

DEPARTAMENTO ADMINISTRATIVO DE CIENCIA, TECNOLOGÍA E INNOVACIÓN - COLCIENCIAS -

CONVOCATORIA PARA LA FORMACIÓN DE CIUDADANOS EN CIENCIA DE DATOS

1. PRESENTACIÓN

EL DEPARTAMENTO ADMINISTRATIVO DE CIENCIA, TECNOLOGÍA E INNOVACIÓN – COLCIENCIAS, como ente rector del Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTel), en alianza con el Ministerio de Tecnologías de la Información y las Comunicaciones, buscan con esta convocatoria proporcionar la formación y certificación teórico-práctica de ciudadanos de datos para contribuir a la solución de problemas o necesidades reales en el ecosistema de tecnologías de la información y las comunicaciones a través de un análisis diferenciador de sus complejos volúmenes de datos que permita impulsar la competitividad del sector TIC nacional en el camino de la economía digital.

Por iniciativa del Ministerio de Tecnologías de la Información y las Comunicaciones y el Departamento Administrativo de Ciencia, Tecnología e Innovación –Colciencias-, diferentes empresas del sector público y del sector privado de Colombia se unieron con el propósito de fortalecer la generación de soluciones de análisis de información. Producto de este acuerdo nace el Centro de Excelencia y apropiación en Big Data y Data Analytics – Alianza CAOBA -, el cual tiene como objetivo generar soluciones soportadas en analítica de datos a sectores industriales, gubernamentales y académicos.

Este centro está constituido por representantes de sectores empresariales verticales, empresas TIC y algunas de las mejores universidades del país; una capacidad combinada de investigación desde las universidades que permite generar una oferta de productos innovadores en tecnología con el objetivo de generar servicios y soluciones aplicables que agreguen valor a los sectores estratégicos del país.

Dentro de los objetivos del Centro de Excelencia y Apropiación en Big Data y Data Analytics –CAOBA se encuentra aportar al fortalecimiento del ecosistema TIC nacional. En el contexto tecnológico actual los sistemas económicos mundiales transitan por un proceso de cambio que promete la transformación radical de sus estructuras y procesos; desde la producción hasta la comercialización. Esta evolución es la cuarta revolución industrial donde la fusión estructural de las TIC en todas las etapas del ciclo económico posiciona favorablemente a Colombia, por lo que se reconoce la importancia de trabajar en la consolidación de una economía digital sólida y robusta.

En el país se han identificado elementos estratégicos, hacia los cuales es necesario orientar los esfuerzos públicos para dinamizar el proceso de simbiosis entre las TIC y las etapas del ciclo económico. Uno de estos elementos estratégicos es la industria TI, líder del camino hacia la digitalización económica; desarrollando y comercializando soluciones que permitan hacer más eficientes y eficaces las etapas del ciclo productivo.

Para garantizar el fortalecimiento de las empresas TIC se identificó que es fundamental promover el equilibrio entre oferta y demanda de productos de base tecnológica. Las capacidades competitivas de las empresas para ofertar y apropiar dichas soluciones TIC, guardan directa relación con las competencias de sus equipos de trabajo; y allí el talento humano altamente capacitado en TI es prioritario. Lo anterior evidencia la necesidad de contar con equipos altamente calificados para abordar las necesidades de cierre de déficit de profesionales TI, proyectado a 2018 en 53.042 profesionales.

En ese sentido los instrumentos de política pública desarrollados por el Gobierno colombiano se orientan a fortalecer la industria TIC en el país, a través de la formación de talento humano altamente capacitado que aporte a la dinamización y crecimiento del sector.

El Plan Nacional de Desarrollo Todos Por un Nuevo País 2014-2018, resalta que es necesario fortalecer la formación de Talento TIC, para de esta manera contribuir al crecimiento y dinamización de la industria TIC en el país. Siguiendo estos lineamientos, el Plan Vive Digital para la Gente 2014-2018 establece en sus líneas estratégicas que “en el corto plazo necesitamos desarrollar competencias en los profesionales TI existentes” para lo cual se cualificará el perfil profesional de ingenieros, técnicos y tecnólogos para que responda a las necesidades de la industria TI. Las necesidades hacen converger habilidades estadísticas, matemáticas, tecnológicas, de análisis y de conocimiento de negocio, los profesionales que cuentan con este perfil son escasos en las industrias y por eso en el mediano plazo se establece el apoyo a la formación de ciudadanos de datos. Así mismo, el Plan Nacional de Ciencia, Tecnología e Innovación para el desarrollo del sector de las Tecnologías de la Información y las Comunicaciones - TIC 2017-2022, establece que se deberá “Preparar, retener y mantener actualizado el talento TIC que necesita el ecosistema.”

La ciencia de datos es fundamental para la concreción de los beneficios de la economía digital. Las oportunidades de mercado se asocian en muchos casos a la asimetría de la información. Con la simbiosis entre los procesos productivos y las TIC, las decisiones y comportamientos de los agentes, variables y objetos que intervienen en el ciclo económico quedan registradas en la red. Estos datos son la materia prima para la reducción de la asimetría y falta de información, ya que organizados y analizados son insumo para predicciones y proyecciones cada vez más precisas que ayudarán al gobierno, los consumidores, y a las firmas a reducir la incertidumbre a la hora de tomar decisiones.

Consciente de la importancia de los datos y de su potencial valor, el Ministerio de Tecnologías de la Información y las Comunicaciones ha liderado iniciativas como Datos Abiertos, para poner a disposición de los ciudadanos los datos del Gobierno colombiano. Estos esfuerzos públicos han llevado a que el país sea reconocido internacionalmente como líder en datos abiertos. Particularmente en la clasificación del país como cuarto a nivel mundial en datos abiertos según la OCDE. Pero para que el valor potencial de los datos en el ámbito público, investigaciones y negocios se materialice, es necesario que

exista talento humano capacitado para analizar y comprender los innumerables datos existentes, además de proponer soluciones a partir de los análisis realizados.

En consecuencia, la presente convocatoria es un instrumento de concreción y materialización de las metas, objetivos y compromisos del Gobierno Nacional expuesta anteriormente. Cofinanciando la formación teórico-práctica y certificación de Ciudadanos de Datos a nivel nacional para contribuir a la solución de problemas o necesidades reales en el ecosistema de tecnologías de la información y las comunicaciones a través de un análisis diferenciador de sus complejos volúmenes de datos que permita impulsar la competitividad del sector TIC nacional en el camino de la economía digital.

Nota:

1. La presente convocatoria es la anunciada en la página web de Colciencias bajo el nombre “Convocatoria para la formación especializada en Analítica de Datos”, la cual tiene como fecha de apertura diciembre de 2017.

2. OBJETIVO

2.1 General

Seleccionar 200 ciudadanos colombianos para acceder a formación teórico-práctica y certificación como *Citizen Data Scientists* para contribuir a la solución de problemas o necesidades reales en el ecosistema de tecnologías de la información y las comunicaciones.

2.2 Específicos

- Fortalecer la productividad y competitividad de la industria TIC nacional a través de la formación y certificación de 200 ciudadanos colombianos como *Citizen Data Scientists*.
- Habilitar de forma transversal las aptitudes (Análisis de datos, Tecnologías de la información, administración de proyectos, Negocios y dominio) para generar un perfil óptimo de ciudadano de datos capaz de enfrentar los nuevos retos de información no estructurada generados por la apropiación de nuevas tecnologías en la economía digital.

3. DIRIGIDA A

La presente convocatoria está dirigida para algunos de los siguientes 3 grupos de personas, para presentar su propuesta se deberá elegir uno de los tres (3) grupos descritos a continuación:

1. Profesional, tecnólogo o técnico en Ingeniería, Administración, Economía, Matemáticas, estadística, con experiencia profesional general y/o específica en tecnologías de información firmada, mínima de 2 años.
2. Profesionales en cualquiera área del conocimiento con posgrados relacionados en TI, con experiencia profesional general y/o específica en tecnologías de información firmada, mínima de 2 años.
3. Profesional, tecnólogo o técnico en cualquiera área del conocimiento con experiencia profesional mínima de 5 años en manejo de herramientas de análisis de datos como Excel avanzado, reportes o consultas en bases de datos (SQL).

Nota:

2. Los estudios deben haber sido realizados en Colombia o convalidados ante el Ministerio de Educación Nacional, para la cual deberá anexarse la resolución de convalidación.
3. Los participantes deberán elegir uno de los tres grupos descritos anteriormente para realizar la inscripción de su propuesta.

4. DEFINICIONES

Para efectos de la presente convocatoria, se adoptan las siguientes definiciones:

- **Ciudadano Colombiano experto en Ciencia de Datos:** Persona con capacidades transversales en estadística, negocio, y tecnología, capacidades que le permiten acceder a grandes cantidades de datos (Big Data) de diferentes fuentes, analizarlos y proyectarlos en el tiempo; permitiendo darle mayor valor a la información.
- **Curso de ciudadano de Datos:** Curso de actualización, formación, profundización en temáticas académicas y/o laborales con énfasis en los dominios de analítica y TI, y con conocimientos básicos en los dominios de negocio y administración de proyectos. Estos cursos tienen una duración de 70 horas.
- **Ciudadano de Datos (Citizen Data Scientist):** En palabras de Gartner, el “*citizen Data Scientist*” es una persona que crea o genera modelos que aprovechan el análisis predictivo o prescriptivo, pero cuya principal función de trabajo se encuentra fuera del ámbito de la estadística y análisis”
- **Entidad Formadora:** La Pontificia Universidad Javeriana en representación del Centro de Excelencia y Apropiación en Big Data y Data Analytics –CAOBA, será la entidad que se encargará de llevar a cabo la formación en ciudadano de datos de los beneficiarios seleccionados en el marco de la presente convocatoria.

- **Población priorizada:** Miembros de población afrocolombiana, raizal, palanquera, indígena, ROM y/o víctimas del conflicto armado.

5. REQUISITOS

A continuación, se presentan los requisitos para cada uno de los grupos mencionados en el numeral 3. Las postulaciones que cumplan con el lleno de los mismos pasarán a ser evaluadas conforme a lo establecido en el numeral 11 de la presente convocatoria.

5.1 Para profesional, tecnólogo o técnico en Ingeniería, Administración, Economía, Matemáticas, estadística o carreras afines a la ingeniería

No	REQUISITOS	DOCUMENTACIÓN REQUERIDA
5.1.1	Ser ciudadano colombiano	Fotocopia legible de la cédula de ciudadanía por ambas caras ampliada al 150%.
5.1.2	Formación	Diploma o acta de grado de la formación profesional, tecnólogo o técnico en Ingeniería, Administración, Economía, Matemáticas, estadística o carreras afines a la ingeniería. En caso de haber realizado los estudios de pregrado en el exterior, presentar la resolución de convalidación del título emitida por el Ministerio de Educación Nacional.
5.1.3	Experiencia profesional	Certificación de experiencia profesional general y/o específica en tecnologías de información firmada, mínima de 2 años, donde se indique cargo, duración y descripción de sus funciones.
5.1.4	Promedio Académico	Contar con un promedio general acumulado (PGA) de mínimo 3,5 en la escala colombiana durante el pregrado o valor equivalente en programas técnicos o tecnológicos: Adjuntar el certificado de notas que indique el PGA expedido por el área responsable de la institución donde realizó los estudios. Este debe incluir la escala de calificación del

		programa en números o su equivalente
5.1.5	Carta Presentación Postulante	Carta de motivación firmada en donde exprese los intereses y motivaciones personales para participar en la convocatoria, además deberá plantearse un problema que el beneficiario se propone abordar a partir de la ciencia de datos así como el diseño de una aplicación futura. Máximo 2 páginas
5.1.6	Carta de compromiso	Carta de compromiso donde el beneficiario se compromete a cumplir con los requisitos académicos y manifieste su interés en capacitarse y certificarse. Anexo No. 2 - Carta de compromiso.

5.2 Para profesionales en cualquiera área del conocimiento con posgrados relacionados en TI

No	REQUISITOS	DOCUMENTACIÓN REQUERIDA
5.2.1	Ser ciudadano colombiano	Fotocopia legible de la cédula de ciudadanía por ambas caras ampliada al 150%.
5.2.2	Formación	Diploma o acta de grado de la formación profesional y de la formación en posgrado relacionado con TI. En caso de haber realizado los estudios de pregrado o posgrado en el exterior, presentar la resolución de convalidación del título emitida por el Ministerio de Educación Nacional.
5.2.3	Experiencia profesional	Certificación de experiencia profesional general y/o específica en tecnologías de información firmada, mínima de 2 años, donde se indique cargo, duración y descripción de sus funciones.
5.2.4	Promedio Académico	Contar con un promedio general acumulado (PGA) de mínimo 3,5 en la escala colombiana durante el

Av. Calle 26 # 57- 83/41 Torre 8 Piso 2 – PBX: (57+1) 6258480, Ext 2081 – Línea gratuita nacional 018000914446 – Bogotá D.C. Colombia
www.COLCIENCIAS.gov.co

		pregrado o valor equivalente en programas técnicos o tecnológicos: Adjuntar el certificado de notas que indique el PGA expedido por el área responsable de la institución donde realizó los estudios. Este debe incluir la escala de calificación del programa en números o su equivalente
5.2.5	Carta Presentación Postulante	Carta de motivación firmada en donde exprese los intereses y motivaciones personales para participar en la convocatoria, además deberá plantearse un problema que el beneficiario se propone abordar a partir de la ciencia de datos así como el diseño de una aplicación futura. Máximo 2 páginas
5.2.6	Carta de compromiso	Carta de compromiso donde el beneficiario se compromete a cumplir con los requisitos académicos y manifieste su interés en capacitarse y certificarse. Anexo No. 2 - carta de compromiso.

5.3 Para profesional, tecnólogo o técnico en cualquiera área del conocimiento con experiencia profesional mínima de 5 años en manejo de herramientas de análisis de datos como Excel avanzado, reportes o consultas en bases de datos (SQL).

No	REQUISITOS	DOCUMENTACIÓN REQUERIDA
5.3.1	Ser ciudadano colombiano	Fotocopia legible de la cédula de ciudadanía por ambas caras ampliada al 150%.
5.3.2	Formación	Diploma o acta de grado de la formación profesional, tecnólogo o técnico en cualquiera área del conocimiento. En caso de haber realizado los estudios de pregrado en el exterior, presentar la resolución de convalidación del título emitida por el Ministerio de Educación Nacional.

5.3.3	Experiencia profesional	Certificación de experiencia profesional mínima de 5 años en tecnologías de información, firmada, donde se indique cargo, duración y describa sus funciones, indicando su participación en proyectos o procesos que involucren el análisis de datos o el uso de herramientas de análisis de datos, consultas o generación de reportes.
5.3.4	Promedio Académico	Contar con un promedio general acumulado (PGA) de mínimo 3,5 en la escala colombiana durante el pregrado o valor equivalente en programas técnicos o tecnológicos: Adjuntar el certificado de notas que indique el PGA expedido por el área responsable de la institución donde realizó los estudios. Este debe incluir la escala de calificación del programa en números o su equivalente
5.3.5	Carta Presentación Postulante	Carta de motivación firmada en donde exprese los intereses y motivaciones personales para participar en la convocatoria, además deberá plantearse un problema que el beneficiario se propone abordar a partir de la ciencia de datos así como el diseño de una aplicación futura. Máximo 2 páginas
5.3.6	Carta de compromiso	Carta de compromiso donde el beneficiario se compromete a cumplir con los requisitos académicos y manifieste su interés en capacitarse y certificarse. Anexo No. 2 - Carta de compromiso.

5.4 Aquellos candidatos que sean miembros de población afrocolombiana, raizal, palenquera, indígena, ROM y/o víctimas del conflicto armado, pueden acreditar dicha condición de la siguiente manera según el caso:

- Presentar una certificación expedida por el Ministerio del Interior, la cual indique que el candidato es miembro reconocido de la población

afrocolombiana, raizal o palenquera, indígena o ROM.

- Presentar una certificación expedida por la Unidad de Víctimas, la cual indique que el candidato es víctima del conflicto armado.

COLCIENCIAS, dando cumplimiento a lo indicado en el numeral 5 del artículo 4 de la Ley 1286 de 2009, garantizando los principios del mérito y la calidad, establece una etapa de ajustes de requisitos que se realizará en el periodo señalado en el Numeral 15. Cronograma.

MECANISMO DEL PROCESO DE SUBSANACIÓN DE REQUISITOS

- Surtido el proceso de revisión de requisitos, COLCIENCIAS habilitará el aplicativo SIGP a las postulaciones que se consideran deben subsanar algún(os) requisito(s) para poder continuar el proceso.
- El documento que debe ser subsanado tendrá la frase “No Cumple” y a su lado la respectiva observación con los motivos de incumplimiento.
- Los postulantes deberán adjuntar al aplicativo el o los documento(s) corregido(s) dentro del periodo y horario establecido para ello.

Nota:

1. COLCIENCIAS y Ministerio TIC podrá solicitar en cualquier momento, información y documentación adicional, complementaria o aclaraciones de los documentos entregados.
2. Los candidatos deberán cumplir con la totalidad de los requisitos para que su postulación sea evaluada, en caso contrario no continuarán al proceso de evaluación.
3. Es responsabilidad de los candidatos consultar en el aplicativo SIGP los documentos que deben ser subsanados dentro del periodo y horario establecido para ello de acuerdo a lo establecido en el numeral 15. CRONOGRAMA.
4. En caso de que se encuentre algún tipo de alteración a los documentos presentados para participar en la convocatoria, el candidato será retirado del proceso de selección y se pondrá en conocimiento de las autoridades competentes la situación presentada.

6. CONDICIONES INHABILITANTES.

- Cada beneficiario podrá presentar sólo una propuesta en cada cohorte en la presente convocatoria. COLCIENCIAS verificará si el beneficiario presenta más de una propuesta; si eso ocurre se anularán una de las solicitudes presentadas, que para este caso será la segunda propuesta presentada.
- Ciudadanos que no sean colombianos.

7. DURACIÓN Y FINANCIACIÓN

Se cofinanciará la formación y certificación de al menos 200 candidatos seleccionados en la presente convocatoria. La fecha de inicio de la formación no podrá ser posterior al tercer trimestre de 2018.

El valor total del curso es de cuatro millones cuatrocientos mil pesos moneda corriente (\$4.400.000), de los cuales Colciencias – Ministerio TIC condonaran cuatro millones de pesos moneda corriente (\$4.000.000) y el candidato deberá aportar un valor de cuatrocientos mil pesos moneda corriente (\$ 400.000). Los requisitos de condonación se describen a continuación:

- El participante deberá presentar y aprobar el examen de certificación que se realizará al finalizar el curso.
- Este examen contempla los temas abordados en los diferentes módulos y el participante tendrá la oportunidad de presentarlo hasta un máximo de tres (3) veces.
- Esta condonación tendrá un valor de \$4.000.000 pesos M/CTE.

Una vez publicada la lista de elegibles, los beneficiarios deben realizar el pago, a la Pontificia Universidad Javeriana, de cuatrocientos mil pesos (\$400.000) M/CTE, no reembolsables, de manera previa al inicio del curso. La universidad remitirá a las personas de la lista de elegibles el instructivo correspondiente para llevar a cabo este proceso.

