

Prometheus at scale with Thanos

«Even though Borgmon remains internal to Google, the idea of treating time-series data as a data source for generating alerts is now accessible to everyone through those open source tools like Prometheus [...]»



Who am I



Agenda

- Old school monitoring and the Prometheus model
- Prometheus issues and alternatives for long term storage
- Thanos history, components and how it works
- Multi-cluster monitoring setup with Thanos



When you search in google...



not



Old school monitoring and the new wave

- Server - client model
- The inevitable change with kubernetes
- Monitoring-as-a-service
- Paid vs open-source



Nagios[®]

ZABBIX



WAVEFRONT



DATADOG



Sysdig



Prometheus

- What is Prometheus?
- History and key moments
- Scrape not push
- The Prometheus metrics format
- Client libraries
- Exporters



KubeCon 2018 Copenhagen



Prometheus issues

- no clustering
- no long term storage
- remote_write and remote_read

```
remote_write:  
  - url: "http://localhost:1234/receive"
```

```
# Configures the queue used to write to remote storage.  
queue_config:  
  # Number of samples to buffer per shard before we start dropping them.  
  [ capacity: <int> | default = 10000 ]
```

- [AppOptics](#): write
- [Chronix](#): write
- [Cortex](#): read and write
- [CrateDB](#): read and write
- [Elasticsearch](#): write
- [Gnocchi](#): write
- [Graphite](#): write
- [InfluxDB](#): read and write
- [IRONdb](#): read and write
- [Kafka](#): write
- [M3DB](#): read and write
- [OpenTSDB](#): write
- [PostgreSQL/TimescaleDB](#): read and write
- [SignalFx](#): write
- [Splunk](#): write
- [TiKV](#): read and write
- [VictoriaMetrics](#): write
- [Wavefront](#): write



Alternatives for long term storage and high availability

- Cortex
- InfluxDB
- M3DB
- Thanos



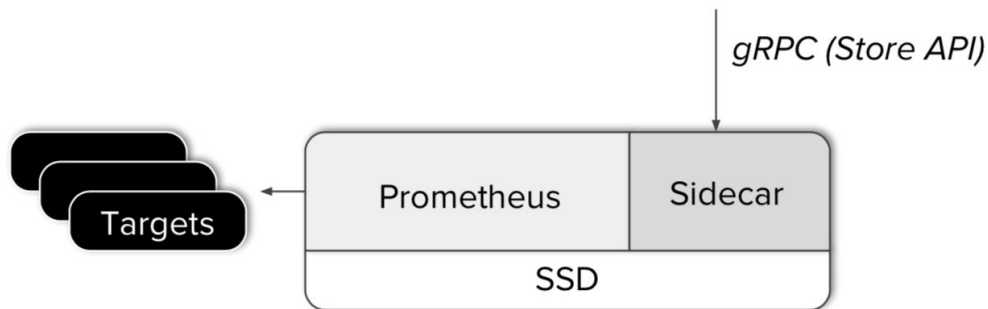
Thanos history

- Prometheus 2.0
- Fabian Reinartz (Google) - Writing a Time Series Database from Scratch:
<https://fabxc.org/tsdb/>
- Bartek Plotka (Improbable)



Thanos sidecar

- Features:
 - Backs up to S3
 - Does not uses remote_write
- What to change*:
 - tsdb.min/max-block-duration = 2h
 - tsdb.retention = 12h

