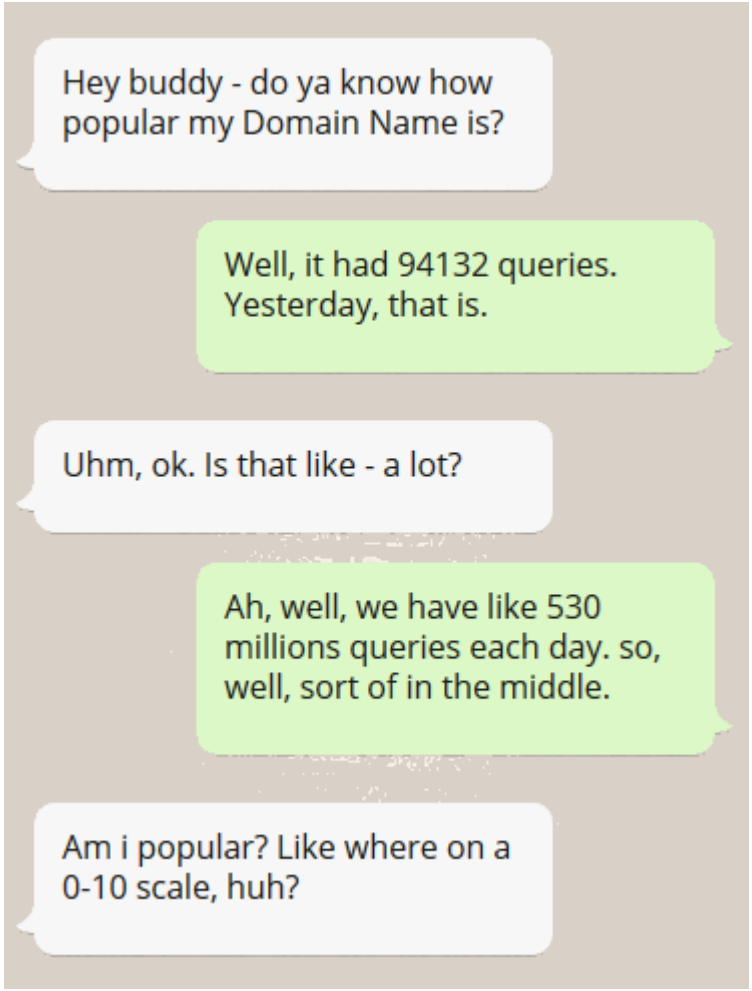
DNS Magnitude

„How popular is this Domain?“

yet another (DNS based) **approach**

Alexander Mayrhofer
Head of R&D

Motivation



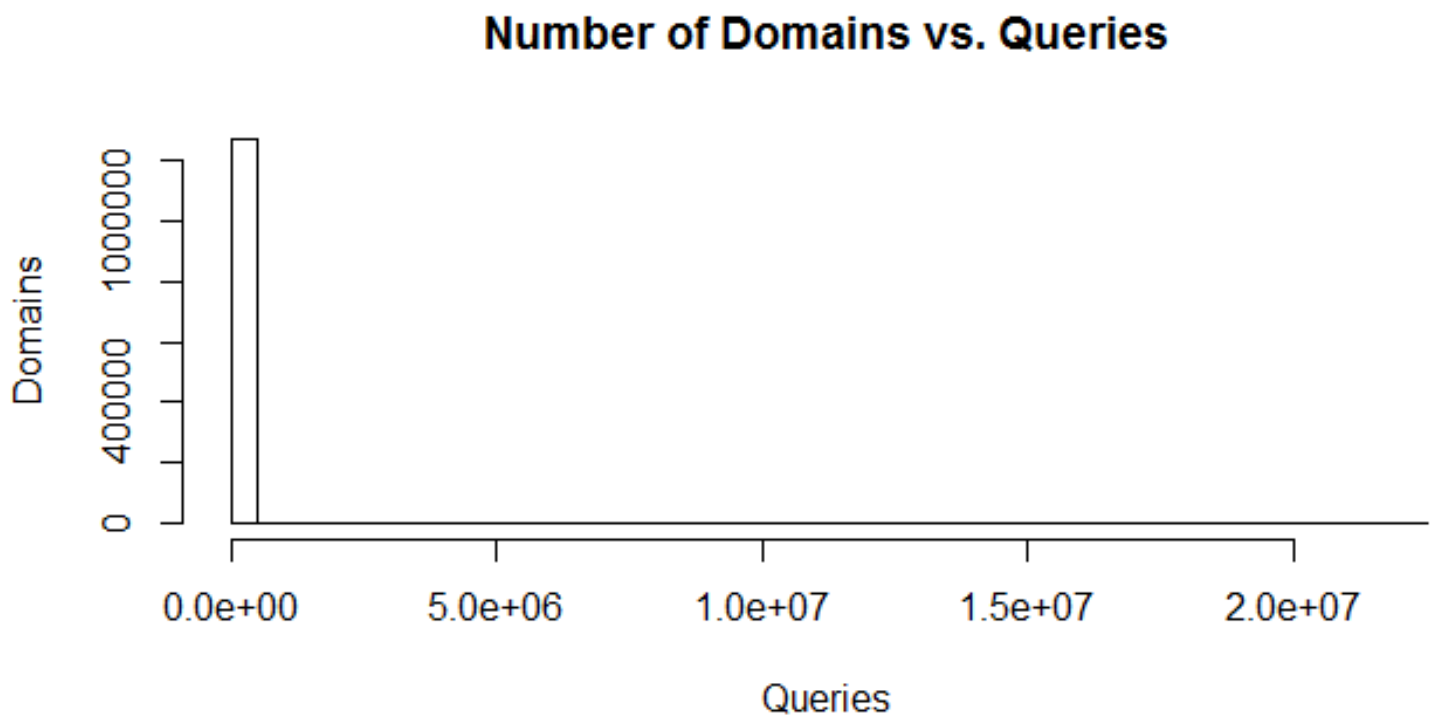
- Single, easy to understand „popularity“ figure
- Based on DNS statistics
  (because that‘s what we have?)
- Copy „Earthquake magnitude“ figures
  (because everybody knows them)
- **„DNS Magnitude“?**

