



Máster en Data Science. URJC

Arquitecturas en la nube

Presupuesto de mejora de servicio TAJO

*Javier Llorente Mañas
Carlos Sánchez Vega*

2017

Índice

0.1	¿Por qué AWS?	2
0.2	Secciones de la propuesta	2
0.3	Descripción de la propuesta	2
0.4	Esquema conceptual de la propuesta definida	4
0.5	Presupuesto del servicio de Amazon S2	5
0.6	Presupuesto del servicio de Amazon S3	6
0.7	Presupuesto del servicio de Amazon Glacier	7
0.8	Presupuesto del servicio de Amazon Redshift	8
0.9	Presupuesto del servicio de Amazon EMR	8
0.10	Presupuesto del servicio de Amazon SES	9
0.11	Presupuesto del servicio de Amazon EML	9
0.12	Presupuesto del servicio de Amazon Quicksight	10
0.13	Desglose total de los servicios	11

0.1 ¿Por qué AWS?

Usar los productos de AWS nos permite abstraernos de la infrestructura y centrarnos en el análisis y procesado de los datos. El objetivo que tenemos, como compañía, es derivar valor de negocio de los datos obtenidos, de tal forma que nos permita tomar las decisiones de negocio lo más rápido posible que se pueda. Tales decisiones se transforman en valor añadido y ventaja competitiva frente a otras empresas.

La empresa Tajo tiene las siguientes necesidades:

- Generación de informes
- Sistema de aprendizaje automático para el envío de ofertas personalizadas

Se puede ver un esquema conceptual, de la arquitectura de la solución diseñada, en la sección de *Esquema conceptual de la propuesta definida*

Nota importante: he supuesto que todos nuestros clientes estarán en Europa, motivo por el cual los presupuestos reflejados serán para servicios en Europa (más concretamente, por cercanía, en Irlanda)

0.2 Secciones de la propuesta

- Descripción de la propuesta
- Esquema conceptual de la propuesta definida
- Presupuesto del servicio de Amazon EC2
- Presupuesto del servicio de Amazon S3
- Presupuesto del servicio de Amazon Glacier
- Presupuesto del servicio de Amazon Redshift
- Presupuesto del servicio de Amazon EMR
- Presupuesto del servicio de Amazon SES
- Presupuesto del servicio de Amazon EML
- Presupuesto del servicio de Amazon Quicksight
- Desglose total de los servicios

0.3 Descripción de la propuesta

Generación de informes

Se generarán los informes, que se alimentarán de los datos generados por la herramienta Amazon Redshift.

Sistema de aprendizaje para el envío de ofertas

El sistema de aprendizaje para el envío de ofertas, es decir, el sistema de recomendación de ofertas, se compone de los siguientes puntos:

1. Un sistema de aprendizaje automático, que aprenda u obtenga patrones de productos que pudieran ser objeto de compra, en base al perfil de usuario y las compras realizadas.
2. Sistema de mailing, para enviar las recomendaciones a los usuarios

Ambas funcionalidades, mostradas anteriormente, tienen un flujo común que mostraremos a continuación:

1. Exportación de la base de datos local a S3.
2. Análisis los datos con Amazon EMR: Amazon Elastic Map Reduce (Amazon EMR) es un servicio web que facilita el procesamiento rápido y rentable de grandes cantidades de datos. Amazon EMR utiliza Hadoop, un “framework” de código abierto, para distribuir los datos y el procesamiento en un clúster de instancias de Amazon EC2 de tamaño variable.
3. “Limpieza” de los datos con Amazon Redshift. Los datos generados por esta herramienta serán útiles para los siguientes puntos:
 - (a) Generación de informes diarios
 - (b) Entrada de datos para Amazon ML, que procesará los datos y se tomarán conclusiones en base al procesado

