



# BIG DATA EN EL SISTEMA DATATUR- ESTADISTICAS TURISTICAS DE TURESPAÑA

15 noviembre 2016



Pablo Burgos Casado Coordinador áreas Informática  
S.G. Gestión Económico-Administrativa y TI de Turespaña





## 2. SISTEMA DATATUR

DATATUR es un Sistema de información de estadísticas turísticas que se compone de diferentes repositorios de información tanto en razón a su origen como al diferente grado de estructuración de la misma. Por citar algunos orígenes AENA (vuelos); AECFA (previsiones de vuelo), RENFE (trenes), Puertos del Estado (barcos), Dirección General de Tráfico (carretera), INE, Policía Nacional, Mº de Trabajo e Inmigración (datos de afiliación a la S.S.), OAG (vuelos). Esta información está a disposición, a través del portal WEB de TURESPAÑA, de todos los ciudadanos (en su zona pública) y de sus suscriptores (en su zona restringida).

Origen Ficheros	Proceso de Carga	Frecuencia de Carga	Nº Ficheros	Formato / Tamaño Ficheros Mb
INE	IPC - IPCA	Mensual	14	Csv / 5
	Índice Costes Laborales	Trimestral	6	Xls / 2
	IPH, ADR y REVPAR	Mensual	8	Xls / 3
	Encuestas de Ocupación (Hotelería, Extrahotelería)	3 x Mensual	8	Xls / 5
	Encuesta de Servicio	Anual	-	-
	Oferta Turística	Anual	-	-
Puertos del Estado	EPA	Trimestral	1	Zip / 35
	Administrativos Puertos	Mensual	8	Xls / 430
Renfe	Administrativos Cruceros	Mensual	2	Xls /
	Administrativos Trenes	Mensual	2	Xls / 103
Turespaña	Frontur Muestral Campo (Aforo y Encuesta)	Mensual	7	Zip / 609
	Frontur Muestral Depurado (Aforo y Encuesta)	2 x Mensual	6	Zip / 8
Mº Empleo y Seguridad Social	Frontur Elevado	2 x Mensual (Año anterior)	5	Varios / 500
	DGT Carreteras Depurado	Mensual	1	Xml / 2,5
	Directorio Hoteles Turespaña	Anual	-	Xml / 5 Gb
	Egatur Muestral	Mensual	200	-
AECFA	Directorio Hoteles Turespaña	Anual	-	-
	Egatur Abreviado, Extendido, Elevado	2 x Mensual (Año Anterior)	3	Xls / 22
AENA	Afilación Seg. Social	Mensual	10	Varios / 10
	Encuesta Coyuntura Laboral	Anual	-	-
Banco de España	Administrativos DGT (No Depurado)	Mensual	73	Xls / 8
C.G.E.F. Secr. General - Estadística	Conoper	Mensual	1	Txt / 5
	Geslot	Quincenal	38	Txt / 50
AECFA	Balanza de Pagos	Mensual	2	Csv / 2
	Policía	Anual	5	Xls / 5



# 2. SISTEMA DATATUR: ENCUESTAS

DATATUR incluye la encuesta de movimientos turísticos en fronteras (Frontur) que recoge datos relativos a la entrada en España de visitantes no residentes en España, la encuesta de gasto turístico (Egatur) que recoge datos relativos al gasto que realizan en España los visitantes no residentes en España, (Familitur) que recoge datos relativos a los viajes que realizan los residentes en España.

The screenshot shows the main interface of the DATATUR website. At the top, there are navigation links for 'Inicio', 'Mapa Web', 'Contacto', 'Otras Webs', and 'Glosario'. Below this is a search bar with the text 'Término a buscar' and a 'Buscar' button. The main content area is divided into several sections:

- Estadísticas:** A sidebar menu with options like 'Ficha de Coyuntura', 'Frontur', 'Informes Dinámicos', 'Series', 'Metodología', 'Microdatos', 'Egatur', 'Familitur', 'Análisis Turístico', and 'Otras Estadísticas'.
- FRONTUR:** A section describing the survey of tourist movements at borders, including a list of monthly reports from 2014 to 2015.
- NOVEDADES:** A section for news, listing several publications from 2016 regarding tourism companies and travel agencies.
- Publicaciones Anuales (\*):** A section for annual reports, mentioning the joint publication of Frontur and Egatur data from 2009 onwards.

This block contains two screenshots of specific report pages from the DATATUR system.

The top screenshot is for the **EGATUR** (Tourist Expenditure) report. It shows a navigation menu on the left and a main content area with the following elements:

- Información Mensual:** A section for monthly reports, with a dropdown for 'Mes' (Month) and 'Año' (Year) set to 2015, and a 'Ver Informe' button.
- Últimos informes, Septiembre - 2015:** A list of report topics such as 'Gasto de los turistas según destino principal', 'según país de residencia', 'según vías de acceso', 'según tipo de alojamiento', 'según forma de organización', and 'según motivo del viaje'.

The bottom screenshot is for the **FAMILITUR** (Family Tourism) report. It follows a similar layout to the EGATUR page:

- Información Mensual:** Similar navigation for monthly reports.
- Últimos informes, Enero - 2015:** A list of report topics such as 'Número de viajes por CCAA de residencia', 'por destinos interno desagregado y emisor', 'por motivo principal', 'según tipo de transporte', and 'por tipo de alojamiento'.



## 2. SISTEMA DATATUR: ANALISIS TURISTICO

En esta sección pueden consultarse :

- **Balantur:** Balance del año turístico en el que se recoge el análisis de la actividad turística en España utilizando para ello distintas fuentes estadísticas oficiales. Se estudian la demanda y la oferta turística, la posición competitiva en el contexto internacional y los ingresos por turismo.
- **Coyuntura:** Boletín trimestral de coyuntura turística con el análisis de los principales indicadores de la actividad turística
- **España en Europa:** Esta publicación recoge una comparativa de los datos más importantes disponibles de los países de la UE sobre su turismo nacional y sus condicionantes principales (población, renta, precios).