Este valor no obtendrá los beneficios regulares de descuento de Educación Continua.

El participante deberá suscribir un pagaré a favor de la Pontificia Universidad en representación del Centro de Excelencia en Big Data y Data Analytics (CAOBA) por un valor de cuatro millones de pesos (\$4.000.000), el cual se hará efectivo en caso de no lograr las certificaciones de asistencia y/o no presentar la prueba de suficiencia correspondiente al valor cofinanciado por el Fondo Francisco José de Caldas - FFJC, salvo que la inasistencia obedezca a motivos de fuerza mayor o caso fortuito, debidamente justificados a la Pontificia Universidad Javeriana, en representación del Centro de Excelencia y Apropriación en Big Data y Data Analytics –CAOBA

8: CONTENIDOS DEL CURSO

- La convocatoria incluirá la formación y certificación de al menos 200 personas en ciudadano de datos, siguiendo como lineamiento las tres esferas de competencia de este perfil y acorde con el perfil "*Citizen Data Scientist*". Ver documento anexo No. 1.

- El programa de formación a ofrecer a los beneficiarios en el marco de la presente convocatoria, será impartido por el Centro de Excelencia y Apropiación en Big Data y Data Analytics –CAOBA, representado por la Pontificia Universidad Javeriana, quien expedirá los certificados a otorgar a los beneficiarios.
- Para el desarrollo del programa de formación y certificación a ofrecer a los beneficiarios en el marco de la presente convocatoria, la Pontificia Universidad Javeriana en representación del Centro de Excelencia y Apropiación en Big Data y Data Analytics, suscribió el contrato de prestación de servicios No. FP44842-466-2017 celebrado entre Fiduciaria la Previsora S.A. – Fiduprevisora S.A.- actuando como vocera y administradora del Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación – Fondo Francisco José de Caldas.
- Los cursos se ofrecerán de manera presencial en las siguientes ciudades:
 - Bogotá D.C.
 - Medellín
 - Cali
 - Manizales
 - Barranquilla
 - Bucaramanga

Para lo cual, la Pontificia Universidad Javeriana en representación del Centro de Excelencia y Apropiación en Big Data y Data Analytics –CAOBA otorgará a los beneficiarios seleccionados en la presente convocatoria, las condiciones óptimas para desarrollar el programa en cada una de estas ciudades, y todo el apoyo logístico necesario para su desempeño y formación (Materiales, salones, conexión a internet, refrigerios)

Los beneficiarios de la presente convocatoria deberán asumir los costos de desplazamiento a las ciudades donde se ofrecerán los programas de formación.

Para la apertura de los cursos es necesario contar con mínimo quince (15) participantes. En la eventualidad en la que un participante sea elegido y su ciudad de residencia no cuente con el mínimo requerido, el participante podrá conservar su cupo y asistir a las clases en otra ciudad, para lo cual deberá asumir sus costos de traslado, manutención y alojamiento.

- El curso Ciudadano de Datos tiene una duración de siete semanas no consecutivas cuyas fechas específicas serán proporcionados por La Pontificia Universidad Javeriana en representación del Centro de Excelencia y Apropiación en Big Data y Data Analytics –CAOBA luego del cierre de la convocatoria, una vez sean seleccionados los aspirantes aprobados. Se dará inicio a los cursos entre el primer,

segundo y tercer semestre de 2018 dependiendo de los resultados de los cohortes y serán dictados los días jueves y viernes de 6:00pm a 9:00pm y los sábados de 8:00am a 12:00m. Las fechas estarán sujetas a los resultados de la convocatoria por número de beneficiarios y por ciudad.

- **Certificados**

1. La Pontificia Universidad Javeriana en representación del Centro de Excelencia en Big Data y Data Analytics (CAOBA), a través de la Dirección de Educación Continua, entregará un certificado de asistencia a aquellos participantes que cumplan con el 80% de las horas programadas del total del curso.
 2. El control de asistencia será realizado por la Pontificia Universidad Javeriana, en representación del Centro de Excelencia en Big Data y Data Analytics (CAOBA), mediante la firma de listados de ingreso y salida. Los casos de inasistencia por situaciones de fuerza mayor, si se presentan, serán resueltos en cada caso por la Universidad Ejecutora y el participante.
2. La Certificación como ciudadano de datos (Citizen Data Scientist), la cual será expedida por la Pontificia Universidad Javeriana en representación del Centro de Excelencia en Big Data y Data Analytics (CAOBA) aprobada la prueba de suficiencia de acuerdo con los parámetros que se establezcan para ello.

Es preciso aclarar que este tipo de certificación no se podrá utilizar para realizar procesos de homologación en programas conducentes a título.

Estructura del Curso de *Citizen Data Scientist*

NOMBRE:	<i>CITIZEN DATA SCIENTIST – CAOBA</i>
NIVEL BLOOM:	APLICAR
OBJETIVO GENERAL	El objetivo del curso es fortalecer en el participante su capacidad de participar en la concepción y el diseño de soluciones a problemas de negocio a partir de la aplicación de modelos de analítica fundamentados en técnicas, herramientas, lenguajes y roles organizacionales apropiados según las restricciones y características de la empresa.

OBJETIVOS EN FORMACIÓN DE COMPETENCIAS POR DOMINIO	<p>Al finalizar el curso los participantes podrán:</p> <p>Negocio: identificar problemas susceptibles de ser resueltos con analítica, en el marco del área de negocio que le corresponda al participante, y cuyas soluciones le generen valor al negocio. Adicionalmente, el participante entenderá metodologías de gestión y desarrollo de proyectos de analítica.</p> <p>Analítica: entender y tener capacidad para seleccionar y aplicar técnicas de limpieza, preparación, análisis, modelamiento a partir de conjuntos de datos, para construir soluciones de analítica descriptiva y predictiva. También, entender estadísticamente el alcance y la limitación de dichas técnicas y de los modelos construidos.</p> <p>Tecnologías de información: aplicar una metodología de analítica para desarrollar fases del ciclo de vida de la analítica de datos utilizando distintas herramientas que permitan realizar procesos de extracción, almacenamiento, perfilamiento, preparación de datos y comunicación de los datos.</p>
OBJETIVOS EN FORMACIÓN DE COMPETENCIAS TRANSVERSALES	<p>Habilidades transversales: fortalecer las capacidades de trabajo en equipo y comunicación de resultados de proyectos mediante la visualización de datos y de resultados de proyectos de analítica.</p>
DURACIÓN:	70 horas presenciales

METODOLOGÍA:	<p>Cada módulo se realizará en varias sesiones que incluirán un componente teórico y un componente práctico. Durante el desarrollo del curso se utilizarán diferentes herramientas y lenguajes que permitirán al participante poner en práctica los conceptos teóricos abordados. El curso no busca certificar en una tecnología particular sino en los conceptos, metodologías y habilidades necesarios para desempeñar el rol de citizen data scientist en una organización.</p> <p>Algunas las Guías y textos de referencias se encuentran en inglés. Se recomienda que el aspirante tenga un buen nivel de comprensión de lectura en este idioma.</p>
REQUISITOS DE INGRESO:	<ul style="list-style-type: none"> • Para poder aprovechar todo el contenido de este curso es necesario que el participante demuestre experiencia o formación en lenguaje SQL y/o herramientas para el análisis de datos como Excel, así como conocimientos en estadística básica. • Cada participante deberá contar con computador portátil para el desarrollo del curso.

Contenido del Programa

MÓDULOS	
MÓDULO 1: 8 horas	Analítica en las organizaciones Temas: <ul style="list-style-type: none"> • Conceptos: minería de datos, analítica de datos, ciencia de datos, Big Data • Fuentes de datos: las grandes V's • Roles organizacionales en la formulación y ejecución de estrategia analítica • Generación de valor en la organización: <ul style="list-style-type: none"> ○ Ideación y planeación estratégica para la identificación de oportunidades ○ Modelos de beneficio ○ Toma de decisiones basada en analítica

MÓDULO 2: 20 horas	Metodologías para desarrollo de proyectos de analítica y ciclo de vida de los datos Temas: <ul style="list-style-type: none"> • De CRISP-DM a ASUM-DM • Ciclo de vida de los datos <ul style="list-style-type: none"> ○ Exploración y visualización de datos ○ Análisis de calidad de datos ○ Herramientas de ETL ○ Métodos de preparación de datos
MÓDULO 3: 32 horas	Construcción de modelos de analítica utilizando herramientas computacionales Temas: <ul style="list-style-type: none"> • Principios de estadística fundamentales para la analítica • Aprendizaje supervisado • Aprendizaje no supervisado • Visualización de datos
MÓDULO 4: 10 horas	Plataformas TI para almacenamiento y procesamiento de grandes volúmenes de datos Temas: <ul style="list-style-type: none"> • Bases de datos (SQL / NoSQL) • Herramientas de procesamiento de datos en Hadoop
Total: 70 horas	

9. PROCEDIMIENTO DE INSCRIPCIÓN

9.1 Ingresar a la página web de Colciencias (www.colciencias.gov.co).

9.2 Ingresar al menú ubicado en la parte superior de la página web – SCIENTI Y SIGP, donde se despliega Sistemas de Información- Plataforma SIGP, allí encontrará disponibles los diferentes formularios en línea para aplicar a las convocatorias de Colciencias. (<http://www.colciencias.gov.co/scienti>).

9.3 Hacer clic en la opción Formulario en línea para registro de entidades y/o persona natural ubicado en la parte derecha, <http://formulariosigp.colciencias.gov.co:7001/EntidadesColciencias/>. En caso de haberla registrado previamente, continuar con el paso número 9.5.

9.4 Realizar el pre-registro de la persona natural con la información solicitada. Colciencias en un periodo de 2 días hábiles remitirá a su correo electrónico de contacto las credenciales de acceso para el ingreso al sistema de registro de entidades y/o persona natural, solo si la persona natural fue avalada, de lo contrario el sistema enviará un correo informando que la persona natural fue rechazada y la razón del

- rechazo.
- 9.5 Ingresar al módulo de entidades y/o persona natural, con sus credenciales y registrar la información de la persona natural. En caso de encontrarse registrada la persona natural en este sistema, no se requiere de un nuevo registro, sin embargo podrá actualizar la información.
- 9.6 Una vez finalizado el registro de la persona natural, Ingresar al link de Formulario de Proyectos y/o programas:
<http://formulariosigp.colciencias.gov.co:7001/FormularioProyectos/>
- 9.7 Diligenciar la información solicitada por el formulario en línea para el registro de inscripción del curso de FORMACIÓN DE CIUDADANOS EN CIENCIA DE DATOS. Para recibir las credenciales de acceso por correo electrónico deberá registrar como mínimo el título, (en cuyo campo se debe seleccionar el nombre de la persona natural que previamente registró en el formulario de entidades y/o persona natural), la entidad (persona natural), convocatoria, y correo electrónico.
- 9.8 Debe adjuntar la totalidad de los documentos que soportan los requisitos. Si para un requisito requiere anexar más de un documento, guárdelos en una sola carpeta y adjúntela como archivo comprimido .ZIP o .RAR en el campo correspondiente al requisito. Si requiere cambiar algún archivo de los que previamente adjuntó, puede hacerlo examinando y adjuntando el documento deseado. Es importante resaltar que el tamaño de cada archivo anexado de cada uno de los requisitos, cuenta con una capacidad máxima de 3MB; para el requisito "Otros adjuntos" tiene una capacidad máxima de 10 MB.
- 9.9 Para validar, hacer clic en el botón validar ubicado en la parte inferior del menú. Si la información registrada cumple con los criterios solicitados por la convocatoria, la validación será exitosa y se activará la opción de enviar formulario, en caso contrario se mostrará en pantalla la información que no cumple con los criterios definidos en la convocatoria.
- 9.10 Una vez se oprime el botón de enviar formulario, el sistema registrará la solicitud y generará su respectivo código de registro que corresponde al radicado de la solicitud.
- 9.11 Una vez haya sido enviada la solicitud, si la convocatoria se encuentra abierta el sistema le permite modificarla, oprimiendo el botón "Modificar Proyecto", para realizar los cambios que consideren. Al dar clic en la opción "Modificar Proyecto" que está en el menú se muestra una pantalla de confirmación con el siguiente mensaje "¿Está seguro de modificar el proyecto?, una vez confirme esta acción el proyecto ya no estará radicado en Colciencias. Para volver a radicarlo debe validar y enviar de nuevo el proyecto una vez lo haya modificado." Cabe aclarar que si no lo vuelve a validar y enviar a Colciencias, esta solicitud no quedará radicada oficialmente en Colciencias y no seguirá a la etapa de revisión de requisitos, así cuente con la pantalla de evidencia de un envío anterior satisfactorio.

9.12 Para realizar seguimiento a la propuesta, deberá ingresar al formulario con el usuario y contraseña enviados mediante el correo electrónico al momento de registrar la propuesta en el sistema, y por el menú de requisitos podrá consultar las observaciones realizadas a los requisitos de la solicitud, con el fin de que sean subsanadas por el proponente.

NOTAS:

1. Se aceptarán únicamente las solicitudes que se presenten a través del formulario en línea con toda la información solicitada en la presente convocatoria.
2. No se tendrá en cuenta para el proceso de evaluación y selección, la información enviada en medios distintos al SIGP (correo postal, fax, correo electrónico otro), ni posterior a la fecha y hora límite establecida. Excepto cuando sea solicitado directamente por Colciencias.
3. Si requieren anexar documentación adicional del proyecto como: tablas, fórmulas, gráficas, anexos, cartas, etc., guárdelos en una sola carpeta y adjúntelas como archivo .ZIP o .RAR (archivo comprimido), en el requisito de Otros adjuntos. Si requieren cambiar algún archivo de los que previamente adjuntó, puede hacerlo examinando y adjuntando el documento deseado.
4. Para remitir toda la documentación relacionada con cada uno de los requisitos, el tamaño máximo de envío es de 3 MB, en caso de requerir información adicional al proyecto, puede utilizar la opción llamada "Otros adjuntos" y contará con un tamaño máximo de 10 Mb.
5. En caso de que se presenten errores en el aplicativo SIGP durante el proceso de inscripción y envío de la solicitud o del cargue de la documentación soporte, y por esta razón después de efectuados varios intentos no pueda enviarse la solicitud de forma normal dentro del plazo establecido, capturar las pantallas con el mensaje de error junto con la fecha y hora en que se presentó el inconveniente que soporte el problema. Remitir antes de la fecha de cierre establecida en el cronograma de los términos de referencia, todas las evidencias de la falla presentada, diligenciando el formulario de PQRDS, que se encuentra en el siguiente link: <http://www.colciencias.gov.co/ciudadano/canal-pqrds>, explicando en el campo "detalle de la solicitud" los inconvenientes ocurridos durante el envío y anexando en el formulario las pantallas con las evidencias del problema que se le presentó. COLCIENCIAS basado en las evidencias verificará el error reportado y le dará respuesta a la solicitud.

10. AUTORIZACION USO DE DATOS PERSONALES

Para todos los efectos legales, la presentación, inscripción o registro de la propuesta, de forma inequívoca y exclusivamente para los fines de la presente convocatoria implica obligatoriamente la autorización expresa del titular de información, para el tratamiento de sus datos personales por parte del Departamento Administrativo de Ciencias, Tecnología e Innovación – COLCIENCIAS y la Pontificia Universidad Javeriana en representación del el Centro de Excelencia y Apropiación Alianza CAOBA, Quien presente, inscriba o registre

una propuesta como persona natural o la persona jurídica en cuyo nombre se presenta la propuesta declara que ha recibido autorización expresa de todas las personas naturales y/o jurídicas vinculadas a esta propuesta y a esta propuesta, para suministrar las informaciones a que hace referencia la ley de manejo de datos, comprometiéndose a responder ante COLCIENCIAS, por cualquier demanda, litigio presente o eventual, reclamación judicial o extrajudicial, formulada por cualquiera de las personas naturales o jurídicas vinculadas a la propuestas y/o el proyecto.

11 CRITERIOS DE EVALUACIÓN

Los beneficiarios inscritos dentro de los plazos establecidos para la presente convocatoria y que cumplan con la totalidad de los requisitos se someterán a los siguientes criterios de evaluación:

No.	Criterios	Descripción del criterio	Asignación del puntaje		Puntaje Máximo
1	Formación	Promedio académico	Mayor a 3,5 y menor o igual a 3.7	10	20
			Mayor a 3.7 y menor o igual a 4.3	15	
			Mayor a 4,3	20	
2	Experiencia	Se evaluará la experiencia profesional adicional a la mínima requerida para los 3 grupos de personas definidas en el numeral No. 3	Experiencia General adicional a la mínima requerida entre 6 y 24 meses	10	20
			Experiencia General adicional a la mínima requerida mayor de 2 años	15	

			Experiencia laboral adicional a la mínima requerida específicamente relacionada con TI mayor a 2 años.	20	
3	Carta de Motivación	Se otorgarán 10 puntos a criterio de los evaluadores. Se evaluará la carta de motivación de cada candidato calificando la afinidad de los intereses y capacidades del beneficiario en relación con el curso de formación en ciudadano de datos ofertado.		10	

4	Evaluación	<p>Los postulantes deberán participar en un proceso evaluativo, en donde se determinarán sus aptitudes y potencial para acceder al curso de formación de ciudadano de datos.</p> <p>Una vez cerrada la convocatoria, aquellos participantes que cumplan con los requisitos establecidos en la presente convocatoria, serán notificados a su correo electrónico registrado en la propuesta para que presenten el examen de conocimiento que se realizará virtualmente en las fechas establecidas en el Numeral No. 15 CRONOGRAMA.</p> <p>Cada participante registrado solo tendrá una oportunidad para presentar el examen que tendrá una duración estimada de dos horas y cuyo objetivo es garantizar que los participantes cuentan con los conocimientos básicos para acceder al curso. El correo de notificación para presentación de la prueba tendrá todas las indicaciones pertinentes de la misma.</p>	40
6	Certificación de participación en programas del MinTIC	<p>Los postulantes que hayan participado como personas naturales o a través de una personería jurídica de alguno de los programas (APPS y Talento TI) del MinTIC pueden anexar una carta certificando su participación en estos programas. Puede utilizar la opción llamada "Otros adjuntos" que contará con un tamaño máximo de 10 Mb para cargar el certificado correspondiente.</p>	5
7	Población priorizada	<p>Miembros de población afrocolombiana, raizal, palenquera, indígena, ROM, en situación de discapacidad y/o víctimas del conflicto armado.</p>	5

Notas:

1. La evaluación será realizada por evaluadores asignados La Pontificia Universidad Javeriana en representación del Centro de Excelencia y Apropiación en Big Data y Data Analytics –CAOBA
2. Para el caso de la certificación de participación en programas del MinTIC, esta deberá adjuntarse al momento de presentar la propuesta en formulario dispuesto para la convocatoria a través del SIGP.
3. El criterio de “Evaluación” cuenta con un puntaje de 40 puntos, por ende se deberá obtener una buena calificación en este criterio, ya que en caso de ser bajo el puntaje, no será suficiente para alcanzar la calificación mínima del 70 así en los otros criterios se llegase a cumplir con el puntaje máximo definido para cada uno.

