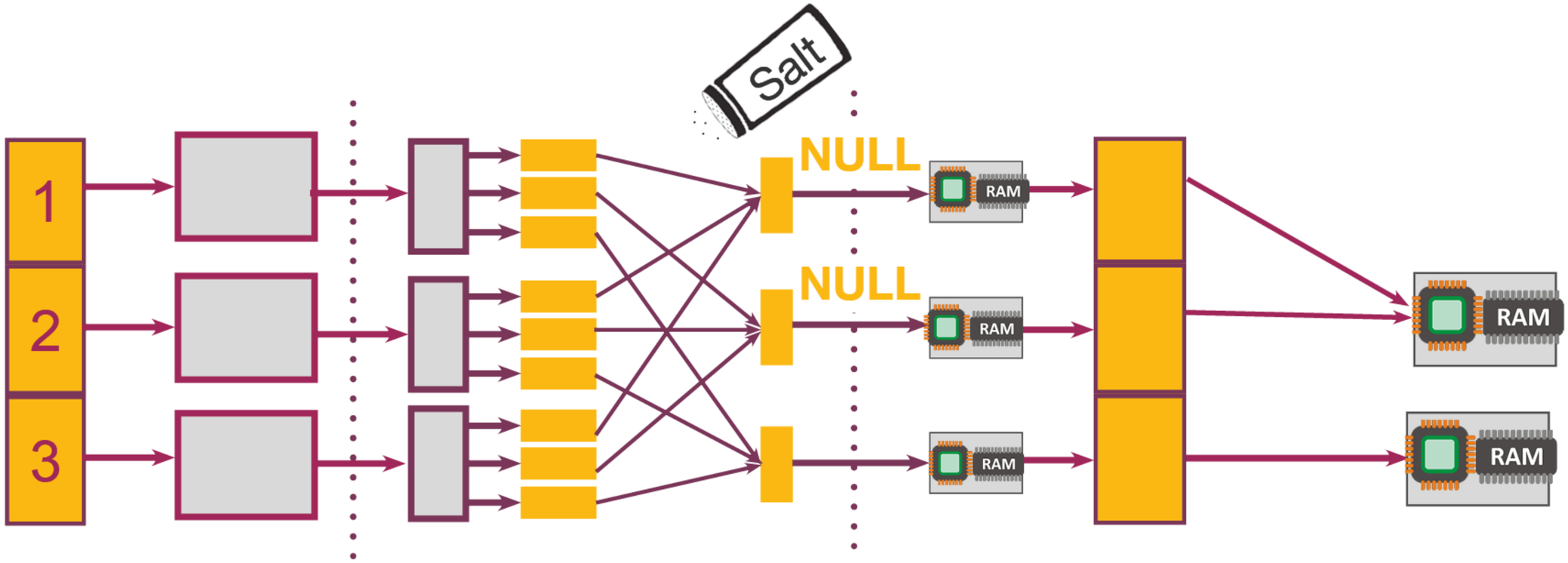
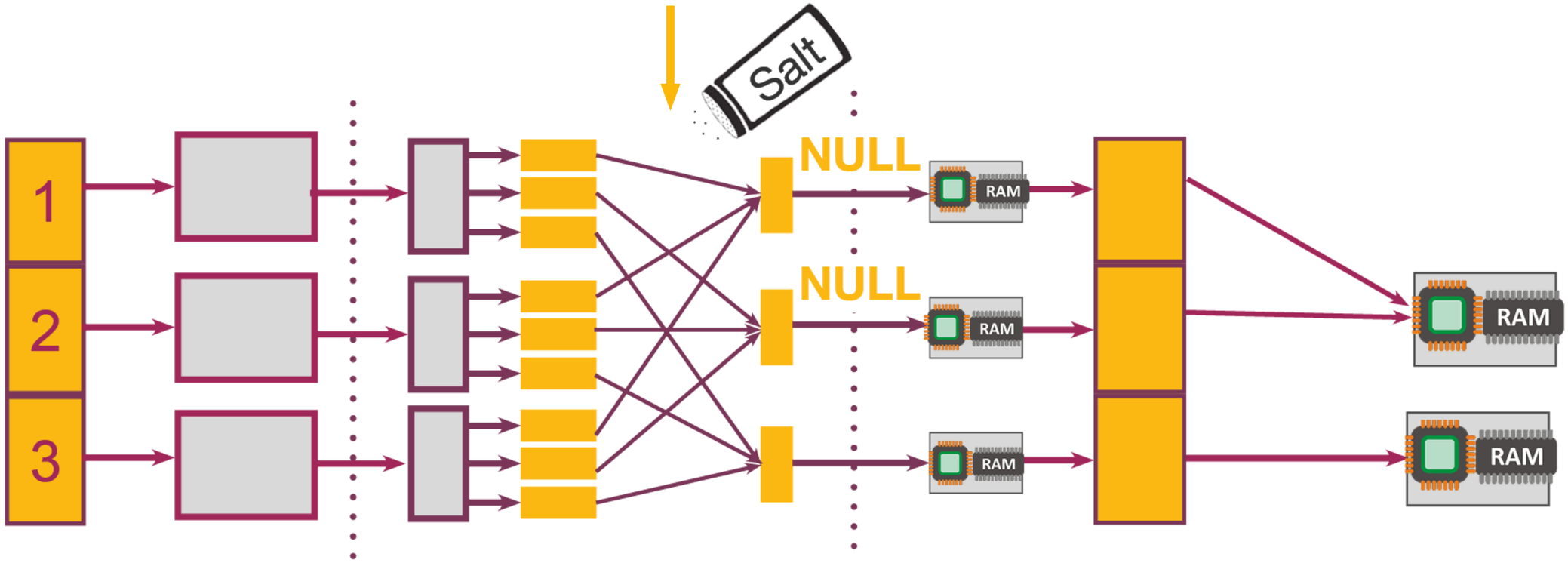
SQL over BigData

Hive Introduction


Salt
1
2
3
NULL
NULL
RAM
RAM
RAM
RAM
RAM

```
read_data()
do_magic(add_salt=True)
...
```



High-Level Languages

Apache Hive

Apache Pig

High-Level Languages

Apache Hive

Apache Pig

procedural (how)

High-Level Languages

Apache Hive

declarative (what)

Apache Pig

procedural (how)

High-Level Languages

Apache Hive

declarative (what)

~SQL

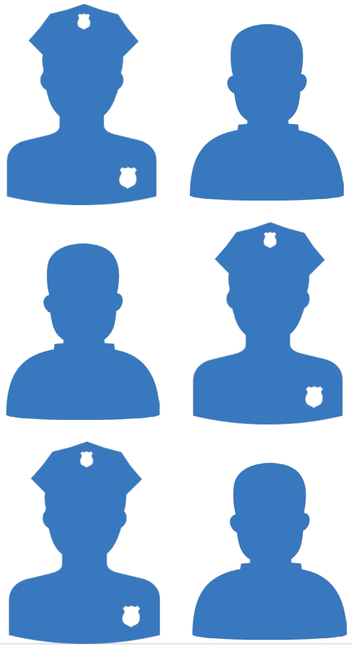
Apache Pig

procedural (how)

High-Level Languages

Apache Hive

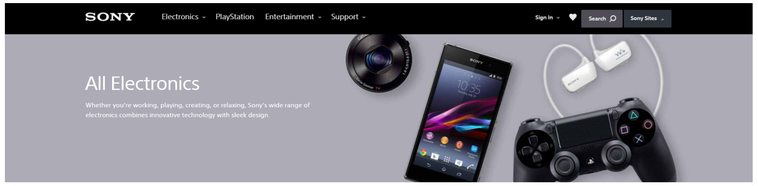
declarative (what)
~SQL



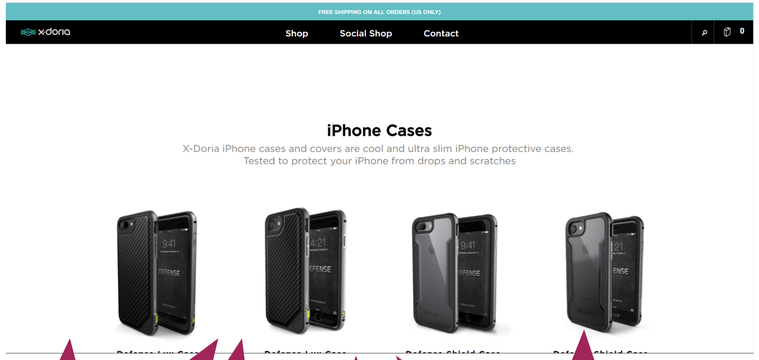
Apache Pig

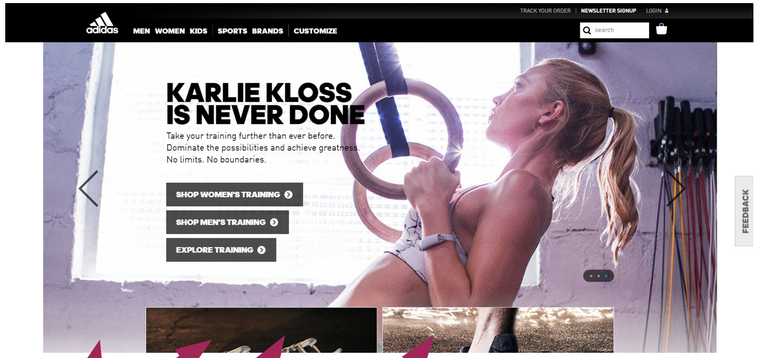
procedural (how)

