

**UNIVERSIDAD EAFIT**  
**ESCUELA DE INGENIERÍA**  
**INGENIERÍA DE SISTEMAS**

**2018**

**Profesor: Edwin Nelson Montoya Múnera – email: emontoya@eafit.edu.co**

## 1. INFORMACIÓN GENERAL DEL CURSO:

Nombre materia:	Tópicos Especiales en Telemática
Código materia:	ST0263
Créditos:	3
Intensidad horaria:	48 horas
Escuela:	Ingenierías
Departamento:	Informática y Sistemas
Programa:	Ingeniería de Sistemas
Prerrequisitos:	ST0255 Telemática
Semestre:	VII

## 2. PRESENTACIÓN:

El presente curso permite a los alumnos obtener los conocimientos, habilidades y actitudes necesarias para el diseño, desarrollo y ejecución de aplicaciones y sistemas computacionales escalables, intensivos en datos e intensivos en computación.

Las aplicaciones y sistemas computacionales modernos, exigen un conocimiento y apropiación alta de la infraestructura tecnológica donde ejecutarán, de tal forma que puedan cumplir diferentes atributos de calidad como Escalabilidad, Alto desempeño en computación y Alto desempeño en el procesamiento de grandes volúmenes de datos.

La Escalabilidad hace referencia al crecimiento de la demanda de recursos en los sistemas como el incremento en el número de usuarios, mayores cantidades de almacenamiento, mejores tiempos de respuesta y sistemas más seguros entre otros. Esta escalabilidad está soportada por las buenas prácticas, patrones de arquitectura, tácticas y herramientas tanto a nivel de aplicación como a nivel de infraestructura TI.

El Alto desempeño en computación hace referencia a la necesidad de contar con infraestructura tecnológica para el procesamiento paralelo y distribuido que ejecutan aplicaciones y programas que resuelven problemas complejos, los cuales requieren la articulación de muchos procesadores y computadores para lograr en tiempos razonables, la ejecución de aplicaciones que ambientes tradicionales demorarían semanas, meses o inclusive años. Estos sistemas de alto desempeño comúnmente son llamados Computación de Alto Desempeño o High-Performance Computing (HPC).

El Big Data, dada la alta proliferación de datos que ya hoy en día se pueden generar desde las redes sociales, miles de sensores, millones de smartphones, el Internet de las Cosas, etc. Requiere contar infraestructuras tecnológicas que permitan no solo almacenar este volumen tan alto de datos, sino que

permita procesarlos en tiempos razonables. Es por ello, que surge todo un campo de Ciencia de los Datos e Ingeniería de los Datos, éste último más relacionado con los Ingenieros de Sistemas.

### 3. OBJETIVO:

#### General:

Desarrollar aplicaciones y sistemas transaccionales, intensivos en datos e intensivos en procesamiento escalables que consideren diferentes atributos de calidad que consideren las particularidades propias de los tipos de sistemas a considerar (transaccionales, intensivos en datos e intensivos en procesamiento). El desarrollo hace referencia a las etapas de un ciclo de vida que considera definición, análisis, diseño, implementación y pruebas utilizando patrones de arquitectura, mejores prácticas, tácticas y herramientas computacionales para el Desarrollo del sistema.

#### Específicos:

- Comprender la fundamentación de los sistemas distribuidos, arquitectura, características, modelos y tipos y los diferentes tipos de sistemas (transaccionales, intensivos en datos e intensivos en computación).
- Desarrollar sistemas y aplicaciones transaccionales escalables con atributos de calidad relacionados a Disponibilidad, Rendimiento y Seguridad.
- Desarrollar sistemas y aplicaciones intensivos en datos que utiliza una infraestructura TI adecuada para Big Data.
- Desarrollar sistemas y aplicaciones intensivos en computación (HPC) que utiliza una infraestructura adecuada para HPC ya sea en Clúster o en Cloud Computing.

