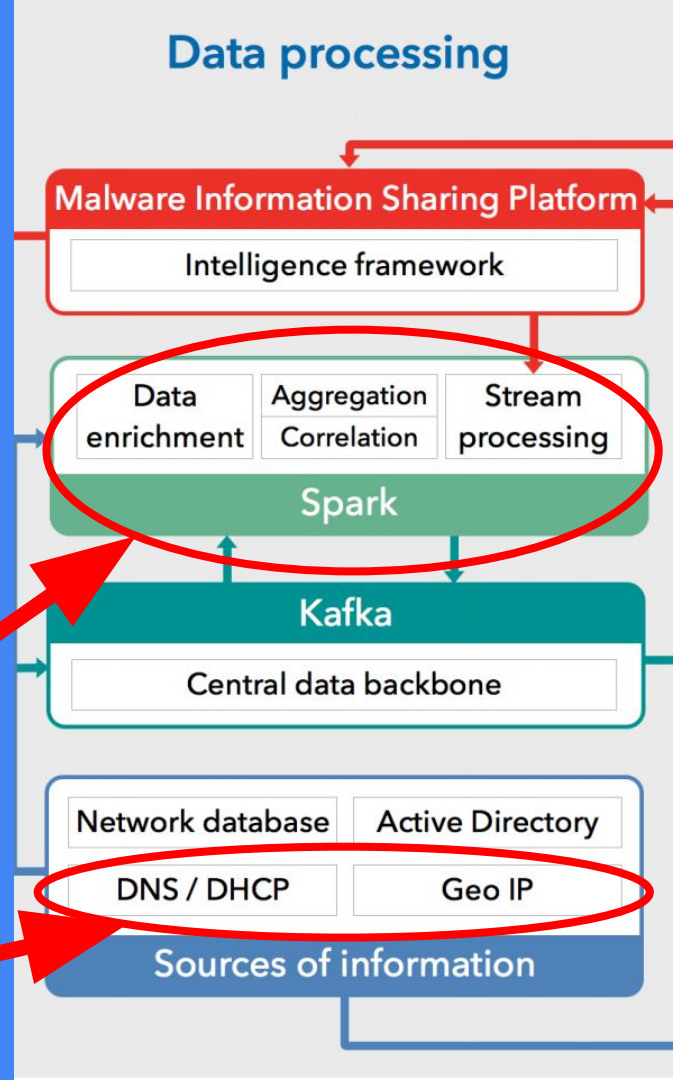


Untitled presentation



A story of data streaming and async I/O

From Scala to Go



Task: Data Enrichment

```
{  
  "proto": "tcp",  
  "service": "ssl",  
  "duration": 4.325414,  
  "timestamp": 1533888289171,  
  "srcip": "2a02:aa12:1500:3580:e575:1b16:a8c4:c800",  
  "dstip": "2001:1458:201:66::100:14",  
}
```

