Scaling Security Operations Center message processing

From Scala to Go
Spark Streaming - not “real” streaming

*discretized stream processing*

records → Receiver → batches (RDDs) → batches processed with tasks → Spark → IDLE TIME

Long-running DNS queries

records processed in batches with short tasks
each batch is a RDD (partitioned dataset)
Complexity for mitigation

- 1st layer cache
- 2nd layer cache (Redis)
- Backup job to pick up slack

Figure 2: DNS data enrichment flowchart
Moving parts

Go impl.

Cache

Spark impl.
Enter Go

```go
func (orig_value []byte) {
    keep, value := p.process.Process(orig_value)
    if keep {
        processed := p.producer.Produce(value)
    }
}(msg.Value)
```
Easy scale-out
Conclusions / stats

Our DNS servers are warning that this host has been sending an UNACCEPTABLE rate of queries for one hour (876 requests/sec).
Extra juicy bit