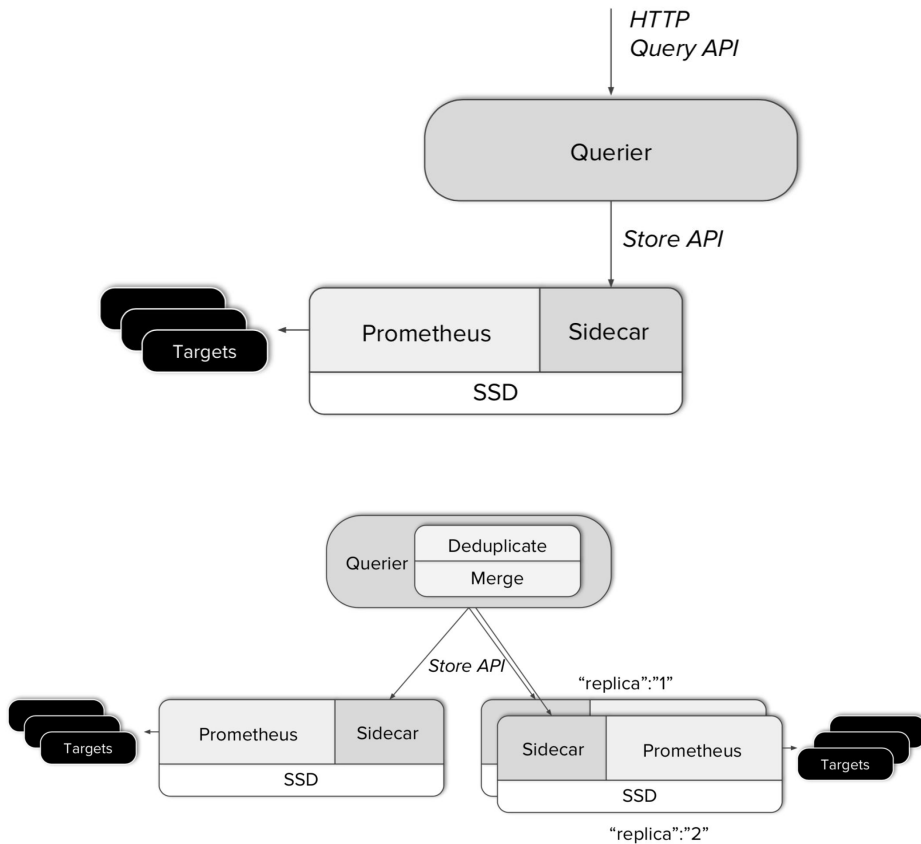


* if we want to also use Thanos Store



Thanos Query

- Issues with Prometheus
- Features:
 - Global data view
 - Deduplication
 - No loss of data during update or downtime
 - Big queries do not lead to scraping downtime



Thanos Ruler

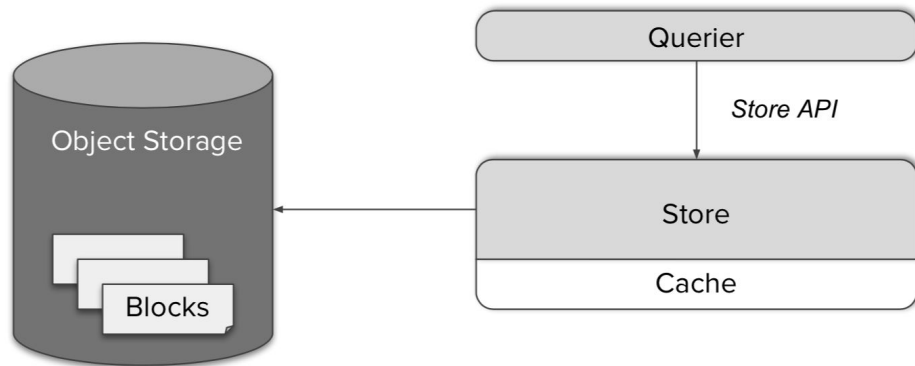
- Another data source
- Alerting and recording rules
- No duplication of alerts



Thanos Store and Compact

Features:

- Compaction
- Downsampling
- Customizable retention
- Cache of the metadata in S3
- Range queries against S3
- Infinite retention is possible





