



**UNIVERSIDAD DE CASTILLA-LA MANCHA**  
**ESCUELA SUPERIOR DE INFORMÁTICA**  
**Tecnologías y sistemas de información**

**ANTEPROYECTO DEL TRABAJO FIN DE GRADO**  
**GRADO EN INGENIERÍA INFORMÁTICA**  
**TECNOLOGÍA ESPECÍFICA DE COMPUTACIÓN**

**Sense: Hub de datos de alto rendimiento y  
escalabilidad para tecnologías ‘Internet of Things’.**

Autor: David Martín García

Director: David Villa Alises

Director: Maria Jose Santofimia Romero

Noviembre, 2015

# Índice

<b>1. INTRODUCCIÓN</b>	<b>2</b>
<b>2. TECNOLOGÍA ESPECÍFICA / INTENSIFICACIÓN / ITINERARIO CURSADO POR EL ALUMNO</b>	<b>3</b>
<b>3. OBJETIVOS</b>	<b>5</b>
<b>4. MÉTODO Y FASES DE TRABAJO</b>	<b>5</b>
<b>5. MEDIOS QUE SE PRETENDEN UTILIZAR</b>	<b>7</b>
5.1. Medios Hardware . . . . .	7
5.2. Medios Software . . . . .	8
<b>6. REFERENCIAS</b>	<b>8</b>

# 1. INTRODUCCIÓN

La sociedad avanza hacia una nueva era en la que internet forma un gran papel. En los últimos años estamos experimentando un crecimiento exponencial en la cantidad de datos volcados en internet así como en el número de dispositivos que crean o consumen tal información. Actualmente cursamos dentro de la intensificación de computación asignaturas como 'Sistemas Multiagentes' o 'Minería de Datos' donde principios como la ubicuidad o el big-data quedan marcados como una premisa de futuro.

Todo está interconectado y tecnologías como la domótica, ocio o la salud sufren esta tendencia. Hoy en día dispositivos como smart watches, smart bands, smart phones, smart tv se conectan a internet para consumir y publicar datos, desde tendencias de uso hasta datos personales de nuestra salud o ejercicio físico, por otro lado, empresas de multitud de ámbitos consumen y publican datos para uso interno, empresas de logística, automatización de entornos domésticos e industriales, generación y distribución de energía eléctrica, automóvil, telecomunicaciones e información. Todos estos campos se ven cada vez más influenciados por esta tecnología.

Solo hay que imaginar una cadena logística, donde una gran cantidad de sensores y actuadores funcionan de forma conjunta para proveer de información, tanto al usuario como al propio funcionamiento de la empresa.

Cada día se ve más presente mercados influenciados por tecnologías M2M (Machine to Machine) donde la interacción es únicamente entre dispositivos. Empresas como Cisco prevén 50 billones de dispositivos conectados a internet, esta ingente cantidad de datos conlleva grandes problemas de almacenamiento, así como un requisito, la optimalidad, ya que el número de conexiones crece exponencialmente, tanto que es necesario nuevos protocolos como IPv6, donde existen 340 sextillones de direcciones. Esto hace intuir que no es una simple moda o tendencia que se disipará en unos años, es una necesidad, y empresas como IBM<sup>1</sup>, Microsoft o Intel<sup>2</sup> apuestan por ello. Otras empresas como Ericson, Siemens Bosch tienen las tecnologías IoT como punta de lanza en las presentaciones de sus tecnologías más novedosas.

Todo esto hace prever que no es solo un cambio de tendencia temporal, es uno de los mayores campos de investigación y negocio en los próximos años.

Por otro lado, el número de hubs de datos ha crecido de manera exponencial, así como su valor en el mercado, dado que existe una gran demanda de este tipo de servicios.

---

<sup>1</sup><http://www.ibmbigdatahub.com/technology/internet-of-things>

<sup>2</sup><https://software.intel.com/es-es/iot/microsoft-azure>

En este punto comienza este proyecto, donde se pondra en marcha un Hub de datos, donde poder publicar y consumir toda esta información. La principal meta de este proyecto es la optimizacion de este proceso, asi como el analisis de los datos, dos de los principales hitos en la mayoria de Hubs actuales. Este hub sera de caracter gratuito y open-source, donde se mezclan tecnologias y lenguajes de forma heterogenea, y donde cada uno es el mejor en su campo, todos ellos en conjunto formaran un conjunto de micro-servicios, interrelacionados entre ellos. Las premisas del proyecto estan principalmente marcadas por la tasa de ingestion de datos asi como la posibilidad de realizar operaciones a un bajo costo, operaciones como el analisis de los datos.