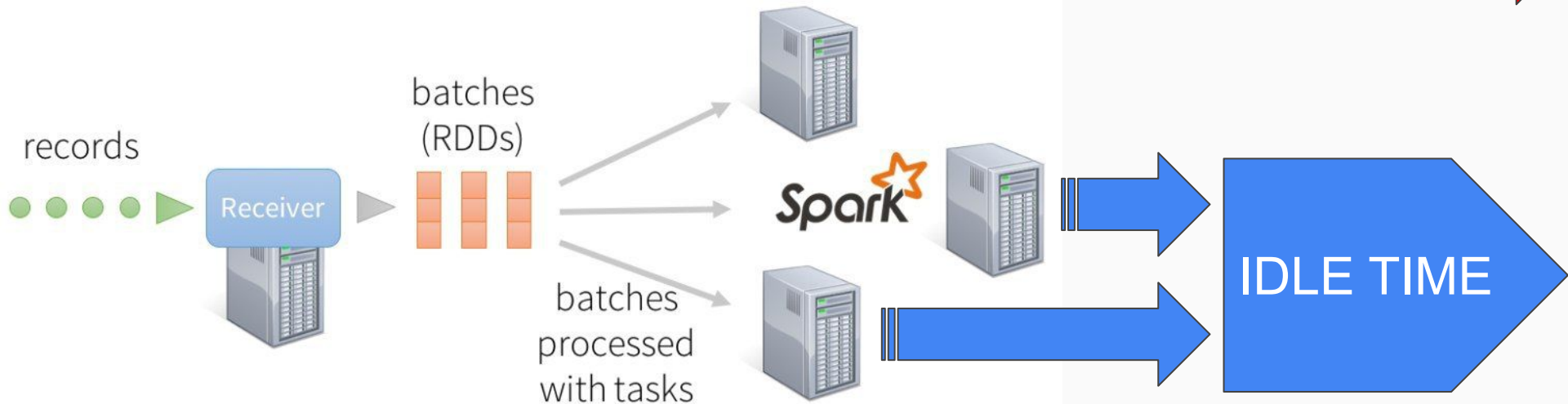


```
{  
  "proto": "tcp",  
  "service": "ssl",  
  "duration": 4.325414,  
  "timestamp": 1533888289171,  
  "srcip": "2a02:aa12:1500:3580:e575:1b16:a8c4:c800",  
  "dstip": "2001:1458:201:66::100:14",  
  "srcip_country": "Switzerland",  
  "srcip_org": "UPC Schweiz GmbH",  
}
```

