



Università degli Studi
Mediterranea
di Reggio Calabria

Dipartimento di
Ingegneria dell'Informazione, delle
Infrastrutture e della
Energia
Sostenibile

Corso di Laurea Magistrale in Ingegneria Informatica e dei Sistemi per le
Telecomunicazioni



Tesina del corso di Sistemi Informatici Aziendali
Anno Accademico 2016/2017

Progetto a cura di:
Barbara Messineo
Demetrio Iero
Vincenzo Longo

INDICE

Parte I

“Valutazione del processo di produzione di energia elettrica, delle risorse consumate per produrla e delle emissioni prodotte negli USA”

<i>Introduzione e motivazioni</i>	pag. 4
<i>Obiettivi dell’analisi</i>	pag. 5
<i>Presentazione dei dati utilizzati</i>	pag. 6
<i>Schemi a stella</i>	pag. 11
<i>Analisi mediante Qlik Sense</i>	pag. 13
<i>Analisi mediante Tableau</i>	pag. 21
<i>Analisi mediante MongoDB</i>	pag. 26

Parte II

“Valutazione di tutte le richieste di accesso HTTP al NASA Kennedy Space Center WWW server in Florida del mese di luglio del 1995”

<i>Universo di riferimento</i>	pag. 36
<i>Sorgente dei dati</i>	pag. 37
<i>Ambiente di lavoro Splunk e risultati ottenuti</i>	pag. 38

Parte III

“Conclusioni e valutazione dei software utilizzati”

<i>Qlik Sense</i>	pag. 44
<i>Tableau</i>	pag. 44
<i>MongoDB</i>	pag. 45
<i>Splunk</i>	pag. 45

Valutazione del processo di produzione di energia elettrica, delle risorse consumate per produrla e delle emissioni prodotte negli USA



“Il solo modo di rinnovarsi è di svecchiare, di dare un'energia pulita.”

Joan Mirò

Introduzione e Motivazioni

Ci troviamo in un'era in cui, ormai, l'energia elettrica è diventata un bene di prima necessità. Osservando la vita quotidiana ci si rende conto, infatti, che molte attività sono legate all'uso di una qualche forma di energia: le automobili sfruttano l'energia prodotta dalla combustione dei carburanti, gli elettrodomestici utilizzano l'energia elettrica, le pile usano energia chimica, le barche a vela sfruttano la spinta del vento (energia eolica) e così via.

Naturalmente per sfruttare l'energia elettrica bisogna prima produrla, perché, di fatto, non è una forma di energia liberamente presente in natura.

La produzione di energia elettrica viene effettuata nelle centrali elettriche, che generalmente erogano grandi potenze a regime di produzione costante dell'ordine delle centinaia o migliaia di Megawatt/h.

Com'è già ben noto, la produzione di energia elettrica avviene a partire da due gruppi di risorse primarie:

- Risorse **rinnovabili** (energia solare, eolica, idroelettrica, geotermica, marina).
- Risorse **non rinnovabili** (petrolio, carbone, gas naturale, energia nucleare).

Gli impianti di produzione tradizionali, che utilizzano risorse esauribili, sono altamente inquinanti perché fanno uso di combustibili fossili che, una volta bruciati, producono un aumento sostanziale di ossidi di azoto (NO_x), anidride solforosa (SO₂) e anidride carbonica (CO₂) all'interno dell'atmosfera terrestre.

Mentre in passato si avevano dubbi su questa correlazione, recentemente l'ONU ha confermato che l'aumento della concentrazione di CO₂ ha influenze aggravanti sull'effetto serra.

Quindi, queste emissioni di scarto sono polveri che contribuiscono all'inquinamento e al surriscaldamento globale; per non parlare, poi, delle centrali nucleari che producono enormi quantità di scorie radioattive che al momento è praticamente impossibile smaltire ad impatto zero sull'ambiente.

Perciò, da diversi anni, per la produzione di elettricità il mondo sta cominciando a muoversi verso l'uso di risorse rinnovabili. A differenza dei combustibili fossili e nucleari, le fonti rinnovabili (vento, sole, acqua e calore della Terra) sono economiche, pulite e sicure. La caratteristica di essere pulite si riferisce al fatto che, durante l'intero processo di produzione di energia elettrica, hanno la capacità di non immettere nell'aria sostanze inquinanti e/o climalteranti; per tale motivo sono alla base del "**Green Thinking**" o politica verde.

La volontà di approfondire, in questo progetto, come gli USA (la superpotenza occidentale) stanno evolvendo, negli anni, il processo di produzione di elettricità (dalla scelta del tipo di risorsa primaria utilizzata, fino ad arrivare alle emissioni prodotte annualmente su tutto il territorio), ha spinto il nostro gruppo a condurre un'analisi di tipo Big Data Analytics per capire se è possibile garantire, in futuro, il fabbisogno di energia elettrica dipendendo sempre meno dalle fonti esauribili, e come gli USA stessi si stiano ponendo nei confronti del "*Green Thinking*".

Un approccio rilevante di questa superpotenza nei confronti delle risorse rinnovabili potrebbe influenzare positivamente anche le scelte degli altri Paesi a favore dell'ecosostenibilità.

Non a caso, la prima centrale termoelettrica fu impiantata proprio negli USA, a New York nel 1882, per rifornire la prima rete di illuminazione pubblica.

Obiettivi dell'analisi

- Capire e confrontare le quantità di energia elettrica prodotta dai vari Stati sia da fonti rinnovabili che da fonti non rinnovabili, negli anni dal 1990 al 2015.
- Capire la quantità di risorse rinnovabili impiegate dai vari Stati per la produzione di energia elettrica, suddivise per tipo di risorsa, negli anni dal 1990 al 2015.
- Analizzare quali Stati stanno effettivamente nel tempo diminuendo l'utilizzo di fonti non rinnovabili in favore di quelle rinnovabili.
- Analizzare le emissioni prodotte dalle centrali elettriche dei vari Stati, individuando quali tra loro risultano essere più attenti nei riguardi dell'ambiente.



Presentazione dei dati utilizzati

Per ottenere le informazioni di interesse sono stati consultati diversi portali di Open Data statunitensi; l'obiettivo della ricerca si è focalizzato su dataset contenenti informazioni attuali e storiche, rimanendo in un intervallo temporale dell'ordine di un ventennio. Tra le varie opzioni, la scelta del nostro gruppo di lavoro è ricaduta sui dati pubblicati sul sito Web governativo americano della U.S. EIA (Energy Information Administration), un'agenzia principale del U.S. Federal Statistical System responsabile della raccolta, dell'analisi e della diffusione delle informazioni riguardanti l'energia e l'interazione tra quest'ultima con l'economia e l'ambiente. La sezione relativa all'energia elettrica è riportata nella Figura 1.

Annual data

Year Range	Description	format
1990 - 2015	Net Generation by State by Type of Producer by Energy Source (EIA-906, EIA-920, and EIA-923) ¹	XLS
1990 - 2015	Fossil Fuel Consumption for Electricity Generation by Year, Industry Type and State (EIA-906, EIA-920, and EIA-923) ²	XLS
1990 - 2014	Existing Nameplate and Net Summer Capacity by Energy Source, Producer Type and State (EIA-860) ^{1, 3}	XLS
2015 - 2019	Proposed Nameplate and Net Summer Capacity by Year, Energy Source, and State (EIA-860) ¹	XLS
1990 - 2014	U.S. Electric Power Industry Estimated Emissions by State (EIA-767, EIA-906, EIA-920, and EIA-923) ⁴	XLS
1990 - 2015	Average Price by State by Provider (EIA-861) ⁵	XLS
1990 - 2015	Number of Retail Customers by State by Sector (EIA-861) ⁵	XLS

Monthly data from Electric Power Monthly

Year Range	Description	format
2001 - Present	Net Generation by State by Type of Producer by Energy Source ¹	XLS
2001 - Present	Fossil Fuel Consumption for Electricity Generation by Year, Industry Type and State ²	XLS

See also:
[Electric Power Monthly](#)

Figura 1. Portale U.S. EIA <https://www.eia.gov/electricity/data/state/>

Tra le tabelle a disposizione, tutte in formato .xls, sono state scelte le seguenti:

- La tabella relativa ai Megawatt/h generati dal 1990 al 2015.
- La tabella relativa al consumo di combustibili fossili per la produzione di elettricità dal 1990 al 2015.
- La tabella relativa alle stime sulle emissioni prodotte dalle industrie di energia elettrica dal 1990 al 2014.

Inoltre, è stato necessario integrare le tabelle sopracitate con un'ulteriore tabella, contenente informazioni aggiuntive riguardo la suddivisione geografica dei 50 Stati.

La tabella relativa agli Stati, come mostrato in Figura 2, contiene i seguenti campi:

- **NAME:** rappresenta il nome di uno Stato senza abbreviazioni.
- **STATE:** rappresenta la sigla dello Stato.
- **CENSUS REGION NAME:** indica l'appartenenza di uno Stato ad una delle 4 Regioni di natura geografica e storico-culturale proposte dallo United States Census Bureau (West, South, Midwest e Northeast).
- **CENSUS DIVISION NAME:** indica l'appartenenza di uno Stato ad una delle 8 Divisioni proposte dallo United States Census Bureau; esse rappresentano una sottocategoria delle Regioni (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain e Pacific).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
NAME	STATE	CENSUS REGION NAME	CENSUS DIVISION NAME											
Alaska	AK	West	Pacific											
Alabama	AL	South	East South Central											
Arizona	AZ	West	Mountain											
Arkansas	AR	South	West South Central											
California	CA	West	Pacific											
Colorado	CO	West	Mountain											
Connecticut	CT	Northeast	New England											
Delaware	DE	South	South Atlantic											
Florida	FL	South	South Atlantic											
Georgia	GA	South	South Atlantic											
Hawaii	HI	West	Pacific											
Idaho	ID	West	Mountain											
Illinois	IL	Midwest	East North Central											
Indiana	IN	Midwest	East North Central											
Iowa	IA	Midwest	West North Central											
Kansas	KS	Midwest	West North Central											
Kentucky	KY	South	East South Central											
Louisiana	LA	South	West South Central											
Maine	ME	Northeast	New England											
Maryland	MD	South	South Atlantic											
Massachusetts	MA	Northeast	New England											
Michigan	MI	Midwest	East North Central											
Minnesota	MN	Midwest	West North Central											
Mississippi	MS	South	East South Central											
Missouri	MO	Midwest	West North Central											
Montana	MT	West	Mountain											
Nebraska	NE	Midwest	West North Central											
Nevada	NV	West	Mountain											
New Hampshire	NH	Northeast	New England											
New Jersey	NJ	Northeast	Mid-Atlantic											

Figura 2. Tabella degli Stati.

La tabella relativa alla generazione di energia elettrica, mostrata in Figura 3, contiene i seguenti campi:

- **YEAR:** indica l'anno di riferimento.
- **STATE:** indica la sigla dello Stato.
- **TYPE OF PRODUCER:** indica la tipologia di generatore industriale che ha prodotto un determinato quantitativo di energia elettrica a partire da una determinata risorsa, come ad esempio un generatore di un impianto idroelettrico.
- **ENERGY SOURCE:** indica la risorsa utilizzata per produrre un certo quantitativo di energia elettrica.
- **GENERATION (Megawatt-hours):** indica il quantitativo annuo di Megawatt/h prodotti da un determinato generatore con una determinata fonte di energia.
- **FIVE_YEARS:** indica un periodo di riferimento temporale di 5 anni per il processo di produzione energetica (ottenuto mediante un'operazione di ETL preliminare sui dati).
- **TEN_YEARS:** indica un periodo di riferimento temporale di 10 anni per il processo di produzione energetica (ottenuto mediante un'operazione di ETL preliminare sui dati).
- **TWENTY_YEARS:** indica un periodo di riferimento temporale di 20 anni per il processo di produzione energetica (ottenuto mediante un'operazione di ETL preliminare sui dati).

YEAR	STATE	TYPE OF PRODUCER	ENERGY SOURCE	GENERATION (Megawatthours)	FIVE_YEARS	TEN_YEARS	TWENTY_YEARS
1990	AK	Total Electric Power Industry	Total	5,599,506	1990-1994	1990-1999	1990-2009
1990	AK	Total Electric Power Industry	Coal	510,573	1990-1994	1990-1999	1990-2009
1990	AK	Total Electric Power Industry	Hydroelectric Conventional	974,521	1990-1994	1990-1999	1990-2009
1990	AK	Total Electric Power Industry	Natural Gas	3,468,261	1990-1994	1990-1999	1990-2009
1990	AK	Total Electric Power Industry	Petroleum	497,116	1990-1994	1990-1999	1990-2009
1990	AK	Total Electric Power Industry	Wind	0	1990-1994	1990-1999	1990-2009
1990	AK	Total Electric Power Industry	Wood and Wood Derived Fuels	151,035	1990-1994	1990-1999	1990-2009
1990	AK	Electric Generators, Electric Utilities	Total	4,493,024	1990-1994	1990-1999	1990-2009
1990	AK	Electric Generators, Electric Utilities	Coal	311,960	1990-1994	1990-1999	1990-2009
1990	AK	Electric Generators, Electric Utilities	Hydroelectric Conventional	974,521	1990-1994	1990-1999	1990-2009
1990	AK	Electric Generators, Electric Utilities	Natural Gas	2,869,638	1990-1994	1990-1999	1990-2009
1990	AK	Electric Generators, Electric Utilities	Petroleum	336,905	1990-1994	1990-1999	1990-2009
1990	AK	Electric Generators, Electric Utilities	Wind	0	1990-1994	1990-1999	1990-2009
1990	AK	Combined Heat and Power, Industrial Power	Total	840,949	1990-1994	1990-1999	1990-2009
1990	AK	Combined Heat and Power, Industrial Power	Natural Gas	596,623	1990-1994	1990-1999	1990-2009
1990	AK	Combined Heat and Power, Industrial Power	Petroleum	93,291	1990-1994	1990-1999	1990-2009
1990	AK	Combined Heat and Power, Industrial Power	Wood and Wood Derived Fuels	151,035	1990-1994	1990-1999	1990-2009
1990	AK	Combined Heat and Power, Commercial Power	Total	265,533	1990-1994	1990-1999	1990-2009
1990	AK	Combined Heat and Power, Commercial Power	Coal	198,613	1990-1994	1990-1999	1990-2009
1990	AK	Combined Heat and Power, Commercial Power	Petroleum	66,920	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Total	79,652,133	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Coal	53,658,115	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Hydroelectric Conventional	10,366,507	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Natural Gas	1,020,714	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Nuclear	12,051,882	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Other Biomass	47,503	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Other Gases	269,476	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Petroleum	138,089	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Wood and Wood Derived Fuels	2,099,847	1990-1994	1990-1999	1990-2009
1990	AL	Electric Generators, Independent Power Producers	Total	27,936	1990-1994	1990-1999	1990-2009
1990	AL	Electric Generators, Independent Power Producers	Other Biomass	25,310	1990-1994	1990-1999	1990-2009
1990	AL	Electric Generators, Independent Power Producers	Wood and Wood Derived Fuels	2,626	1990-1994	1990-1999	1990-2009
1990	AL	Electric Generators, Electric Utilities	Total	76,231,696	1990-1994	1990-1999	1990-2009

Figura 3. Tabella relativa alla produzione di energia elettrica.

La tabella relativa al consumo di combustibili fossili, mostrata in Figura 4, contiene i seguenti campi:

- **YEAR:** indica l'anno di riferimento.
- **STATE:** indica la sigla dello Stato.
- **TYPE OF PRODUCER:** indica la tipologia di generatore industriale che ha prodotto un determinato quantitativo di energia elettrica a partire da una determinata risorsa.
- **ENERGY SOURCE (UNITS):** indica la risorsa utilizzata per produrre energia elettrica, con una precisazione sull'unità di misura utilizzata nelle rispettive misure.
- **CONSUMPTION for ELECTRICITY:** indica il quantitativo annuo di una risorsa consumata per la produzione di energia elettrica.
- **FIVE_YEARS:** indica un periodo di riferimento temporale di 5 anni per il consumo di combustibili fossili da parte delle industrie di energia elettrica (ottenuto mediante un'operazione di ETL preliminare sui dati).
- **TEN_YEARS:** indica un periodo di riferimento temporale di 10 anni per il consumo di combustibili fossili da parte delle industrie di energia elettrica (ottenuto mediante un'operazione di ETL preliminare sui dati).
- **TWENTY_YEARS:** indica un periodo di riferimento temporale di 20 anni per il consumo di combustibili fossili da parte delle industrie di energia elettrica (ottenuto mediante un'operazione di ETL preliminare sui dati).

YEAR	STATE	TYPE OF PRODUCER	ENERGY SOURCE (UNITS)	CONSUMPTION for ELECTRICITY	FIVE_YEARS	TEN_YEARS	TWENTY_YEARS
1990	AK	Total Electric Power Industry	Coal (Short Tons)	404 871	1990-1994	1990-1999	1990-2009
1990	AK	Total Electric Power Industry	Petroleum (Barrels)	961 837	1990-1994	1990-1999	1990-2009
1990	AK	Total Electric Power Industry	Natural Gas (Mcf)	42 764 948	1990-1994	1990-1999	1990-2009
1990	AK	Electric Generators, Electric Utilities	Coal (Short Tons)	290 182	1990-1994	1990-1999	1990-2009
1990	AK	Electric Generators, Electric Utilities	Petroleum (Barrels)	657 706	1990-1994	1990-1999	1990-2009
1990	AK	Electric Generators, Electric Utilities	Natural Gas (Mcf)	34 366 142	1990-1994	1990-1999	1990-2009
1990	AK	Combined Heat and Power, Commercial Power	Coal (Short Tons)	114 689	1990-1994	1990-1999	1990-2009
1990	AK	Combined Heat and Power, Commercial Power	Petroleum (Barrels)	112 409	1990-1994	1990-1999	1990-2009
1990	AK	Combined Heat and Power, Industrial Power	Petroleum (Barrels)	191 722	1990-1994	1990-1999	1990-2009
1990	AK	Combined Heat and Power, Industrial Power	Natural Gas (Mcf)	8 398 806	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Coal (Short Tons)	22 242 353	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Petroleum (Barrels)	249 875	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Natural Gas (Mcf)	10 401 790	1990-1994	1990-1999	1990-2009
1990	AL	Total Electric Power Industry	Other Gases (Billion BTU)	3 473	1990-1994	1990-1999	1990-2009
1990	AL	Electric Generators, Electric Utilities	Coal (Short Tons)	22 010 160	1990-1994	1990-1999	1990-2009
1990	AL	Electric Generators, Electric Utilities	Petroleum (Barrels)	132 744	1990-1994	1990-1999	1990-2009
1990	AL	Electric Generators, Electric Utilities	Natural Gas (Mcf)	4 090 291	1990-1994	1990-1999	1990-2009
1990	AL	Combined Heat and Power, Electric Power	Coal (Short Tons)	16 681	1990-1994	1990-1999	1990-2009
1990	AL	Combined Heat and Power, Electric Power	Natural Gas (Mcf)	320 132	1990-1994	1990-1999	1990-2009
1990	AL	Combined Heat and Power, Industrial Power	Coal (Short Tons)	215 512	1990-1994	1990-1999	1990-2009
1990	AL	Combined Heat and Power, Industrial Power	Petroleum (Barrels)	117 131	1990-1994	1990-1999	1990-2009
1990	AL	Combined Heat and Power, Industrial Power	Natural Gas (Mcf)	5 991 367	1990-1994	1990-1999	1990-2009
1990	AL	Combined Heat and Power, Industrial Power	Other Gases (Billion BTU)	3 473	1990-1994	1990-1999	1990-2009
1990	AR	Total Electric Power Industry	Coal (Short Tons)	11 884 395	1990-1994	1990-1999	1990-2009
1990	AR	Total Electric Power Industry	Petroleum (Barrels)	480 098	1990-1994	1990-1999	1990-2009

Coal includes anthracite, bituminous coal, subbituminous coal, lignite, waste coal, and synthetic coal. As of 2011, coal also includes synthesis gas derived from coal.

Other Gases includes blast furnace gas, propane gas, and other manufactured and waste gases derived from fossil fuels. As of 2011, Other Gases does not include propane.

Petroleum includes distillate fuel oil (all diesel and No. 1, No. 2, and No. 4 fuel oils), residual fuel oil (No. 5 and No. 6 fuel oils and bunker C fuel oil), jet fuel, kerosene, petroleum coals, waste oil, and beginning in 2011, propane.

Figura 4. Tabella relativa al consumo di combustibili fossili per la produzione di energia elettrica.

La tabella relativa alle stime sulla produzione di emissioni da parte delle industrie di energia elettrica, mostrata in Figura 5, contiene i seguenti campi:

- **Year:** indica l'anno di riferimento.
- **State:** indica la sigla dello Stato.
- **Producer Type:** indica la tipologia di generatore industriale che ha prodotto un determinato quantitativo di energia elettrica a partire da una determinata risorsa.
- **Energy Source:** indica la risorsa utilizzata da un certo tipo di industria di produzione di energia elettrica.
- **CO₂ (Metric Tons):** indica il quantitativo annuo delle tonnellate di CO₂ derivanti da un certo tipo di industria di produzione di energia elettrica.
- **SO₂ (Metric Tons):** indica il quantitativo annuo delle tonnellate di SO₂ derivanti da un certo tipo di industria di produzione di energia elettrica.
- **NO_x (Metric Tons):** indica il quantitativo annuo delle tonnellate di NO_x derivanti da un certo tipo di industria di produzione di energia elettrica.
- **Five_Years:** indica un periodo di riferimento temporale di 5 anni per le emissioni relative al processo di produzione energetica (ottenuto mediante un'operazione di ETL preliminare sui dati).
- **Ten_Years:** indica un periodo di riferimento temporale di 10 anni per le emissioni relative al processo di produzione energetica (ottenuto mediante un'operazione di ETL preliminare sui dati).
- **Twenty_Years:** indica un periodo di riferimento temporale di 20 anni per le emissioni relative al processo di produzione energetica (ottenuto mediante un'operazione di ETL preliminare sui dati).

emission_annual_modified.xls [Visualizzazione protetta] - Excel

File Home Inserisci Layout di pagina Formule Dati Revisione Visualizza Che cosa si vuole fare?

