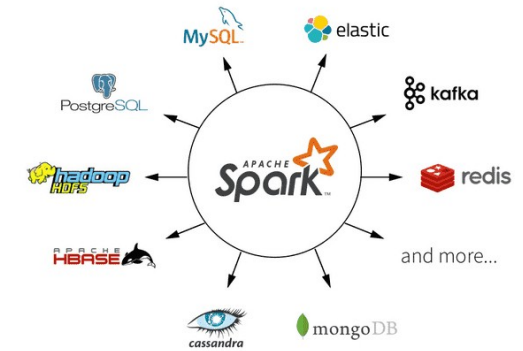


Big Data & DataViz

Ingestion, Processing, Visualisation



Master 1 Data Engineer

Réalisé par : Lisa OULD OULHADJ

Rayane HAROUN

Yacine BOUFALA

Plan de présentation

- I. Introduction
- II. Déroulement du projet
 - 1. Présentation du sujet
 - 2. Présentation de l'architecture
 - 3. Le choix des technos
 - 4. Présentation Dashboard
 - 5. Difficultés rencontrées
- III. Conclusion



I. Introduction

Comment développer une pipeline big data qui sera capable de traiter une forte volumétrie de données à intervalles réguliers (batch) ou des données en temps réel (streaming) ?

The slide features decorative curved lines in shades of blue and green, appearing as if they are part of a larger circular design, framing the central text.

II. Déroulement du projet

1. Présentation des sources de données



Santé

A	B	C	D	E	F	G	H	I	J	K	
ID	dep	sexe	Date	Hospitalisation	Reanimation	HospConv	Radio	Deces	code_region	incid_reanimation	
0	1	0	14/03/2021	183	19	111.0	2234	538	84	8	
1	1	1	14/03/2021	88	13	52.0	1083	319	84	8	
2	1	2	14/03/2021	94	6	58.0	1132	216	84	8	
3	3	0	14/03/2021	73	13	44.0	1676	502	84	8	
4	3	1	14/03/2021	40	8	25.0	778	286	84	8	
5	3	2	14/03/2021	32	5	18.0	881	207	84	8	
6	7	0	14/03/2021	108	11	86.0	1644	439	84	8	
7	7	1	14/03/2021	46	10	32.0	737	264	84	8	
8	7	2	14/03/2021	59	1	51.0	875	172	84	8	
9	15	0	14/03/2021	25	2	17.0	453	103	84	8	
10	15	1	14/03/2021	17	0	13.0	218	57	84	8	
11	15	2	14/03/2021	8	2	4.0	235	43	84	8	
12	26	0	14/03/2021	220	40	122.0	2226	648	84	8	
13	26	1	14/03/2021	126	30	68.0	1101	362	84	8	
14	26	2	14/03/2021	89	9	52.0	1087	273	84	8	
15	38	0	14/03/2021	626	66	309.0	5614	1423	84	8	
16	38	1	14/03/2021	319	45	150.0	2735	822	84	8	
17	38	2	14/03/2021	303	20	157.0	2847	593	84	8	
18	42	0	14/03/2021	352	41	229.0	4229	1073	84	8	
19	42	1	14/03/2021	178	34	118.0	2125	619	84	8	
20	42	2	14/03/2021	173	7	110.0	2073	446	84	8	
21	43	0	14/03/2021	75	5	43.0	1045	212	84	8	
22	43	1	14/03/2021	38	3	26.0	500	122	84	8	
23	43	2	14/03/2021	35	2	15.0	529	88	84	8	
24	63	0	14/03/2021	227	29	122.0	1943	574	84	8	
25	63	1	14/03/2021	110	22	59.0	903	315	84	8	
26	63	2	14/03/2021	112	7	62.0	1025	256	84	8	
27	69	0	14/03/2021	888	161	495.0	11493	2492	84	8	
28	69	1	14/03/2021	456	110	248.0	5641	1452	84	8	
29	69	2	14/03/2021	425	50	243.0	5717	1022	84	8	
30	73	0	14/03/2021	102	15	67.0	2261	604	84	8	

