



Search without the Database

paul reinheimer

@preinheimer 1



Paul Reinheimer



Mansef Group

We're Hiring - Here in Montréal

Case Study

Several Million unique visitors per
day

Very few "hot" search terms

Moderate dataset

MySQL Database backend



Problem

- ▶ MySQL Slow
 - ▶ No Stemming
 - ▶ No word positioning analysis
 - ▶ Short words
- ▶ Additional Business Requirements
- ▶ Fast is good



Getting Data In

id	title	year	rating
1	TMNT	1988	9.5
2	Goonies	1989	9.0
3	The Postman	1992	3.5
4	Alien	1994	8.5

id	title	year	rating	studio	category
1	TMNT	1988	9.5	x	action
2	Goonies	1989	9.0	y	adventure
3	The Postman	1992	3.5	z	boring
4	Alien	1994	8.5	a	awesome

film_id	tag_id
1	2
2	3
3	4
4	5

id	director
1	blue
2	red
3	green
4	orange

film_id	description
1	Four awesome turtles...
2	Friends, pirates, goonies.
3	Nap
4	They tried to turn around

id	title	year	rating	studio	category
1	TMNT	1988	9.5	x	action
2	Goonies	1989	9.0	y	adventure
3	The Postman	1992	3.5	z	boring
4	Alien	1994	8.5	a	awesome

id	actor
1	Mel Gibson
2	Shredder
3	Sigourney Weaver

id	tag
1	kids
2	awesome
3	sci-fi

id	producer
1	1
2	2
3	3
4	4

film_id	tag_id	id	director	description
1	1	1	1	1
2	1	2	1	1
3	2	3	2	2
4	3	4	3	3
	4		4	4

id	title	year	rating	studio	category
1	TMNT	1988	9.5	x	action
2	Goonies	1989	9.0	y	adventure
3	The Postman	1992	3.5	z	boring
4	Alien	1994	8.5	a	awesome

id	actor	id	tag	id	producer
1	1	1	1	1	1
2	1	2	2	2	1
3	2	3	3	3	2
	3	4	4	4	3
	4				4

```

SELECT
  `dvd_scenes`.`scene_id`,
  `dvds`.`dvd_id`,
  `dvds`.`name`,
  `dvds`.`name`          AS sort_title,
  `dvds`.`featured`      AS premium,
  `dvds`.`description`,
  `dvds`.`date_released`,
  `dvds`.`channel_id`,
  `dvds`.`studio_id`,
  `dvd_scenes_dyn`.`times_viewed` AS sub_order,
  `dvd_scenes_dyn`.`rating`      AS rating,
  `channels`.`name`            AS channel_name,
  `studios`.`name`             AS studio_name,
  `channels_keywords`.`keywords` AS channel_keywords,
  GROUP_CONCAT(DISTINCT `categories`.`name` ORDER BY `categories`.`name` DESC SEPARATOR ', ') AS categories_list,
  GROUP_CONCAT(DISTINCT `models`.`name` ORDER BY `models`.`name` DESC SEPARATOR ', ') AS models_list
FROM `dvd_scenes`
  INNER JOIN dvds
    ON dvds.dvd_id = dvd_scenes.dvd_id
  INNER JOIN dvd_scenes_dyn
    ON dvd_scenes_dyn.id = dvd_scenes.scene_id
  LEFT JOIN scenes_categories
    ON dvd_scenes.scene_id = scenes_categories.scene_id
  LEFT JOIN categories
    ON `scenes_categories`.`category_id` = categories.category_id
  LEFT JOIN studios
    ON studios.studio_id = dvds.studio_id
  LEFT JOIN scenes_models
    ON scenes_models.scene_id = dvd_scenes.scene_id
  LEFT JOIN `channels`
    ON channels.channel_id = dvds.channel_id
  LEFT JOIN `channels_keywords`
    ON dvds.channel_id = channels_keywords.channel_id
  LEFT JOIN `models`
    ON scenes_models.model_id = models.model_id
WHERE dvds.is_active = 1
  AND dvd_scenes.is_active = 1
  AND dvds.date_released < UNIX_TIMESTAMP()
  AND dvds.date_released > 0
  AND dvds.channel_id NOT IN (21,26)
GROUP BY scene_id