Spark Streaming - not “real” streaming

Spark Streaming
discretized stream processing

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

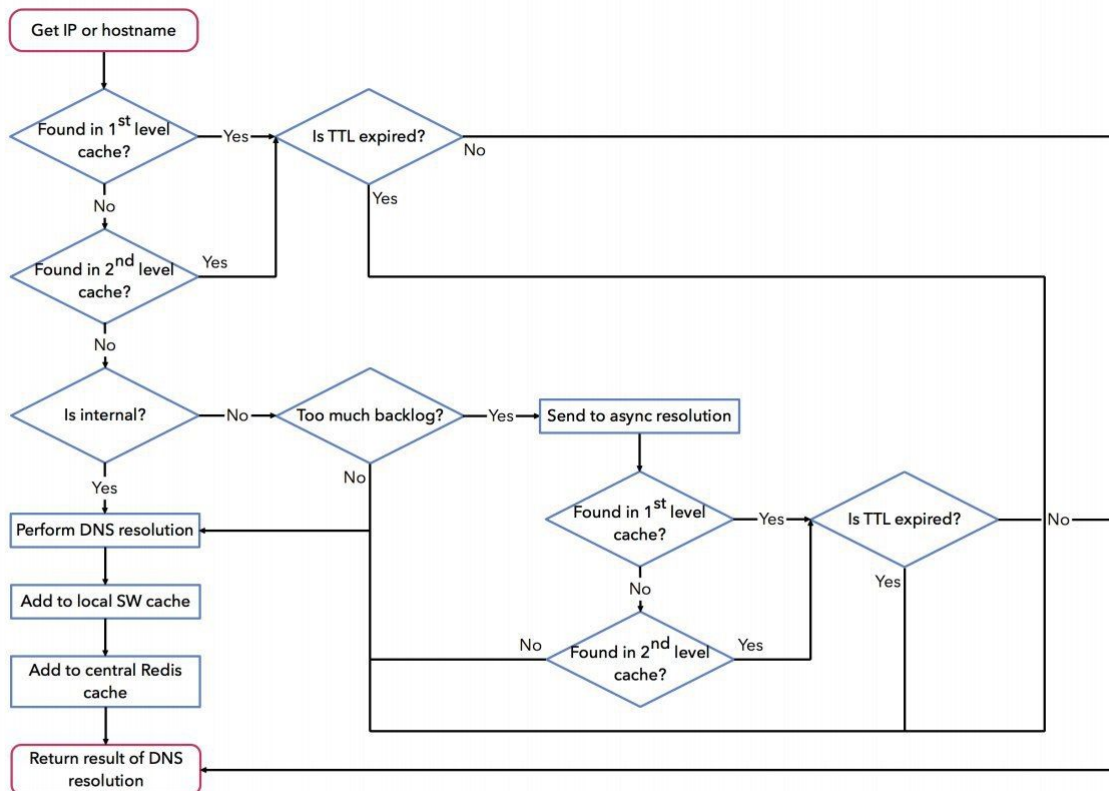
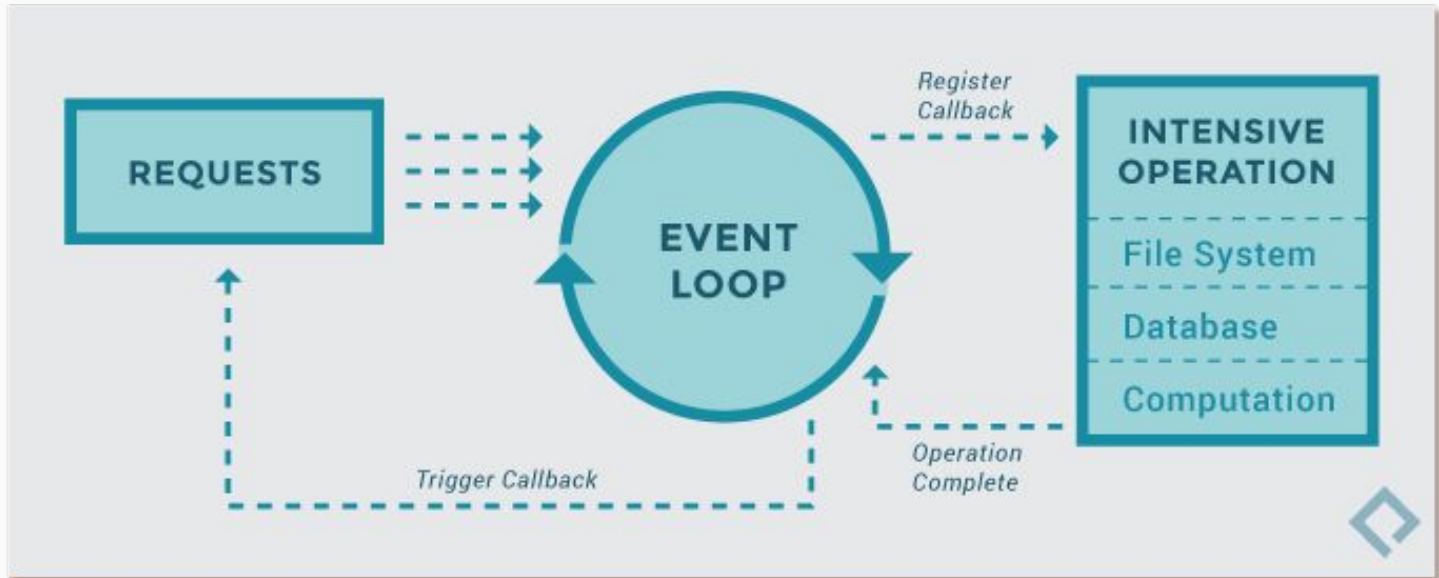