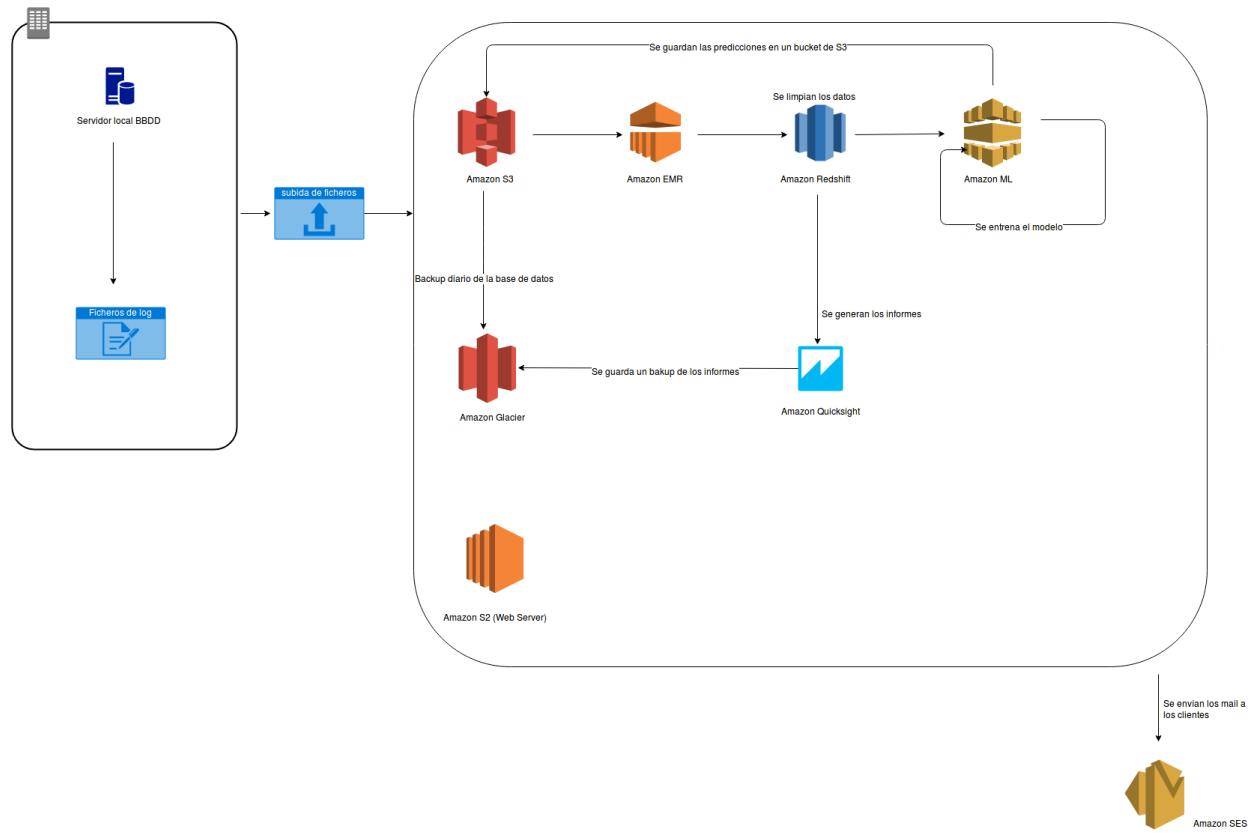
En cuanto al resto de necesidades, usaremos las siguientes herramientas:

- **Servidor Web:** usaremos Amazon S2 para alojar el servidor para alojar la aplicación web de la empresa. Entre las ventajas más notables es que permite escalar recursos rápidamente según las peticiones que tenga la aplicación web.
- **Generación de informes:** se usará la herramienta Amazon Quicksight. Dicho sistema se alimentará, en nuestro caso, de los datos “limpiados” en Amazon Redshift. Además, se podrá acceder a los informes desde dispositivos móviles, lo cual será útil para la directiva o empleados que necesiten acceder a los informes fuera de la oficina.
- **Sistema de aprendizaje:** se creará un sistema de recomendación de ventas utilizando Machine Learning. Machine Learning es la tecnología que encuentra patrones en los datos (en nuestro caso, los históricos de venta de usuarios y sus perfiles) para ser usados para crear predicciones, que nos darán nuevos datos de utilidad (en el caso que nos concierne, recomendaciones sobre potenciales compras). La herramienta de Machine Learning que usaremos de Amazon es Amazon ML.
- **Backup diario de la base de datos:** a pesar de que en el enunciado no se cita

explícitamente que se necesite un sistema de almacenamiento de backup, se cree que es recomendable. Por ello, se ha escogido Amazon Glacier. Gracias a Amazon Glacier podremos guardar un backup diario de la base de datos y de los informes generados por Amazon Quicksight