```



Import Tips

Group Concat

id	tag
1	kids
2	awesome
3	sci-fi
4	adventure

film_id	tag_id
1	1, 4
2	2, 1
3	3, 1
4	4, 2



kids, adventure

Retrieve Repeatedly

```
SELECT `name`, `name` AS `name2` FROM `table`  
      WHERE `clause` = 'value';
```

Filter Results

```
SELECT `name`, `name` AS `name2` FROM  
`table` WHERE `clause` = 'value';
```

```
if(isset($_GET['noprem'])) {  
    $searchClient->setFilter('p', array(0));  
}elseif(isset($_GET['premonly'])) {  
    $searchClient->setFilter('p', array(1));  
}
```



Sphinx

<http://www.flickr.com/photos/waldenpond/4237758518/>

sphinx.conf

source

index

Indexing Information

```
indexer --all --rotate
```

Indexing Information

Data

Data

Appending Information

Data

Testing

search puppies

Starting the Demon

searchd

Extra Tips

- ▶ Schedule Index regeneration and rotation
- ▶ Rotate your logs
- ▶ Make sure it keeps running



Querying with PHP

Sphinx Object

```
$searchClient = newSphinxClient ($sphinx_host,  
                                $sphinx_port);
```

Weights

```
$weights = array(  
    'name' => 2,  
    'categories_list' => 3,  
    'models_list' => 2,  
);  
$searchClient->SetFieldWeights($weights);
```

SPH_SORT_EXPR

```
$cl->SetSortMode ( SPH_SORT_EXPR,  
"@weight + ( user_karma + ln(pageviews) )*0.1" );
```

Sort Method

```
switch ($sortMethod) {
  case 'relevancy':
    $sortMode = sprintf('@weight DESC,sub_order %s', $sortway);
    break;
  case 'name':
    $sortMode = sprintf('sort_title %s', $sortway);
    break;
  case 'date':
    $sortMode = sprintf('date_released %s', $sortway);
    break;
  case 'views':
    $sortMode = sprintf('sub_order %s', $sortway);
    break;
  case 'rate':
    $sortMode = sprintf('rating %s', $sortway);
    break;
}
```

Assign Sort & Pagination

```
$searchClient->SetSortMode(SPH_SORT_EXTENDED,  
    $sortMode);  
$searchClient->SetLimits($start, $num_per_page);
```

Run Search

```
$searchClient->AddQuery($_GET['q'], 'dvds');  
$searchClient->AddQuery($_GET['q'], 'scenes');  
  
//Run Query  
$results = $searchClient->RunQueries();
```