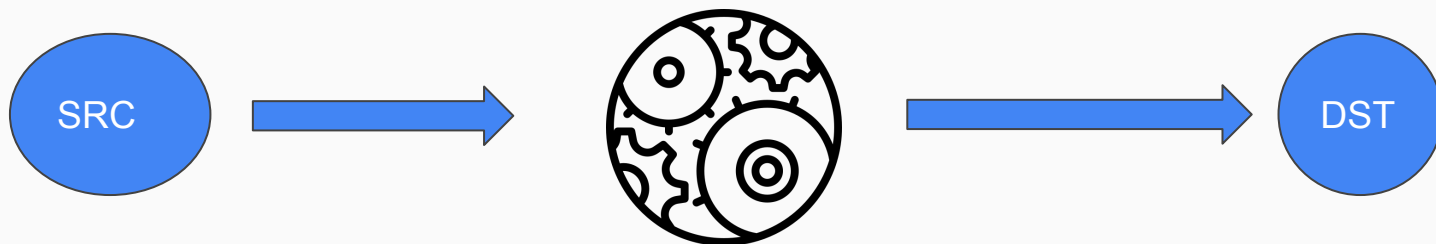
Figure 2: DNS data enrichment flowchart

Advantages of being asynchronous



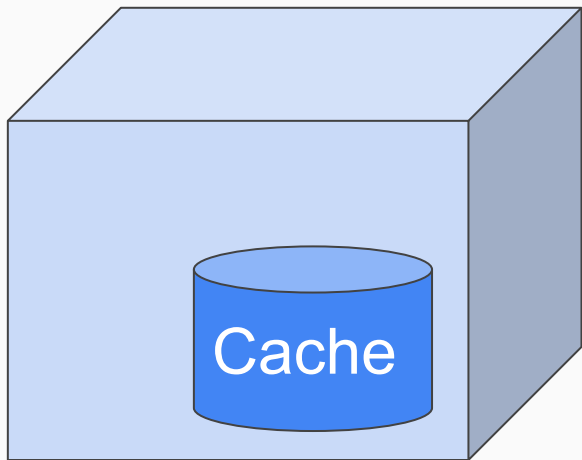
Enter Go

```
go func(orig_value []byte) {  
    keep, value := p.process.Process(orig_value)  
    if keep {  
        processed := p.producer.Produce(value)  
    }  
}(msg.Value)
```