- **Herramienta de mailing:** se usará Amazon SES

0.4 Esquema conceptual de la propuesta definida



0.5 Presupuesto del servicio de Amazon S2

Services Estimate of your Monthly Bill (\$ 166.17)

Choose region: Europe (Ireland) Inbound Data Transfer is Free and Outbound Data Transfer is 1 GB free per region per month

Compute: Amazon EC2 Instances:

Description	Instances	Usage	Type	Billing Option	Monthly Cost
Linux	1	100 % Utilized/h	Linux on m4.xlarge	On-Demand (No Co.	\$ 162.51

Compute: Amazon EC2 Dedicated Hosts:

Description	Number of Hosts	Usage	Type	Billing Option
Add New Row				

Storage: Amazon EBS Volumes:

Description	Volumes	Volume Type	Storage	IOPS	Baseline Throughput	Snapshot Storage
Add New Row						

Elastic IP:

Number of Additional Elastic IPs:

Elastic IP Non-attached Time: Hours/Month

Number of Elastic IP Remaps: Per Month

Data Transfer:

Inter-Region Data Transfer Out: GB/Month

Data Transfer Out: GB/Month

Data Transfer In: GB/Month

VPC Peering Data Transfer: GB/Month

Intra-Region Data Transfer: GB/Month

Public IP/Elastic IP Data Transfer: GB/Month

Elastic Load Balancing:

Number of Elastic LBs:

Total Data Processed by all ELBs: GB/Month

Services

- [Amazon EC2](#)
- [Amazon S3](#)
- [Amazon Route 53](#)
- [Amazon CloudFront](#)
- [Amazon RDS](#)
- [Amazon DynamoDB](#)
- [Amazon ElastiCache](#)
- [Amazon CloudWatch](#)
- [Amazon SES](#)
- [Amazon SNS](#)
- [Amazon Elastic Transcoder](#)
- [Amazon WorkSpaces](#)
- [Amazon WorkDocs](#)
- [AWS Directory Service](#)
- [Amazon Redshift](#)
- [Amazon Glacier](#)
- [Amazon SQS](#)
- [Amazon SWF](#)
- [Amazon Elastic MapReduce](#)
- [Amazon Kinesis Streams](#)
- [Amazon CloudSearch](#)
- [AWS Snowball](#)
- [AWS Direct Connect](#)

0.6 Presupuesto del servicio de Amazon S3

Amazon EC2

Amazon S3

Amazon Route 53

Amazon CloudFront

Amazon RDS

Amazon DynamoDB

Amazon ElastiCache

Amazon CloudWatch

Amazon SES

Amazon SNS

Amazon Elastic Transcoder

Amazon WorkSpaces

Amazon WorkDocs

AWS Directory Service

Amazon Redshift

Amazon Glacier

Amazon SQS

Amazon SWF

Amazon Elastic MapReduce

Amazon Kinesis Streams

Amazon CloudSearch

AWS Snowball

AWS Direct Connect

Amazon VPC

Amazon SimpleDB

 Amazon S3 is storage for the Internet. It is designed to make web-scale computing easier for developers. For more information, see the [Amazon S3 Product Page](#).

Standard Storage:

Storage:	<input type="text" value="1"/> TB	<input type="button" value="▼"/>
PUT/COPY/POST/LIST Requests:	<input type="text" value="1000"/> Requests	
GET and Other Requests:	<input type="text" value="10000"/> Requests	

Standard - Infrequent Access Storage:

Storage:	<input type="text" value="0"/> GB	<input type="button" value="▼"/>
PUT/COPY/POST/LIST Requests:	<input type="text" value="0"/> Requests	
GET and Other Requests:	<input type="text" value="0"/> Requests	
Lifecycle Transitions:	<input type="text" value="0"/> Transitions	
Data Retrieval:	<input type="text" value="0"/> GB	<input type="button" value="▼"/>

Reduced Redundancy Storage:

