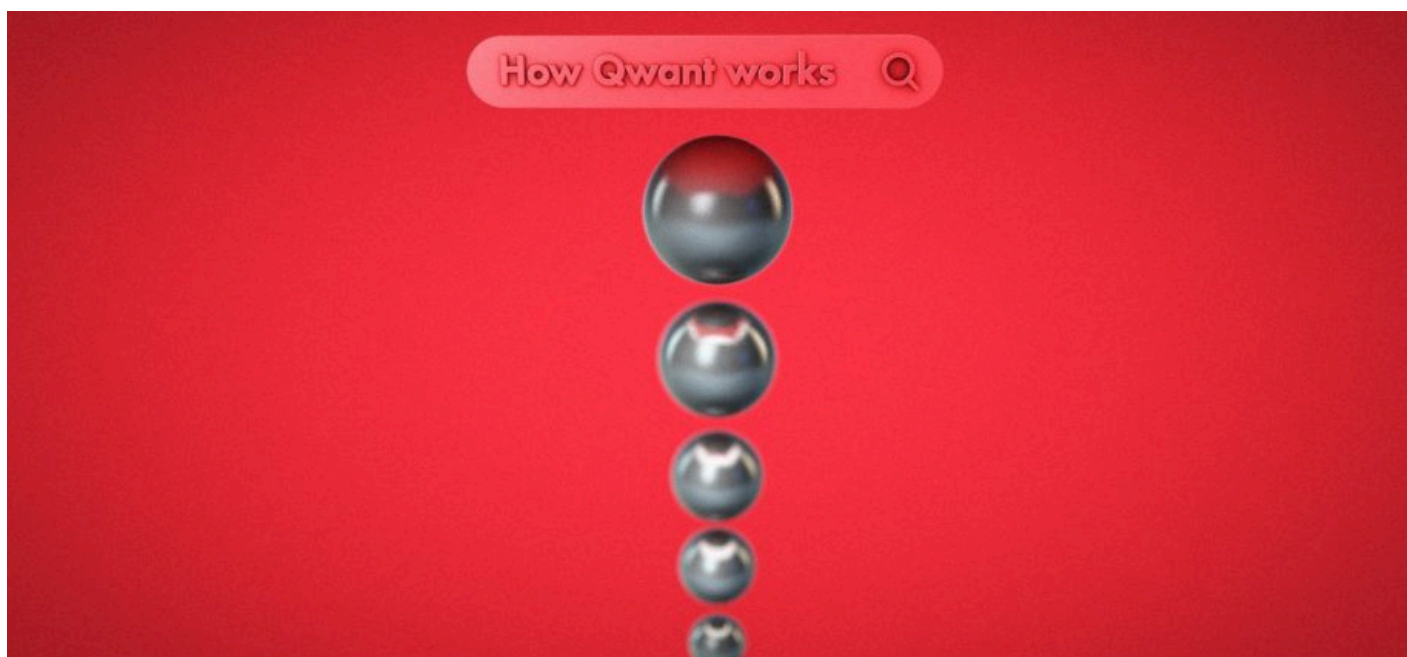


Better Web, Qwant News

How Microsoft tools strengthen Qwant



On the occasion of VivaTech 2019, Qwant and Microsoft announced together a partnership to support Qwant's strong growth while continuing to respect the privacy of its users.

Team Qwant 14 June 2019 ⌚ 6min

On the occasion of VivaTech 2019, Qwant and Microsoft announced together a partnership to support Qwant's strong growth while continuing to respect the privacy of its users. Thus, Qwant will be able to rent to Microsoft computing and storage power that will be used on data to index the Web.

Let one thing be clear: the interest for Microsoft is to provide computing power to Qwant to meet its exponential growth. That's all. 🗨️ [@JP_Courtois](#) [#VivaTech](#)
pic.twitter.com/W7cHzwyleY

— Qwant (@Qwant_FR) [May 17, 2019](#)

This announcement is the result of your interest in our services: the growth of requests processed by our servers is constantly increasing, and at this rate our infrastructures will soon be underwater and slow to respond! So we decided to review our entire infrastructure to focus our servers on protecting your data, and migrate some of what is not sensitive to cloud machines. Since we are continually looking to improve our quality, this strategy will also allow us to increase our Web coverage with additional means for our indexing.

