


Performance Engineer

Hedging

* proactive

* rather than wait for "tail latency"

* proactively duplicate request

Overload Protection

* prevent server overload

* monitor & share info to be used for detection

Observability

* directions for detecting & localizing problems

* React to problem

autoconfig

* prevent problems by determining optimal configs

Chaos Engineering

* introduce / create problems to proactively detect bugs in code

"Three pillars of observability"

metrics (series)

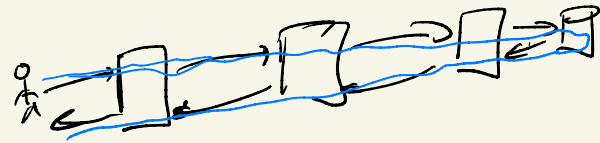
timeseries data capturing
resource use over
time

logs

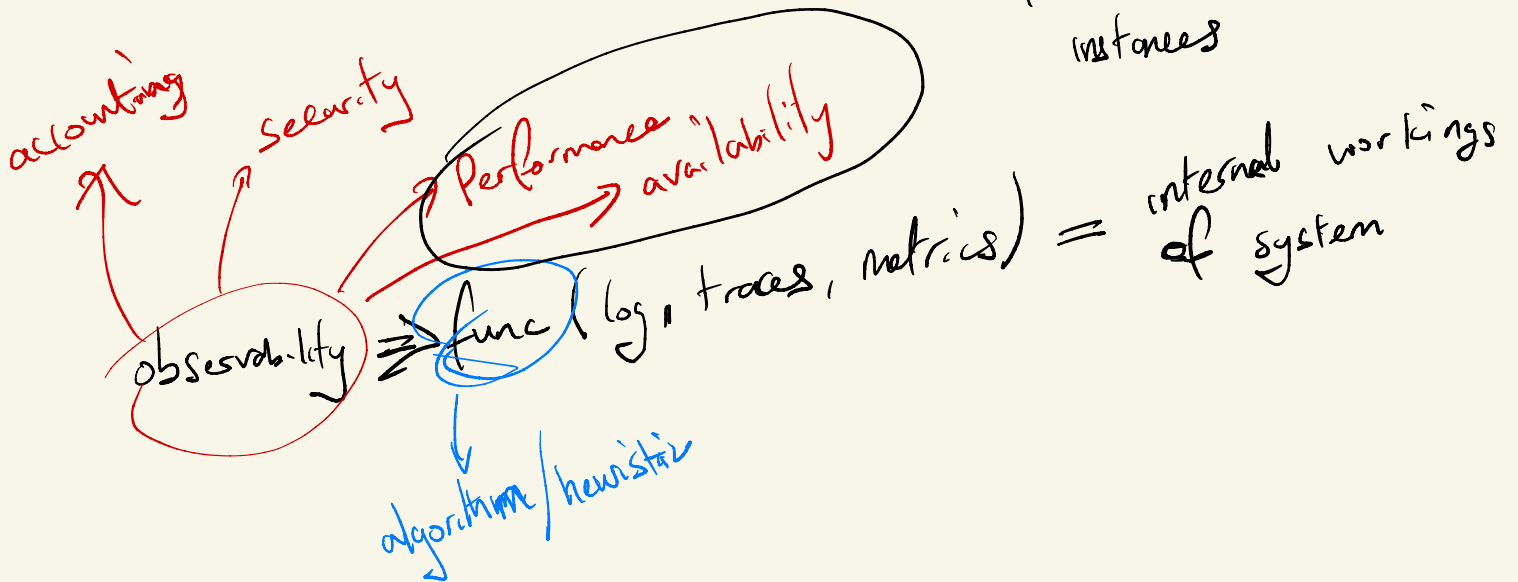
recording of event in
the system

distributed tracing

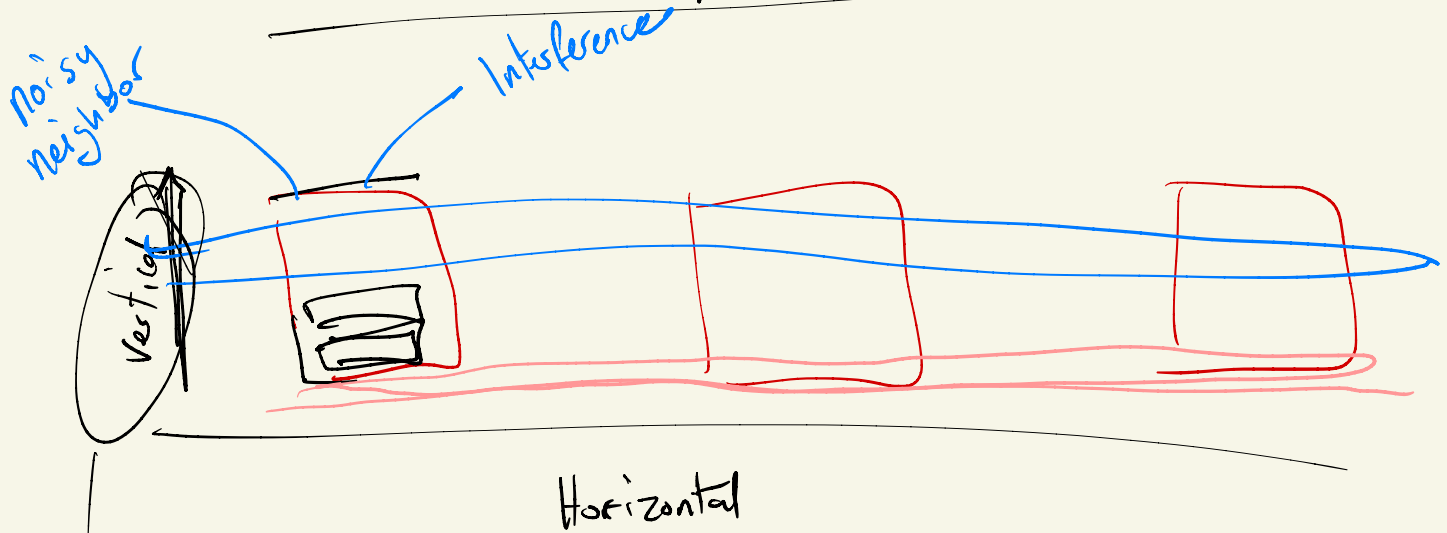
* Traces = follows the
flow of execution
across instance boundaries