1. Présentation des sources de données



Emploi

B	C	D	E	F	G	H	I	J	K
libelle_zone_d_emploi	code_region	region	code_zone_d_emploi	zone_d_emploi	Date	effectifs_salaries_brut	effectifs_salaries_cvs	masse_salariale_brut	masse_salariale_cvs
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	31/03/1998	821	806	2893087	2829823
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/06/1999	795	811	2955897	2903472
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/09/2000	848	875	3187211	3262555
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	31/03/2002	957	954	3613175	3570590
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/06/2003	1014	1010	3819669	3747364
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	31/12/2005	1002	996	4301553	4289172
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	31/03/2006	1002	997	4176864	4151645
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/06/2008	1268	1261	6028984	5949849
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/06/2009	1339	1332	6187898	6120766
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/09/2010	1314	1348	6324586	6366036
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/06/2011	1493	1487	7365087	7285491
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	31/03/2012	1400	1402	7005538	7213990
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/06/2012	1272	1266	7044778	6944017
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/09/2012	1245	1281	7203062	7155713
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	31/03/2013	1373	1371	7437650	7679130
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/06/2013	1402	1391	8114887	7959301
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	31/12/2013	1491	1474	7999281	7984903
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/06/2015	1483	1461	8568927	8271665
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/09/2015	1428	1472	8494057	8461876
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/09/2017	1401	1460	8706665	8747322
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	31/12/2017	1471	1444	8678398	8826419
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	31/03/2018	1472	1445	8722210	8962642
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/09/2019	1469	1544	9675477	9763416
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	31/12/2020	1587	1541	9388785	9572204
0101 Côte sous le vent		1 Guadeloupe	101	Côte sous le vent	30/06/2021	1525	1548	8512936	8132396
0102 Est Grande Terre		1 Guadeloupe	102	Est Grande Terre	31/03/1998	3388	3286	12843971	12837759
0102 Est Grande Terre		1 Guadeloupe	102	Est Grande Terre	30/06/1998	3409	3290	13632561	12994737
0102 Est Grande Terre		1 Guadeloupe	102	Est Grande Terre	31/12/1998	3219	3302	13782591	13728669
0102 Est Grande Terre		1 Guadeloupe	102	Est Grande Terre	30/06/2000	3486	3398	14491072	13904952
0102 Est Grande Terre		1 Guadeloupe	102	Est Grande Terre	30/09/2000	3342	3467	13359794	14020767

