

Codename Cassandra: Backend



A modular approach for using
Wiki* GLAM's data



Good architectures for future proven applications



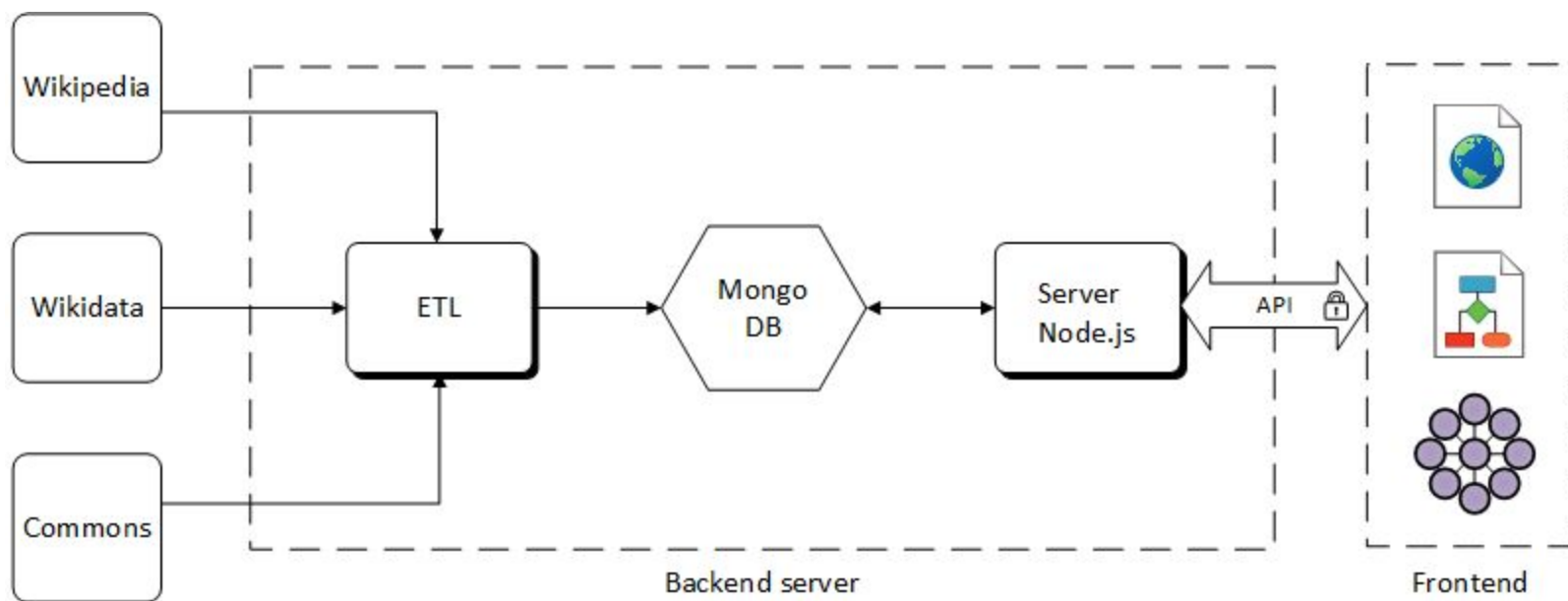
Like in house building, foundations are important:
they are invisible but enable all the features

A modular data driven architecture

Styles change over time, interfaces may become obsolete and when software age, tends to not work anymore!

Having data only drowned in visualizations makes difficult changes and new features.

To afford all the needs of tomorrow a step at a time, a **modular architecture**, with **data as protagonist**.



Backend architecture

GET

An ETL service extract data from Wiki*, parsing and loading them to the database

STORE

A MongoDB instance store all the data collected

SERVE

A node.js server exposes REST API to interact with the database

Make it modular!

GET

Data are taken via SQL queries or API: it's simple to change an existing rule or to add a new one

If data format change we also change with little effort

STORE

We choose MongoDB because is a noSQL database (so is schemaless)

Old schemas and decisions aren't blocking anymore

SERVE

Data are served via HTTP REST API: it's easier for who write applications, it's fast to adapt to new data

Infinite frontends are possible with zero changes here

API

- **Modularity:** it's easy to add new ones and change the existing
- **Security:** authentication and encrypted connection
- **Applications:** you can build any kind of service (websites, dashboards, reports, visualizations) or integrate with existing ones