### COYUNTURA DE MERCADOS COMPETIDORES

#### 19. Llegadas de turistas no residentes (en miles)

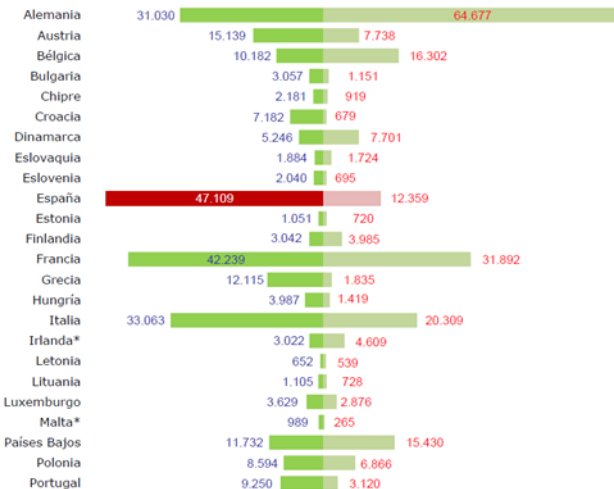
	España	Francia	Italia	Portugal	Grecia	Croacia	Chipre	Tónex	Marruecos	Turquia	Egipto	EEUU	Méjico	Cuba	República Dominicana	China	
<b>AÑO 2014</b>																	
<b>MES</b>																	
Abril	5.262,1	-	3.988,0	838,2	728,0	522,6	181,0	525,8	855,0	2.573,1	859,9	6.484,4	2.328,9	287,2	445,0	-	
Mayo	6.094,4	-	4.568,0	1.023,8	1.651,7	876,7	293,2	605,4	760,3	3.863,9	768,2	6.298,3	2.386,9	195,9	385,9	-	
Junio	6.582,9	-	5.092,0	966,2	2.697,5	1.672,0	342,2	798,7	871,1	4.530,1	785,7	6.027,2	2.501,0	181,3	442,0	-	
Julio	8.296,4	-	6.045,0	1.040,0	4.222,9	2.744,7	382,0	245,0	1.495,8	4.952,4	885,8	7.342,7	2.635,2	211,6	514,6	-	
Agosto	9.070,0	-	5.996,0	1.202,9	4.858,4	3.378,6	373,1	1696,5	1.420,6	6.535,6	997,7	8.381,6	2.374,1	199,8	424,7	-	
Septiembre	7.023,2	-	5.224,0	1.085,4	3.643,7	1.337,1	316,6	646,0	757,6	5.496,0	883,9	6.125,3	1.799,8	151,9	282,3	-	
Octubre	5.894,3	-	4.262,0	957,4	1.840,6	539,8	251,5	488,0	394,6	4.293,3	1.002,9	6.264,9	2.101,4	187,4	323,5	-	
Noviembre	3.429,7	-	2.951,0	504,1	668,2	134,0	81,4	315,4	613,3	2.264,4	898,2	5.809,6	2.553,7	259,7	392,3	-	
Diciembre	3.263,7	-	2.555,0	394,8	537,7	117,8	56,9	350,6	759,5	1.740,5	781,6	6.213,0	3.305,3	332,2	534,3	-	
<b>AÑO 2015</b>																	
<b>MES</b>																	
Enero	3.169,8	-	2.837,0	379,8	606,1	77,8	41,8	340,1	604,8	1.762,0	677,5	5.453,5	2.571,3	370,8	483,0	-	
Febrero	3.282,3	-	2.454,0	429,1	509,2	93,4	50,7	283,1	614,3	1.564,9	640,2	4.708,9	2.471,9	361,7	478,8	-	
Marzo	4.156,2	-	3.265,0	622,3	613,1	193,8	97,5	400,7	727,0	2.017,6	-	5.870,7	2.843,4	404,4	537,6	-	
Abril	5.407,1	-	3.863,0	898,2	-	539,7	201,5	404,1	821,2	2.626,7	-	-	2.528,8	330,5	462,0	-	
Mayo	6.504,9	-	5.109,0	1.106,5	-	1.028,8	307,4	507,3	763,7	3.775,0	-	-	2.539,4	237,3	417,0	-	
Junio	6.672,9	-	5.494,0	1.058,7	-	1.719,2	337,0	534,6	781,2	4.349,7	-	-	2.736,9	218,6	480,4	-	
Julio	8.782,7	-	-	1.131,4	-	3.095,1	414,5	-	1.713,7	-	-	-	2.991,4	266,8	568,3	-	
Agosto	9.226,0	-	-	-	-	-	-	771,7	-	-	-	-	-	-	465,8	-	
Septiembre	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
<b>Porcentajes de variación</b>																	
<b>AÑO 2014</b>																	
<b>MES</b>																	
Abril	13,2	17,0	6,7	22,7	30,6	19,1	11,4	1,6	13,0	6,4	-21,9	15,5	25,3	4,8	14,5	2,3	
Mayo	5,7	-4,6	-2,6	13,8	12,6	-12	6,1	-2,9	5,0	3,9	-20,7	10,7	26,4	2,5	17,2	1,2	
Junio	4,5	0,7	3,8	11,1	13,7	12,1	11,0	-3,2	-4,0	9,7	-20,5	9,3	21,4	-1,4	11,7	-1,2	
Julio	5,9	-3,8	-0,2	7,9	29,4	1,4	5,7	-2,0	14,0	3,4	15,8	4,6	20,6	-0,4	9,4	0,3	
Agosto	8,8	2,7	-0,2	10,9	25,0	6,2	5,9	7,9	-7,0	11,9	76,7	7,1	14,1	6,7	15,8	1,1	
Septiembre	8,1	-	-2,0	10,2	23,0	1,0	-11,5	-11,2	-2,0	3,0	193,5	1,7	10,2	-2,9	13,2	-2,1	
Octubre	7,8	2,6	7,6	15,0	18,9	15,2	-6,1	-12,8	-7,6	0,0	79,5	6,1	22,9	11,5	14,0	-0,7	
Noviembre	2,9	-2,2	5,3	13,8	53,5	11,5	-0,1	-15,1	-1,4	1,3	33,5	7,7	25,6	10,9	9,3	1,3	
Diciembre	5,6	3,7	-0,6	14,4	33,5	22,3	3,7	-18,8	-3,9	8,9	15,3	4,8	30,6	8,6	9,0	8,9	
<b>AÑO 2015</b>																	
<b>MES</b>																	
Enero	3,6	1,8	3,6	13,6	49,9	28,5	2,8	-16,3	0,3	11,8	5,5	1,8	6,8	15,9	7,5	5,1	
Febrero	5,3	3,6	5,6	9,1	56,2	21,3	-33,1	-12,1	-1,8	2,7	3,8	1,6	9,2	14,4	7,2	-3,7	
Marzo	6,7	3,8	2,4	16,8	34,3	24,8	25,7	-4,4	0,0	2,6	10,6	0,2	4,0	12,5	7,3	4,2	
Abril	2,8	-0,7	-1,2	7,4	-	3,0	11,3	-23,2	-4,0	2,1	7,4	-	8,6	15,1	3,8	6,3	
Mayo	6,8	9,6	11,8	8,4	-	16,7	4,9	-16,2	0,5	-2,3	16,5	-	6,4	21,1	8,1	4,3	
Junio	1,4	-	-	7,9	10,1	-	2,8	-1,5	-33,0	-10,3	-4,0	4,4	-	9,4	20,6	8,7	1,8
Julio	6,3	-	-	8,9	-	12,8	6,5	-28,0	14,6	-	-	-	-	13,5	26,1	10,4	0,8
Agosto	1,6	-	-	-	-	-	-	-84,5	-	-	-	-	-	-	-	9,7	-
Septiembre	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