### 4. COMPETENCIAS GENERICAS:

#### Instrumentales:

- Capacidad de análisis y síntesis
- Capacidad de organización y planificación
- Comunicación oral y escrita
- Conocimiento de una lengua extranjera
- Resolución de problemas
- Toma de decisiones

#### Personales:

- Trabajo en equipo
- Habilidades en las relaciones interpersonales
- Razonamiento Crítico

#### Sistemáticas:

- Aprendizaje autónomo

Otras:

- Capacidad de aplicar los conocimientos teóricos en la práctica
- Experiencia previa

## 5. COMPETENCIAS ESPECIFICAS:

- Especifica requisitos no funcionales relacionados con los atributos de calidad de Disponibilidad, Rendimiento y Seguridad en sistemas informáticos escalables.
- Diseña aplicaciones y sistemas computacionales que tienen en cuenta atributos de calidad como Disponibilidad, Rendimiento y Seguridad.
- Selecciona la arquitectura de sistema y software adecuada
- Desarrolla sistemas y aplicaciones computacionales que consideran diferentes atributos de calidad teniendo en cuenta la disponibilidad de tecnologías de hardware, software, redes.
- Comprende y usa la infraestructura TI para ejecutar aplicaciones Big Data.
- Desarrolla aplicaciones distribuidas intensivas en datos basadas en Map/Reduce y Spark en diferentes lenguajes de desarrollo.
- Desarrolla aplicaciones paralelas y distribuidas intensivas en cómputo ejecutando en diferentes ambientes de computo como Clusters o GPU

## 6. RESULTADOS DE APRENDIZAJE:

- Desarrolla sistemas y aplicaciones escalables teniendo en cuenta atributos de calidad como Disponibilidad, Rendimiento y Seguridad.
- Desarrolla sistemas y aplicaciones big data con las principales tecnologías actuales para el procesamiento de altos volúmenes de datos.
- Desarrolla sistemas y aplicaciones de alto desempeño con las principales tecnologías para clústeres de alto desempeño.

## 7. CONTENIDOS:

### UNIDAD No 1 - SISTEMAS COMPUTACIONALES ESCALABLES (6 semanas, 18 horas)

#### Competencias:

- Competencias genéricas:
  - a. todas las generales del curso
- Competencias específicas:
  - Desarrollar sistemas y aplicaciones computacionales escalables que consideran diferentes atributos de calidad teniendo en cuenta la disponibilidad de tecnologías de hardware, software, redes así como las mejores practicas de ingeniería para DevOps

### Resultados de aprendizaje:

- Especifica requisitos no funcionales relacionados con los atributos de calidad de Disponibilidad, Rendimiento y Seguridad en sistemas informáticos escalables.
- Diseña aplicaciones y sistemas computacionales que tienen en cuenta atributos de calidad como Disponibilidad, Rendimiento y Seguridad.
- Selecciona la arquitectura de sistema adecuada
- Implementa sistemas y aplicaciones computacionales escalables que consideran diferentes atributos de calidad teniendo en cuenta la disponibilidad de tecnologías de hardware, software, redes.
- Apropia mejores practicas de ingeniería para DevOps

### Temas:

#### 1.1 Conceptos fundamentales

- Escalabilidad
- Arquitectura de Sistemas
- Estilos de arquitectura
- Arquitectura Cliente/Servidor
- Arquitectura P2P
- Atributos de Calidad

#### 1.2 Despliegue de aplicaciones y servicios

- On-premise (data center)
- Cloud Computing
- Virtualización
- Contenedores, Micro-servicios, Docker, Kubernetes.
- DevOps: CI, CD, CD

#### 1.3 Diseño de aplicaciones escalables

##### 1.3.1 Disponibilidad

- Definición
- Tácticas de Disponibilidad
- Patrones de Arquitectura
- Mejores prácticas

##### 1.3.2 Rendimiento

- Definición
- Tácticas de Rendimiento
- Patrones de arquitectura
- Mejores prácticas
  - Performance Tuning
  - Cache
  - Colas

##### 1.3.3 Seguridad

- Definición
- Servicios de seguridad
  - Confidencialidad
  - Integridad

- Disponibilidad
- Autenticación
- Autorización
- No repudio
- Tácticas de seguridad

## UNIDAD No 2 - COMPUTACIÓN Y BIG DATA (6 semanas, 18 horas)

### Competencias:

- Genericas:
  - todas las de curso
- Especificas:
  - Comprende y usa la infraestructura TI para ejecutar aplicaciones Big Data.
  - Desarrolla aplicaciones distribuidas intensivas en datos basadas en Map/Reduce y Spark en diferentes lenguajes de desarrollo.

