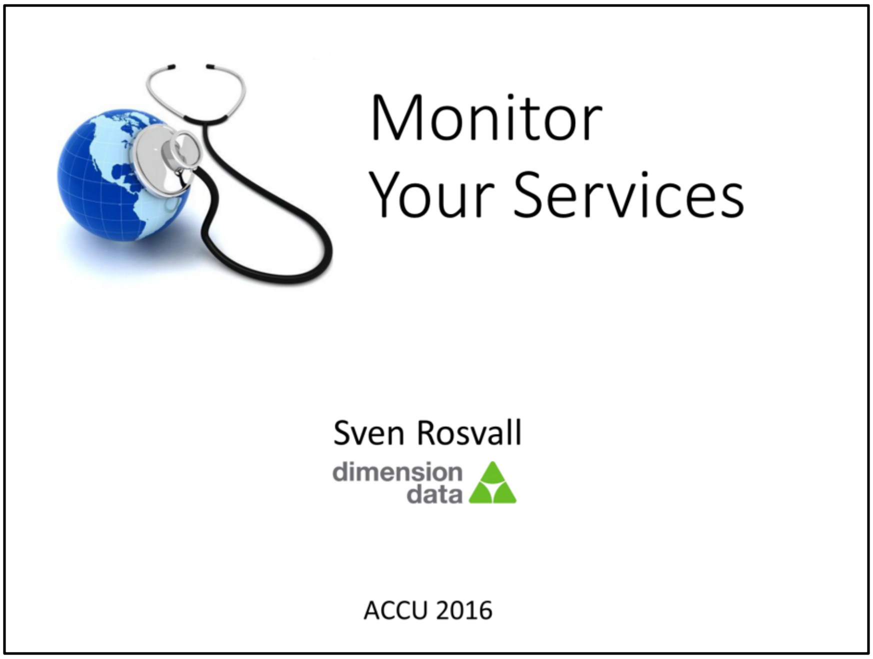


Monitor Your Services

Sven Rosvall

dimension data

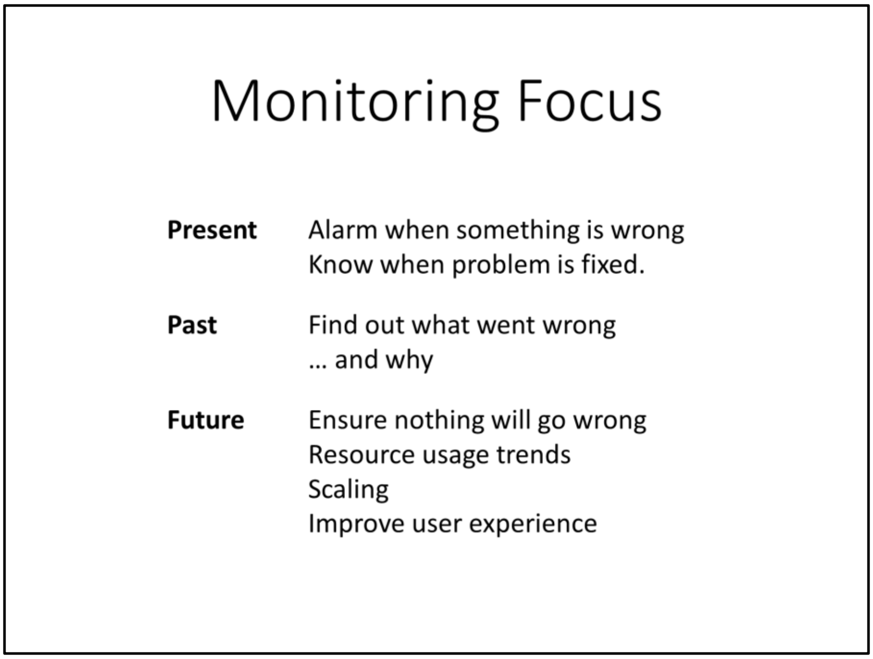
ACCU 2016

Me
- Quality interest
- Good code
- Reliable code / service
- Worked for Amazon, Microsoft, Dimension Data

Service:
- Public or internal service.
- Parts of the Talk may also apply to boxed software.

KNOW

IMPROVE



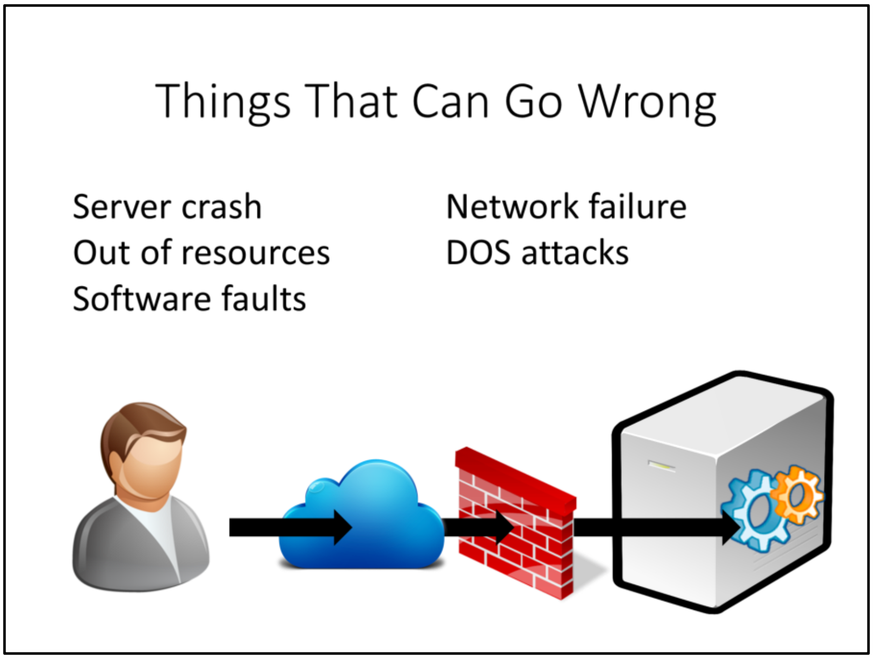
Monitoring Focus

Present	Alarm when something is wrong Know when problem is fixed.
Past	Find out what went wrong … and why
Future	Ensure nothing will go wrong Resource usage trends Scaling Improve user experience

Present: Know problems before customer notices.

Service Appears Wrong

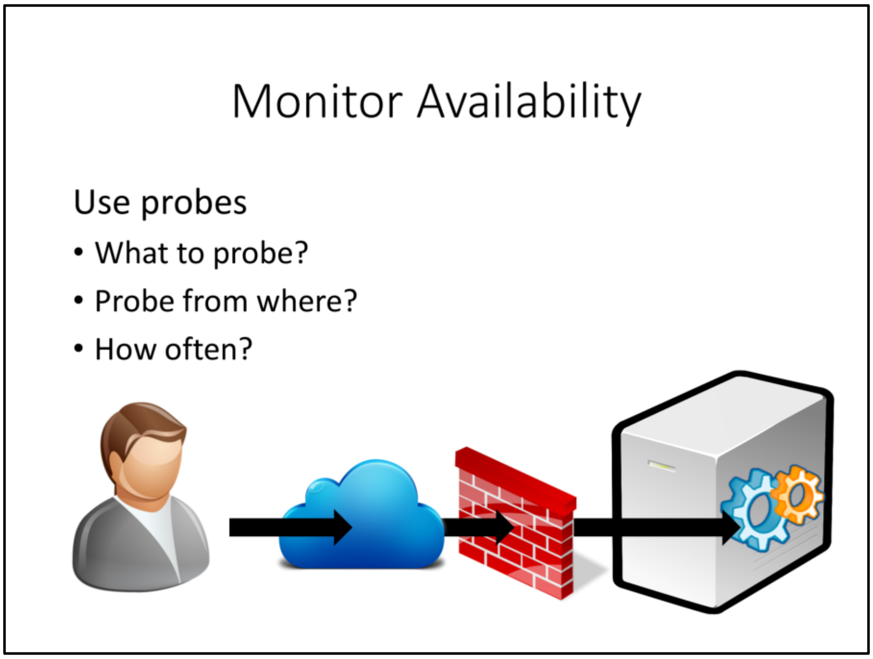
- No response from service
- Intermittently available
- Slow responses
- Incorrect responses
- Incorrect behaviour



Things That Can Go Wrong

Server crash
Out of resources
Software faults

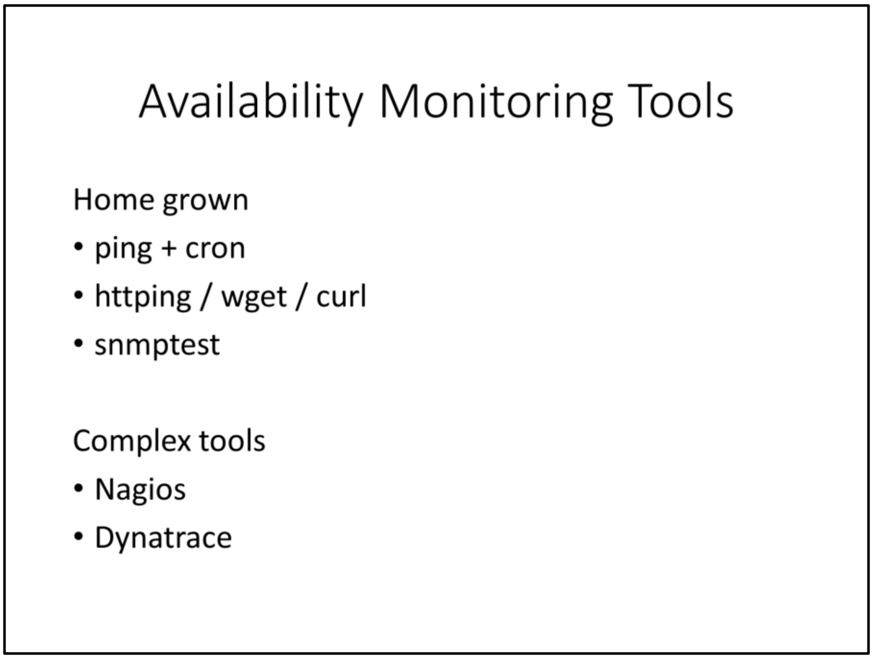
Network failure
DOS attacks



Monitor Availability

Use probes

- What to probe?
- Probe from where?
- How often?



Availability Monitoring Tools

Home grown

- ping + cron
- httping / wget / curl
- snmptest

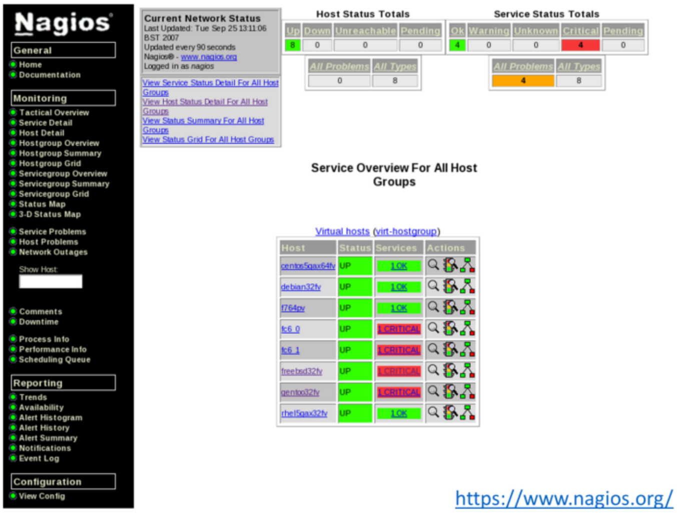
Complex tools

- Nagios
- Dynatrace

Ping (ICMP) is very shallow. Some servers respond to ping even when they have crashed. Ping does not go through some firewalls.