1. Présentation des sources de données



Emploi

id	type	natureObjet	referenceCPV	datePublicationDonnees	Montant	formePrix	nomAcheteur	Latitude Ach	Longitude Ac	categorieEta	distanceAcheteurEtablissement	Latitude Etablissement	Longitude Etablissement
20202020S1	Marché	Fournitures	Cartes accréditives	17/08/2020	0.01	Ferme	Etablissements et organ	48.82685427	2.129822351	GE	23.635675145997322	48.8769868607	2.33753841595
2,0211E+15	Marché	Travaux	Pose de clôtures	09/03/2021	0.01	Révisable	Ministère de la Transitio	49.44150688	1.093732732	PME	64.26344479631022	48.8642753222	1.08005278559
20192019S1	Marché	Services	Services de cafétéria	25/07/2019	0.01	Ferme et actualisabl	Chambres de Commerce	45.18217821	5.720990897	NC	109.41532685594025	45.7041579754	4.88121408321
2,0192E+15	Marché	Services	Services de traitement et d'e	21/05/2019	0.01	Ferme	Ministère des armées - Centre National de Soutien ETI					48.7728906673	2.07603849225
2,0192E+15	Marché	Services	Services de contrôle techniq	24/12/2019	0.01	Révisable	Ministère de la Transitio	49.44150688	1.093732732	NC	124.08684248693082	48.7759550031	1.99412790053
20192019S1	Marché	Services	Test et évaluation	14/06/2019	0.01	Ferme	Chambre de Commerce	47.39807735	0.696151819	PME	0.0	47.3980773574	0.696151819266
2,0181E+15	Marché	Services	Services topographiques	05/11/2018	0.01	Révisable	Ministère de la Transitio	49.44150688	1.093732732	PME	0.0	49.4415068863	1.09373273216
20182019F5	Marché	Fournitures	Appareils d'hémodialyse	11/01/2018	0.01	Ferme	CHRU NANCY	48.69006710	6.175948324	NC	435.662264724046	48.8365394869	2.2389245247
2,0201E+15	Marché	Fournitures	Équipement de contrôle de l	27/05/2020	0.01	Révisable	Ministère des armées - Plate-forme commissariat			PME		48.5716343863	7.76783353204
20192019S1	Marché	Fournitures	Lait	01/06/2019	0.01	Révisable	CHI COMPIEGNE NOYON	49.39926582	2.853836007	GE	449.4574801630858	45.8310057435	4.77089761479
20192019S1	Marché	Services	Test et évaluation	14/06/2019	0.01	Ferme	Chambre de Commerce	47.39807735	0.696151819	PME	0.0	47.3980773574	0.696151819266
2,0211E+15	Marché	Services	Services topographiques	04/05/2021	0.01	Révisable	Ministère de la Transitio	49.44150688	1.093732732	PME	381.84043718453955	47.2321613301	-1.54797539458
2,0201E+15	Marché	Services	Services de gestion de projet	15/04/2020	0.01	Révisable	Ministère de la Transitio	49.44150688	1.093732732	NC	582.2077632642673	45.7731043316	4.85202749669
20202020F0	Marché	Fournitures	Produits pharmaceutiques	05/03/2020	0.01	Ferme et actualisabl	CH DU MANS	47.98894335	0.200024013	PME	254.0944621440304	48.8129348783	2.34310649123
20202020S1	Marché	Services	Services de traduction	10/08/2020	0.01	Ferme	CHU SAINT ETIENNE	45.43019604	4.378086977	NC	0.0	45.4301960415	4.37808697748
2,0201E+15	Marché	Services	Services d'essais techniques	12/11/2020	0.01	Révisable	Ministère de la Transitio	49.44150688	1.093732732	NC	340.7273609599376	48.111571436	-1.68168588816
2,0212E+15	Marché	Travaux	Ouvrages ferroviaires	31/03/2021	0.01	Ferme et actualisabl	Ministère des armées - Etablissement du service d'ETI					43.5360639608	5.39837823449
20192019S1	Marché	Services	Services d'assurance tous ris	25/06/2019	0.01	Ferme	Ministère de la Culture -	48.82827064	2.362358986	NC	6.073971962706373	48.8769868607	2.33753841595