## 2. TECNOLOGÍA ESPECÍFICA / INTENSIFICACIÓN / ITINERARIO CURSADO POR EL ALUMNO

Cuadro 1: Tecnología Específica cursada por el alumno

Marcar la tecnología cursada	
	Tecnologías de la Información
*	<b>Computación</b>
	Ingeniería del Software
	Ingeniería de Computadores

Cuadro 2: Justificación de las competencias específicas abordadas en el TFG

Competencia	Justificación
Capacidad para adquirir, obtener, formalizar y representar el conocimiento humano en una forma computable para la resolución de problemas mediante un sistema informático en cualquier ámbito de aplicación, particularmente los relacionados con aspectos de computación, percepción y actuación en ambientes entornos inteligentes.	Al tratar con valores de cualquier forma, como por ejemplo sensores, debemos comprender las diferentes magnitudes así como una posible forma normal para poder comparar entre los distintos valores.
Capacidad para desarrollar y evaluar sistemas interactivos y de presentación de información compleja y su aplicación a la resolución de problemas de diseño de interacción persona computadora.	Desarrollo de entorno gráfico donde poder monitorizar los datos así como las distintas opciones y incidencias que se detecten, además de la gestión de perfiles de usuario así como de los distintos metadatos asociados a la entrada de datos.
Capacidad para conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.	Análisis de los datos obtenidos así como detección de valores anormales y obtención de posibles respuestas ante tal estímulo.
Capacidad para evaluar la complejidad computacional de un problema, conocer estrategias algorítmicas que puedan conducir a su resolución y recomendar, desarrollar e implementar aquella que garantice el mejor rendimiento de acuerdo con los requisitos establecidos.	Diseño e implementación de algoritmo de agrupamiento de datos. Este algoritmo agregará los diferentes datos con el fin de mantener controlado el tamaño máximo de la base de datos.

### **3. OBJETIVOS**

Desarrollo de un hub para el almacenamiento de datos con caracter temporal orientado al concepto IoT ('Internet of Thing').

Se debe plantear un arquitectura de microservicios de alto rendimiento. Se plantea una parte central que permita el almacenamiento, analisis y tratamiento de los datos, asi como servicios en torno a esta para las notificaciones a los distintos dispositivos que requieran de estos datos.

Debe existir servicios de notificación, analisis, autenticación ademas de la gestion y almacenamiento de los datos asi como su consulta.

Tambien se requiere de una interfaz grafica de gestion que se desarrolla mediante una aplicación web, asegurando asi su total compatibilidad con la gran mayoría de dispositivos existente.

Debe tener compatibilidad con los estandares de comunicacion REST e ZeroC ICE, teniendo ademas en cuenta MQTT como posible candidato.

Se debe disponer de una API para la comunicacion con los distintos dispositivos externos, asi como un protocolo de comunicación para la parte interna entre servicios.

Se tendra el cuenta el rendimiento como principal punto de seleccion en la infraestructura asi como en cada parte del proyecto, debido a esto se preve un desarrollo multilenguaje como punto principal para la busqueda de este objetivo.

Tambien se desarrollara el aprovisionamiento de servidores asi como la disposicion de un entorno de desarrollo de facil instalación para cualquier desarrollador.

TODO Incluir grafico preliminar de la arquitectura

### **4. MÉTODO Y FASES DE TRABAJO**

Para llevar a cabo este proyecto se ha llevado a cabo un análisis de las diversas metodologías existentes.

La eleccion de una metodologia agil es fiel reflejo de la naturaleza de este TFG, donde a fecha inicial no se define totalmente los requisitos con total exactitud y donde la planificacion a corto plazo cobra un mayor sentido.

Tras obtener las ventajas e inconvenientes de cada una se ha optado por scrumban, una mezcla entre Kanban y Scrum, donde se pactaran diferentes hitos dentro del plazo máximo para desa-

rollar el proyecto, cumpliendo con diferentes sprints, pero a su vez haciendo uso de Kanban como planificación de las diferentes tareas.

Con esto tenemos las ventajas de ambos, por un lado controlamos el progreso en casa sprint, y nos permite a su vez gestionar las tareas mediante un tablón Kanban, obteniendo la mínima latencia desde el desarrollo hasta la integración con su versión de producción. Además nos permite seguir las diferentes fases del software, desde el diseño, hasta su puesta al público, pasando por las fases de codificación, test, aceptación y calidad.