Dynatrace a.k.a. Gomez

Nagios



Current Network Status
Last Updated: Tue Sep 25 13:11:06 BST 2007
Updated every 90 seconds
Nagios® - www.nagios.org
Logged in as *nagios*

View Service Status Detail For All Host Groups
View Host Status Detail For All Host Groups
View Status Summary For All Host Groups
View Status Grid For All Host Groups

Host Status Totals

Up	Down	Unreachable	Pending
8	0	0	0

All Problems	*All Types*
0	8

Service Status Totals

Ok	Warning	Unknown	Critical	Pending
4	0	0	4	0

All Problems	*All Types*
4	8

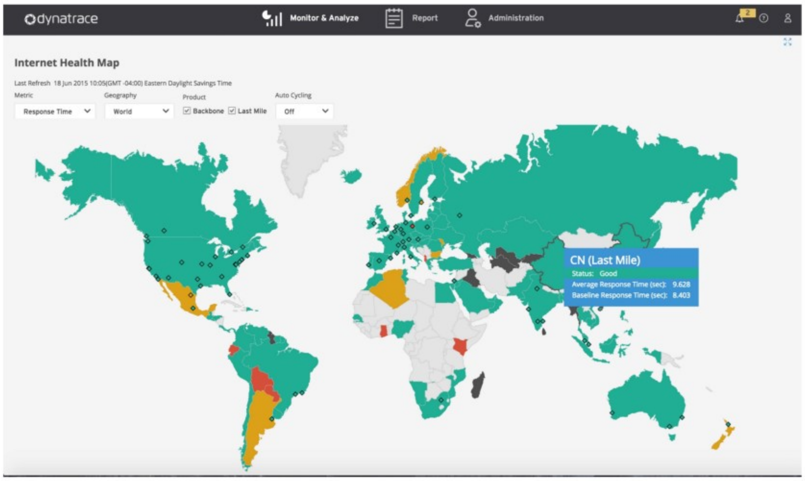
Service Overview For All Host Groups

Virtual hosts (virt-hostgroup)

Host	Status	Services	Actions
centos5gax64fv	UP	1 OK	
debian32fv	UP	1 OK	
f764pv	UP	1 OK	
fc6_0	UP	1 CRITICAL	
fc6_1	UP	1 CRITICAL	
freebsd32fv	UP	1 CRITICAL	
gentoo32fv	UP	1 CRITICAL	
rhel5gax32fv	UP	1 OK	

https://www.nagios.org/

Dynatrace



Internet Health Map

Last Refresh 18 Jun 2015 10:05(GMT -04:00) Eastern Daylight Savings Time

Metric: Response Time

Geography: World

Product: ☑ Backbone ☑ Last Mile

Auto Cycling: Off

CN (Last Mile)

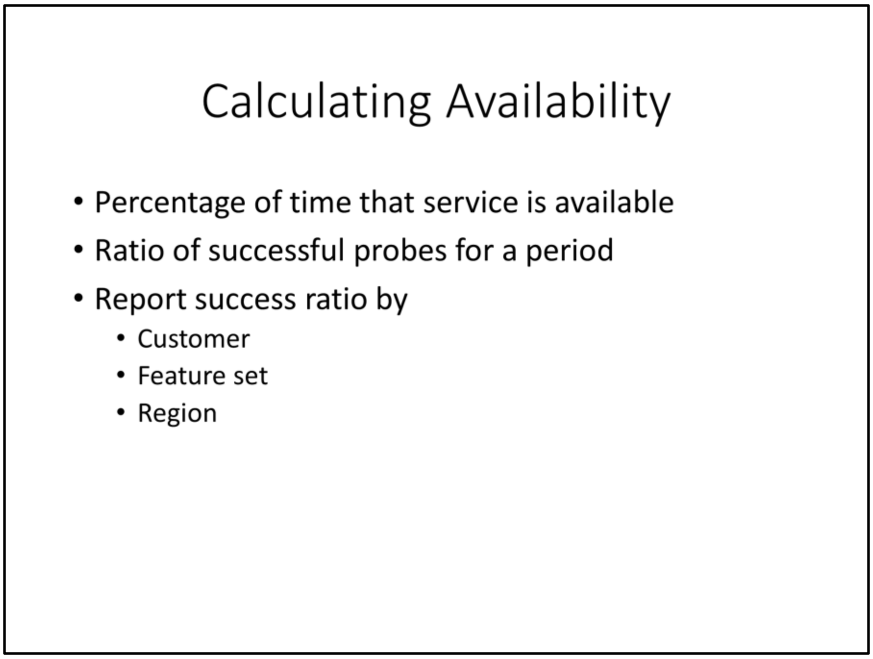
Status: Good

Average Response Time (sec): 9.628

Baseline Response Time (sec): 8.403

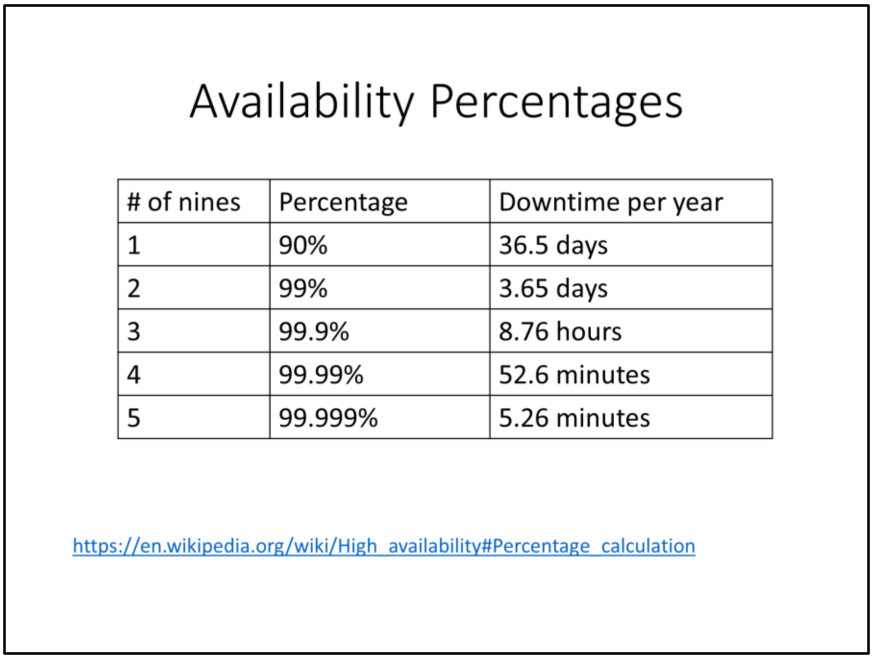
http://www.dynatrace.com/

Formerly known as Gomez.



Calculating Availability

- Percentage of time that service is available
- Ratio of successful probes for a period
- Report success ratio by
  - Customer
  - Feature set
  - Region

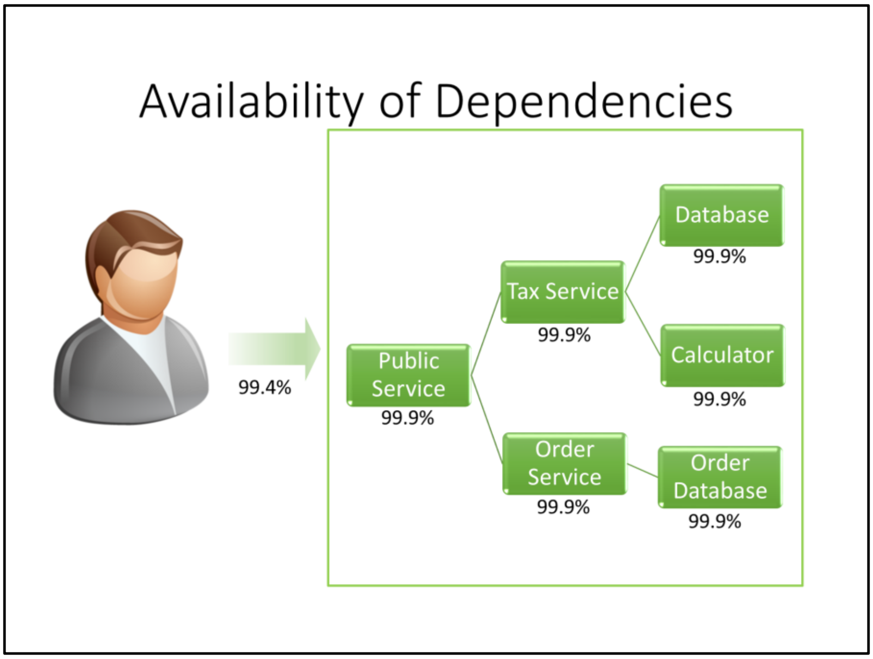


Availability Percentages

# of nines	Percentage	Downtime per year
1	90%	36.5 days
2	99%	3.65 days
3	99.9%	8.76 hours
4	99.99%	52.6 minutes
5	99.999%	5.26 minutes

https://en.wikipedia.org/wiki/High_availability#Percentage_calculation

Some examples of SLAs:
https://aws.amazon.com/ec2/sla/
https://azure.microsoft.com/en-us/support/legal/sla/virtual-machines/v1_0/
http://cloud.dimensiondata.com/saas-solutions/about/legal/service-level-agreement



Availability of Dependencies

99.4%

Public Service 99.9%

Tax Service 99.9%

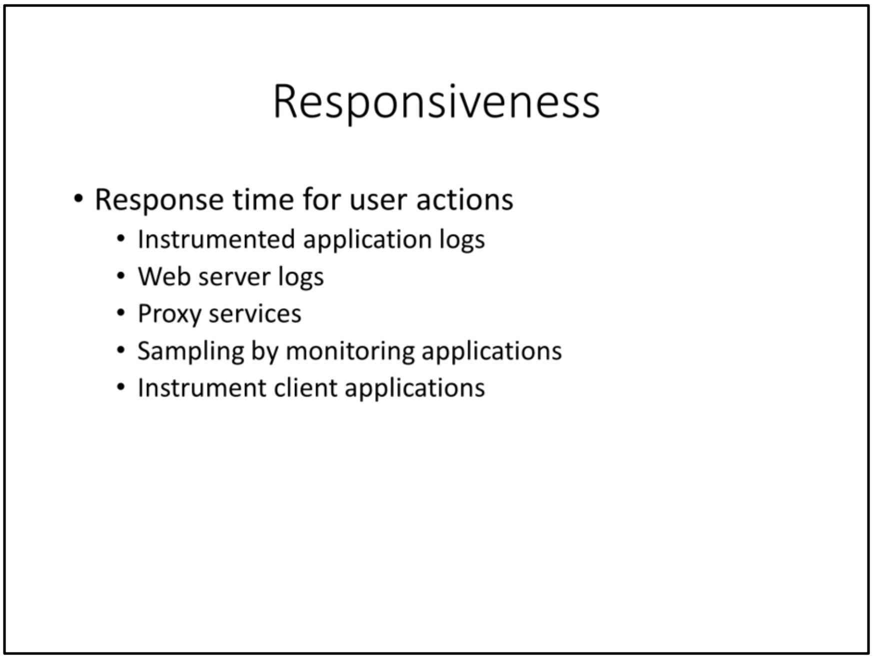
Database 99.9%

Calculator 99.9%

Order Service 99.9%

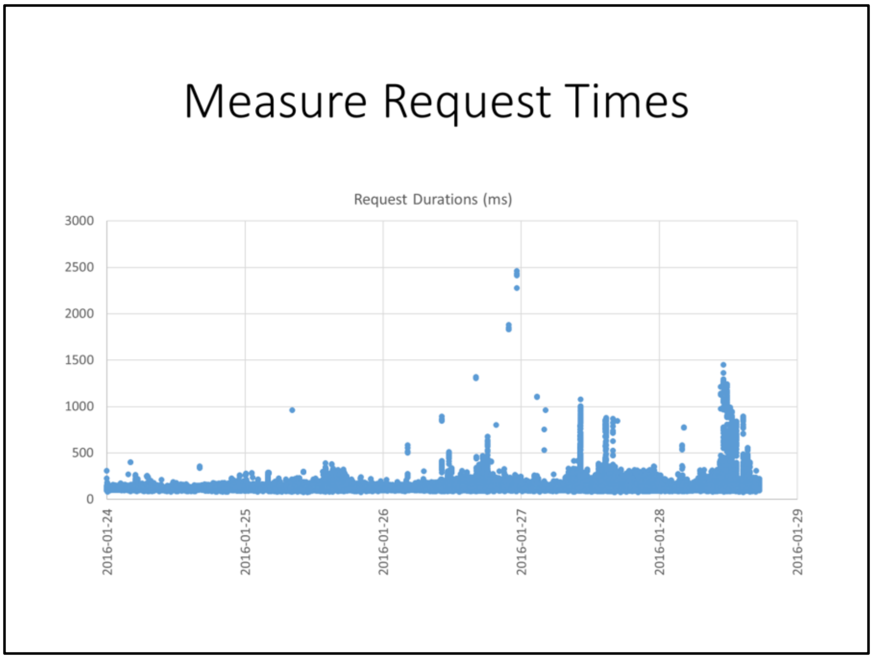
Order Database 99.9%

Each service has 99.9% availability.
Customer experiences 99.4%



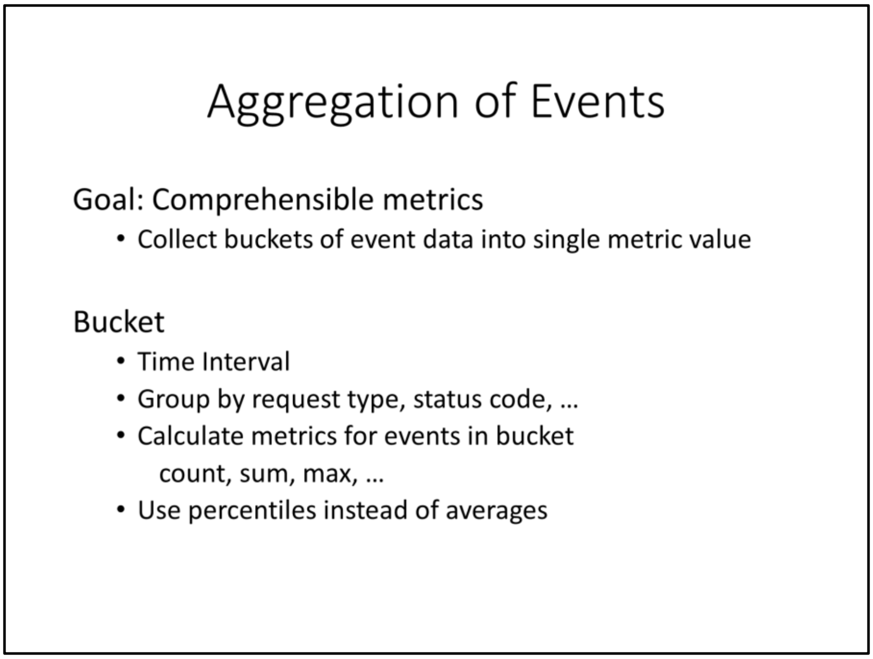
Responsiveness

- Response time for user actions
  - Instrumented application logs
  - Web server logs
  - Proxy services
  - Sampling by monitoring applications
  - Instrument client applications



