



About me

Open source

- Flink committer
- Member of the project management committee (PMC)
- Volunteer on the Apache privacy committee

Career

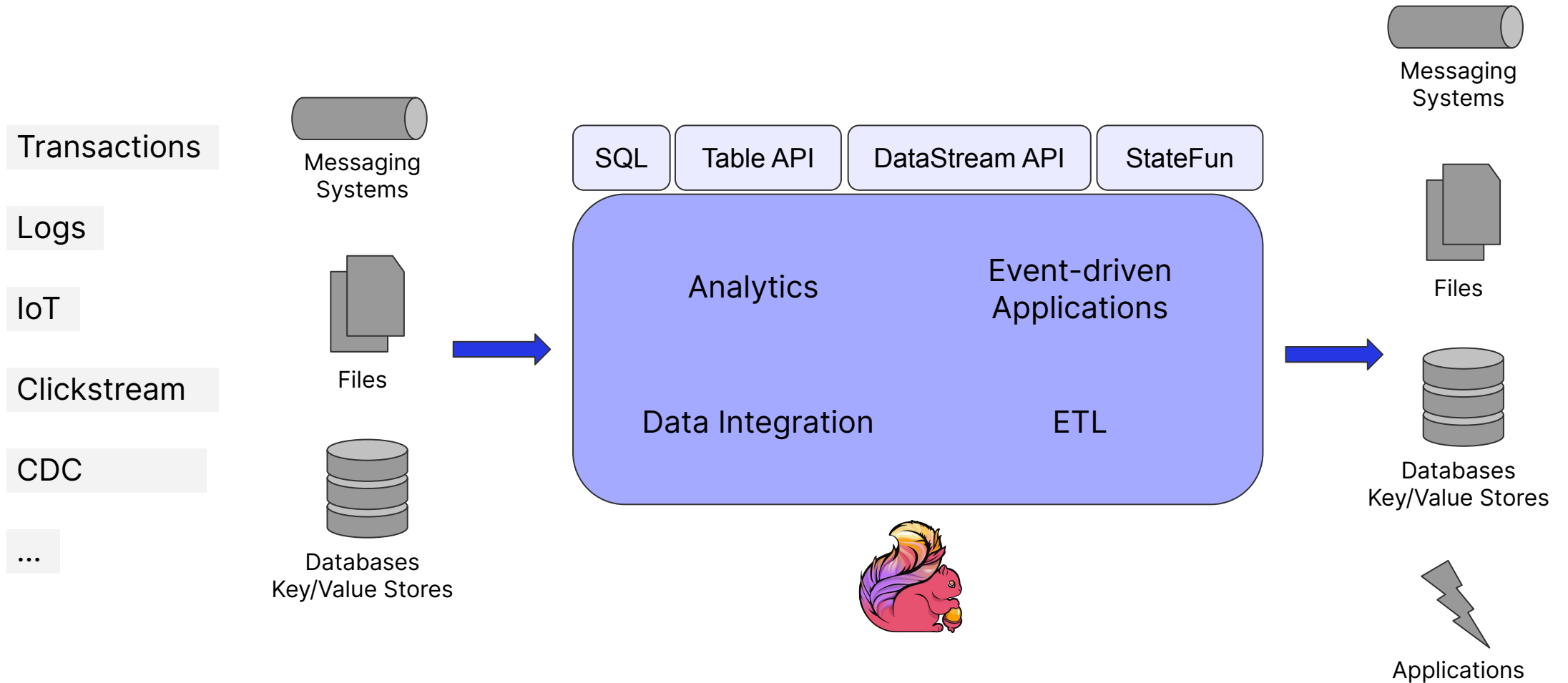
- Product Lead Streaming & Engagement @ ING
- Product Manager SDK & SQL team @ Ververica
- Product Manager @ Immerok