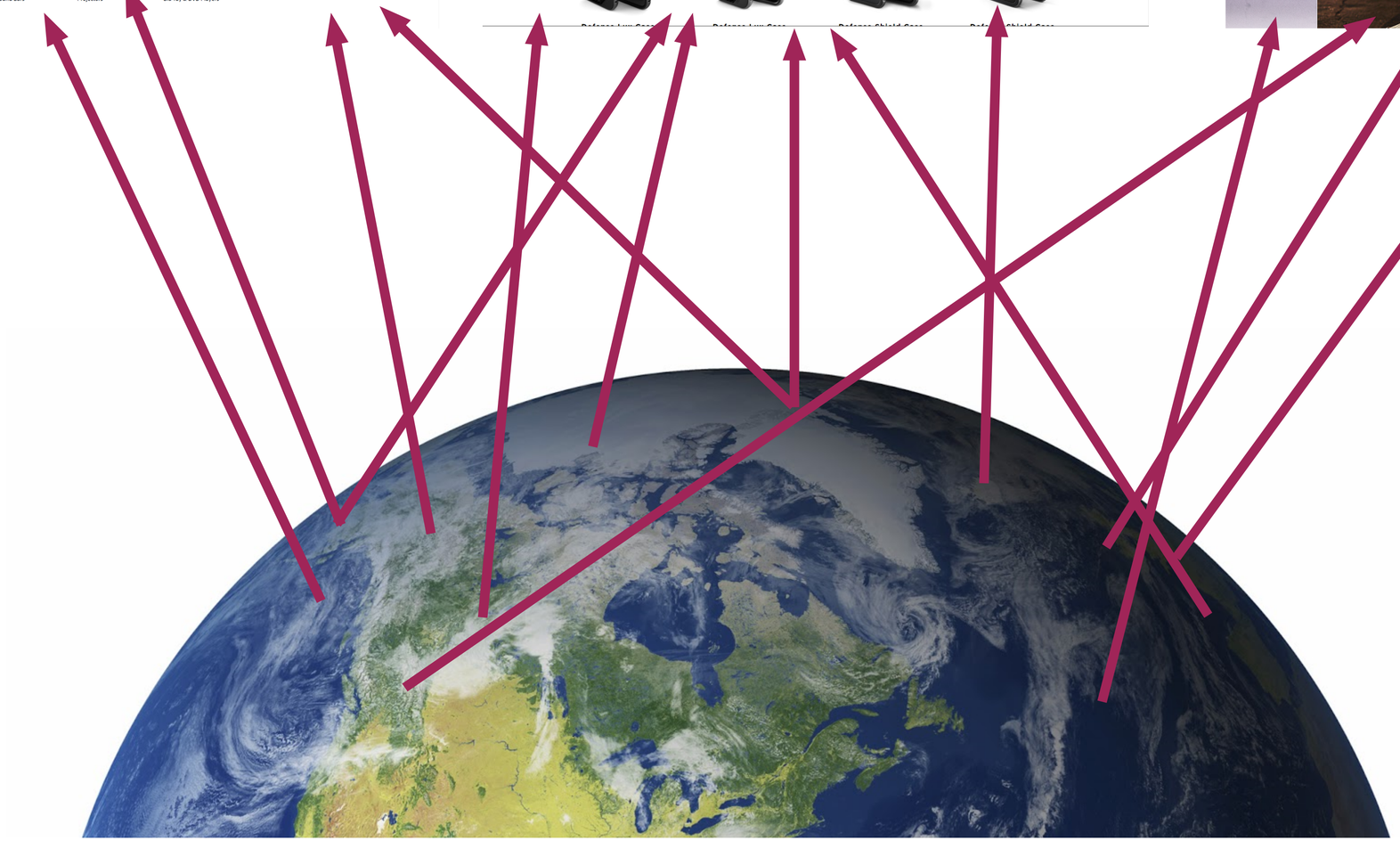
January 1, 200… October 1, 2008 July 1, 2013

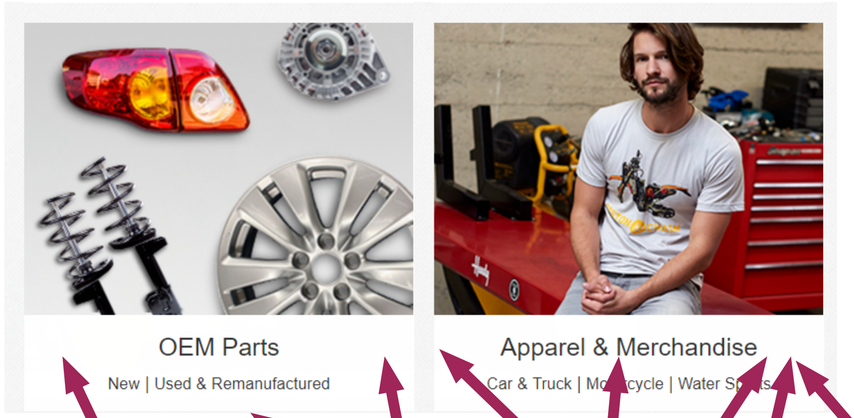

SONY
Electronics
PlayStation
Entertainment
Support
Sign In
Search
Sony Sites
All Electronics
Whether you're working, playing, creating, or relaxing, Sony's wide range of electronics combines innovative technology with sleek design.

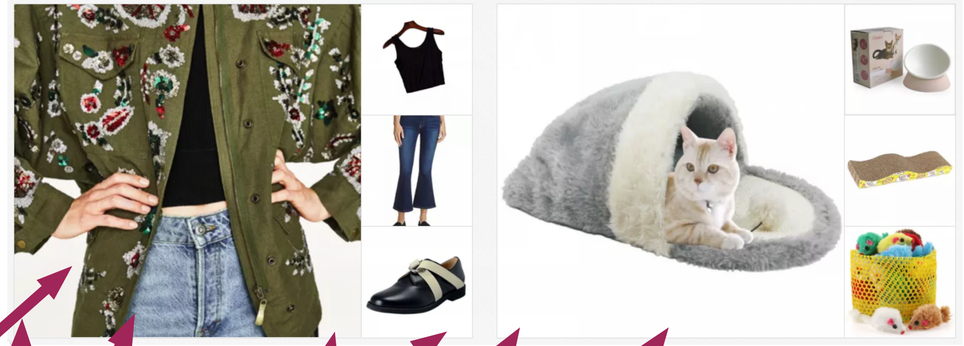

Televisions & Home Theater
Televisions
Home Theater & Sound Bars
Projectors
Blu-ray & DVD Players
ACCESSORIES
For Televisions
For Home Theater