Measure Request Times

NA GetServer



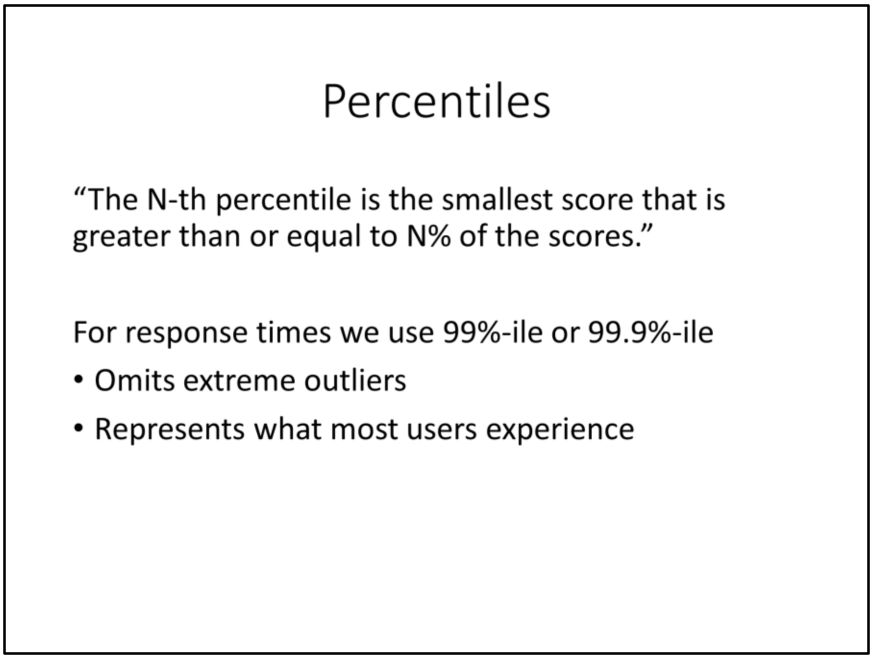
Aggregation of Events

Goal: Comprehensible metrics

- Collect buckets of event data into single metric value

Bucket

- Time Interval
- Group by request type, status code, …
- Calculate metrics for events in bucket
  count, sum, max, …
- Use percentiles instead of averages



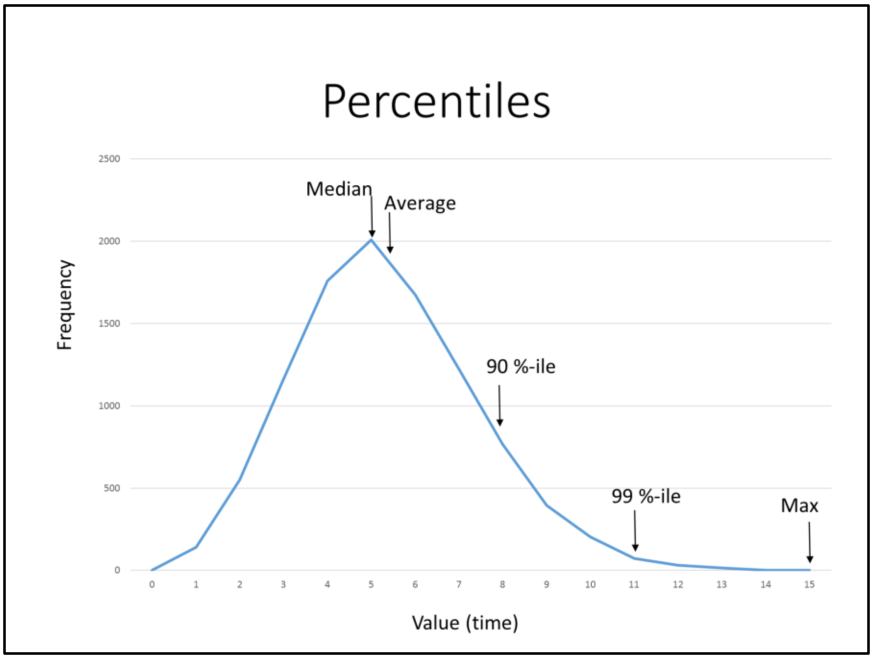
Percentiles

"The N-th percentile is the smallest score that is greater than or equal to N% of the scores."

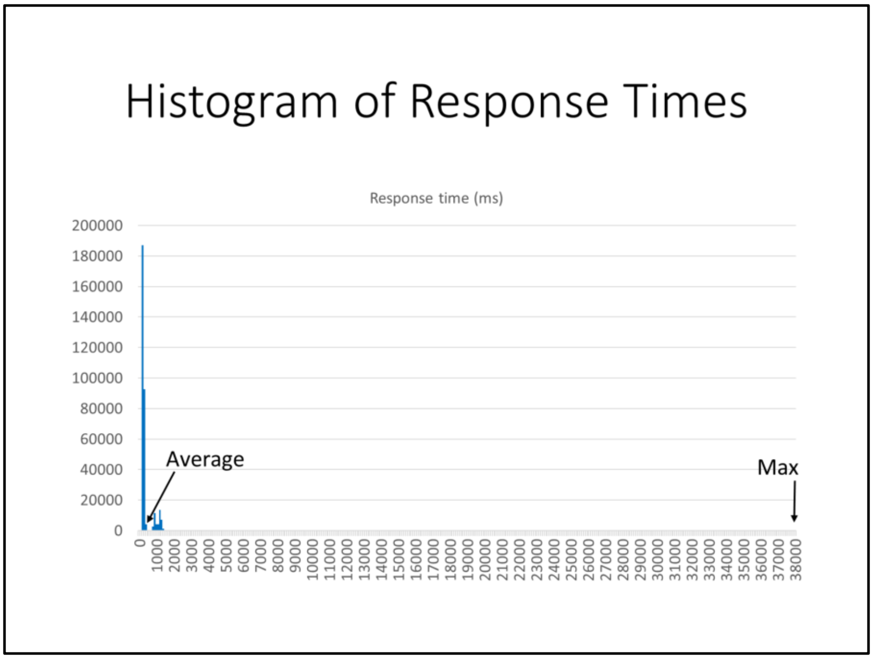
For response times we use 99%-ile or 99.9%-ile

- Omits extreme outliers
- Represents what most users experience

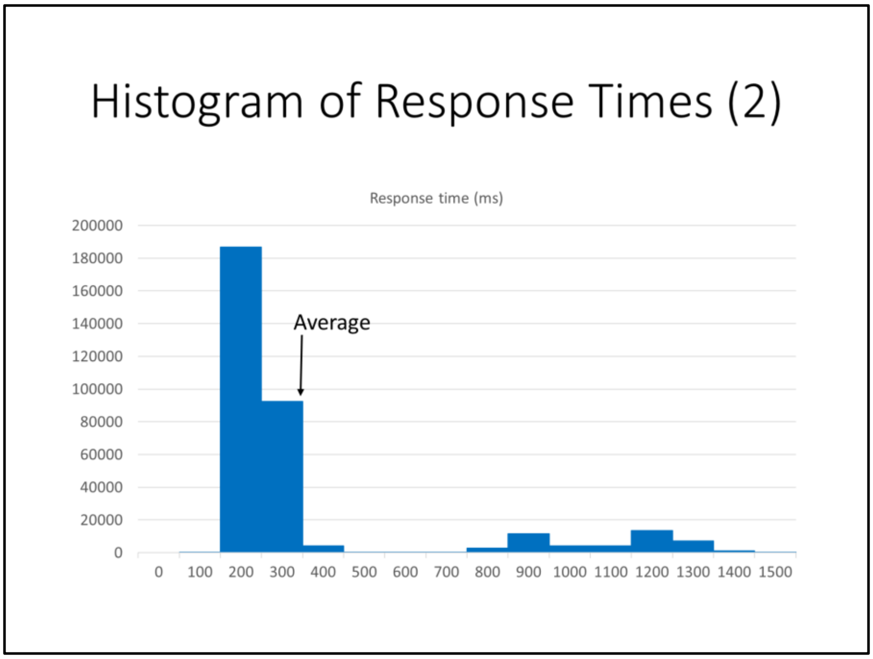
There is no definitive definition of percentiles.
Think of car speeds. Average speed makes sense for a single car that has different speeds during the journey. Total distance divided by total time. Easy to calculate and reason about.
Measure speed of 10 cars in one area. 9 are sticking to the speed limit of 30 mph, but one is doing 100 mph. The average speed for this group of cars is 37 mph. Does this mean that the whole group is driving too fast? 90%-ile is 30 mph, which means that the vast majority (90%) is sticking to speed limits.



Percentiles

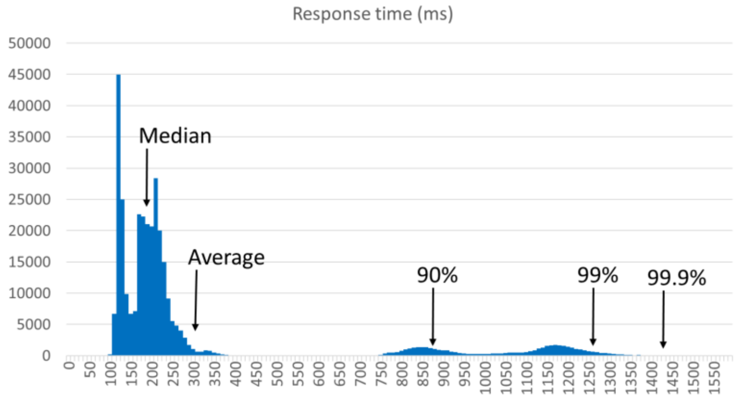


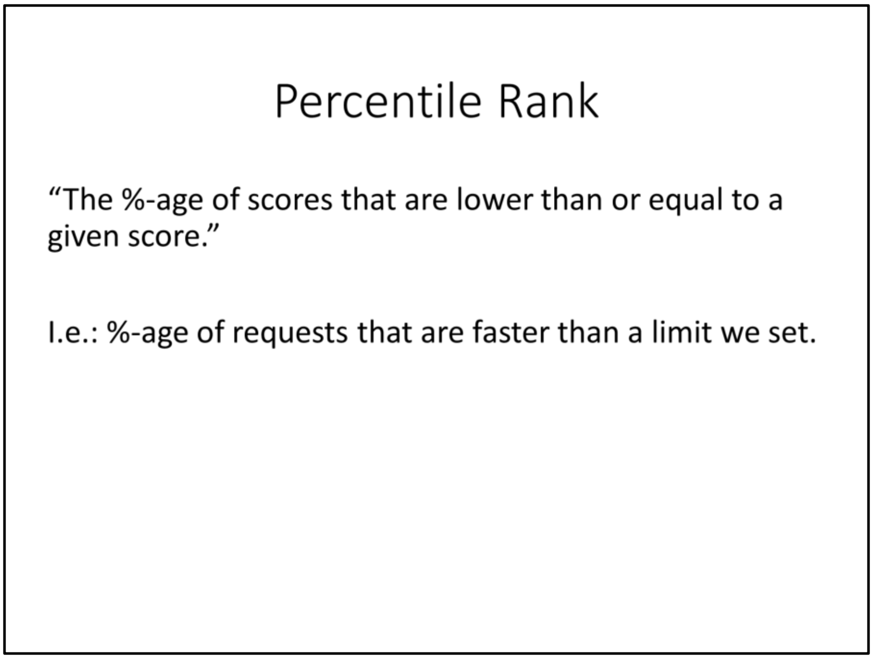
Histogram of Response Times



Histogram of Response Times (2)

Histogram of Response Times (3)





Percentile Rank

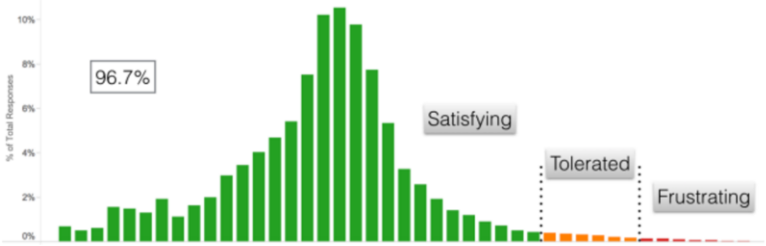
"The %-age of scores that are lower than or equal to a given score."