2,0191E+15	Marché	Services	Services de réparation et d'e	03/04/2019	0.01	Révisable	Ministère de la Transitio	49.44150688	1.093732732	GE	160.35820969812542	49.1546776305	-0.327460237934
20202020F2	Marché	Fournitures	Tétines, téterelles et articles	23/12/2020	0.01	Révisable	HOSPICES CIVILS DE LYO	45.74804931	4.825675844	PME	60.77598344914019	45.4301960415	4.37808697748
20192019S2	Marché	Services	Services de réparation, d'ent	21/10/2019	0.01	Ferme	CHU SAINT ETIENNE	45.43019604	4.378086977	NC	68.0078815387777	45.7345274274	4.91192378092
20202020F0	Marché	Fournitures	Lait	29/01/2020	0.01	Ferme	CHI COMPIEGNE NOYON	49.39926582	2.853836007	GE	84.10557338393673	48.8951908356	2.28680331405
2,0181E+15	Marché	Travaux	Entretien des routes nationa	03/12/2018	0.01	Révisable	Ministère de la Transitio	49.44150688	1.093732732	PME	82.44926551775377	49.712052838	0.399604094774
20202020F1	Marché	Fournitures	Livres, brochures et dépliant	05/10/2020	0.01	Ferme et actualisabl	CHU SAINT ETIENNE	45.43019604	4.378086977	PME	0.0	45.4301960415	4.37808697748
20192019F1	Marché	Fournitures		29/05/2019	0.01	Ferme	Organisme de formation	48.87278476	2.312530916	NC			
2,0181E+15	Marché	Services	Services de relations publicu	10/12/2018	0.01	Révisable	Ministère de la Transitio	49.44150688	1.093732732	NC			
20192019S1	Marché	Services	Services d'assurance respons	24/06/2019	0.01	Ferme	Ministère de la Culture -	48.82827064	2.362358986	NC	11.279714859560194	48.9296428406	2.35946866986
20202008DS	Marché	Services	Services de gardiennage	11/05/2020	0.01	Ferme	VILLE D'OYONNAX	46.26100681	5.652412985	PME	17.498520891477092	46.1836294549	5.51446848032
2,0201E+15	Marché	Fournitures	Étiquettes	09/06/2020	0.01	Révisable	Ministère des armées - Plate-forme commissariat			PME		47.3030353752	-1.56014129347
2,0201E+15	Marché	Fournitures	Équipements d'exposition	09/07/2020	0.01	Révisable	Ministère des armées - Plate-forme commissariat			PME		48.9228776124	2.24687029827
2,0202E+15	Marché	Travaux	Travaux de peinture	14/05/2020	0.01	Révisable	Ministère de la Justice -	48.80227889	2.117165301	PME	39.26613771785153	48.7202231352	2.46254422066
2,0191E+15	Marché	Services	Services de relations publicu	27/03/2019	0.01	Révisable	Ministère de la Transitio	49.44150688	1.093732732	PME	148.50926824664427	48.840040811	2.29268840406
2,0201E+15	Marché	Fournitures	Rideaux, tentures, lambrequ	29/04/2020	0.01	Révisable	Plate-forme commissariat Ouest (Rennes)			PME		48.9547049297	4.41685816506
20202020S1	Marché	Services	Prestations de services pour	03/12/2020	0.01	Révisable	CHR ORLEANS	47.88246650	1.916609618	PME	78.95644623724121	47.6608848758	2.59497464873
2,0201E+15	Marché	Services	Services de conseils environn	17/02/2020	0.01	Révisable	Ministère de la Transitio	49.44150688	1.093732732	PME	848.8373928682636	43.4510080445	5.85936021559
2020yDmaYI	Marché	Services	Services d'assurance domm	06/01/2021	0.01	Ferme	Angers Loire habitat	47.47662847	-0.55608945	NC	122.47183473282674	47.2246985502	-1.63426478543
2,0191E+15	Marché	Services	Services de réparation et d'e	01/11/2019	0.01	Révisable	Services du Premier min	45.02773512	3.896089317	GE	137.34798144967115	45.7708803723	4.88922582274
20192019S1	Marché	Fournitures	Lait	01/06/2019	0.01	Révisable	CHI COMPIEGNE NOYON	49.39926582	2.853836007	GE	449.4574801630858	45.8310057435	4.77089761479