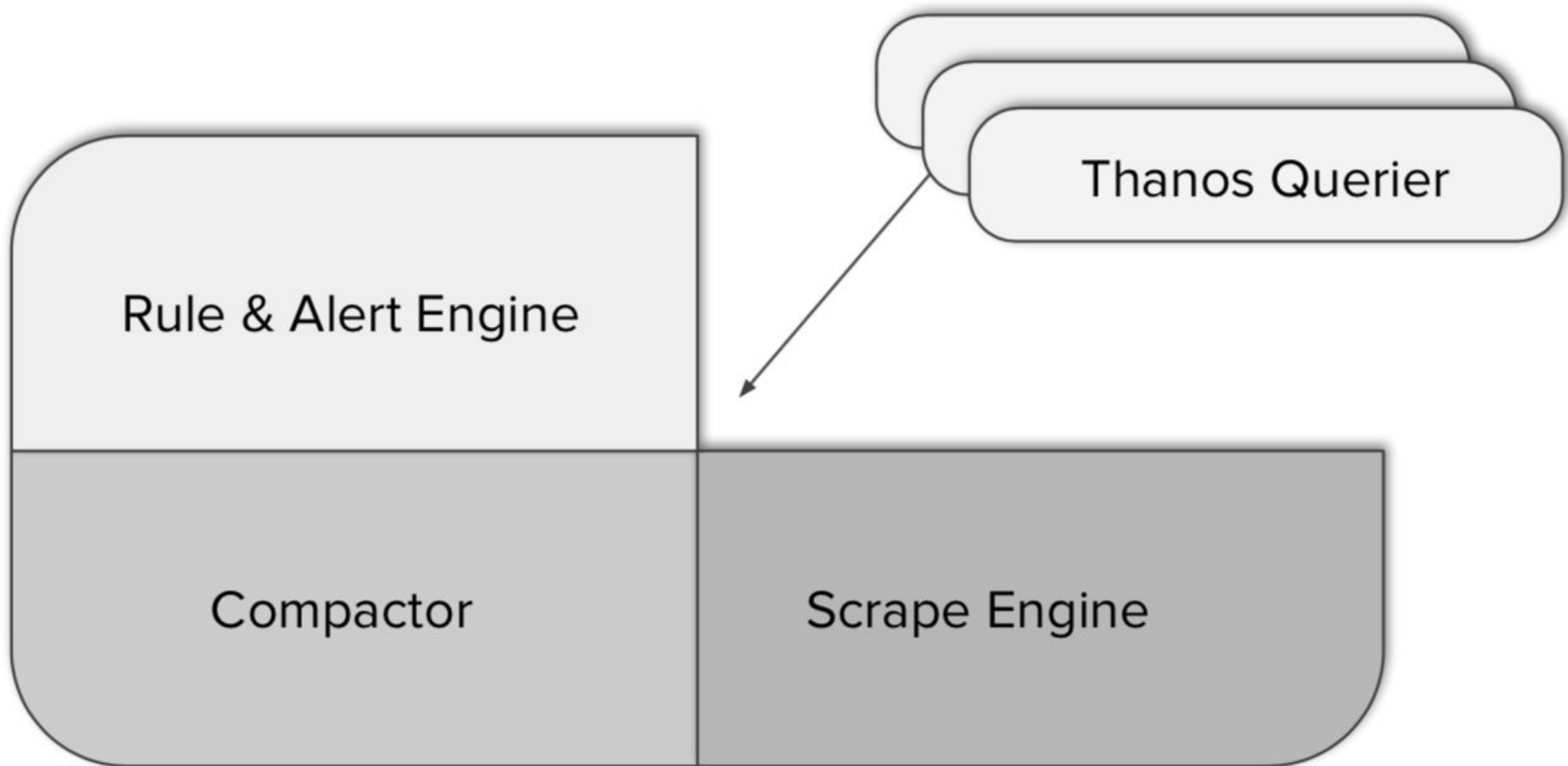
Prometheus

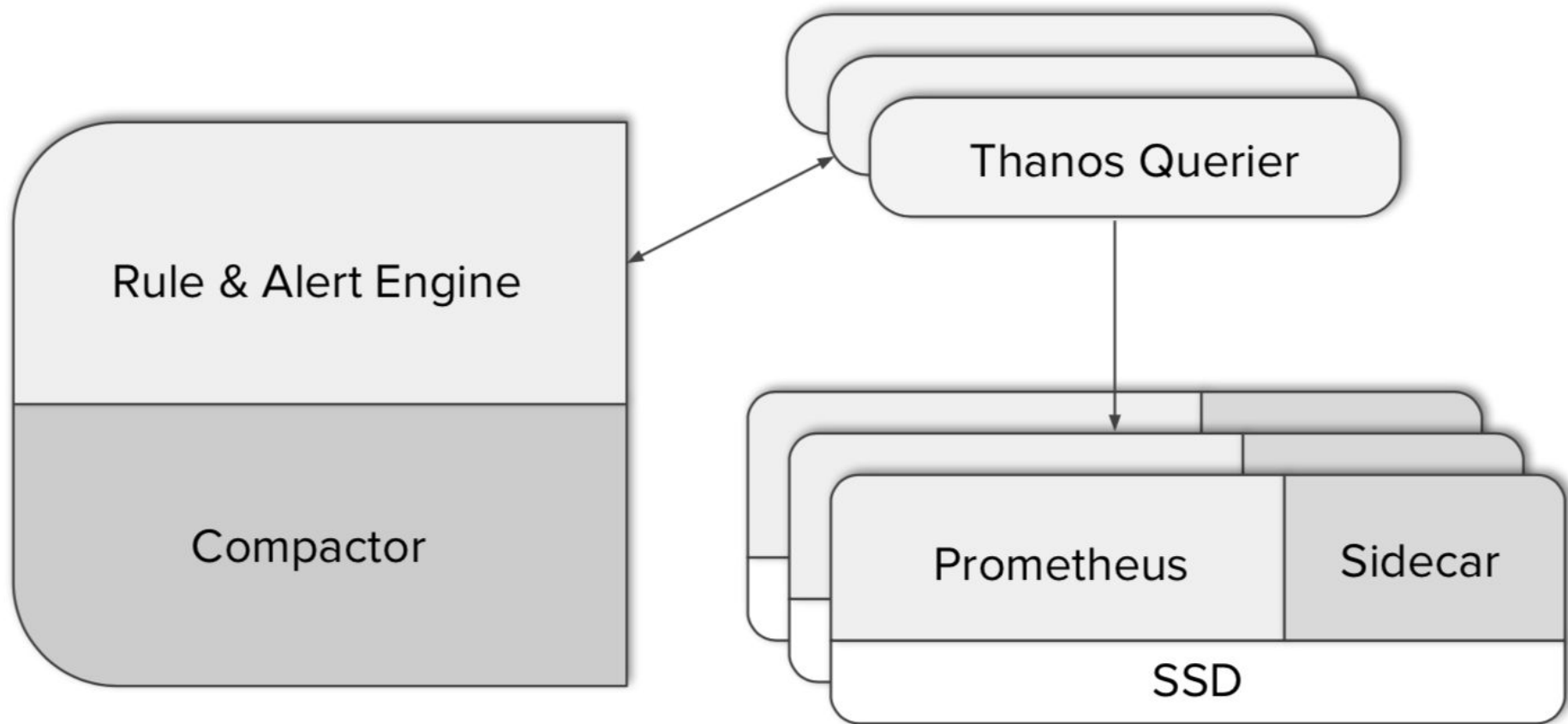
Rule & Alert Engine