@MartijnVisser82
martijnvisser@apache.org

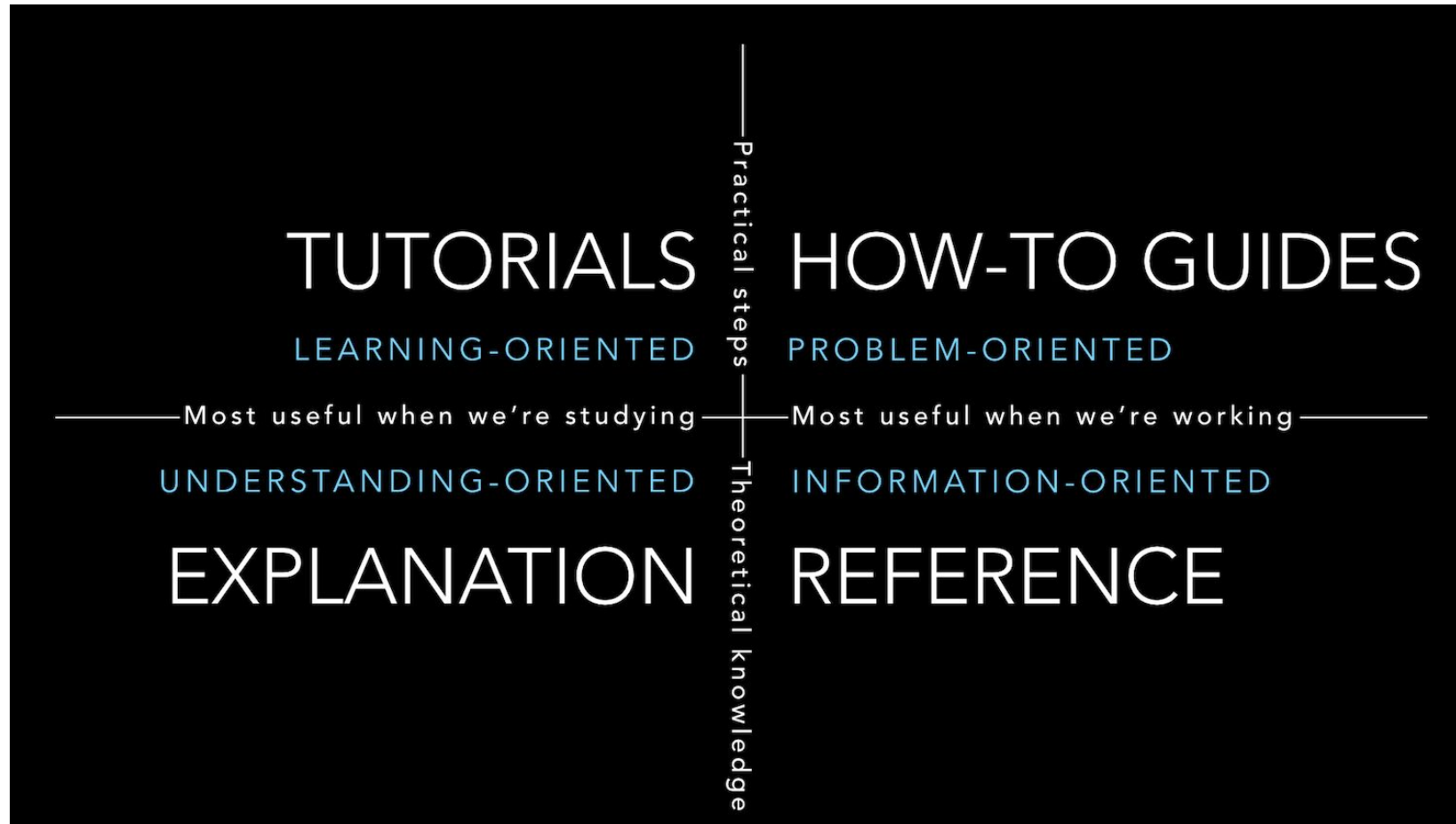


What is Apache Flink used for?





Communities create documentation with a purpose



[1] <https://documentation.divio.com/>



How do you know that your documentation works as intended?