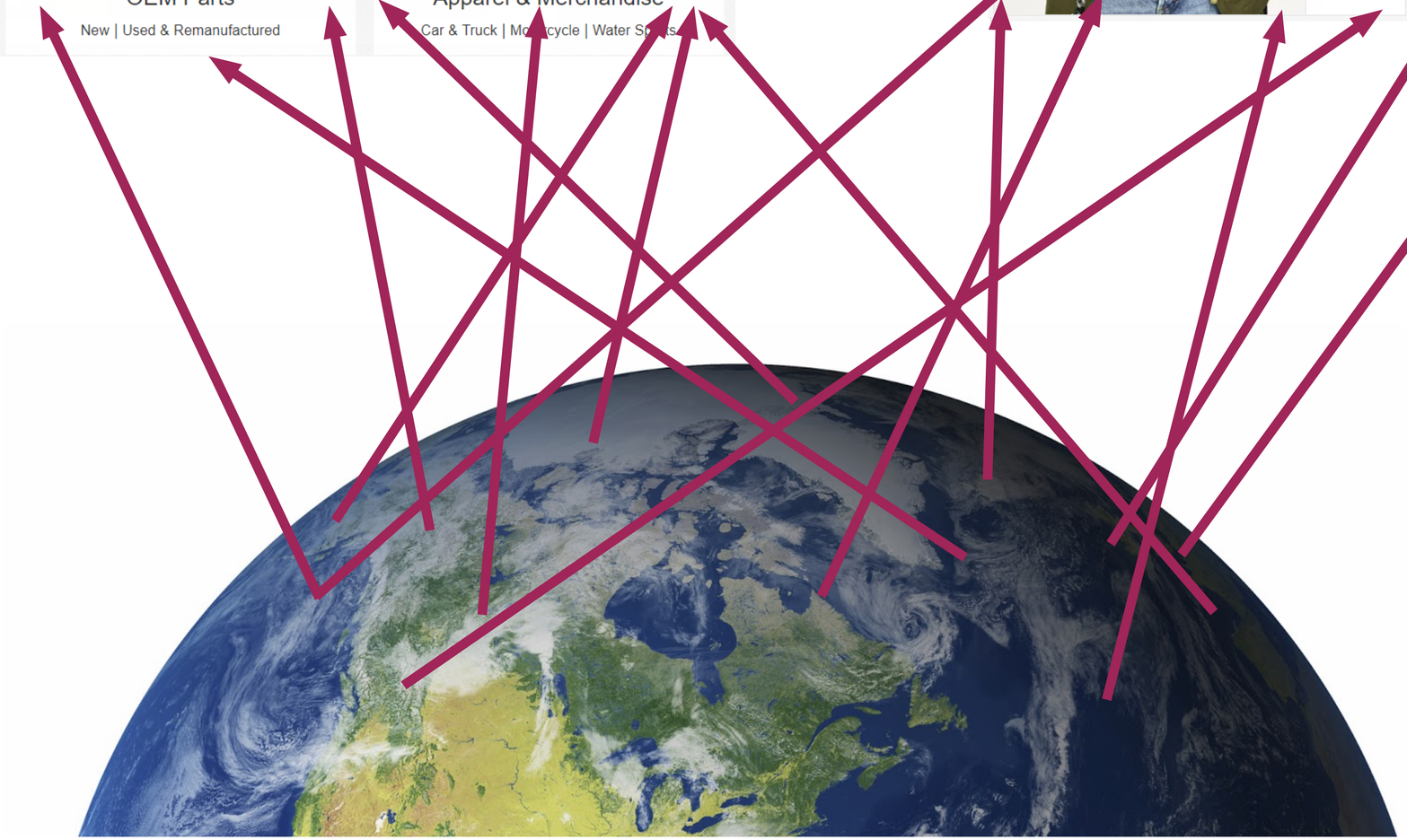

FREE SHIPPING ON ALL ORDERS (US ONLY)
xdoria
Shop
Social Shop
Contact
0
iPhone Cases
X-Doria iPhone cases and covers are cool and ultra slim iPhone protective cases.
Tested to protect your iPhone from drops and scratches

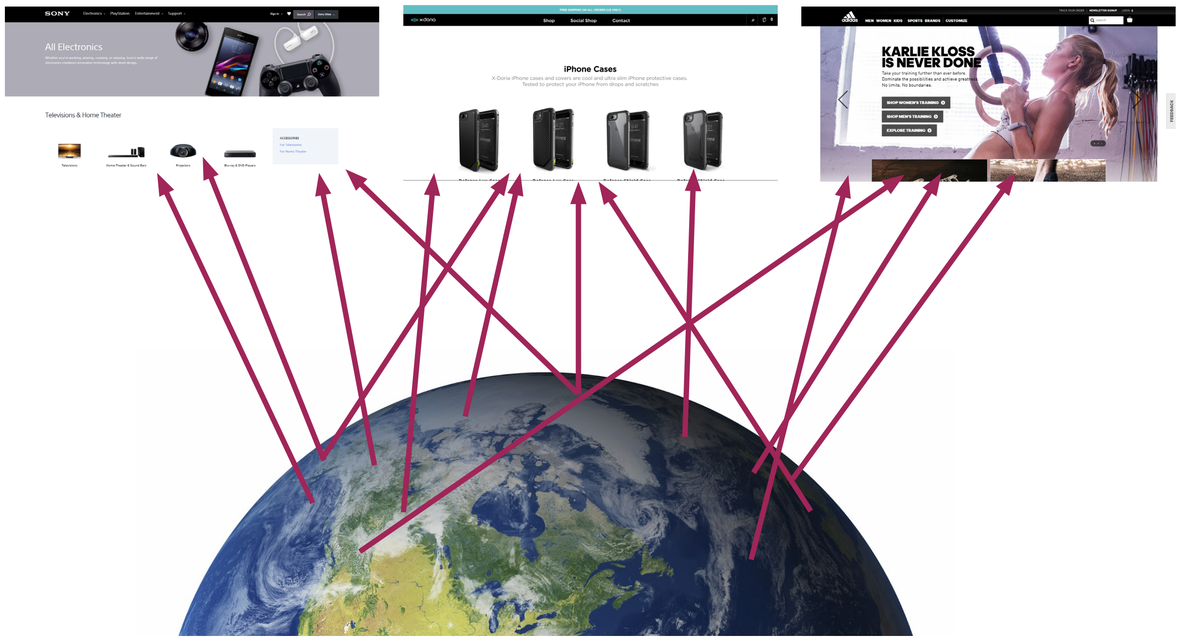

adidas
MEN
WOMEN
KIDS
SPORTS
BRANDS
CUSTOMIZE
TRACK YOUR ORDER
NEWSLETTER SIGNUP
LOGIN
search
KARLIE KLOSS
IS NEVER DONE
Take your training further than ever before.
Dominate the possibilities and achieve greatness.
No limits. No boundaries.
SHOP WOMEN'S TRAINING
SHOP MEN'S TRAINING
EXPLORE TRAINING
FEEDBACK



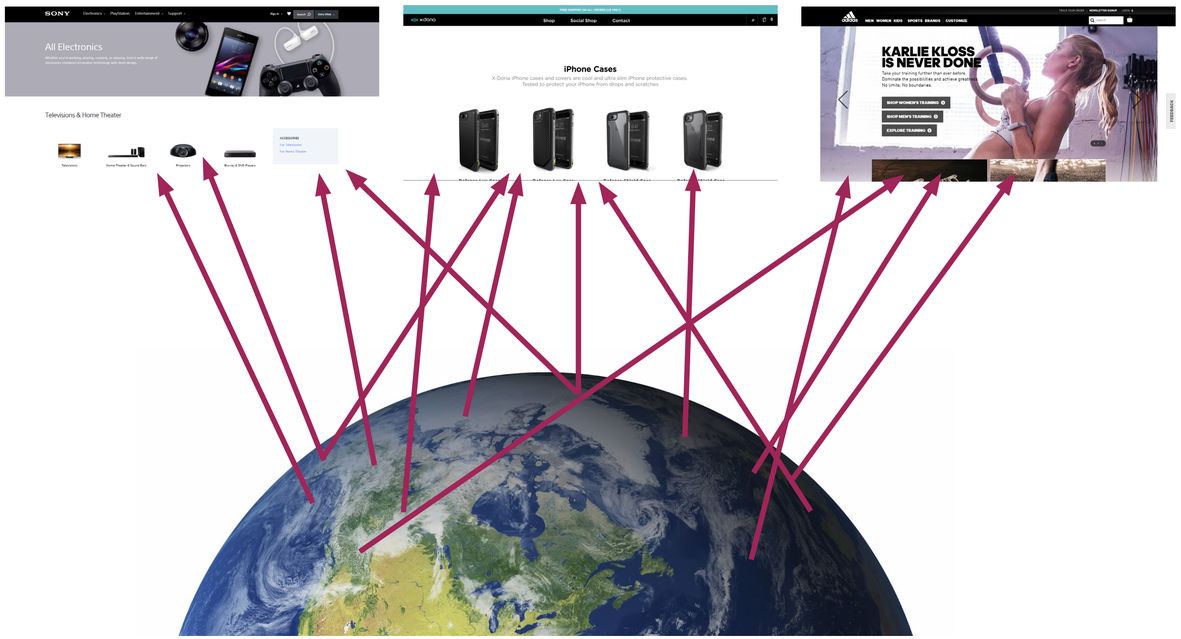

OEM Parts
New | Used & Remanufactured
Apparel & Merchandise
Car & Truck | Motorcycle | Water Sports