### BALANZA PAGOS INGRESOS Y PAGOS POR TURISMO

España -con 47.109 millones de euros- fue el país de la Unión que más ingresos por turismo obtuvo en el año 2013, seguido de Francia e Italia.

Los residentes en Alemania, con un total de 64.677 millones de euros, continuaron siendo los que más gastaron en turismo emisor, seguidos de los británicos con 39.574 millones de euros.

#### Ingresos y pagos por turismo (en millones de C)



Ingresos

Pagos

España

Ingresos

Pagos



## 2. SISTEMA DATATUR: OTRAS ESTADISTICAS

- Balanza de Pagos: (Banco de España) información sobre ingresos y pagos de turismo y viajes.
- Compañías de Bajo Coste (Datos de vuelos aportados por AENA).
- Empleo Turístico: a partir de datos obtenidos del Ministerio de Empleo y Seguridad Social (datos mensuales de afiliación a la Seguridad Social) y del INE (datos trimestrales de la EPA).
- Encuestas de Ocupación: (INE) datos nacionales, desglosados según CCAA y provincias, para Hoteles, Apartamentos Turísticos, “Campings” y Turismo Rural.
- Índices e indicadores INE: evolución de los precios de los productos y servicios relacionados con el turismo y la hostelería (IPC), de la evolución de precios específicos de los alojamientos hoteleros (IPH) y de la evolución de los ingresos registrados en los alojamientos hoteleros (IIH).

**Estadísticas** Inicio > Estadísticas > Otras Estadísticas > Empleo Turístico > Encuesta Población Activa > Trimestrales

**Trimestrales**

- Ficha de Coyuntura
- Frontur
- Egatur
- Familitur
- Análisis Turístico
- Otras Estadísticas
  - Balanza de Pagos
  - Compañía Bajo Coste
  - Cuenta Satélite
  - Empleo Turístico
  - Encuestas Ocupación
  - Habitur
  - Índices e indicadores
  - Movimiento Viajeros
  - Museo del Prado
  - Ocupatur

**Afilación Seguridad Social**

- Nota Informativa
- Anuales
- Mensuales
- Metodología
- Ficha
- Series

**Encuesta Población Activa**

- Nota Informativa
- Anuales
- Trimestrales
- Metodología
- Ficha
- Informes / Publicaciones

A continuación se muestran enlaces para la ejecución de informes en tiempo real sobre Población Activa en base trimestral:

**Población Activa. Información Trimestral**

Trimestre 3 Año 2015 [Ver Informe](#)

**Últimos informes. Trimestre 3 - 2015. Resultados 1 - 31.31.**

- Activos según edad por actividades de la industria turística.
- Activos según el nivel de formación por actividades de la industria turística.
- Activos según comunidad autónoma por actividades de la industria turística.
- Activos según nacionalidad por actividades de la industria turística.
- Activos según situación profesional por actividades de la industria turística.
- Ocupados según edad por actividades de la industria turística.

Calendario de disponibilidad: 21

**¡NUEVO!**  
FICHAS EJECUTIVAS  
MERCADOS EMISORES

Estadísticas turísticas

### BALANZA DE PAGOS

Ingresos por turismo (millones de euros).  
Año 2015

	Ingresos	Tasa de Variación Interanual	Diferencia
<b>Total</b>	<b>40.082</b>	<b>3,1</b>	<b>1.217</b>
Enero 2015	3.054	2,4	72
Febrero 2015	2.594	2,7	67
Marzo 2015	3.082	2,6	79
Abril 2015	3.244	2,6	81
Mayo 2015	4.261	5,0	204
Junio 2015	4.965	1,9	93
Julio 2015	6.418	5,3	322
Agosto 2015	6.851	3,0	199
Septiembre 2015	5.613	1,8	100

(\*) Fuente: Banco de España.