Finance +

2. Présentation de l'architecture

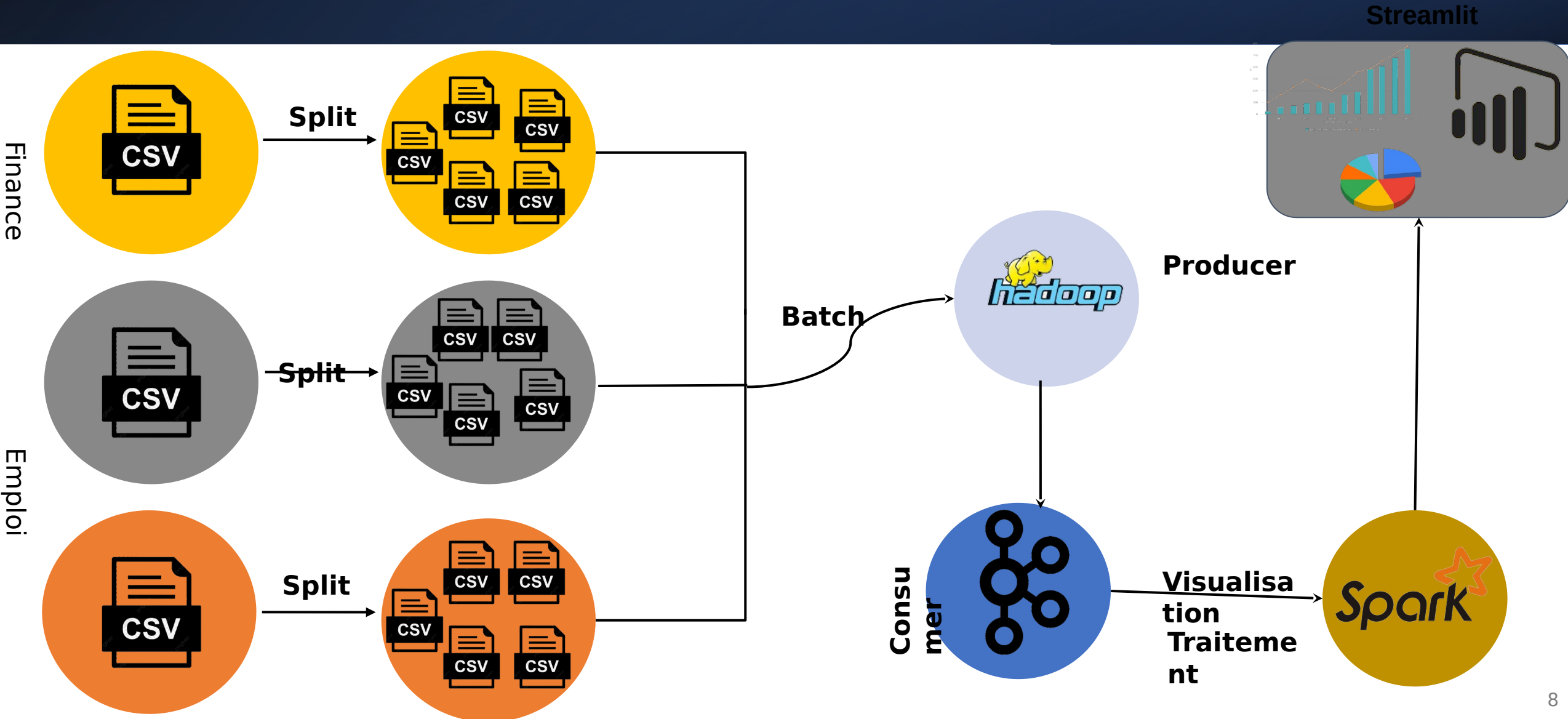


Image Docker

```
version: '2'

services:
  zookeeper:
    image: wurstmeister/zookeeper:3.4.6
    expose:
      - "2181"


  kafka:
    image: wurstmeister/kafka:2.11-2.0.0
    depends_on:
      - zookeeper
    ports:
      - "9092:9092"
    expose:
      - "9093"
    environment:
      KAFKA_ADVERTISED_LISTENERS: INSIDE://kafka:9093,OUTSIDE://localhost:9092
      KAFKA_LISTENER_SECURITY_PROTOCOL_MAP: INSIDE:PLAINTEXT,OUTSIDE:PLAINTEXT
      KAFKA_LISTENERS: INSIDE://0.0.0.0:9093,OUTSIDE://0.0.0.0:9092
      KAFKA_ZOOKEEPER_CONNECT: zookeeper:2181
      KAFKA_INTER_BROKER_LISTENER_NAME: INSIDE
    volumes:
      - ./work:/home/projet-BD/work






  spark-master:
    image: bde2020/spark-master:3.0.1-hadoop3.2
    ports:
      - "8080:8080"
      - "7077:7077"
    links:
      - "kafka:kafka-server"
    volumes:
      - ./work:/home/projet-BD/work
    environment:
      - "SPARK_LOCAL_IP=spark-master"

  spark-worker:
    image: bde2020/spark-worker:3.0.1-hadoop3.2
    depends_on:
      - spark-master
    environment:
      - SPARK_MASTER=spark://spark-master:7077
      - SPARK_WORKER_CORES=2
      - SPARK_WORKER_MEMORY=3G
      - SPARK_DRIVER_MEMORY=2G
      - SPARK_EXECUTOR_MEMORY=2G
    volumes:
      - ./work:/home/projet-BD/work

  py:
    image: python
    stdin_open: true
```



 image4
RUNNING

	image4_py_1	python	RUNNING
	image4_zookeeper_1	wurstmeister/z...	RUNNING
	image4_kafka_1	wurstmeister/k...	RUNNING PORT: 9092
	image4_spark-master_1	bde2020/spark-...	RUNNING PORT: 7077
	image4_spark-worker_1	bde2020/spark-...	RUNNING