Storage:	<input type="text" value="0"/> GB	<input type="button" value="▼"/>
PUT/COPY/POST/LIST Requests:	<input type="text" value="0"/> Requests	
GET and Other Requests:	<input type="text" value="0"/> Requests	

Storage Management:

Inventory:	<input type="text" value="0"/> Per Month	<input type="button" value="▼"/>	Million Objects
Analytics:	<input type="text" value="0"/> Per Month	<input type="button" value="▼"/>	Million Objects
Object Tagging:	<input type="text" value="0"/> Per Month	<input type="button" value="▼"/>	10,000 Tags

Glacier Requests:

Lifecycle Transition Requests into Glacier:	<input type="text" value="0"/> Requests
---	---

Transfer Acceleration:

Inter-Region Acceleration Data Transfer Out:	<input type="text" value="0"/> GB/Month	<input type="button" value="▼"/>
Acceleration Data Transfer Out:	<input type="text" value="0"/> GB/Month	<input type="button" value="▼"/>
Accelerated Data Transfer In - over US, EU or JP locations:	<input type="text" value="0"/> GB/Month	<input type="button" value="▼"/>
Accelerated Data Transfer In - over all other locations:	<input type="text" value="0"/> GB/Month	<input type="button" value="▼"/>

Data Transfer:

Inter-Region Data Transfer Out:	<input type="text" value="0"/> GB/Month	<input type="button" value="▼"/>
Data Transfer Out:	<input type="text" value="500"/> GB/Month	<input type="button" value="▼"/>
Data Transfer In:	<input type="text" value="10"/> GB/Month	<input type="button" value="▼"/>
Data Transfer Out to CloudFront:	<input type="text" value="0"/> GB/Month	<input type="button" value="▼"/>

0.7 Presupuesto del servicio de Amazon Glacier

Get Started with AWS: [Learn more about our Free Tier](#) or [Sign Up for an AWS Account »](#)

FREE USAGE TIER: New Customers get free usage tier for first 12 months

Services		Estimate of your Monthly Bill (\$ 774.17)	
Choose region: Europe (Ireland) Inbound Data Transfer is Free and Outbound Data Transfer is 1 GB free per region per month			
 Amazon Glacier is an extremely low-cost storage service that provides secure and durable storage for data archiving and backup. It is optimized for data that is infrequently accessed and for which retrieval times of several hours are suitable. Clear Form			
Storage: Storage: <input type="text" value="100"/> TB			
Retrieval: Data Retrieval (Standard Tier): <input type="text" value="0"/> GB Data Retrieval (Expedited Tier): <input type="text" value="0"/> GB Data Retrieval (Bulk Tier): <input type="text" value="0"/> GB Standard Retrieval Request: <input type="text" value="0"/> Requests Expedited Retrieval Request: <input type="text" value="0"/> On-Demand Requests/Provisioned Capacity Unit Bulk Retrieval Request: <input type="text" value="0"/> Requests			
Requests: UPLOAD Requests: <input type="text" value="2000"/> Requests			
Data Transfer: Inter-Region Data Transfer Out: <input type="text" value="0"/> GB/Month Data Transfer Out: <input type="text" value="10"/> GB/Month Data Transfer In: <input type="text" value="10"/> GB/Month			
Reset All			



0.8 Presupuesto del servicio de Amazon Redshift

Get Started with AWS: [Learn more about our Free Tier](#) or [Sign Up for an AWS Account »](#)