12. PROCEDIMIENTO DE EVALUACIÓN

Los candidatos inscritos dentro de los plazos establecidos para la presente convocatoria y que cumplan con los requisitos, serán evaluados por la Pontificia Universidad Javeriana en representación del Centro de Excelencia y Apropiación en Big Data y Data Analytics – CAOBA teniendo en cuenta los criterios de evaluación descritos el numeral 11 CRITERIOS DE EVALUACIÓN.

Toda información proporcionada es de carácter confidencial y no será utilizada para ningún fin diferente a la realización de la evaluación. Los expertos evaluadores estarán cobijados por cláusulas de confidencialidad y de no conflicto de interés.

De acuerdo con el CRONOGRAMA de la convocatoria, se concederá un término para solicitar aclaraciones y/o modificaciones a la publicación preliminar del banco de elegibles, una vez éstas sean resueltas, se procederá a la publicación del banco definitivo de propuestas elegibles.

En caso de empate, se seleccionará en primera posición, aquel postulante que presente un mayor puntaje en el criterio de evaluación No. 4 (**Evaluación**). Si el empate persiste, se tendrá en cuenta el mayor puntaje en el criterio No. 2 (**Experiencia**). Si aún continúa existiendo empate, se tendrá en cuenta aquel postulante que presente un mayor puntaje en el criterio de evaluación No.1 (**formación**). De mantenerse el empate se tendrá en cuenta aquel postulante que presente un mayor puntaje en el criterio de evaluación No. 3 (**Carta de motivación**). De mantenerse el empate se tendrá en cuenta aquel postulante que presente un mayor puntaje en el criterio de evaluación No. 5 (**Población priorizada**). De mantenerse el empate se tendrá en cuenta aquel postulante que presente un mayor puntaje en el criterio de evaluación No. 6 (**Certificación de participación en programas del MinTIC**). Si agotados estos criterios de desempate aún persiste la situación de empate, se dará prioridad a la primera postulación presentada.

Nota: Los procesos de evaluación se harán una vez cerrado cada uno de las cohortes establecidas en la presente convocatoria.

13. BANCO DE ELEGIBLES

Las propuestas que superen 70 puntos en la evaluación ingresarán al banco de elegibles y se asignarán los cupos hasta su agotamiento en estricto orden descendente.

Los resultados de la convocatoria se publicarán en la página web de Colciencias (<http://www.colciencias.gov.co>) en las fechas establecidas en el Cronograma.

La inclusión de un candidato en el banco de elegibles, no implica obligación ni compromiso alguno de COLCIENCIAS, Ministerio TIC, la Pontificia Universidad Javeriana en representación del Centro de Excelencia y Apropiación en Big Data y Data Analytics – CAOBA, de asignar recursos, ni genera derecho a recibir apoyos económicos para quienes hayan presentado las aplicaciones correspondientes.

Notas:

1. Surtida la publicación definitiva del banco de elegibles, COLCIENCIAS informará mediante comunicación escrita y/o electrónica a los beneficiarios que hayan sido seleccionados para ser formados, indicando las instrucciones y condiciones para iniciar el proceso de formación con la Pontificia Universidad Javeriana en representación del Centro de Excelencia y Apropiación en Big Data y Data Analytics – CAOBA
2. A partir de la fecha de comunicación, se contarán cinco (5) días hábiles como plazo máximo para que el beneficiario remita a COLCIENCIAS la totalidad de la documentación. Vencido el término, en caso que el beneficiario no remita la documentación, se entenderá que desiste de su interés para ser beneficiario de la presente convocatoria y se procederá a financiar la siguiente propuesta del banco definitivo de elegibles.
3. Una vez allegada dicha documentación a COLCIENCIAS se remitirá a la Pontificia Universidad Javeriana en representación del Centro de Excelencia y Apropiación en Big Data y Data Analytics – CAOBA, quienes serán los encargados de dar inicio al proceso de formación y certificación.

14. ACLARACIONES

Una vez publicados los resultados preliminares del banco de elegibles, los interesados podrán presentar solicitudes de aclaraciones y comentarios por un periodo de tres (3) días hábiles. Por fuera de este término se considera que las reclamaciones son extemporáneas.

Las peticiones y reclamaciones se deben presentar exclusivamente a través del formulario electrónico que se encuentra disponible en el enlace <http://www.colciencias.gov.co/ciudadano/canal-pqrds>, con el asunto "Convocatoria para la formación en ciencia de datos".

15. CRONOGRAMA

ACTIVIDAD	FECHA LIMITE
PRIMERA COHORTE	
Apertura de la convocatoria	22 de diciembre de 2017
Autoconsulta permanente de requisitos (si aplica)	22 de diciembre de 2017 hasta el 23 de enero de 2018 hasta las 04:00 p.m. hora colombiana
Cierre parcial de la convocatoria	30 de enero de 2018 hasta las 05:00 p.m. hora colombiana
Periodo de revisión de requisitos (si aplica)	30 de enero de 2018 hasta el 05 de febrero de 2018
Periodo de ajuste de requisitos	06 de febrero de 2018 hasta el 08 de febrero de 2018 hasta las 05:00 p.m. hora colombiana.
Presentación evaluación de conocimiento - virtual	Del 15 al 18 de febrero de 2018
Publicación del banco preliminar de propuestas elegibles	21 de febrero de 2018
Período de solicitud de aclaraciones del banco preliminar de elegibles	Del 21 al 23 de febrero de 2018
Respuesta a solicitud de aclaraciones	26 de febrero de 2018
Publicación del banco definitivo de propuestas elegibles	28 de febrero de 2018
SEGUNDA COHORTE	
Apertura de la convocatoria	31 de enero de 2018
Autoconsulta permanente de requisitos (si aplica)	31 de enero de 2018 hasta el 23 de febrero de 2018 hasta las 04:00 p.m. hora colombiana
Cierre de la convocatoria	02 de marzo de 2018 hasta las 04:00 p.m. hora colombiana
Periodo de revisión de requisitos (si aplica)	02 de marzo de 2018 hasta el 08 de marzo de 2018
Periodo de ajuste de requisitos	09 de marzo de 2018 hasta el 13 de marzo de 2018 hasta las 05:00 p.m. hora colombiana.
Presentación evaluación de conocimiento - virtual	Del 15 al 18 de marzo de 2018
Publicación del banco preliminar de propuestas elegibles	23 de marzo de 2018
Período de solicitud de aclaraciones del banco preliminar de elegibles	Del 26 al 28 de marzo de 2018
Respuesta a solicitud de aclaraciones	05 de abril de 2018
Publicación del banco definitivo de propuestas elegibles	12 de abril de 2018

Nota:

1. Durante el periodo de Autoconsulta permanente de requisitos, el candidato podrá revisar si sus documentos se han adjuntado correctamente al aplicativo. Posteriormente, en el periodo de revisión de requisitos, COLCIENCIAS verificará que los documentos adjuntos presentan la información requerida de forma consistente; COLCIENCIAS reportará en el SIGP las inconsistencias en la información de modo tal que el candidato pueda ajustarla a través del aplicativo durante el periodo establecido en el Cronograma. La etapa denominada Autoconsulta permanente de requisitos, no eximen de responsabilidad a los interesados en esta convocatoria de la consulta permanente de la información a fin de entregar la propuesta completa. Los beneficios otorgados por COLCIENCIAS para facilitar la presentación de las propuestas, no desplazan la responsabilidad de los aspirantes de consultar y adjuntar correctamente los documentos necesarios para ser habilitados, evaluados o seleccionados en esta convocatoria.
2. Si en la primera cohorte se cumple con la cantidad máxima de beneficiarios (200), los elegibles en la segunda cohorte quedaran en el Banco de Elegibles publicado en las fechas establecidas en el numeral 15 – Cronograma.
3. Los postulantes podrán aplicar en la segunda cohorte, en caso que no cumplan con el puntaje mínimo en la primera cohorte al cual se presentaron.

16. MODIFICACIONES

COLCIENCIAS podrá modificar el contenido de los términos de referencia, así como su cronograma, por las causas establecidas en el Instructivo Trámite de Adendas a Convocatorias de CTel (documento disponible en la página web de Colciencias, Micrositio Calidad), esto es por fuerza mayor o caso fortuito, por causas imprevisibles no atribuibles a COLCIENCIAS y por necesidad de la Entidad acordes con el interés general, garantizando que con la modificación no se vulnerará la expectativa prevista de los posibles participantes.

17. PROPIEDAD INTELECTUAL

Dado que la presente Convocatoria es exclusiva para seleccionar los beneficiarios para acceder a formación teórico-práctica y certificación como Citizen Data Scientists, no hay lugar a la generación de creaciones intelectuales, ello en virtud del objetivo planteado en la presente convocatoria.

En el evento en el cual se llegare a generar creaciones intelectuales susceptibles de protección vía propiedad intelectual en el marco de la formación y certificación objeto de la Convocatoria, la titularidad será definida entre el beneficiario y la Pontificia Universidad Javeriana en representación del Centro de Excelencia y Apropiación en Big Data y Data Analytics –CAOBA, acorde a la normatividad relacionada con propiedad intelectual vigente.

18. VEEDURÍAS CIUDADANAS

Las veedurías ciudadanas establecidas de conformidad con la Ley 850 de 2003, podrán desarrollar su actividad durante la presente convocatoria, conforme a lo estipulado en dicha normativa.

19. ANEXOS

- Anexo 1: Perfil "Citizen Data Scientist" CAOBA
- Anexo 2: Carta de compromiso

20. ACEPTACIÓN DE TÉRMINOS Y VERACIDAD

Con la inscripción, los interesados aceptan las características, requisitos y condiciones de la presente convocatoria así como lo dispuesto en los presentes términos de referencia para el desarrollo de la misma y la entrega del recurso. Una vez presentada la propuesta no será posible alegar desconocimiento de lo escrito en estos términos de referencia ni de sus anexos. El momento para referirse a lo establecido en los términos de referencia de la convocatoria y sus anexos es durante la etapa de reclamaciones y antes de publicación del banco preliminar.

De igual forma declaran que la información suministrada es veraz y corresponde a la realidad. En caso de encontrarse alguna incoherencia y/o inconsistencia en la información o documentación suministrada, COLCIENCIAS podrá en cualquier momento rechazar la propuesta o si es del caso declarar la pérdida del beneficio, sin perjuicio de las acciones legales correspondientes.

21. MAYOR INFORMACIÓN**DEPARTAMENTO ADMINISTRATIVO DE CIENCIA, TECNOLOGÍA E INNOVACIÓN
- COLCIENCIAS -**

Av. Calle 26 N° 57- 83 / 41 Torre 8 Piso 2, Bogotá D.C.
<http://www.colciencias.gov.co>

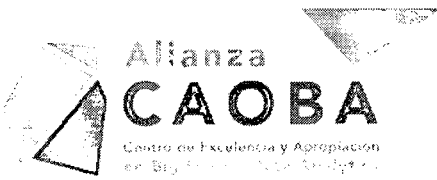
Centro de Contacto
Teléfono: (+57 - 1) 6258480 Extensión 2081
Línea gratuita nacional: 018000 914446
Horario: Lunes a jueves de 8:00 a.m. a 5:00 p.m. y viernes de 7:00 a.m. a 4:00 p.m.
(jornada continua).

En caso de inquietudes o comentarios sobre la presente convocatoria, favor diligenciar el formulario electrónico que se encuentra disponible en el enlace <http://www.colciencias.gov.co/ciudadano/canal-pqrds>, con el asunto "Convocatoria formación en ciencia de datos – 2017".

V.o Bo. Director Técnico
Vo Bo. Secretaría General



CÉSAR OCAMPO
Director General



Perfil Alianza Caoba

Reporte técnico

CAOBA - Universidad de los Andes
ACUA-M-R-PerfilCaoba-P001
Julio 2017

Índice

1. Introducción	3
2. Problemática	4
2.1. Características de los proyectos de <i>Big data analytics</i> y <i>data Analytics</i>	4
2.2. Perfiles existentes	5
2.3. Necesidad de un nuevo perfil	6
3. Antecedentes	6
3.1. Otras propuestas de perfiles	6
3.2. Metodología de ASUM de IBM	7
3.3. Taxonomía de Bloom	9
3.4. Conclusión	10
4. Marco conceptual para el perfil Big data y Data Analytics	11
4.1. Fundamentos del marco conceptual	12
4.2. Formulación matemática del perfil	13
4.3. Perfil Núcleo - P0	14
4.3.1. El núcleo del negocio	15
4.3.2. El núcleo en TI	15
4.3.3. El núcleo en analítica	16
4.3.4. El núcleo en administración de proyectos	16
4.3.5. El núcleo transversal	16
4.4. Perfil de Negocio - P1	16
4.5. Perfil en TI - P2	17
4.6. Perfil en Analítica - P3	17
4.7. Perfil en Administración de proyectos - P4	18

5. Validación de la propuesta de los perfiles CAOBA	20
6. Formas de representación de los perfiles CAOBA	21
6.1. Estrella de Bloom	21
6.2. Los perfiles profesionales según la estrella de Bloom	21
6.3. Los perfiles Caoba según la estrella de Bloom	24
6.3.1. Perfil Núcleo - P0	24
6.3.2. Perfil de Negocio - P1	25
6.3.3. Perfil en TI - P2	26
6.3.4. Perfil en Analítica - P3	27
6.3.5. Perfil en Administración de proyectos - P4	28
7. Uso del Perfil	29
7.1. Oferta de formación académica	29
7.1.1. Cursos ofrecidos por la Universidad de los Andes	29
7.1.2. Programas y Cursos ofrecidos por la Pontificia Universidad Javeriana	35
7.1.3. Programas y Cursos ofrecidos por la Universidad ICESI	37
7.2. Perfil para la demanda académica	38
8. Conclusiones	39

Cuadro 1: Habilidades técnicas

Habilidades o capacidades técnicas	Más importantes para las empresas	Las más difíciles de encontrar
Analytics/Business Intelligence/Big Data/Data Scientist	47.1 %	46.1 %
Data/Information Architecture	28.5 %	26 %
Functional Area Knowledge	23.8 %	24.8 %
Enterprise Architect	22.3 %	23.2 %
Application/Solution Architecture	23.2 %	21.4 %

1. Introducción

Actualmente el mundo experimenta grandes cambios con el avance de las tecnologías de la información y la generación continua de datos. Esto ha creado nuevas oportunidades para los negocios en temáticas como la analítica de datos (DA) y *Big data* (BD), para las cuales se requieren profesionales con conocimientos y habilidades específicas, aplicados a determinados campos de acción.

La analítica vista como los procesos de transformación de datos en información, ha permeado la sociedad y está cambiando la forma de hacer negocios, la interacción entre personas, empresas, entre otras. En este nuevo contexto, el trabajo en equipo es un factor clave que ayuda en la efectividad de los proyectos.

El desarrollo de proyectos de analítica, en todas sus formas (desde la descriptiva hasta la prescriptiva), es un reto y una tendencia a nivel mundial, donde las habilidades que se requieren para estos perfiles son cada vez más difíciles de encontrar. Esto se evidencia en el estudio realizado por SIM International & Leon Kappelman [10], en el cual se preguntó a 785 empresas acerca de sus necesidades en el tema y la dificultad para encontrar esos recursos. El cuadro 1 muestra los porcentajes más relevantes, ordenado por su dificultad. Para el caso de *Analytics/Business Intelligence/Big data/Data Scientist* se observa que es considerada para el 47 % de las empresas como una de las habilidades técnicas más importantes y en un 46 % difícil de encontrar en los perfiles profesionales. En contraposición por ejemplo, las habilidades en *Application/Solution Architecture* son importantes para el 23.2 % de las empresas, pero tienen sólo un 21.4 % de dificultad para encontrarlas en el mercado laboral.

Así mismo, en el cuadro 2 se observa un listado de habilidades transversales que son requeridas en los perfiles que se están buscando y la dificultad para encontrarlas en el mercado, ordenado por la dificultad. Para este caso, el 45.2 % de las empresas consideran el liderazgo como una de las más importantes y con un 43.8 % de las más difíciles de encontrar; un 17.2 % de las empresas consideran la comunicación escrita como una de las menos importantes entre el listado, demostrando así las preferencias que tienen sobre ciertas habilidades. Se puede observar que las habilidades más difíciles de encontrar no necesariamente son las más importantes para las empresas, como lo es el caso de *Pensamiento Holístico* donde es considerada más difícil de encontrar que importante para las empresas.

En respuesta a lo anterior, en los últimos años se han definido nuevos perfiles profesionales, que recopilan conjuntos de habilidades que se requieren para la implementación de los proyectos que surgen en las áreas de BD&DA. Lograr una formación en estos temas, constituye un reto, y para enfrentarlo, los países generan nuevos centros de formación y las universidades incluyen en sus

Cuadro 2: Habilidades transversales

Habilidades transversales	Mas importantes para las empresas	Las más difíciles de encontrar
Liderazgo/Proponer liderazgo	45.2 %	43.8 %
Pensamiento del Sistema / Pensamiento Holístico	27 %	33.5 %
Análisis de negocios	39 %	31 %
Innovación	28 %	27.4 %
Solución de problemas	25.7 %	21.7 %
Comunicación escrita	17.2 %	20.5 %
Toma de decisiones	19.7 %	20.5 %
Comunicación oral	19 %	19.4 %
Colaboración con otros	27.6 %	17.5 %

currículos cursos pertinentes que intentan suplir la alta demanda que se presenta en la actualidad [12]. Sin embargo, las exigencias son cada vez mayores y aún no se comprende bien el rol que debe jugar este profesional, lo que conlleva a la dificultad para encontrarlos, contratarlos y para formarlos.

Esto motiva el trabajar en la definición unificada de un perfil, con el fin de entender las capacidades que incluye y su interacción con otros perfiles profesionales, lo cual es el principal objetivo del presente documento.

2. Problemática

Actualmente se buscan perfiles profesionales para el desarrollo de proyectos de *Big Data (BD)* y *Data Analytics (DA)*, pero surge como problemática el hecho de que las empresas no pueden definir claramente el perfil que requieren debido a la novedad de estas temáticas, no existe un lenguaje común, y cada vez se dificulta más conocer un perfil acorde a las necesidades que se buscan para el desarrollo de estos proyectos.

Para una mejor comprensión de la problemática, se presenta a continuación las características que presentan los proyectos de *Big Data (BD)* y *Data Analytics (DA)*, los diferentes perfiles que han sido propuestos hasta el momento y se concluye con la necesidad de un nuevo perfil.

2.1. Características de los proyectos de Big data analytics y data Analytics

El término *Big data* se refiere a gigantescas cantidades de información que deben ser capturadas y analizadas, con el fin de que las empresas aprovechen la información que poseen de su negocio (información de clientes, histórico de ventas etc.), en su toma de decisiones estratégicas. Sin embargo, muchas veces las empresas no poseen entre su personal los profesionales que desarrollen y realicen la analítica requerida sobre este tipo de datos [6].