I.e.: %-age of requests that are faster than a limit we set.

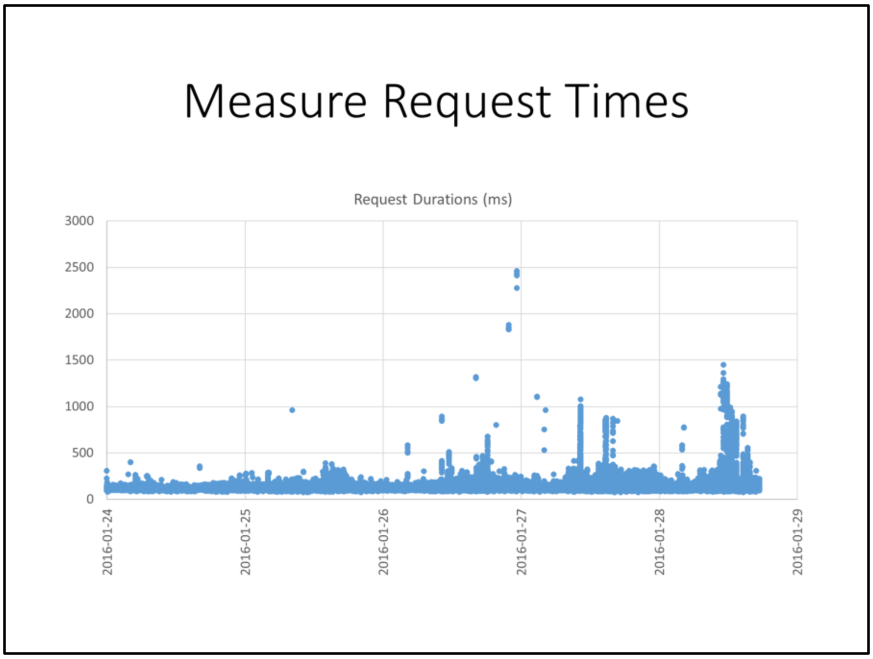
Apdex

Application Performance Index

$$\text{Apdex} = \frac{\text{Satisfying + Tolerated/2}}{\text{Total}}$$



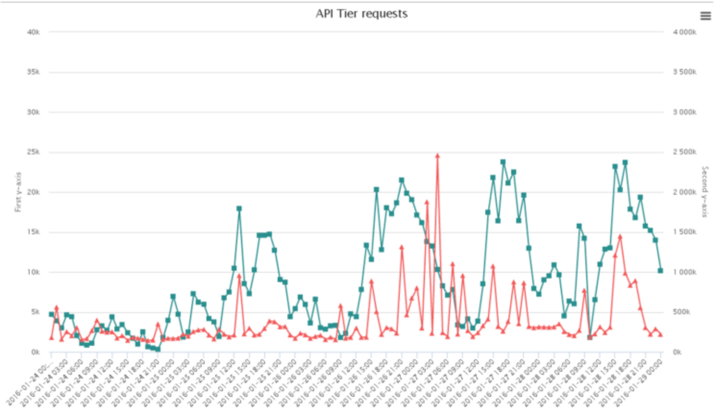
http://www.apdex.org/index.php/alliance/specifications/

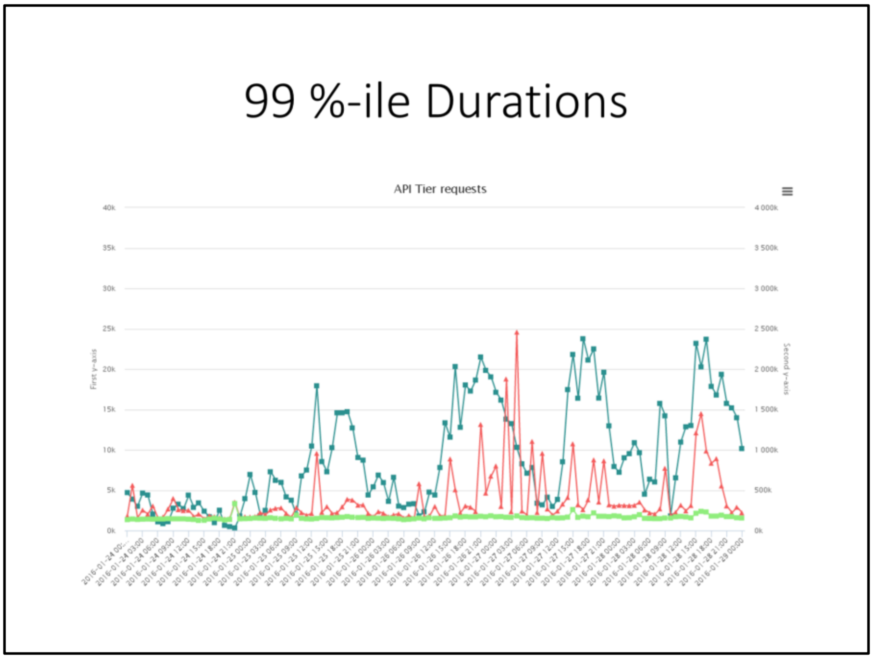


Measure Request Times

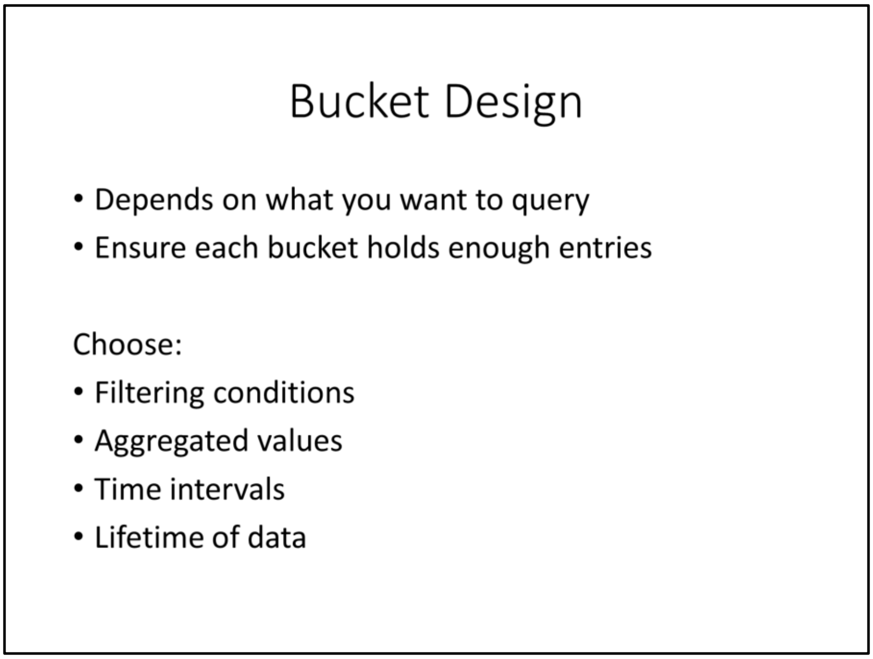
NA GetServer

Request Count and Max Durations





99 %-ile Durations

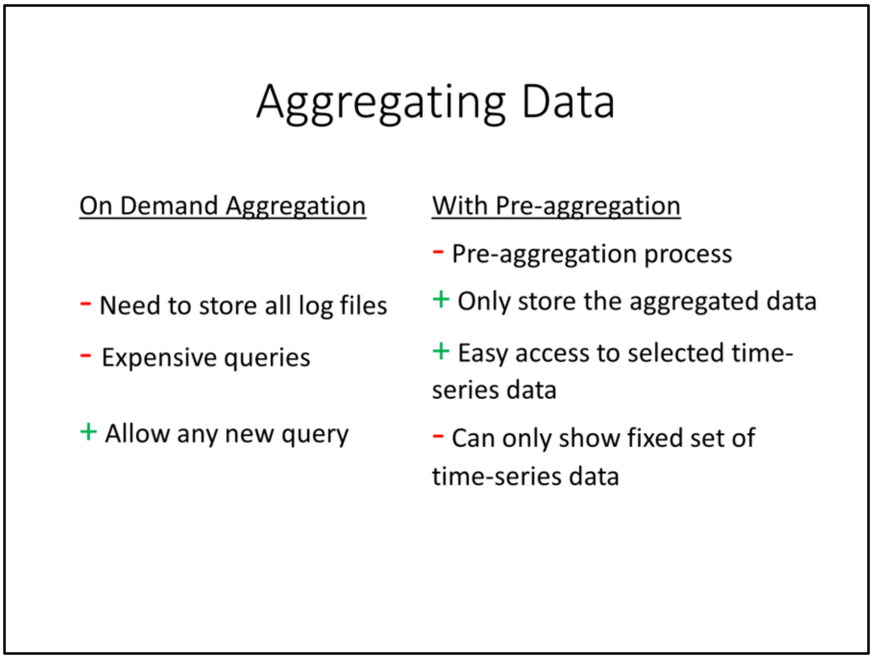


Bucket Design

- Depends on what you want to query
- Ensure each bucket holds enough entries

Choose:

- Filtering conditions
- Aggregated values
- Time intervals
- Lifetime of data



Aggregating Data

On Demand Aggregation

- Need to store all log files
- Expensive queries

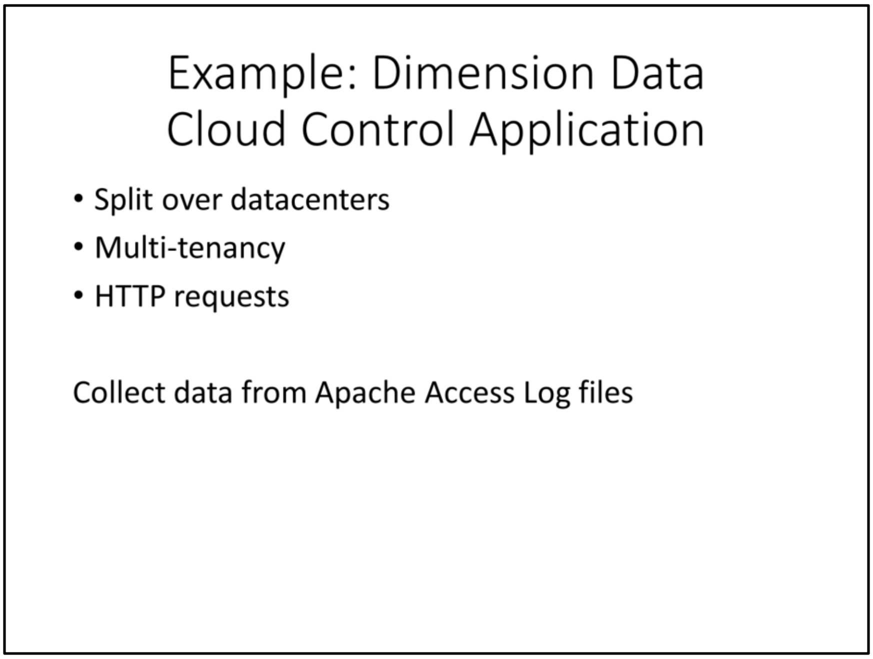
+ Allow any new query

With Pre-aggregation

- Pre-aggregation process

+ Only store the aggregated data
+ Easy access to selected time-series data

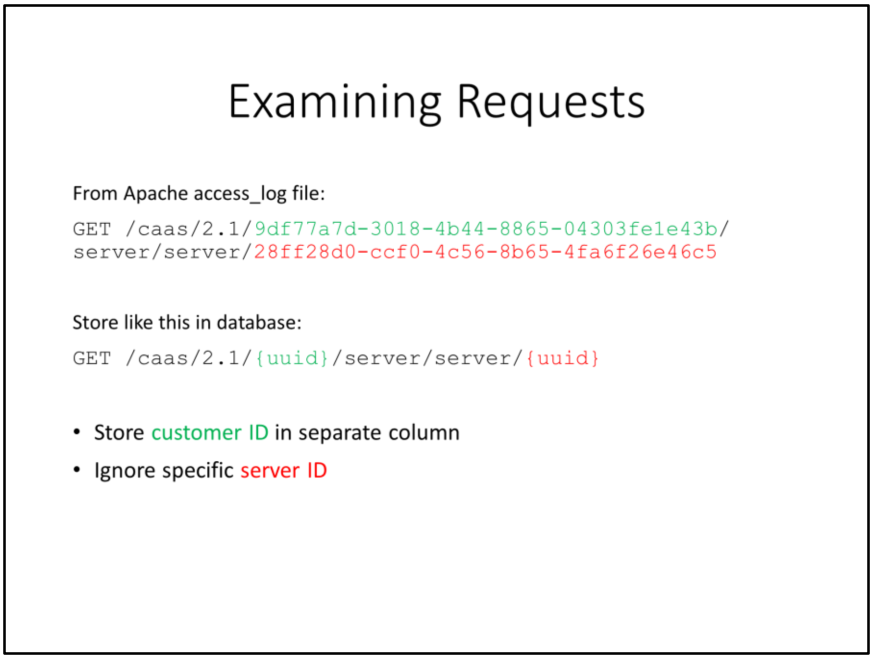
- Can only show fixed set of time-series data