VISUALIZZAZIONE PROTETTA Attenzione: I file provenienti da Internet possono contenere virus. A meno che non sia necessario modificare il file, è consigliabile restare in Visualizzazione protetta. Abilita modifica

Year	State	Producer Type	Energy Source	CO2 (Metric Tons)	SO2 (Metric Tons)	NOx (Metric Tons)	Five_Years	Ten_Years	Twenty_Years	
27797	2009	WY	Total Electric Power Industry	Natural Gas	460 031	2	1 129	2005-2009	2000-2009	1990-2009
27798	2009	WY	Total Electric Power Industry	Other Gases	0	4	1 880	2005-2009	2000-2009	1990-2009
27799	2009	WY	Total Electric Power Industry	Other	0	0	46	2005-2009	2000-2009	1990-2009
27800	2009	WY	Total Electric Power Industry	Petroleum	42 155	25	39	2005-2009	2000-2009	1990-2009
27801	2010	AK	Electric Utility	All Sources	2 985 504	787	12 823	2010-2014	2010-2014	2010-2014
27802	2010	AK	Electric Utility	Coal	243 080	23	378	2010-2014	2010-2014	2010-2014
27803	2010	AK	Electric Utility	Natural Gas	2 122 340	10	5 925	2010-2014	2010-2014	2010-2014
27804	2010	AK	Electric Utility	Petroleum	620 084	754	6 520	2010-2014	2010-2014	2010-2014
27805	2010	AK	IPP NAICS-22 Cogen	All Sources	325 615	549	742	2010-2014	2010-2014	2010-2014
27806	2010	AK	IPP NAICS-22 Cogen	Coal	325 615	549	742	2010-2014	2010-2014	2010-2014
27807	2010	AK	Commercial Non-Cogen	All Sources	529	0	11	2010-2014	2010-2014	2010-2014
27808	2010	AK	Commercial Non-Cogen	Petroleum	529	0	11	2010-2014	2010-2014	2010-2014
27809	2010	AK	Commercial Cogen	All Sources	733 998	1 622	1 648	2010-2014	2010-2014	2010-2014
27810	2010	AK	Commercial Cogen	Coal	723 229	1 615	1 615	2010-2014	2010-2014	2010-2014
27811	2010	AK	Commercial Cogen	Natural Gas	789	0	2	2010-2014	2010-2014	2010-2014
27812	2010	AK	Commercial Cogen	Petroleum	9 980	7	31	2010-2014	2010-2014	2010-2014
27813	2010	AK	Industrial Non-Cogen	All Sources	20 582	49	418	2010-2014	2010-2014	2010-2014
27814	2010	AK	Industrial Non-Cogen	Other Biomass	0	35	4	2010-2014	2010-2014	2010-2014
27815	2010	AK	Industrial Non-Cogen	Petroleum	20 582	14	414	2010-2014	2010-2014	2010-2014
27816	2010	AK	Industrial Cogen	All Sources	58 824	8	386	2010-2014	2010-2014	2010-2014
27817	2010	AK	Industrial Cogen	Natural Gas	46 319	0	140	2010-2014	2010-2014	2010-2014
27818	2010	AK	Industrial Cogen	Petroleum	12 505	8	246	2010-2014	2010-2014	2010-2014
27819	2010	AK	Total Electric Power Industry	All Sources	4 125 052	3 015	16 028	2010-2014	2010-2014	2010-2014
27820	2010	AK	Total Electric Power Industry	Coal	1 291 924	2 187	2 735	2010-2014	2010-2014	2010-2014
27821	2010	AK	Total Electric Power Industry	Natural Gas	2 169 448	10	6 067	2010-2014	2010-2014	2010-2014
27822	2010	AK	Total Electric Power Industry	Other Biomass	0	35	4	2010-2014	2010-2014	2010-2014
27823	2010	AK	Total Electric Power Industry	Petroleum	663 680	783	7 222	2010-2014	2010-2014	2010-2014
27824	2010	AL	Electric Utility	All Sources	66 802 261	185 193	66 526	2010-2014	2010-2014	2010-2014
27825	2010	AL	Electric Utility	Coal	61 222 905	185 128	51 011	2010-2014	2010-2014	2010-2014
27826	2010	AL	Electric Utility	Natural Gas	5 502 247	21	5 434	2010-2014	2010-2014	2010-2014
27827	2010	AL	Electric Utility	Other Biomass	0	0	1	2010-2014	2010-2014	2010-2014
27828	2010	AL	Electric Utility	Petroleum	77 409	44	80	2010-2014	2010-2014	2010-2014
27829	2010	AL	IPP NAICS-22 Non-Cogen	All Sources	8 161 593	43	817	2010-2014	2010-2014	2010-2014
27830	2010	AL	IPP NAICS-22 Non-Cogen	Natural Gas	8 147 527	41	794	2010-2014	2010-2014	2010-2014
27831	2010	AL	IPP NAICS-22 Non-Cogen	Petroleum	14 066	2	23	2010-2014	2010-2014	2010-2014
27832	2010	AL	IPP NAICS-22 Cogen	All Sources	1 824 498	978	1 206	2010-2014	2010-2014	2010-2014
27833	2010	AL	IPP NAICS-22 Cogen	Coal	210 874	900	454	2010-2014	2010-2014	2010-2014

State Emissions NOTES

Pronto

Figura 5. Tabella relativa alle emissioni prodotte dalle industrie di energia elettrica.

Schemi a stella

Dopo aver trovato, validato e ripulito i dati, il nostro gruppo si è dedicato all'attività di formalizzazione di un modello multidimensionale che rappresentasse il Case Study preso in esame.

In tale modello sono stati individuati 3 fatti:

- *PRODUZIONE*, relativo all'energia elettrica prodotta nei vari anni con le diverse fonti di energia.
- *CONSUMO_RISORSE*, relativo all'utilizzo dei combustibili fossili per la produzione di elettricità.
- *EMISSIONI_PRODOTTE*, relativo alle emissioni prodotte dalle centrali elettriche.

Per il fatto *PRODUZIONE* è stato realizzato lo schema a stella mostrato in Figura 6; in particolare, esso è composto da 4 dimensioni (Tempo, Stato, Risorsa e Produttore) ed una sola misura (Megawatt/h). Per questo fatto sono state identificate le seguenti gerarchie:

- Quinquennio → Decennio → Ventennio (per la dimensione Tempo);
- Stato → Divisione → Regione (per la dimensione Stato);

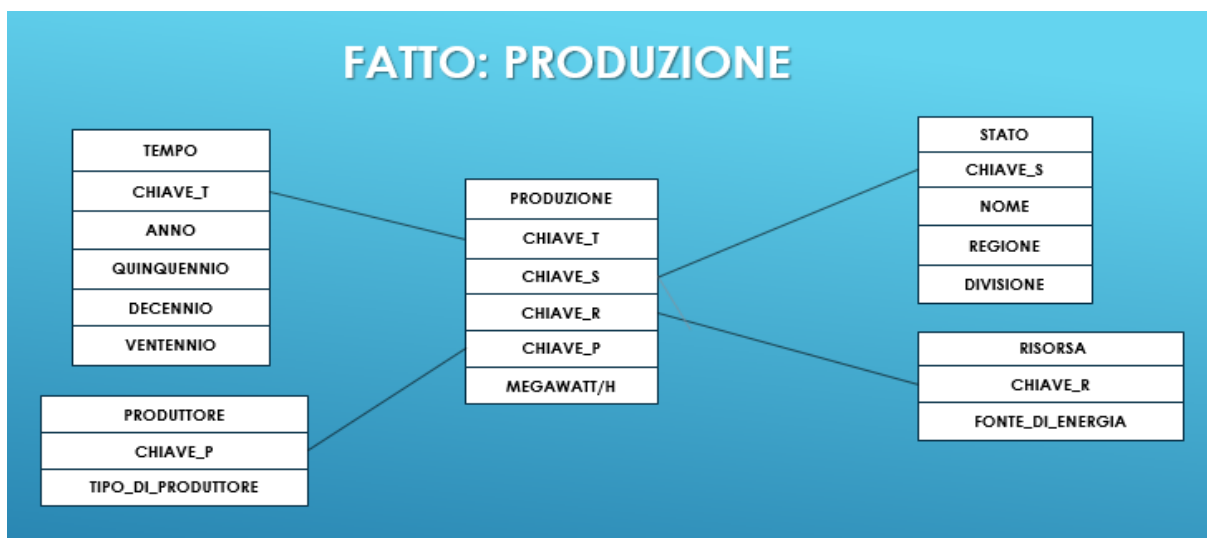


Figura 6. Schema a stella relativo al fatto *PRODUZIONE*.

Per il fatto *CONSUMO_RISORSE* è stato realizzato lo schema a stella mostrato in Figura 7; in particolare, esso è composto da 4 dimensioni (Tempo, Stato, Risorsa e Produttore) ed una sola misura (Unità consumate). Per questo fatto sono state identificate le seguenti gerarchie:

- Quinquennio → Decennio → Ventennio (per la dimensione Tempo);
- Stato → Divisione → Regione (per la dimensione Stato);

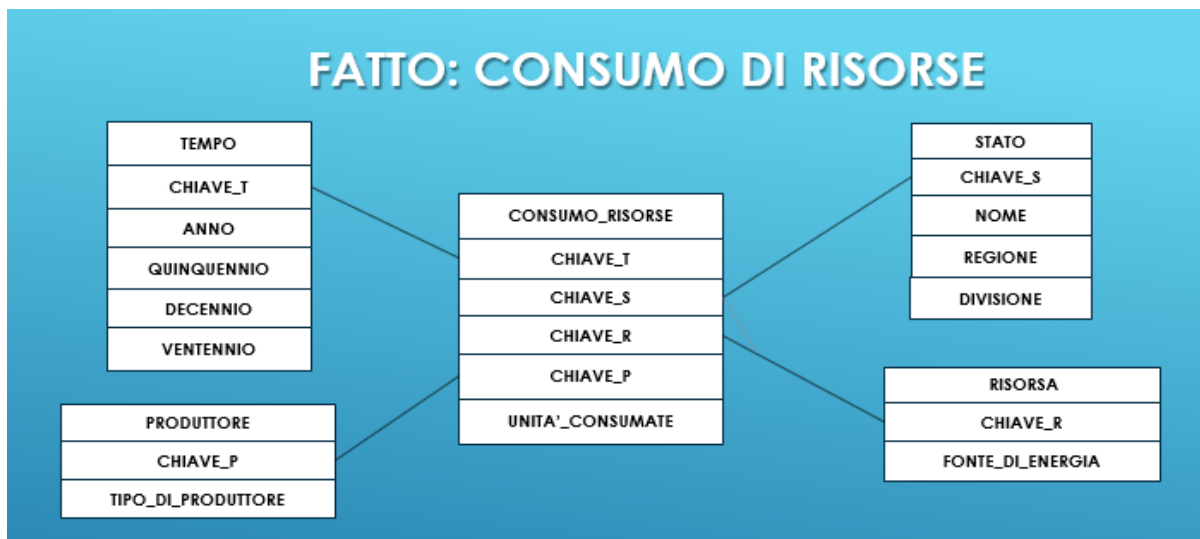


Figura 7. Schema a stella relativo al fatto CONSUMO_RISORSE.

Per il fatto EMISSIONI_PRODOTTE è stato realizzato lo schema a stella mostrato in Figura 8; in particolare, esso è composto da 4 dimensioni (Tempo, Stato, Risorsa e Produttore) e da 3 misure (CO2 (Metric Tons), SO2 (Metric Tons) e NOx (Metric Tons)). Per questo fatto sono state identificate le seguenti gerarchie:

- Quinquennio → Decennio → Ventennio (per la dimensione Tempo);
- Stato → Divisione → Regione (per la dimensione Stato);

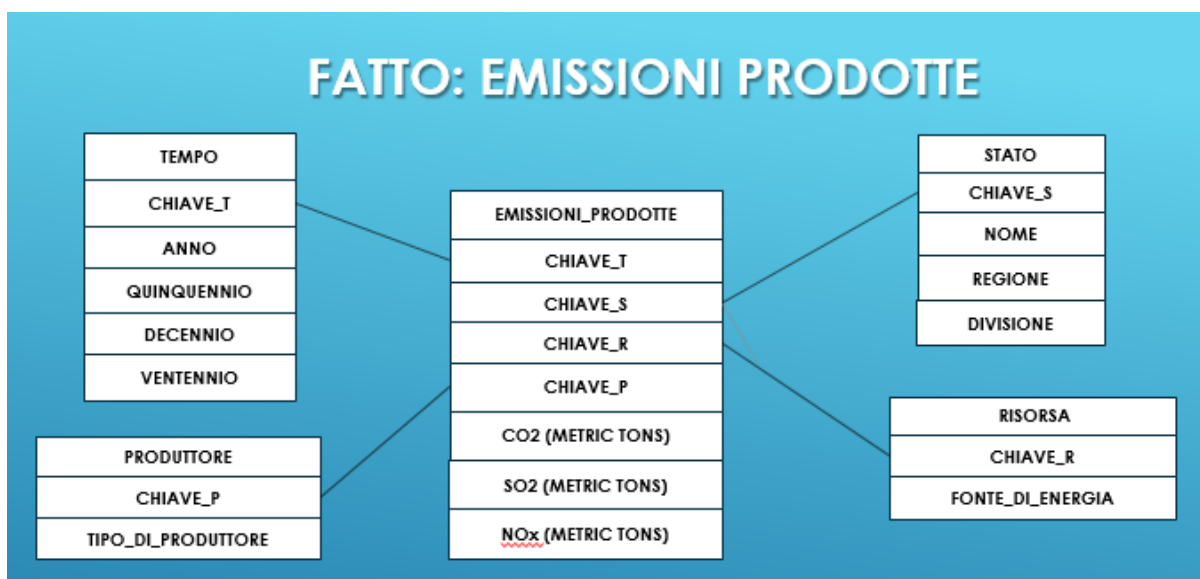


Figura 8. Schema a stella relativo al fatto EMISSIONI_PRODOTTE.

Analisi mediante Qlik Sense

Dopo aver installato il tool Qlik Sense e creato un'app relativa al nostro Case Study all'interno di quest'ultimo, abbiamo importato le tabelle esaminate nelle pagine precedenti mediante la procedura guidata proposta da Qlik Sense stesso. Una volta inserite e collegate opportunamente le tabelle tra loro, secondo gli schemi a stella prodotti in precedenza, ciascun membro del gruppo si è fatto carico di estrapolare delle dashboard che riuscissero a fornire dei dati in merito agli obiettivi proposti durante la presentazione del Case Study.

Considerando il fatto che abbiamo a che fare con dati storici, si è scelto di usare dei grafici interattivi, sfruttando l'utilizzo di operazioni di drill-down sia per la dimensione spaziale che temporale, in modo da avere, dapprima, una visione d'insieme per le varie tematiche e, successivamente, mediante dei click all'interno dei vari grafici, lasciare libero l'utilizzatore di zoomare su determinate regioni di interesse.

Nelle Figure da 9 a 15 sono riportati i grafici relativi all'analisi sul consumo di combustibili fossili; in particolare, sono stati affiancati in un'unica dashboard i dati su base temporale (a sinistra) e quelli su base geografica (a destra); in questo modo è possibile percorrere parallelamente le due gerarchie, dato che un'operazione di drill-down su un grafico si ripercuoterà anche sull'altro, in modo da visualizzare risultati più mirati.

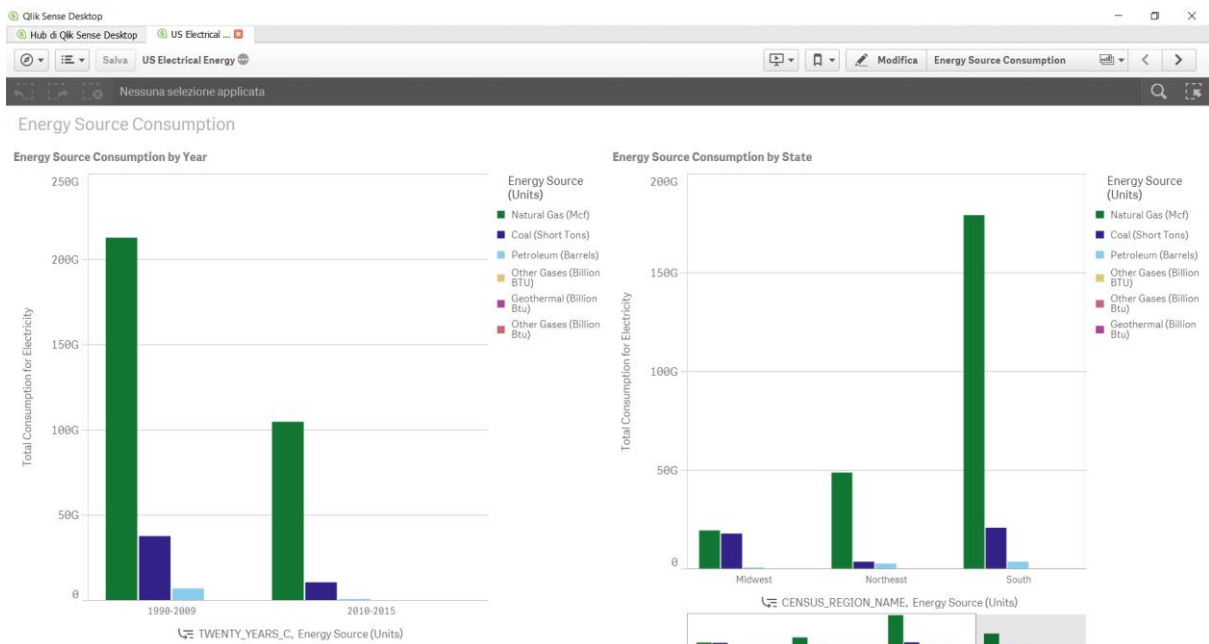


Figura 9. Grafici relativi al consumo di combustibili fossili per Regione in 20 anni.

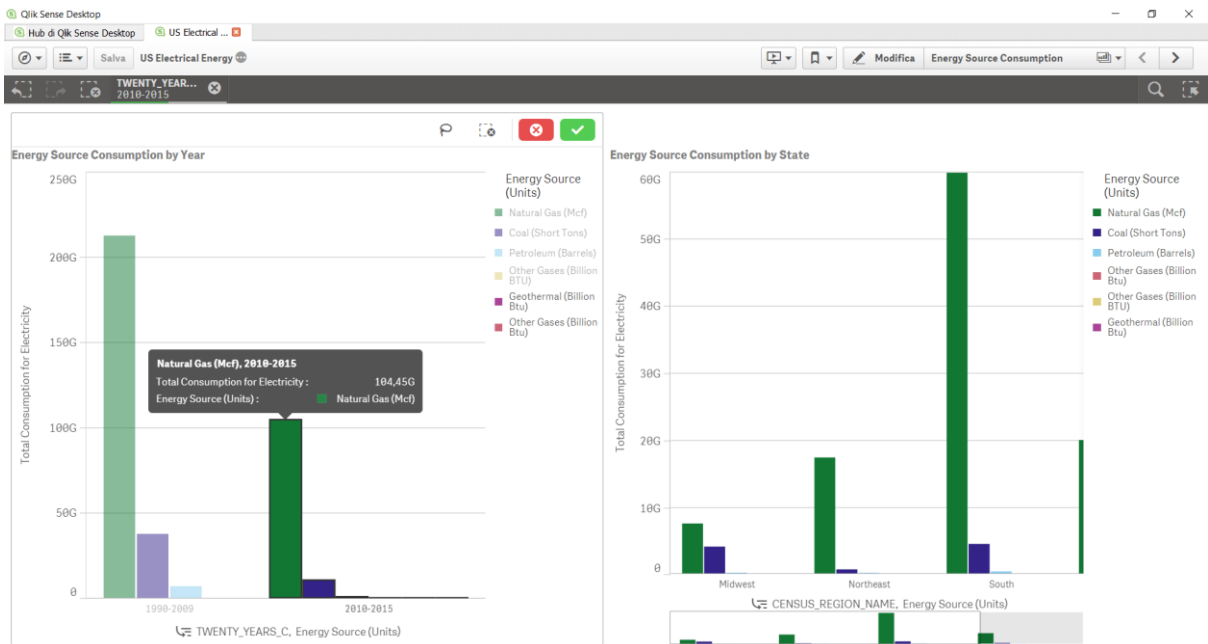


Figura 10. Selezione risultati relativi agli anni dal 2010 al 2015.

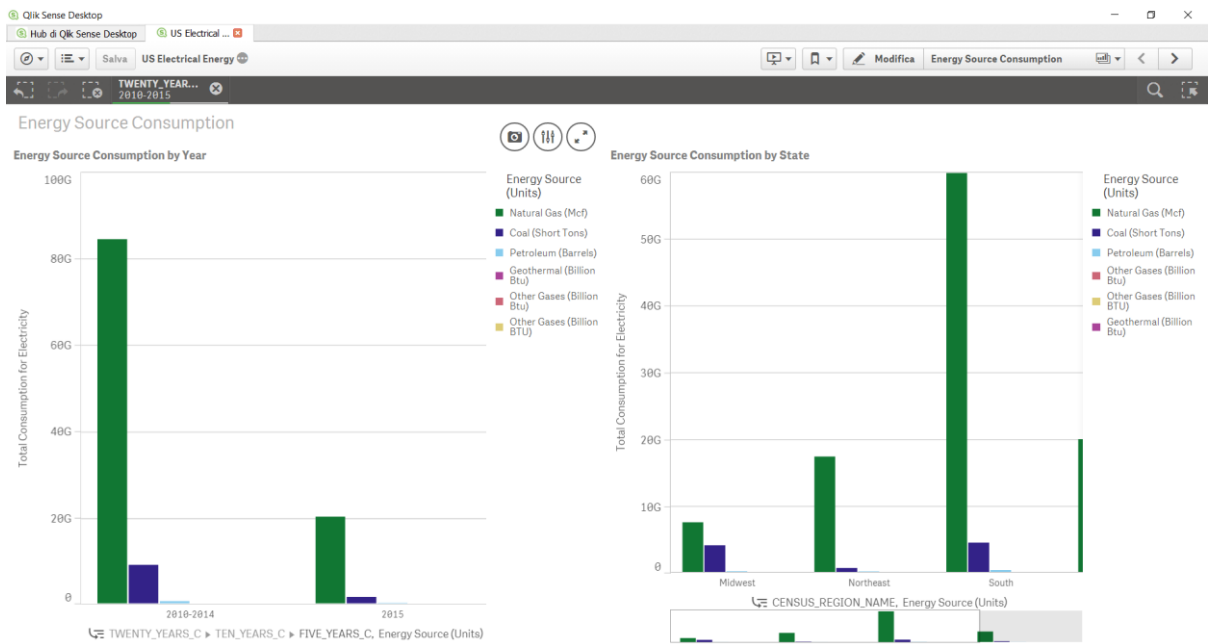


Figura 11. Visualizzazione grafici dopo il drill-down per decennio.