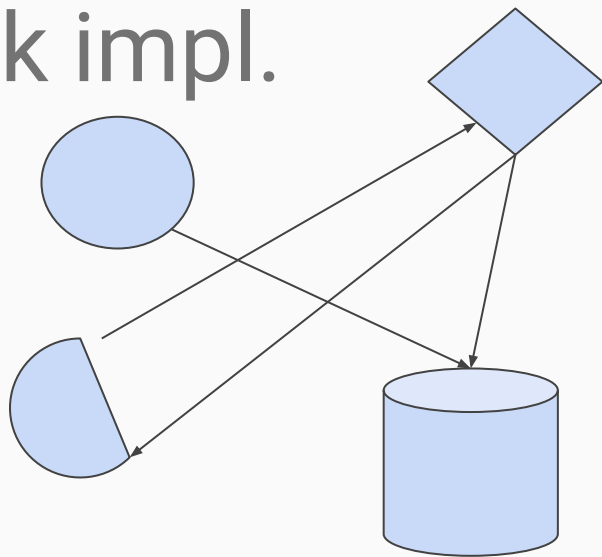


Moving parts

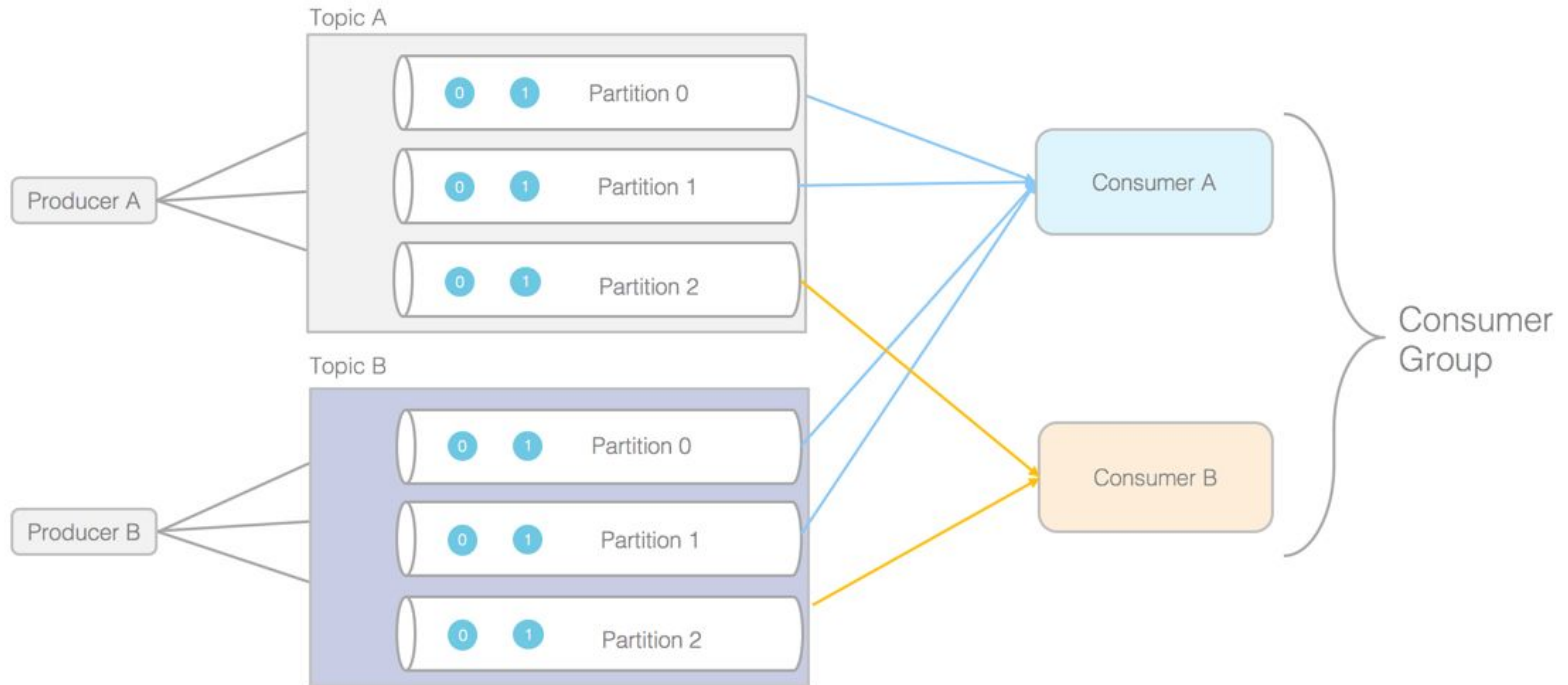
Go impl.



Spark impl.



Easy scale-out



Conclusions / stats

```
2019/04/12 11:57:54 Acker for bro_ssh:0 processed 323 entries in 32
2019/04/12 11:57:54 Acker for bro_conn:1 processed 59845 entries in
2019/04/12 11:57:54 Acker for csl_netlog:0 processed 418845 entries
2019/04/12 11:57:54 Acker for bro_dpd:0 processed 1979 entries in 395
```

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

Timestamp diff (all) ▾

2019-05-10 06:16:00

bro_conn bro_conn_enr:	1 s
bro_dpd bro_dpd_enr:	1 s
bro_ftp bro_ftp_enr:	9 s
bro_http bro_http_enr:	0 ns
bro_irc bro_irc_enr:	0 ns
bro_kerberos bro_kerberos_enr:	1 s
bro_known_certs bro_known_certs_enr:	0 ns
bro_known_hosts bro_known_hosts_enr:	1 s
bro_known_services bro_known_services_enr:	0 ns
bro_radius bro_radius_enr:	0 ns
bro_rdp bro_rdp_enr:	0 ns
bro_sip bro_sip_enr:	2 s
bro_smtp bro_smtp_enr:	0 ns
bro_snmp bro_snmp_enr:	2 s
bro_software bro_software_enr:	2 s
bro_ssh bro_ssh_enr:	0 ns
bro_ssl bro_ssl_enr:	1 s
bro_syslog bro_syslog_enr:	0 ns
csl_login csl_login_enr:	1 s
csl_netlog csl_netlog_enr:	1 s

Extra juicy bit