### Resultados de aprendizaje:

- Desarrolla sistemas y aplicaciones big data con las principales tecnologías actuales para el procesamiento de altos volúmenes de datos.

### Temas:

- 2.1 Infraestructura TI para Big Data
  - Hadoop y su ecosistema
  - Spark y su ecosistema
  - Computación en nube para Big Data
- 2.2 Almacenamiento Big Data
  - Arquitectura HDFS
- 2.3 Procesamiento Distribuido para Big Data
  - Map/Reduce
  - Spark
- 2.4 Bases de Datos SQL y NoSQL
- 2.5 Introducción a la Analítica de datos
  - Aprendizaje de máquina
    - Supervisado
    - No supervisado
  - Minería de texto
- 2.6 Proyectos de Big Data
  - Marcos de trabajo: ASUM-DM
  - Definición de problema
  - Disponibilidad de datos
  - ETL (Extract-Transform-Load)
  - Procesamiento
  - Aplicaciones

## UNIDAD DIDACTICA No 3 – COMPUTACIÓN DE ALTO DESEMPEÑO (4 semanas, 12 horas)

### Competencias:

- Genericas:
  - todas las de curso
- Especificas:
  - Apropia una infraestructura de TI para ejecutar aplicaciones intensivas en Computación
  - Desarrolla aplicaciones paralelas y distribuidas intensivas en cómputo ejecutando en diferentes ambientes de computo como Clusters o GPU

### Resultado de aprendizaje:

- Desarrolla sistemas y aplicaciones de alto desempeño con las principales tecnologías para clústeres de alto desempeño, tanto desde la infraestructura como desde las herramientas de desarrollo de software.

### Temas:

#### 1.1 Infraestructura

- Introducción al HPC
- Arquitecturas de Supercómputo
- Centros de Datos para HPC
- Hardware para HPC
- Software para HPC

#### 1.2 Desarrollo de aplicaciones

- Lenguajes de Programación y Compiladores para HPC
- Metodología PCAM
- Estrategias de Paralelización
- Programación en MPI
- CUDA y MIC
- Aplicaciones

## 8. ESTRATEGIAS DE ENSEÑANZA APRENDIZAJE

Se utilizan estrategias basadas en:

1. Lecturas y síntesis por parte de los estudiantes para adquirir los conocimientos fundamentales de los Ambientes de Despliegue de aplicaciones, Aplicaciones Escalables, Atributos de Calidad, Sistemas Big Data y Computación de Alto Desempeño.
2. Síntesis de temas de la materia por parte del profesor a modo de presentación. Resolución de dudas de las lecturas.
3. Laboratorios informáticos de ambientes de despliegue, aplicaciones escalables, sistemas big data y de alto desempeño.
4. Aprendizaje basado en proyectos, mediante la realización durante todo el curso, de 4 proyectos individual y en equipo, donde aplican los conocimientos, habilidades y actitudes desarrolladas en la materia.

El desarrollo de este curso combina dos perfiles: Desarrollador de Software y Desarrollador de Sistemas, los cuales se integran bajo el enfoque DevOps

Para el perfil de Desarrollador de Software, se hará énfasis en el análisis, diseño e implementación de Aplicaciones, haciendo énfasis en la etapa de Diseño e Implementación.

Para el perfil de Desarrollador de Sistemas, se hará énfasis en Diseño y arquitectura del sistema, Despliegue de Aplicaciones en Infraestructuras TI, DevOps y atributos de calidad relacionados con la Escalabilidad, Big Data y Computación de alto desempeño en entornos de data centers propios o en nube.

Estos dos perfiles son complementarios, y por ello, las actividades evaluativas contemplarán trabajo en grupo que favorezca la complementariedad que favorezca la integración de cada uno de ellos.