But: ASF projects should not use Google Analytics at all

Project Websites

Can I use Google Analytics?

The Apache Software Foundation discourages the use of Google Analytics.

Court decisions around Google Analytics have changed several times in the past years and made its use uncertain. Several countries in the EU declared [Google Analytics cannot be used compliant to the GDPR.](#)

To avoid legal risks, ASF projects shall not use Google Analytics at all.

[1] <https://privacy.apache.org/faq/committers.html>



How can we get answers to questions like:

1. Are the readers looking at our Java APIs or our Python APIs documentation?
2. Can the readers of our documentation find what they are looking for?
3. Is our content engaging?
4. What is the last page they see before leaving the documentation?
5. What are users downloading via our documentation?



Matomo: A privacy friendly alternative to Google Analytics

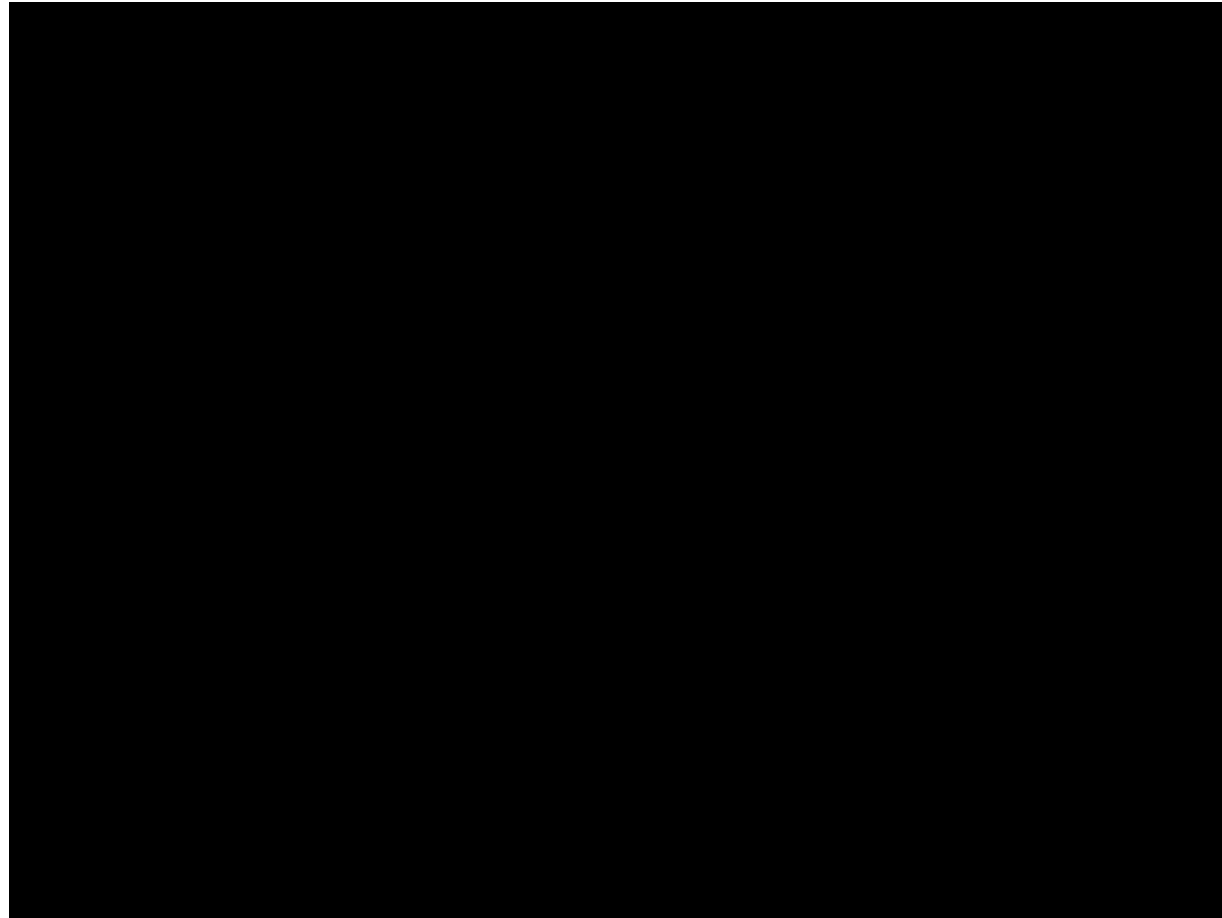
1. Discussion Flink Dev mailing list
2. Discussion on Privacy mailing list
 - a. Privacy impact
 - b. Self-hosted vs SaaS solution
 - c. Make data available for everyone
3. Created local demonstration
4. Agreed to setup Matomo and implement it for Flink
5. If successful, we could share it with other ASF projects