FREE USAGE TIER: New Customers get free usage tier for first 12 months

Services		Estimate of your Monthly Bill (\$ 401.57)														
Choose region: Europe (Ireland) <input type="button" value="▼"/>																
 Amazon Redshift is a fast, fully managed, petabyte-scale data warehouse service that makes it simple and cost-effective to efficiently analyze all your data using your existing business intelligence tools. <input type="button" value="Clear Form"/>																
Amazon Redshift On-Demand Nodes: <table border="1"> <thead> <tr> <th>Description</th> <th>Nodes</th> <th>Node Type</th> <th>Usage</th> </tr> </thead> <tbody> <tr> <td>: limpieza datos</td> <td>1</td> <td>dc1.large - 0.16TB SSD</td> <td>100 % Utilized/<input type="button" value="▼"/></td> </tr> <tr> <td colspan="4"> <input type="button" value="Add New Row"/> </td> </tr> </tbody> </table>					Description	Nodes	Node Type	Usage	: limpieza datos	1	dc1.large - 0.16TB SSD	100 % Utilized/ <input type="button" value="▼"/>	<input type="button" value="Add New Row"/>			
Description	Nodes	Node Type	Usage													
: limpieza datos	1	dc1.large - 0.16TB SSD	100 % Utilized/ <input type="button" value="▼"/>													
<input type="button" value="Add New Row"/>																
Amazon Redshift Reserved Nodes: <table border="1"> <thead> <tr> <th>Description</th> <th>Nodes</th> <th>Node Type</th> <th>Billing Option</th> </tr> </thead> <tbody> <tr> <td colspan="4"> <input type="button" value="Add New Row"/> </td> </tr> </tbody> </table>					Description	Nodes	Node Type	Billing Option	<input type="button" value="Add New Row"/>							
Description	Nodes	Node Type	Billing Option													
<input type="button" value="Add New Row"/>																
Additional Backup Storage: <table border="1"> <thead> <tr> <th>Description</th> <th>Storage</th> </tr> </thead> <tbody> <tr> <td colspan="2"> <input type="button" value="Add New Row"/> </td> </tr> </tbody> </table>					Description	Storage	<input type="button" value="Add New Row"/>									
Description	Storage															
<input type="button" value="Add New Row"/>																
<small>Backup storage up to 100% of provisioned Storage for Active Nodes is provided for Free</small>																

Amazon EC2 Amazon S3 Amazon Route 53 Amazon CloudFront Amazon RDS Amazon DynamoDB Amazon ElastiCache Amazon CloudWatch Amazon SES Amazon SNS Amazon Elastic Transcoder Amazon WorkSpaces Amazon WorkDocs AWS Directory Service Amazon Redshift Amazon Glacier Amazon SQS

0.9 Presupuesto del servicio de Amazon EMR

Get Started with AWS: [Learn more about our Free Tier](#) or [Sign Up for an AWS Account »](#)

FREE USAGE TIER: New Customers get free usage tier for first 12 months

Services		Estimate of your Monthly Bill (\$ 773.16)																	
Choose region: Europe (Ireland) <input type="button" value="▼"/>																			
 Amazon Elastic MapReduce is a web service that enables businesses, researchers, data analysts, and developers to easily and cost-effectively process vast amounts of data. <input type="button" value="Clear Form"/>																			
Cluster Size: <table border="1"> <thead> <tr> <th>Cluster Name</th> <th>Instances</th> <th>Usage</th> <th>Hadoop Distribution</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>ofertas/clientes</td> <td>1</td> <td>100 % Utilized/<input type="button" value="▼"/></td> <td>MapR M3</td> <td>c1.medium</td> </tr> <tr> <td colspan="5"> <input type="button" value="Add New Row"/> </td> </tr> </tbody> </table>					Cluster Name	Instances	Usage	Hadoop Distribution	Type	ofertas/clientes	1	100 % Utilized/ <input type="button" value="▼"/>	MapR M3	c1.medium	<input type="button" value="Add New Row"/>				
Cluster Name	Instances	Usage	Hadoop Distribution	Type															
ofertas/clientes	1	100 % Utilized/ <input type="button" value="▼"/>	MapR M3	c1.medium															
<input type="button" value="Add New Row"/>																			

Amazon EC2 Amazon S3 Amazon Route 53 Amazon CloudFront Amazon RDS Amazon DynamoDB Amazon ElastiCache



0.10 Presupuesto del servicio de Amazon SES

Get Started with AWS: [Learn more about our Free Tier](#) or [Sign Up for an AWS Account »](#)