### ENCUESTA DE POBLACIÓN ACTIVA

Tabla 1. Datos absolutos

Año 2014. Datos definitivos.

	Total	16-29 Años	30-44 Años	45 Años y Más
<b>Total Actividades Industria Turística</b>	<b>2.623.707</b>	<b>618.419</b>	<b>1.127.744</b>	<b>877.544</b>
Hostelería	1.722.183	441.941	724.660	555.583
Servicios de alojamiento	404.795	77.062	180.184	147.551
Servicios de comidas y bebidas	1.317.389	364.879	544.475	408.033
Transporte de viajeros	280.844	24.978	115.817	140.049
Otras Actividades Industria Turística	620.681	151.500	287.268	181.911
<b>Total Actividades Sector Servicios</b>	<b>14.719.719</b>	<b>2.356.765</b>	<b>6.540.103</b>	<b>5.822.850</b>
<b>Total Actividades a Nivel Nacional</b>	<b>22.954.568</b>	<b>3.920.061</b>	<b>10.049.179</b>	<b>8.985.328</b>

CNAE 09

Los datos anuales de la EPA se obtienen como media de los cuatro trimestres del año.

Fuente: Turespaña. - Elaboración propia a partir de los datos de la Encuesta de Población Activa (INE).

Nota informativa: Los datos se han obtenido en base a la EPA que publicó el INE en abril de 2014, calculados con la nueva base de población que incorpora la información actualizada de los Censos de Población y Viviendas de 2011.

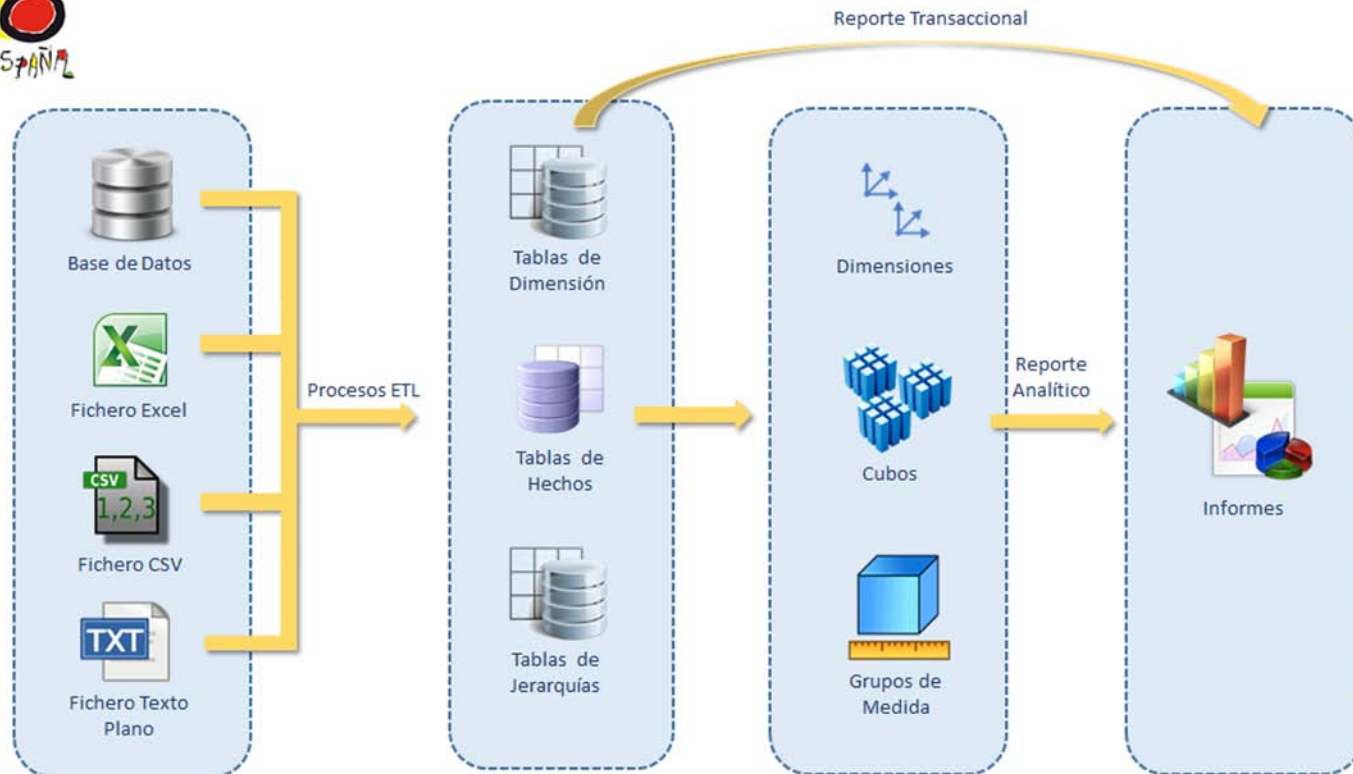
### 3. DATATUR Tecnología empleada

Se utiliza tecnología de Microsoft (base de datos Sql Server 2008R2)

El número de BBDD Relacionales que se alimentan en este proceso es de 12.