The screenshot displays the Matomo dashboard interface. The top navigation bar includes 'matomo', 'Dashboard', 'All Websites', and 'Tag Manager'. The main content area is titled 'Visits Log' and shows two visitor sessions. The first session, dated Thursday, December 9, 2021, at 12:08:37, shows a visitor from IP 172.20.0.0 in the United Kingdom. The 'Actions' list for this session includes: 'Apache Flink: How To Contribute', 'Apache Flink: Stateful Computations over Data Streams', 'Apache Flink: Stateful Functions - Event-driven Applications on Apache Flink', 'Apache Flink: Powered by Flink', 'Apache Flink: Downloads', and 'Apache Flink: Downloads - apache-flink-1140'. The second session, dated Thursday, December 9, 2021, at 11:10:16, shows a visitor from IP 172.20.0.0 in the United Kingdom. The 'Actions' list for this session includes: 'Apache Flink: What is Apache Flink? - Architecture' and 'Apache Flink: What is Apache Flink? - Applications'. The bottom of the screenshot shows a video player interface with a red progress bar and a timestamp of 8:11 / 11:33.



What can you do with Matomo?

- You can see it yourself at <https://analytics.apache.org>





Example 1: Are readers looking at the Java APIs or the Python APIs?

Pages						
PAGE URL	PAGEVIEWS	UNIQUE PAGEVIEWS	BOUNCE RATE	AVG. TIME ON PAGE	EXIT RATE	AVG. PAGE LOAD TIME
flink	205,189	167,504	43%	00:01:45	32%	3.94s
flink-docs-release-1.15	75,630	62,068	43%	00:01:46	27%	3.32s
docs	36,849	30,888	48%	00:01:59	31%	3.18s
dev	14,580	12,030	47%	00:01:56	29%	2.79s
table	6,627	5,481	49%	00:01:53	30%	3.78s
datastream	5,247	4,350	47%	00:02:07	32%	2.07s
python	1,275	1,045	41%	00:01:41	20%	3.03s
configuration			38%	00:01:44	19%	1.18s
dataset	181	151	36%	00:02:02	24%	3.9s

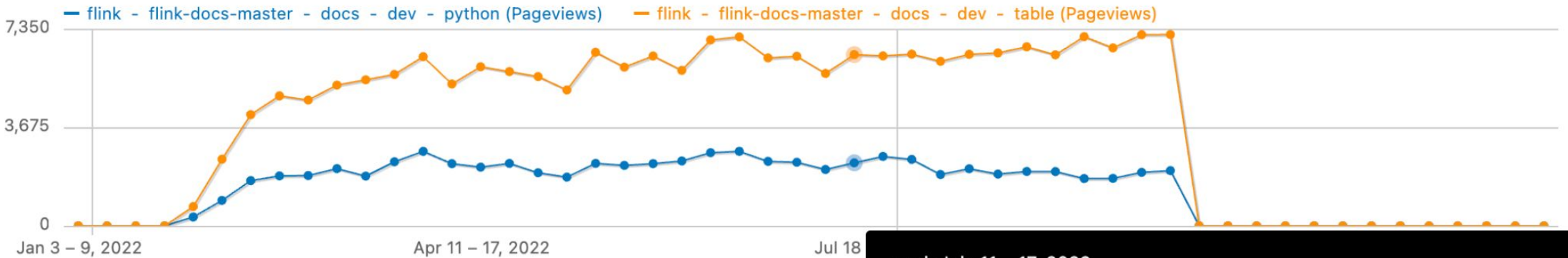


Open Row Evolution
See how the metrics for this row changed over time



Example 1: Are these results changing over a period of time?