Se utilizará una metodología de Aprendizaje Basada en Proyectos y Aprendizaje autónomo. Esto implica que el estudiante jugará un rol muy importante en el proceso de aprendizaje.

Para la parte de fundamentación:

- Lectura de notas/presentaciones del tema por parte del estudiante.
- Síntesis por parte de los estudiantes de los conceptos utilizando diferentes herramientas como Mapas Conceptuales, Wikis y Blogs entre otros.
- Resolución de dudas por parte del profesor.

Para la parte de Desarrollo de Software:

- Requisito: Entender los conceptos fundamentales del desarrollo de software, programación, apropiación de frameworks modernos web/móvil.
- Realización de aplicaciones en cada uno de los Proyectos de la materia.

Para la parte de Desarrollo de Sistemas:

- Requisito: Entender los conceptos fundamentales y tener habilidades de organización, trabajo en equipo por proyectos, gestión de proyectos, análisis de necesidades a partir de problemáticas u oportunidades de negocio y concepción global de un sistema, como la integración de aplicaciones, software base e infraestructura de TI.
- Despliegue de las aplicaciones en diferentes infraestructuras tecnológica de acuerdo al tipo de sistema: transaccional-escalable, big data e intensivo en computación, bajo el enfoque DevOps

## 9. RECURSOS DIDÁCTICOS

Para el desempeño adecuado del curso se requiere:

- Aulas de clase, donde se realiza las presentaciones magistrales y trabajo autónomo del estudiante.
- Salas de computadores donde se realizan los laboratorios, hay espacio para el desarrollo de los proyectos y resolución de dudas.
- Data Center Académico, que provee toda la infraestructura tecnológica para la realización de los laboratorios y proyectos.

- Cloud Computing, se utiliza infraestructura en nube para apoyar laboratorios y la realización de los proyectos en las modalidades:
- IaaS: Infraestructura como un Servicio: Amazon, MS Azure, IBM Bluemix
- PaaS: Plataforma como un Servicio: Heroku, Amazon, Azure, IBM Bluemix
- APIaaS: Servicios API como un Servicio: Amazon, Azure, IBM Watson
- SaaS: Software como un servicio: Github, Travis, etc.

## 10. CRITERIOS Y PROCEDIMIENTOS DE EVALUACIÓN

### Desempeño comprensivo: (45%)

Realización de tres (3) MOOCs, uno por unidad y un Examen Final sobre fundamentos.

- (1) 10% MOOC de Introducción a DevOps
- (2) 10% MOOC de un tema seleccionado de BigData
- (3) 10% MOOC de HPC
- (4) Examen Final (15%) conceptos fundamentales de DevOps, Sistemas Escalables, Big Data y HPC

### Desempeño práctico: (55%)

Realización de cuatro (4) proyectos:

**Proyecto 1 (10%):** Trabajo individual. Desarrollo de una aplicación web utilizando un framework moderno backend y frontend, utilizando mejores prácticas DevOps para integración y entrega continua distinguiendo los diferentes ambientes: Desarrollo, Pruebas y despliegue en Infraestructura TI (data center académico) propia y en Nube (diferentes alternativas). Trabajo Individual.

**Proyecto 2 (15%):** Trabajo en equipo. Desarrollo de una aplicación web escalable utilizando un framework web moderno, utilizando mejores prácticas DevOps para integración y entrega continua distinguiendo los diferentes ambientes de desarrollo, pruebas y despliegue en una infraestructura TI propia. La infraestructura TI debe ser analizada, diseñada e instalada por el equipo de desarrollo. Considera la incorporación de atributos de calidad relacionados con los sistemas escalables como Disponibilidad, Rendimiento y Seguridad. Desempeño individual ya que por estudiante se trabaja un atributo de calidad de forma autónoma y Desempeño en grupo porque el proyecto se aborda entre varios estudiantes basados en la complementariedad de conocimientos y habilidades, abordamiento de problemas complejos, bajo restricciones en tiempo.