Hoy en día el *Big data* está cambiando la manera como trabajan las empresas, pues se genera mucha información y oportunidades para aprovecharla, en la búsqueda de ventajas competitivas.

Sin embargo, todavía no son muchas las empresas que, aunque sienten la necesidad, se deciden a enfrentar estos retos y aplicar esta nueva disciplina, lo que genera la oportunidad de formar profesionales en esta temática.

Los datos que poseen las empresas son complejos de manejar y además se generan rápidamente lo que complica el análisis de la información. De allí surge la necesidad de implementar proyectos en BD y DA con el fin de apoyarse en los datos para la toma de decisiones (acertadas). Por esta razón, los proyectos de BD y DA son estratégicos en las empresas, involucran grupos de profesionales y requieren de una metodología que les permita trazar un plan apropiado de trabajo.

El uso de tecnologías (TI) apropiadas es fundamental para el desarrollo de estos proyectos que están orientados a extraer información valiosa de los datos, y así encontrar oportunidades para el negocio partiendo del análisis de la información. Debe hacer posible identificar ideas de negocio en los datos y fijar metas claras para el proyecto.

Actualmente grandes empresas como: SAS, Ibm y Oracle ofrecen como servicios soluciones en *Big data* y *data analytics* para beneficiar a las empresas con el manejo adecuado de sus datos.

2.2. Perfiles existentes

Teniendo en cuenta la problemática anterior, se han creado en los últimos años perfiles para suplir esta necesidad. Estos cumplen con habilidades y competencias (cuadro 1) para suplir cargos que demandan profesionales con las siguientes características: Manejo de herramientas de Big Data para procesar información, manejo de lenguajes de programación, dominio de bases de datos relacionales y no relacionales, uso de herramientas de visualización, habilidades en analítica, uso de herramientas de *Machine Learning* [13].

A continuación se mencionan cargos representativos, que existen en la actualidad.

Chief Data Officer (CDO): Es un miembro de la dirección ejecutiva que se encarga de liderar la gestión de datos y analítica asociada con el negocio, por lo tanto es el responsable de los diferentes equipos que están especializados en datos en la empresa [9, 7].

Data Scientist (científico de los datos): Interpreta grandes volúmenes de datos para transformarlos en información útil para la empresa. Se caracteriza por tener habilidades en matemáticas, estadística, programación ser creativo y tener habilidades comunicativas que le permitan exponer los resultados de su trabajo para ayudar a la empresa en su toma de decisiones. (Ingeniero del futuro) [1, 9, 14, 5].

Citizen Data Scientist: En palabras de Gartner "El *citizen Data Scientist* es una persona que crea o genera modelos que aprovechan el análisis predictivo o prescriptivo, pero cuya principal función de trabajo se encuentra fuera del ámbito de la estadística y análisis" [8].

Data Engineer: Es el responsable de entregar al científico de datos los datos que requiere. Tiene gran conocimiento en bases de datos, arquitecturas de cluster, lenguajes de programación y sistemas de procesamiento de datos [1, 9].

Data Steward (administrador de datos): Es el responsable de mantener la calidad, disponibilidad y seguridad de los datos. Posee conocimientos en los procesos del negocio e identifica cómo son usados dentro de la empresa [1, 9].

Business Data Analyst (analista de datos): Participa en el análisis de los datos con el fin de recolectar las necesidades del cliente para sustentarlas de manera clara al científico de datos

[1, 9].

Data Artist:Es un experto en *Business Analytics* y tiene delegada la tarea de mostrar de manera sencilla (gráficos, infografías y herramientas visuales), los resultados del análisis de los datos para comprender grandes volúmenes de información [1, 9].

Estadístico:Sus funciones se centran en obtener, analizar e interpretar datos cualitativos y cuantitativos usando los métodos estadísticos existentes (e.g., medias, histogramas, desviaciones, regresiones, test-X, test-Z) [15].

Administrador de bases de datos:Tiene conocimientos fuertes en el manejo de bases de datos, típicamente relacionales [15].

Los perfiles vistos anteriormente poseen diversos significados y no se ha determinado un estándar para ellos [15, 9, 5, 14]. Muchos de ellos se sobrelapan y no existe una definición común lo que impide una claridad en los nuevos cargos que se requieren con tales conocimientos.

2.3. Necesidad de un nuevo perfil

Con el fin de obtener un perfil completo para desarrollar proyectos en *BD* y *DA*, se crea la necesidad de definir un nuevo perfil que incluya las habilidades necesarias para realizar este tipo de proyectos.

En la ejecución de proyectos de *BD* y *DA*, se debe contar con el apoyo de un **equipo de trabajo** capaz de comprender los diferentes ejes de conocimiento de un proyecto. Es decir, no basta con el conocimiento del *Data Scientist* o del Ingeniero de Datos por separado, siendo roles diferentes dentro de los perfiles orientados a datos, sino que se necesita la integración de conocimientos en un equipo armónico, en el que se dispongan de habilidades múltiples para el desarrollo satisfactorio de un proyecto.

3. Antecedentes

El problema de definir perfiles para proyectos de *BD* y *DA* se ha abordado en la literatura desde diferentes perspectivas[11, 3]. Este problema puede ser analizado usando descripciones de perfiles, metodologías y aproximaciones pedagógicas que aporten a una conceptualización del mismo. Estos enfoques son descritos en esta sección y son utilizados para identificar los fundamentos que permiten definir un perfil para desarrollar este estilo de proyectos.

3.1. Otras propuestas de perfiles

El modelo que se presenta en la figura 1 muestra la propuesta de Gartner para conformar un Centro de competencia en Inteligencia de Negocios (BICC), el cual parte de la idea de especializar las habilidades de acuerdo a áreas de interés (e.g., IT en la figura), resaltando las habilidades en temas de datos, gobierno y comunicación. Por otro lado resalta una serie de conceptos a considerar, tales como la integración y manejo de la información, herramientas y aplicaciones, organización y procesos, y las necesidades del negocio a ser abordadas. Lo anterior, en lo que refiere a las cualidades analíticas, en negocio y TI que se requieren para el desarrollo de un proyecto de analítica.

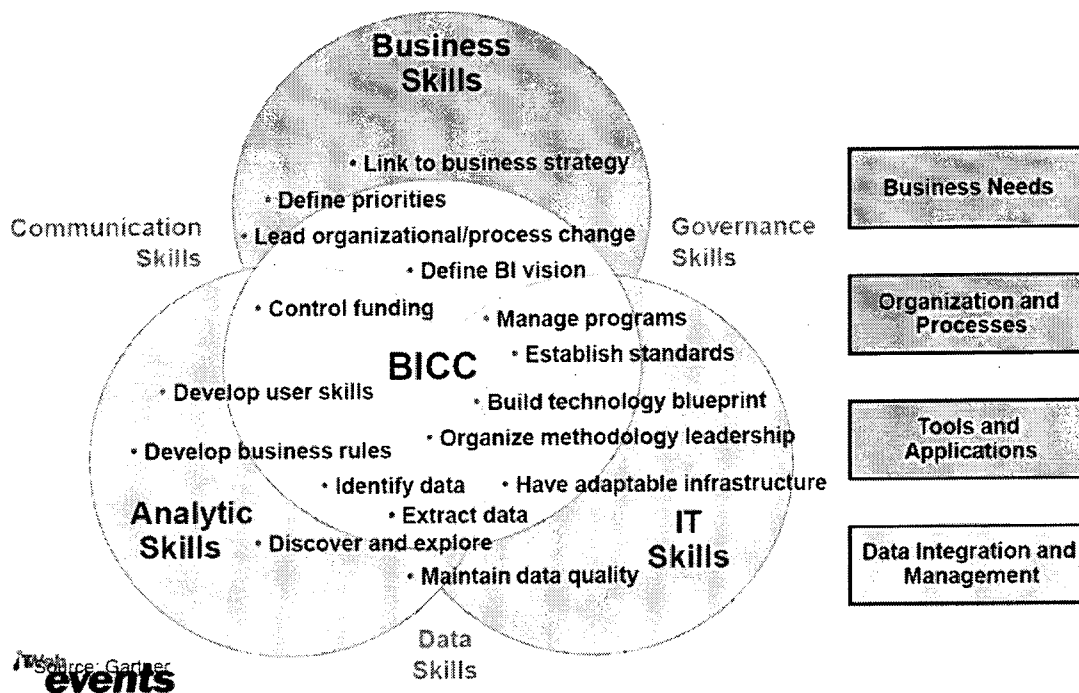


Figura 1: Propuesta de Gartner para adaptar las Competencias BICC

3.2. Metodología de ASUM de IBM

La metodología ASUM-DM de IBM, la cual se encuentra establecida con base en la metodología CRISP DM, aun si tiene algunas similitudes con metodologías reconocidas para realizar proyectos de minería de datos, enfatiza varias de las nuevas prácticas en la ciencia de datos como el uso de volúmenes de datos muy grandes, la incorporación de análisis de texto en el modelado predictivo y la automatización de algunos procesos.

La figura 2 muestra las fases de la metodología ASUM-DM.

Esta metodología se concentra en 5 grupos de fases globales: (1)Analizar, (2)Diseñar, (3)Configurar y Construir, (4)Desplegar, (5)Operar y Optimizar.

Dentro de cada uno de los grupos, se despliegan las 10 fases que componen la metodología ASUM-DM. A continuación se habla brevemente sobre cada elemento, para un mejor entendimiento, tomando como punto de partida la figura 2, en la cual se evidencia el agrupamiento de cada uno de los elementos dentro de las fases previamente mencionadas:

- **Comprensión del Negocio:** Esta primera etapa establece las bases para abordar de forma exitosa el proyecto ya que se orienta al entendimiento de un negocio, con el fin de definir proyectos y soluciones alienados con las estrategias de las organizaciones.
- **Enfoque Analítico:** Implica traducir el problema de negocio a un problema técnico. Una vez que el problema de negocio de la empresa ha sido claramente establecido, el científico de

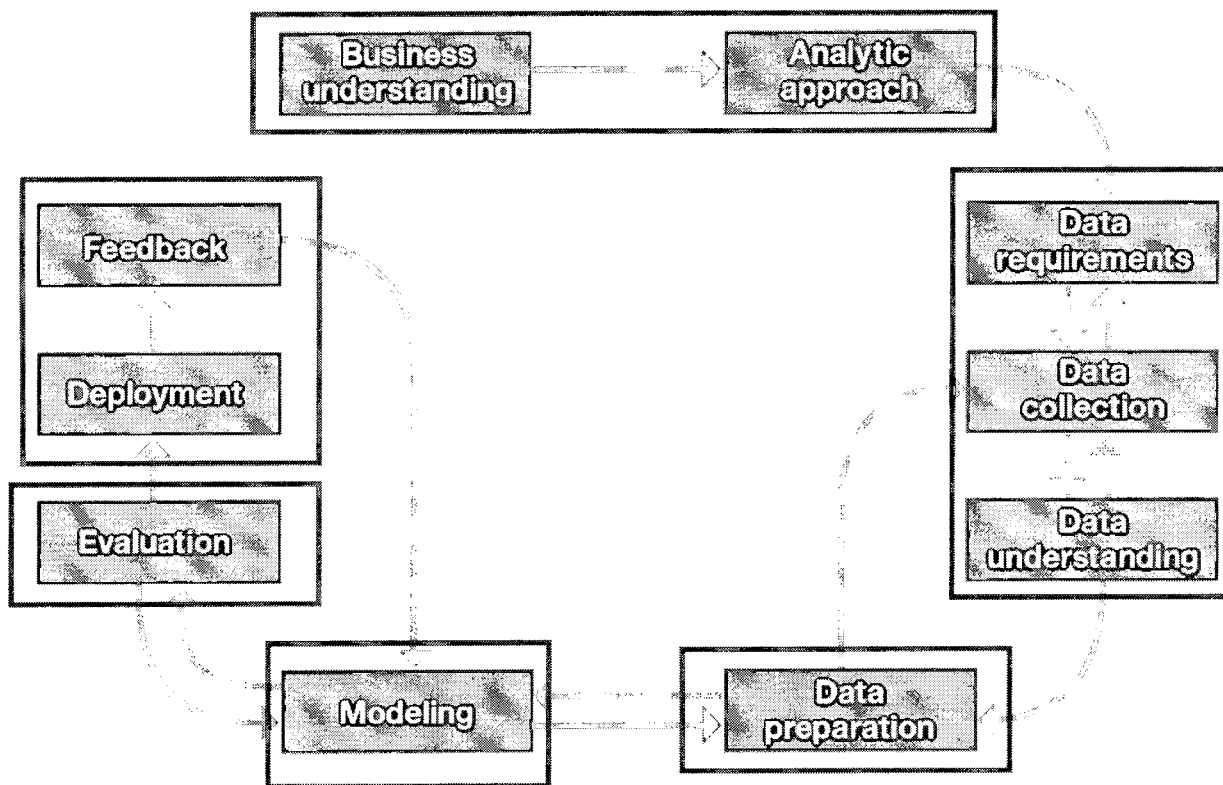


Figura 2: Metodología ASUM-DM de IBM. Recuperada de <http://www.ibmbigdatahub.com/blog/why-we-need-methodology-data-science>

datos puede definir el enfoque analítico para resolver el problema. Esta etapa implica expresar el problema en el contexto de las técnicas estadísticas y de aprendizaje automático, para que la organización pueda identificar y seleccionar las más adecuadas para el proyecto.

- **Requisitos de los datos:** La elección del enfoque analítico determina los requisitos de los datos, ya que los métodos analíticos que se utilizan requieren un contenido de datos, formatos y representaciones particulares, guiados por el conocimiento del dominio.
- **Recolección de los datos:** El científico de datos identifica y reúne los recursos de datos necesarios, relevantes para el dominio del problema. Al encontrar brechas en la recopilación de datos, el científico de datos podría necesitar revisar los requisitos de datos y recopilar más información.
- **Entendimiento de los datos:** Técnicas de visualización o estadística pueden ayudar a un científico de datos a comprender el contenido de los datos, evaluar su calidad y tener hallazgos iniciales de interés para el proyecto.
- **Preparación de los datos:** La etapa de preparación de datos comprende aquellas actividades para construir el conjunto de datos que se utilizará en la etapa de modelado. Estos incluyen la limpieza de datos y otras técnicas de análisis de datos para satisfacer la necesi-

dad de tener un conjunto robusto para la construcción de modelos apropiados para abordar el problema.

- **Modelamiento:** A partir de la primera versión del conjunto de datos preparado, la etapa de modelado se centra en el desarrollo de modelos predictivos de acuerdo con el enfoque analítico previamente definido. Con modelos predictivos, los científicos utilizan un conjunto de formación (datos históricos en los que se conoce el resultado del interés) para construir el modelo. Esta etapa suele ser iterativa, lo que lleva a refinamientos en la preparación de los datos y la especificación del modelo.
- **Evaluación:** El científico de datos evalúa el modelo para entender su calidad y asegurarse de que aborda adecuada y completamente el problema del negocio. La evaluación del modelo implica el cálculo de diversas medidas de diagnóstico.
- **Despliegue:** Después de que se ha desarrollado un modelo con resultados satisfactorios en su evaluación, se despliega en el entorno de producción o en un entorno de prueba comparable.
- **Retroalimentación:** Al recolectar los resultados del modelo implementado, la organización obtiene retroalimentación sobre el rendimiento del modelo y observa cómo afecta su entorno de despliegue.

Una vez descritos los niveles que componen la metodología, cabe aclarar que esta es una metodología planteada para el desarrollo del proyecto en un **equipo de trabajo**. Esto es un elemento fundamental para el cumplimiento de los objetivos planteados por el mismo. Por esta razón, implica un conocimiento base de cada uno de los integrantes del grupo en las diferentes áreas del conocimiento de un proyecto para tener un lenguaje común.

3.3. Taxonomía de Bloom

En el contexto de la definición de un perfil, es necesario definir un esquema de valoración de conocimiento que se adquiere o requiere en procesos de formación y transferencia. Para esto, se propone utilizar la taxonomía de Bloom, definida en el contexto de aprendizaje de las personas. Esta taxonomía es jerárquica y se evidencian diferentes niveles de conocimiento [2].

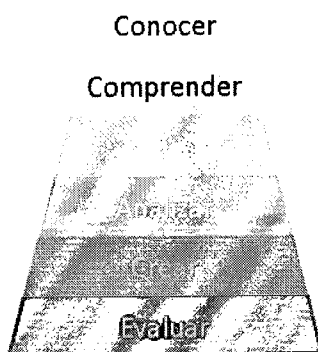


Figura 3: Pirámide de conocimiento. Taxonomía de Bloom

Como se puede observar en la figura 3, dicha pirámide consta de los siguientes niveles, los cuales brevemente se explican según la teoría y la Real Academia Española (RAE):

- **Conocer:** Se refiere a la capacidad de recordar hechos específicos y universales, métodos y procesos, esquemas, estructuras o marcos de referencia.
Según la RAE, la definición del verbo *Conocer* corresponde a averiguar por el ejercicio de las facultades intelectuales la naturaleza, cualidades y relaciones de las cosas.
- **Comprender:** Es la capacidad de comprender qué se le está comunicando. Es decir, se trata del entendimiento demostrativo de hechos e ideas por medio de la organización, la comparación, la interpretación, las descripciones, entre otros.
Según la RAE, *Comprender* es entender, alcanzar o penetrar algo.
- **Aplicar:** Se refiere al uso del conocimiento nuevo. Es decir que este se basa en la resolución de problemas en nuevas situaciones aplicando el conocimiento adquirido, hechos, técnicas y reglas en un modo diferente.
Según la RAE, se da la definición del verbo *Aplicar*: Emplear, administrar o poner en práctica un conocimiento, medida o principio, a fin de obtener un determinado efecto o rendimiento en alguien o algo.
- **Analizar:** Consiste en descomponer un problema, planear, proponer nuevas maneras de ejecutar y/o hacer, aplicando sus conocimientos. Examina y divide la información en partes identificando motivos o causas. Hace inferencias y encuentra evidencia para apoyar generalizaciones. Según la RAE, se da la definición del verbo *Analizar*: Someter algo a un análisis.
- **Crear:** Es el proceso de trabajar con partes, elementos y organizarlos u ordenarlos para formar un todo claro. Compila la información de manera diferente combinando elementos en un nuevo patrón o proponiendo soluciones alternativas.
Según la RAE, el verbo *Crear* corresponde a establecer, fundar, introducir por vez primera algo; hacerlo nacer o darle vida, en sentido figurado.
- **Evaluar:** Se refiere a la capacidad para evaluar, a través de los procesos de análisis y síntesis. Presenta y defiende opiniones haciendo juicios sobre la información, la validez de las ideas o la calidad del trabajo sobre la base de un conjunto de criterios.
Según la RAE, se da la definición del verbo *Evaluar*: Señalar el valor de algo. Estimar, apreciar, calcular el valor de algo.