Esta metodología nos permite adaptarnos a cualquier cambio en los requisitos o en el ritmo de trabajo, ya que por la propia naturaleza del TFG, debe compaginarse con el resto de actividades académicas.

TODO: DEFINIR HITOS A GRANDES RASGOS, es por hacerme una idea del tamaño Corregir si o si.....

- 1º hito
  - Definición de la arquitectura del proyecto
  - Análisis de las diferentes partes
  - Realización de la documentación inicial y puesta a punto de los diferentes servicios involucrados en el desarrollo
- 2º hito
  - Analisis de rendimiento de las diferentes propuestas, así como la obtención de una primera cifra del rendimiento esperado del proyecto en cada una de sus partes.
  - Codificación del núcleo de micro-servicios
  - Obtención de la primera versión mínimamente funcional
- 3º hito
  - Desarrollo de la interfaz web
  - Desarrollo del backend de análisis de datos.
- 4º hito
  - Integración con dispositivos.

## 5. MEDIOS QUE SE PRETENDEN UTILIZAR

### 5.1. Medios Hardware

Para la realización de este proyecto se van a utilizar los siguientes medios hardware, se estructuran en los diversos entornos disponibles:

**Desarrollo:** Entorno de desarrollo, donde se realizan las tareas de codificación.

Cuadro 3: Características del entorno de desarrollo.

MacBook Pro 2014 Mid.	
CPU	Intel i5 4308U @ 2.8Ghz.
RAM	8GB DDR3L 1600Mhz.
Almacenamiento	SSD M2 128GB.
SO	Debian 8.2.

**Staging** Servidor local donde se realizaran las distintas pruebas a la plataforma para comprobar su funcionamiento, además realiza las funciones de servidor público donde se alojara la plataforma.

Cuadro 4: Características del entorno de staging.

CPU	Intel i5 4690K @ 5Ghz.
RAM	16GB DDR3 1833Mhz.
Almacenamiento	SSD 120GB + RAID0 2TB (2 discos de 1TB).
SO	Debian 8.2.

**Producción** Servidor dedicado donde se encuentra el entorno público, este entorno es el que disfrutaran los usuarios.

Cuadro 5: Características del entorno de staging.

CPU	Intel Xeon E5 2620V2.
RAM	32GB DDR3 ECC.
Almacenamiento	3x SSD 256GB (RAID0) + 100GB (Backups).
SO	Debian 8.3.



## 5.2. Medios Software

- Control de versiones: Git.
- Comunicación entre las distintas partes: Slack.
- Técnicas de diseño e implementación software: TDD.
- Servicios de integración continua: TravisCI ó CircleCI.
- Lenguaje: Elixir, C, Python, Ruby, Javascript, Html5.
- Entorno de desarrollo: Emacs.

Además, se prevé utilizar las siguientes tecnologías en la capa de persistencia:

- Base de datos: NoSQL con escalabilidad en horizontal así como un tiempo medio de respuesta bajo en las inserciones. Se prevé utilizar Cassandra.
- Persistencia temporal: Redis.

## 6. REFERENCIAS

En esta sección se incluirán todas las referencias bibliográficas, ordenadas alfabéticamente por el primer apellido del primer autor, de las obras de las cuales se haya realizado alguna cita en los apartados anteriores. Las referencias deberán contener datos básicos como nombre y apellidos de los autores, título de la obra, evento al que pertenece, páginas, fecha y lugar de celebración (si se tratara de artículos de congreso), ISBN, editorial y ciudad (si se tratara de libro), nombre de revista, páginas, volumen y número (si se tratara de revista), etc.

Se empleará un formato de referencia reconocido en el ámbito académico como ACM<sup>34</sup>. Otros formatos aconsejables son, por ejemplo, IEEE, AMA, APA y AMA.

A continuación una sección de «Referencias» con ejemplos de referencias con formato ACM para:

- Enable things to talk, Springer [vKSLM13].

### Referencias

[vKSLM13] Alessandro Bassi Martin Bauer Martin Fiedler Thorsten Kramp Rob van Kraenburg Sebastian Lange and Stefan Meissner. *Enabling Things to Talk*, chapter Introduction to the Internet of Things, pages 1–10. Springer Berlin Heidelberg, 2013.

---

<sup>3</sup><http://www.acm.org/sigs/publications/proceedings-templates>

<sup>4</sup><http://www.cs.ucy.ac.cy/~chryssis/specs/ACM-refguide.pdf>