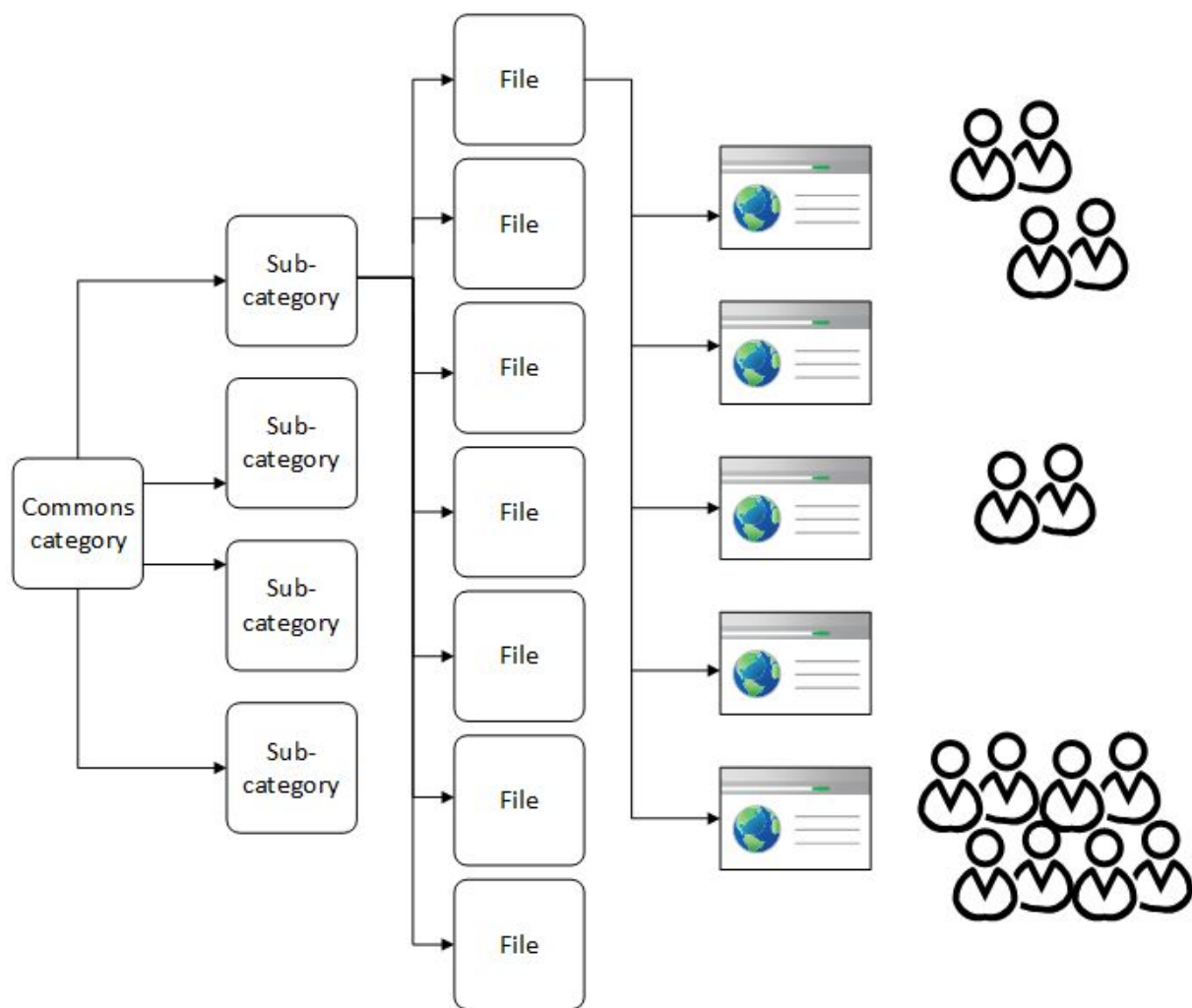
Category tree	
Lista delle categorie con nome e numero di file presenti.	
/category	GET

Uploads	
Numero di file caricati raggruppati per giorno di caricamento.	
/file/upload-date	GET

Which data Wiki*world provide?



Let's have a look on
what are the data that flow here...



3 collections of data

CATEGORY

- category name
- category parent
- category level
- n° files inside

FILE

- file name
- upload username
- upload date
- file size
- category

USAGE

- file name
- wiki project
- page url
- visits

Conclusions



So what are the abilities of this architecture
with that data?

Answer to (simple) questions

- Which is the file most used?
- Which wiki project use the higher number of files from this category?
- Which is the Wikipedia single page with the highest number of files?