Today, to design the search engine you know, we manage more than 400 servers that we own and that we have installed in a data center in the Paris region. They are used for several purposes: the front (what you see), crawl (the discovery and updating of content), the storage of texts and images, indexing, mapping, news, our internal tools and many others. It goes without saying that almost two-thirds of this machine park is reserved for the proper functioning of Qwant Search (whether production or pre-production). However, our technical resources are not infinitely extensible, which is an obstacle to our development while the purchase and renewal of equipment is very expensive. Thus, we will take advantage of the on-demand resources of Microsoft Azure that we will rent to absorb and increase our development. At the same time, we will take advantage of this flexibility to continue to invest in our sovereign infrastructure.

To understand what will happen with this partnership, I suggest you follow the journey of the web pages in our servers.

25 machines visit 12,000 web pages per second

The first step to creating a search engine is to discover what makes the internet, what content it consists of. Indeed, apart from domain names and IP addresses,

nothing is standardized. You can create the site with the structure and content you want without limitation. For us to know about it, we need to run crawlers, or crawlers in our jargon. Today it is about 25 machines that are dedicated to this task, they are able to visit 12,000 web pages per second, or 1 billion per day. In addition to discovering new sites, bots must update content already visited, a key case for news sites.

At the exit of the crawler the data extracted from the web are raw, they must be analyzed before entering our databases. This is where we use computer analysis programs that we call microservices. They are used to determine, for example: the language of the page, its semantic domain, but also approximate graph scores. We will use Azure's Kubernetes service to easily deploy and operate production analytics and research microservices. These will improve the relevance of the engine in the future.

20 billion indexed documents

When the crawled data is analyzed and formatted, it must be indexed. That is to say, to structure them as in the indexes found at the end of the books, to be able to find them. The index in a book gives the correspondence between a word and the page number where it appears. For the internet the application is almost identical except for the number of documents we have to index. Today, our server fleet allows us to reach 20 billion indexed documents, or several hundred terabytes of data.

However, our crawlers discover more documents than can be indexed because of the physical limit of the number of servers available. The crawler downloads nearly 10 gigabytes per second. So currently we prioritize indexing based on the language and graph score of the document, for example. Web pages that are not indexed are stored in an archiving system, this represents nearly 2 petabytes of data. By using, Azure's storage solutions we will be able to free up space at home to increase the size of our indexes and subsequently increase our quality and coverage of the internet. In addition, we will be able to use Azure resources to test new indexing

methods or languages before moving them into production. This is a necessary flexibility to deliver functional services easily.