Condiciones:

- Se puede utilizar una de las aplicaciones desarrolladas en el Proyecto 1, o definida por el equipo.
- Cada miembro del equipo trabajará y responderá por un Atributo de Calidad y un aspecto del ciclo DevOps
- Se va a utilizar el acercamiento DevOps con Integración Continua, Entrega Continua y Despliegue continuo



**Proyecto 3 (10%):** Trabajo en equipo. Desarrollo de una aplicación de analítica en Big Data utilizando el ambiente computacional adecuado para big data. La infraestructura TI para Big Data debe ser analizada, diseñada y usada por el equipo de desarrollo. Considera un ciclo de vida de proyectos big data, la necesidad de unas técnicas básicas de analítica de datos y el desarrollo de aplicaciones en las diferentes etapas de un proyecto Big Data: ETL, Procesamiento y Aplicaciones. Utilizará infraestructura TI y/o en Nube apropiada para el desarrollo y ejecución de proyectos Big Data. Provee un desempeño grupal, basado en la complementariedad, perfiles específicos dentro del ecosistema de Big Data y restricciones en tiempo.

**Proyecto 4 (10%):** Trabajo en equipo. Desarrollo de una aplicación HPC (High-Performance Computing) o de alto desempeño, utilizando las mejores prácticas y arquitecturas para el análisis, diseño, desarrollo y evaluación de aplicaciones de alto desempeño. Dichas aplicaciones ejecutarán en una infraestructura TI basada en un clúster de alto desempeño académico. Provee un desempeño grupal, basado en la complementariedad, perfiles específicos dentro del ecosistema de HPC y restricciones en tiempo.

- Los Proyectos 2, 3 y 4 se evalúan así:
  - 50% de acuerdo a los entregables y funcional, y todos los miembros del equipo reciben la misma nota.
  - 25% Sustentación Individual, se le asigna a cada miembro del equipo de forma independiente.
  - 25% Suma Promedio de todas las Sustentaciones Individuales.

## Seguimiento: (10%)

Cada unidad tendrá un seguimiento del 5%, el cual se evaluará mediante la participación en clase, aporte a preguntas-respuestas y participación en Slack y Piazza.

## 11. BIBLIOGRAFÍA

### Sistemas Escalables:

- [1] Viktor Farcic, "The DevOps 2.0 ToolKit: Automating the Continuous Deployment Pipeline with Containerized Microservices", 2018. Leanpub. URL: <http://leanpub.com/the-devops-2-toolkit>
- [2] Len Bass; Ingo Weber; Liming Zhu, "DevOps: A Software Architect's Perspective", 2015, URL: <http://ezproxy.eafit.edu.co:2288/book/software-engineering-and-development/9780134049885>
- [3] Shailesh Kumar Shivakumar, Architecting High Performing, Scalable and Available Enterprise Web Applications, 2014  
URL: <http://proquestcombo.safaribooksonline.com.ezproxy.eafit.edu.co/book/software-engineering-and-development/enterprise/9780128022580/firstchapter>
- [4] Len Bass; Paul Clements; Rick Kazman, Software Architecture in Practice 3ed, 2015  
URL: <http://proquestcombo.safaribooksonline.com.ezproxy.eafit.edu.co/book/software-engineering-and-development/9780132942799/firstchapter>

## Big Data:

- [5] Thomas Erl; Wajid Khattak; Paul Buhler. Big Data Fundamentals: Concepts, Drivers & Techniques. 2016. Weblink en EAFIT: <http://ezproxy.eafit.edu.co:2288/book/databases/business-intelligence/9780134291185>.
- [6] Big Data: Principles and best practices of scalable real-time data systems. Nathan Marz with James Warren. 2015.
- [7] Zachary Radtka & Donald Miner. Hadoop with python. 2016.

## HPC:

- [8] Thomas Rauber and Gudula Runger, “Parallel Programming for Multicore and Cluster Systems”, 2<sup>nd</sup> ed, 2013.
- [9] Zachary Radtka & Donald Miner. Hadoop with python. 2016.
- [10] S. K. Prasad, A. Gupta, A. L. Rosenberg, A. Sussman, C. C. Weems. “Topics in Parallel and Distributed Computing: Introducing Concurrency in Undergraduate Courses”, 2015.
- [11] G. Zaccane, “Python Parallel Programming Cookbook”, 2015.