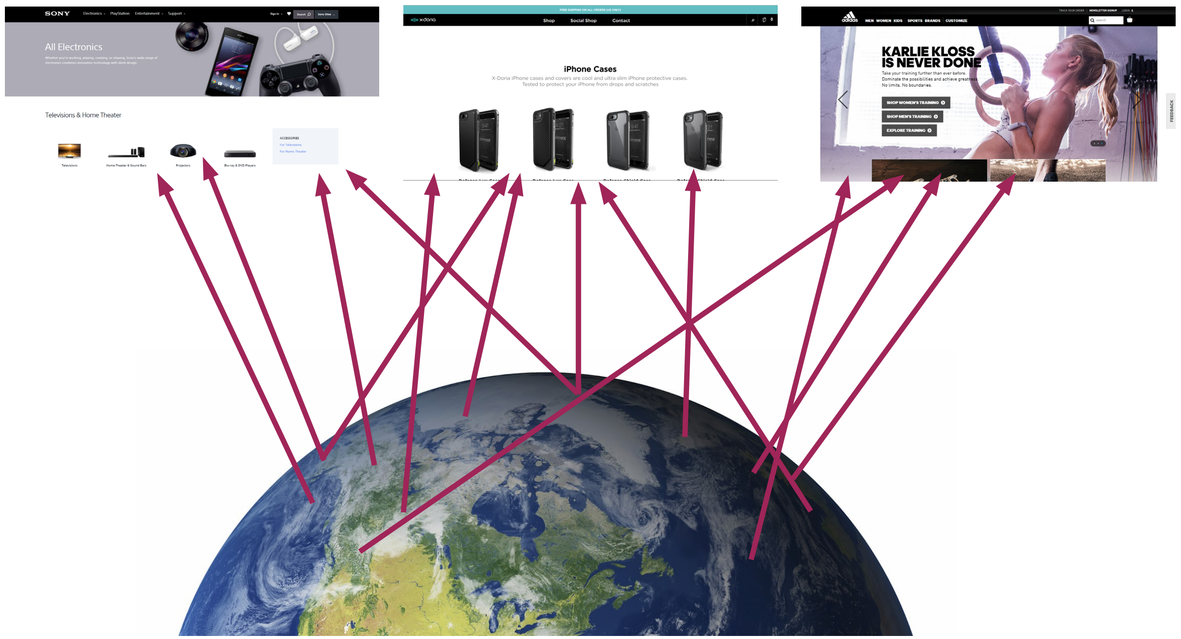




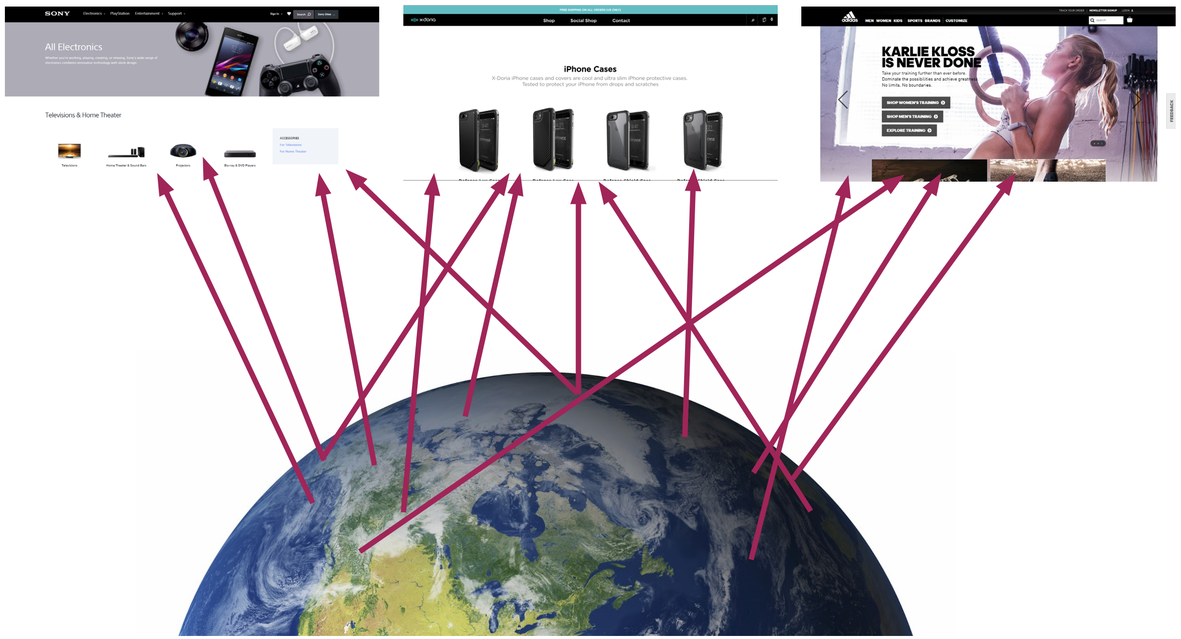

SONY
All Electronics
Televisions & Home Theater
Televisions
Home Theater & Sound Bars
Projectors
Blu-ray & DVD Players
ACCESSORIES
For Televisions
For Home Theater
Shop
Social Shop
Contact
iPhone Cases
X-Doria iPhone cases and covers are cool and ultra slim iPhone protective cases.
Tested to protect your iPhone from drops and scratches.
KARLIE KLOSS
IS NEVER DONE
Take your training further than ever before.
Dominate the possibilities and achieve greatness.
No limits. No boundaries.
SHOP WOMEN'S TRAINING
SHOP MEN'S TRAINING
EXPLORE TRAINING
FEEDBACK

1. most popular regions

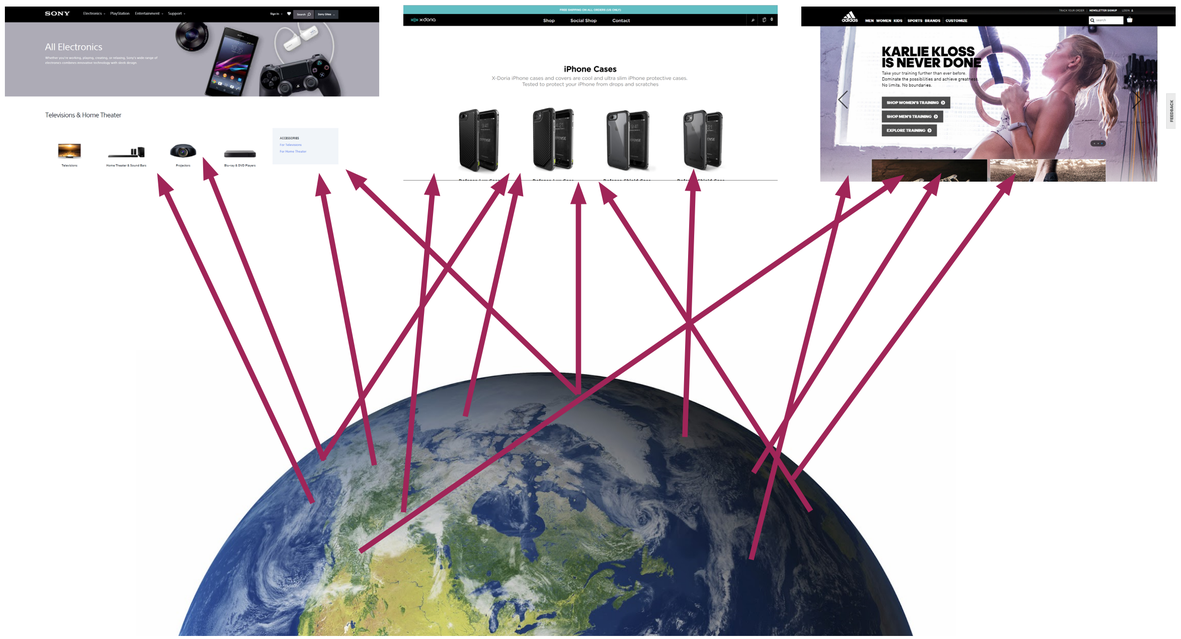




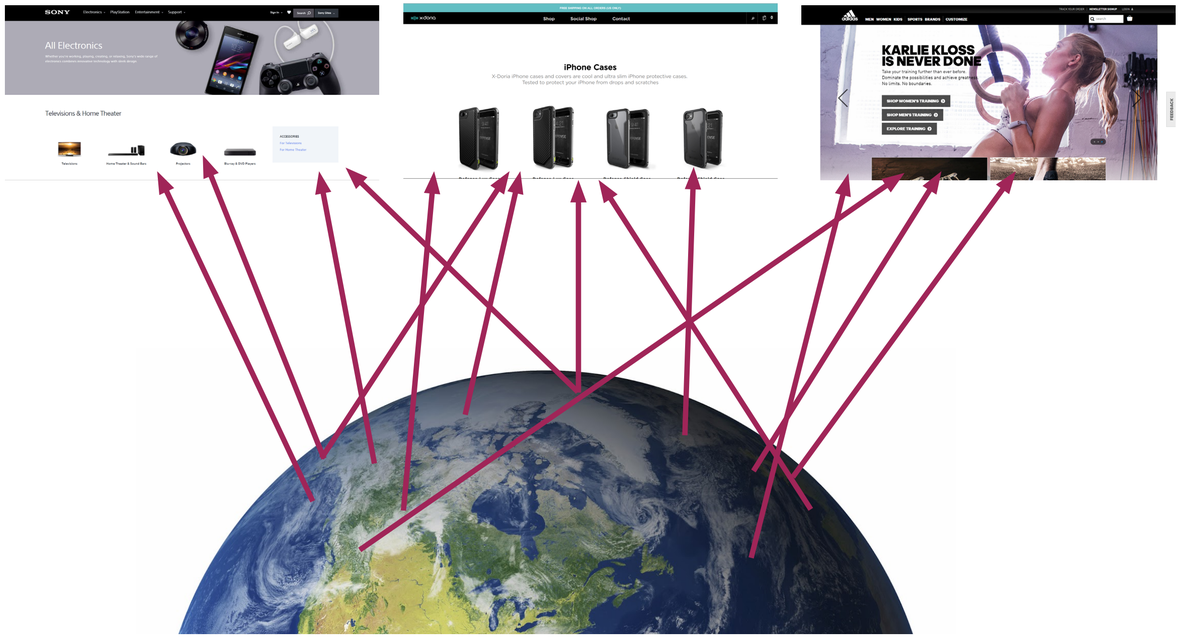
1. most popular regions
2. real users vs bots distri-
   bution



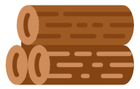
1. most popular regions
2. real users vs bots distribution
3. male vs female audience (per region)



1. most popular regions
2. real users vs bots distri-
   bution
3. male vs female audience
   (per region)
4. average customer age
   (per region)



Web-service access logs



user personal data

Geobase

[ro]bot database

1. most popular regions
2. real users vs bots distri-bution
3. male vs female audience (per region)
4. average customer age (per region)

127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10--[23/Aug/2010:03:50:59+0000]"POST/wordpress3/wp-admin/admin-ajax.php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3 (KHTML, like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715 Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"

127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10--[23/Aug/2010:03:50:59+0000]"POST/wordpress3/wp-admin/admin-ajax.php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3 (KHTML, like Gecko) Chrome/6.0.472.25 Safari/534.3”

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715 Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"