Por lo pronto, conociendo las definiciones impartidas tanto por la teoría de los niveles de conocimiento de la taxonomía de Bloom como por la RAE, es importante recalcar cómo se relacionan los niveles en la jerarquía para la definición de los alcances de un equipo en el desarrollo de un proyecto de BD y DA.

3.4. Conclusión

De acuerdo a lo presentado en las secciones anteriores, se resaltan los siguientes puntos:

- Es importante que en un proyecto en BD y DA se maneje un esquema de trabajo en equipo. La interacción entre los conocimientos y experticias de cada uno de los miembros del grupo, al igual que el seguimiento de una metodología, son factores claves para el éxito de estos proyectos.
- El trabajo en equipo dado que cada miembro del grupo, dispone de habilidades y tiene conocimientos en distintas áreas, requiere de un vocabulario, conocimientos y habilidades comunes, para un trabajo armónico, eficiente y eficaz dentro de los proyectos.
- Todos y cada uno de los integrantes del equipo poseen un conjunto de habilidades transversales que les permiten trabajar de manera conjunta, coherente y completa, pues entre todos deben llevar a buen término el proyecto.
- Una primera aproximación a la definición de perfil, para un integrante de un equipo, es que debe tener las capacidades y conocimientos necesarios y suficientes para participar de manera efectiva en un proyecto de BD y DA. Implica experticia en su área de conocimiento para tomar las decisiones y acciones adecuadas para el buen desarrollo del proyecto. Adicionalmente, debe conocer lo suficiente de los otros dominios para entender las decisiones que allá se están tomando y las implicaciones que ellas tienen en su propio dominio.
- Las aproximaciones metodológicas están enfocadas en su mayoría a las tareas dentro del proyecto, pero no a su administración. Sin embargo, es fundamental la inclusión de esta perspectiva de "Administración del Proyecto" en pro de llevar a cabo una gestión organizada, planificada y coherente con el proyecto, de acuerdo a la disponibilidad de recursos, tiempo, entre otros.

Estos puntos son incorporados en la propuesta presentada por el grupo CAOBA-Uniandes en las siguientes secciones.

1. Marco conceptual para el perfil Big data y Data Analytics

En la presente sección se describe una propuesta de marco conceptual que el grupo CAOBA-Uniandes tiene para definir un perfil en *Big Data and Data analytics* (BD y DA), el cual quiere, a su vez, servir de marco de referencia con respecto a las habilidades y conocimientos necesarios para el desarrollo de proyectos en el tema, para la definición de perfiles profesionales, de cursos de formación, o de requerimientos de una empresa proveedora de soluciones.

Esta propuesta tiene tres partes importantes. La primera de ellas corresponde a las bases del marco conceptual para definir perfiles. La segunda es una formulación matemática, que permite expresar el marco de referencia en términos de los conocimientos y habilidades necesarios para desarrollar un proyecto de BD y DA y la tercera, corresponde a la definición de perfiles específicos para la formación de profesionales en los temas de *Big Data* y *Data analytics*.

La propuesta se creó basada en la idea de definir cualidades del equipo de trabajo en términos de diferentes dominios (e.g., negocio, tecnologías de información), propuestos en el BICC de Gartner. En particular, los dominios propuestos involucrados en este estilo de proyecto de *Big Data* y *Data analytics* son: negocio, tecnologías de información (TI), analítica, administración de proyectos y habilidades transversales. Estos dominios permiten expresar un perfil completo y se describen a continuación.

Negocio: Definición y comprensión de la temática específica del proyecto, que está definida por las necesidades y reglas del negocio.

TI: Administración de la tecnología de TI, despliegue de la solución y manejo del ciclo de vida de los datos, al igual que temas transversales de los mismos.

Analítica: Selección y análisis de los datos de forma apropiada, al igual que selección y construcción de los modelos adecuados para la solución.

Administración de proyectos: Gestión del proyecto, al igual que la gestión del equipo de trabajo colaborativo.

Habilidades transversales: Conjunto de habilidades que apoyan el desarrollo efectivo de un proyecto de BD&DA.

4.1. Fundamentos del marco conceptual

Los tres elementos fundamentales para el desarrollo efectivo de proyectos en *Big Data* y *Data analytics* son: el aspecto metodológico (ver sección 3.2), los conocimientos y habilidades en los diferentes dominios, a nivel del equipo y de cada integrante del mismo, implícitamente abordado en la metodología ASUM-DM y la existencia de un lenguaje común para facilitar la comunicación al interior del grupo, al igual que la comprensión y participación en las decisiones relacionadas con el proyecto. Algunos de estos elementos han sido abordados por propuestas como la metodología ASUM-DM, las características que deben tener los centros de competencia de inteligencia de negocios y la taxonomía de Bloom. Aunque estos elementos permiten definir el trabajo a realizar en proyectos de BD y DA, dado que no fueron propuestos para la definición de perfiles, carecen de elementos importantes, tales como la administración de proyectos, la forma de explicitar el trabajo en equipo y de medir el conocimiento y habilidades que debe tener el equipo, y cada uno de sus integrantes.

Por lo anterior, se realiza una propuesta que integra los aspectos faltantes mencionados previamente y que se describen a continuación.

1. Aspecto metodológico: Basado en la metodología de ASUM-DM, como se muestra en la figura 4, se hace énfasis en la administración de proyectos y en la etapa de modelamiento. A nivel de modelamiento, se extiende el alcance de proyectos de analítica predictiva, y se incluye la analítica descriptiva. En particular, se abarcan soluciones de visualización como lo son los tableros de control (*Dashboards*), la visualización de cuadros de mando y en general, resultados de modelos descriptivos.
2. Valoración de conocimiento y habilidades: Para definir el conocimiento y las habilidades que posee cada integrante del grupo se usa la taxonomía de Bloom. Esta taxonomía permite definir qué nivel de conocimiento/habilidad tiene cada integrante del grupo, en cada dominio para participar activamente en el desarrollo del proyecto, como se muestra en la figura 5. De igual manera, al explicitar el nivel de conocimientos y habilidades de cada integrante, se tendrá también la valoración del equipo de trabajo.

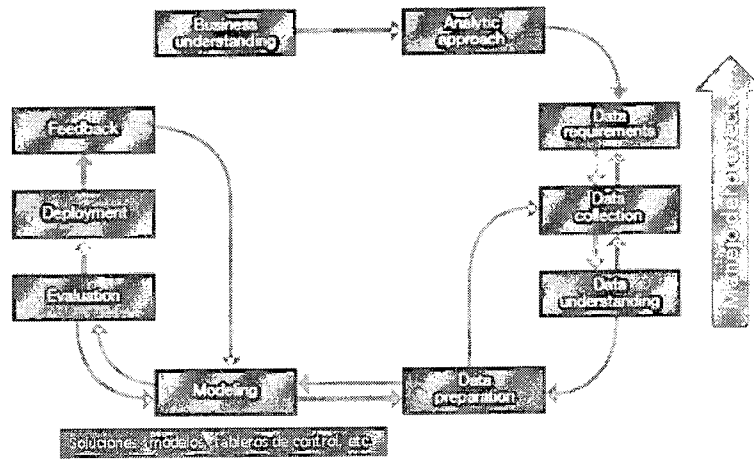


Figura 4: Metodología ASUM-DM adoptada por CAOBA

3. Lenguaje común: Es necesario definir un núcleo para que el equipo trabaje de manera efectiva, con un lenguaje común y estándar que facilite el entendimiento del proyecto y las decisiones asociadas al mismo, en todos los dominios: TI, Negocio, Analítica y Administración de proyectos. De igual manera, el núcleo debe incluir un conjunto de habilidades transversales, para garantizar la participación activa de todos los integrantes del equipo.

Por otro lado, vale la pena aclarar que en términos de habilidades transversales el nivel de Bloom requerido es de usar y analizar, ya que en esos temas no se está pretendiendo avanzar la frontera del conocimiento.

Desde el punto de vista de los integrantes, no se espera que todos tengan toda la experticia planteada, sino que cada uno aporte la suya al equipo completo. Sin embargo, sí se requiere de habilidades y conocimientos compartidos, comunes por todos los miembros para llevar a buen término el proyecto.

4.2. Formulación matemática del perfil

Con el fin de facilitar el entendimiento del perfil que poseen las personas que hacen parte de un equipo de trabajo en BD y DA, se propone una representación de acuerdo a lo ilustrado en la ecuación 1.

$$Perfil_{BD\&A} = P_0 + \sum_{i=1}^4 C_i * P_i, \text{ donde } C_i \in (0, 1)^{\exists i} | C_i \neq 0 \quad (1)$$

La ecuación del perfil BD&A se define en términos de un perfil núcleo (**P0**), más la sumatoria de uno ó más de los perfiles especializados (**Negocio -P1 a Administración -P4**). Para comprender mejor la ecuación es necesario identificar el significado de cada perfil de la siguiente manera.

Perfil núcleo (P0) considerado el núcleo-visión global, facilita la comunicación entre los miembros del grupo, permite manejar un lenguaje común y tener una comprensión de conceptos en los

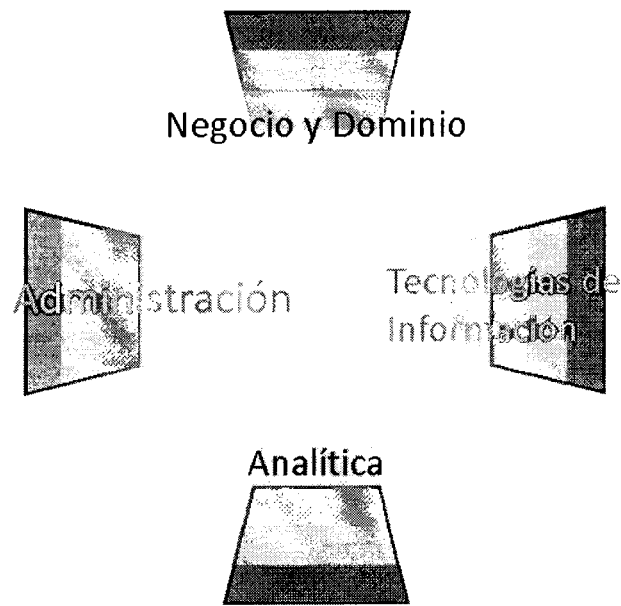


Figura 5: Metodología ASUM-DM adpotada por CAOBA

diferentes dominios, para participar en las decisiones del proyecto. Por otra parte, los perfiles especializados, **Perfil de negocio**, **Perfil en TI**, **Perfil en analítica** y **Perfil en administración de proyectos** se detallan en las siguientes secciones.

A manera de ejemplo, la ecuación 1 es utilizada para representar dos perfiles de BD y A con fortalezas en algunas áreas.

Perfil profesional 1: Un profesional con estudios de pregrado de matemática pura, que quiere especializarse en el dominio de analítica. Este perfil puede ser expresando en el formalismo como se muestra en la ecuación No: 2.

$$\text{Perfil } BD\&A = P_0 + P_3 \tag{2}$$

Perfil profesional 2: Un ingeniero de sistemas con maestría en analítica, con fuertes conocimientos en analítica y tecnologías de la información. Este perfil se representa en el formalismo como se muestra en la ecuación No: 3.

$$\text{Perfil } BD\&A = P_0 + P_2 + P_3 \tag{3}$$

De acuerdo a los ejemplos se puede hacer énfasis en que el formalismo utilizado para representar un perfil BD y DA permite describir proyectos o profesionales en BD y DA con diferentes énfasis.

4.3. Perfil Núclco - P0

Este perfil, ilustrado en la figura 6, representa la participación de manera efectiva en el proyecto, ya que conoce y comprende las decisiones de los cuatro dominios (**P1 a P4**) y en esos dominios, tiene un conjunto fundamental de conocimientos y habilidades.

El tener conocimientos y habilidades en los 2 primeros niveles de Bloom permite al profesional con el perfil núcleo estar en la capacidad de conocer la idea del negocio, las posibles herramientas en TI a usar, los modelos matemáticos existentes y los procedimientos para la administración de proyectos que junto con habilidades (transversales) como trabajo en equipo, comunicación, responsabilidad, liderazgo, disciplina, y pasión entre otras le permiten comprender las decisiones y las consecuencias de las decisiones que se toman durante el proyecto (desde la concepción hasta el cierre).

A continuación se mencionan en detalle los conocimientos que poseen en cada uno de los dominios.

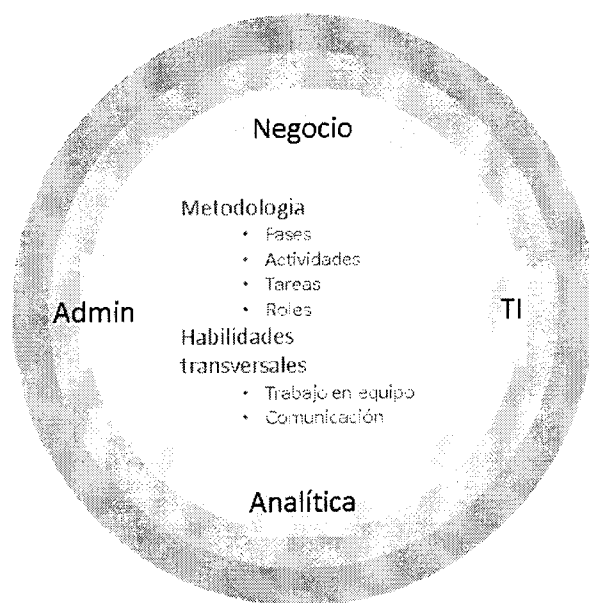


Figura 6: Representación del perfil P0

4.3.1. El núcleo del negocio

Conoce y comprende acerca del negocio, su participación en el proyecto se caracteriza por conocer acerca de la misión y visión de la empresa con el fin de alinear el proyecto con los objetivos de la organización.

4.3.2. El núcleo en TI

Identifica las tecnologías de información existentes que se ajustan a los requerimientos solicitados por parte del cliente y es capaz de identificar sus ventajas, desventajas e implicaciones al usarlas adecuadamente. Adicionalmente, conoce y comprende los procesos asociados a la calidad de datos y a las implicaciones legales y éticas intrínsecas al uso de datos en los proyectos de tecnologías de información.

4.3.3. El núcleo en analítica

Conoce y comprende las aproximaciones analíticas que pueden ser implementadas en este estilo de proyectos teniendo como guía el objetivo para el cual han sido diseñadas, y comprendiendo las ventajas y desventajas que implica su uso en estos proyectos.

4.3.4. El núcleo en administración de proyectos

Conoce y comprende los conceptos de la gestión de proyectos que es parte fundamental para lograr los objetivos que se han planteado para un proyecto, conoce el presupuesto, los recursos y el cronograma con el cual está dirigido el proyecto, esto le permite participar de manera efectiva.

4.3.5. El núcleo transversal

Contiene el conjunto de habilidades blandas requeridas para el desarrollo efectivo de este estilo de proyectos, tales como creatividad, curiosidad, pasión, disciplina, honestidad, liderazgo, actualización permanente, disciplina de documentación, comunicación efectiva en español y muy deseable en inglés, poder de abstracción, negociación, toma de decisiones, definir criterios, evaluar, comparar, escoger, trabajo en equipo, tenacidad y tolerancia a la frustración.

4.4. Perfil de Negocio - P1

Este perfil se caracteriza porque el profesional es capaz de actuar y tomar decisiones sobre el proyecto con base en el conocimiento del negocio y de la organización.

Las habilidades que lo caracterizan corresponden en conocer y comprender la idea de negocio, lo que le permite realizar acciones y tomar decisiones sobre el proyecto de forma adecuada manteniendo los intereses que busca la empresa en la ejecución del proyecto.

Por otro lado, en términos de conocimientos domina temáticas como las que se presentan a continuación:

- Áreas: contexto del negocio, aspectos legales, misión y visión de la empresa, planeación (corto y largo plazo) e identifica las amenazas y oportunidades que se puedan presentar en la ejecución del proyecto debido a su experiencia del negocio.
- Las metodologías y tecnologías: DOFA, VRIN, DELTA, BMM y arquitecturas empresariales.

Sus conocimientos le permiten hacer análisis de los costos, beneficios y herramientas que ayuden en el desarrollo del proyecto.

En la figura 7 se puede ver el diagrama del conocimiento que posee un perfil P1, donde el nivel de conocimiento en el dominio del negocio es completo y en las demás áreas su nivel de conocimiento es básico según se muestra en detalle en la definición del perfil núcleo, explícitamente al inicio de la presente sección (sección 4.3).

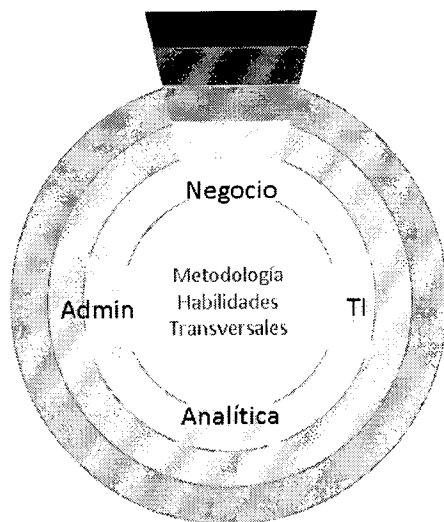


Figura 7: Representación del perfil P1: Dominio del Negocio

4.5. Perfil en TI - P2

El profesional en este perfil, realiza acciones y toma decisiones sobre el proyecto con base en el conocimiento de las tecnologías de información.

Este perfil es experto en tecnologías de información (TI) que apoyen proyectos de *Big Data* y *Data analytics*, como se puede observar en la figura (8). Adicionalmente es responsable de la administración de la infraestructura, en cuanto a hardware y software, y el despliegue de la solución. Cuenta con habilidades en manejo de datos, en términos de su ciclo de vida y calidad y tiene liderazgo en la escogencia de lenguajes, herramientas y metodologías en temas relacionados con TI.

En cuanto a las demás áreas su nivel de conocimiento es básico según la definición del perfil núcleo descrito en la sección 4.3.

En términos de los conocimientos domina temas tales como:

- Áreas: sistema operacional, computación en la nube, computación de alto rendimiento, desarrollo de aplicaciones, lenguajes de programación, análisis sintáctico, bases de datos SQL, bases de datos NO SQL, consolidación de fuentes, arquitecturas de información-analítica.
- Metodologías y tecnologías: *Map/reduce*, *Text Mining*, *Data Mining*, *Web Mining*, *Social Mining*, repositorios y versionamiento, ETL, Sistemas de recomendación.
- Herramientas: Python, java, unix, spark, posgress, hive, mongo, hadoop y github.

4.6. Perfil en Analítica - P3

En este perfil el profesional realiza acciones y toma decisiones sobre el proyecto con base en el conocimiento del área analítica.