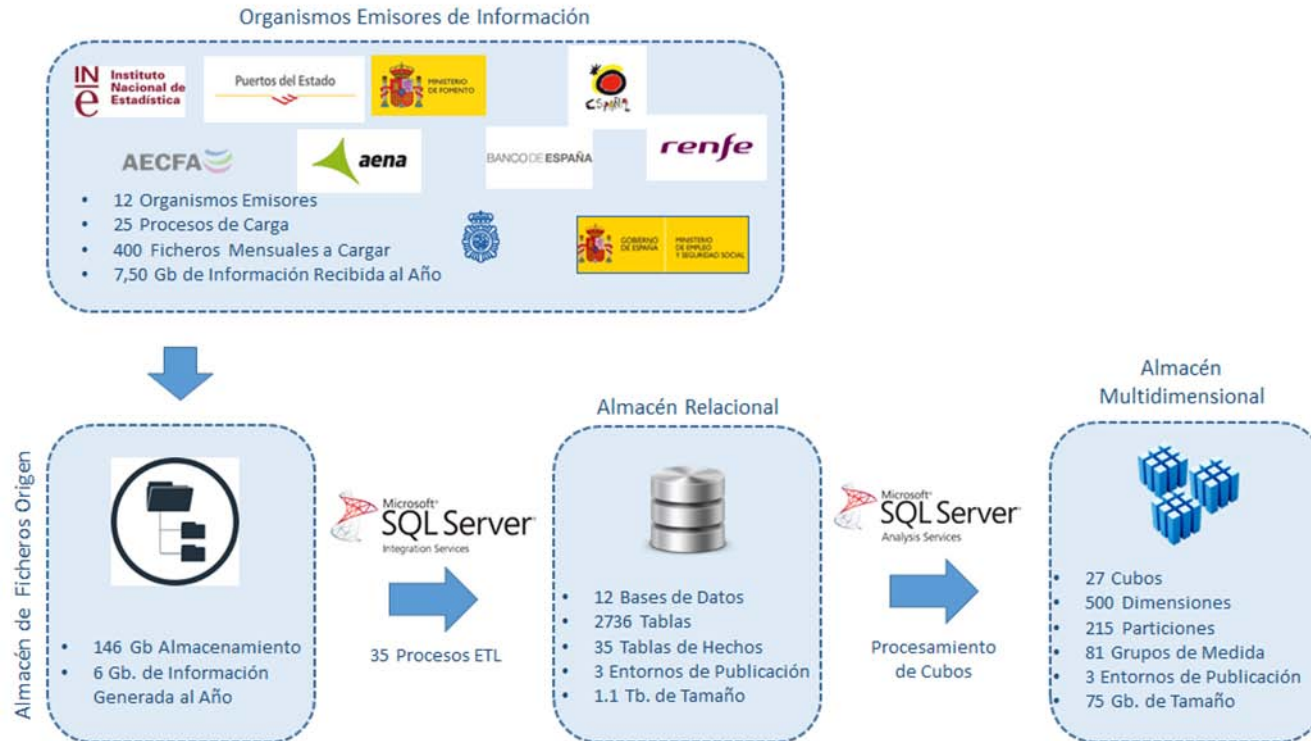
El número de BBDD de AS (Analysis Services) es de 12, que contiene a su vez 27 cubos.

La información es cargada en las BBDD por medio de 35 IS (Integration services).



## 4. Almacenamiento y explotación de la información

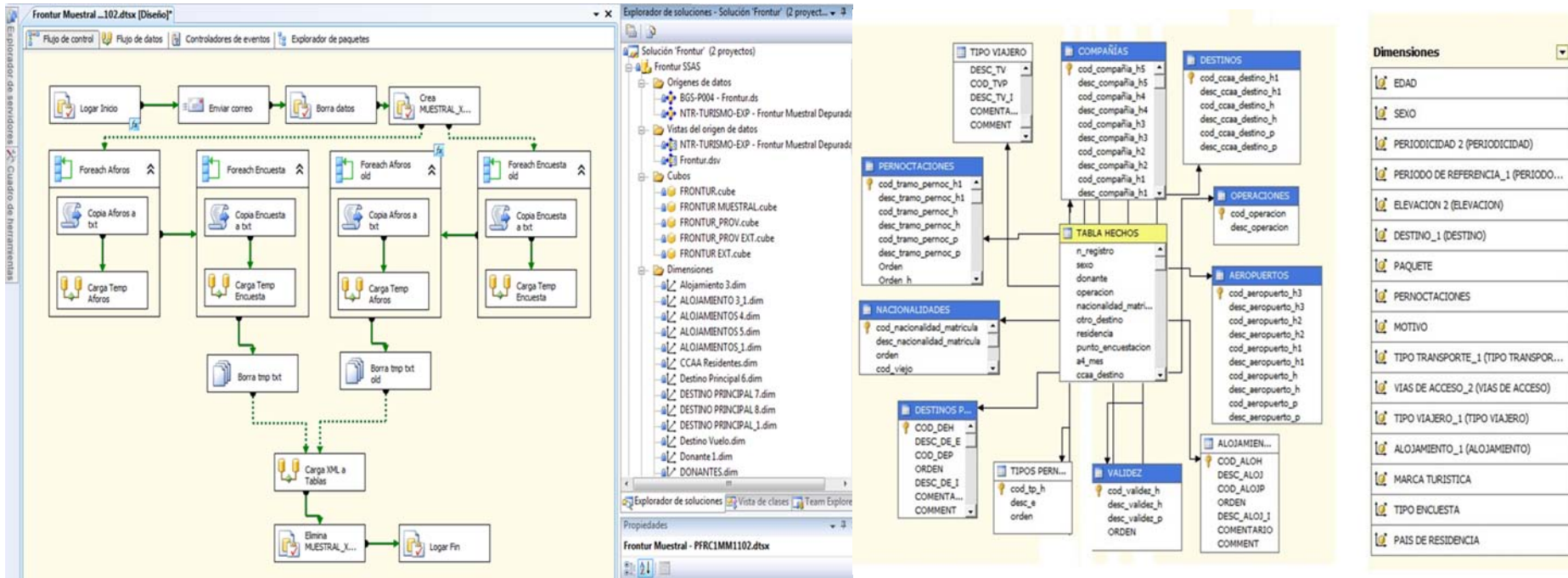
Se recibe información de 12 organismos, con unos 400 ficheros mensuales a cargar para la generación de la estadística. El sistema DATATUR se alimenta a través de 35 procesos ETL y supera el Tb de información almacenada en BBDD. Da lugar a la explotación de 27 cubos dentro del sistema de BI.