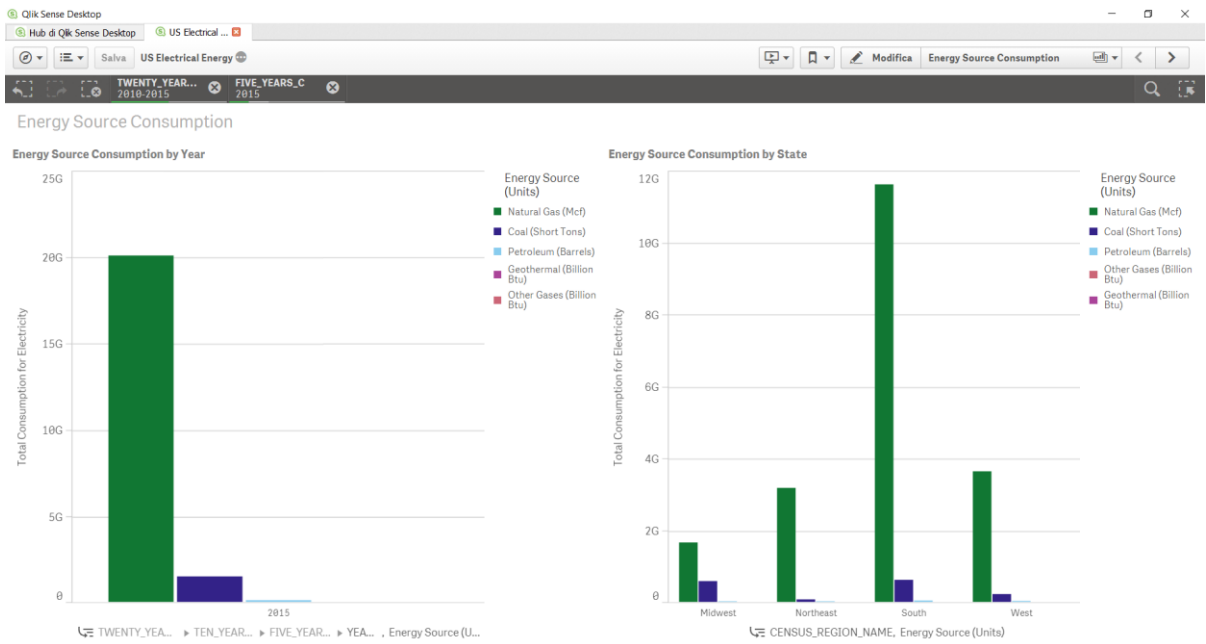


Figura 12. Grafici dopo il secondo drill-down per selezionare l'anno 2015.

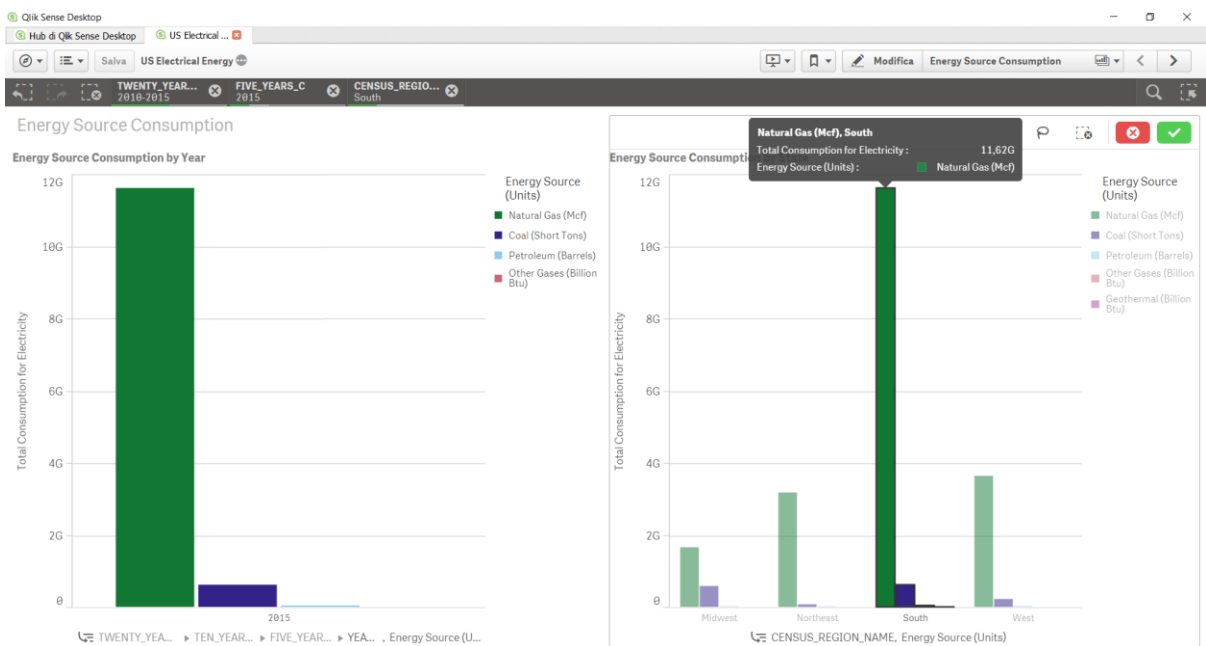


Figura 13. Selezione dei risultati relativi alla Regione South per l'anno 2015.

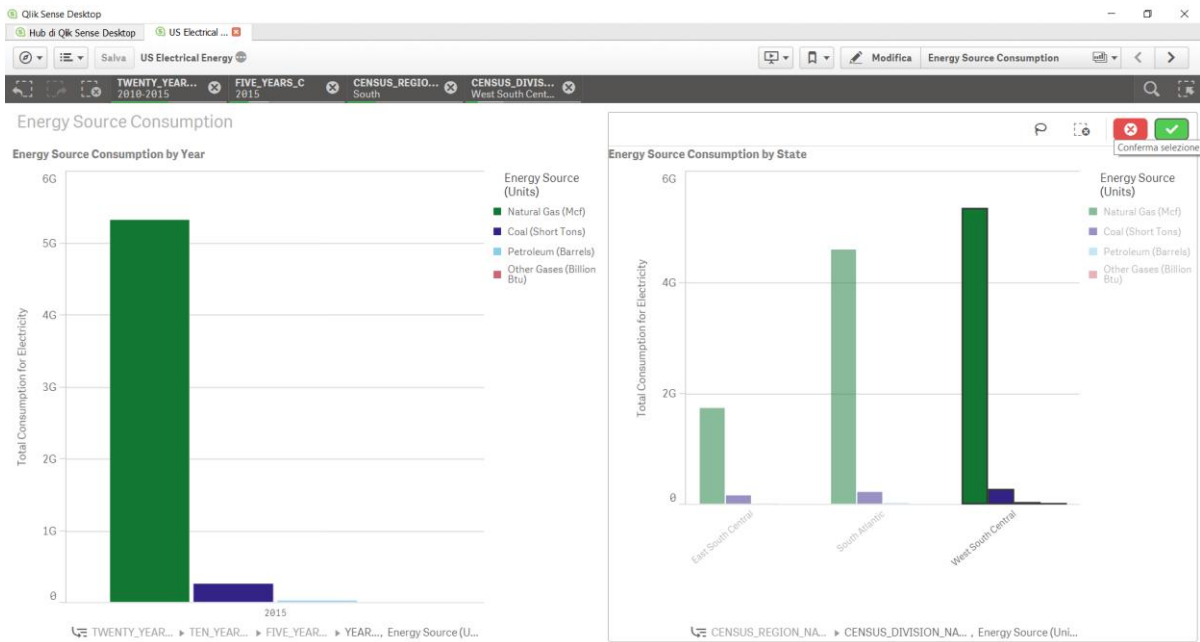


Figura 14. Selezione dei risultati relativi alla Divisione West South Central per l'anno 2015.

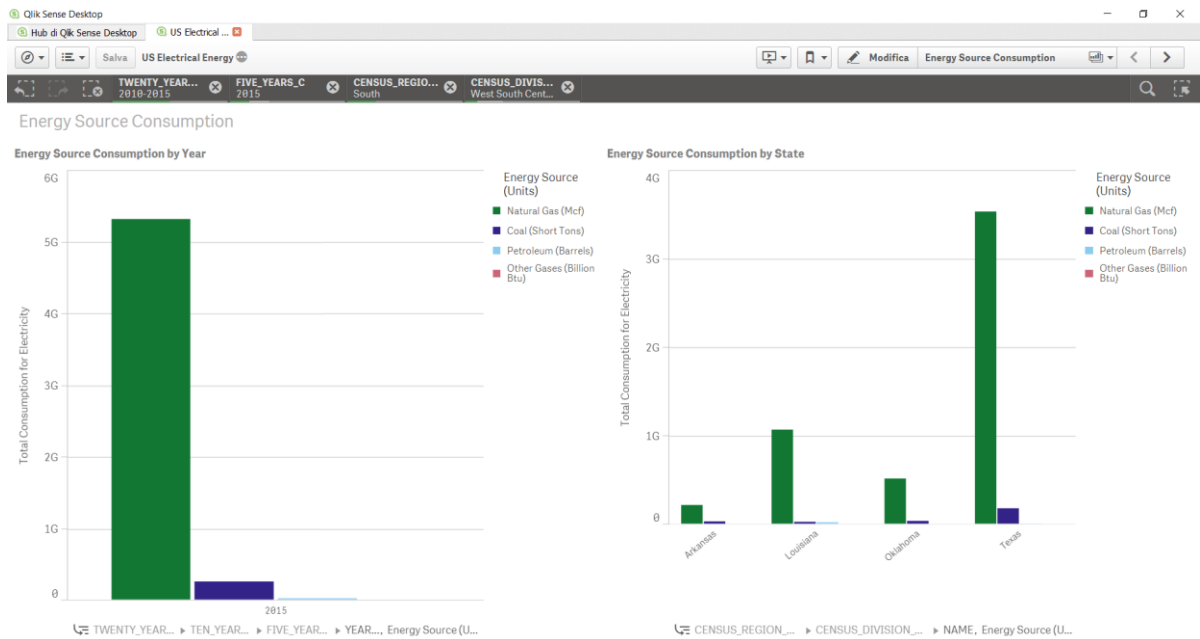


Figura 15. Grafici relativi agli Stati del West South Central per l'anno 2015 in merito al consumo di combustibili fossili.

Nella Figura 16 viene mostrata la dashboard relativa alla produzione di energia elettrica su base temporale; in particolare, sono state messi a confronto i grafici relativi alle fonti di energia rinnovabili (a sinistra) e quelli relativi alle fonti di energia non rinnovabili (a destra), sui quali è possibile muoversi lungo la gerarchia temporale.

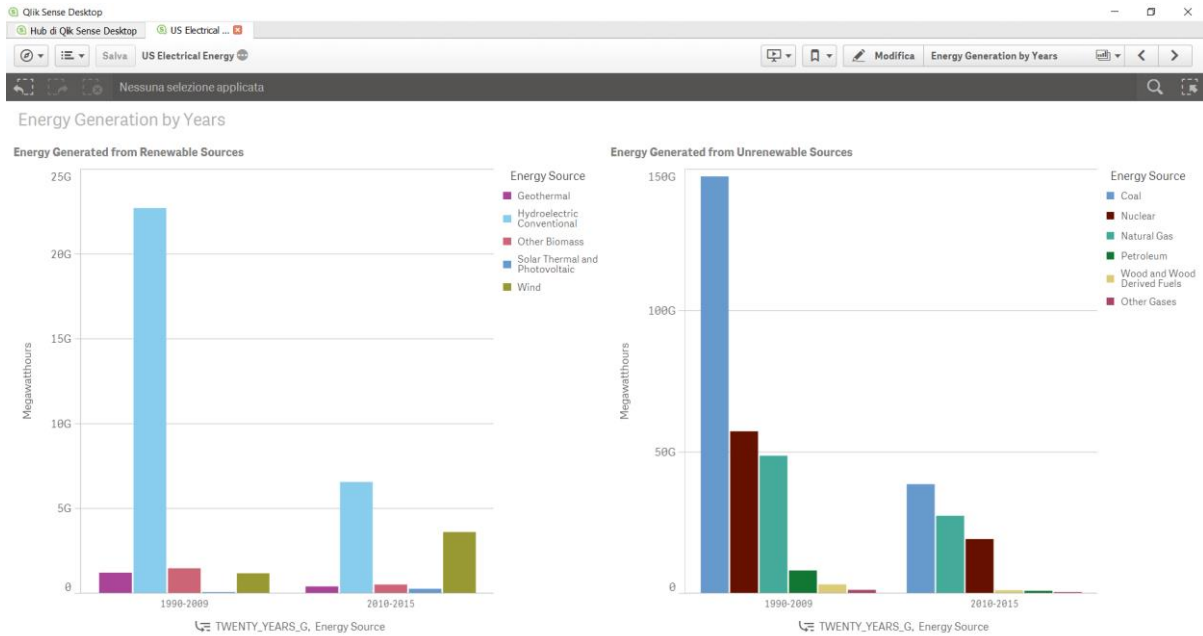


Figura 16. Grafici relativi alla produzione di energia elettrica da fonti rinnovabili e non rinnovabili su base temporale.

Nella Figura 17 è riportato il grafico risultante da una operazione di drill-down per selezionare l'anno 2015.

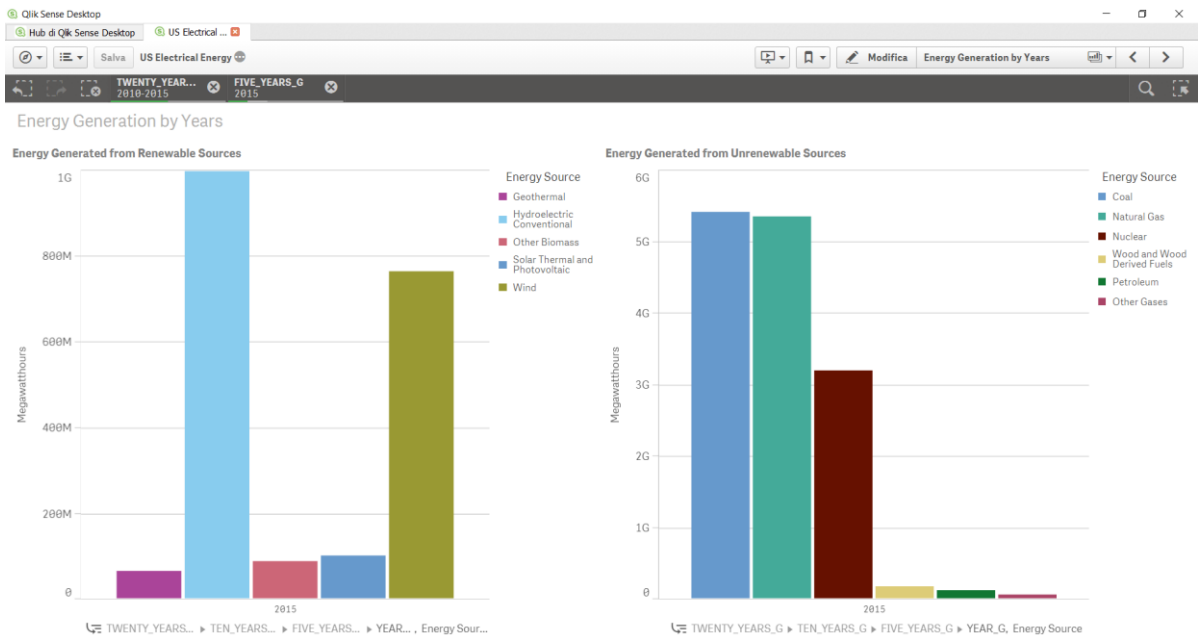


Figura 17. Grafici relativi alla produzione di energia elettrica per l'anno 2015.

Nella Figura 18 viene mostrata la dashboard relativa alla produzione di energia elettrica su base geografica; anche in questo caso, sono state messi a confronto i grafici relativi alle fonti di energia rinnovabili (a sinistra) e quelli relativi alle fonti di energia non rinnovabili (a destra), sui quali è possibile muoversi lungo la gerarchia geografica.

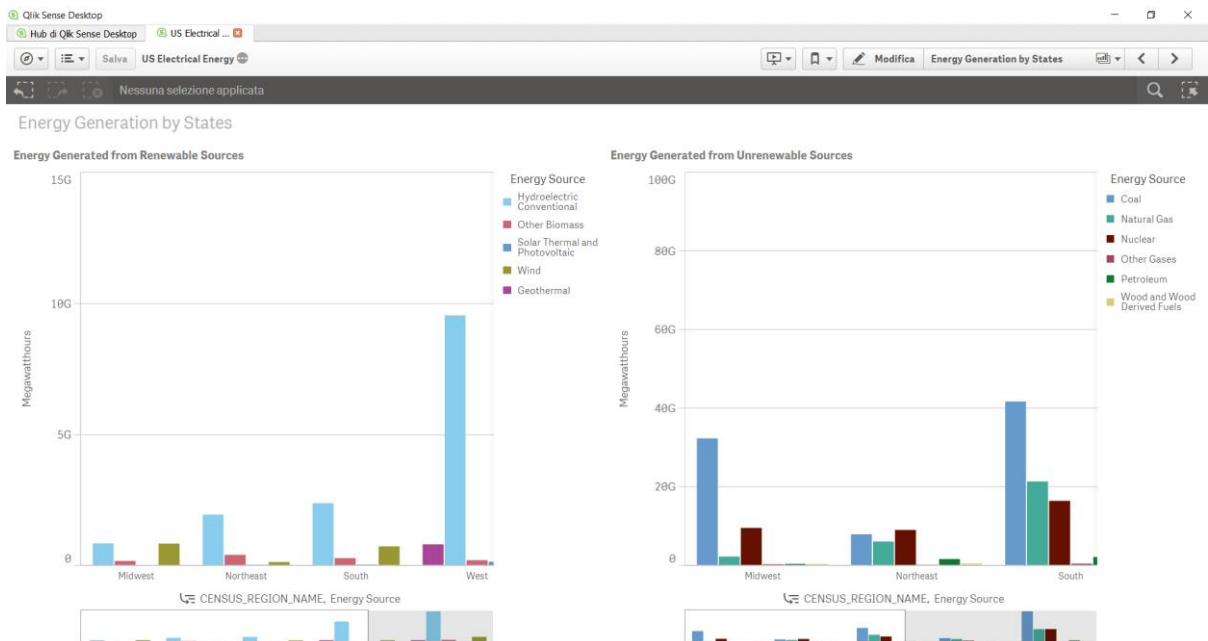


Figura 18. Grafici relativi alla produzione di energia elettrica da fonti rinnovabili e non rinnovabili su base geografica.

Nella Figura 19 è riportato il grafico risultante da una operazione di drill-down per selezionare gli Stati appartenenti alla Divisione West South Central.

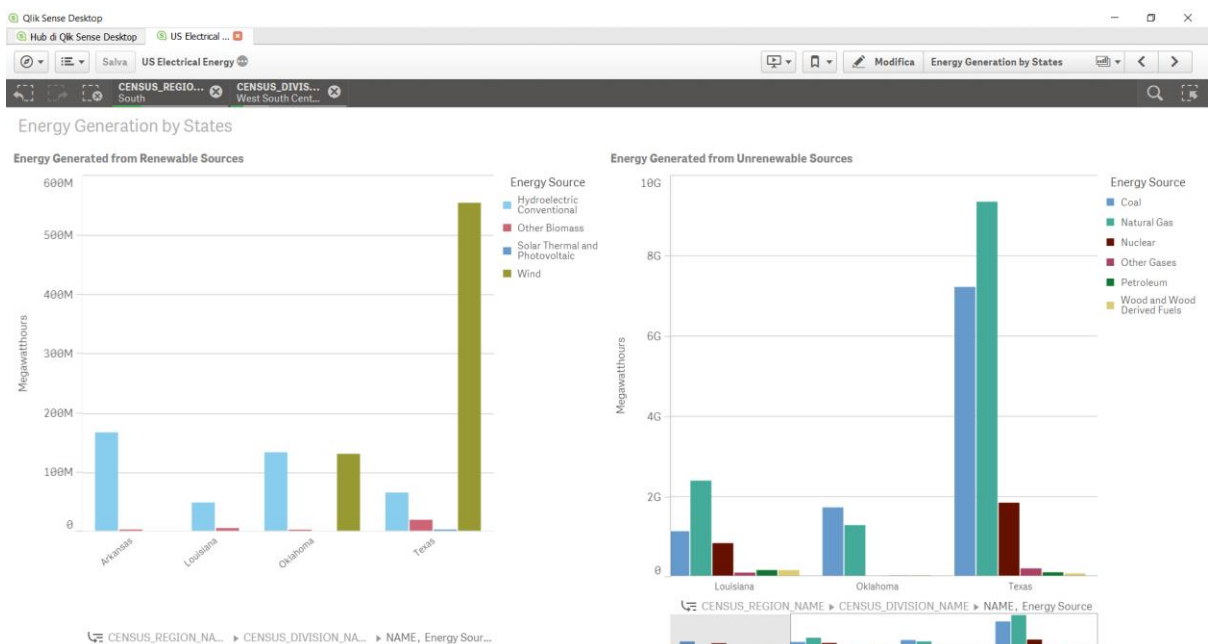


Figura 19. Grafici relativi alla produzione di energia elettrica per gli Stati della Divisione West South Central.

Nelle Figure da 20 a 23 sono riportati i grafici relativi alle emissioni prodotte negli anni dalle varie centrali elettriche, confrontati per sostanza emessa (CO₂, SO₂ e NO_x) e suddivisi su base geografica. In questo caso, attraverso operazioni di drill-down, è possibile arrivare ad una particolareggiata suddivisione per Stato; di volta in volta, i cruscotti mostrati in basso nella dashboard aggiorneranno i totali per ogni sostanza. Attraverso quest'unica dashboard, dunque, è possibile avere un'idea delle zone degli Stati Uniti in cui si riscontrano alte quantità di inquinamento, per poi condurre, eventualmente, ulteriori analisi diagnostiche, oppure per promuovere campagne mirate di sensibilizzazione all'utilizzo di fonti rinnovabili per la produzione di energia elettrica.