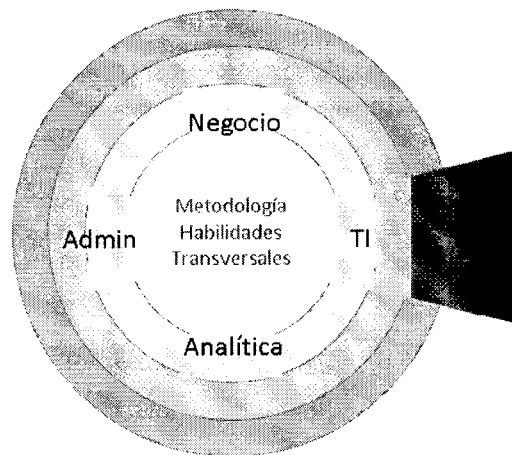


Figura 8: Representación del perfil P2: Tecnologías de información

En la figura 9 se pueden ver los niveles de conocimiento y habilidades que posee donde su fortaleza está en el dominio de analítica y en los demás dominios su nivel es un perfil núcleo (P0).

Este perfil tiene habilidades de manejo de los datos fuentes, en términos de descubrimiento, exploración, selección, preparación y visualización, en la selección de modelos analíticos, la comunicación de los resultados de la solución construida y la selección de herramientas y metodologías apropiadas para la temática analítica.

En términos de los conocimientos domina temas tales como:

- Áreas: Álgebra Lineal, probabilidad, estadística, reducción de dimensión, optimización, simulación, metaheurísticas.
- Metodologías y tecnologías: Modelos de Regresión, modelos de clasificación, *Machine learning*, métodos gráficos, teoría de grafos y cadenas de Markov. Las herramientas que domina son: R, matLab, SAS y stata.

4.7. Perfil en Administración de proyectos - P4

En este perfil el profesional realiza acciones y toma decisiones sobre el proyecto con base en el conocimiento del área de administración de proyectos. Como se puede observar en la figura 10, su fortaleza está en el dominio de Administración de proyectos, en el cual se ve definido el perfil.

Este perfil tiene mayores habilidades en el conocimiento de gestión de proyectos en cuanto a la planeación, tiempos, manejo de recursos y estructuración de las acciones a realizar en pro de cumplir los acuerdos realizados con los distintos actores involucrados en el proyecto. Específicamente, conoce y comprende de forma detallada cómo gestionar un proyecto, los recursos y el presupuesto disponible, y además es capaz de organizar con el uso de herramientas y conocimiento en metodologías la planeación, ejecución, monitoreo y ajuste de tareas antes, durante y después del desarrollo del proyecto. Incentiva y monitorea adicionalmente el trabajo en equipo.

Por otro lado, en términos de los conocimientos dominan temas tales como:

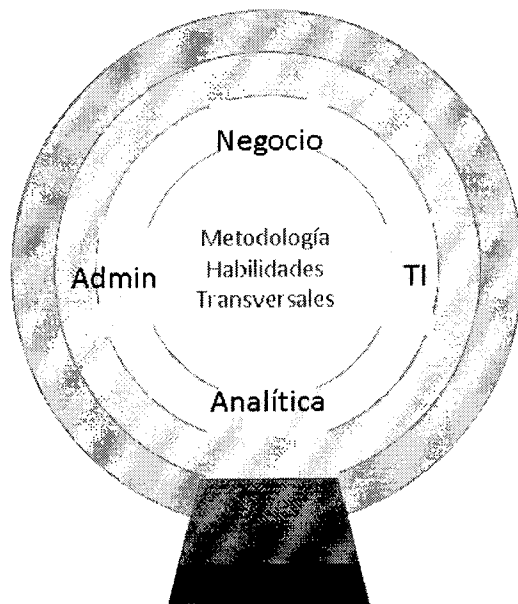


Figura 9: Representación del perfil P3: Analítica

- Áreas: Definir presupuestos, identificación de fuentes de datos pertinentes, mercadeo, perfilamiento de necesidades, clientes, usuarios y procesos a partir de las decisiones basadas en información, definición y aseguramiento de procesos de calidad de información, negociación y administración del talento humano.
- Metodologías y tecnologías: PMBOK, PMP y DAD.
- Herramientas: MS-Project y T-Metrics.

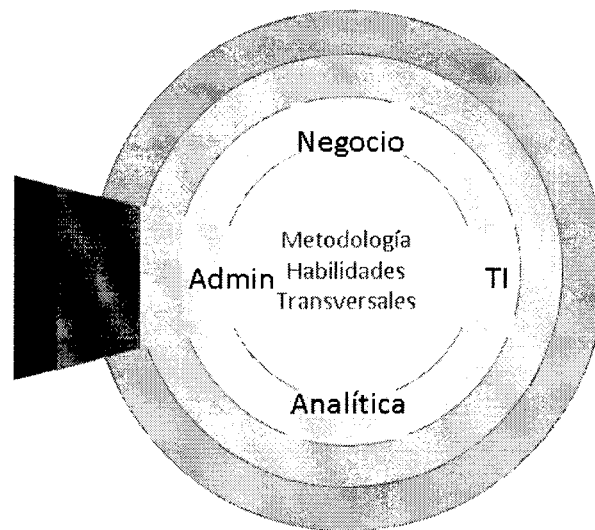


Figura 10: Representación del perfil P4: Administración de Proyectos

5. Validación de la propuesta de los perfiles CAOBA

La metodología ASUM-DM utilizada para realizar proyectos de *Big Data* y *Data Analytics* contiene un conjunto de fases y actividades, descritas en la sección 3.2, que apoyan el desarrollo completo de este tipo de proyectos. Es por esto, que con el objetivo de validar la completitud de la propuesta del perfil CAOBA se utiliza como base tal metodología.

El proceso seguido para la validación consistió en:

- Tomar cada actividad definida en ASUM-DM y:
 - Identificar cuál o cuáles de los perfiles propuestos participan en dicha actividad
 - Determinar cual de los perfiles identificados en el paso anterior, debe ser el responsable de dicha actividad, dejando a los otros perfiles como colaboradores
- Comprobar que todas las actividades ASUM-DM tienen un perfil responsable
- Comprobar que cada perfil es responsable de al menos una de las actividades de ASUM-DM

El resultado mostró que, efectivamente, todas las actividades tienen un responsable y que todo perfil tiene explícitas sus responsabilidades, de donde se puede concluir que los perfiles CAOBA propuestos son válidos para el desarrollo de proyectos de *Big Data* y *Data Analytics*.

Un extracto del resultado de este ejercicio se muestra en la figura 11. La figura muestra para cada actividad los perfiles participantes con una 'x' y el perfil responsable de la actividad resaltado en amarillo.

	Habilidad	Administración	Negocio	TI	Analítica
Prepare for Implementation		x	x	x	x
Conduct Readiness Assessment			x	x	
Conduct Project Kick-off		x			x
Understand Business	Determine Business Needs	x	x	x	x
	Link to business strategy	x	x		x
	Understand data flow			x	
	Assess Situation		x	x	x
	Define priorities	x	x	x	x
	Define Technical Goals (DM, Dboard, ...)		x	x	x
	Define project vision	x	x	x	x
	Create Project Plan	x	x	x	x
	Create Business Understanding Report	x	x		
	Develop business rules	x	x		

Figura 11: Extracto del ejercicio de validación de los perfiles propuestos vs. la metodología ASUM-DM

6. Formalismo gráfico de representación de los perfiles CAOBA

Para poder representar, analizar y comparar los elementos del perfil CAOBA vistos hasta el momento, es necesario encontrar un formalismo de representación gráfica del perfil que permita expresar los dominios disciplinares y nivel de logro de habilidades y conocimientos en cada uno de ellos. El grafismo propuesto, ilustrado en la figura 12, se basa en un grafo polar y es denominado estrella de Bloom. La estrella de Bloom hace explícitas las fortalezas, en términos de habilidades y conocimientos presentes en los proyectos de *Big Data* y *Data Analytics*.

6.1. Estrella de Bloom

La estrella de Bloom presentada en la figura 12 está compuesta por cinco ejes, que corresponden a los dominios de conocimiento involucrados en los proyectos de *Big Data* y *Data Analytics*: negocio, TI, analítica, administración de proyectos y habilidades transversales. Cada uno de los ejes distingue los 6 niveles de conocimiento de la taxonomía de Bloom (3.3), mediante una escala de colores: el nivel central corresponde a 'Conocer' y el nivel externo corresponde a 'Evaluar'. Con la definición en cada eje de un nivel de conocimiento se conforma un polígono de cinco lados que representa un perfil de conocimientos y habilidades en el área de *Big Data* y *Data Analytics*. Este puede ser el perfil deseado de una persona o el perfil necesario para el desarrollo de un proyecto o el perfil de un curso en el área de *Big Data* y *Data Analytics*, como se verá en las secciones siguientes.

6.2. Los perfiles profesionales según la estrella de Bloom

La validez de la estrella de Bloom está directamente relacionada con la facilidad para expresar perfiles existentes y describir espacios de formación. En esta sección se muestra el perfil de científico de datos y del *citizen Data Scientist* en términos de lo Estrella de Bloom.

La representación gráfica del *Data Scientist* se puede ver en la figura 13, que representa una persona con conocimientos y habilidades máximos en cada eje de la estrella. teniendo en cuenta

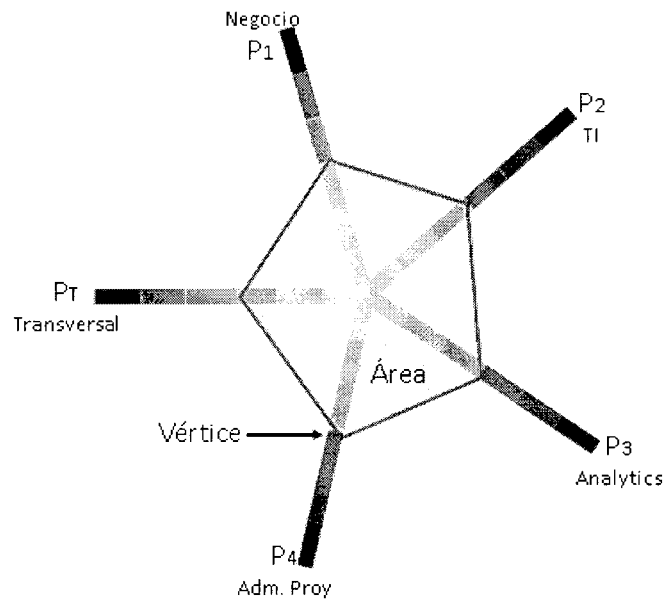


Figura 12: Estrella de Bloom

que el área de su pentágono es la máxima para la representación del conocimiento usando la estrella de Bloom, se concluye que este perfil es el más completo usando esta representación.

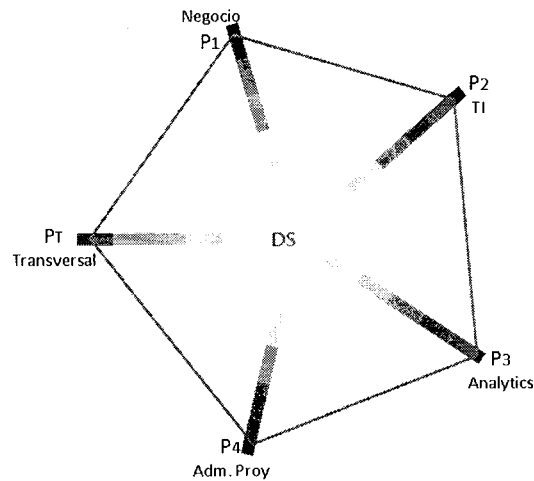


Figura 13: Estrella de Bloom para el perfil de científico de datos

En palabras de Gartner "El *citizen Data Scientist* es una persona que crea o genera modelos que aprovechan la analítica predictiva o prescriptiva, pero cuya función de trabajo principal se encuentra fuera del ámbito de la estadística y el análisis" [8].

La figura 14 muestra una representación de la ubicación del científico de datos y el *Citizen Data Scientist* según el autor Carlos Guadián [8].

La figura 15 representa los niveles de conocimiento que debe poseer un *citizen Data Scientist*.

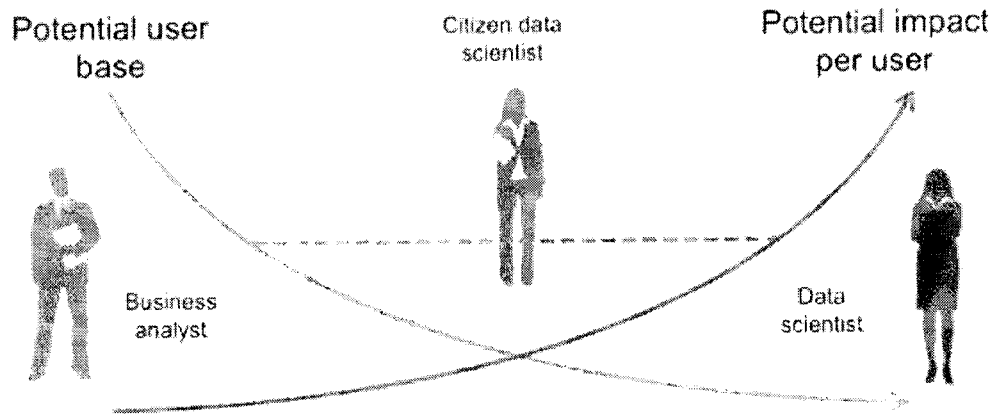


Figura 14: Citizen Data Scientist y Data Scientist

Como se observa en los vértices del pentágono, en los ejes de negocio, habilidades transversales y administración de proyectos tiene conocimientos básicos, lo que le permite conocer los temas sin llegar a la capacidad de crear o evaluar; y los dos vértices restantes del pentágono muestran un mayor nivel de conocimientos en los ejes de TI y analítica, lo que permite aplicar y analizar conocimientos de estas áreas que son su fortaleza.

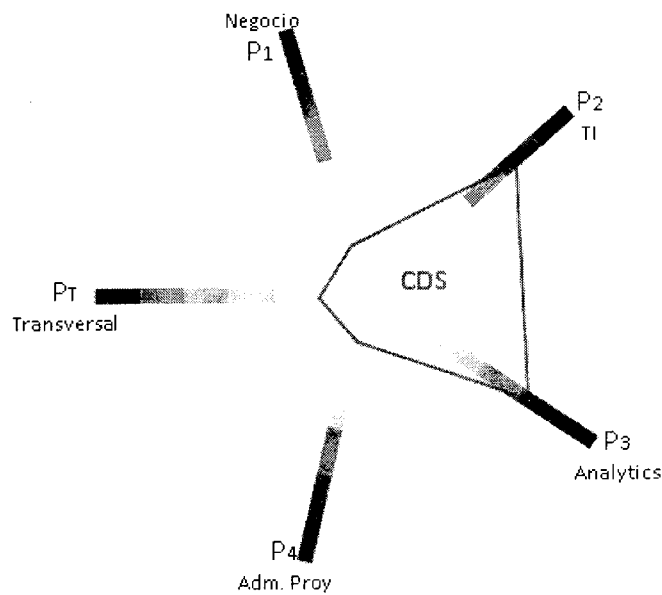


Figura 15: Estrella de Bloom para el perfil de *Citizen Data Scientist*

6.3. Los perfiles Caoba según la estrella de Bloom

A continuación, se puede observar la descripción de los perfiles mencionados y explicados en la sección 4, utilizando como representación la estrella de Bloom.

6.3.1. Perfil Núcleo - P0

El perfil núcleo (P0) incluye las habilidades y conocimientos que debe poseer cada uno de los integrantes de un proyecto de *Big Data* y *Data Analytics* y que le debe permitir participar de manera efectiva en uno de tales proyectos. Implica el uso de un vocabulario común, habilidades para el trabajo en equipo y de comunicación, el conocimiento de la metodología y la capacidad de comprender las decisiones que se tomen en los cinco ejes de los proyectos de *Big Data* y *Data Analytics*.

La figura 16 muestra gráficamente este perfil, descrito de manera detallada en la sección 4.3

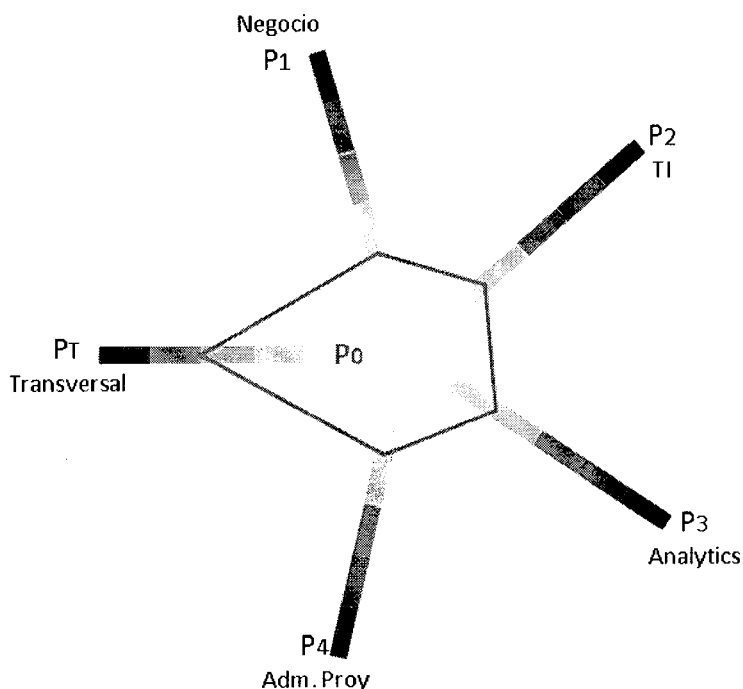


Figura 16: Estrella de Bloom para el perfil CAOBA-Núcleo (P0)

6.3.2. Perfil de Negocio - P1

El perfil de negocio (P1) aumenta el núcleo con los conocimientos y habilidades con respecto al negocio en el cual se está desarrollando el proyecto de *Big Data* y *Data Analytics*. De esta manera, el nivel de conocimiento en el eje de negocio debe ir hasta el nivel de 'Evaluar' y se conservan los niveles del perfil núcleo en los otros ejes de la estrella de Bloom.

La figura 17 muestra gráficamente este perfil, descrito de manera detallada en la sección 4.4.

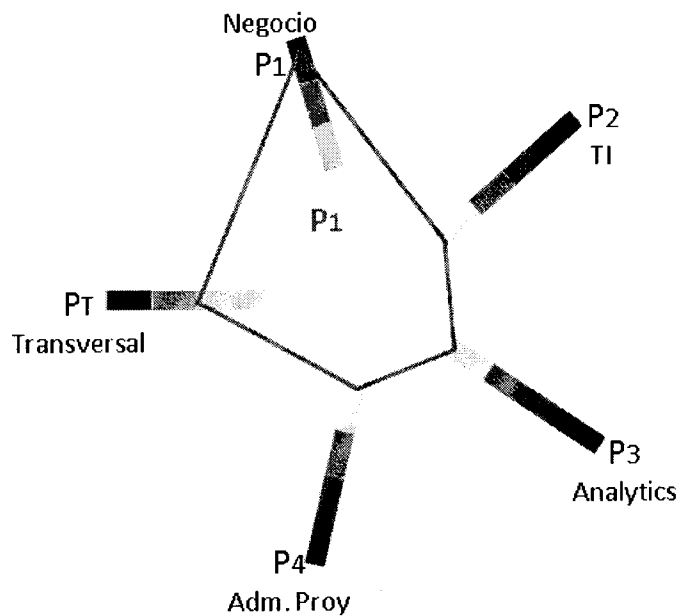


Figura 17: Estrella de Bloom para el perfil CAOBA-Negocio (P1)

6.3.3. Perfil en TI - P2

El perfil en TI (P2) aumenta el núcleo con los conocimientos y habilidades con respecto a las tecnologías de información involucradas en el proyecto de *Big Data* y *Data Analytics*. De esta manera, el nivel de conocimiento en el eje de TI debe ir hasta el nivel de 'Evaluar' y se conservan los niveles del perfil núcleo en los otros ejes de la estrella de Bloom.