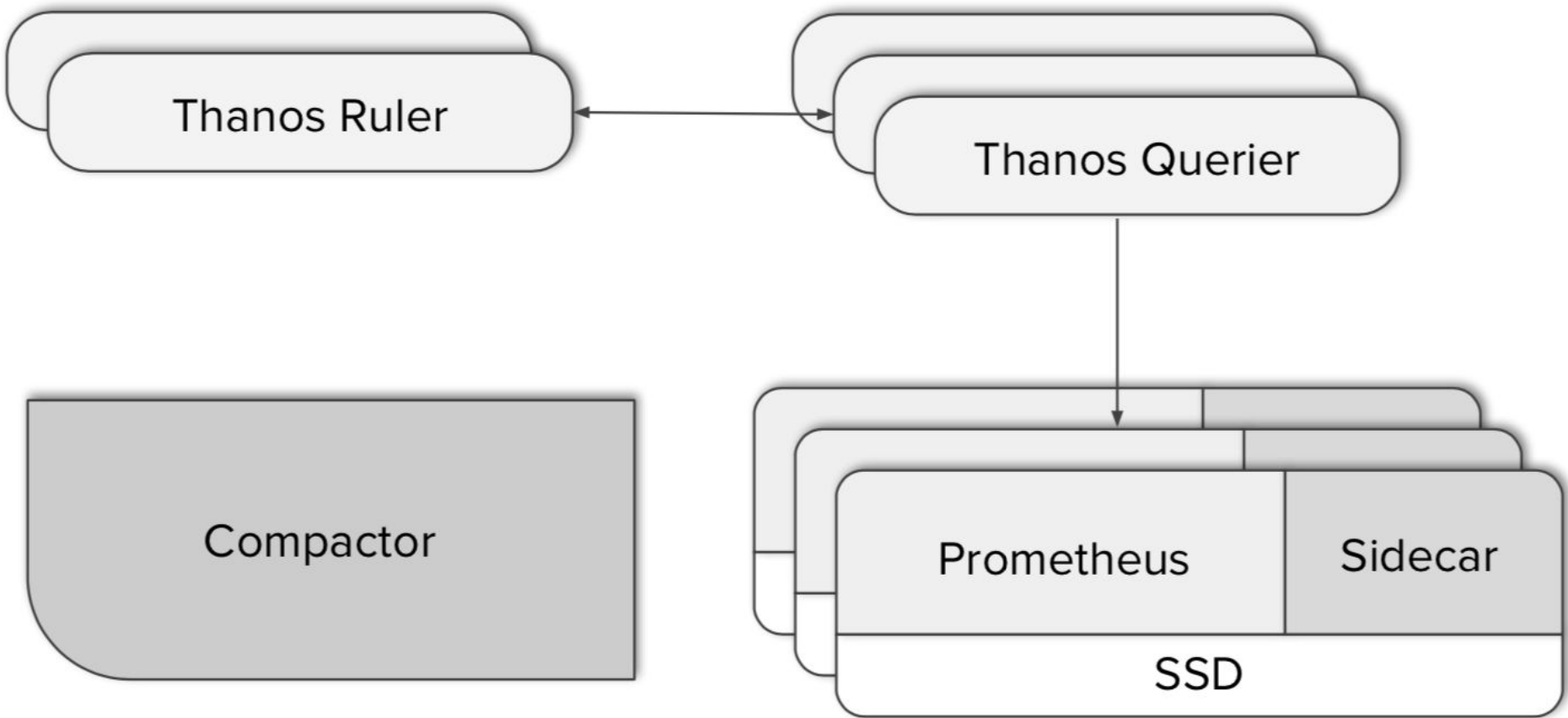
Querier

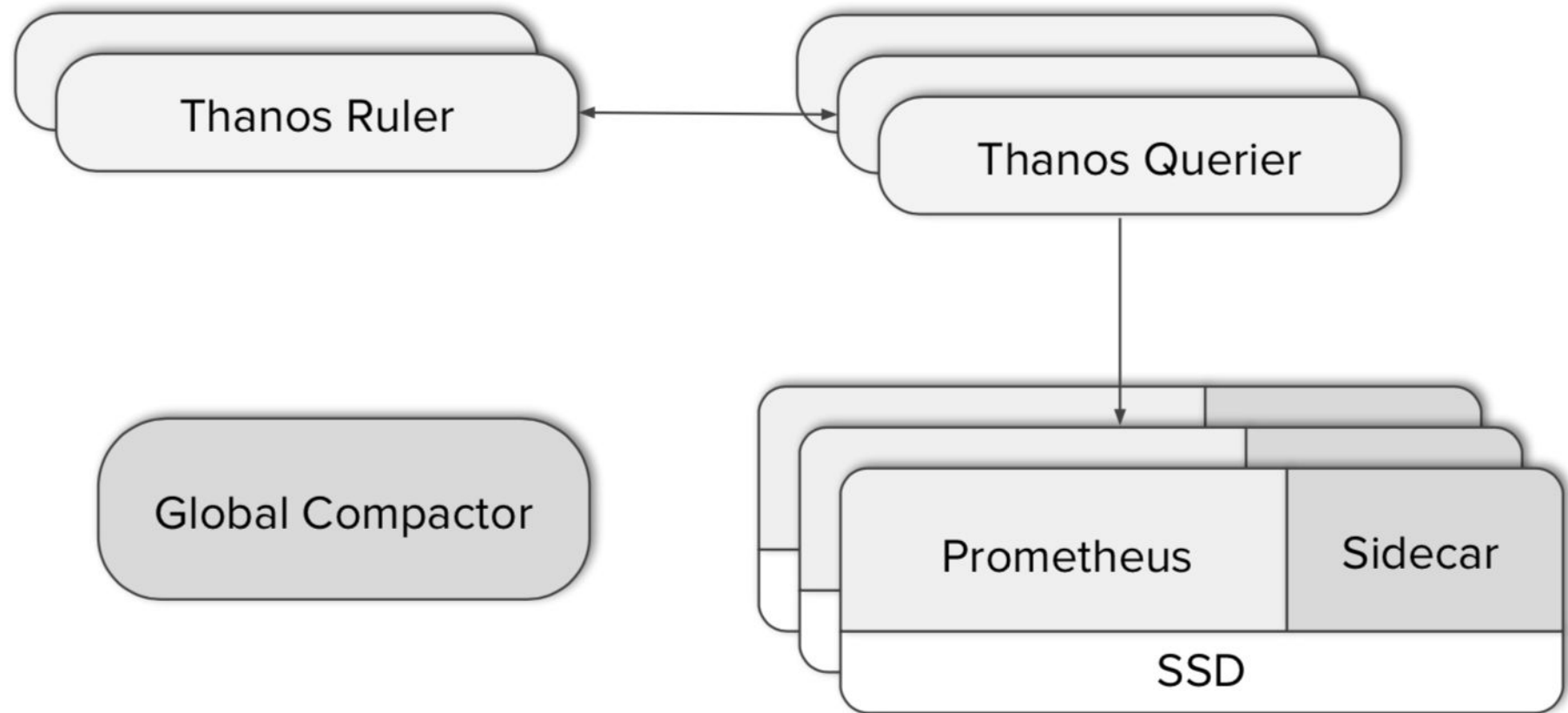
Compactor

Scrape Engine