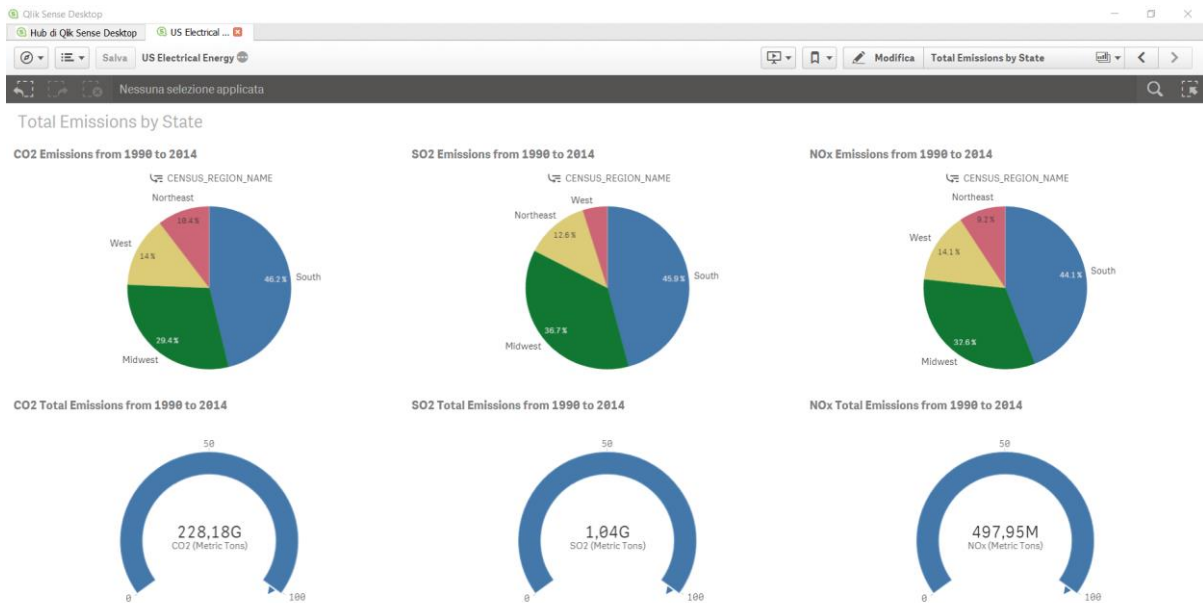


Figura 20. Grafici relativi alle emissioni prodotte dalle centrali elettriche su scala Regionale.

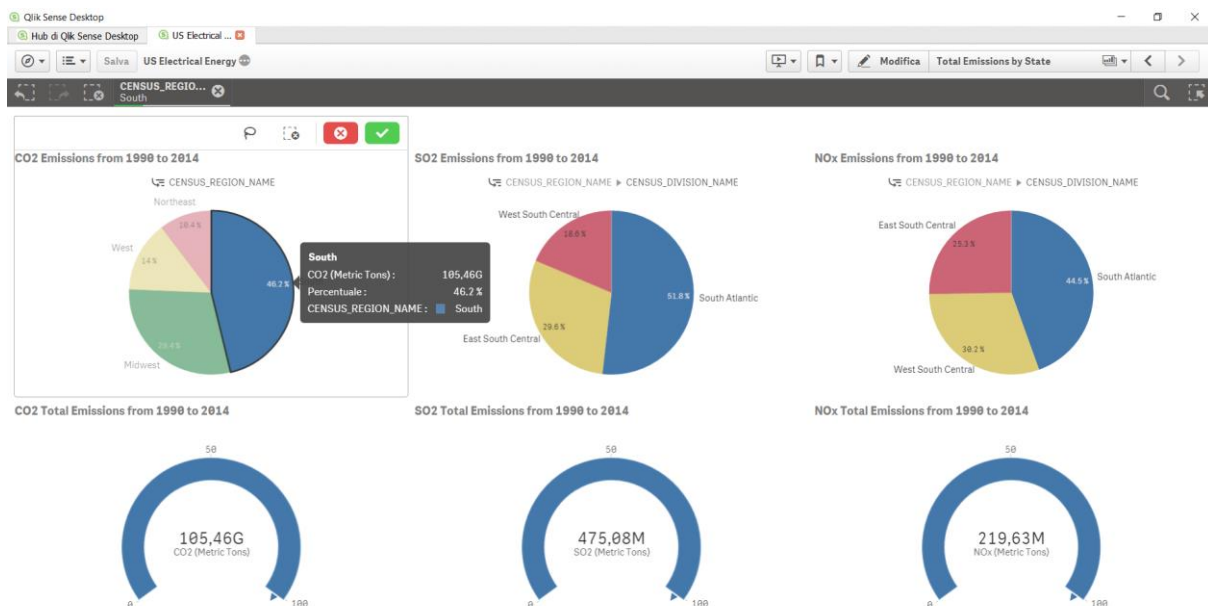


Figura 21. Selezione della Regione South per il primo livello di drill-down su base geografica.

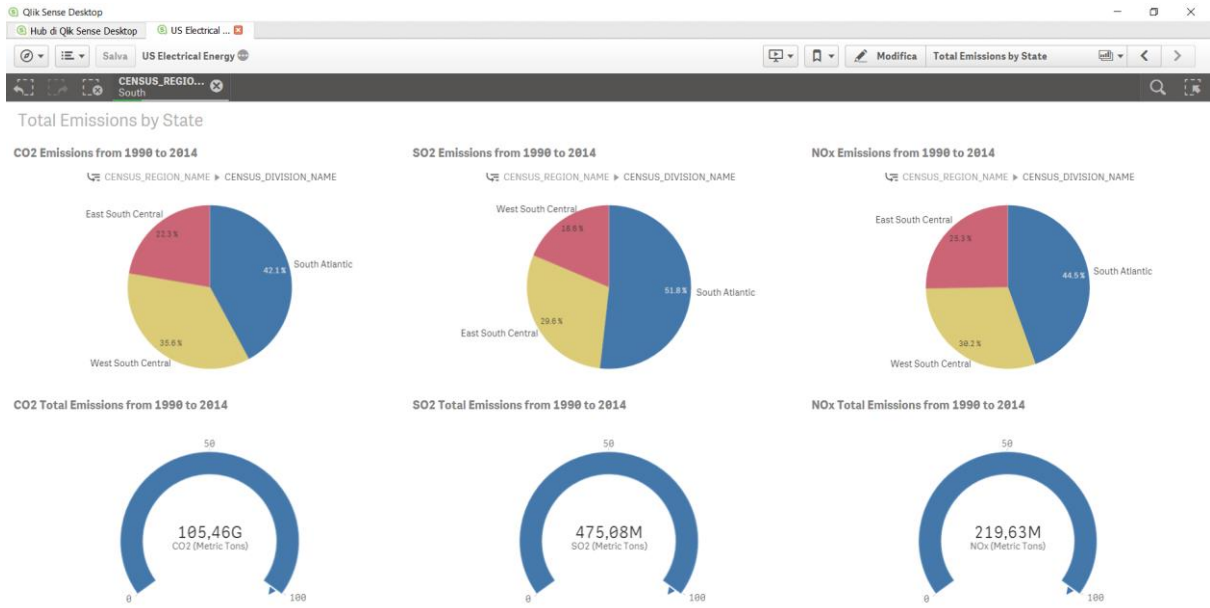


Figura 22. Grafici relativi alle emissioni prodotte dalle centrali elettriche su scala delle Divisioni South.

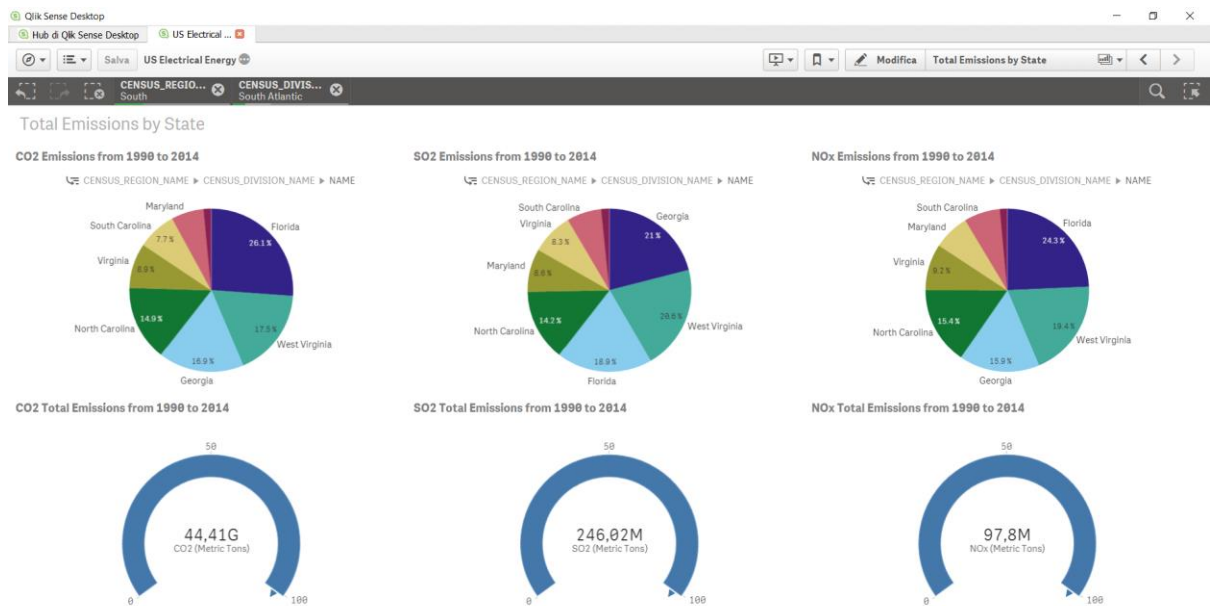


Figura 23. Grafici relativi alle emissioni prodotte dalle centrali elettriche degli Stati del South Atlantic.

Analisi mediante Tableau

In maniera del tutto analoga a Qlik Sense sono stati realizzati i grafici corrispondenti in Tableau. In particolare, in questo tool sono stati creati 3 Book separati, in modo da importare solo le tabelle necessarie per una determinata analisi e snellire le operazioni di calcolo necessarie a produrre i grafici finali.

Il primo Book, mostrato nelle Figure 24 e 25, è relativo ai consumi di combustibili fossili per la produzione di energia elettrica e contiene due Worksheet, uno contenente i grafici secondo la gerarchia temporale e l'altro contenente i grafici secondo la gerarchia geografica, chiamate rispettivamente Time e Location al momento della loro creazione all'interno di Tableau.

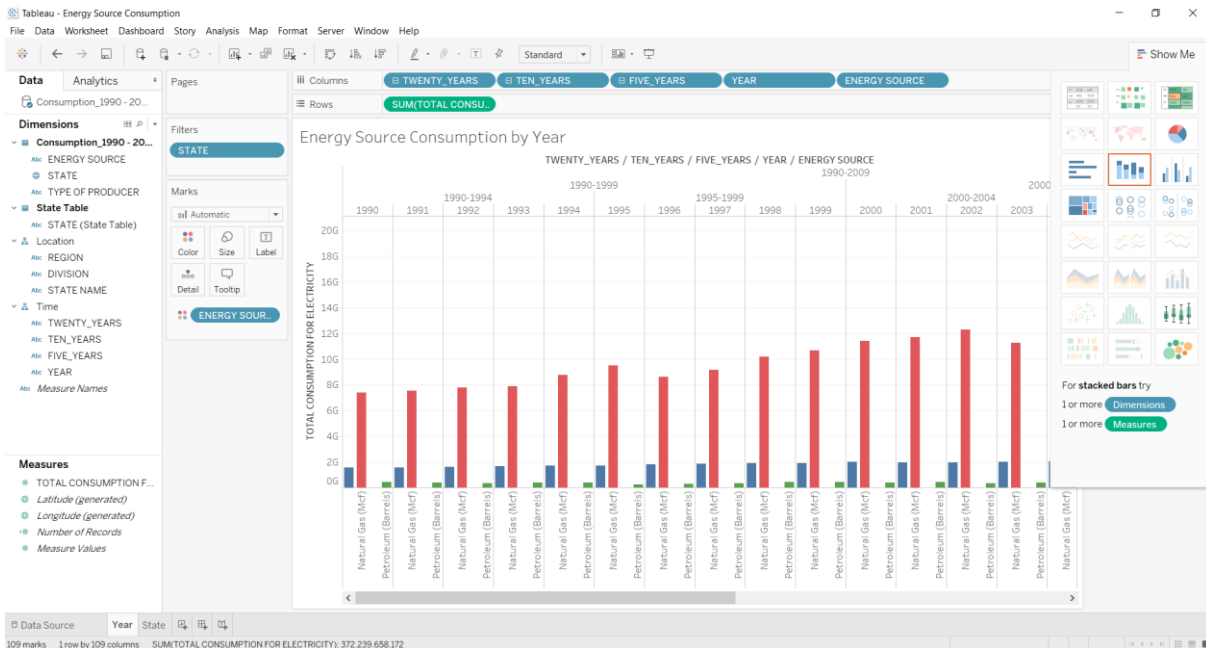


Figura 24. Grafici relativi al consumo di combustibili fossili per la produzione di elettricità su scala temporale.

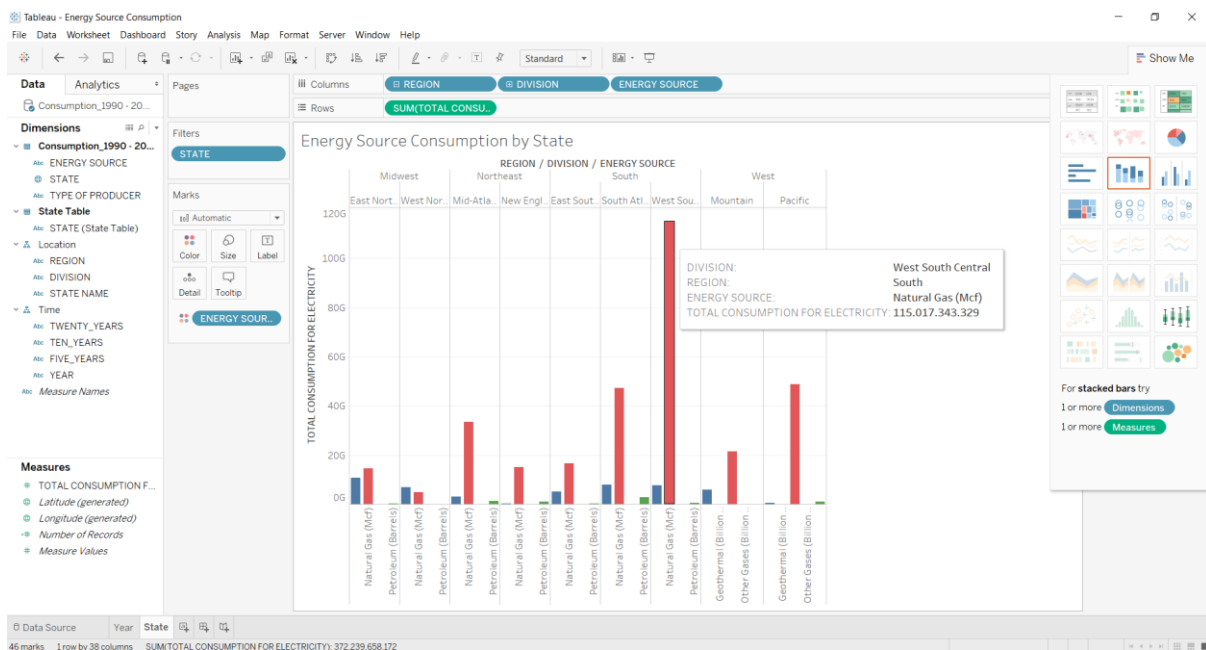


Figura 25. Grafici relativi al consumo di combustibili fossili per la produzione di elettricità su scala geografica.

Il secondo Book, mostrato nelle Figure da 26 a 29, è relativo alla produzione di energia elettrica; in particolare, sono stati creati 4 Worksheet:

- **YearR**: relativo alla produzione di energia elettrica utilizzando fonti rinnovabili, basato su scala temporale (Figura 26).
- **YearU**: relativo alla produzione di energia elettrica utilizzando fonti non rinnovabili, basato su scala temporale (Figura 27).
- **StateR**: relativo alla produzione di energia elettrica utilizzando fonti rinnovabili, basato su scala geografica (Figura 28).
- **StateU**: relativo alla produzione di energia elettrica utilizzando fonti non rinnovabili, basato su scala geografica (Figura 29).

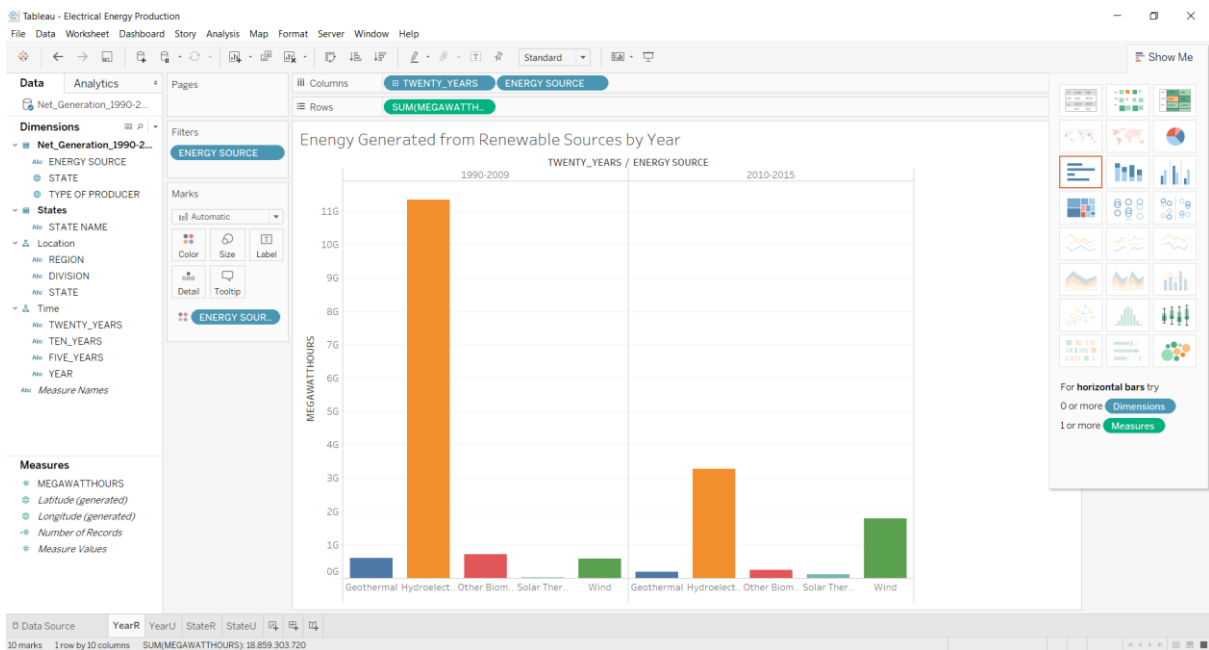


Figura 26. Grafico relativo alla produzione di energia elettrica utilizzando fonti rinnovabili, basato su scala temporale.

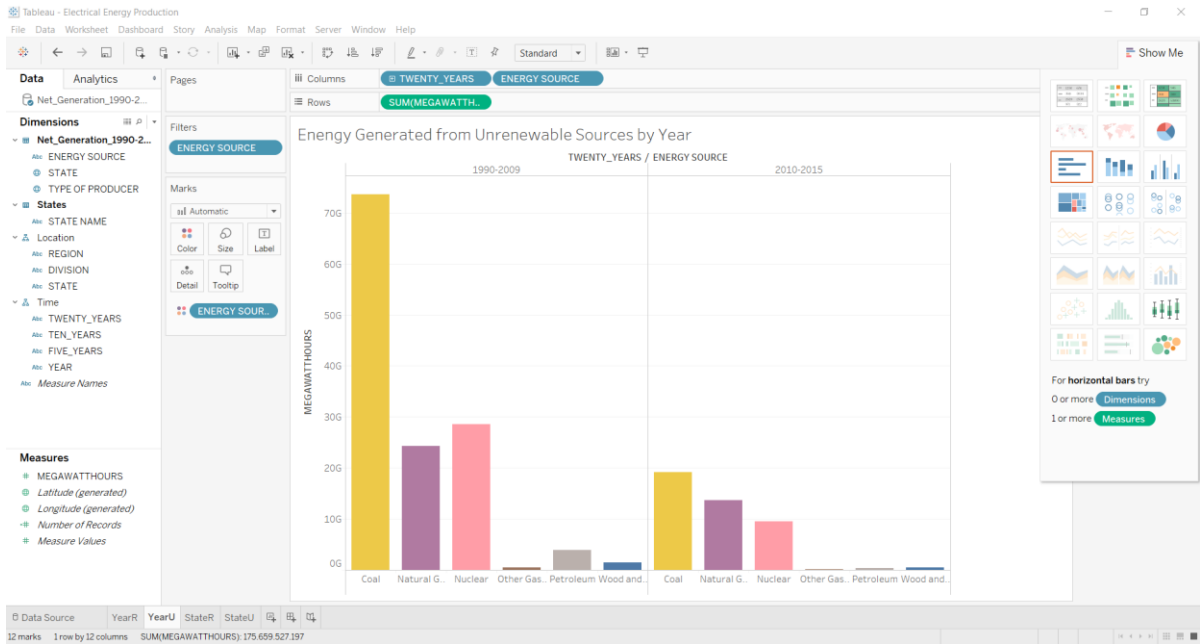


Figura 27. Grafico relativo alla produzione di energia elettrica utilizzando fonti non rinnovabili, basato su scala temporale.