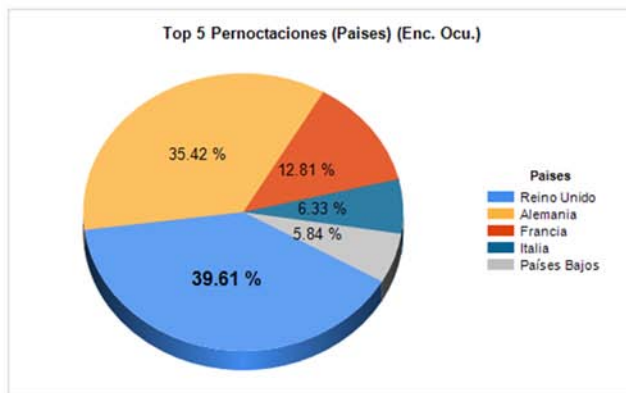
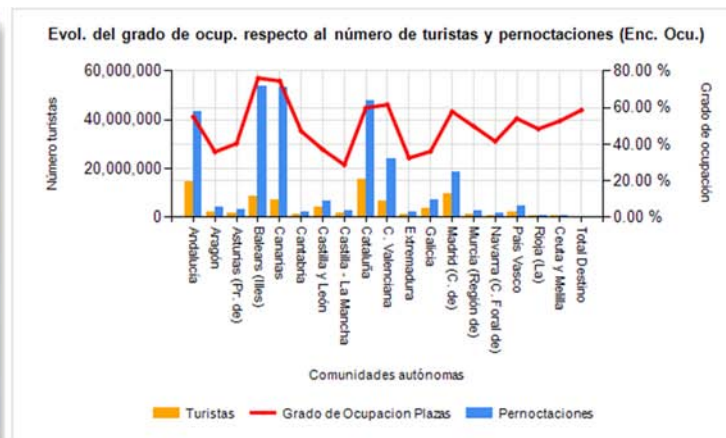
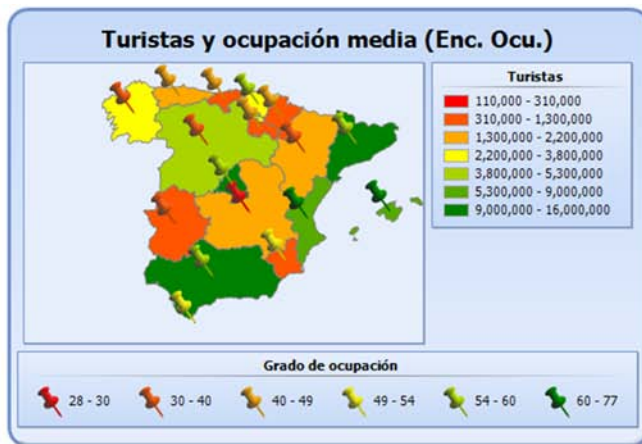
# 5. Procesos de carga y explotación

Los ficheros recibidos se procesan y se publican en 3 entornos diferentes. El entorno de preproducción, el de producción interna (acceso restringido) y el de producción externa (publicación estadística en el portal con acceso libre). El sistema de datos estadísticos multidimensional se explota mediante técnicas de BI (Business Intelligence).



## 6. Análisis de la información Turística

Ante el creciente volumen de información surge la necesidad de explotarla de forma ágil y visual. Una de las posibilidades es explotarla a través de cuadros de mando, donde los mapas temáticos y los indicadores gráficos cobran protagonismo..



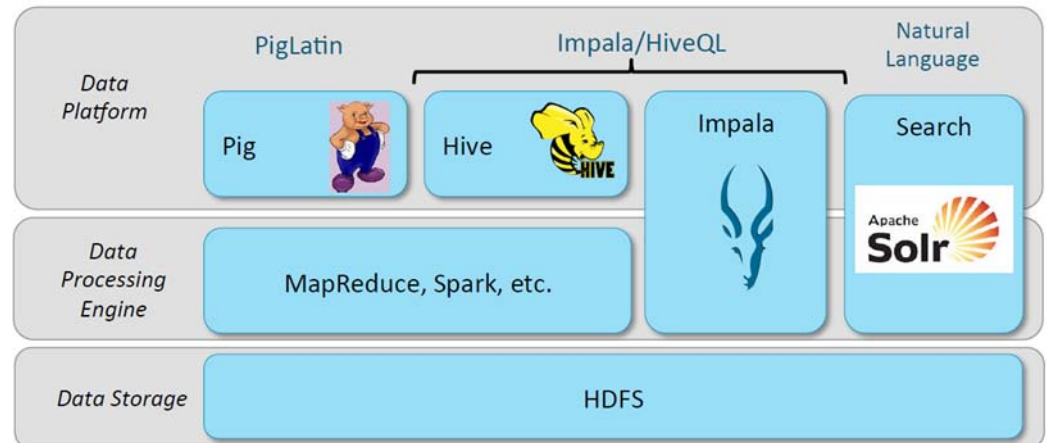
**Gastos total por tipo de turista y medio de transporte (Egatur)**

	Crucero	Ligeros	Línea Regular	Pesados	Vuelo Chárter
Excursionistas	490.92	67.66	260.09	87.79	695.14
Transeúntes	418.36	71.08	122.08	56.67	457.27
Turista	1,492.38	586.85	1,119.46	1,138.48	1,060.71
Turista transeúnte	1,410.97	411.00	465.87	1,476.78	758.31

# 7. El camino lógico hacia el BIG DATA

En torno a Big Data están naciendo aplicaciones y se está profundizando en su desarrollo a partir de diferentes aproximaciones. La necesidad de gestionar y analizar un volumen de datos descomunal a la mayor velocidad posible considerando su extraordinaria variedad (3 V's del Big data) hacen que las organizaciones se planteen esta alternativa. Las previsiones de Gartner apuntaban en el pasado en la misma dirección: actualmente el 65% de las aplicaciones con funciones analíticas avanzadas funcionan sobre Hadoop en 2015.

