

Reservoir Sampling

- ▶ Load shedding (discarding observations) can reduce processing requirements, but introduce bias
- ▶ Reservoir sampling reduces load shed bias
- ▶ The first n stream elements comprise the original reservoir
- ▶ Subsequent elements are candidates based on an evaluation of their location (temporal index) in the stream
- ▶ Modified approach increases likelihood that more recently added observations are in the reservoir; new data has greater salience than older observations
- ▶ Downside: requires computation