Result Object

Search

dvds

scenes

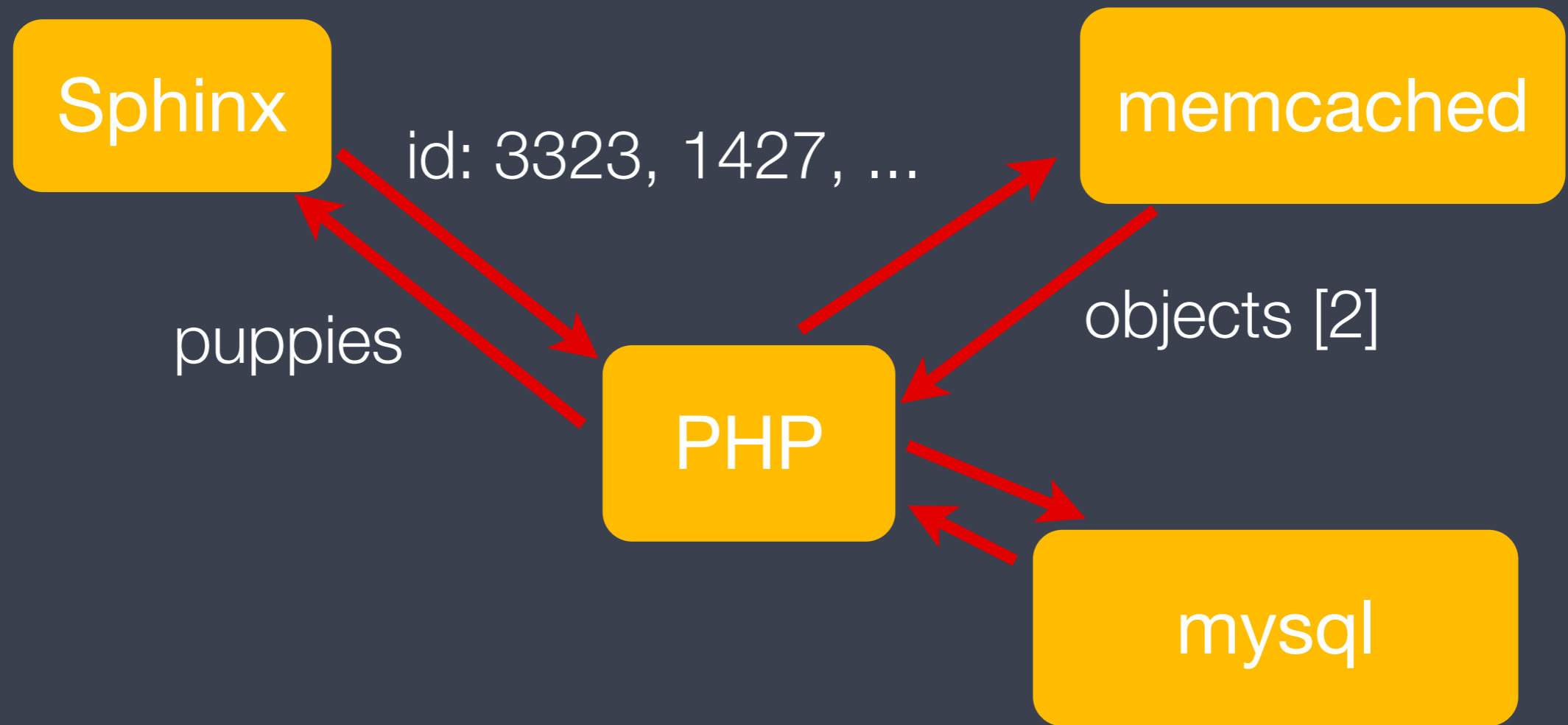