DNS Data Exploration

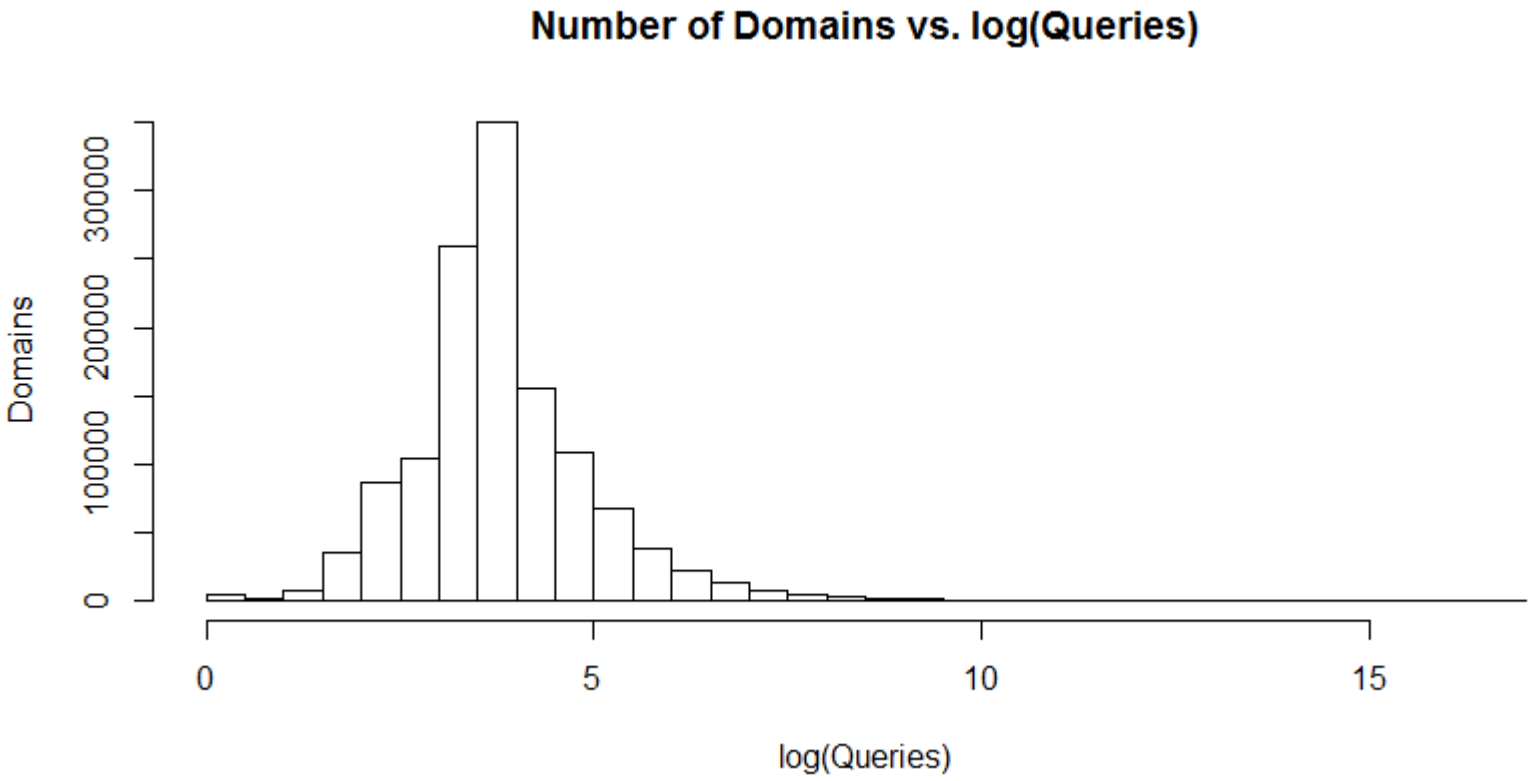
- Basis: DNS „query impact“ of a domain
  - Assumption: Popular (..) domain -> higher query rate
- Single day: ~450 million queries
  - About 20% NXDOMAINS (not considered)
  - Queries for almost all existing domains
  - Problem: Extremely high disparity

„queries by domain“ disparity

- Top 1% of domains: **62%** of queries



Logarithmic Scale?



- Looks more „natural“!
- Earthquake magnitudes use logarithmic scales too

Limit Scale to 0-10?

- Definition: Magnitude 10 = all queries on single Domain
  - Example: $0 < \ln(Q_{Dx}) < 16.91$
  - Scale to ln(totalqueries)
- Hence:

$$mag_{Dx} = \frac{ln(Q_{Dx})}{ln(\sum_{k=1}^{n} Q_{Dk})} * 10$$

First try… Queries-based

- Dominated by infrastructure domains
- TTL has a big impact!

domain	queries	query_mag	
anexia.at	22124665	**8.678725**	<- ISP, low TTL (120s!)
univie.ac.at	20824366	**8.647643**	<- auth. Servers for .at
telekom.at	3573045	**7.743087**	<- ISP
ns.at	3398512	**7.717387**	<- auth. Servers for .at
nessus.at	3031900	**7.658810**	<- Registrar
chello.at	1613822	**7.335218**	<- ISP
internic.at	1391180	**7.259037**	<- Registrar
at	1240702	**7.200293**	<- zone apex
t-systems.at	1055778	**7.117468**	<- ISP
inode.at	1027223	**7.103398**	<- ISP

How to get around TTL impact?