FREE USAGE TIER: New Customers get free usage tier for first 12 months

Services		Estimate of your Monthly Bill (\$ 774.17)
Choose region: Europe (Ireland) Inbound Data Transfer is Free and Outbound Data Transfer is 1 GB free per region per month		
 Amazon Simple Email Service (Amazon SES) is a highly scalable and cost-effective bulk and transactional email-sending service for businesses and developers. Clear Form		
FREE TIER: Amazon EC2 customers who use Amazon SES service pay no charges for the first 2000 Email Messages Per Day		
Messages: Email Messages: <input type="text" value="2000"/> Per Day		
Attachments: Data Transfer Out (Attachments): <input type="text" value="0"/> GB/Month		
Dedicated IP Addresses: Number of Dedicated IP Addresses: <input type="text" value="0"/> Per Month		
Data Transfer: Data Transfer Out: <input type="text" value="0"/> GB/Month Data Transfer In: <input type="text" value="1"/> GB/Month		
Mail Chunks: Quantity: <input type="text" value="0"/> Mail Chunks		

0.11 Presupuesto del servicio de Amazon EML

Amazon Machine Learning cobra una tarifa horaria por el tiempo informático utilizado para crear modelos predictivos y, a continuación, se paga por la cantidad de predicciones generadas. Por tanto, a priori no podríamos crear un presupuesto con una cifra de gasto exacta aunque sí que podríamos hacer una estimación en base al tiempo necesario para procesar las predicciones.

Supondremos que, para el caso de TAJO, utilizaremos 20 horas de tiempo de procesamiento informático para generar sus modelos y obtiene 890 000 predicciones en lote.

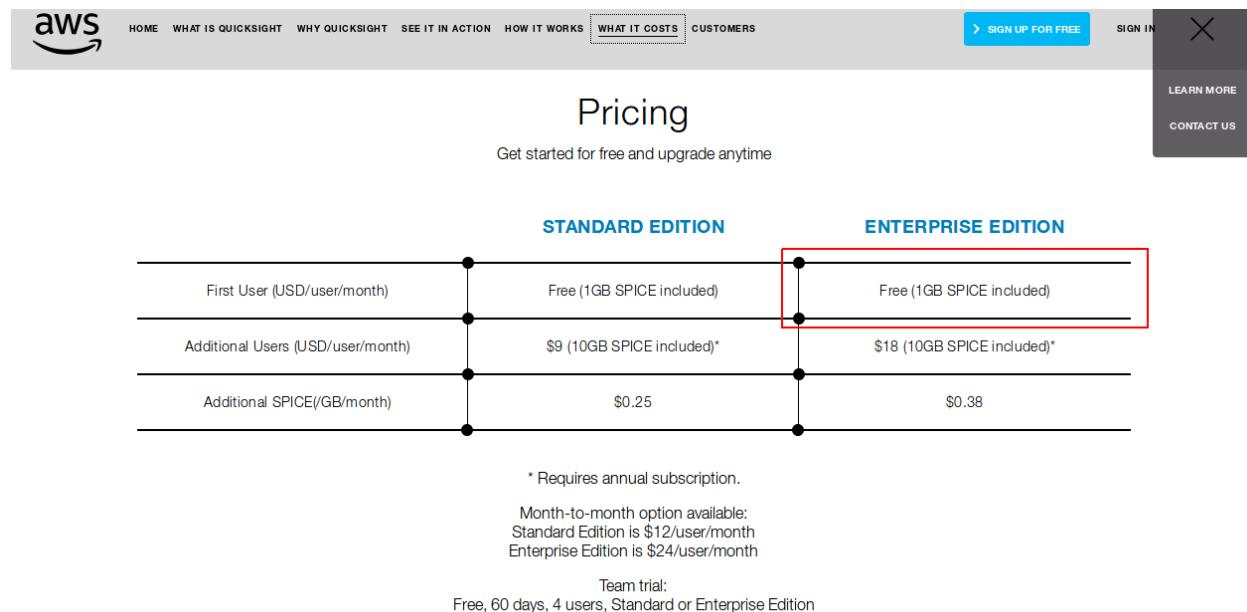
Cuotas de procesamiento informático

- **Cuotas de procesamiento informático.**
 - El precio del procesamiento informático es de 0,42 USD por hora.
 - Cuotas totales de procesamiento informático = Tiempo de procesamiento informático (h) * 0,42 USD
 - **Cuotas totales de procesamiento informático = 20 h * 0,42 USD/h = 8,40 USD**
- **Cuotas mensuales de predicciones**