Nettoyage des données

ID	B	C	D	E	F	G	H	I	J	K															
	libelle_zone_d_emploi	code_region	region	code_zone_d_emploi	zone_d_emploi	Date	effectifs_salaries_brut	effectifs_salaries_cvs	masse_salariale_brut	masse_salariale_cvs															
60	0102 Est Grande Terre		1	Guadeloupe	102	Est Grande Terre	32021		6480	6325	39597042	39739351													
121	0104 RvDgion Pointoise		1	Guadeloupe	104	RvDgion Pointoise	32021		58534	58737	436631751	453046487													
260	3211 Creil		32	Hauts-de-France	3211	Creil	32021		60338	61135	422070846	437447830													
491	0301 Est-littoral		3	Guyane	301	Est-littoral	32021		27472	27716	199580519	206678636													
555	0302 Ouest-Guyanais		3	Guyane	302	Ouest-Guyanais	32021		2817	2802	15662955	16149512													
700	3203 Arras		32	Hauts-de-France	3203	Arras	32021		49702	50242	320765283	333347738													
1049	1103 Etampes		11	Véle-de-France	1103	Etampes	32021		17340	17367	134781342	137087595													
1212	3215 Lens		32	Hauts-de-France	3215	Lens	32021		91825	92773	587012834	608805693													
1223	1106 Marne-la-VallvDe		11	Véle-de-France	1106	Marne-la-VallvDe	32021		158005	158667	1127542534	1143953733													
1271	1107 Meaux		11	Véle-de-France	1107	Meaux	32021		33766	33987	220869177	226992006													
1343	1108 Melun		11	Véle-de-France	1108	Melun	32021		44070	44378	318954996	326380823													
1521	1111 Rambouillet		11	Véle-de-France	1111	Rambouillet	32021		14643	14728	109241283	111798922													
1653	1113 Saclay		11	Véle-de-France	1113	Saclay	32021		186791	187562	1793145164	1783622059													
1791	1115 Versailles-Saint-Quentin		11	Véle-de-France	1115	Versailles-Saint-Quent	32021		237345	236754	2512323726	2461436590													
1914	4415 Romilly-sur-Seine		44	Grand Est	4415	Romilly-sur-Seine	32021		15117	15243	110750991	118075140													
1973	2403 Chartres		24	Centre-Val de Loire	2403	Chartres	32021		51490	51928	377738717	389882577													
2149	2406 Chinon		24	Centre-Val de Loire	2406	Chinon	32021		10193	10242	75631709	82345661													
2274	4404 Chauxmont		44	Grand Est	4404	Chauxmont	32021		3737	3775	21134687	22039998													
2343	2409 Montargis		24	Centre-Val de Loire	2409	Montargis	32021		25326	25514	168653155	175486426													
2394	4408 Forbach		44	Grand Est	4408	Forbach	32021		18022	18146	107123538	113028766													
2469	2411 Pithiviers		24	Centre-Val de Loire	2411	Pithiviers	32021		13414	13576	92486465	97314185													
2530	4412 Nancy		44	Grand Est	4412	Nancy	32021		140041	141037	967753169	1000754669													
2586	4414 Remiremont		44	Grand Est	4414	Remiremont	32021		20042	20489	125745556	129517803													
2591	2413 Tours		24	Centre-Val de Loire	2413	Tours	32021		139365	139920	983541067	1021609161													
2654	4416 Saint-Avoid		44	Grand Est	4416	Saint-Avoid	32021		27310	27619	187304905	197176705													
2705	2415 Vierzon		24	Centre-Val de Loire	2415	Vierzon	32021		8678	8729	55075170	57618748													
2760	4419 Sarrebourg		44	Grand Est	4419	Sarrebourg	32021		16360	16516	104360597	111779919													
2828	4420 Sarreguemines		44	Grand Est	4420	Sarreguemines	32021		21711	21926	146374234	152873682													
2884	4421 Sedan		44	Grand Est	4421	Sedan	32021		1701	1721	11241265	12074866													
2917	0060 Nevers		24	Centre-Val de Loire	60	Nevers	32021		1585	1591	11005527	11330332													
3135	2702 Auxerre		27	Bourgogne-Franche-Comté	2702	Auxerre	32021		39786	40260	251691087	2611134351													
3226	4409 Haguenau		44	Grand Est	4409	Haguenau	32021		65340	65987	468318728	490556222													
											C D E F G H I J K L M N O P Q														
											code_an	ancien	re_q	libelle_zone	code_zone_c	zone_d_emploi	code_zone_c	zone_d_emploi	annee	trimestre	denier_jour	effectifs_sal	effectifs_salaries	masse_salariale_b	masse_salariale_cvs
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	1998	1	31/03/1998	821	806	2893087	2823823
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	1999	2	30/06/1999	795	811	2955897	2903472
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2000	3	30/09/2000	848	875	3187211	3262555
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2002	1	31/03/2002	957	954	3613175	3570590
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2003	2	30/06/2003	1014	1010	3819669	3747364
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2005	4	31/12/2005	1002	996	4301553	4289172
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2006	1	31/03/2006	1002	997	4176864	4151645
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2008	2	30/06/2008	1268	1261	6028984	5949849
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2009	2	30/06/2009	1339	1332	6187898	6120766
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2010	3	30/09/2010	1314	1348	6324586	6366036
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2011	2	30/06/2011	1493	1487	7365087	7285491
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2012	1	31/03/2012	1400	1402	7005538	7213990
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2012	2	30/06/2012	1272	1266	7044778	6944017
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2012	3	30/09/2012	1245	1281	7203062	7155713
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2013	1	31/03/2013	1373	1371	7437650	7679130
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2013	2	30/06/2013	1402	1391	8114887	7959301
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2013	4	31/12/2013	1491	1474	7999281	7984903
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2015	2	30/06/2015	1483	1461	8568927	8271665
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2015	3	30/09/2015	1428	1472	8494057	8461876
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2017	3	30/09/2017	1401	1460	8706665	8747322
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2017	4	31/12/2017	1471	1444	8678398	8826419
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2018	1	31/03/2018	1472	1445	8722210	8962642
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2019	3	30/09/2019	1469	1544	9675477	9763416
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2020	4	31/12/2020	1587	1541	9388785	9572204
											1	Guadeloupe	0101	CvVte s	101	CvVte sous le	101	CvVte sous le	2021	2	30/06/2021	1525	1548	8512936	8132396
											1	Guadeloupe	0102	Est Gra	102	Est Grande T	102	Est Grande T	1998	1	31/03/1998	3388	3286	12843971	12837759
											1	Guadeloupe	0102	Est Gra	102	Est Grande T	102	Est Grande T	1998	2	30/06/1998	3408	3290	13632561	13594761
											1	Guadeloupe	0102	Est Gra	102	Est Grande T	102	Est Grande T	1998	4	31/12/1998	3219	3202	13782691	13782691
											1	Guadeloupe	0102	Est Gra	102	Est Grande T	102	Est Grande T	2000	2	30/06/2000	3486	3398	14491072	13904952
											1	Guadeloupe	0102	Est Gra	102	Est Grande T	102	Est Grande T	2000	3	30/09/2000	3342	3467	13359794	14020767
											1	Guadeloupe	0102	Est Gra	102	Est Grande T	102	Est Grande T	2001	2	30/06/2001	3927	3861	16644504	16061329
											effectifs +														

Producer

```
class Producer:
    def __init__(self):
        self.server = 'kafka:9093'
        self.producer = KafkaProducer(bootstrap_servers = self.server)
        self.spark = SparkSession.builder.master("local[1]").appName('projetBD').getOrCreate()
        self.topic_sante = "projetBD_sante"
        self.topic_emploi = "projetBD_emploi"
        self.topic_finance = "projetBD_finance"

        self.minLignes = 0
        self.maxLignes = 0
        self.pas = 100
        self.run = True

    def senderProducer(self, path, topic):
        while self.run:
            try:
                # recuperer les donnée d'un dataset et l'envoyer !
                data = self.spark.read.format("csv").option("header", "true").load(path, inferSchema=True)
                self.producer.send(topic, dumps(data).encode('utf-8'))
                self.producer.flush()

                self.minLignes += self.pas
                self.maxLignes += self.pas
                sleep(1)

            except: self.shutdown()

        finally:
            self.producer.close()
            self.shutdown()
```