- TTL expiration triggers query from same source IP address
- Approach: Count unique resolvers rather than queries
  - No matter if they query a domain once or 1000 times per day
- New basis: Number of distinct src IP addresses per domain

Hosts based top10 – Better…

- TTL effect seems reduced
- Still dominated by infrastructure zones

	domain	queries	hosts	query_mag	host_mag	
1	univie.ac.at	20824366	394542	8.647643	**9.401667**	
2	telekom.at	3573045	223838	7.743087	**8.988109**	
3	chello.at	1613822	183470	7.335218	**8.843006**	
4	nessus.at	3031900	167832	7.658810	**8.778005**	
5	inode.at	1027223	134049	7.103398	**8.614014**	
6	regdns5.at	830090	132637	6.994053	**8.606288**	<- TTL 10800
7	ns.at	3398512	128279	7.717387	**8.581912**	
8	google.at	724264	124449	6.924069	**8.559796**	<- TTL 10800
9	anexia.at	22124665	118241	8.678725	**8.522460**	<- TTL 120
10	nic.at	623485	118055	6.847181	**8.521311**	<- TTL 900

DNS Magnitude

- Current working definition

$$mag_{Dx} = \frac{ln(H_{Dx})}{ln(H_{tot})} * 10$$

Go for services? Web:

- A/AAAA record and www.% or origin
  - Total 44M queries, 397k hosts (1 day)

	domain	queries	hosts	query_mag	host_mag
1	google.at	398699	105154	7.323973	**8.968340**
2	ebay.at	234151	72845	7.021699	**8.683625**
3	tripadvisor.at	209471	48626	6.958443	**8.370149**
4	airbnb.at	254649	48373	7.069360	**8.366103**
5	yelp.at	146933	41204	6.757051	**8.241693**
6	groupon.at	125715	36463	6.668477	**8.146886**
7	vistaprint.at	110861	29375	6.597066	**7.979238**
8	gmx.at	59330	27845	6.242019	**7.937751**
9	transfermarkt.at	88722	27689	6.470549	**7.933394**
10	kriesi.at	82103	27248	6.426516	**7.920942**

Some examples („web“ based)

- amazon.at 7,8 (13)
- orf.at 6,5 (240)
- google.at 9,0 (1)
- nic.at 6,1 (489)
- phosaigon.at 3,5 (39118)
- „nearlyunused“.at 0,6 (632673)
- post.at 6,8 (100)

Current (early) applications

- Internal „BI“ panel

Domain	Flags	Mag	Transaction
[redacted]at €7.50	SPT	2.5	billwithdraw-z
[redacted]or.at €1.50	SPT	3.2	billwithdraw-z
[redacted].at €30.28	SPT	4.5	billwithdraw-z
[redacted]at €7.50	SPT	2.8	billwithdraw-z
[redacted]bau.at €15.00	SPT	3.9	billwithdraw-z
restaura[redacted].at €30.00	SPT	3.6	billwithdraw-z
[redacted]music.at €7.50	SPT	2.8	billwithdraw-z
[redacted]ety.at	SPT		billwithdraw-z
[redacted]at €6.00	SPT	3.4	billwithdraw-z
[redacted]at €10.00		3.3	billwithdraw