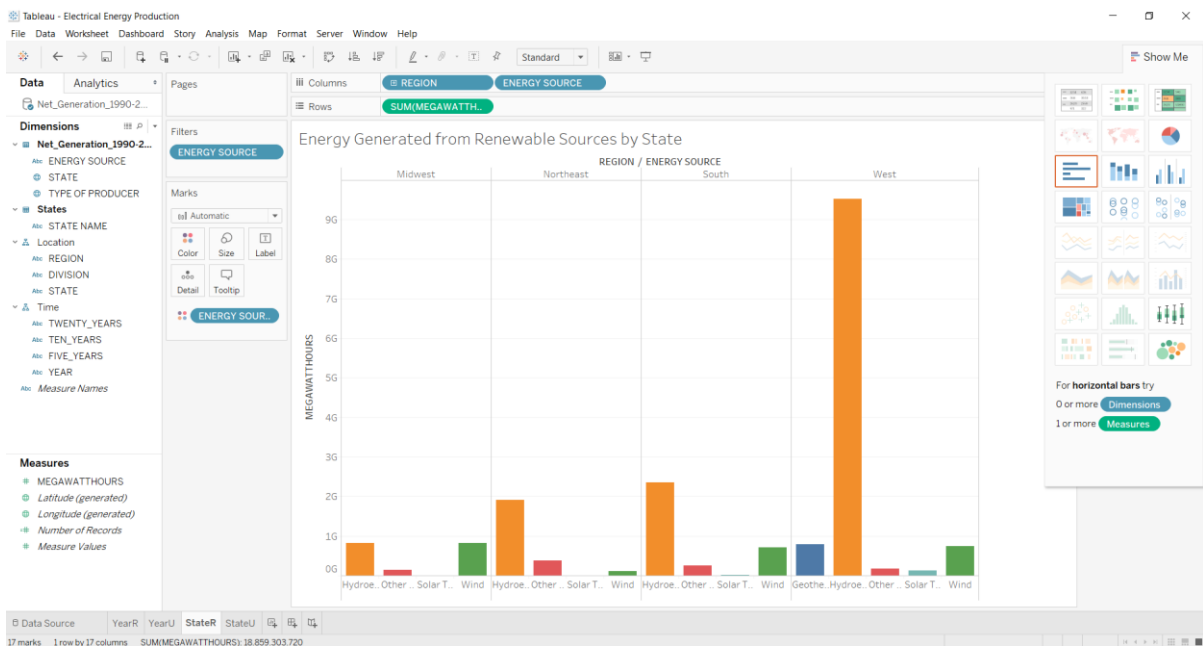


Figura 28. Grafico relativo alla produzione di energia elettrica utilizzando fonti rinnovabili, basato su scala geografica.

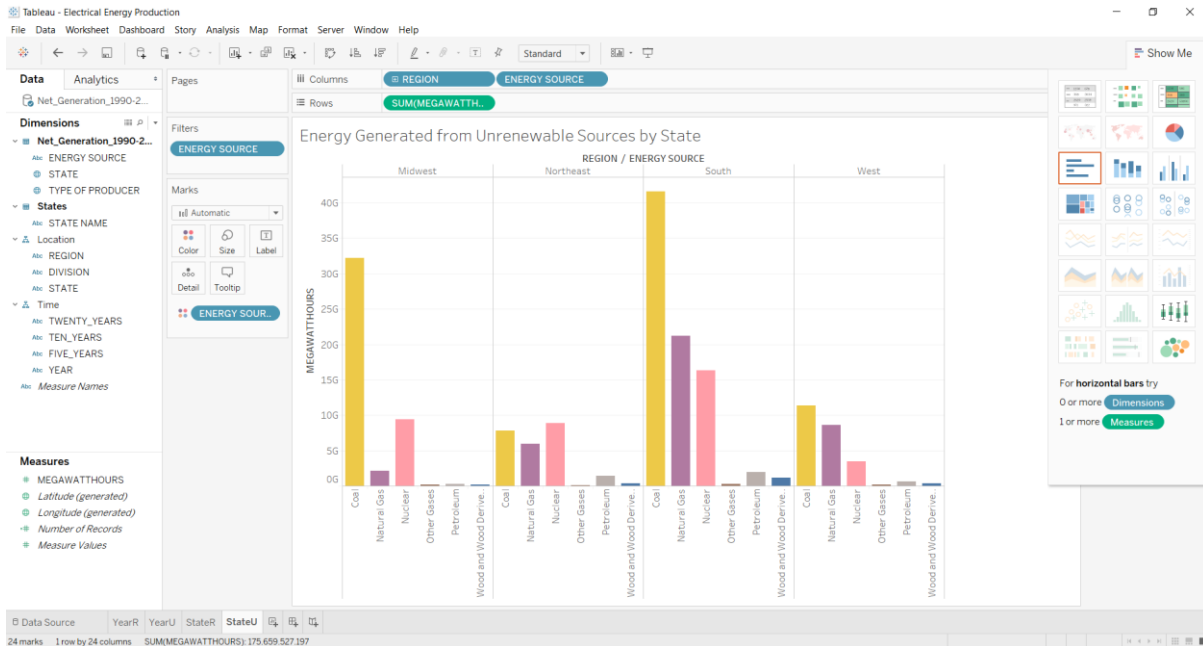


Figura 29. Grafico relativo alla produzione di energia elettrica utilizzando fonti non rinnovabili, basato su scala geografica.

Il terzo Book, dedicato alle emissioni prodotte dalle centrali elettriche, rappresenta una sostanziale differenza rispetto a quanto visto in Qlik Sense. Infatti, come è possibile osservare nelle Figure da 30 a 32, attraverso Tableau è possibile creare delle mappe di calore, utilizzando coordinate generate automaticamente dal tool stesso e integrate con una dimensione geografica (nel caso in esame la sigla degli Stati). In questo modo è stato possibile rappresentare gli Stati più “green”, ovvero quelli che producono meno emissioni, creando un impatto visivo sicuramente migliore rispetto ai grafici a torta realizzati in Qlik Sense. Inoltre, passando sopra ad uno Stato con il cursore, è possibile vedere i dettagli delle emissioni per quest’ultimo.

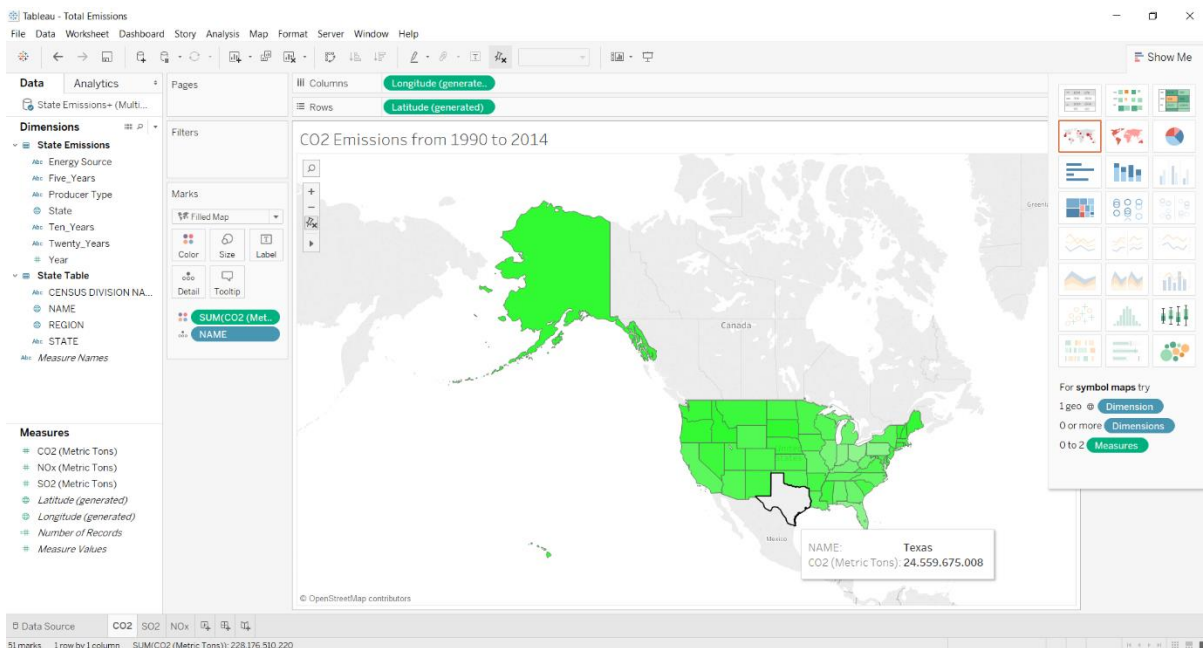


Figura 30. Mappa di calore relativa alle emissioni di CO₂ prodotte dalle centrali elettriche dei vari Stati.

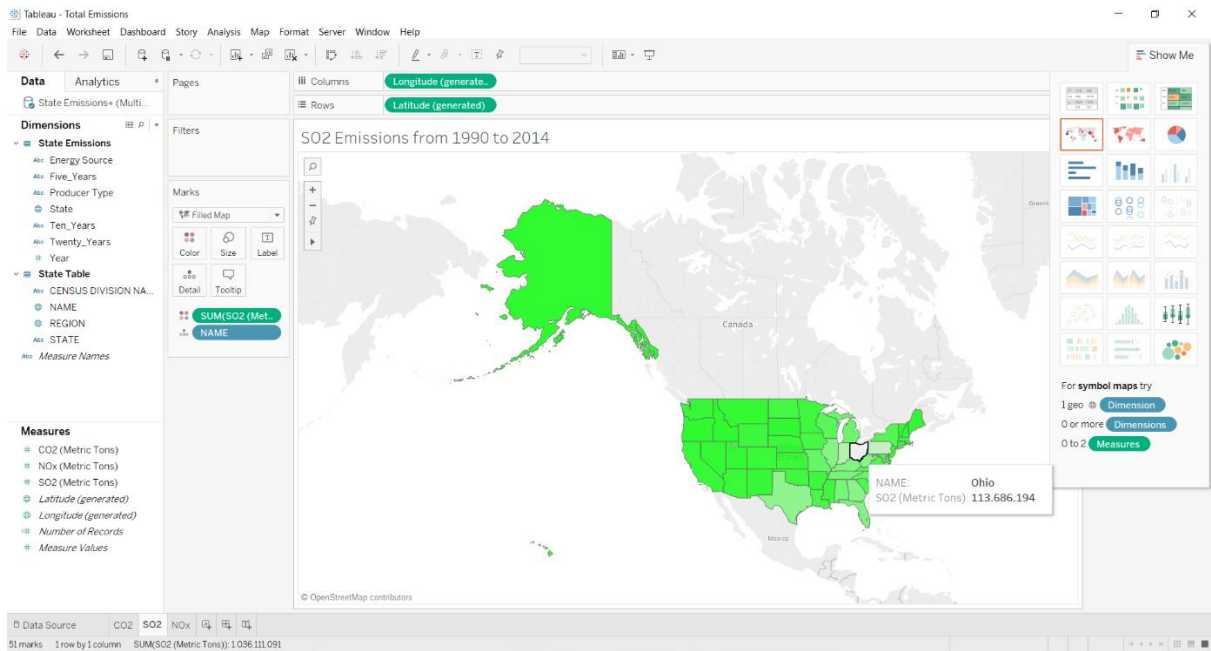


Figura 31. Mappa di calore relativa alle emissioni di SO_2 prodotte dalle centrali elettriche dei vari Stati.

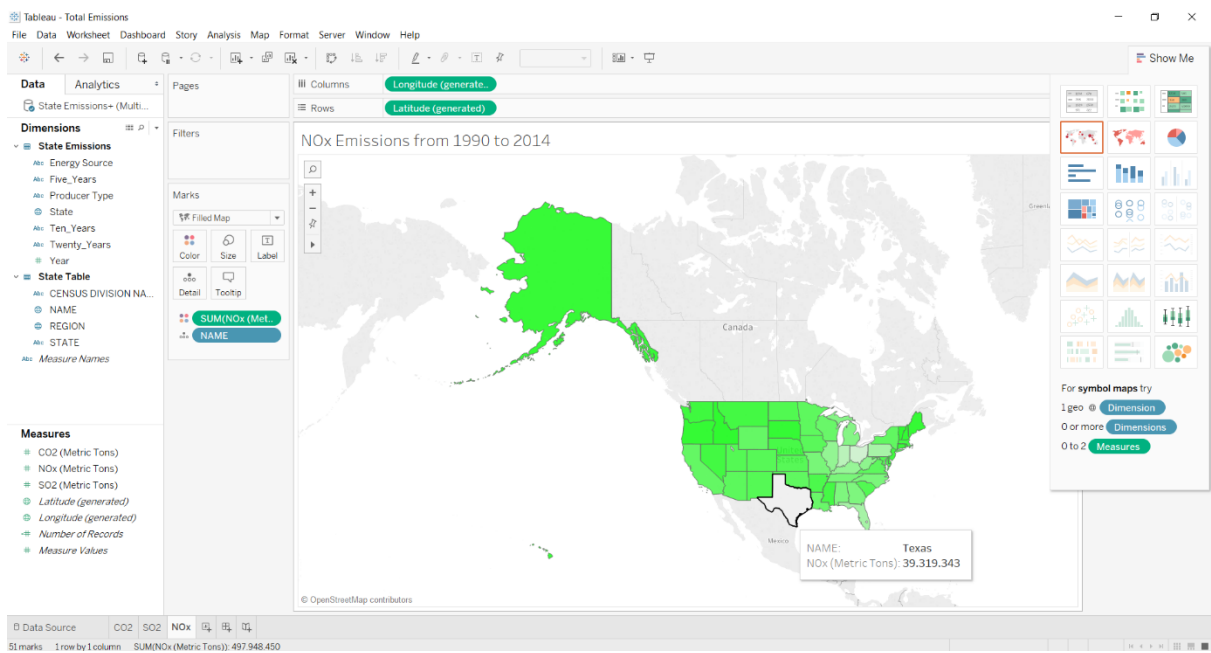


Figura 32. Mappa di calore relativa alle emissioni di NO_x prodotte dalle centrali elettriche dei vari Stati.

Analisi mediante MongoDB

A differenza di Qlik Sense e Tableau, MongoDB non permette di avere delle dashboard per visualizzare sotto forma di mappe, grafici o istogrammi i risultati di un'analisi di tipo Big Data Analytics.

Infatti, esso nasce principalmente come software per la gestione di DBMS di tipo NoSQL orientato ai documenti ed è interfacciabile soltanto da terminale. Tuttavia, mediante l'utility MongoChef, è possibile ricavare informazioni e stamparle su una console testuale a partire dai dati importati all'interno di MongoDB, utilizzando opportuni script in Javascript che realizzano delle query sulla base di dati.

Dopo aver installato MongoDB, tramite il terminale abbiamo creato il database documentale *energydb* ed importato al suo interno le tabelle mediante l'utility *mongoimport*, trasformate in file CSV (Comma Separated Value), in quanto MongoDB può ricevere in ingresso soltanto dati di tipo JSON, CSV e TSV.

Per rendere più immediato l'avvio del server di MongoDB abbiamo creato uno script bash, avviato il quale è possibile connettersi al database tramite l'applicazione MongoChef.

All'interno di MongoChef, successivamente, abbiamo realizzato degli script per interrogare il database ed estrarre le informazioni necessarie a condurre le stesse analisi proposte con Qlik Sense e Tableau.

Il primo script realizzato è *ResourcesConsumptionByState.js* (Listato 1); esso è relativo al consumo di combustibili fossili da parte dei singoli Stati. L'esecuzione di tale script all'interno di MongoChef produce la creazione di 4 Tab nella console IntelliShell, uno per ognuna delle risorse considerate (Coal, Natural Gas, Other Gases e Petroleum), come è possibile vedere in Figura 33, ordinati per quantità decrescente.

```
1. //Resources Consumption by State
2.
3. db.consumption.aggregate(
4.     {$match : {"ENERGY SOURCE (UNITS)" : "Coal (Short Tons)"},
5.     {$match: {"STATE": {$not: {$eq:"US-TOTAL"}}}},
6.     {$match: {"STATE": {$not: {$eq:"US-Total"}}}},
7.     {$group : {_id : "$STATE", Coal:{$sum:"$CONSUMPTION for ELECTRICITY"}, }},
8.     {$sort: {Coal: -1}}
9. )
10. db.consumption.aggregate(
11.     {$match : {"ENERGY SOURCE (UNITS)" : "Natural Gas (Mcf)"},
12.     {$match: {"STATE": {$not: {$eq:"US-TOTAL"}}}},
13.     {$match: {"STATE": {$not: {$eq:"US-Total"}}}},
14.     {$group : {_id : "$STATE", Natural_Gas:{$sum:"$CONSUMPTION for ELECTRICITY"}, }},
15.     {$sort: {Natural_Gas: -1}}
16. )
17. db.consumption.aggregate(
18.     {$match : {"ENERGY SOURCE (UNITS)" : "Other Gases (Billion BTU)"},
19.     {$match: {"STATE": {$not: {$eq:"US-TOTAL"}}}},
20.     {$match: {"STATE": {$not: {$eq:"US-Total"}}}},
21.     {$group : {_id : "$STATE", Other_Gases:{$sum:"$CONSUMPTION for ELECTRICITY"}, }},
22.     {$sort: {Other_Gases: -1}}
23. )
24. db.consumption.aggregate(
25.     {$match : {"ENERGY SOURCE (UNITS)" : "Petroleum (Barrels)"},
26.     {$match: {"STATE": {$not: {$eq:"US-TOTAL"}}}},
27.     {$match: {"STATE": {$not: {$eq:"US-Total"}}}},
28.     {$group : {_id : "$STATE", Petroleum:{$sum:"$CONSUMPTION for ELECTRICITY"}, }},
29.     {$sort: {Petroleum: -1}}
30. )
```

Listato 1. Codice all'interno dello script *ResourcesConsumptionByState.js*

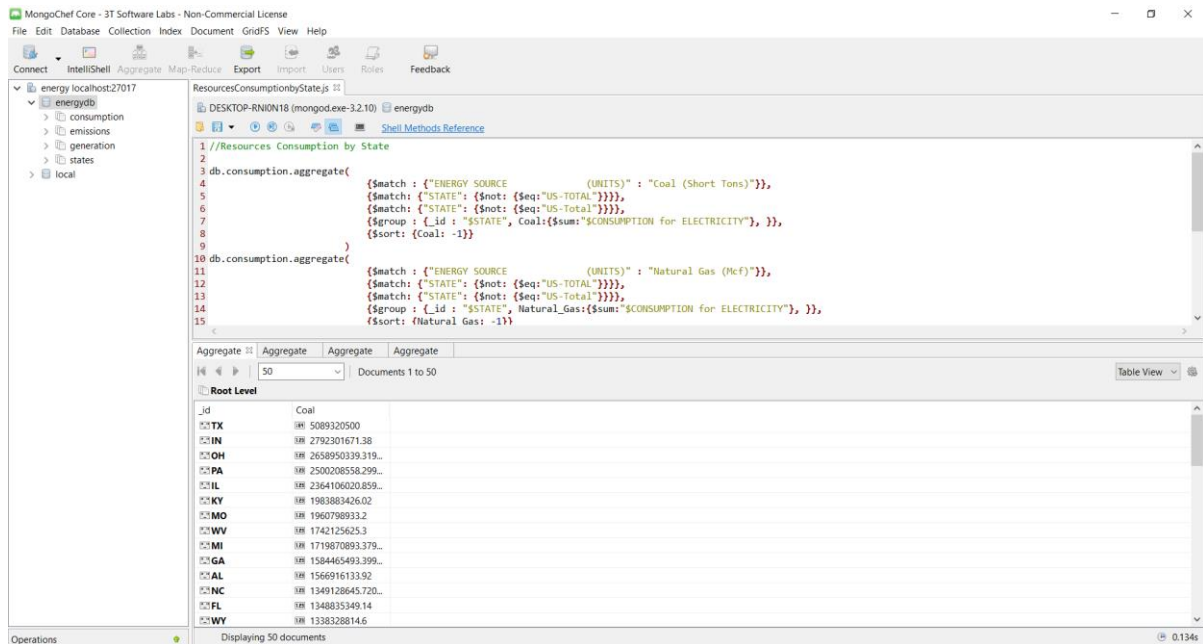


Figura 33. Risultato dell'esecuzione dello script `ResourcesConsumptionByState.js` in MongoChef.

Il secondo script realizzato è `ResourcesConsumptionByYear.js` (Listato 2); esso è relativo al consumo di combustibili fossili nei vari anni dal 1990 al 2015. L'esecuzione di tale script all'interno di MongoChef produce la creazione di 4 Tab nella console IntelliShell, uno per ognuna delle risorse considerate (Coal, Natural Gas, Other Gases e Petroleum), come è possibile vedere in Figura 34, ordinati per quantità decrescente.

```

1. //Resources Consumption by Year
2.
3. db.consumption.aggregate(
4.     {$match : {"ENERGY SOURCE (UNITS)" : "Coal (Short Tons)"},
5.     {$match : {"STATE": {$not : {$eq:"US-TOTAL"}}}},
6.     {$match : {"STATE": {$not : {$eq:"US-Total"}}}},
7.     {$group : {_id : "$YEAR", Coal:{$sum:"$CONSUMPTION for ELECTRICITY"}, }},
8.     {$sort : {Coal: -1}}
9. )
10. db.consumption.aggregate(
11.     {$match : {"ENERGY SOURCE (UNITS)" : "Natural Gas (Mcf)"},
12.     {$match : {"STATE": {$not : {$eq:"US-TOTAL"}}}},
13.     {$match : {"STATE": {$not : {$eq:"US-Total"}}}},
14.     {$group : {_id : "$YEAR", Natural_Gas:{$sum:"$CONSUMPTION for ELECTRICITY"}, }},
15.     {$sort : {Natural_Gas: -1}}
16. )
17. db.consumption.aggregate(
18.     {$match : {"ENERGY SOURCE (UNITS)" : "Other Gases (Billion BTU)"},
19.     {$match : {"STATE": {$not : {$eq:"US-TOTAL"}}}},
20.     {$match : {"STATE": {$not : {$eq:"US-Total"}}}},
21.     {$group : {_id : "$YEAR", Other_Gases:{$sum:"$CONSUMPTION for ELECTRICITY"}, }},
22.     {$sort : {Other_Gases: -1}}
23. )
24. db.consumption.aggregate(
25.     {$match : {"ENERGY SOURCE (UNITS)" : "Petroleum (Barrels)"},
26.     {$match : {"STATE": {$not : {$eq:"US-TOTAL"}}}},
27.     {$match : {"STATE": {$not : {$eq:"US-Total"}}}},
28.     {$group : {_id : "$YEAR", Petroleum:{$sum:"$CONSUMPTION for ELECTRICITY"}, }},
29.     {$sort : {Petroleum: -1}}
30. )

```

Listato 2. Codice all'interno dello script `ResourcesConsumptionByYear.js`

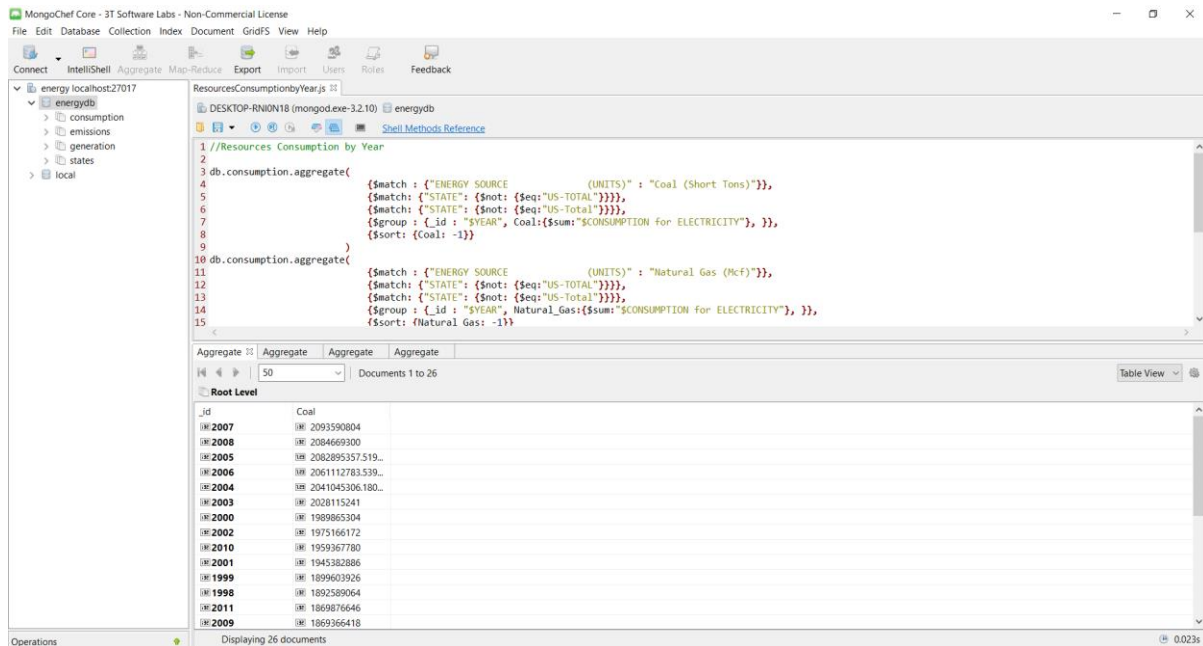


Figura 34. Risultato dell'esecuzione dello script ResourcesConsumptionByYear.js in MongoChef.