Strengthening our graph calculation capacity

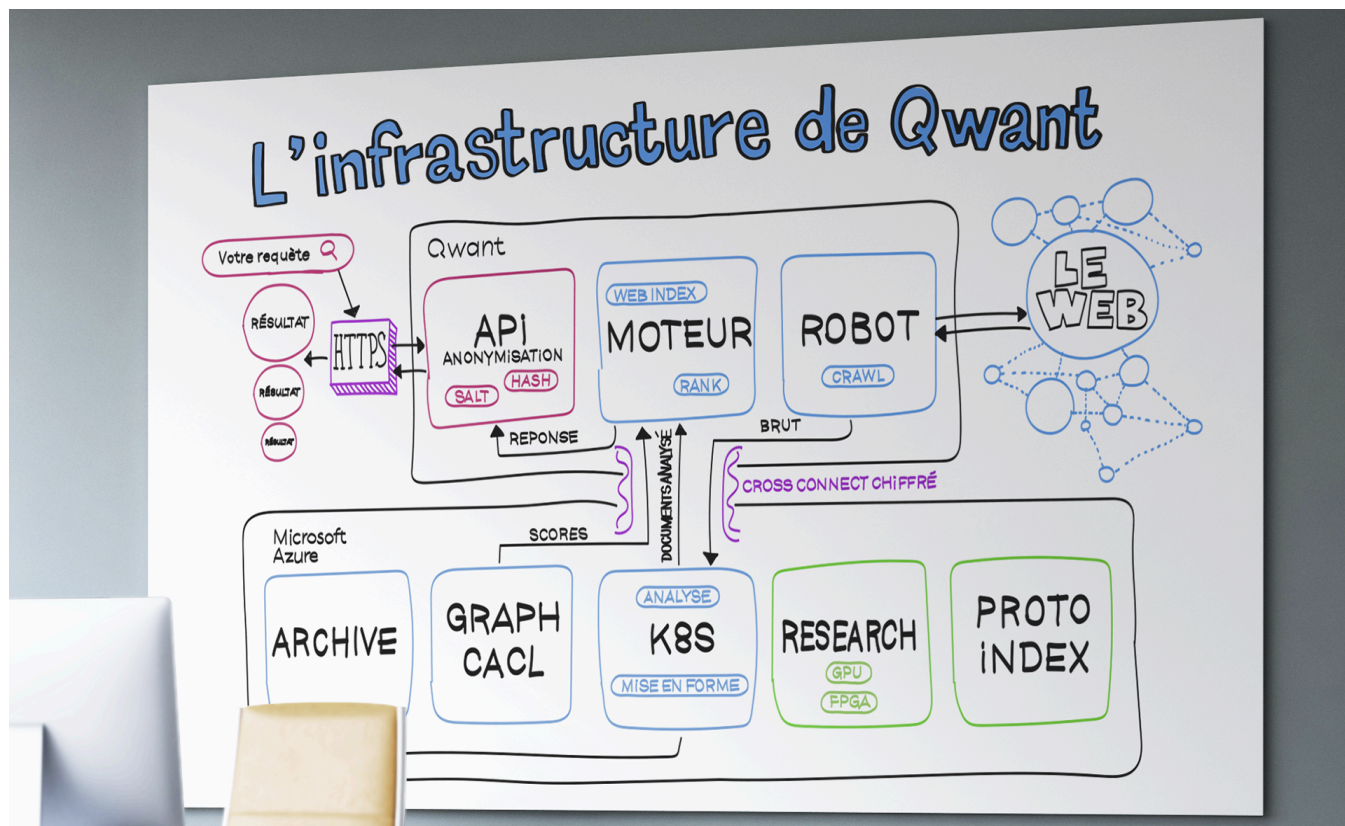
When the index is ready we can easily find all the documents that are related to one or more keywords. However, let's take the example of a "kitten" that will bring up tens of millions of web pages from our indexes. To reorder all this, we need to order these pages, using graph scores and machine learning methods.

Calculating graph scores is very expensive, on the one hand because the web graph represents terabytes of data. In addition, because it requires several servers working in parallel to extract it. We will be able to use Azure resources on demand to calculate the graph or train our Learning To Rank models.

A team of specialized researchers to improve the index

For nearly a year and a half, we have strengthened our team of specialized researchers (in imaging, high-performance computing, language processing, cryptography...) to improve the Qwant search engine. We already use very powerful graphics cards in Nvidia's DGX servers. These servers are used to train machine learning models, for machine translation, image search, graph algorithm optimization. We also work on document analysis technologies or information retrieval using FPGAs.

Today, an advanced project is the image search that integrates with our crawler and must move to the scale of production. We have to store hundreds of millions of images, or several hundred terabytes of images. Here, the object storage that Azure offers will be very useful to increase the size of our image index and facilitate model learning.



User does not interact with Microsoft

Finally, let's talk about you, who use us. When you connect to Qwant, you will always be exclusively in contact with machines owned, installed and administered by us. You will never be connected to machines in the Azure cloud, and your personal data will never be sent there. We use Azure cloud for Qwant's only "back office". In addition, and for the record, as soon as your connection is established, we anonymize your data, in particular your IP which is salted and chopped. That is, we add noise and a clipping of this IP to make it anonymous when we ask to display advertisements or need to store data in our logs. Only this anonymous data circulates in our internal network.

Relative posts



Better Web, Privacy

#SaferInternetDay: Six tools to protect children on the Internet

Team Qwant

9 February 2021 ⌚ 3 min



Better Web

#BetterWeb 03: Guillaume Chaslot – AlgoTransparency to break filter bubbles

Team Qwant

2 April 2021 ⌚ 1 min



Better Web

Qwant is involved in the Online Child Protection Laboratory

admin

20 December 2022 ⌚ 4 min



The search engine where you are the user, not the product!

Discover the service

Services

Search

Junior

App Android

App iOS

About Qwant

Why Qwant?

Better Web

Join Us

Help Center

Contact Us

Press

Follow us !



[Privacy](#)

[Terms of Service](#)

[Legal notices](#)

[Ranking](#)

© 2026 Qwant. All right reserved.