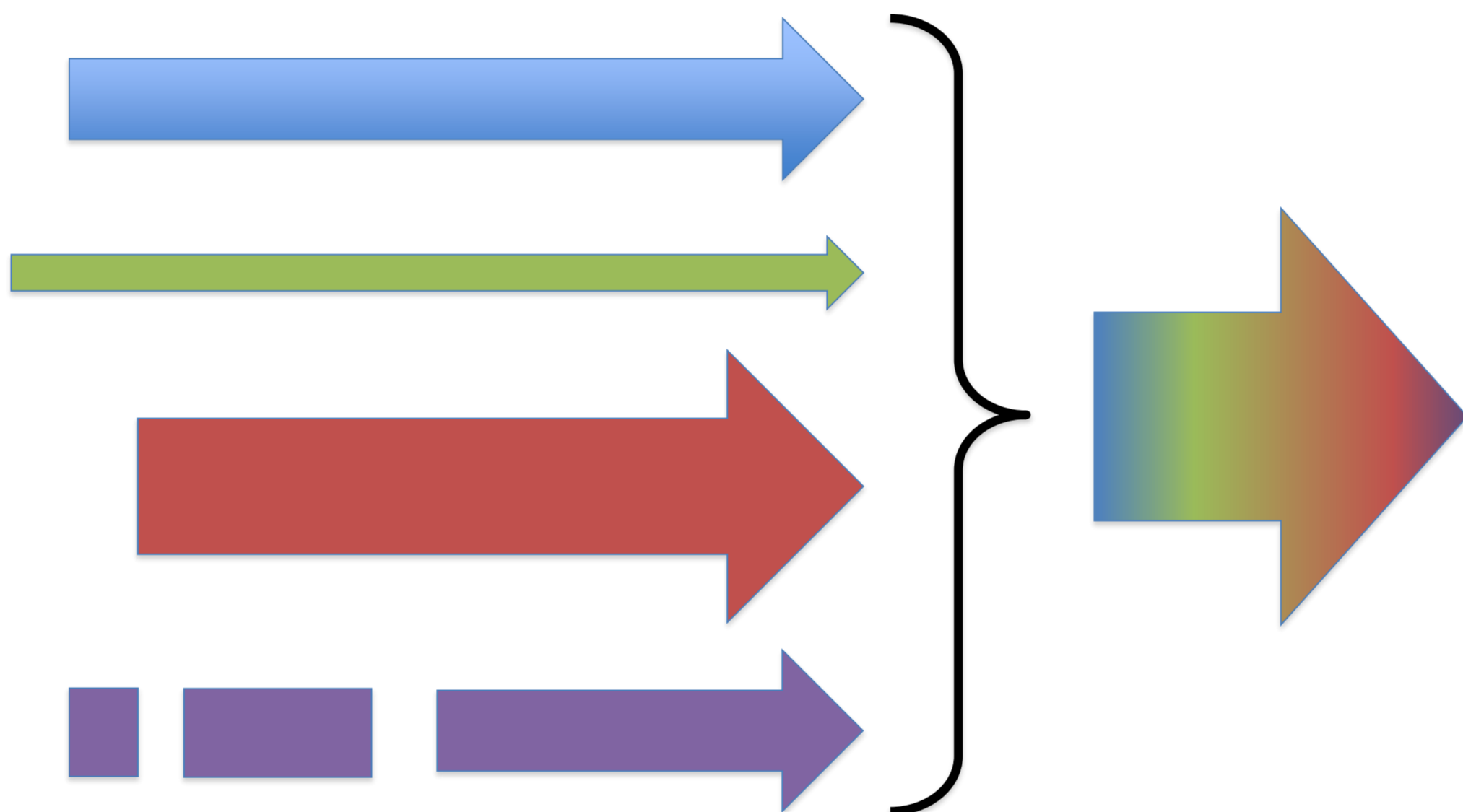


# StreamConverge: Spark Streaming over Heterogeneous Datasets

## MARKET NEED

- Complex systems often include many different streaming data formats produced by different system outputs, sensors and sub-systems.
- Often there is a need to integrate these heterogeneous data streams in real-time in an efficient, scalable and fault tolerant way.



## TECHNOLOGY SOLUTION

We have implemented a heterogeneous streaming data integration demonstrator on **Apache Spark Streaming** which is:

- Tolerant to out-of-order event arrival
- Merges and integrates events by key
- Scalable and efficient for high velocity, high volume data

## THE FUTURE OF THIS TECHNOLOGY SOLUTION

- The technology can be used across many different scenarios where heterogeneous streaming data is present such as
  - Integration of back end streaming machine generated data such as system logs, sensor data, telemetry data
  - Customer data integration such as social media streams, customer touch points, website interactions etc.



## RESEARCH TEAM

Dr Guangyu Wu, UCD

Dr Oisín Boydell, UCD

Dr Brian MacNamee, UCD