IPv4
194.0.2.235

IPv6
2001:0db8:0000:0042:0000:8a2e:0370:7334

127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10--[23/Aug/2010:03:50:59+0000]"POST/wordpress3/wp-admin/admin-ajax.php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3 (KHTML, like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715 Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"

	IPv4	IPv6
4-5 B	194.0.2.235	2001:0db8:0000:0042:0000:8a2e:0370:7334
7+ B		

127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10--[23/Aug/2010:03:50:59+0000]"POST/wordpress3/wp-admin/admin-ajax.php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3 (KHTML, like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715 Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"

127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10 - - [23/Aug/2010:03:50:59+0000] "POST /wordpress3/wp-admin/admin-ajax.php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3 (KHTML, like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715 Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"


```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200
2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10 - - [23/Aug/2010:03:50:59+0000]"POST/wordpress3/wp-admin/admin-ajax.
php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php"
"Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3
(KHTML, like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22  - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-"
"Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715
Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"
```

HTTP - HyperText Transfer Protocol

127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10 - - [23/Aug/2010:03:50:59 +0000] "POST /wordpress3/wp-admin/admin-ajax.php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3 (KHTML, like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715 Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"

127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "**GET /apache_pb.gif HTTP/1.0**" 200 2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10 - - [23/Aug/2010:03:50:59 +0000] "**POST /wordpress3/wp-admin/admin-ajax.php HTTP/1.1**" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3 (KHTML, like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "**GET /ct/ HTTP/1.1**" 200 - "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715 Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"


```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /cooking HTTP/1.0" 200 2326
"http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10 - - [23/Aug/2010:03:50:59 +0000] "POST /wordpress3/wp-admin/admin-ajax.
php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php"
"Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3
(KHTML, like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /software/development/python3 HTTP/1.1"
200 - "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715
Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"
```


```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200
2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10 - - [23/Aug/2010:03:50:59 +0000] "POST /wordpress3/wp-admin/admin-ajax.
php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php"
"Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3
(KHTML, like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22  - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-"
"Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715
Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"
```

2xx - all good

3xx - redirect

4xx - client-side problem

5xx - server-side problem

Apache

HTTP Server

2xx - all good

3xx - redirect

4xx - client-side problem

5xx - server-side problem

2xx - all good

3xx - redirect

4xx - client-side problem

5xx - server-side problem


```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326
"http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10 - - [23/Aug/2010:03:50:59 +0000] "POST /wordpress3/wp-admin/admin-
ajax.php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php"
"Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3 (KHTML,
like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-"
"Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715
Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"
```


```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326
"http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10 - - [23/Aug/2010:03:50:59 +0000] "POST /wordpress3/wp-admin/admin-
ajax.php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php"
"Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3 (KHTML,
like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-"
"Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715
Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"
```

Apache Common Log Format

Combined Log Format = Common Log Format + 2 fields


```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326
"http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10 - - [23/Aug/2010:03:50:59 +0000] "POST /wordpress3/wp-admin/admin-
ajax.php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php"
"Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3 (KHTML,
like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-"
"Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715
Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"
```

127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

123.65.150.10 - - [23/Aug/2010:03:50:59 +0000] "POST /wordpress3/wp-admin/admin-ajax.php HTTP/1.1" 200 2 "http://www.example.com/wordpress3/wp-admin/post-new.php" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_4; en-US) AppleWebKit/534.3 (KHTML, like Gecko) Chrome/6.0.472.25 Safari/534.3"

83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-" "Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715 Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"

Combined Log Format example:

```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326
"http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"
```

…

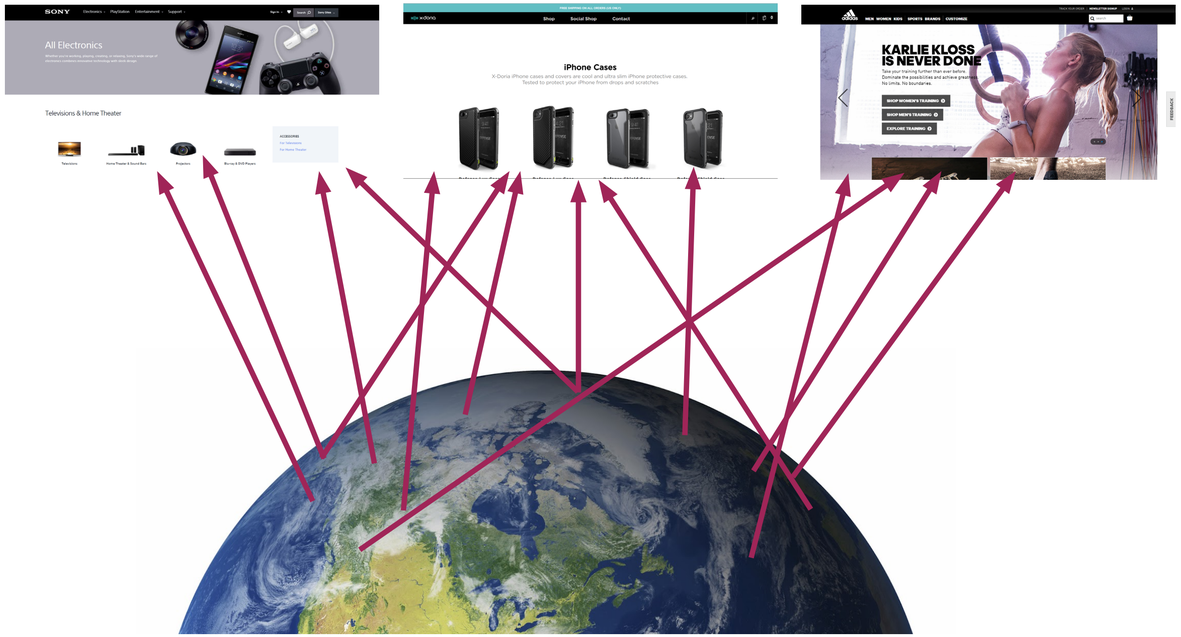
custom configuration:

```
56.167.169.126    20150601052153http://newsru.com/7691562  193 102 Opera/5.0
(compatible; MSIE 9.0; Windows NT 6.1; Win64; x64; Trident/5.0; .NET CLR 3.5.30729;)
75.208.40.166     20150601052421http://news.yandex.ru/8023677 178 405 Opera/5.0
(Windows; U; MSIE 9.0; Windows NT 8.0; Win64; x64; Trident/5.0; .NET4.0E; en)
```

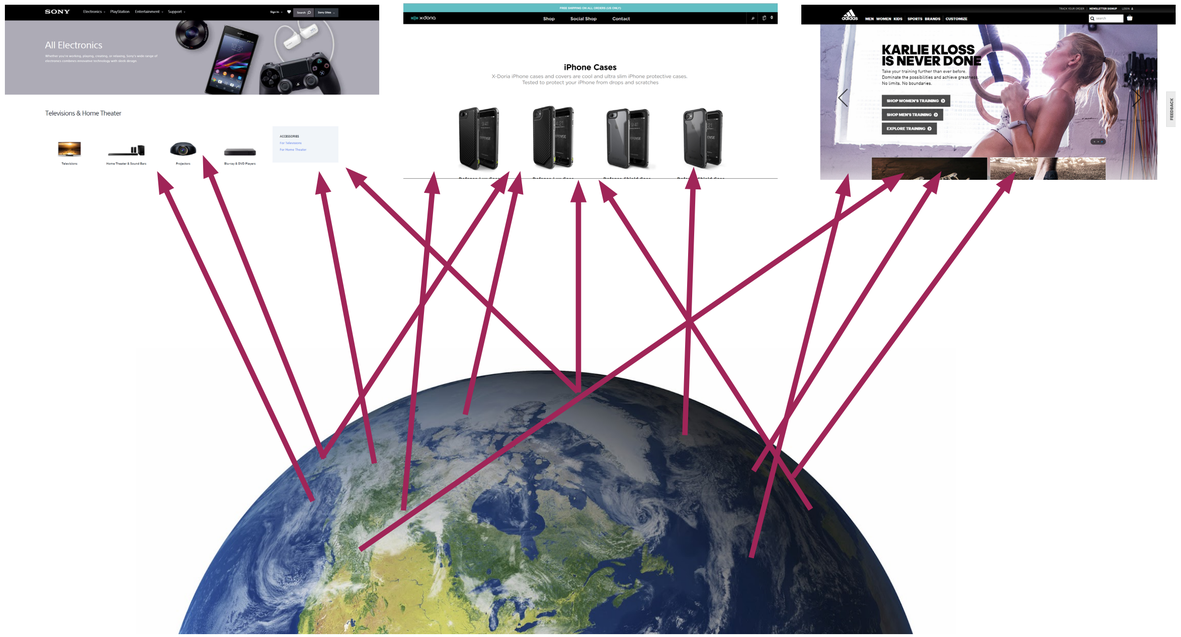
+ cookies:

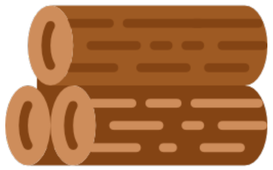
```
83.0.11.22 - - [02/Aug/2009:12:31:30 +0200] "GET /ct/ HTTP/1.1" 200 - "-"
"Mozilla/5.0 (Windows; U; Windows NT 5.1; pl; rv:1.9.1.1) Gecko/20090715
Firefox/3.5.1" "c1=1; c2=2; PHPSESSID=6c4513f22852a235b8988da822f89d04"
```