Example: Dimension Data Cloud Control Application

- Split over datacenters
- Multi-tenancy
- HTTP requests

Collect data from Apache Access Log files



Examining Requests

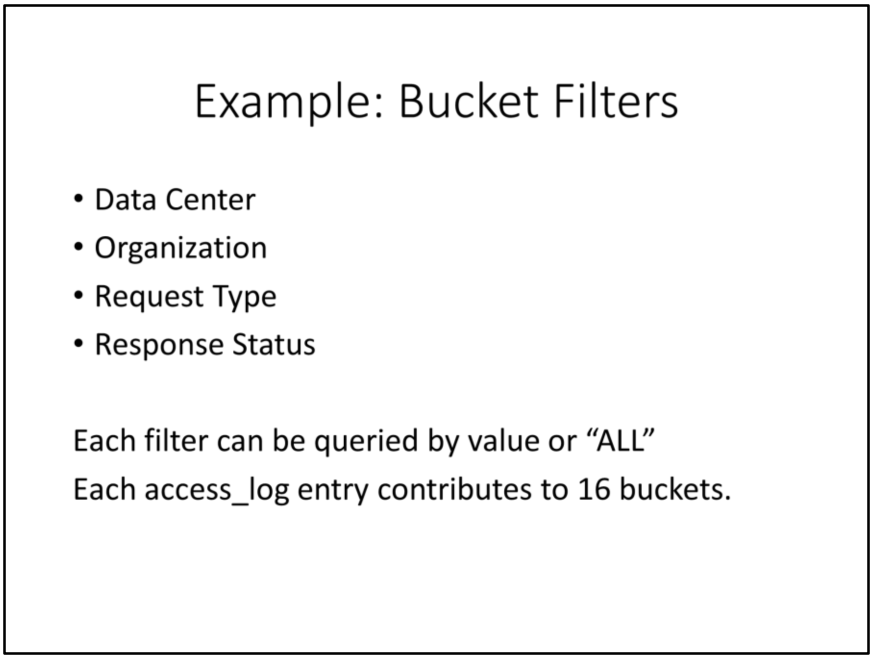
From Apache access_log file:

```
GET /caas/2.1/9df77a7d-3018-4b44-8865-04303fe1e43b/
server/server/28ff28d0-ccf0-4c56-8b65-4fa6f26e46c5
```

Store like this in database:

```
GET /caas/2.1/{uuid}/server/server/{uuid}
```

- Store customer ID in separate column
- Ignore specific server ID

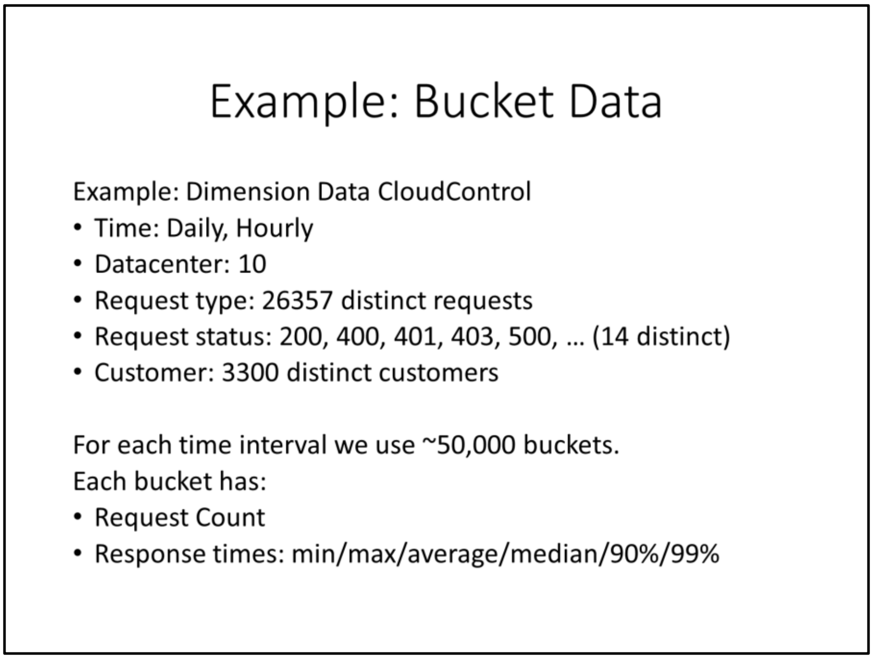


Example: Bucket Filters

- Data Center
- Organization
- Request Type
- Response Status

Each filter can be queried by value or “ALL”

Each access_log entry contributes to 16 buckets.



Example: Bucket Data

Example: Dimension Data CloudControl

- Time: Daily, Hourly
- Datacenter: 10
- Request type: 26357 distinct requests
- Request status: 200, 400, 401, 403, 500, … (14 distinct)
- Customer: 3300 distinct customers

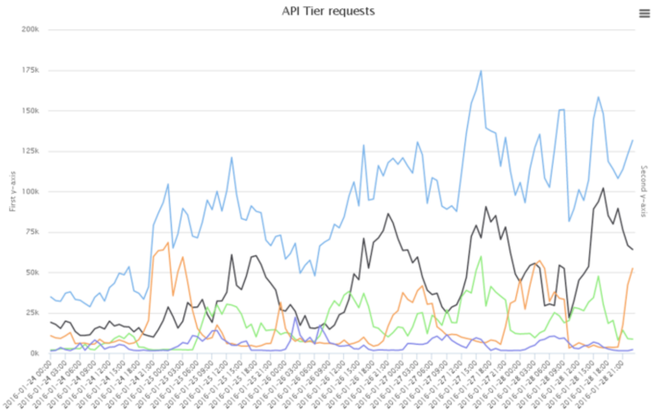
For each time interval we use ~50,000 buckets.

Each bucket has:

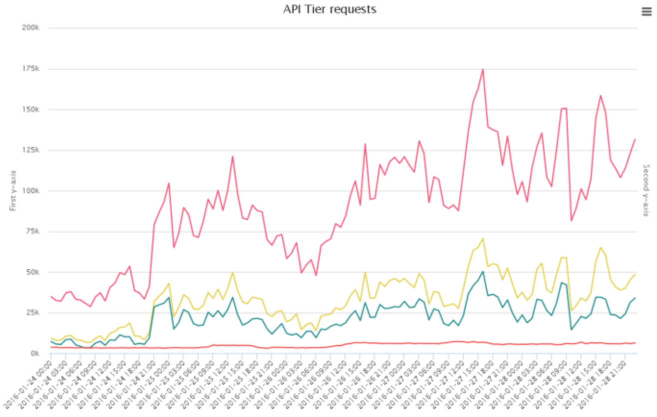
- Request Count
- Response times: min/max/average/median/90%/99%

Will reduce to 500 distinct requests

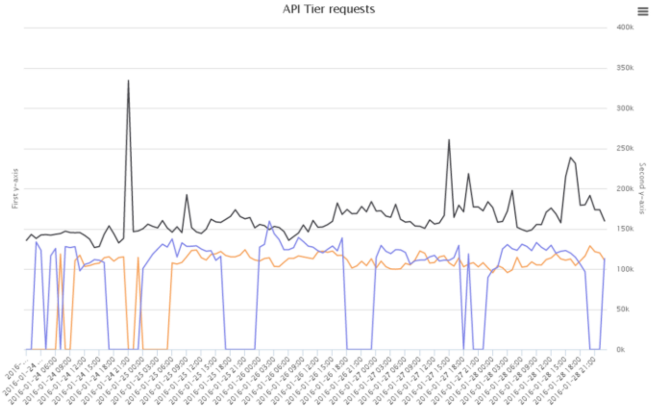
Example: Request count by Datacentre



Example: Request count per request type



Example: GetServer response time 99%-ile by datacenter



Example: Storage

- 5-6 GB zipped access_log files per month
- ~40 GB unzipped access_log files per month
- ~120 million requests per month

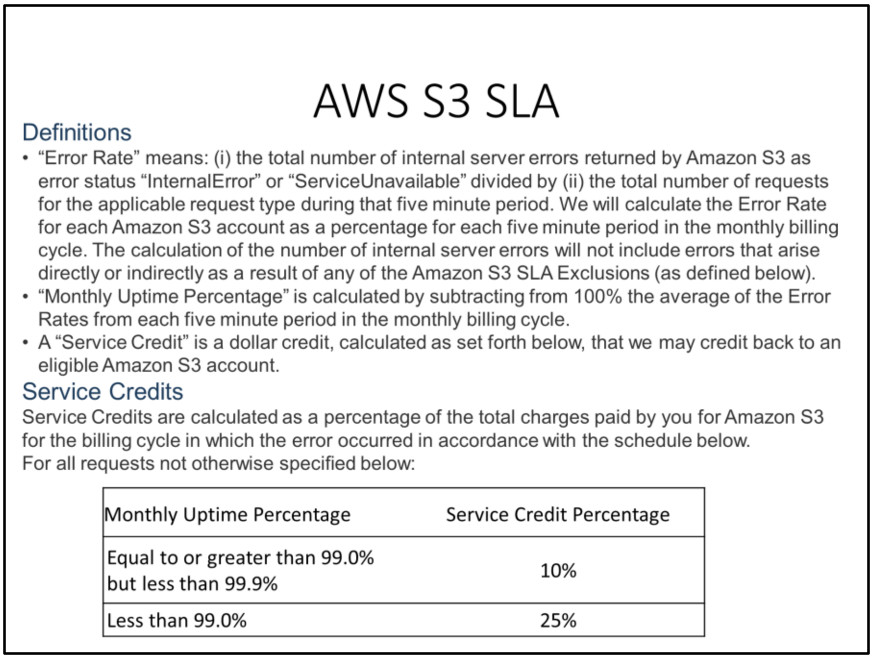
Aggregated database contents

- Daily: 200 MB per month
- Hourly: 1 GB per month

Service Level Agreement

Promise to your users

- Availability
- Reliability
- Responsiveness



AWS S3 SLA

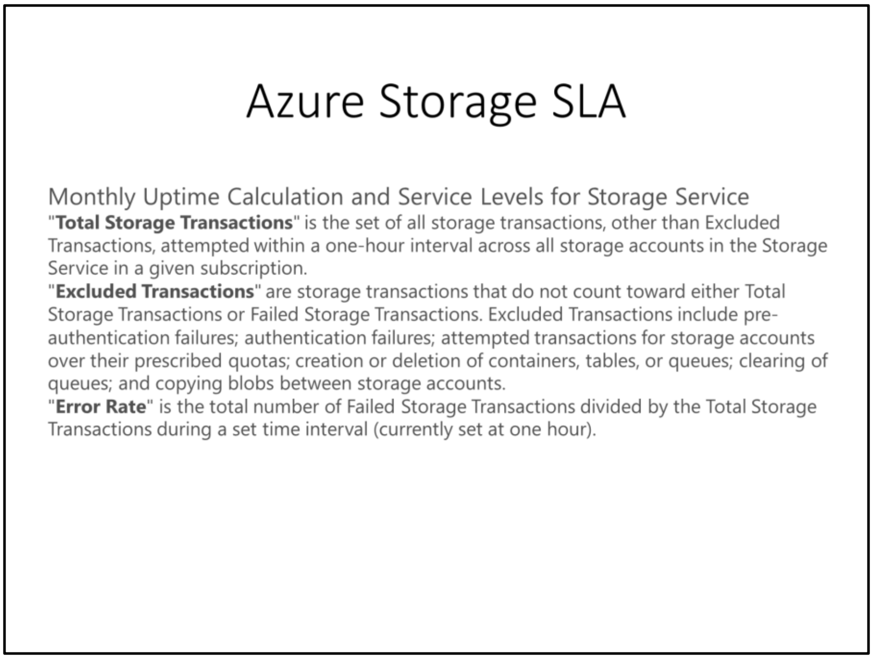
Definitions

- “Error Rate” means: (i) the total number of internal server errors returned by Amazon S3 as error status “InternalError” or “ServiceUnavailable” divided by (ii) the total number of requests for the applicable request type during that five minute period. We will calculate the Error Rate for each Amazon S3 account as a percentage for each five minute period in the monthly billing cycle. The calculation of the number of internal server errors will not include errors that arise directly or indirectly as a result of any of the Amazon S3 SLA Exclusions (as defined below).
- “Monthly Uptime Percentage” is calculated by subtracting from 100% the average of the Error Rates from each five minute period in the monthly billing cycle.
- A “Service Credit” is a dollar credit, calculated as set forth below, that we may credit back to an eligible Amazon S3 account.