stats

matches

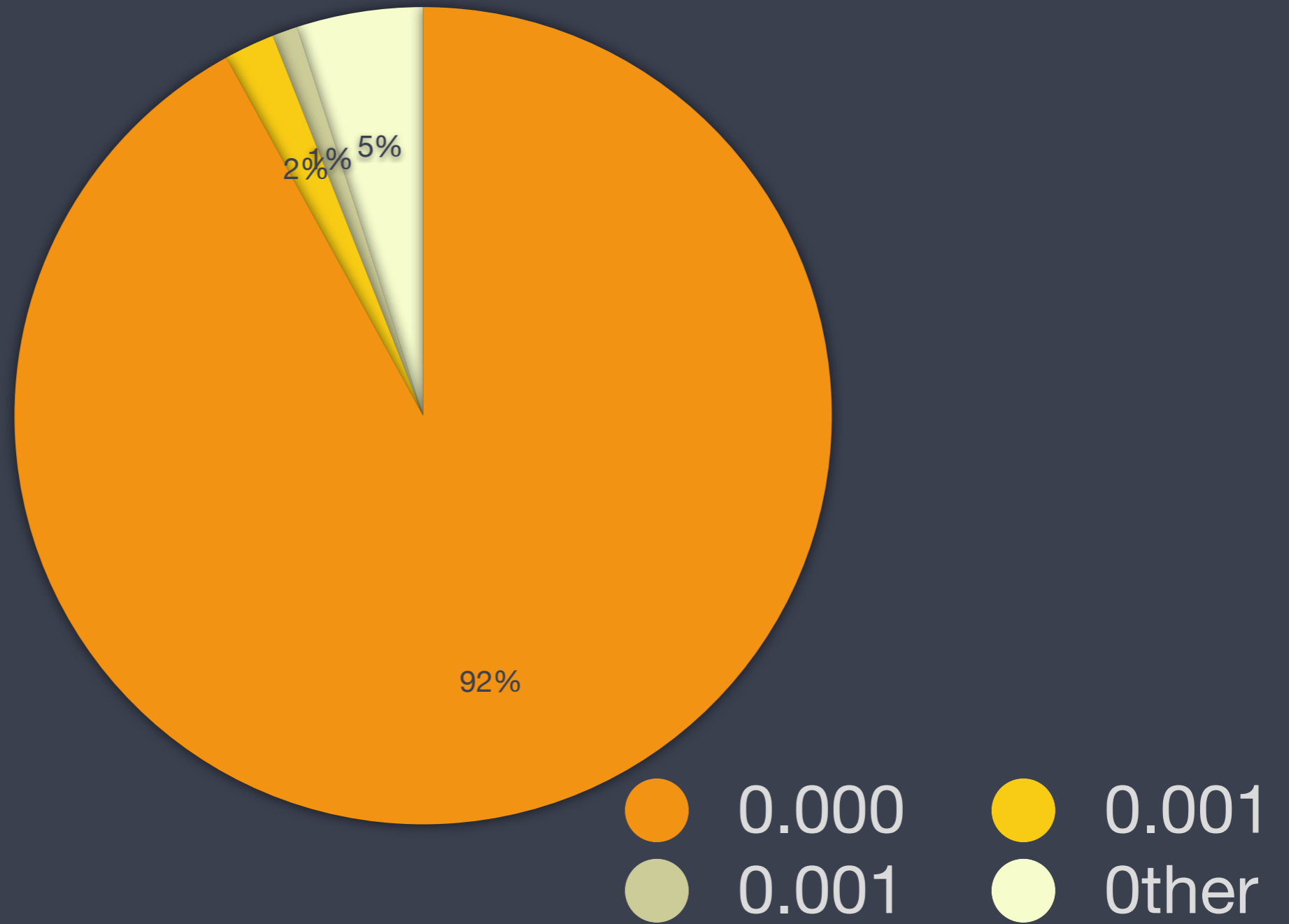
[3329]