1. Most popular regions



1. Most popular regions

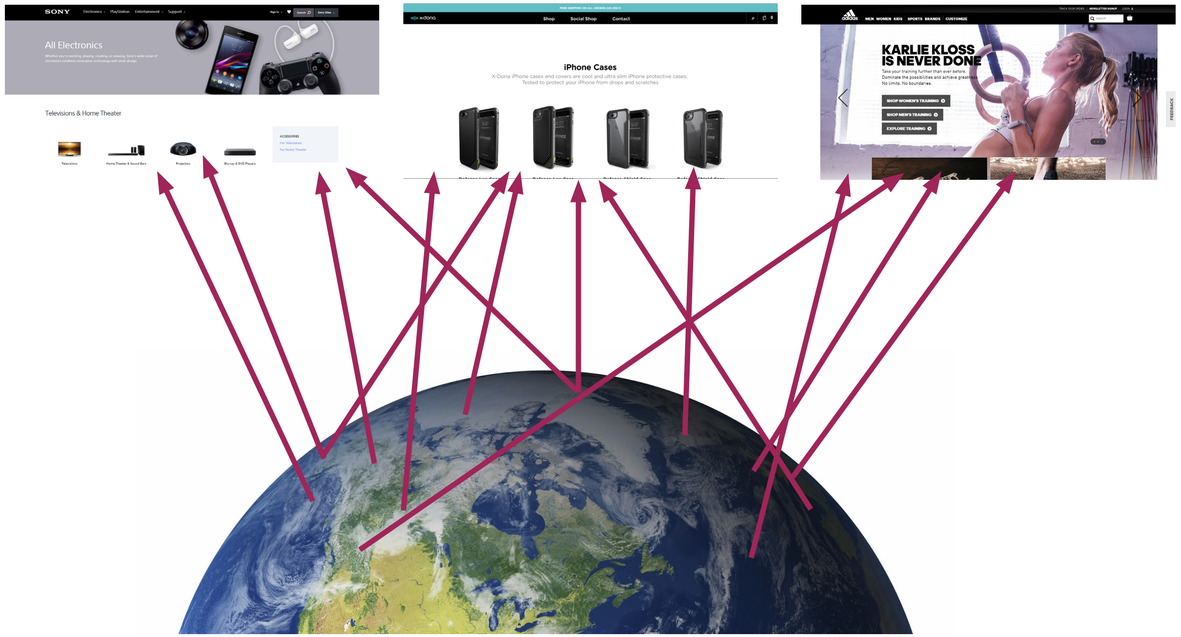


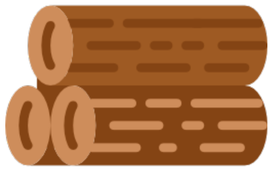


Web-service access logs

Geobase

1. Most popular regions





Web-service access logs

Geobase

IPv4: 109.188.67.224

↓

area: Moscow City Center

↓

city: Moscow

↓

country: Russia

↓

Earth

109.188.67.224, Moscow City Centre, Moscow, Russia, Earth

109.188.67.221, Moscow City Centre, Moscow, Russia, Earth

...

1. Most popular regions

Web-service access logs

Geobase

`format: ip, request, status_code, ...`

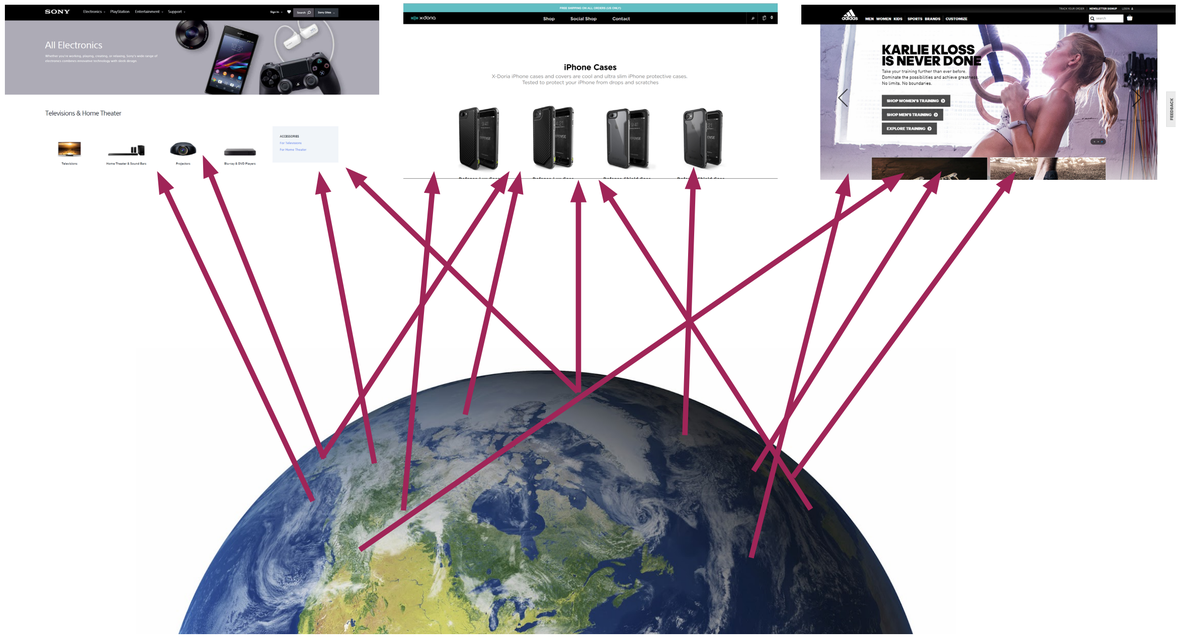
`format: ip, region`$_{city}$`, region`$_{country}$`,...`

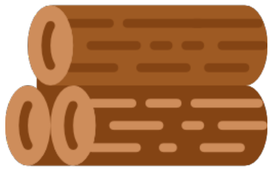
`Join(ip)` `+ WordCount(region) +`

`TOP(100)`
`Sort + LIMIT`

format: region, hit_count

2. Real users vs bots distribution

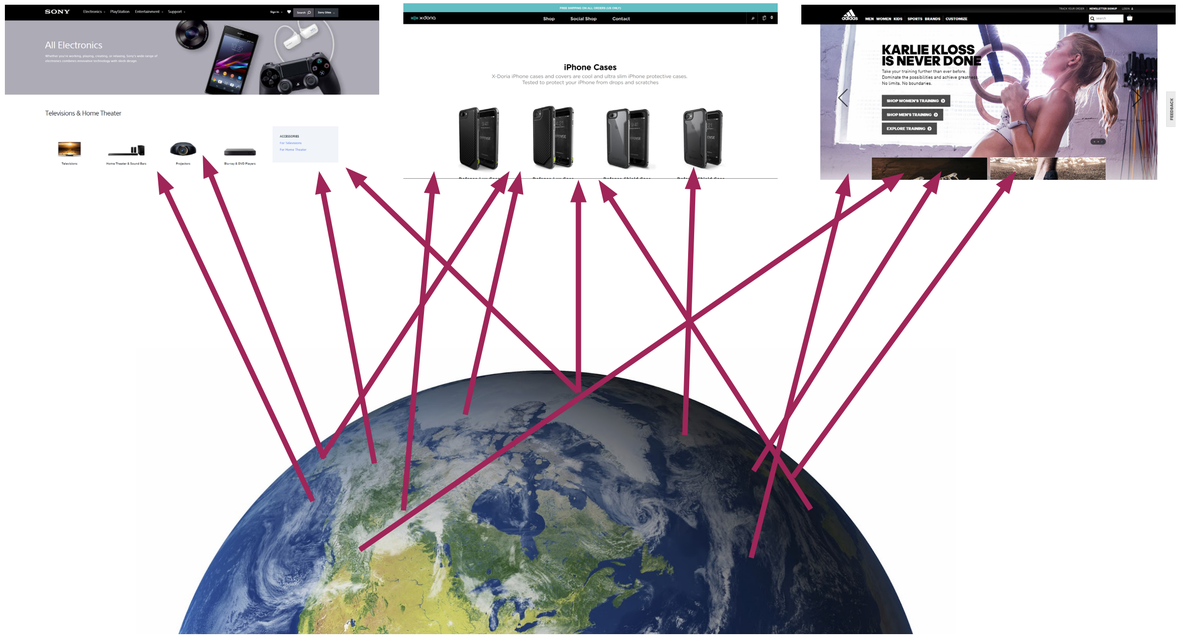




Web-service access logs

Geobase

2. Real users vs bots distribution

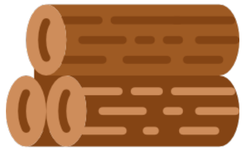


Web-service
access logs

bot name + $user\_agent_1$ + ip_1
$user\_agent_2$ + ip_2
... ...

[ro]bot database

2. Real users vs bots distribution

Web-service access logs

format: ip, request, user_agent, ...