Service Credits

Service Credits are calculated as a percentage of the total charges paid by you for Amazon S3 for the billing cycle in which the error occurred in accordance with the schedule below.
For all requests not otherwise specified below:

Monthly Uptime Percentage	Service Credit Percentage
Equal to or greater than 99.0% but less than 99.9%	10%
Less than 99.0%	25%



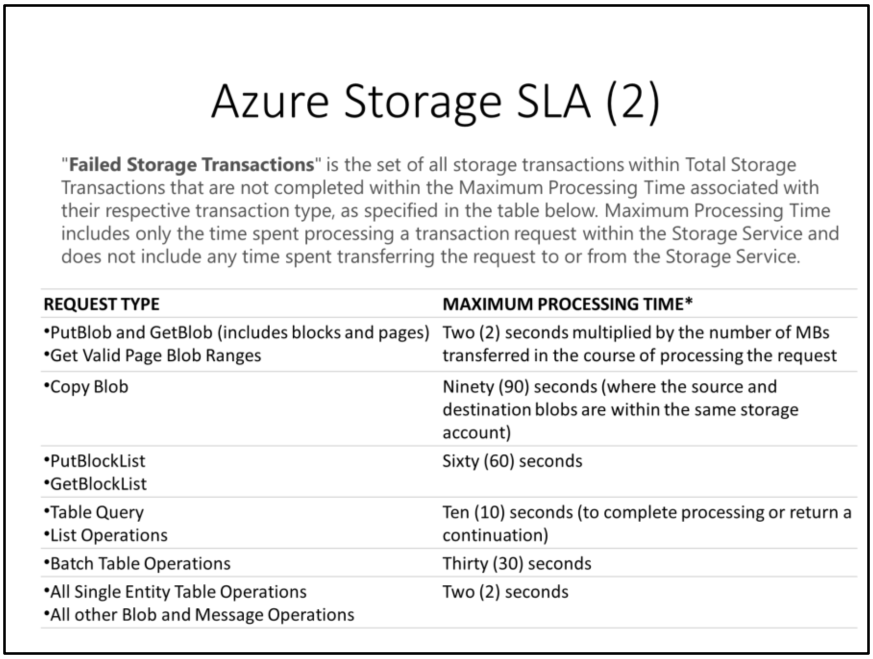
Azure Storage SLA

Monthly Uptime Calculation and Service Levels for Storage Service

"**Total Storage Transactions**" is the set of all storage transactions, other than Excluded Transactions, attempted within a one-hour interval across all storage accounts in the Storage Service in a given subscription.

"**Excluded Transactions**" are storage transactions that do not count toward either Total Storage Transactions or Failed Storage Transactions. Excluded Transactions include pre-authentication failures; authentication failures; attempted transactions for storage accounts over their prescribed quotas; creation or deletion of containers, tables, or queues; clearing of queues; and copying blobs between storage accounts.

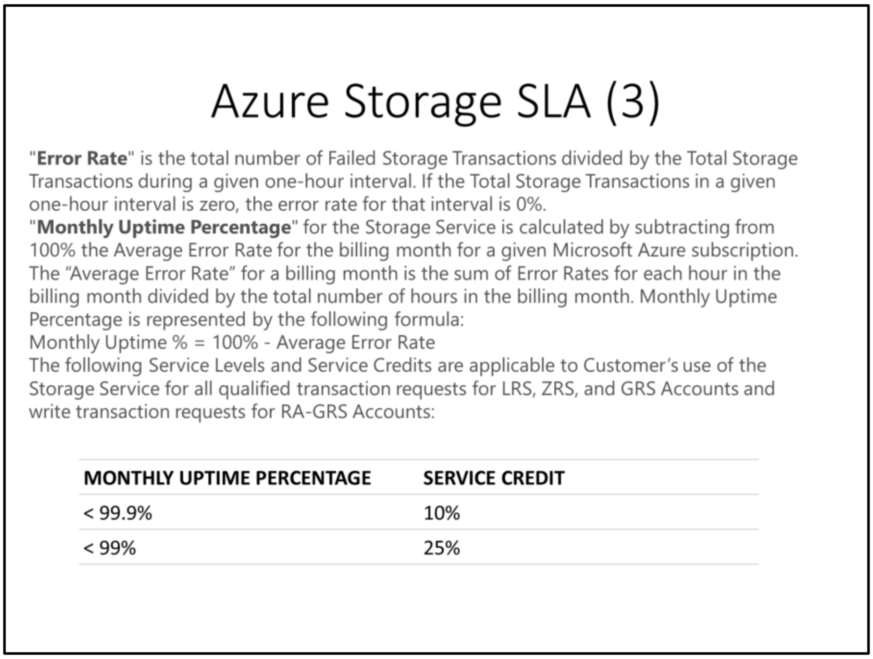
"**Error Rate**" is the total number of Failed Storage Transactions divided by the Total Storage Transactions during a set time interval (currently set at one hour).



Azure Storage SLA (2)

"**Failed Storage Transactions**" is the set of all storage transactions within Total Storage Transactions that are not completed within the Maximum Processing Time associated with their respective transaction type, as specified in the table below. Maximum Processing Time includes only the time spent processing a transaction request within the Storage Service and does not include any time spent transferring the request to or from the Storage Service.

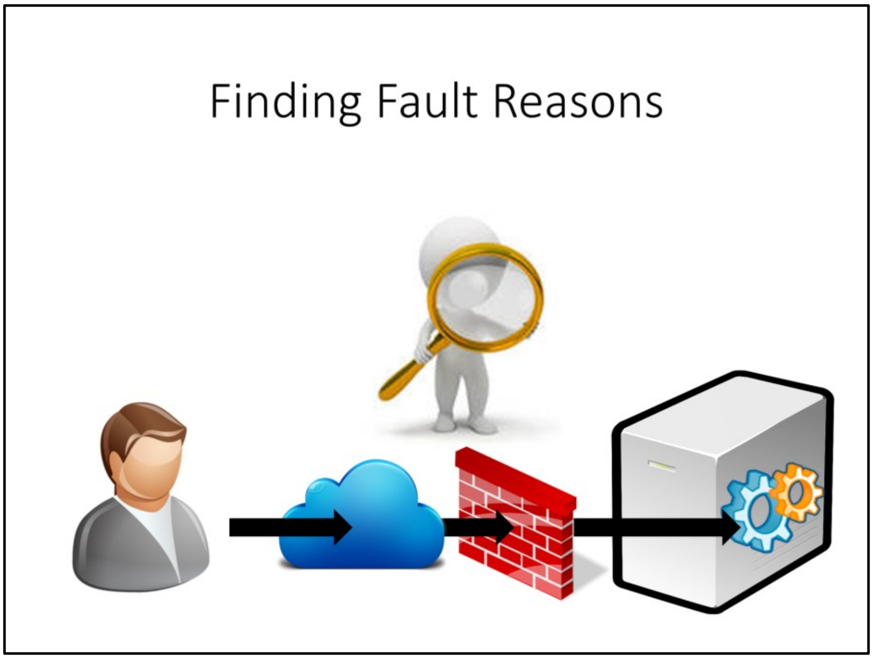
REQUEST TYPE	MAXIMUM PROCESSING TIME*
•PutBlob and GetBlob (includes blocks and pages) •Get Valid Page Blob Ranges	Two (2) seconds multiplied by the number of MBs transferred in the course of processing the request
•Copy Blob	Ninety (90) seconds (where the source and destination blobs are within the same storage account)
•PutBlockList •GetBlockList	Sixty (60) seconds
•Table Query •List Operations	Ten (10) seconds (to complete processing or return a continuation)
•Batch Table Operations	Thirty (30) seconds
•All Single Entity Table Operations •All other Blob and Message Operations	Two (2) seconds



Azure Storage SLA (3)

"**Error Rate**" is the total number of Failed Storage Transactions divided by the Total Storage Transactions during a given one-hour interval. If the Total Storage Transactions in a given one-hour interval is zero, the error rate for that interval is 0%.
"**Monthly Uptime Percentage**" for the Storage Service is calculated by subtracting from 100% the Average Error Rate for the billing month for a given Microsoft Azure subscription. The "Average Error Rate" for a billing month is the sum of Error Rates for each hour in the billing month divided by the total number of hours in the billing month. Monthly Uptime Percentage is represented by the following formula:
Monthly Uptime % = 100% - Average Error Rate
The following Service Levels and Service Credits are applicable to Customer's use of the Storage Service for all qualified transaction requests for LRS, ZRS, and GRS Accounts and write transaction requests for RA-GRS Accounts:

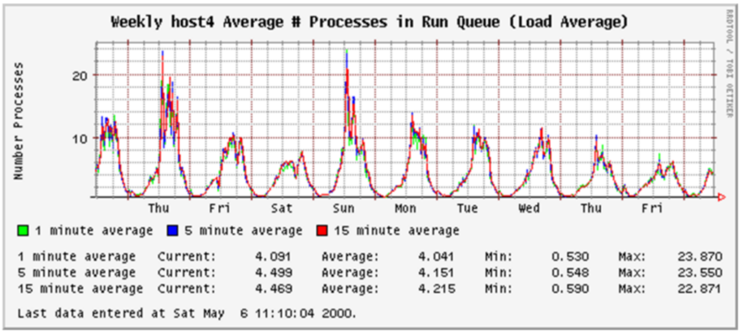
MONTHLY UPTIME PERCENTAGE	SERVICE CREDIT
< 99.9%	10%
< 99%	25%



Finding Fault Reasons

Air industry has long experience in investigating crashes.
Manage to find root cause from small pieces of evidence.
Corrective actions to reduce risk of reoccurrence.
Flight recorders have improved over the years.