La figura 18 muestra gráficamente este perfil, descrito de manera detallada en la sección 4.5

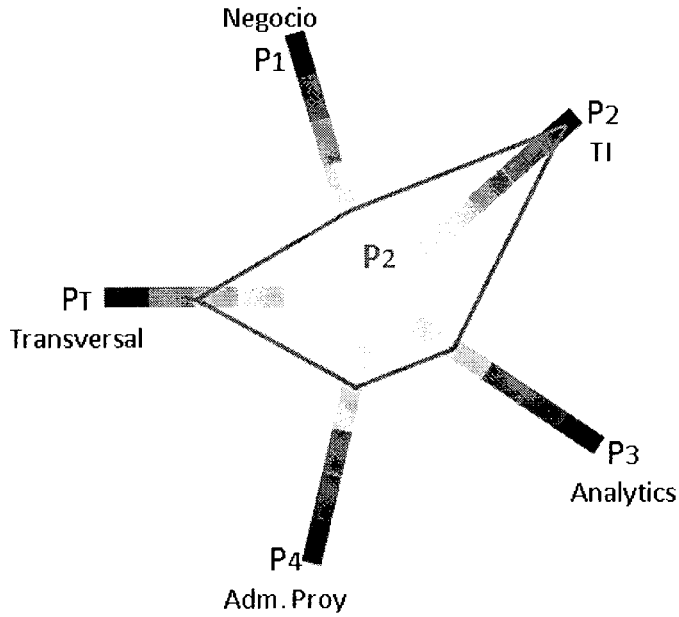


Figura 18: Estrella de Bloom para el perfil CAOBA- TI (P2)

6.3.4. Perfil en Analítica - P3

El perfil en Analítica (P3) aumenta el núcleo con los conocimientos y habilidades con respecto a estadística, modelaje y análisis requeridos para el desarrollo de un proyecto de *Big Data* y *Data Analytics*. De esta manera, el nivel de conocimiento en el eje de Analytics debe ir hasta el nivel de 'Evaluar' y se conservan los niveles del perfil núcleo en los otros ejes de la estrella de Bloom.

La figura 19 muestra gráficamente este perfil, descrito de manera detallada en la sección 4.6

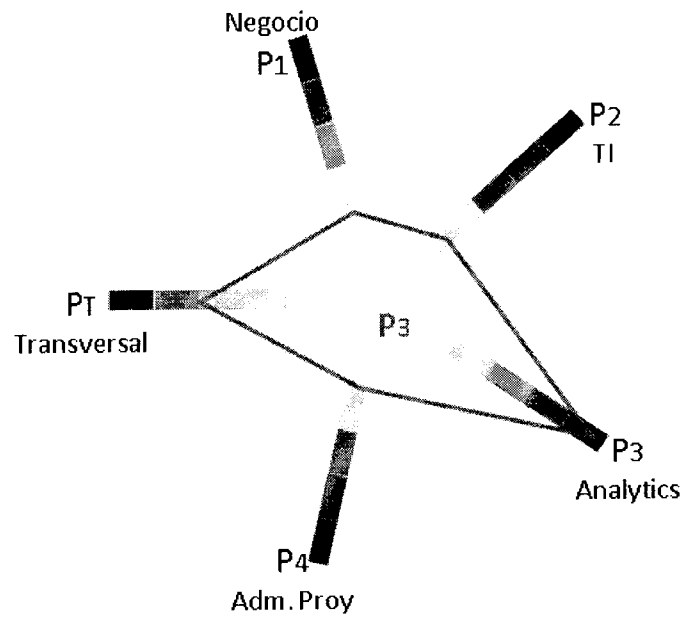


Figura 19: Estrella de Bloom para el perfil CAOBA-Analytics (P3)

6.3.5. Perfil en Administración de proyectos - P4

El perfil en Administración de proyectos (P4) aumenta el núcleo con los conocimientos y habilidades con respecto a la administración de proyectos requeridos para llevar a buen fin un proyecto de *Big Data* y *Data Analytics*. De esta manera, el nivel de conocimiento en el eje de Administración debe ir hasta el nivel de 'Evaluar' y se conservan los niveles del perfil núcleo en los otros ejes de la estrella de Bloom.

La figura 20 muestra gráficamente este perfil, descrito de manera detallada en la sección 4.7

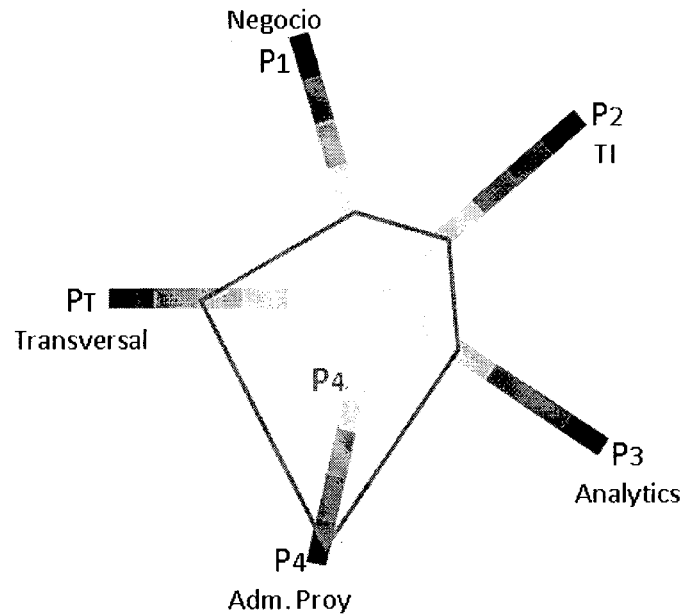


Figura 20: Estrella de Bloom para el perfil CAOBA-Administración de Proyectos (P4)

7. Uso del Perfil

La definición de la estrella de Bloom y la expresión de los diferentes perfiles mostrados en la sección 6 permite plantear que este formalismo puede ser utilizado como marco de referencia para describir las habilidades y conocimientos presentes en los proyectos de *Big Data* y *Data Analytics*.

En esta sección se ilustra entonces el uso de la estrella de Bloom, en la descripción de perfiles requeridos por las empresas y de cursos de formación ofrecidos por Universidades.

7.1. Oferta de formación académica

En la actualidad las universidades cuentan con cursos dentro de sus programas de maestría o de educación continuada, que permiten fortalecer determinadas áreas del conocimiento, relacionadas con el desarrollo eficiente y eficaz de proyectos de *Big Data* y *Data Analytics*.

El objetivo de esta oferta es ofrecer un conjunto de habilidades, técnicas y conocimientos a ser adquiridos por un estudiante durante un curso, que representa el perfil de salida del curso. Ahora bien, los cursos tienen una duración y dedicación para lograr esos objetivos, pero entonces exigen que los estudiantes tengan, antes de empezar, un cierto nivel de conocimientos y habilidades, lo que representa el perfil de entrada al curso.

Todos estos elementos pueden ser representados mediante la estrella de Bloom, como lo muestra la figura 21, donde

- Perfil de entrada: Son los conocimientos previos que debe tener una persona antes de tomar un curso, representado por el polígono PI, de color azul, de la figura 21.
- Perfil de salida: Indica el nivel de habilidades y conocimientos que se adquieren y fortalecen durante el curso, representado por el polígono rojo (PO) de figura 21.

Siguiendo este marco de referencia, a continuación se muestran algunos de los cursos ofrecidos por algunas de las universidades pertenecientes a la Alianza CAOBA. Se debe resaltar que este ejercicio puede ser realizado por cualquier institución académica y que puede ser realizado a nivel de curso, o de programa académico completo.

7.1.1. Cursos ofrecidos por la Universidad de los Andes

a) **Big Data e Ingeniería de Información: Desafíos y oportunidades (Educación continuada Uniandes)**

Este es un curso de divulgación que se dicta en modalidad de educación continuada, con duración de 12 horas teóricas. Por ser de divulgación, no requiere de conocimientos previos en el tema y debido a su duración el nivel de salida se restringe al nivel de 'Conocer'.

Los objetivos que busca cumplir este curso son [4]:

- Comprender dimensiones de cambio y oportunidades que se presentan a través de la convergencia tecnológica y la información.

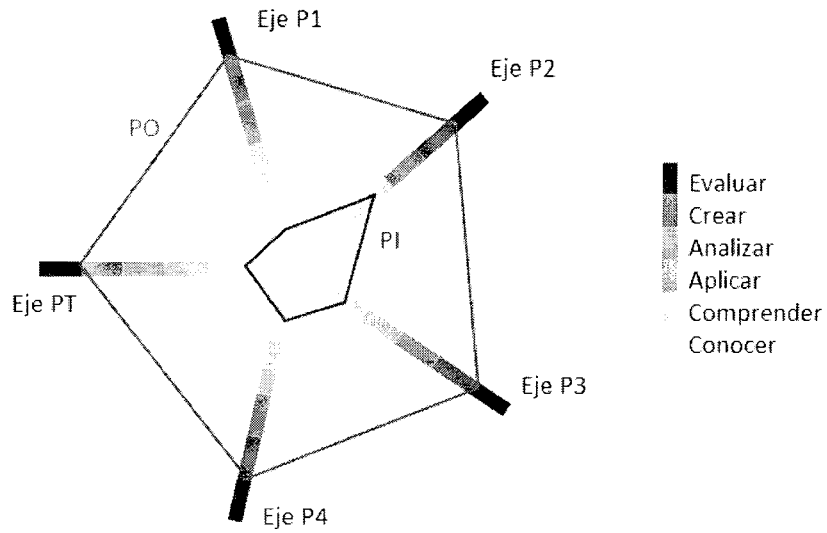


Figura 21: Representación del perfil de entrada y de salida de una oferta académica

- Visualizar el impacto y sus ámbitos de aplicación en todas las áreas del quehacer: la biología, la física, la geociencia, la medicina, la ingeniería, entre otras.
- Comprender los desafíos que desde el punto de vista tecnológico implica una solución basada en Big Data.

El contenido responde preguntas como ¿Qué es big data? ¿Por qué surge? ¿Cómo es la tecnología para construir soluciones de big data.

La figura 22 muestra este curso mediante el formalismo de la estrella de Bloom.

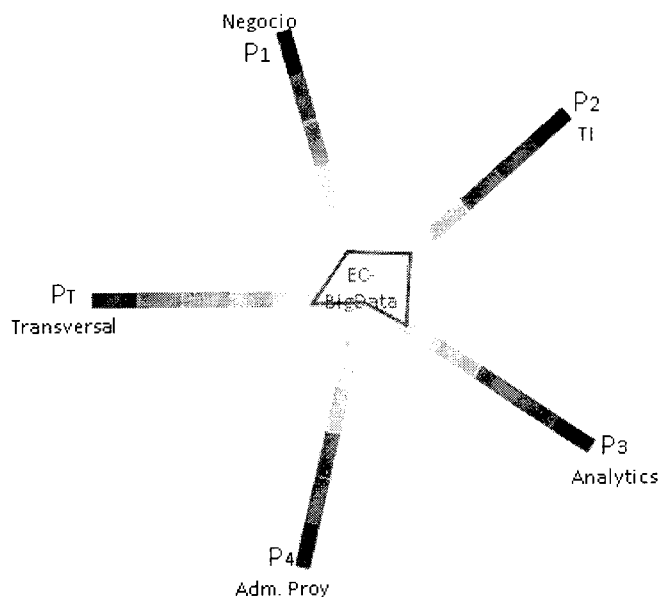


Figura 22: Perfil de entrada y salida del curso Big Data e Ingeniería de Información: Desafíos y oportunidades

b) Análisis de información sobre Big Data (Maestría en Ingeniería de Información)

Este curso cuenta con 45 horas teóricas presenciales, talleres, laboratorios y exámenes. Se enfoca en los ejes área de TI y de analítica, y exige conocimientos relativamente avanzados en TI y básicos en los otros ejes.

Los objetivos que se cumplen en este curso son [4]:

- Identificar las oportunidades de transformación y generación de procesos de generación de valor basadas en el análisis de información, proveniente tanto de fuentes internas como externas a la organización.
- Comprender, definir y evaluar arquitecturas orientadas por datos (*Scalable Data-Driven Architectures*), en particular aquellas que involucran requerimientos de alta escalabilidad de procesamiento y almacenamiento
- Integrar metodologías y tecnología para el descubrimiento y entendimiento de información basado en fuentes altamente escalables. Ejemplos de ellas son *Linked Data, Social Data, Sentiment Analysis, Online Stream Analysis, Web Intelligence*.
- Integrar metodologías y tecnología de análisis de información apropiadas para escenarios de datos no estructurados o semi-estructurados, en los cuales se enfrenta el proceso de análisis en condiciones de la alta escalabilidad y son necesarios procesos de adaptación y reacción en tiempo real.
- Desarrollar una solución que permita generar valor y diferenciación a partir de procesos de análisis de información sobre big data.

La figura 23 muestra este curso mediante el formalismo de la estrella de Bloom.

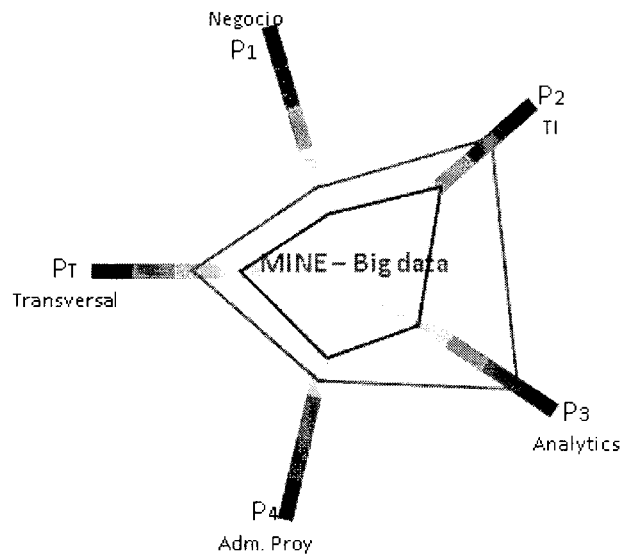


Figura 23: Perfil de entrada y salida del curso Análisis de información sobre Big Data

c) **Ciencia de Datos Aplicada (Maestría en Ingeniería de Información)**

Este curso requiere de 45 horas presenciales (30 de ellas prácticas), talleres semanales y de la implementación de un proyecto final utilizando datos de una organización real.

Se enfoca en los ejes de Negocio, analítica y TI, exigiendo un nivel relativamente alto en el eje de analítica.

Los objetivos del curso son [4]:

- Desarrollar habilidades para la selección y uso de herramientas y técnicas apropiadas para preparar, analizar datos y comunicar información que pueda ser accionable en organizaciones.
- Preparar y analizar fuentes de datos con el fin de extraer información útil para una organización.
- Utilizar de forma apropiada técnicas computacionales y de estadística para generar información.
- Definir un modelo de beneficio para la organización al utilizar la información seleccionada y analizada previamente.
- Comunicar efectivamente los beneficios de realizar un proyecto con los hallazgos identificados, con el fin de que la organización comprenda la utilidad de la información e inicie acciones enfocadas a su uso.

La figura 24 muestra este curso mediante el formalismo de la estrella de Bloom.

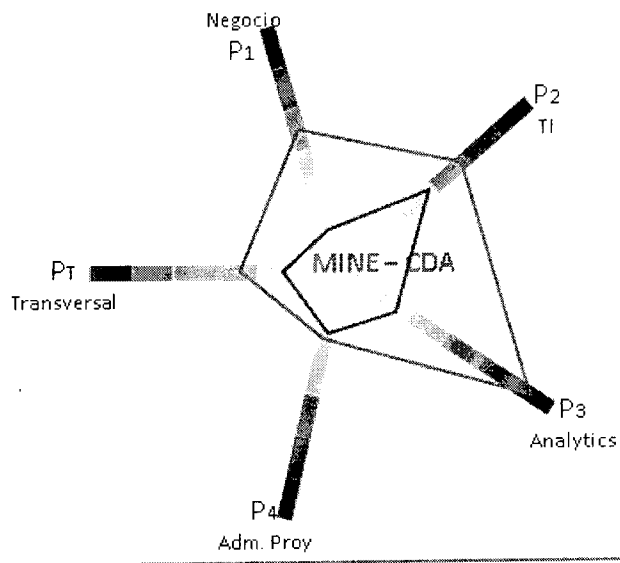


Figura 24: Perfil de entrada y salida del curso Ciencia de Datos Aplicada

d) **Business Analytics (Maestría en Tecnologías de Información para el Negocio)**

Este curso requiere de 45 horas presenciales, talleres quincenales, evaluaciones y la realización de un proyecto final sobre un caso de estudio real.

Se enfoca en los ejes de Negocio y analítica, exigiendo un nivel relativamente alto en los dos ejes.

Los objetivos del curso son [4]:

- Explicar a los ejecutivos de la organización la importancia de utilizar *Business Analytics* en las organizaciones para generar ventajas competitivas.
- Identificar oportunidades para las empresas desde la perspectiva de *Business Analytics*, que les permitan ser competitivas. Identificar requerimientos analíticos que apoyen decisiones tácticas y estratégicas de la organización y que incluyan Minería de Datos en la construcción de la solución.
- Utilizar metodologías y herramientas para construir modelos de análisis de datos que le permitan responder a un requerimiento analítico alineado con los objetivos y metas de la organización.
- Entender el proceso de despliegue de una solución de *Business Analytics* en una organización.

La figura 25 muestra este curso mediante el formalismo de la estrella de Bloom.

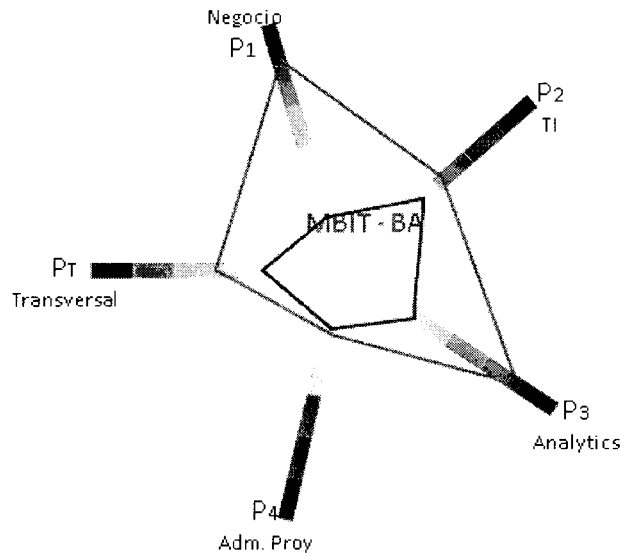


Figura 25: Perfil de entrada y salida del curso Business Analytics

e) **Visual Analytics (Maestría en Ingeniería de Sistemas y Computación)**

Este curso requiere de 45 horas presenciales, talleres quincenales, evaluaciones y la realización de un proyecto final sobre un caso de estudio real.