- Class qui permet de lancer Kafka et de créer nos différents topic puis envoi les données dans les topic grâce à la fonction `senderProducer()` et démarre des thread iot pour récupérer les petits fichiers des trois dossiers (Sante, Travail, Finance) grâce à une fonction `runProducer`

Split

```
class Splitter:
    def __init__(self):
        self.Lignes = 100
        self.fileId = 0
        self.nameFile = "file-"

    def split_csv(self, source_filepath, dest_folder):
        if self.Lignes <= 0: raise Exception('records_per_file must be > 0')

        with open(source_filepath, 'r') as source:
            read = reader(source)

            records_exist = True

            while records_exist:
                i = 0
                target_filepath = path.join(dest_folder, f'{self.nameFile}{self.fileId}.csv')

                with open(target_filepath, 'w') as target:
                    writ = writer(target)
                    while i < self.Lignes:
                        if i == 0: writ.writerow(next(read))
                        try:
                            writ.writerow(next(read))
                            i += 1
                        except:
                            records_exist = False
                            break
                    if i == 0:
```

- Permet de partitionné un fichier csv en plusieurs petit fichier csv(fichier de 100 ligne dans notre cas) en précisant la source et le dossier destinataire des fichiers

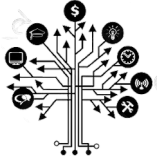
IOT

```
class IOT:
    def __init__(self):
        self.run = True
        self.timer = 1 # le temps en (s)
        self.fileCount = 1

    def codeExec(self, pathFile, fileId, newFolder, linux=True):
        if linux: # for linux
            return "sudo docker cp {}/file-{}.csv sparkmaster:{}".format(pathFile, fileId, newFolder)
        else: # for windows
            return "docker cp {}/file-{}.csv sparkmaster:{}".format(pathFile, fileId, newFolder)

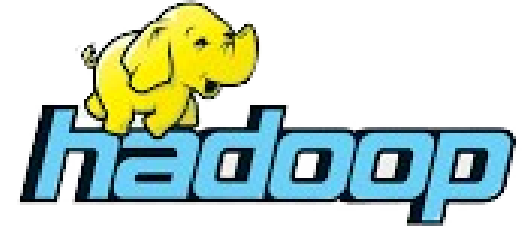
    def iotSender(self, folder, newFolder):
        while self.run:
            if len(listdir(folder))+1 != self.fileCount :
                try:
                    run(self.codeExec(folder, self.fileCount, newFolder, system()), shell=True, check=True)
                    print("Data file {} Sent evry {}s -- 200 OK".format(self.fileCount, self.timer))
                    self.fileCount += 1
                    sleep(self.timer)
                except:
                    self.run = False
                    print("Bad authentification happend !");
            else:
                self.run = False
                print("All files sent !");

if __name__ == "__main__":
    """ Creation des threads """
    sante_iot = Thread(target = IOT().iotSender, args=(dicoPath['SANTE'], dicoToPath['SANTE'],))
    empoie_iot = Thread(target = IOT().iotSender, args=(dicoPath['EMPLOIE'], dicoToPath['EMPLOIE'],))
    finance_iot = Thread(target = IOT().iotSender, args=(dicoPath['FINANCE'], dicoToPath['FINANCE'],))
```