- El precio mensual de las predicciones en lote es de 0,10 USD por 1 000 predicciones.
- **Cuotas totales de predicciones = (0,10 USD/1 000) * 890 000 = 89,00 USD**
- **Cuotas totales mensuales**
 - **Cuotas totales mensuales Cuotas de procesamiento informático + cuotas de predicciones = 8,40 USD + 89,00 USD = 97,40 USD**

0.12 Presupuesto del servicio de Amazon Quicksight

En cuanto al sistema de generación de informes, podremos usar la versión gratuita, como se muestra a continuación:



The screenshot shows the AWS Quicksight Pricing page. At the top, there's a navigation bar with links for HOME, WHAT IS QUICKSIGHT, WHY QUICKSIGHT, SEE IT IN ACTION, HOW IT WORKS, WHAT IT COSTS (which is highlighted), and CUSTOMERS. There are also buttons for SIGN UP FOR FREE, SIGN IN, LEARN MORE, and CONTACT US. Below the navigation, the word "Pricing" is centered, followed by the subtext "Get started for free and upgrade anytime".

The main content is a table comparing the Standard Edition and Enterprise Edition. The table has three columns: "STANDARD EDITION" and "ENTERPRISE EDITION" (with a red box around it). The rows represent different usage scenarios:

	STANDARD EDITION	ENTERPRISE EDITION
First User (USD/user/month)	Free (1GB SPICE included)	Free (1GB SPICE included)
Additional Users (USD/user/month)	\$9 (10GB SPICE included)*	\$18 (10GB SPICE included)*
Additional SPICE(/GB/month)	\$0.25	\$0.38

* Requires annual subscription.
 Month-to-month option available:
 Standard Edition is \$12/user/month
 Enterprise Edition is \$24/user/month

Team trial:
 Free, 60 days, 4 users, Standard or Enterprise Edition

0.13 Desglose total de los servicios

Estimate of Your Monthly Bill		
<input checked="" type="checkbox"/> Show First Month's Bill (include all one-time fees, if any)		
 Below you will see an estimate of your monthly bill. Expand each line item to see cost breakout of each service. To save this bill and input values, click on 'Save and Share' button. To remove the service from the estimate, jump back to the service and clear the specific service's form.		
Export to CSV		Save and Share
<input type="checkbox"/> Amazon EC2 Service (Europe)	\$ 186.79	
Compute:	\$ 162.51	
Elastic IPs:	\$ 3.66	
Elastic LBs:	\$ 20.50	
Data Processed by Elastic LBs:	\$ 0.12	
<input type="checkbox"/> Amazon S3 Service (Europe)	\$ 23.58	
Standard Storage:	\$ 23.56	
Standard Put/List Requests:	\$ 0.01	
Standard Other Requests:	\$ 0.01	
<input type="checkbox"/> Amazon SES Service (Europe)	\$ 0.00	
Email Messages:	\$ 0.00	
<input type="checkbox"/> Amazon Redshift Service (Europe)	\$ 219.60	
Compute:	\$ 219.60	
<input type="checkbox"/> Amazon Glacier Service (Europe)	\$ 409.71	
Storage:	\$ 409.60	
Upload Requests:	\$ 0.11	
<input type="checkbox"/> Amazon Elastic MapReduce Service (Europe)	\$ 130.30	
Compute:	\$ 130.30	
<input type="checkbox"/> AWS Data Transfer In	\$ 0.00	
Europe (Ireland) Region:	\$ 0.00	
<input type="checkbox"/> AWS Data Transfer Out	\$ 45.81	
Europe (Ireland) Region:	\$ 45.81	
<input type="checkbox"/> AWS Support (Basic)	\$ 0.00	
Support for all AWS services:	\$ 0.00	
Free Tier Discount:	\$ -241.62	
Total Monthly Payment:		\$ 774.17

El importe total de los servicios sería el importe de todos los servicios + servicio de Amazon ML. Es decir:

- **Importe de todos los servicios** (mostrados en la imagen superior) = 774,17 USD
- **Importe del servicio de Amazon ML**: 97,40 USD

El importe total de la mejora de servicio es de 774,17 USD + 97,40 USD = 871,57 USD / mensuales