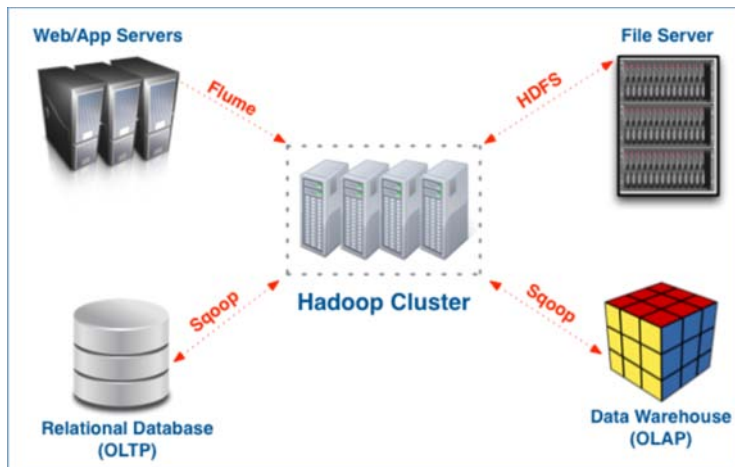
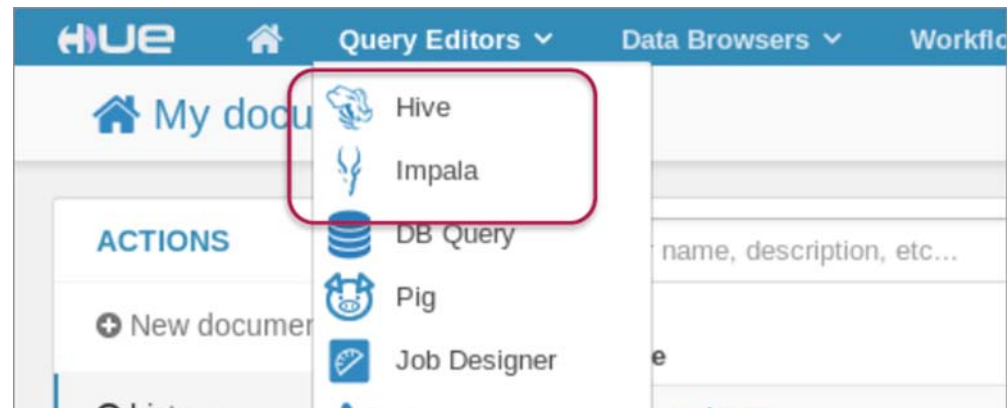
3V - Volume, Variety, Velocity



# 8. Distribuciones Hadoop y Big data

La distribución de Cloudera (CDH) fue la primera en aparecer en el mercado, combinando Big Data y Hadoop. CDH no solo incluye el núcleo de Hadoop (HDFS, MapReduce...) sino que también integra diversos proyectos de Apache (HBase, Mahout, Pig, Hive, etc.).

## Distribuciones Hadoop



**Choose a database** (points to the DATABASE dropdown)

**Explore schema and sample data** (points to the Table name list)

**Enter, edit, save and execute queries** (points to the query editor area)

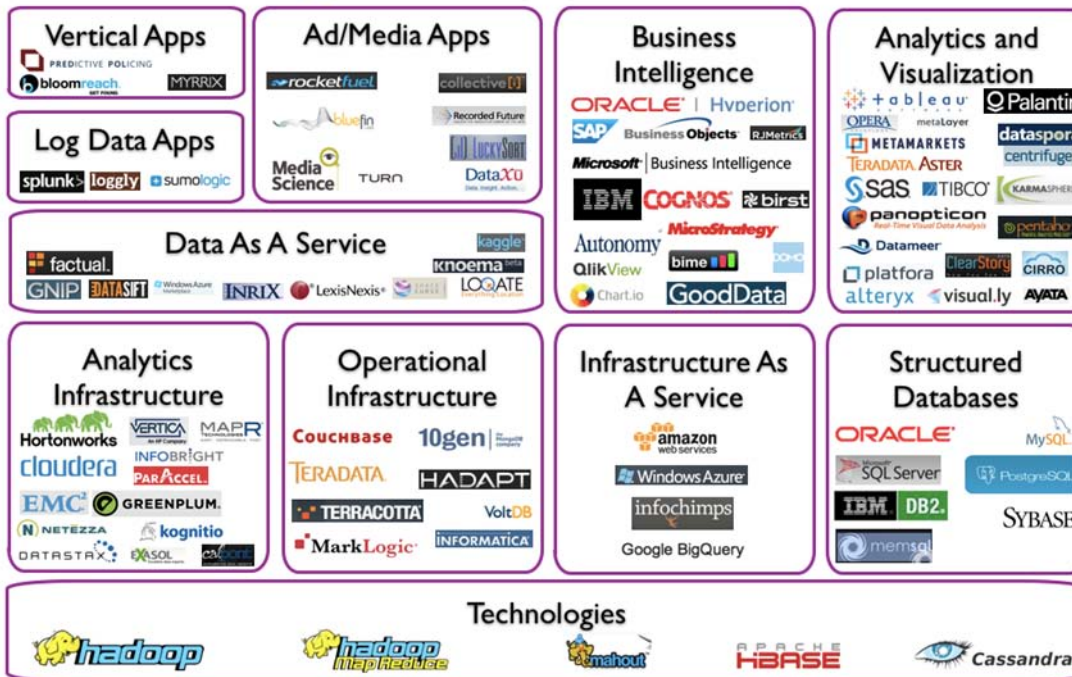
**View results, logs, reports, etc.** (points to the results table)

cust_id	fname	lname	address	city	state	zipcode
0	1000002	Marilyn	Ham	25831 North 25th Street	Concord	CA 94522
1	1000006	Gerard	Franks	356 Turner Street	Pioneer	CA 95666
2	1000010	Mason	Preston	2656 West 13th Street	Redwood Valley	CA 95470
3	1000012	Pamela	Cortez	2279 North Mulberry Avenue	San Francisco	