captures instances & flows
for localizing problem to
instances



Dimensions of observability



interactions
btw different
entities on a
machine

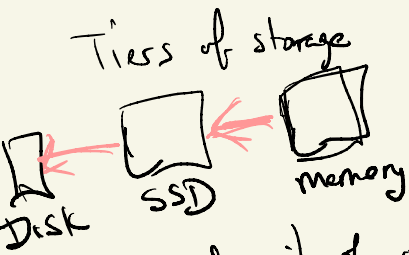
Noisy Neighbor : another service on the same node
uses resources in manner that impacts
your ability to use resource hence bad
perf

	vertical	horizontal
logs	✓	
traces		✓
metrics	✓	

which events
 & which resources

vertical = issues due to interactions on a machine
 horizontal = issues due to interactions across machines

Data Storage / Sampling

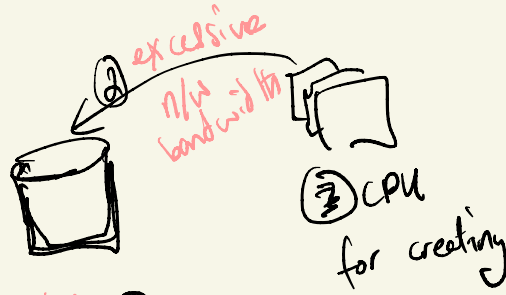


④ get rid of old or irrelevant data

⑤ sample data as you collect

⑥ Turn off observability data collection

limited storage ①



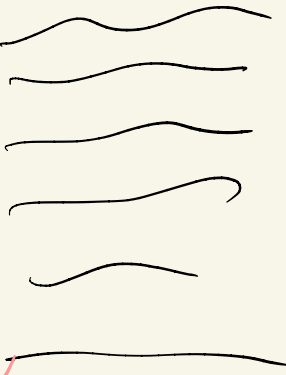
Bottlenecks in sending data to a central DB

Heterogeneity in services leads to variety in the quality of captured data

different sampling → to no data

Logs

stream of random line/event



unstructured
stream of events

fake
print
statement



① what are fundamental
templates

② as a stream of templates

print ("log data about cpu: %u @ time: %s")

identify unusual templates

identify anomalous
spike in relative
of templates

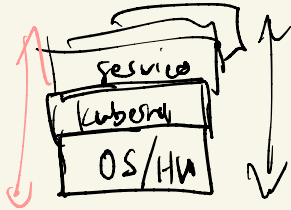
Trace / microservice
based errors

Metrics

V. Logs

Sieve $\Rightarrow$ reduce overheads or speed up processing

unstructured
microservice / trace level



explosion in metrics

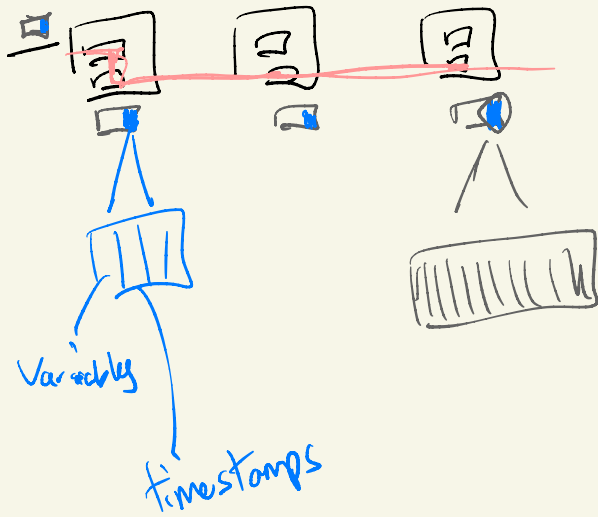
k-shape

$\Downarrow$
reduce #
of metrics
collected / analyze

call-graph to help causality

$\Downarrow$
restrict the pair of
variables / metrics
that granger causality

Trace



each instance/process involved in
servicing a request

or each function that's required
to process a request