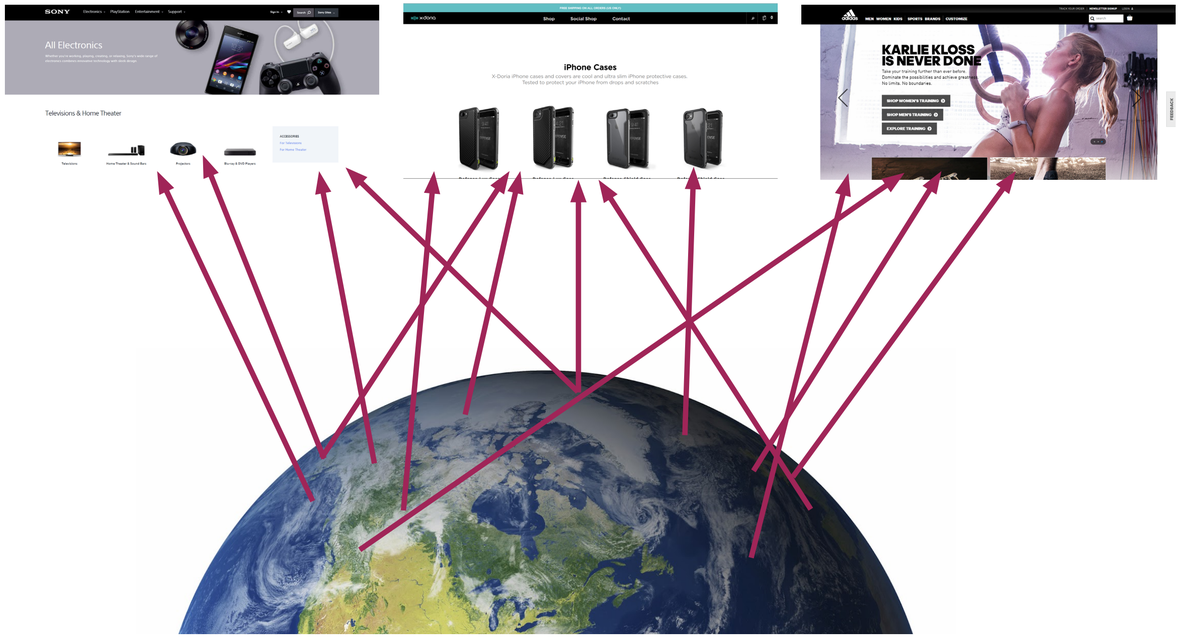
[ro]bot database

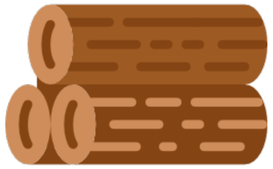
format: bot_name, user_agents, ips

Join(ip, user_agent) + WordCount(request / user, bot)

format: region, user_hits, bot_hits

3. Male vs female audience (per region)

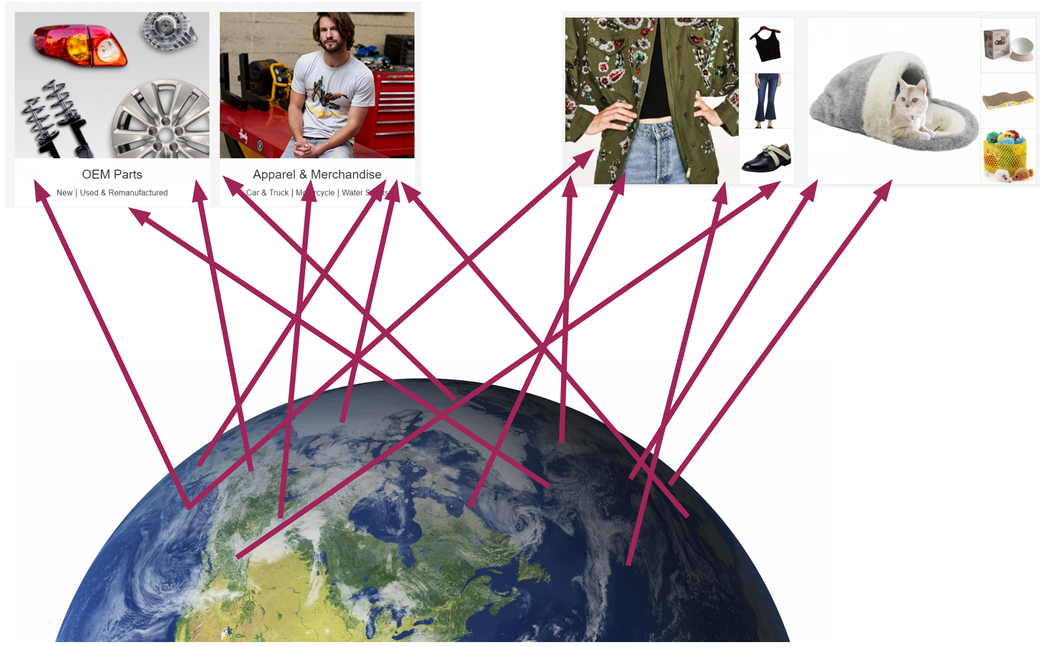




Web-service access logs

Geobase

3. Male vs female audience (per region)



	age	gender	occupation	zipcode
user_id				
1	24	M	technician	85711
2	53	F	other	94043
3	23	M	writer	32067
4	24	M	technician	43537
5	33	F	other	15213
6	42	M	executive	98101
7	57	M	administrator	91344
8	36	M	administrator	05201
9	29	M	student	01002
10	53	M	lawyer	90703

Web-service access logs

Geobase

User personal data

3. Male vs female audience (per region)

Web-service access logs

Geobase

User personal data

`format: ip, request, user_agent, ...`

`format: user_id, gender, age, ...`

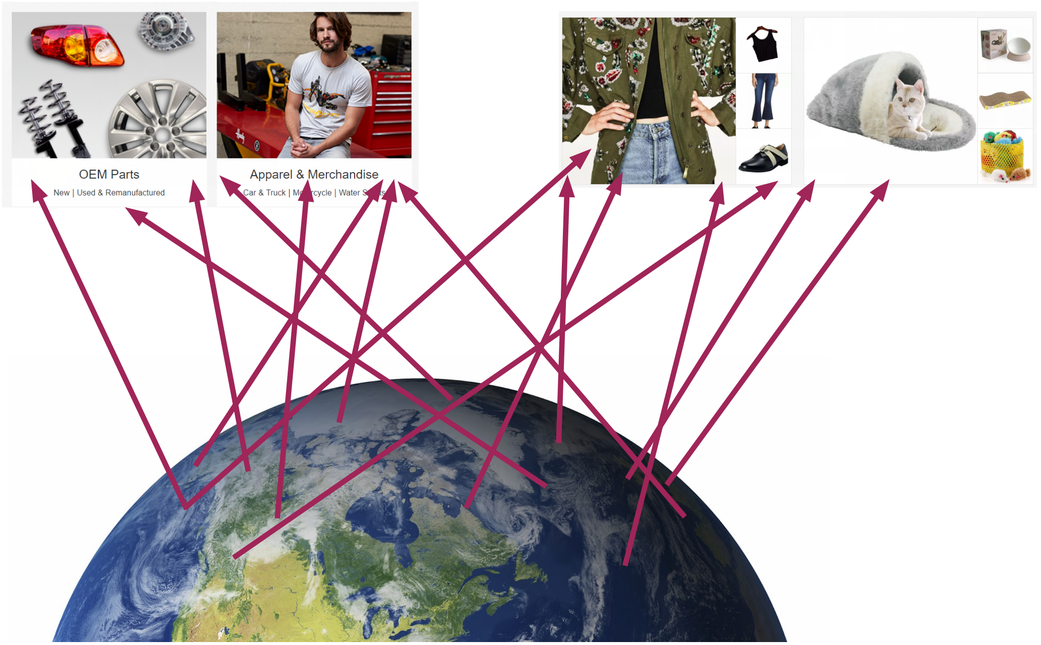
`format: ip, region`$_{city}$`, region`$_{country}$`, ...`

`Join(ip)` + `Join(user_id)` + `WordCount(region/gender)`

user_id = ip + user_agent

format: region,
male_hits, female_hits

4. Average customer age (per region)



	age	gender	occupation	zipcode
user_id				
1	24	M	technician	85711
2	53	F	other	94043
3	23	M	writer	32067
4	24	M	technician	43537
5	33	F	other	15213
6	42	M	executive	98101
7	57	M	administrator	91344
8	36	M	administrator	05201
9	29	M	student	01002
10	53	M	lawyer	90703

Web-service access logs

Geobase

User personal data

4. Average customer age (per region)

Web-service access logs

Geobase

User personal data

format: ip, request, user_agent, ...

format: user_id, gender, age, ...

format: ip, $region_{city}$, $region_{country}$, ...

