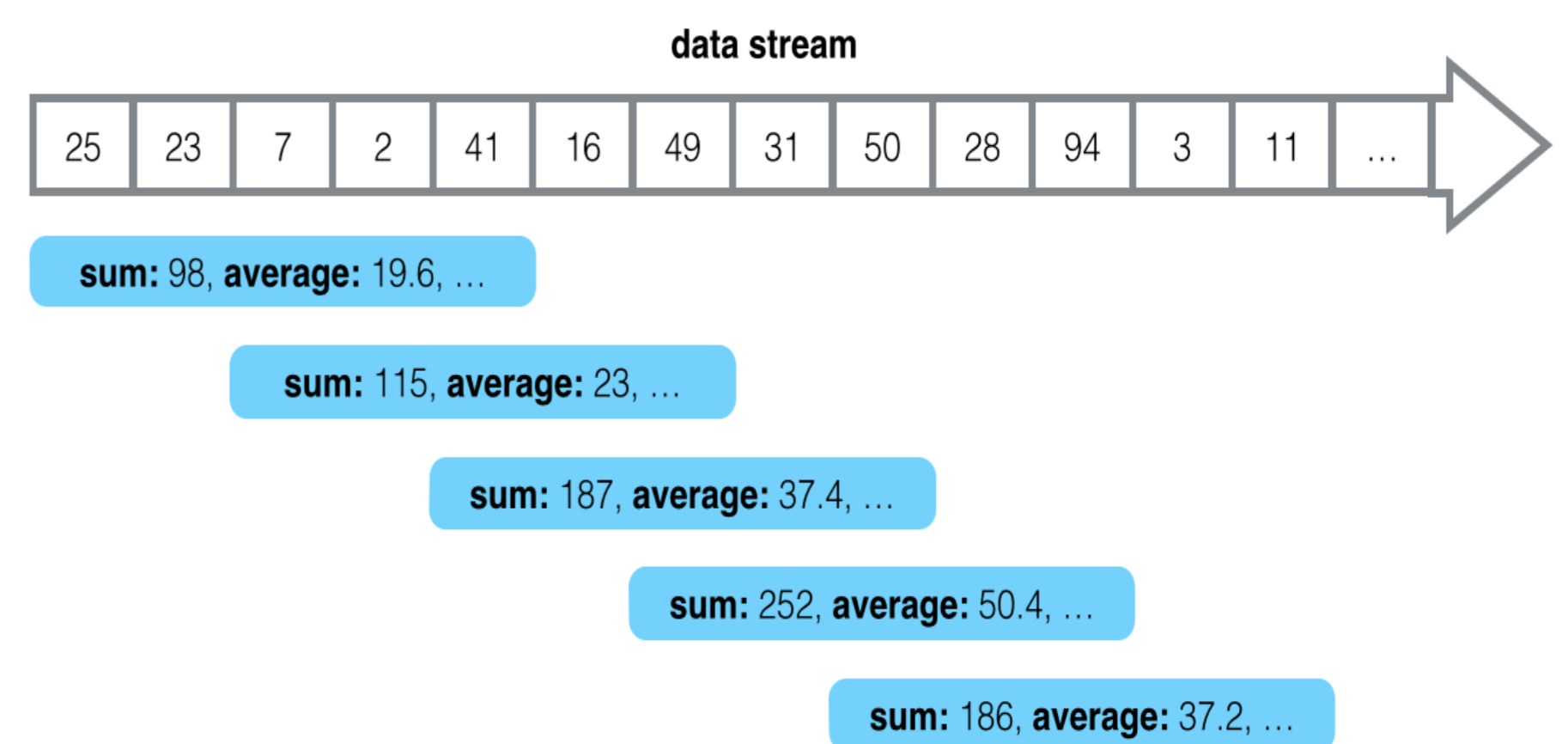
TECHNOLOGY SOLUTION

A number of popular parallel distributed platforms for data stream processing exist, such as **Storm**, **Storm Trident** and **Spark Streaming**. This project evaluates and compares these different platforms for the task of continuous computation of live statistical metrics over streaming data. Prototype implementations were developed and these were tested under different conditions, datasets and hardware/system configurations.



APPLICABILITY

The stream processing platforms that we evaluated in this project suit different real-time or near real-time metric computation scenarios. This evaluation provides insight into which platforms suit particular needs as well as the trade-off between throughput and latency.



RESEARCH TEAM

- Dr. Guangyu Wu, UCD
- Dr. Oisin Boydell, UCD
- Hodei Iraola, UCD
- Dr. Brian MacNamee, UCD