somedomain.at 6.5

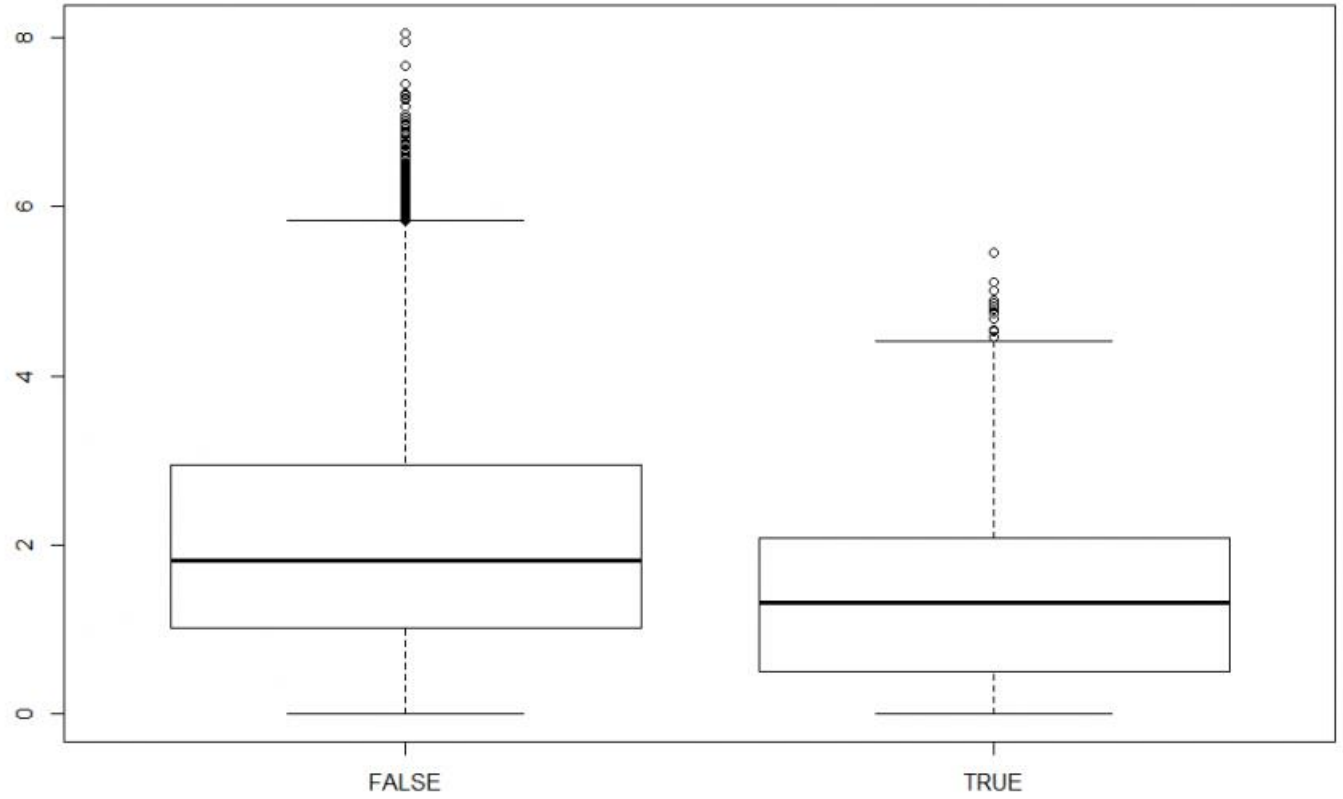
3.7

NXDOMAINs

Top NXDOMAINs

Rank	Domain	Magnitude
#1	[redacted]ac.at	6,4
#2	sex[redacted].at	6,4
#3	[redacted].at	6,3
#4	[redacted].at	6,2
#5	[redacted].at	6,1
#6	id[redacted].at	6,0
#7	[redacted].ac.at	6,0
#8	[redacted].co.at	6,0
#9	[redacted].at	6,0
#10	installati[redacted].at	5,9

Application – Delete propensity



- Correlation lower than expected
- But no domain deleted with mag > 5.8!
- Delete Prediction: Input to a neural network (WIP)

Tools used

- ENTRADA/Hadoop (Storage)
- Impala (SQL-Queries)
- R (prototyping)
  - PHP for production (shhh, don‘t tell anybody! ;)
- Results stored in Redis
- Airflow for Orchestration

- ~300 lines of code in total

Further work

- Refine algorithm (a-z query clients, „long tail“ scale)
- NZRS work, Alexa 1M, Umbrella Top 1M list
- Study impact of DNS parameters
  - TTL
  - Prefetching
  - Future: QNAME minimization?
- ISP recursive resolvers – better vantage point?

Thanks for listening!

- Questions? Suggestions?
- alexander.mayrhofer@nic.at