In seguito, sono stati realizzati gli script *EnergyGenerationUnrenewableByYear.js* (Listato 3) e *EnergyGenerationRenewableByYear.js* (Listato 4); essi sono relativi all'energia generata, rispettivamente, da fonti non rinnovabili e rinnovabili nei vari anni dal 1990 al 2015. L'esecuzione di tali script all'interno di MongoChef produce la creazione di vari Tab nella console IntelliShell, uno per ognuna delle risorse considerate, come è possibile vedere in Figura 35 ed in Figura 36, ordinati per quantità decrescente.

```

1. //Energy Generated from Unrenewable Sources by Year
2.
3. db.generation.aggregate(
4.     {$match : {"ENERGY SOURCE" : "Coal"}},
5.     {$group : {_id : "$YEAR", Coal:{$sum:"$GENERATION (Megawatthours)"}, }}
6. )
7.
8. db.generation.aggregate(
9.     {$match : {"ENERGY SOURCE" : "Natural Gas"}},
10.    {$group : {_id : "$YEAR", Natural_Gas:{$sum:"$GENERATION (Megawatthours)"}, }}
11. )
12.
13. db.generation.aggregate(
14.    {$match : {"ENERGY SOURCE" : "Nuclear"}},
15.    {$group : {_id : "$YEAR", Nuclear:{$sum:"$GENERATION (Megawatthours)"}, }}
16. )
17.
18. db.generation.aggregate(
19.    {$match : {"ENERGY SOURCE" : "Other Gases"}},
20.    {$group : {_id : "$YEAR", Other_Gases:{$sum:"$GENERATION (Megawatthours)"}, }}
21. )
22.
23. db.generation.aggregate(
24.    {$match : {"ENERGY SOURCE" : "Petroleum"}},
25.    {$group : {_id : "$YEAR", Petroleum:{$sum:"$GENERATION (Megawatthours)"}, }}
26. )
27.
28. db.generation.aggregate(
29.    {$match : {"ENERGY SOURCE" : "Wood and Wood Derived Fuels"}},
30.    {$group : {_id : "$YEAR", Wood_and_Wood_Derived_Fuels:{$sum:"$GENERATION (Megawatthours)"}, }}
31. )

```

Listato 3. Codice all'interno dello script EnergyGenerationUnrenewableByYear.js

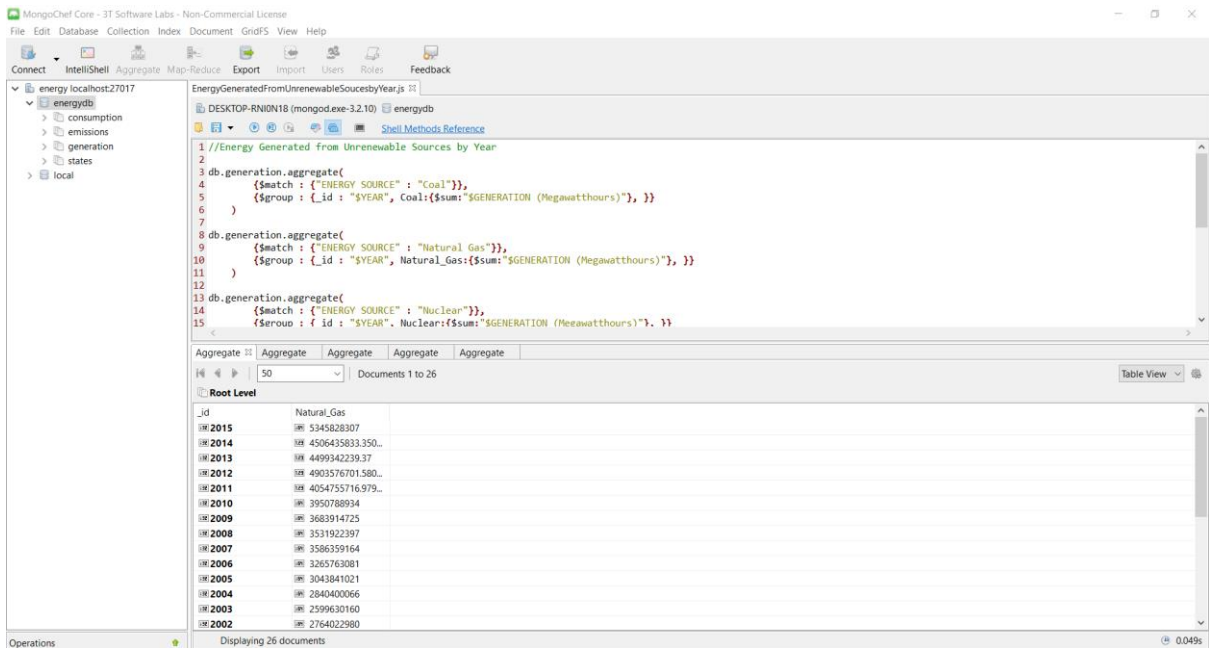


Figura 35. Risultato dell'esecuzione dello script EnergyGenerationUnrenewableByYear.js in MongoChef.

```

1. //Energy Generated from Renewable Sources by Year
2.
3. db.generation.aggregate(
4.     {$match : {"ENERGY SOURCE" : "Geothermal"}},
5.     {$group : {_id : "$YEAR", Geothermal:{$sum:"$GENERATION (Megawatthours)"}, }}
6. )
7.
8. db.generation.aggregate(
9.     {$match : {"ENERGY SOURCE" : "Hydroelectric Conventional"}},
10.    {$group : {_id : "$YEAR", Hydroelectric:{$sum:"$GENERATION (Megawatthours)"}, }}
11. )
12.
13. db.generation.aggregate(
14.    {$match : {"ENERGY SOURCE" : "Other Biomass"}},
15.    {$group : {_id : "$YEAR", Other_Biomass:{$sum:"$GENERATION (Megawatthours)"}, }}
16. )
17.
18. db.generation.aggregate(
19.    {$match : {"ENERGY SOURCE" : "Solar Thermal and Photovoltaic"}},
20.    {$group : {_id : "$YEAR", Photovoltaic:{$sum:"$GENERATION (Megawatthours)"}, }}
21. )
22.
23. db.generation.aggregate(
24.    {$match : {"ENERGY SOURCE" : "Wind"}},
25.    {$group : {_id : "$YEAR", Wind:{$sum:"$GENERATION (Megawatthours)"}, }}
26. )

```

Listato 4. Codice all'interno dello script EnergyGenerationRenewableByYear.js

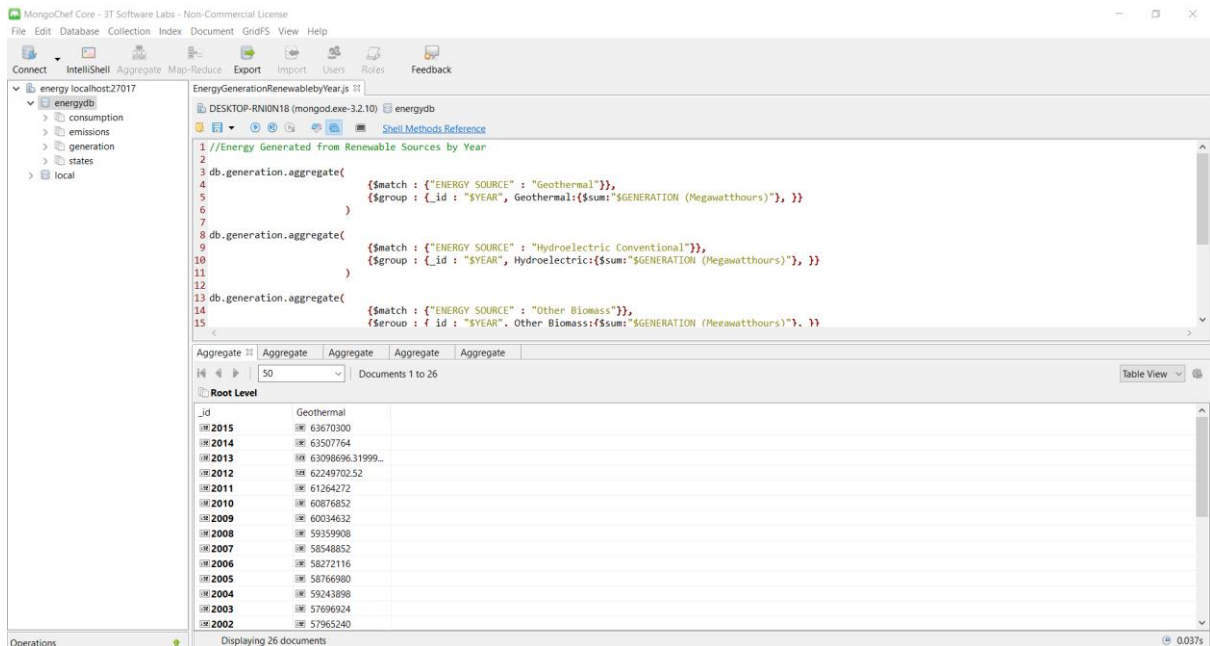


Figura 36. Risultato dell'esecuzione dello script *EnergyGenerationRenewableByYear.js* in MongoChef.

Successivamente, sono stati realizzati gli script *EnergyGenerationUnrenewableByState.js* (Listato 5) e *EnergyGenerationRenewableByState.js* (Listato 6); essi sono relativi all'energia generata, rispettivamente, da fonti non rinnovabili e rinnovabili dalle centrali elettriche dei vari Stati. L'esecuzione di tali script all'interno di MongoChef produce la creazione di vari Tab nella console IntelliShell, uno per ognuna delle risorse considerate, come è possibile vedere in Figura 37 ed in Figura 38, ordinati per quantità decrescente.

```

1. //Energy Generated from Unrenewable Sources by State
2.
3. db.generation.aggregate(
4.     {$match : {"ENERGY SOURCE" : "Coal"}},
5.     {$match : {"STATE" : {$not : {$eq:"US-TOTAL"}}}},
6.     {$match : {"STATE" : {$not : {$eq:"US-Total"}}}},
7.     {$group : {_id : "$STATE", Coal:{$sum:"$GENERATION (Megawatthours)"}, }},
8.     {$sort: {Coal: -1}}
9. )
10.
11. db.generation.aggregate(
12.     {$match : {"ENERGY SOURCE" : "Natural Gas"}},
13.     {$match : {"STATE" : {$not : {$eq:"US-TOTAL"}}}},
14.     {$match : {"STATE" : {$not : {$eq:"US-Total"}}}},
15.     {$group : {_id : "$STATE", Natural_Gas:{$sum:"$GENERATION (Megawatthours)"}, }},
16.     {$sort: {Natural_Gas: -1}}
17. )
18.
19. db.generation.aggregate(
20.     {$match : {"ENERGY SOURCE" : "Nuclear"}},
21.     {$match : {"STATE" : {$not : {$eq:"US-TOTAL"}}}},
22.     {$match : {"STATE" : {$not : {$eq:"US-Total"}}}},
23.     {$group : {_id : "$STATE", Nuclear:{$sum:"$GENERATION (Megawatthours)"}, }},
24.     {$sort: {Nuclear: -1}}
25. )
26.
27. db.generation.aggregate(
28.     {$match : {"ENERGY SOURCE" : "Other Gases"}},
29.     {$match : {"STATE" : {$not : {$eq:"US-TOTAL"}}}},
30.     {$match : {"STATE" : {$not : {$eq:"US-Total"}}}},
31.     {$group : {_id : "$STATE", Other_Gases:{$sum:"$GENERATION (Megawatthours)"}, }},
32.     {$sort: {Other_Gases: -1}}
33. )

```

```

34.
35. db.generation.aggregate(
36.     {$match : {"ENERGY SOURCE" : "Petroleum"}},
37.     {$match: {"STATE": {$not: {$eq:"US-TOTAL"}}}},
38.     {$match: {"STATE": {$not: {$eq:"US-Total"}}}},
39.     {$group : {_id : "$STATE", Petroleum:{$sum:"$GENERATION (Megawatthours)"}, }},
40.     {$sort: {Petroleum: -1}}
41. )
42.
43. db.generation.aggregate(
44.     {$match : {"ENERGY SOURCE" : "Wood and Wood Derived Fuels"}},
45.     {$match: {"STATE": {$not: {$eq:"US-TOTAL"}}}},
46.     {$match: {"STATE": {$not: {$eq:"US-Total"}}}},
47.     {$group : {_id : "$STATE", Wood_and_Wood_Derived_Fuels:{$sum:"$GENERATION (Megawatthours)"}, }},
48.     {$sort: {Wood_and_Wood_Derived_Fuels: -1}}
49. )

```

Listato 5. Codice all'interno dello script EnergyGenerationUnrenewableByState.js

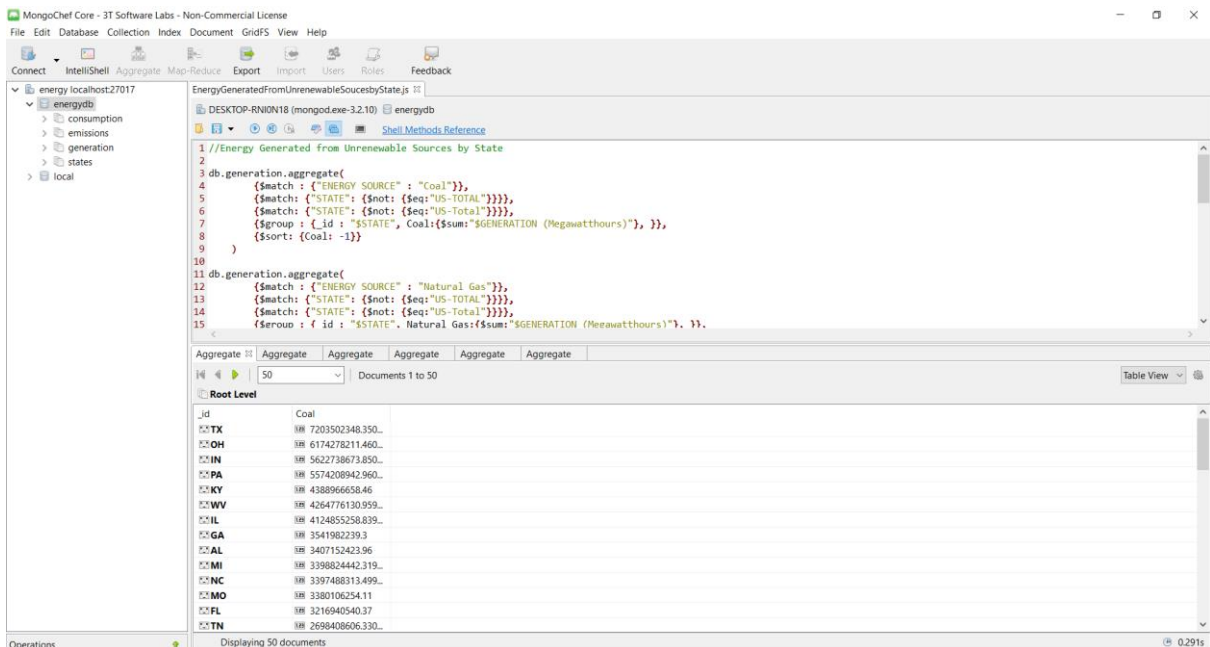


Figura 37. Risultato dell'esecuzione dello script EnergyGenerationUnrenewableByState.js in MongoChef.

```

1. //Energy Generated from Renewable Sources by State
2.
3. db.generation.aggregate(
4.     {$match : {"ENERGY SOURCE" : "Geothermal"}},
5.     {$match: {"STATE": {$not: {$eq:"US-TOTAL"}}}},
6.     {$match: {"STATE": {$not: {$eq:"US-Total"}}}},
7.     {$group : {_id : "$STATE", Geothermal:{$sum:"$GENERATION (Megawatthours)"}, }},
8.     {$sort: {Geothermal: -1}}
9. )
10.
11. db.generation.aggregate(
12.     {$match : {"ENERGY SOURCE" : "Hydroelectric Conventional"}},
13.     {$match: {"STATE": {$not: {$eq:"US-TOTAL"}}}},
14.     {$match: {"STATE": {$not: {$eq:"US-Total"}}}},
15.     {$group : {_id : "$STATE", Hydroelectric:{$sum:"$GENERATION (Megawatthours)"}, }},
16.     {$sort: {Hydroelectric: -1}}
17. )
18.

```

```

19. db.generation.aggregate(
20.     {$match : {"ENERGY SOURCE" : "Other Biomass"}},
21.     {$match: {"STATE": {$not: {$eq:"US-TOTAL"}}}},
22.     {$match: {"STATE": {$not: {$eq:"US-Tot1"}}}},
23.     {$group : {_id : "$STATE", Other_Biomass:{$sum:"$GENERATION (Megawatthours)"}, }},
24.     {$sort: {Other_Biomass: -1}}
25. )
26.
27. db.generation.aggregate(
28.     {$match : {"ENERGY SOURCE" : "Solar Thermal and Photovoltaic"}},
29.     {$match: {"STATE": {$not: {$eq:"US-TOTAL"}}}},
30.     {$match: {"STATE": {$not: {$eq:"US-Tot1"}}}},
31.     {$group : {_id : "$STATE", Photovoltaic:{$sum:"$GENERATION (Megawatthours)"}, }},
32.     {$sort: {Photovoltaic: -1}}
33. )
34.
35. db.generation.aggregate(
36.     {$match : {"ENERGY SOURCE" : "Wind"}},
37.     {$match: {"STATE": {$not: {$eq:"US-TOTAL"}}}},
38.     {$match: {"STATE": {$not: {$eq:"US-Tot1"}}}},
39.     {$group : {_id : "$STATE", Wind:{$sum:"$GENERATION (Megawatthours)"}, }},
40.     {$sort: {Wind: -1}}
41. )

```

Listato 6. Codice all'interno dello script EnergyGenerationRenewableByState.js

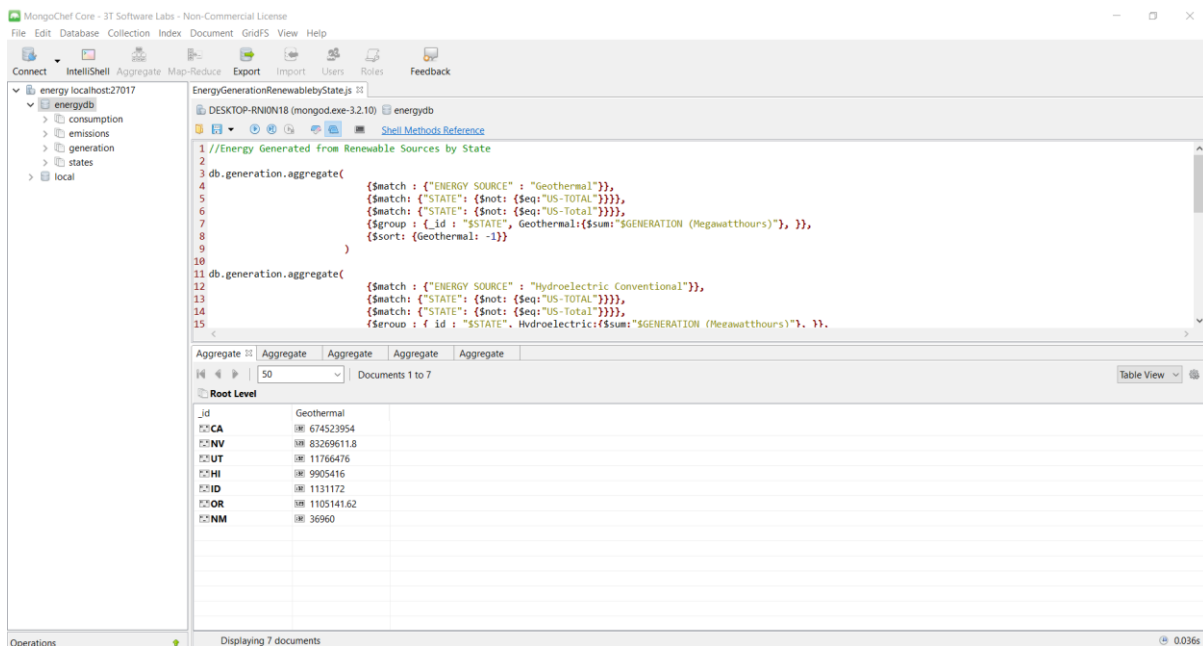


Figura 38. Risultato dell'esecuzione dello script EnergyGenerationRenewableByState.js in MongoChef.