Evolution of multiple rows



week July 11 - 17, 2022

- 2,350 flink - flink-docs-master - docs - dev - python (Pageviews)
- 6,384 flink - flink-docs-master - docs - dev - table (Pageviews)

🔗 🖼️ 💬 📅 Week

Page URL: Comparing 2 rows

- flink - flink-docs-master - docs - dev - python (Pageviews) from 0 to 73,219
- flink - flink-docs-master - docs - dev - table (Pageviews) from 0 to 202,998

» [Pick another row to compare](#)

Available metrics



Example 2: Can readers find what they are looking for?

Site Search Keywords

KEYWORD	SEARCHES	SEARCH RESULTS PAGES	% SEARCH EXITS
sink	41	1.2	2%
checkpoint	36	1.1	6%
sql	35	1.2	6%
join	33	1.4	3%
window	33	1.2	3%
cdc	31	1.3	23%
kafka	30	1.1	3%
yarn	29	1.1	3%
state	27	1.4	4%
watermark	25	1.5	4%

1-10 of 500 [Next >](#)

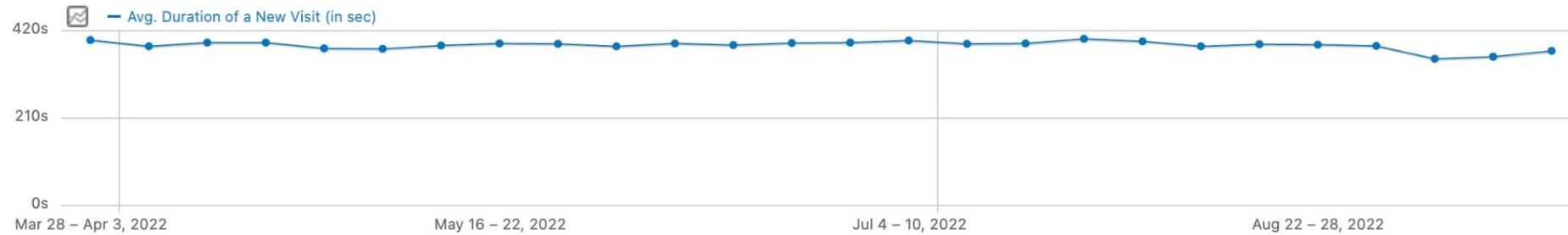
Pages Following a Site Search

DESTINATION PAGE	CLICKED IN SEARCH RESULTS	PAGEVIEWS	AVG. PAGE LOAD TIME
flink/flink-docs-release-1.15/docs/d...	68	782	2.61s
flink/flink-docs-release-1.15/docs/d...	37	326	1.78s
flink/flink-docs-master/docs/deploy...	36	908	2.33s
flink/flink-docs-release-1.15/docs/d...	35	268	2.49s
flink/flink-docs-stable/index	30	2,791	1.91s
flink/flink-docs-release-1.15/docs/c...	30	622	4.46s
flink/flink-docs-release-1.15/docs/d...	30	346	1.72s
flink/flink-docs-release-1.15/index	28	4,774	3.77s
flink/flink-docs-release-1.15/docs/d...	25	227	1.69s
flink/flink-docs-release-1.15/docs/d...	24	199	1.5s
flink/flink-docs-release-1.15/zh/index	23	1,112	1.53s
flink/flink-docs-release-1.15/docs/d...	21	242	18.39s



Example 3: Is our content engaging?