Try to answer to hard questions

- Why those are the most used file?
- Why many high quality files are not used?
- Which is the Wikipedia field I have improved with this group of files?

filename	size	time	year	cat
spelterini zürich city.jpg		788866 2008-04-01T05:27:16.000Z	2008	Historical_images_of Rathausbrücke Zürich
hundert jahre bilder aus der stadt zürich - das polytechnikum 1865.jpg		110242 2010-02-27T00:29:38.000Z	2010	Historical_images_of buildings_in Zürich
prozessionsachse murerplan.jpg		594863 2006-04-19T14:21:46.000Z	2006	Historical_images_of Münsterbrücke Zürich
destruction of icons in zurich 1524.jpg		1338249 2010-11-28T02:55:36.000Z	2010	Historical_images_of Wasserkirche, Zürich
rennwegtor.jpg		380879 2008-03-13T09:52:58.000Z	2008	Rennwegtor
bild_fraumuenster.jpg		1056174 2006-04-05T11:20:32.000Z	2006	Historical_images_of Kornhaus, Zürich
zentralbibliothek zürich - zürich bahnhofplatz - 000006617_1.jpg		2976084 2013-03-22T10:19:43.000Z	2013	Historical_black_and_white_photographs_of Zürich Main Station
herrliberger_elgg_1740.jpg		1291944 2008-04-06T07:49:21.000Z	2008	Eigentliche_Vorstellung_der Adelichen Schlösser im Zürich Gebieth
rathaus_zuerich1780.jpg		6355824 2009-10-20T08:03:05.000Z	2009	Historical_images_of Rathaus Zürich
regensberg - ulrich von regensberg 20100524_18-27-34_shiftn.jpg		3819734 2010-05-25T00:03:40.000Z	2010	Kloster Oetenbach
rathaussaal_zuerich.jpg		1240673 2007-05-31T12:25:53.000Z	2007	Historical_images_of Rathaus Zürich
herrliberger_kyburg1740.jpg		1236757 2008-04-05T09:40:34.000Z	2008	Vorstellung_Loblichen Standts Zürich Schlösser oder so genante Ausere
zürich_bullinger.jpg		1230385 2008-03-25T16:36:57.000Z	2008	Wellenberg
zürich - oerlikon - gygerkarte_1667.jpg		388169 2008-09-20T00:24:10.000Z	2008	Kloster St. Martin (Zürich)
lokomotive_rhein_1867.jpg		581159 2008-09-28T09:52:31.000Z	2008	Historical_black_and_white_photographs_of Zürich Main Station
coop_provisorium.jpg		8545266 2016-07-31T05:10:34.000Z	2016	Papierwerd
spelterini limmtraum.jpg		877076 2007-12-28T11:54:50.000Z	2007	Historical_photographs_of Gedecktes Brüggli
bild_maschinenhalle_escher_wyss_1875.jpg		1639628 2014-03-07T08:42:09.000Z	2014	Historical_images_of buildings_in Zürich
herrliberger_mur_1741.jpg		1040644 2008-04-06T08:33:02.000Z	2008	Eigentliche_Vorstellung_der Adelichen Schlösser im Zürich Gebieth
dübendorf - dübelstein - gerold edlibach_1489.jpg		4172197 2008-09-19T22:28:34.000Z	2008	Historical_images_of Fraumünster, Zürich
zürich_kantonsschule.jpg		288948 2006-09-06T07:47:55.000Z	2006	Historical_images_of buildings_in Zürich
zürich - wasserkirche - bürgerbibliothek_1719_img_1196.jpg		3583580 2008-10-22T18:36:27.000Z	2008	Historical_images_of Wasserkirche, Zürich
barfüsser_arkade4.jpg		1482433 2008-07-17T15:11:17.000Z	2008	Historical_images_of Barfüsserkloster, Zürich
mainstation-zuerich.jpg		422200 2007-12-16T14:37:40.000Z	2007	Historical_black_and_white_photographs_of Zürich Main Station
uetliberg_leu.jpg		1337713 2008-03-22T07:01:36.000Z	2008	Kloster Selnau
zürich_silberysen.jpg		1326331 2008-03-29T17:32:06.000Z	2008	Grendeltor
herrliberger_kloster_kappel_1741.jpg		1156147 2008-04-05T19:50:09.000Z	2008	Vorstellung_Loblichen Standts Zürich So genante Ausere Amtheusser
paradeplatz_ca_1888.jpg		283773 2006-09-26T16:24:33.000Z	2006	Historical_images_of St. Peter Church in Zürich
herrliberger - schloss knonau.jpg		262325 2009-05-06T22:18:23.000Z	2009	Eigentliche_Vorstellung_der Adelichen Schlösser im Zürich Gebieth
zürich - lindenhof - stützmauer img_1382.jpg		9769149 2011-06-15T21:37:40.000Z	2011	Schweizer Heimatwerk
	avg size best	avg date		
	1,89391492207845	Sept 2008		
	avg size			
	4,18647827148438			
	avg size best new			
	5,46696829795837			

Made in - synapta

<https://synapta.it>

@synapta

CC-BY

alessio.melandri@synapta.it

@AlessioMelandri

Keep in touch!