Infine, sono stati realizzati gli script *TotalEmissionsByYear.js* (Listato 7) e *TotalEmissionsByState.js* (Listato 8); essi sono relativi alle emissioni, rispettivamente, prodotte dalle centrali elettriche negli anni dal 1990 al 2014 e suddivise per i vari Stati. L'esecuzione di tali script all'interno di MongoChef produce la creazione di un'unica Tab nella console IntelliShell, contenente i valori relativi a CO₂, SO₂ e NO_x, come è possibile vedere in Figura 39 ed in Figura 40, ordinati per quantità decrescente relativa alla CO₂.


```

1. //Total Emissions Produced by Year
2.
3. db.emissions.aggregate(
4.     {$match: {"State": {$not: {$eq:"US-TOTAL"}}}},
5.     {$group : {_id : "$Year", CO2:{$sum:"$CO2\n(Metric Tons)"}, SO2:{$sum:"$SO2\n(Metric Tons)"}, NOx:{$sum:"$N
   Ox\n(Metric Tons)"},}},
6.     {$sort: {CO2: -1}}
7. )

```

Listato 7. Codice all'interno dello script TotalEmissionsByYear.js

The screenshot shows the MongoChef interface with the following data displayed in the 'Aggregate' window:

Root Level	CO2	SO2	NOx
2007	1018812944	36166660	14599922
2005	1017532652	41358002	15844310
2006	9955671004	38094124	15197710
2004	9947926232	41235070	16572096
2008	9936048444	31319192	13320892
2000	9883337780	47615104	22524448
2003	9780377200	42583082	18126880
2002	9695852360	43525580	20774524
2001	9674426892	44697314	21159456
2010	9554383660	21601800	9964096
1999	9465209184	51373350	23821888
1998	9406398068	53857756	25834626
2011	9148282920	19379960	9625644
2009	9078030512	23881296	9581800

Figura 39. Risultato dell'esecuzione dello script TotalEmissionsByYear.js in MongoChef.

```

1. //Total Emissions Produced by State
2.
3. db.emissions.aggregate(
4.     {$match: {"State": {$not: {$eq:"US-TOTAL"}}}},
5.     {$group : {_id : "$State", CO2:{$sum:"$CO2\n(Metric Tons)"}, SO2:{$sum:"$SO2\n(Metric Tons)"}, NOx:{$sum:"$
   NOx\n(Metric Tons)"},}},
6.     {$sort: {CO2: -1}}
7. )

```

Listato 8. Codice all'interno dello script TotalEmissionsByState.js

The screenshot shows the MongoChef interface with a script named 'TotalEmissionsByState.js' being executed. The script uses the \$group operator to aggregate emissions data by state. The results are shown in a table view below the script editor.

```

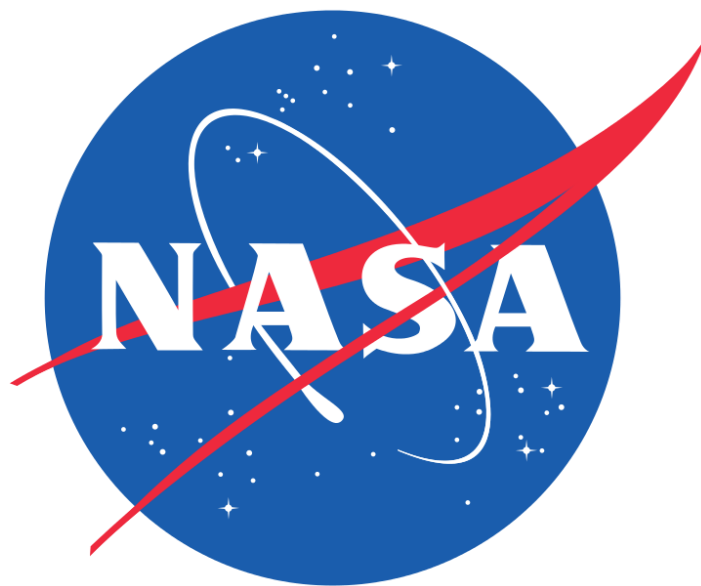
1 //Total Emissions Produced by State
2
3 db.emissions.aggregate(
4   { $match: { "state": { $not: { $eq: "US-TOTAL" } } } },
5   { $group : { _id : "$State", CO2:{$sum:"$CO2\n(Metric Tons)"}, SO2:{$sum:"$SO2\n(Metric Tons)"}, NOx:{$sum:"$NOx\n(Metric Tons)"} } },
6   { $sort: { CO2: -1 } }
7 )
8

```

State	CO2	SO2	NOx
TX	24559675008	55315826	39319343
OH	11900941776	113686194	32355844
PA	11667091264	86750773	24030964
FL	11607948396	46465345	23781576
IN	11174657200	76187614	27958670
IL	8925297292	49772577	19936149
KY	8427359096	49030628	20707504
WV	7773842316	50712280	18933041
GA	7503824120	51684868	15593925
MI	7468093812	34110839	17146203
AL	7376461888	42150365	15637819
MO	6779008316	31050926	14483740
NC	6637174284	34940902	15092357
CA	5716902500	5401521	10460433

Figura 40. Risultato dell'esecuzione dello script TotalEmissionsByState.js in MongoChef.

Valutazione di tutte le richieste di accesso HTTP al NASA Kennedy Space Center WWW server in Florida del mese di luglio del 1995



National Aereounatics and Space Administration

Universo di riferimento ed obiettivi dell'analisi

La **National Aeronautics and Space Administration**, istituita nel 1958 dal presidente degli Stati Uniti Dwight D. Eisenhower, ben nota con il suo acronimo **NASA**, è l'agenzia governativa responsabile del programma spaziale degli Stati Uniti e per la ricerca aereospaziale civile e militare. Il quartier generale è situato a Washington D.C. e il suo motto è: "**Ad astra per aspera**" che significa letteralmente "*verso le stelle vincendo le difficoltà*".

La fondazione della NASA avvenne a seguito del lancio del primo satellite artificiale da parte dell'Unione Sovietica (Sputnik 1) il 4 ottobre 1957. I membri del Congresso degli Stati Uniti, allarmati dal possibile pericolo per la sicurezza nazionale e per la possibile perdita della leadership tecnologica, chiesero al presidente Eisenhower un'azione immediata. Dopo alcuni mesi si decise, pertanto, la creazione di una nuova agenzia federale civile per le attività spaziali che rilevasse le attività della vecchia agenzia aeronautica NACA (National Advisory Committee for Aeronautics).

Negli ultimi anni l'attività della NASA si è focalizzata principalmente sulle esplorazioni di Marte, proposta frequentemente e descritta con dovizia di dettagli dai mass-media; tuttavia, la NASA ha rappresentato molto di più di ciò che viene pubblicizzato ed i suoi programmi sono cambiati nel tempo.

A partire dalla sua fondazione, è stata responsabile della maggior parte delle scoperte in campo spaziale, come, ad esempio, il programma **Mercury** per verificare la possibilità del volo spaziale, **Gemini** per realizzare i primi rendez-vous e docking nello spazio, **Skylab** per la messa in opera della prima stazione spaziale americana che si disintegrò cadendo nell'atmosfera, **Apollo** per l'esplorazione lunare e quella più recente dello **Space Shuttle** con le sue missioni esplorative.

La nostra attività di analisi, tramite il tool **Splunk**, si è andata ad intrecciare con il programma Space Shuttle.

In particolare, abbiamo voluto valutare i log contenenti tutte le richieste di accesso HTTP pervenute al NASA Kennedy Space Center WWW Server in Florida nel luglio del 1995.

La volontà di approfondire gli accessi in quel particolare periodo storico è duplice: da una parte la quasi contemporanea nascita del web e dall'altra una serie di importanti missioni spaziali avvenute sempre in quello stesso arco temporale.

Infatti, il 27 giugno del 1995 ha avuto luogo il lancio dello Shuttle relativo alla missione STS-71, durante la quale si ebbe il primo aggancio con la stazione spaziale sovietica Mir, esattamente 2 giorni dopo; oltre a questa, il 13 luglio 1995 ha avuto luogo la missione STS-70, relativa al lancio ed alla messa in opera di un TDRS-G (Tracking and Data Relay Satellite-G) all'interno dell'orbita geostazionaria, adoperato per le reti di tracciamento avanzate e le comunicazioni satellitari.

Gli **obiettivi** dell'analisi che ci siamo prefissati di raggiungere sono i seguenti:

- Individuare il giorno di picco degli accessi al server per valutarne le cause;
- Quantificare la mole di dati scaricati durante il giorno in cui si è rilevato il picco;
- Studiare da chi sono pervenute le richieste HTTP più numerose;
- Valutare come il mondo ha reagito alle informazioni trasmesse dalla NASA nel periodo di riferimento, analizzando in particolare quali sono state le risorse multimediali e le pagine più richieste.

Sorgente dei dati

I dati per l'analisi effettuata (in formato .zip contenente i log) sono stati reperiti presso il sito Web governativo americano ITA (Internet Traffic Archive), una repository responsabile di supportare un ampio accesso alle tracce del traffico della rete Internet, sponsorizzata dall'ACM SIGCOMM, un forum professionale per le discussioni circa le reti di computer e di comunicazione.

Come è possibile vedere in Figura 41, i dati presentano i seguenti campi:

- *Host*, il quale rappresenta l'indirizzo IP oppure il dominio della macchina che ha inviato la richiesta HTTP.
- *Timestamp*, il quale contiene la data e l'ora in cui è stata effettuata la richiesta HTTP.
- *Request*, contenente il corpo della richiesta HTTP.
- *HTTP Reply Code*, contenente il codice della risposta HTTP in merito alla richiesta effettuata.
- *Bytes in the reply*, contenente un valore numerico che rappresenta la dimensione in byte della risposta ad una determinata richiesta.

NASA-HTTP

Description

These two traces contain two month's worth of all HTTP requests to the NASA Kennedy Space Center WWW server in Florida.

Format

The logs are an ASCII file with one line per request, with the following columns:

1. **host** making the request. A hostname when possible, otherwise the Internet address if the name could not be looked up.
2. **timestamp** in the format "DAY MON DD HH:MM:SS YYYY", where **DAY** is the day of the week, **MON** is the name of the month, **DD** is the day of the month, **HH:MM:SS** is the time of day using a 24-hour clock, and **YYYY** is the year. The timezone is -0400.
3. **request** given in quotes.
4. **HTTP reply code**.
5. **bytes in the reply**.

Measurement

The first log was collected from 00:00:00 July 1, 1995 through 23:59:59 July 31, 1995, a total of 31 days. The second log was collected from 00:00:00 August 1, 1995 through 23:59:59 August 31, 1995, a total of 7 days. In this two week period there were 3,461,612 requests. Timestamps have 1 second resolution. Note that from 01/Aug/1995:14:52:01 until 03/Aug/1995:04:36:13 there are no accesses recorded, as the Web server was shut down, due to Hurricane Erin.

Privacy

The logs fully preserve the originating host and HTTP request. Please do not however attempt any analysis beyond general traffic patterns.

Acknowledgements

The logs were collected by Jim Dumoulin of the Kennedy Space Center, and contributed by Martin Arlitt (mfa126@cs.usask.ca) and Carey Williamson (carey@cs.usask.ca) of the University of Saskatchewan.

Publications

This is one of six data sets analyzed in an upcoming paper by M. Arlitt and C. Williamson, entitled "Web Server Workload Characterization: The Search for Invariants", to appear in the proceedings of the 1996 ACM SIGMETRICS Conference on the Measurement and Modeling of Computer Systems, Philadelphia, PA, May 23-26, 1996. An [extended version](#) of this paper is available on-line; see also the [DISCUS home page](#) and the group's [publications](#).

Related

Permission has been granted to make four of the six data sets discussed in "Web Server Workload Characterization: The Search for Invariants" available. The four data sets are: [Calgary-HTTP](#), [ClarkNet-HTTP](#), [NASA-HTTP](#), and [Saskatchewan-HTTP](#).

Figura 41. Portale ITA <http://ita.ee.lbl.gov/html/contrib/NASA-HTTP.html>

Ambiente di lavoro Splunk e risultati ottenuti

Dopo aver installato il tool, per prima cosa è stato necessario inserire al suo interno i dati precedentemente scaricati, tramite la procedura guidata della Web app di Splunk; una volta caricati, essi si presentavano nella forma riportata in Figura 42.

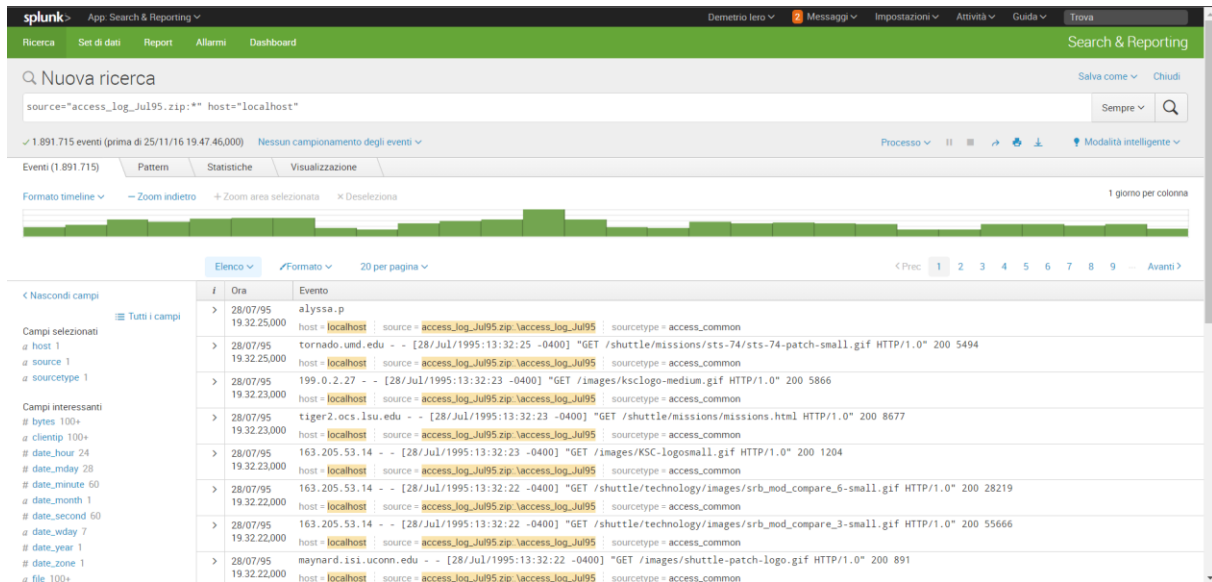


Figura 42. Panoramica relativa ai dati di log subito dopo l'inserimento all'interno di Splunk.

Per poter effettuare le analisi proposte, essendo i campi generati di default suddivisi in maniera grossolana, è stato necessario effettuare un'operazione di *parsing* dei dati, al fine di poter tirare fuori dei campi più specifici. Alla fine di tale procedimento, come riportato in Figura 43, sono stati estratti 7 campi:

- `_ora`, contenente il timestamp della richiesta.
- `IP_client`, contenente l'indirizzo IP o il dominio della macchina che ha effettuato la richiesta.
- `byte_of_response`, contenente la dimensione in byte della risposta alla richiesta.
- `method`, il quale specifica il metodo usato dalla richiesta HTTP.
- `protocol_version`, il quale indica la versione del protocollo HTTP utilizzata.
- `http_reply_code`, contenente il codice della risposta relativo ad una richiesta specifica.
- `resource`, contenente l'URI della risorsa richiesta.

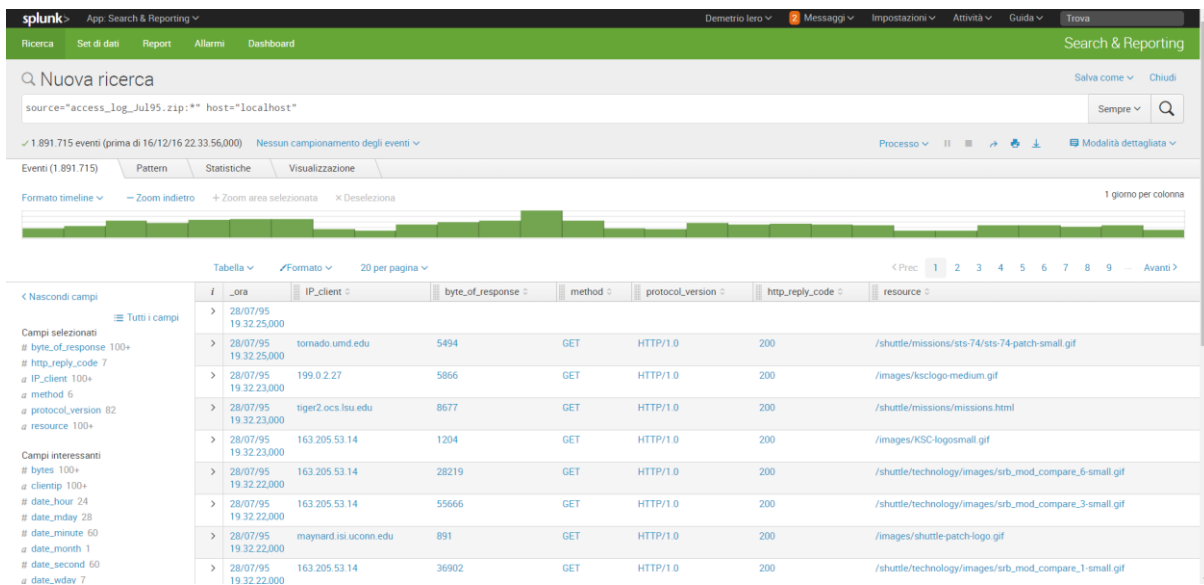


Figura 43. Panoramica relativa ai dati di log dopo aver effettuato il parsing dei campi.

Dopo aver effettuato l'estrazione dei campi desiderati, attraverso l'app di base fornita da Splunk "Search and Reporting" è stato possibile realizzare delle dashboard, utili per trarre conoscenza dai dati in esame. Non è stato possibile utilizzare altri tipi di app presenti all'interno dello store di Splunk, in quanto i dati a nostra disposizione non erano compatibili con nessuna di esse.

Innanzitutto, è stato realizzato un grafico a linee per individuare il giorno in cui sono state effettuate più richieste; quest'ultimo è risultato essere il 13 luglio, come mostrato in Figura 44, giorno in cui vi è stato il lancio dello Shuttle per la missione STS-70.

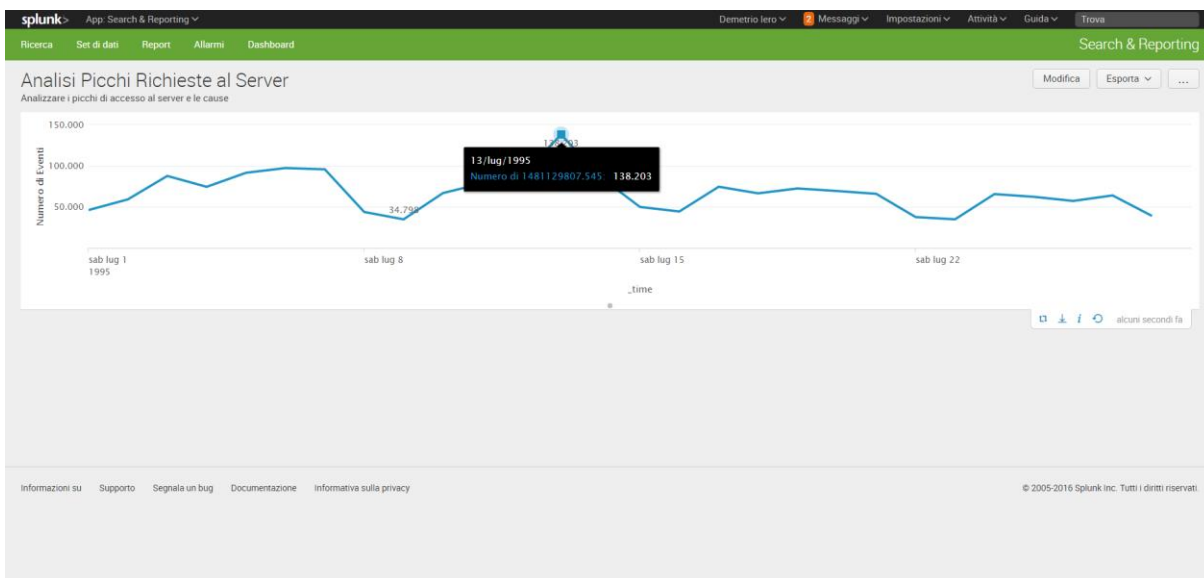


Figura 44. Dashboard relativa alle richieste HTTP effettuate nel luglio del 1995.

Una volta individuato il giorno in cui si è riscontrato il picco di richieste, abbiamo proseguito con il calcolare il totale dei byte inviati ai richiedenti nell'arco della giornata, per stimare nell'intorno di quali momenti della giornata si è registrato un maggiore afflusso di attività. Come è possibile evincere dalla Figura 45, le fasce orarie in cui si è registrata una maggiore attività sono risultate essere quella pomeridiana e quella serale.