Join(ip) + Join(user_id) + **Average**(age)

user_id = ip + user_agent

format: region, average_age

4. Average customer age (per region)

Web-service access logs

Geobase

User personal data

`format: ip, request, user_agent, ...`

`format: user_id, gender, age, ...`

`format: ip, region`$_{city}$`, region`$_{country}$`, ...`

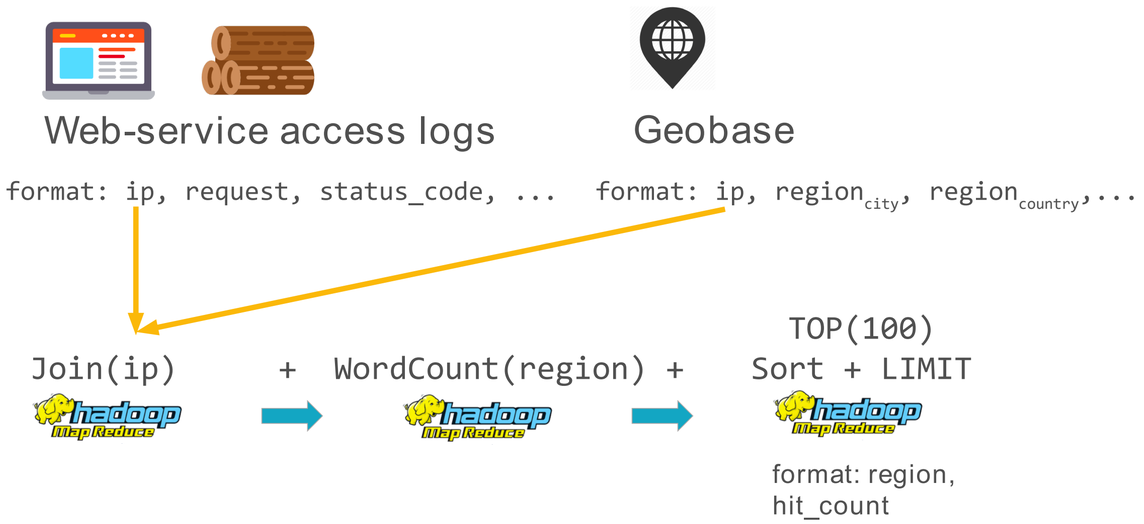
`Join(ip)` + `Join(user_id)` + `Average(age)`

user_id = ip + user_agent

format: region, average_age

use Combiner for optimisation (see: the first course)

1. Most popular regions



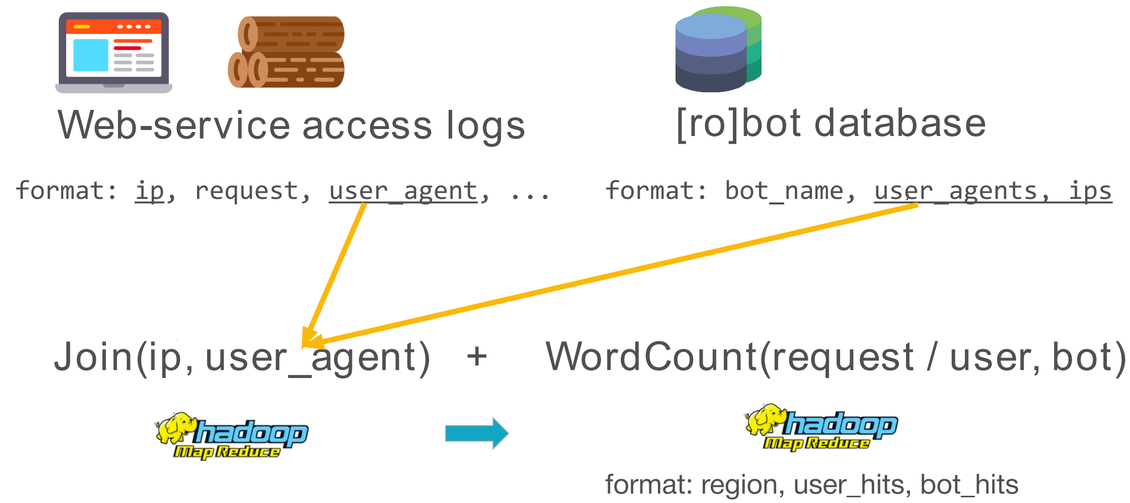
WordCount

```
SELECT regioncity, COUNT(1) AS hit_count
FROM access_log JOIN geo_base
ON (access_log.host = geo_base.host)
GROUP BY regioncity ORDER BY hit_count LIMIT 100
```

join

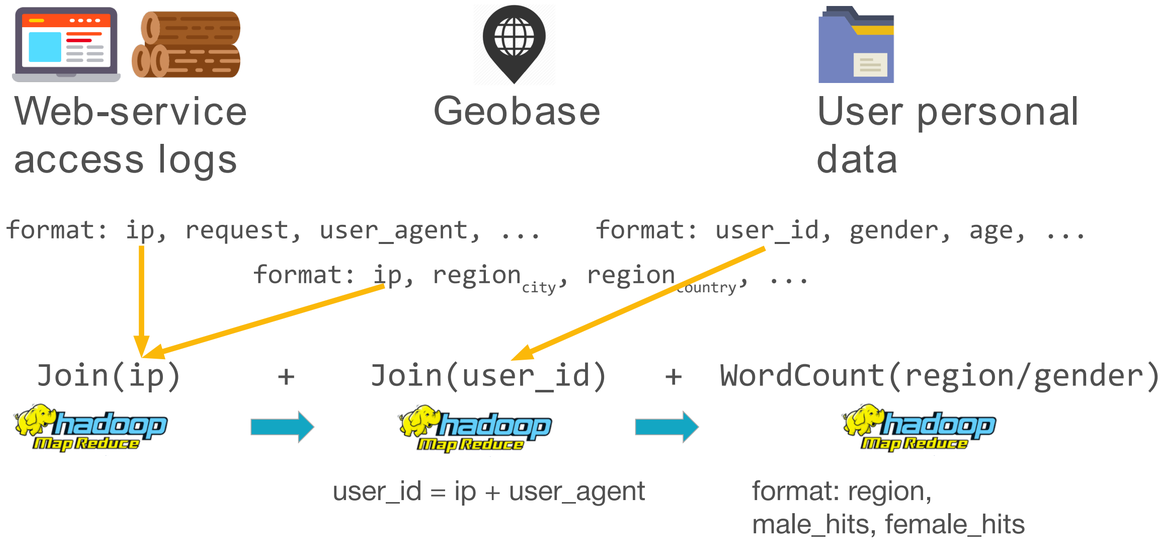
TOP(100)

2. Real users vs bots distribution




```
SELECT request,
    SUM(IF(robot.bot_name IS NULL, 1, 0)) as user_hit_count,
    SUM(IF(robot.bot_name IS NOT NULL, 1, 0)) as bot_hit_count
FROM access_log LEFT OUTER JOIN robot ON (
    access_log.host = robot.host
    AND access_log.user_agent = robot.user_agent
)
GROUP BY request
```

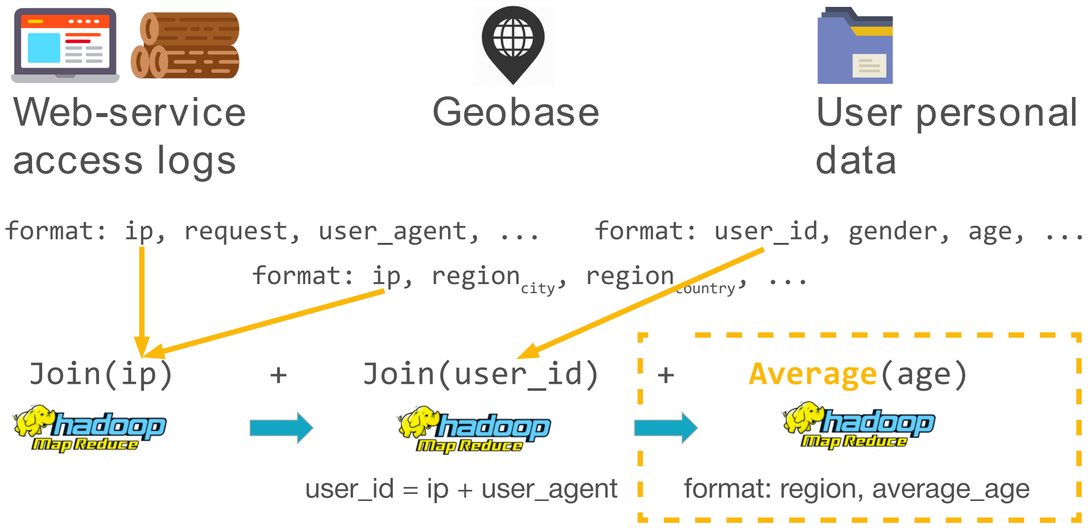
3. Male vs female audience (per region)




```
SELECT region_city,
    SUM(IF(user.gender = “M”,1,0)) as male_hit_count,
    SUM(IF(user.gender = “F”,1,0)) as female_hit_count
FROM access_log
    JOIN geo_base ON (access_log.host = geo_base.host)
    JOIN user ON (access_log.host = user.host
        AND access_log.user_agent = user.user_agent
    )
GROUP BY region_city
```

two joins

4. Average customer age (per region)



Java source code:

https://goo.gl/faOeTh

https://goo.gl/DCeVPk

Hive QL


```
SELECT region2, AVG(user.age)
FROM access_log
    JOIN geo_base ON (access_log.host = geo_base.host)
    JOIN user ON (access_log.host = user.host
        AND access_log.user_agent = user.user_agent
    )
GROUP BY region2
```

Summary

Summary

- You can **list** common log fields of Apache HTTP Server and **explain** what they are used for

Summary

- You can **list** common log fields of Apache HTTP Server and **explain** what they are used for
- You can **list** several high-level programming languages and **explain** what the difference between them

Summary

- You can **list** common log fields of Apache HTTP Server and **explain** what they are used for
- You can **list** several high-level programming languages and **explain** what the difference between them
- You should be able to **explain and execute** HiveQL queries to solve several business use cases