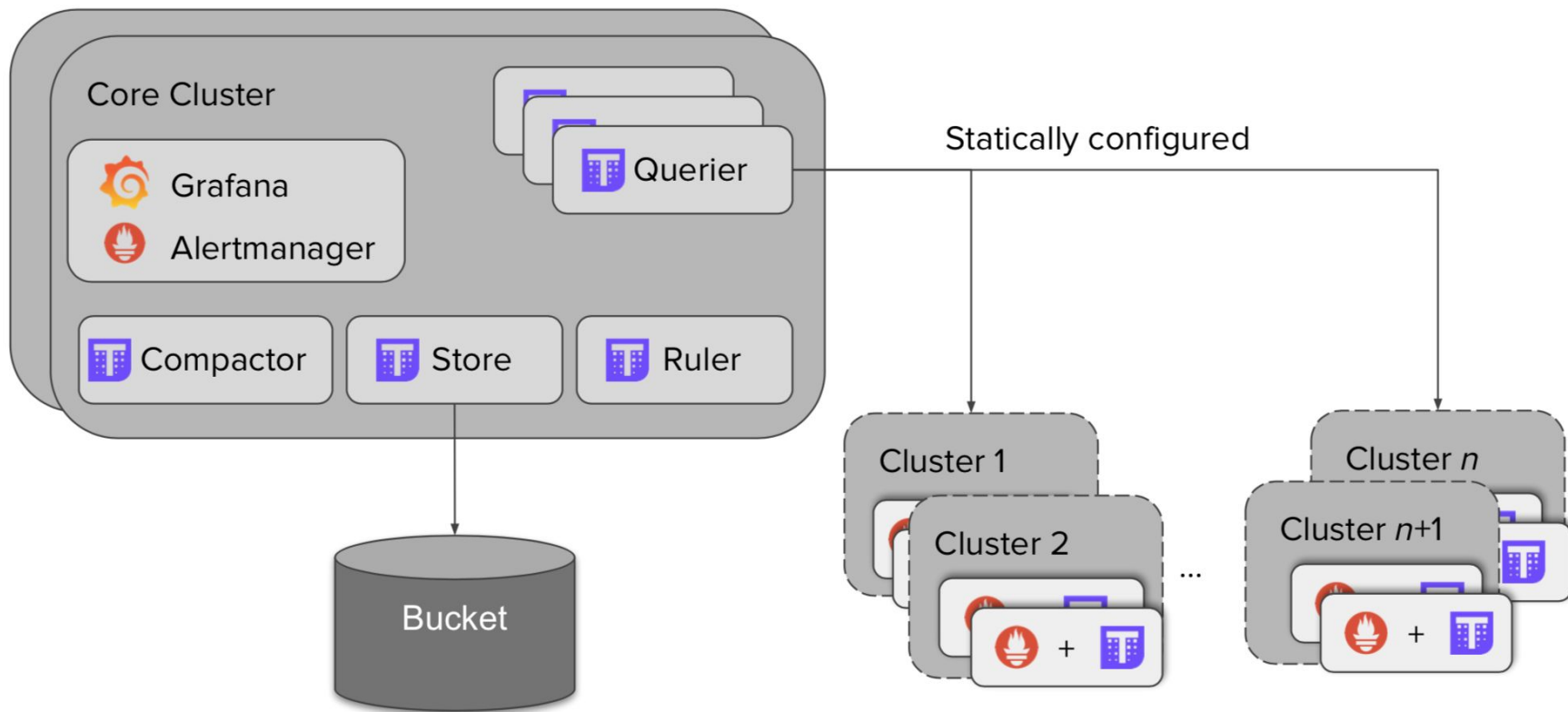








Multi-cluster monitoring setup



Q&A

References:

https://www.youtube.com/watch?v=Fb_IYX01IX4&t=1231s

<https://www.slideshare.net/BartomiejPotka/thanos-global-durable-prometheus-monitoring>