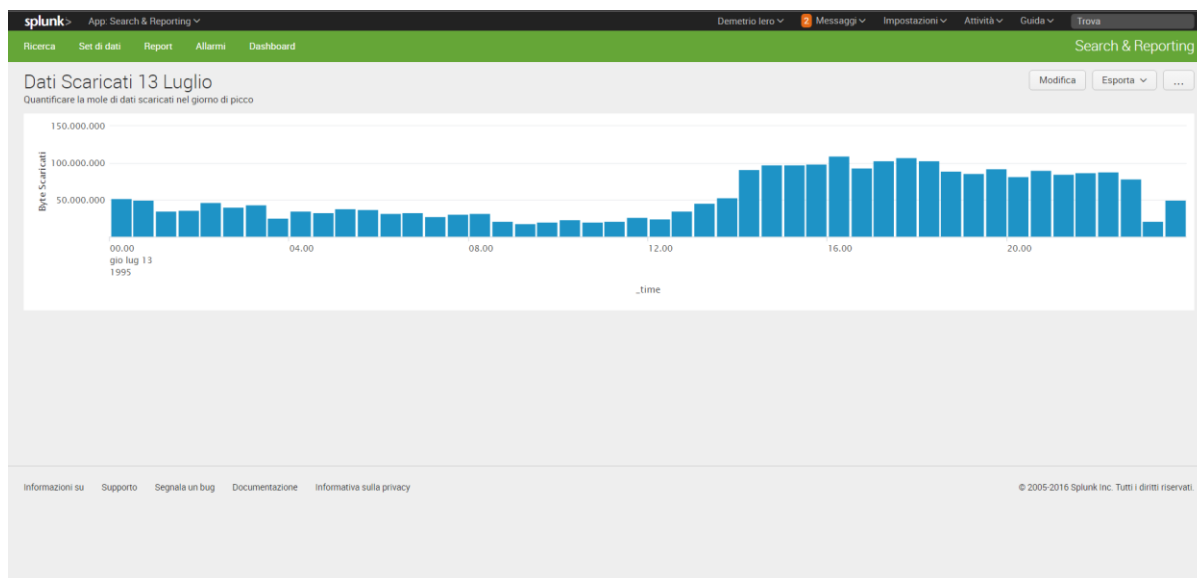


Figura 45. Totale dei byte inviati dal server della NASA agli host richiedenti il giorno 13 luglio 1995.

Successivamente, ci siamo occupati di analizzare quante richieste effettuate il 13 luglio provenissero da host anonimi e quante, invece, provenissero da host aventi un nome di dominio ben preciso. Le richieste singole, considerate poco rilevanti ai fini dell'analisi, sono state raggruppate all'interno della categoria *other*, come è possibile vedere in Figura 46.

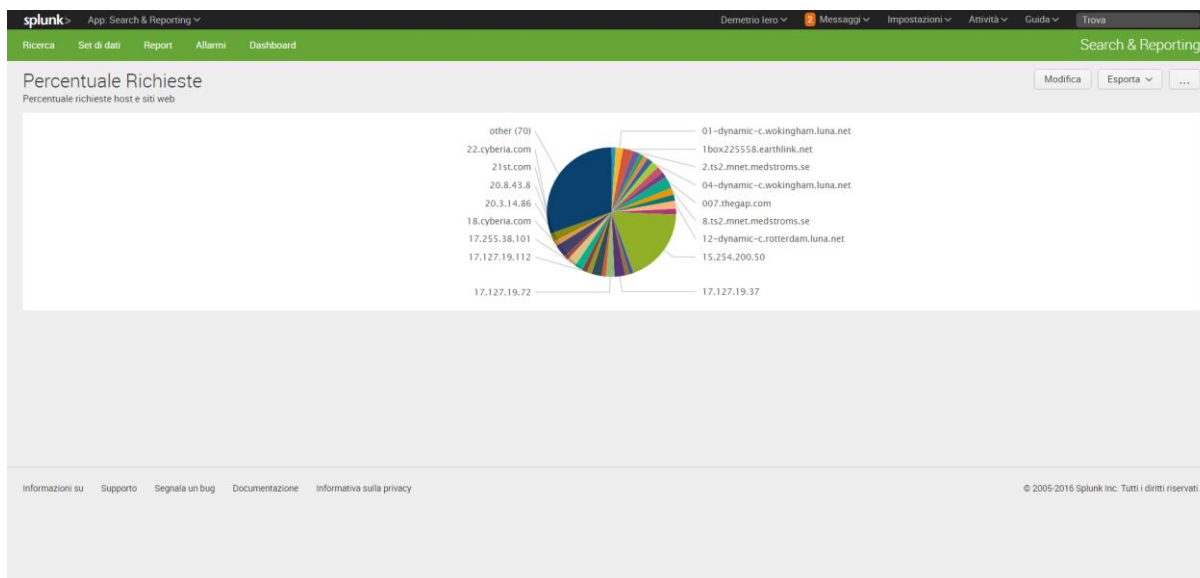


Figura 46. Dashboard relativa alla percentuale di richieste effettuate dai vari host il 13 luglio 1995.

In seguito, per curiosità indotta da quest'ultima analisi, siamo andati a ricercare tramite il sito Web <https://who.is/> a cosa corrispondessero gli IP dei 3 host anonimi che hanno effettuato la maggior parte delle richieste HTTP. In particolare, abbiamo scoperto che l'indirizzo IP **15.254.200.20** (Figura 47) appartiene ad un server della Hewlett-Packard Company (HP) situato a Palo Alto (CA) ed ha effettuato ben 286 richieste; l'indirizzo IP **28.8.43.8** (Figura 48) appartiene, invece, ad un server della Computer Science Corporation (CSC-48) nella città di Falls Church (VA) ed ha effettuato 50 richieste. Infine, con 41 richieste effettuate, abbiamo evidenziato l'IP **17.127.19.37** (Figura 49) il quale appartiene ad un server della Apple Inc. situato a Cupertino (CA).

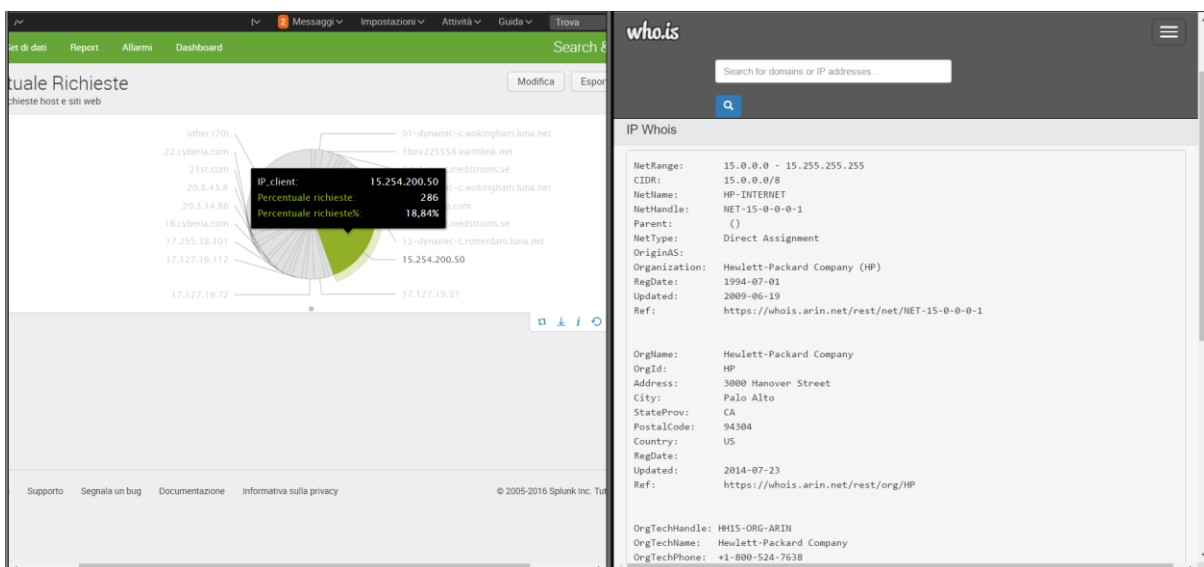


Figura 47. Richieste effettuate dal server della HP.

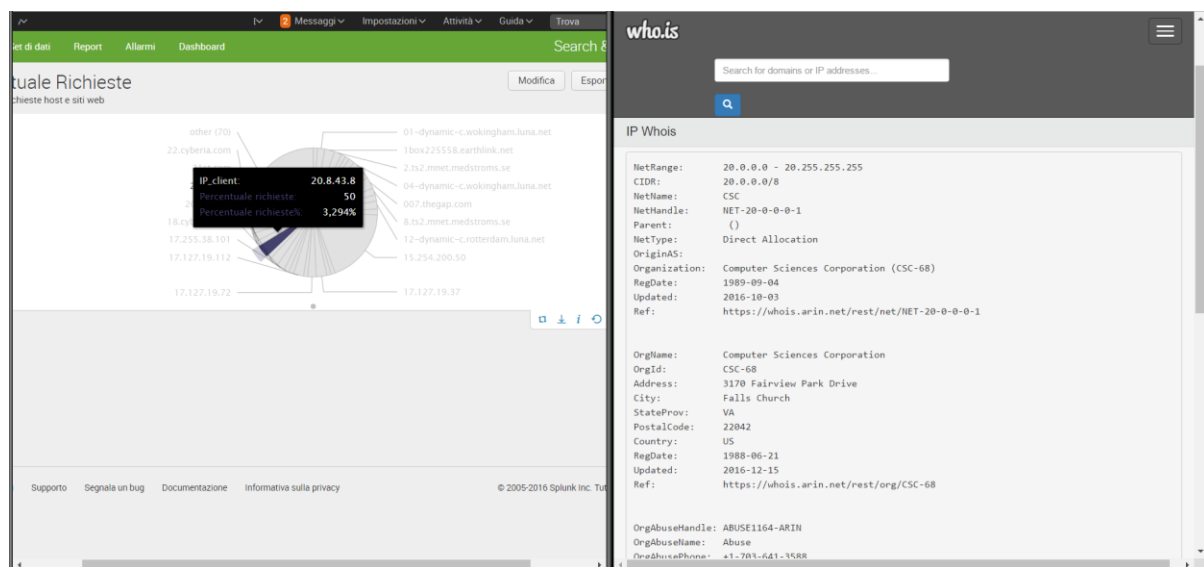


Figura 48. Richieste effettuate dal server della Computer Science Corporation.

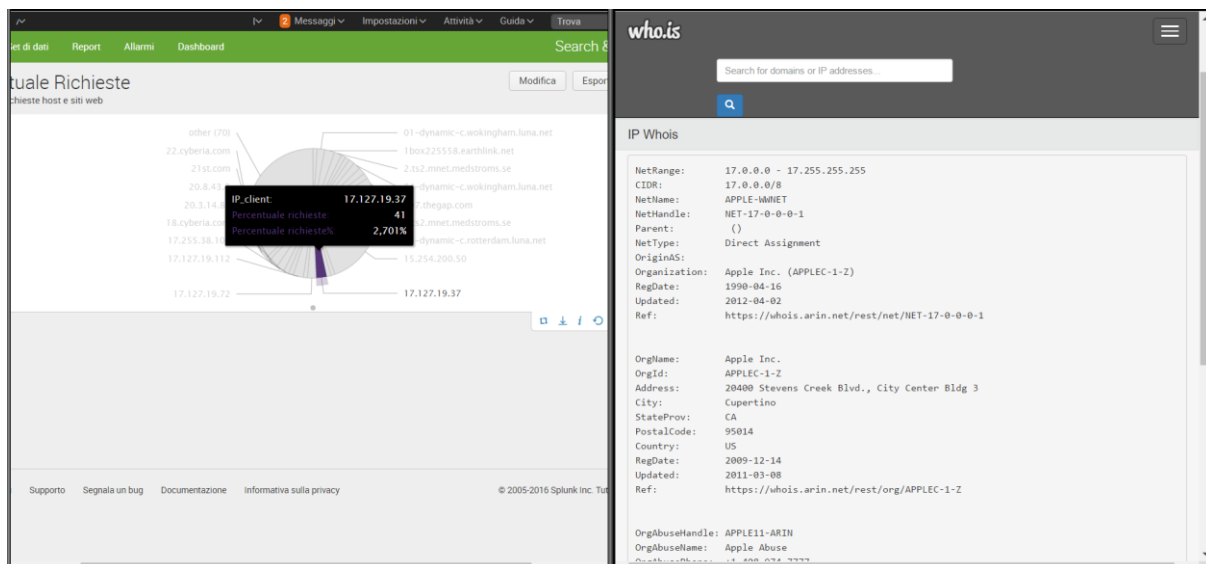


Figura 49. Richieste effettuate dal server della Apple Inc.

L'ultima analisi effettuata attraverso Splunk, riguardante i contenuti richiesti il 13 luglio 1995 dai vari host, è mostrata in Figura 50. In particolare, la dashboard si compone di 3 report, uno riguardante le risorse video, uno riguardante le immagini e, infine, uno riguardante le pagine Web che sono state maggiormente richieste. Anche in questo caso, i risultati poco rilevanti, che hanno ottenuto una percentuale bassa di richieste da parte degli utenti, sono stati raggruppati all'interno della categoria *other*.

Da questa analisi è emerso che, oltre ai contenuti della missione all'ordine del giorno, sono stati richiesti anche file multimediali relativi a missioni precedenti alla STS-70 e pagine Web contenenti approfondimenti su queste ultime. Ciò ci fa concludere che, comunque, il lancio della missione STS-70 è riuscito ad incuriosire gli utenti e ad invogliarli nell'apprendimento di qualcosa in più sulla storia della NASA, trascendendo il superficiale intento prettamente divulgativo e documentale relativo alla missione attuale, ipotizzato, in prima battuta, a monte dell'analisi effettuata.

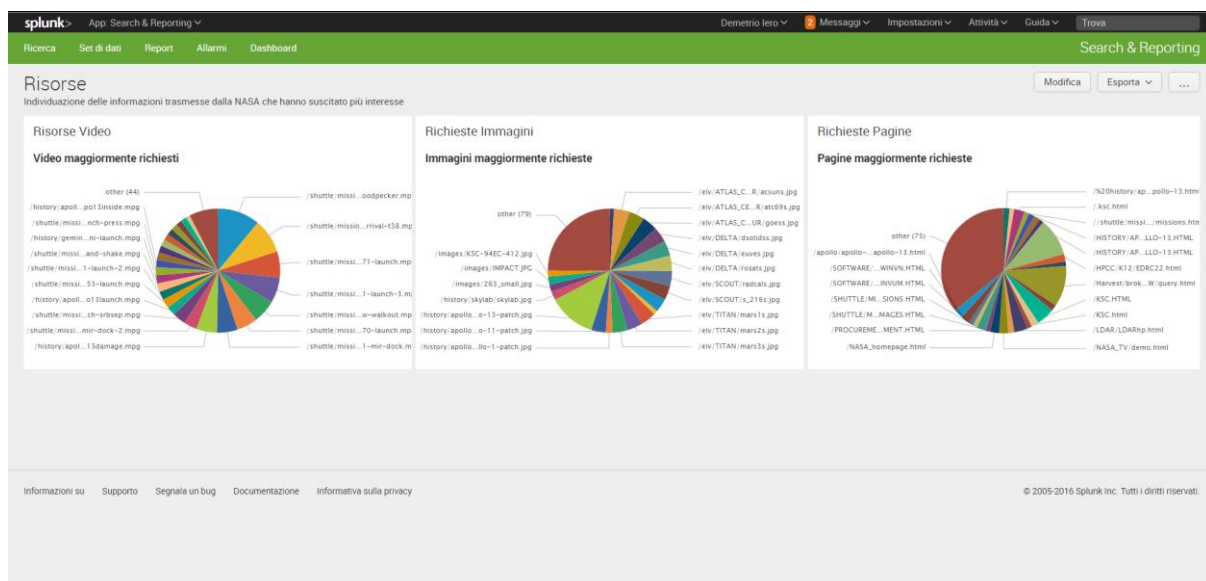


Figura 50. Dashboard relativa alle risorse multimediali e Web richieste il 13 luglio 1995.

Conclusioni e valutazione dei software utilizzati



Qlik Sense

Alla fine della nostra esperienza, è risultata subito chiara ed unanime la decisione secondo cui questo software sia perfetto per ottenere dashboard interattive molto curate, sia sotto l'aspetto dei contenuti che sotto l'aspetto grafico. Non richiede particolari approfondimenti preliminari all'utilizzo e fornisce molte operazioni guidate sia per l'inserimento dei dati all'interno dell'applicazione, sia per la creazione dei vari grafici. Inoltre, qualora si volesse ottenere qualcosa di più particolare, come ad esempio le gerarchie temporali e spaziali della nostra analisi, necessarie per la componente interattiva dei grafici, è possibile, per i meno esperti, consultare le guide online messe a disposizione sul sito del produttore oppure dalla community. Inoltre, è un'applicazione che lavora *in memory*, utilizzando prevalentemente la RAM, aumentando così di gran lunga l'efficienza di calcolo dei risultati, indispensabile per qualsiasi tool di Big Data Analytics. Rimane, comunque, una realtà standalone commerciale (gratuita per gli studenti); non è possibile, pertanto, integrarla con software di terze parti tramite librerie.

- Usabilità: ★★★★★
- Efficienza nell'uso delle risorse: ★★★★★
- Resa grafica dei risultati: ★★★★★
- Personalizzazione: ★★★★★
- Integrazione con altri software: ★★★★★
- Documentazione di supporto: ★★★★★
- Valutazione dell'esperienza complessiva: ★★★★★

Tableau

Rispetto al suo diretto rivale commerciale Qlik Sense, è risultato essere un po' meno immediato da utilizzare, soprattutto nelle fasi iniziali, il che lo rende maggiormente indicato per utilizzatori aventi abilità tecniche maggiori. Si sono rivelate molto utili le video guide proposte all'interno del sito dell'azienda produttrice, che contribuiscono a rendere la curva di apprendimento decisamente meno ripida, sebbene per la maggior parte delle funzionalità principali siano rese immediate nell'utilizzo dall'ambiente di lavoro, a nostro avviso realizzato egregiamente. Una piccola pecca è relativa all'efficienza, in quanto, se ci si trova a lavorare con molte tabelle in join tra loro, i tempi di attesa dei risultati si sono rivelati essere nettamente superiori a quelli riscontrati in Qlik Sense. L'interattività con i grafici è quasi assente; infatti, effettuando un'operazione di drill-down, ad esempio, essa comporterà un'estensione statica del grafico, il che lo rende inferiore a Qlik Sense. Tuttavia, è possibile utilizzare dati di posizione generati in real-time, integrabili con eventuali altre coordinate geografiche presenti nei dati in analisi, finalizzate alla creazione di mappe geografiche e di calore estremamente ben fatte e di impatto grafico notevole. Anche in questo caso, l'integrazione con altri software risulta assente, ponendo limiti alle varietà di utilizzo applicativo.

- Usabilità: ★★★★★
- Efficienza nell'uso delle risorse: ★★★★★
- Resa grafica dei risultati: ★★★★★
- Personalizzazione: ★★★★★
- Integrazione con altri software: ★★★★★
- Documentazione di supporto: ★★★★★
- Valutazione dell'esperienza complessiva: ★★★★★

MongoDB

Una premessa risulta necessaria prima di dare una valutazione ingenerosa a questo software: siamo di fronte ad un'applicazione di gestione di una base di dati NoSQL di tipo documentale e non ad un software dedicato ad analisi di tipo descrittivo e diagnostico. La creazione del database e la gestione delle collezioni al suo interno tramite riga di comando, un'utility, MongoChef, dall'interfaccia grafica molto minimalista (che fa un po' il verso ad un IDE Javascript ma con una pessima gestione delle attività di debug sul codice) ed un query language limitato, sicuramente non portano l'usabilità complessiva a livelli molto alti. In aggiunta a ciò, la documentazione e le varie guide sparse per il Web non facilitano di molto la vita dei neofiti che, come noi, si avvicinano per la prima volta ad un database NoSQL, con l'intenzione di effettuare analisi su di esso. Tuttavia, seppur lavori soltanto con poche tipologie di dati, risulta essere ben ottimizzato nel consumo di risorse sia per le attività di ETL che per le operazioni effettuate sulle varie collezioni. Un altro punto a favore di MongoDB è che, essendo un software open source, sono presenti in rete diverse librerie che lo rendono personalizzabile ed integrabile con i principali linguaggi di programmazione, accrescendone le possibilità di utilizzo sia nell'ambito della gestione di database documentali e sia nelle attività di Big Data Analytics.

- Usabilità: ★☆☆☆☆
- Efficienza nell'uso delle risorse: ★★★★★
- Resa grafica dei risultati: ★☆☆☆☆
- Personalizzazione: ★★★★★
- Integrazione con altri software: ★★★★★
- Documentazione di supporto: ★★★★★
- Valutazione dell'esperienza complessiva: ★★★★★

Splunk

Secondo la nostra opinione, a valle dell'esperienza effettuata in questo corso, Splunk risulta sicuramente essere uno dei software più potenti per ciò che riguarda l'analisi dei dati di log nel campo della Cyber Security, partendo da streaming di dati provenienti da diverse sorgenti (ad esempio, diversi server sparsi per il mondo), fino ad arrivare a file di log generati da una singola macchina di cui si vogliono analizzare i tentativi di violazione di accesso. Malgrado non sia possibile integrarlo direttamente con altre applicazioni, Splunk presenta un'enorme scalabilità, grazie alla vasta gamma di app messe a disposizione da terzi all'interno dello store, la maggior parte delle quali accessibili gratuitamente. Se ciò, però, non dovesse risultare sufficiente, è possibile creare delle proprie app da installare manualmente all'interno di Splunk, il che lo rende altamente personalizzabile. A livello di reportistica, presenta un'ampia gamma di grafici e diagrammi da poter realizzare, dai classici diagrammi a torta o a barre, fino ad arrivare a mappe di calore interattive abbastanza elaborate. Tuttavia, ciò che lo rende difficile da utilizzare è una documentazione di supporto non molto curata, soprattutto per chi si avvicina per la prima volta a questo tipo di analisi. Inoltre, a livello didattico, la scarsa conoscenza preliminare del formato dei dati reperiti in rete per effettuare analisi di training, come quella mostrata in questa tesina, fa in modo che l'apprezzabile capacità di Splunk di poter elaborare qualsiasi tipo di file di log diventi un'arma a doppio taglio, soprattutto se non si conoscono bene le tecniche necessarie ad estrarre nuovi campi, come, ad esempio, il parsing tramite l'uso delle espressioni regolari.

Un'ultima pecca che riguarda questo software è relativa alla gestione dei dati caricati in esso, in quanto, se intercorrono periodi di inattività lunghi sui dati (dell'ordine di una settimana), essi vengono rimossi automaticamente ed è necessario ricaricarli (con tempi di attesa non banali se sono consistenti); se, inoltre, si volessero togliere dei dati inseriti erroneamente, la procedura per fare ciò risulta parecchio macchinosa.

- Usabilità: ★★★★★
- Efficienza nell'uso delle risorse: ★★★★★
- Resa grafica dei risultati: ★★★★★
- Personalizzazione: ★★★★★
- Integrazione con altri software: ★★★★★
- Documentazione di supporto: ★★★★★
- Valutazione dell'esperienza complessiva: ★★★★★