Average Visit Duration



Frequency Overview

8 returning visits

17 min 33s average visit duration for returning visitors

13.4 actions per returning visit

30% returning visits have bounced (left the website after one page)

107 actions by the returning visits

69,070 new visits

6 min 10s average visit duration for new visitors

4 actions per new visit

42% new visits have bounced (left the website after one page)

275,369 actions by the new visits



Example 4: What is the last page they see before leaving?





Exit pages

EXIT PAGE URL	EXITS	UNIQUE PAGEVIEWS	EXIT RATE	AVG. PAGE LOAD TIME
flink	1,798,109	5,794,107	31%	3.18s
/index	161,627	478,285	34%	4.07s
zh	80,559	282,847	28%	2.77s
/downloads.html	55,574	109,345	51%	1.74s
news	49,604	97,233	51%	6.09s
/flink-architecture.html	34,787	80,693	43%	2.21s
/usecases.html	22,355	46,800	48%	3.4s
2021	14,189	26,317	54%	4.98s
features	9,274	14,310	65%	6.69s
2018	6,369	9,412	68%	8.5s
03	5,345	7,458	72%	8.41s
01	5,345	7,458	72%	8.41s
/end-to-end-exactly-once-apache-flink.html	5,300	7,393	72%	8.23s



Example 5: What are users downloading via our documentation?

Downloads

DOWNLOAD URL	UNIQUE DOWNLOADS	DOWNLOADS
 www.apache.org	73,462	86,016
 /dyn/closer.lua/flink/flink-1.15.0/flink-1.15.0-bin-scala_2.12.tgz	7,044	8,591
 /dyn/closer.lua/flink/flink-1.14.4/flink-1.14.4-bin-scala_2.11.tgz	5,771	6,886
 /dyn/closer.lua/flink/flink-1.13.6/flink-1.13.6-bin-scala_2.12.tgz	5,452	6,285
 /dyn/closer.lua/flink/flink-1.13.6/flink-1.13.6-bin-scala_2.11.tgz	5,345	6,165

Conversions Overview

- Your best converting countries are: [China](#) , [United States](#)  and [Hong Kong SAR China](#) 
- Your top converting keywords are: [flink](#), [flink官网](#) and [flink 下载](#)
- Your best converting websites referrers are: [nightlies.apache.org](#), [link.csdn.net](#) and [blog.csdn.net](#)
- Your best converting entry page is: [flink](#)
- Returning visitors conversion rate is **3.04%**, New visitors conversion rate is **2.3%**

 [Show Visits Log segmented by this Goal](#)

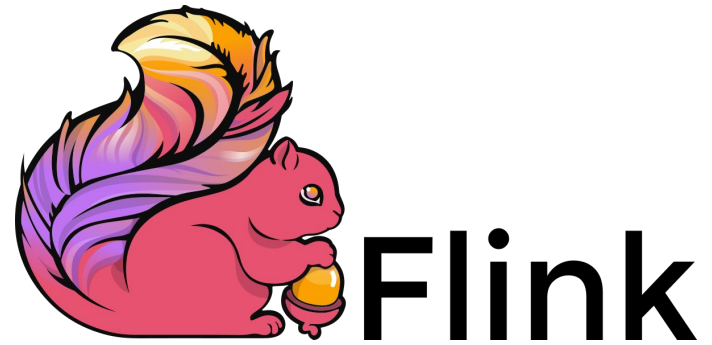


Take away for using web analytics in your ASF project

1. Determine its goal
2. Implement
3. Analyze and conclude
4. Improve

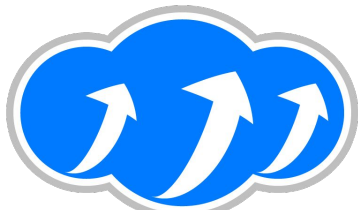


Apache Flink was the first ASF project to start with Matomo





There are now 31 Apache Software Foundation projects onboarded





Reach out if you want to be onboarded

- Via #privacy-matomo in the ASF Slack
- Via privacy@apache.org