3. Choix des technos

- Hadoop
- Spark
- Kafka
- Streamlit



1. Présentation du Dashboard

```
Data_Rea=pd.DataFrame(Rea,columns=['Reanimation'])
Data_DateSante=pd.DataFrame(Date_Sante,columns=['Date'])
df_Sante = pd.concat([Data_Hospi,Data_Rea,Data_DateSante], axis=1)
HospiReaParDate=df_Sante.groupby(df_Sante.Date)

#Lecture
#st.write(df_Sante)
#st.write(HospiReaParDate)

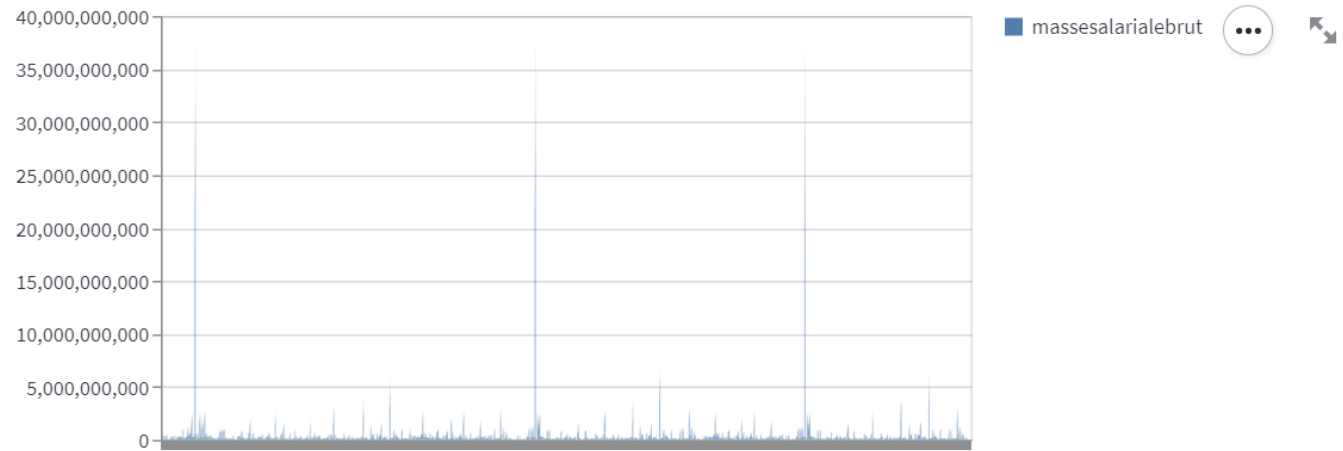
#Travail
effectifssalariesbrut=()
effectifssalariescv=()
massesalarialebrut=()
massesalarialecv=()
DateTravail=()

#List
effectifssalariesbrut=Travail.effectifs_salaries_brut.tolist()
effectifssalariescv=Travail.effectifs_salaries_cv.tolist()
massesalarialebrut=Travail.masse_salariale_brut.tolist()
massesalarialecv=Travail.masse_salariale_cv.tolist()
DateTravail=Travail.Date.tolist()

#DataFrame
Data_effectifssalariesbrut=pd.DataFrame(effectifssalariesbrut,columns=['effectifssalariesbrut'])
Data_effectifssalariescv=pd.DataFrame(effectifssalariescv,columns=['effectifssalariescv'])
Data_massesalarialebrut=pd.DataFrame(massesalarialebrut,columns=['massesalarialebrut'])
Data_massesalarialecv=pd.DataFrame(massesalarialecv,columns=['massesalarialecv'])
Data_DateTravail=pd.DataFrame(DateTravail,columns=['Date'])
df_Travail = pd.concat([Data_effectifssalariesbrut,Data_effectifssalariescv,Data_massesalarialebrut,Data_mas

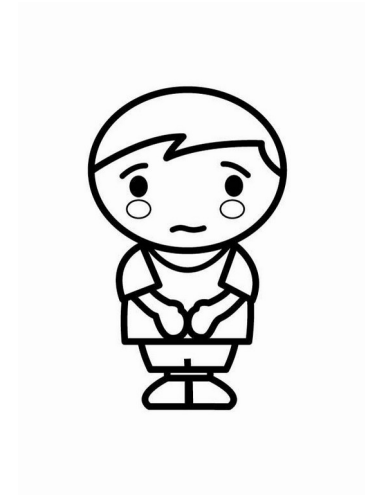
#Lecture
st.write(df_Travail)
st.bar_chart(Data_massesalarialebrut)
```

	effectifssalariesbrut	effectifssalariescv	massesalarialebrut	massesalar
15086	3288486	3281859	37023801921	3631
1405	3259061	3275137	33792303259	3627
25943	3262542	3244219	35879647246	3618
1403	3273608	3251772	35738063603	3595
15087	3203575	3207972	37683906555	3576
15085	3217479	3211268	36144927268	3542
25942	3201715	3206310	37426273114	3539
1401	3226460	3231879	36987650461	3503
25941	3206935	3223225	32554091644	3492
1409	3200673	3190878	35428776951	3477



4. Difficultés rencontrées

- Les installations
- Problème des versions
- Manque d'API gratuite et puissante pour faire du streaming
- Connecteur Kafka-Spark
- Connecteur Spark-MongoDB *
- Pas de Streaming Dans PowerBI

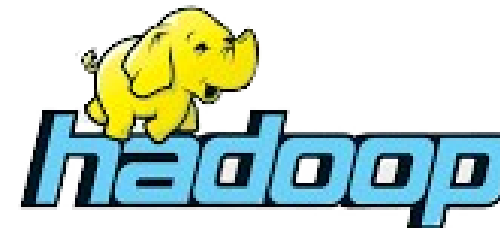
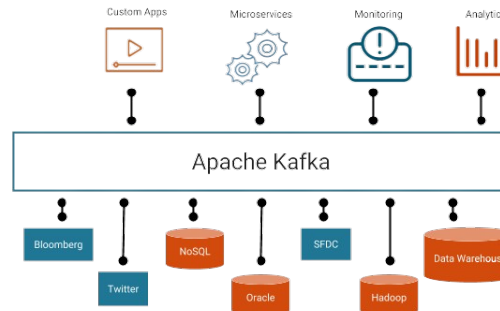


Axe d'améliorations

- Des visualisation et des traitements spark plus pertinents.



Conclusion





**MERCI DE
VOTRE
ATTENTION**