System Monitoring



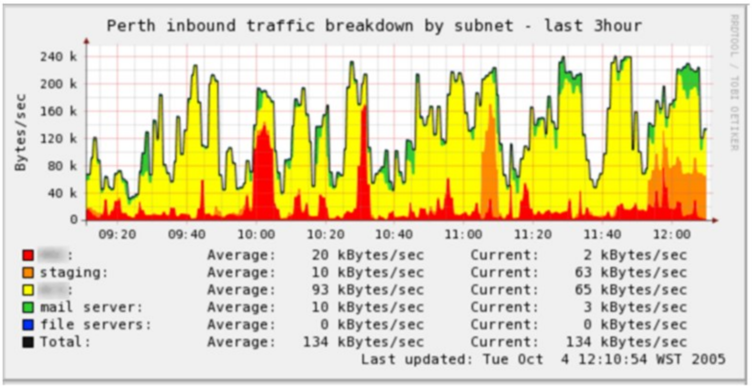
```
1 minute average  Current:     4.091    Average:     4.041    Min:     0.530    Max:     23.870
5 minute average  Current:     4.499    Average:     4.151    Min:     0.548    Max:     23.550
15 minute average Current:     4.469    Average:     4.215    Min:     0.590    Max:     22.871

Last data entered at Sat May  6 11:10:04 2000.
```

Screenshot from Nagios

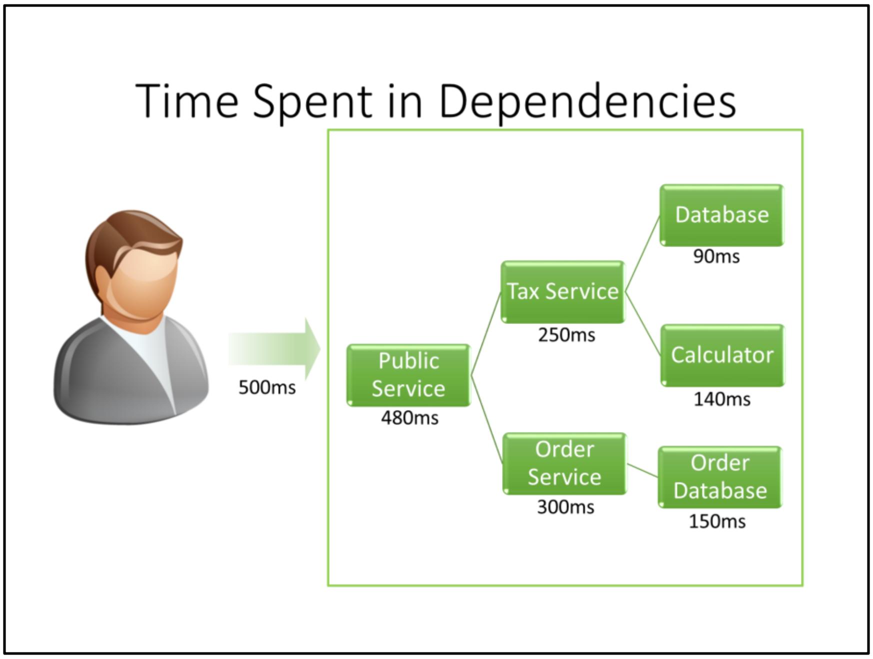
See also www.icinga.org, a fork of Nagios with a modern style UI.

Network Monitoring



```
🟥 [redacted]:          Average:   20 kBytes/sec     Current:    2 kBytes/sec
🟧 staging:             Average:   10 kBytes/sec     Current:   63 kBytes/sec
🟨 [redacted]:          Average:   93 kBytes/sec     Current:   65 kBytes/sec
🟩 mail server:         Average:   10 kBytes/sec     Current:    3 kBytes/sec
🟦 file servers:        Average:    0 kBytes/sec     Current:    0 kBytes/sec
⬛ Total:               Average:  134 kBytes/sec     Current:  134 kBytes/sec
                                        Last updated: Tue Oct  4 12:10:54 WST 2005
```

Screenshot from Cacti



Time Spent in Dependencies

500ms

Public Service
480ms

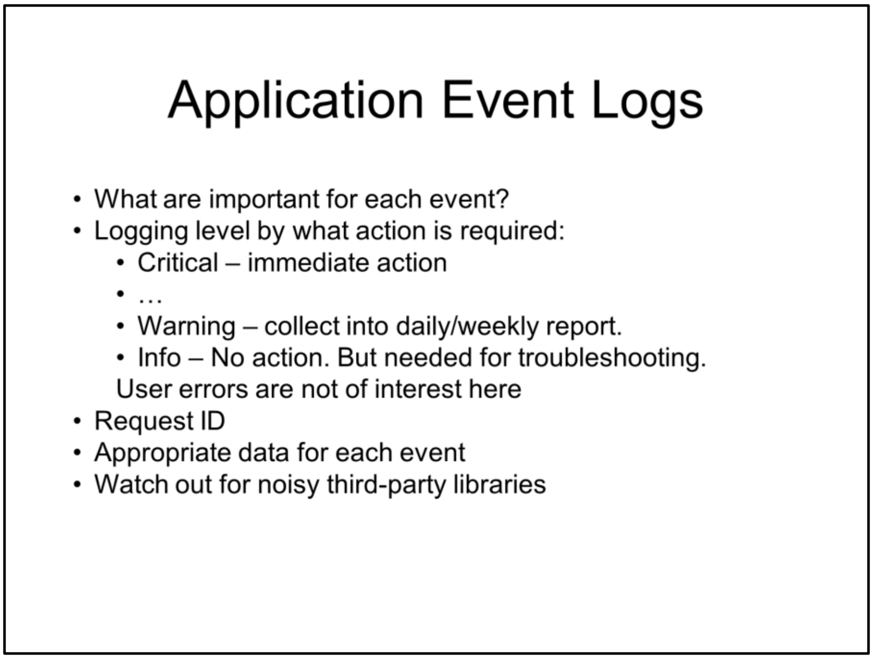
Tax Service
250ms

Database
90ms

Calculator
140ms

Order Service
300ms

Order Database
150ms



Application Event Logs

- What are important for each event?
- Logging level by what action is required:
  - Critical – immediate action
  - …
  - Warning – collect into daily/weekly report.
  - Info – No action. But needed for troubleshooting.

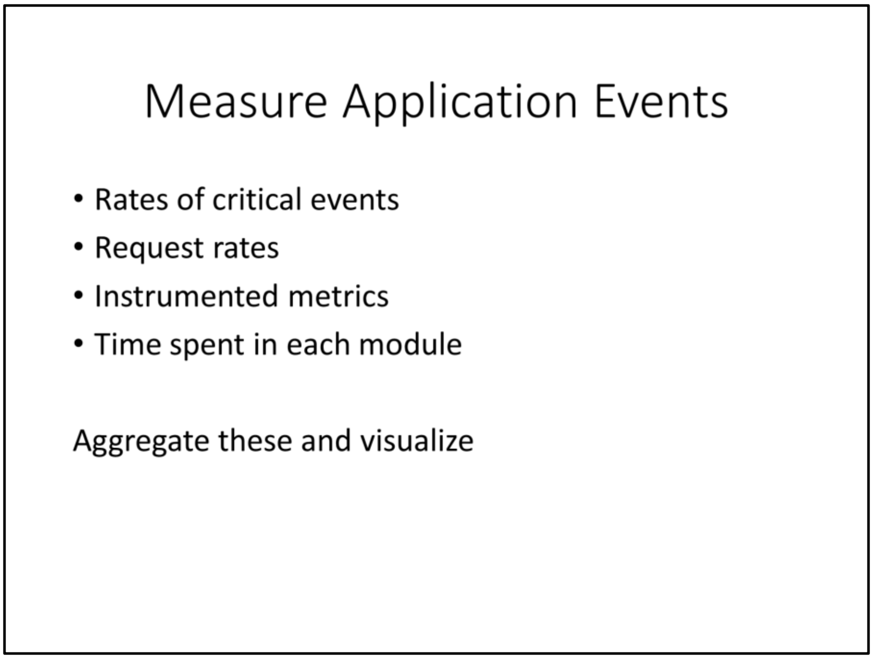
  User errors are not of interest here
- Request ID
- Appropriate data for each event
- Watch out for noisy third-party libraries

(You expected me to talk a lot about this, right?)

Application Event Log Store

- Text files
- Syslog
- Windows Events
- Database

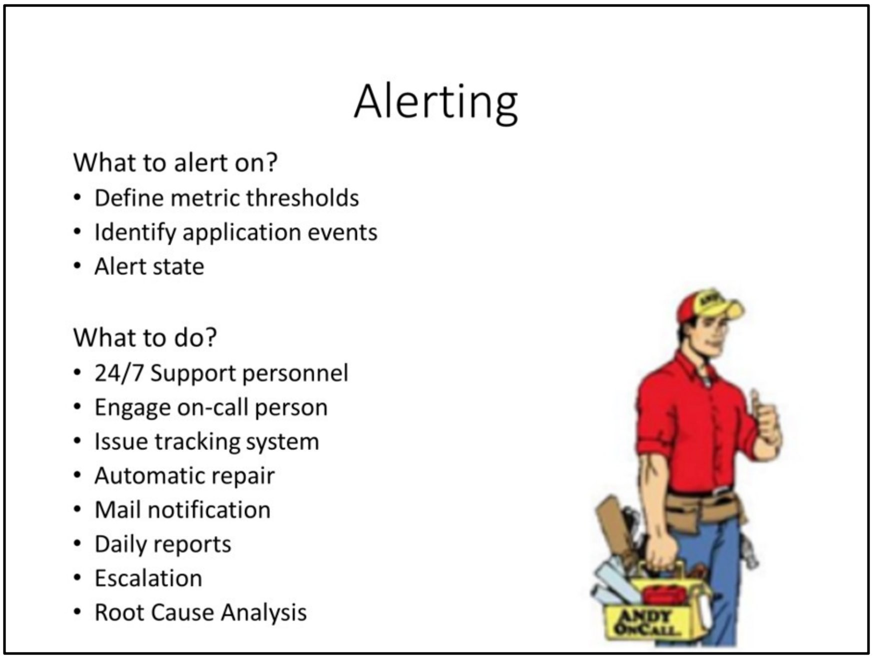
- Store in accessible locations
- Live access



Measure Application Events

- Rates of critical events
- Request rates
- Instrumented metrics
- Time spent in each module

Aggregate these and visualize



Alerting

What to alert on?

- Define metric thresholds
- Identify application events
- Alert state

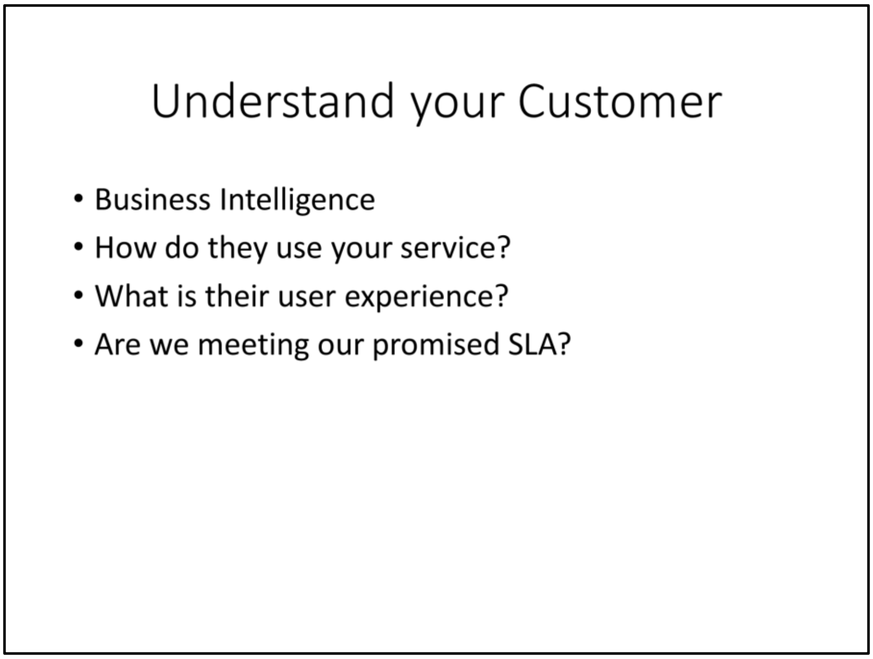
What to do?

- 24/7 Support personnel
- Engage on-call person
- Issue tracking system
- Automatic repair
- Mail notification
- Daily reports
- Escalation
- Root Cause Analysis

ANDY
ONCALL

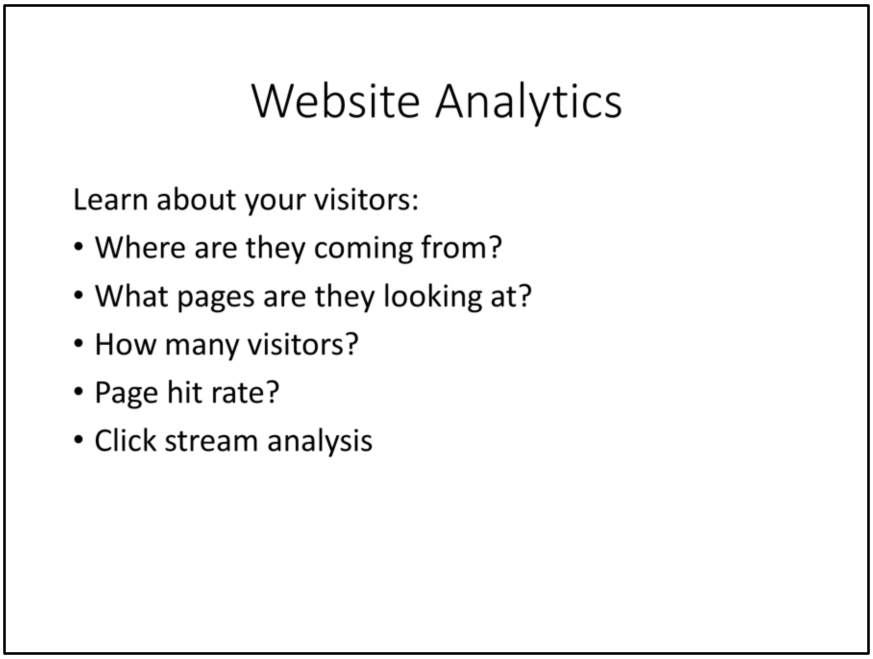
Monitoring for the Future

- Critical resources?
- Trend analysis
- Improve customer experience



Understand your Customer

- Business Intelligence
- How do they use your service?
- What is their user experience?
- Are we meeting our promised SLA?