[1428]

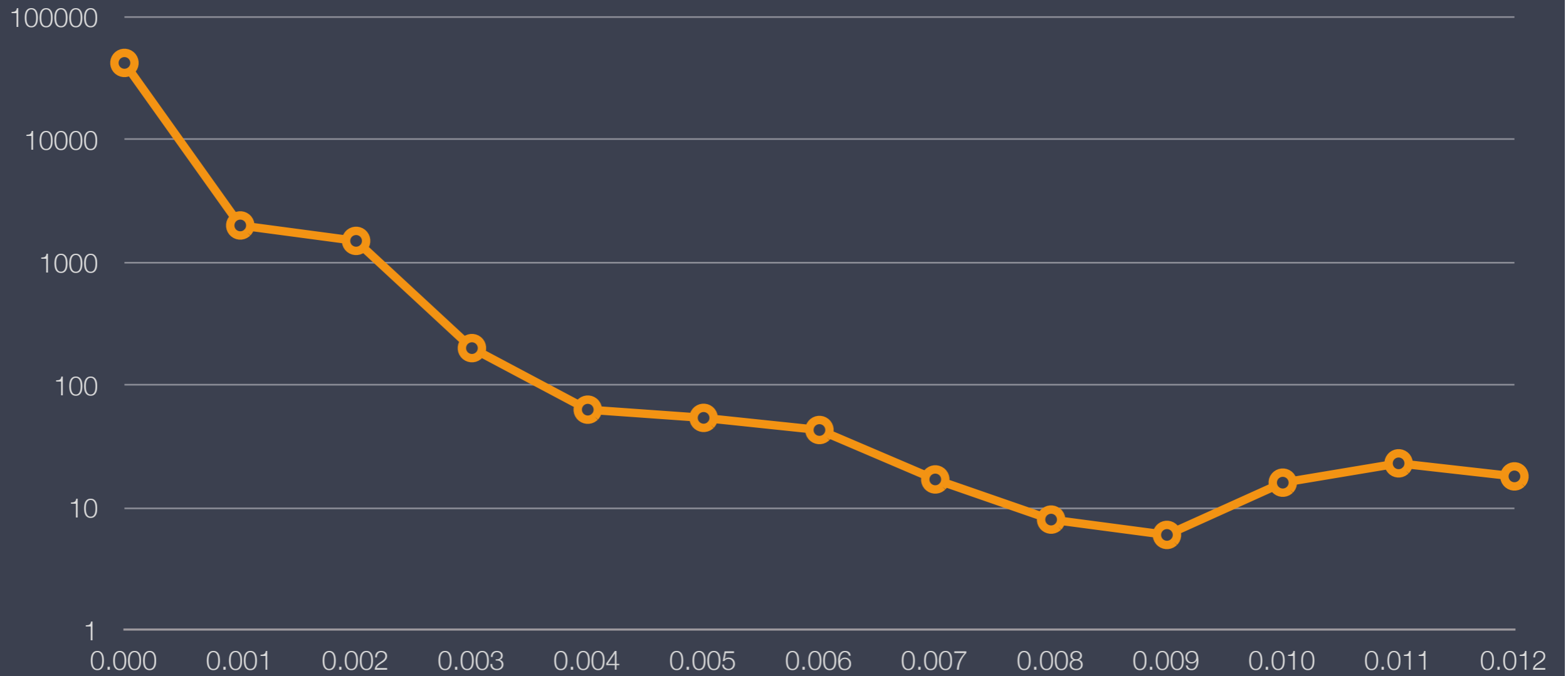
Results



Stats



Performance

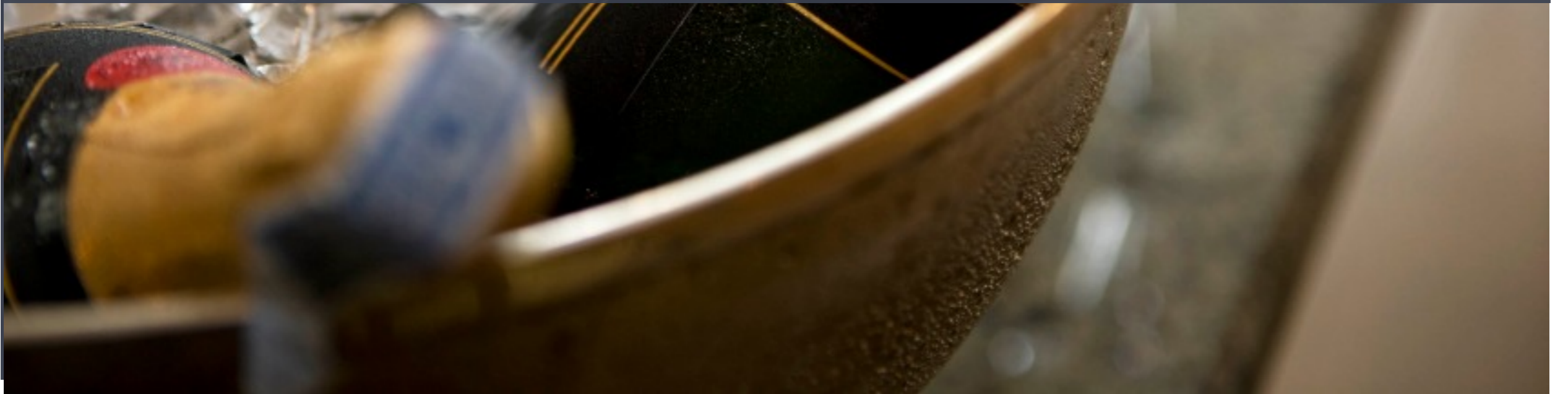


Other Stats

- ▶ 4.5 Million Unique searches/day
 - ▶ Also providing related film feature
- ▶ 8 Servers 10% CPU



No Good Deed goes Unpunished



Ajax

Search:

Ajax

Search: puppies

DVDs

Puppies on Patrol

Puppies grand adventure

Scenes

Ice cream parlour puppies

Actors

patricia puppies

Stats

- ▶ Increased time on site for tour
- ▶ Spread traffic over more content

Feedback & Contact

firstname@preinheimer.com

<http://blog.preinheimer.com>

@preinheimer

<http://wonderproxy.com>

Joind.in

<http://joind.in/talk/view/1383>

Write a comment

Comment

Great Talk! You could improve it by...

Useful Feedback

Mark as private?

Rating

5 Stars!

Submit Comment