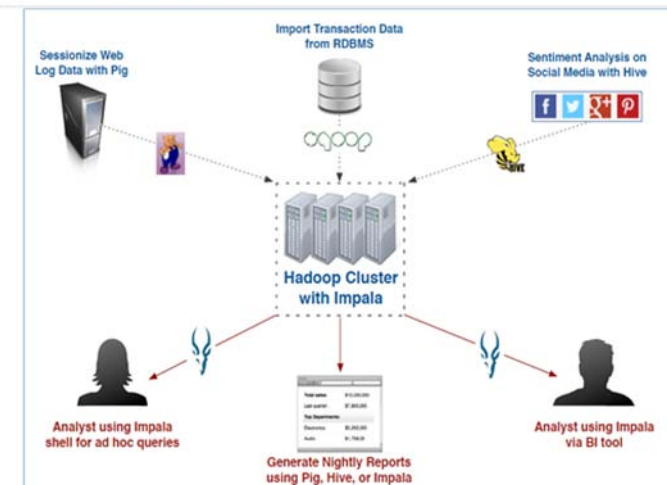
# 9. Necesidad del Big data para estadísticas Turísticas

La necesidad de analizar distintas fuentes de información, incluidas las nuevas posibilidades surgidas con redes sociales con vistas a análisis de los sentimientos (sentiment analysis) (ej. Comentarios sobre experiencias turísticas en España en redes sociales en otro país) o realizar modelos predictivos acerca de los potenciales turistas en base a información histórica hacen necesaria la evolución hacia modelos de Big Data.

## Big Data Landscape

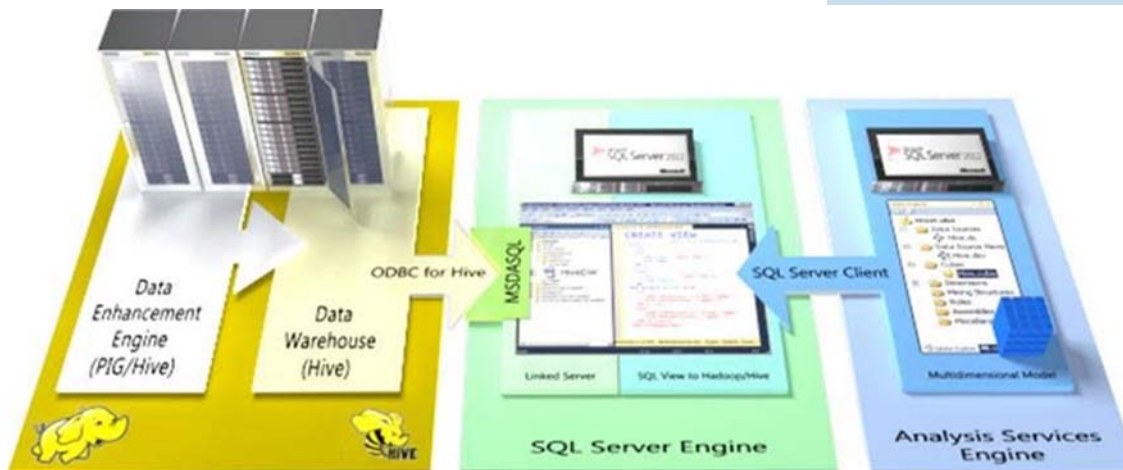
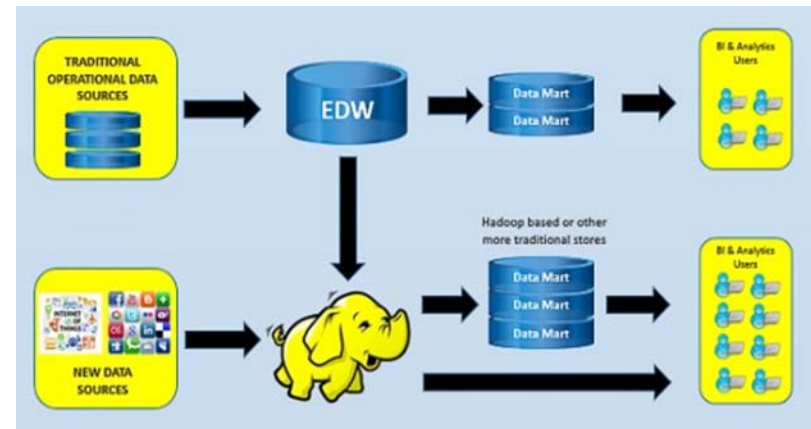


### Analysis Workflow Example



# 10. Nuevo sistema de Bigdata en Turespaña

Contempla la creación de un cluster Hadoop, el empleo de conectores ODBC desde SQL server y la posible explotación estadística con SAS. Se ha recibido formación en la distribución de Cloudera (CDH), combinando Big Data y Hadoop. Incluye el núcleo del cluster Hadoop con HDFS, y el empleo de las herramientas existentes (presente Hive o Impala- futuro Spark).





Gracias  
por su  
atención!!!