Website Analytics

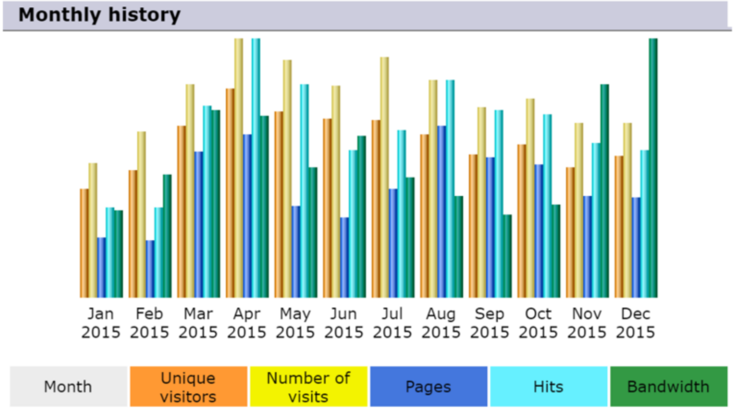
Learn about your visitors:

- Where are they coming from?
- What pages are they looking at?
- How many visitors?
- Page hit rate?
- Click stream analysis

Relates to Business Intelligence.

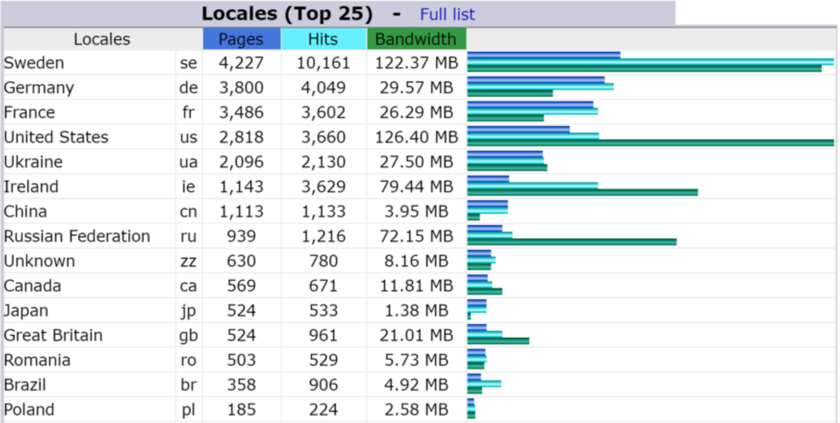
??? Maybe move these slides to the end.

AWStats (1)



Monthly history

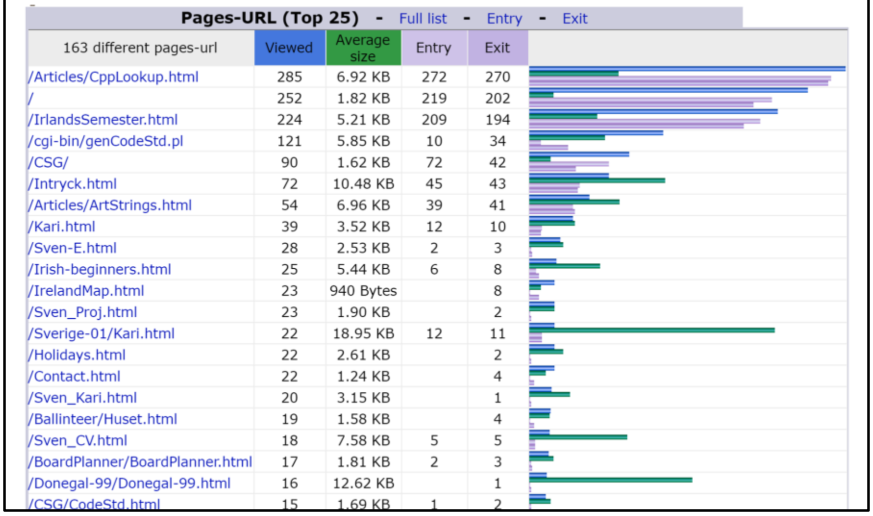
AWStats (2)



Locales (Top 25) - Full list

Locales		Pages	Hits	Bandwidth	
Sweden	se	4,227	10,161	122.37 MB	
Germany	de	3,800	4,049	29.57 MB	
France	fr	3,486	3,602	26.29 MB	
United States	us	2,818	3,660	126.40 MB	
Ukraine	ua	2,096	2,130	27.50 MB	
Ireland	ie	1,143	3,629	79.44 MB	
China	cn	1,113	1,133	3.95 MB	
Russian Federation	ru	939	1,216	72.15 MB	
Unknown	zz	630	780	8.16 MB	
Canada	ca	569	671	11.81 MB	
Japan	jp	524	533	1.38 MB	
Great Britain	gb	524	961	21.01 MB	
Romania	ro	503	529	5.73 MB	
Brazil	br	358	906	4.92 MB	
Poland	pl	185	224	2.58 MB	

AWStats (3)

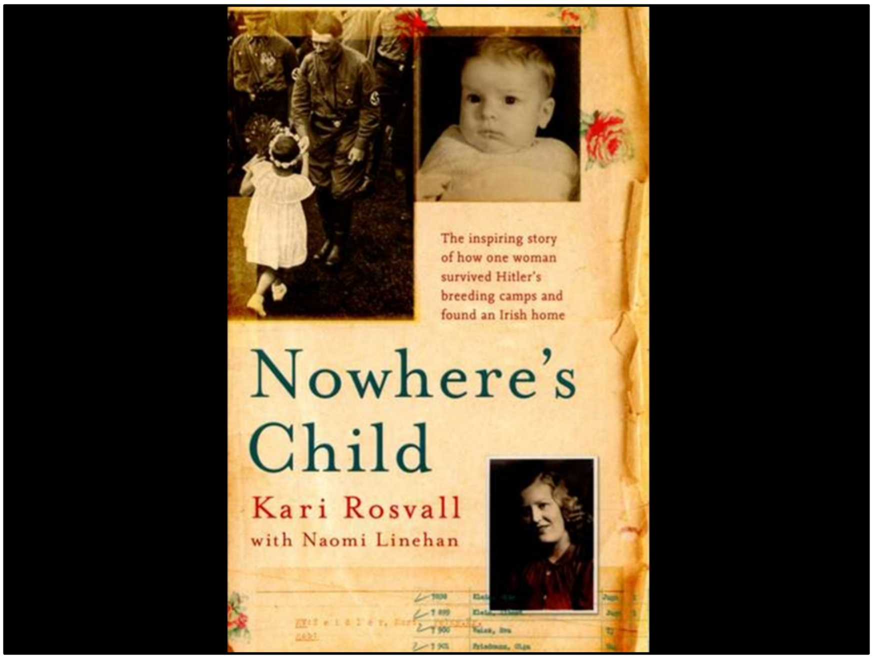


Pages-URL (Top 25) - Full list - Entry - Exit

163 different pages-url	Viewed	Average size	Entry	Exit
/Articles/CppLookup.html	285	6.92 KB	272	270
/	252	1.82 KB	219	202
/IrlandsSemester.html	224	5.21 KB	209	194
/cgi-bin/genCodeStd.pl	121	5.85 KB	10	34
/CSG/	90	1.62 KB	72	42
/Intryck.html	72	10.48 KB	45	43
/Articles/ArtStrings.html	54	6.96 KB	39	41
/Kari.html	39	3.52 KB	12	10
/Sven-E.html	28	2.53 KB	2	3
/Irish-beginners.html	25	5.44 KB	6	8
/IrelandMap.html	23	940 Bytes		8
/Sven_Proj.html	23	1.90 KB		2
/Sverige-01/Kari.html	22	18.95 KB	12	11
/Holidays.html	22	2.61 KB		2
/Contact.html	22	1.24 KB		4
/Sven_Kari.html	20	3.15 KB		1
/Ballinteer/Huset.html	19	1.58 KB		4
/Sven_CV.html	18	7.58 KB	5	5
/BoardPlanner/BoardPlanner.html	17	1.81 KB	2	3
/Donegal-99/Donegal-99.html	16	12.62 KB		1
/CSG/CodeStd.html	15	1.69 KB	1	2

Common to split articles into several pages to count users who read the whole article.

Most popular pages on my website.
Note that my wife’s page is more popular than mine.

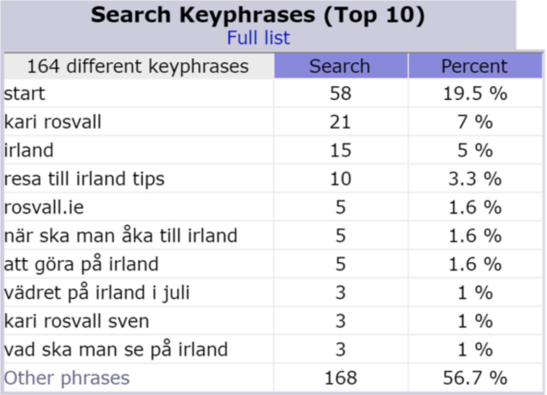


The inspiring story of how one woman survived Hitler's breeding camps and found an Irish home

Nowhere's Child

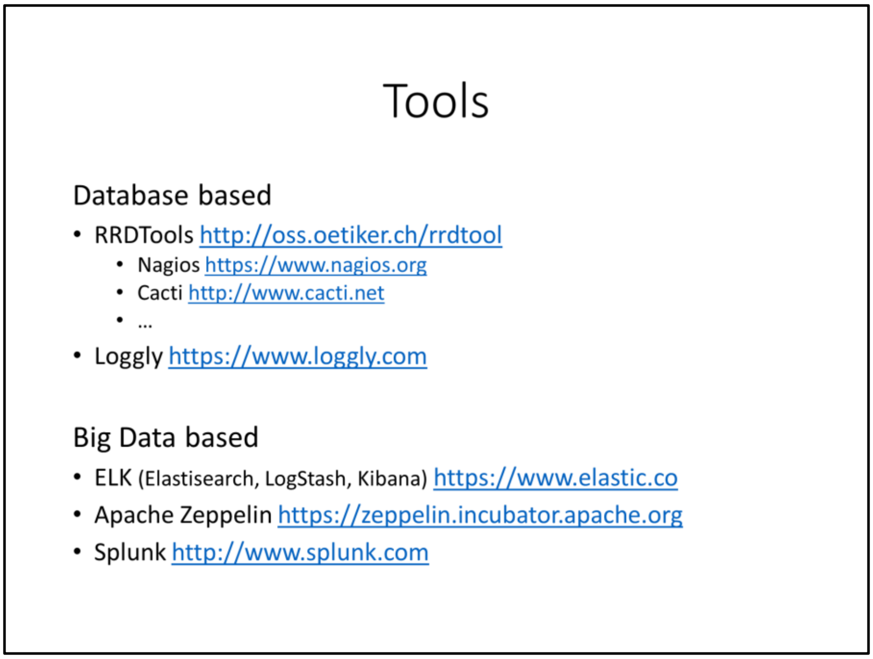
Kari Rosvall
with Naomi Linehan

AWStats (4)



Search Keyphrases (Top 10)
Full list

164 different keyphrases	Search	Percent
start	58	19.5 %
kari rosvall	21	7 %
irland	15	5 %
resa till irland tips	10	3.3 %
rosvall.ie	5	1.6 %
när ska man åka till irland	5	1.6 %
att göra på irland	5	1.6 %
vädret på irland i juli	3	1 %
kari rosvall sven	3	1 %
vad ska man se på irland	3	1 %
Other phrases	168	56.7 %



Tools

Database based

- RRDTools http://oss.oetiker.ch/rrdtool
  - Nagios https://www.nagios.org
  - Cacti http://www.cacti.net
  - …
- Loggly https://www.loggly.com

Big Data based

- ELK (Elastisearch, LogStash, Kibana) https://www.elastic.co
- Apache Zeppelin https://zeppelin.incubator.apache.org
- Splunk http://www.splunk.com