Se enfoca en los ejes de TI y analítica, exigiendo un nivel relativamente alto en el eje de analítica.

Visual Analytics, o Analítica visual, es el conjunto de conocimiento que permite utilizar técnicas de visualización interactiva con algoritmos y métodos de análisis de datos con el fin de apoyar el razonamiento analítico para la toma de decisiones. La Analítica visual es utilizada en áreas muy diversas que cubren ciencias, ingeniería, negocios y gobierno. Encontramos ejemplos de áreas de aplicación en análisis y toma de decisiones en sistemas urbanos, exploración petrolera, banca, seguridad, emergencias y desastres naturales, monitoreo de salud para manejo de epidemias, entre otros.

El contenido del curso posee tanto bases teóricas de esta área interdisciplinaria, análisis de casos de aplicación, como la realización de miniproyectos de aplicación de técnicas de analítica visual en situaciones reales [4].

La figura 26 muestra este curso mediante el formalismo de la estrella de Bloom.

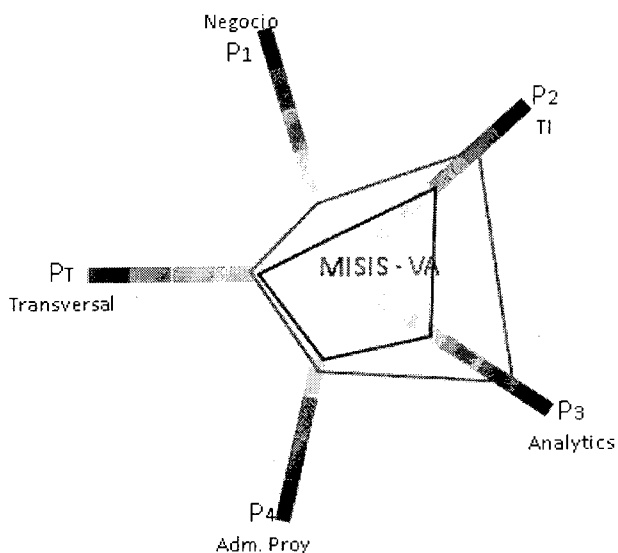


Figura 26: Perfil de entrada y salida del curso Visual Analytics

7.1.2. Programas y Cursos ofrecidos por la Pontificia Universidad Javeriana

a) Diplomado en Analítica de Datos

Este es un programa de Diplomado, con duración de 111 horas, que incluye sesiones de teoría y de práctica usando herramientas comerciales y también de código abierto (*open source*).

El objetivo general que busca cumplir este programa es promover en los participantes una actitud analítica basada en datos para resolver problemas organizacionales y fortalecer la toma de decisiones gracias a la generación de modelos descriptivos y predictivos a partir de datos estructurados y narrativos.

Los objetivos específicos del programa son:

- Fortalecer la capacidad de identificar necesidades y oportunidades de negocio y traducirlas en términos de técnicas y métodos de analítica de datos.
- Presentar las diferentes técnicas de analítica de datos contemplando necesidades organizacionales de índole descriptivo y predictivo.
- Reconocer nuevas tendencias y necesidades en la analítica de datos producto de la diversidad y alto volumen de datos disponibles actualmente.

La figura 27 muestra este curso mediante el formalismo de la estrella de Bloom.

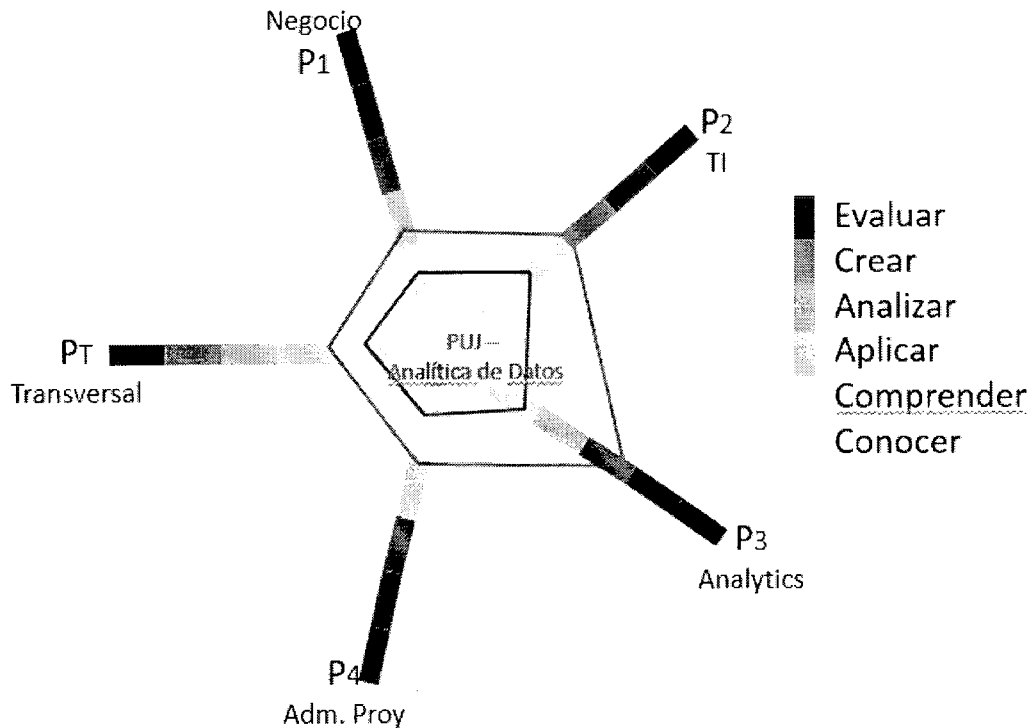


Figura 27: Perfil de entrada y salida del Diplomado en Analítica de Datos

b) **Curso de Analítica de Texto**

Este curso tiene una duración de 60 horas, con sesiones de teoría y de práctica usando herramientas comerciales y también de código abierto (*open source*).

El objetivo principal de este curso es promover en los participantes una actitud analítica basada en datos que originalmente se encuentran en texto narrativo para resolver problemas organizacionales y fortalecer la toma de decisiones, gracias a la identificación de patrones y a la generación de modelos a partir de este tipo de datos.

Los objetivos específicos del curso son:

- A partir del análisis de casos de éxito, reconocer la utilidad de la analítica de los textos narrativos para fortalecer el entendimiento empresarial.
- Identificar las principales técnicas de obtención y pre-procesamiento de texto narrativo empleados para estructurarlos y así permitir realizar análisis sobre ellos.
- Reconocer las técnicas descriptivas y predictivas más empleadas en la analítica de texto y cómo deben ser interpretados sus resultados.

La figura 28 muestra este curso mediante el formalismo de la estrella de Bloom.

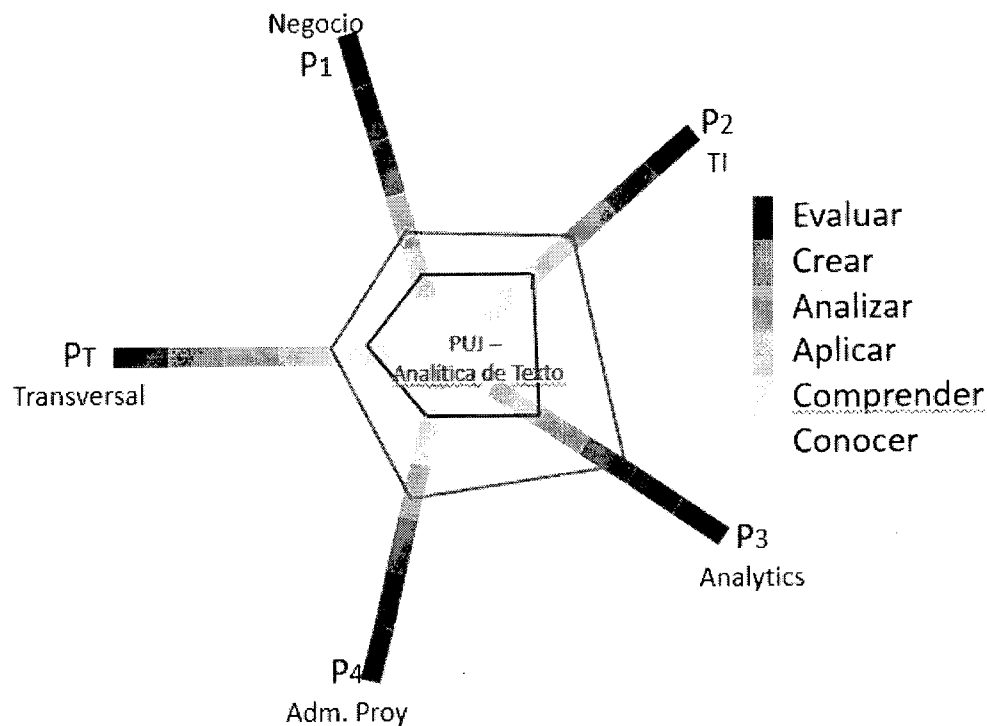


Figura 28: Perfil de entrada y salida del Diplomado en Analítica de Datos

7.1.3. Programas y Cursos ofrecidos por la Universidad ICESI

a) Diplomado en Analítica de Datos y Big Data

Este es un programa de Diplomado, con duración de 120 horas, organizadas en 30 sesiones de 4 horas cada una.

Los objetivos del programa son:

- Identificar el potencial de generar valor a partir de un conjunto de datos resultantes de la operación del negocio.
- Formular preguntas relevantes para el negocio, que puedan ser respondidas a partir del análisis de los datos disponibles.
- Aplicar las técnicas de analítica más adecuadas para construir un modelo de solución.
- Evaluar los resultados obtenidos a partir de diferentes modelos analíticos.
- Aplicar técnicas de visualización para comunicar efectivamente los resultados de los modelos analíticos.
- Reconocer las implicaciones en la infraestructura computacional que se derivan del tratamiento de grandes volúmenes de información en proyectos de analítica de datos.
- Aplicar la metodología ASUM-DM al desarrollo de proyectos de analítica de datos.

La figura 29 muestra este curso mediante el formalismo de la estrella de Bloom.

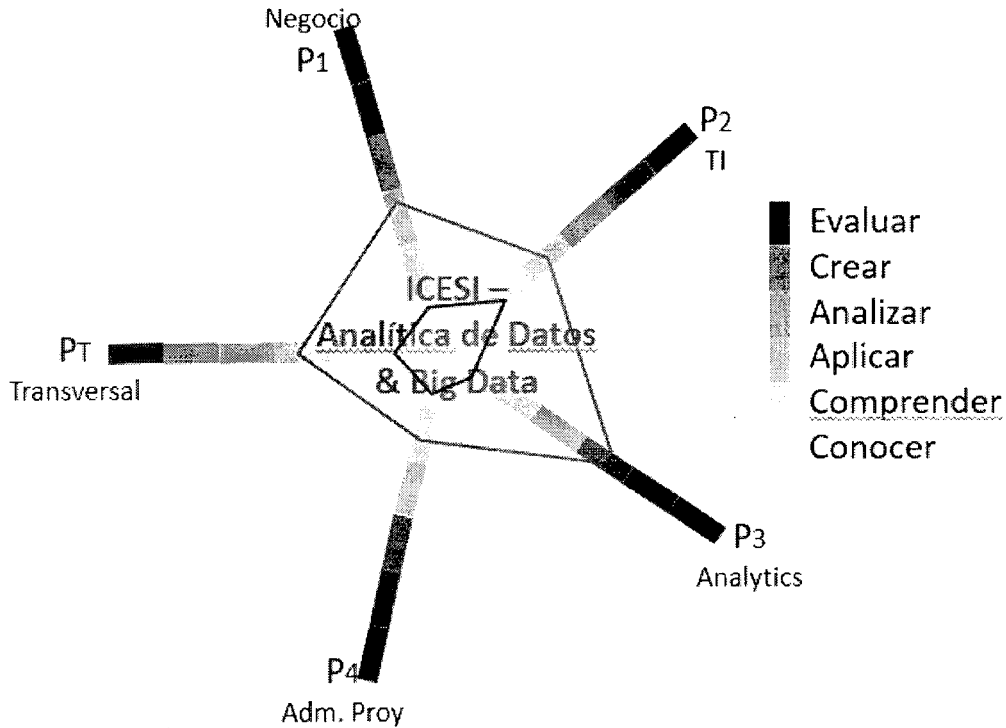


Figura 29: Perfil de entrada y salida del Diplomado en Analítica de Datos

7.2. Perfil para la demanda académica

Al igual que las instituciones académicas pueden expresar su oferta en términos de la estrella de Bloom, las empresas pueden usarla, por ejemplo, para:

- Definir sus necesidades de formación general de la empresa, con base en planes estratégicos de evolución de la empresa
- Determinar las necesidad de formación específica para el desarrollo de un proyecto, con base en los objetivos del mismo.
- Determinar el personal necesario, y su perfil, para completar el equipo de trabajo necesario para el desarrollo de un proyecto. Esto se logra con base en el perfil requerido para el proyecto y los perfiles de las personas que están actualmente en el equipo de trabajo.

8. Conclusiones

Las principales conclusiones de este trabajo son:

- La propuesta de perfiles especializados en cada uno de los dominios permite expresar claramente las habilidades y conocimientos presentes en un proyecto de *Big Data* y *Data Analytics*, ya que fue validada contra la metodología ASUM-DM.
- Es más completa que otras propuestas de perfiles para este tipo de proyectos, pues incluye explícitamente los ejes de administración de proyectos y de las habilidades suaves (*soft skills*) que son necesarias para el desarrollo exitoso de estos proyectos.
- El formalismo gráfico propuesto, llamado la Estrella de Bloom, es poderoso en el sentido que expresa de manera clara y sintética el nivel o los niveles de habilidades y conocimientos de interés, como por ejemplo:
 - El perfil de una persona
 - El perfil de una propuesta de formación (perfil de salida) de un curso, o de un programa de formación, mediante la unión de los perfiles de salida de cada uno de los cursos.
 - El perfil requerido para acceder a la propuesta de formación (perfil de entrada), que cuando se ponen juntos en una misma estrella permite visualizar el énfasis de la propuesta de formación y el esfuerzo que se requiere para alcanzar los objetivos. En este sentido facilita la concepción de los programas o cursos ofrecidos.
 - El perfil deseado por una empresa para afrontar con éxito el desarrollo de un proyecto de *Big Data* y *Data Analytics*. Esta demanda puede ser a nivel de la empresa completa o a nivel del equipo de trabajo para un proyecto en particular, o para la contratación de personal especializado.

Referencias

- [1] BIGDATACoeE Barcelona. Los 7 perfiles clave de los profesionales del big data. <http://www.bigdatabcn.com/es/los-7-perfiles-clave-los-profesionales-big-data/>, 2016. Accedido 08-02-2017.
- [2] M. D.; Furst E. J.; Hill W. H.; Krathwohl D. R. Bloom, B. S.; Engelhart. *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay Company, New York, 1956.
- [3] Giovanni Cock Gómez et al. Big data en las empresas: una nueva era de la información. B.S. thesis, Universidad Militar Nueva Granada, 2015.
- [4] Universidad de los Andes. Universidad de los andes. <https://www.uniandes.edu.co/> , 2017. Accedido 20-02-2017.
- [5] Oxford dictionaries. Data scientist. https://en.oxforddictionaries.com/definition/data_scientist, 2016.
- [6] Revista Dinero. Big data: la mina de oro. <http://www.dinero.com/edicion-impresa/tecnologia/articulo/el-poder-economico-del-big-data-su-desarrollo-colombia/210853>, 2015. Accedido 08-02-2017.
- [7] Jane Griffin. The role of the chief data officer. *Information Management*, 18(2):28, 2008.
- [8] Carlos Guadián. Puedes ser un ciudadano científico de datos (citizen data scientist). <http://www.k-government.com/2016/03/02/ciudadano-cientifico-datos/>, 2017. Accedido 13-02-2017.
- [9] Cristina Juan. Los perfiles profesionales más buscados de big data. <http://comunidad.iebschool.com/iebs/software-de-gestion/profesionales-mas-buscados-big-data/>, 2016. Accedido 08-02-2017.
- [10] Leon Kappelman. It trends symposium. https://c.ymcdn.com/sites/www.simnet.org/resource/collection/7A70E28BA-4E88-B958-C86941C704C3/IT_Trends_SIMposium_2014_Keynote_Intro_Slides.pdf, 2015. Accedido 15-03-2017.
- [11] David López García et al. Análisis de las posibilidades de uso de big data en las organizaciones. Master's thesis, Universidad de cantabria, 2013.
- [12] Maribel tirados. La escasez de profesionales big data hace subir los salarios. <http://www.bigdatahispano.org/noticias/la-escasez-de-profesionales-big-data-hace-subir-los-salarios/>, 2014. Accedido 08-02-2017.
- [13] Auribox Training. Curso de big data con hadoop y mongo db-auribox training. www.auriboxtraining.com, 2016. Accedido 20-02-2017.
- [14] Wil MP Van der Aalst. Data scientist: The engineer of the future. In *Enterprise Interoperability VI*, pages 13–26. Springer, 2014.
- [15] Rayo Ángel. Las habilidades y el perfil de un profesional big data. <http://www.bit.es/knowledge-center/habilidades-y-perfil-de-un-profesional-big-data/>, 2016. Accedido 08-02-2017.

DEPARTAMENTO ADMINISTRATIVO DE CIENCIA, TECNOLOGIA E INNOVACIÓN
- COLCIENCIAS -

CONVOCATORIA PARA LA FORMACIÓN DE CIUDADANOS EN CIENCIA DE DATOS

ANEXO 2 - CARTA DE COMPROMISO

(Ciudad). (Día) de (Mes) de 201_

Señores

COLCIENCIAS

Av. Calle 26 # 57-41 Torre 8 pisos 2 al 6
Bogotá D.C.

Asunto: Aval y compromiso del (nombre persona natural).

Respetados señores,

Yo, _____, identificado(a) con cédula de ciudadanía No. _____, declaro bajo juramento, la veracidad de la información y documentación que entrego para postularme como posible beneficiario del Curso "Ciudadano de Datos" (Citizen Data Scientist) – convocatoria para la formación de ciudadanos en ciencia de datos.

De la misma forma, mediante la presente me permito establecer los compromisos que enuncio a continuación:

1. Contrapartida:

Se establece el compromiso que en caso de que la propuesta resulte financiable por COLCIENCIAS – Ministerio TIC, EL BENEFICIARIO aportará los recursos como contrapartida y que han sido diligenciados en el Sistema Integral de Gestión de Proyectos –SIGP–, cuyo valor asciende a \$400.000 (no reembolsables), y pagar dicha suma a la Pontificia Universidad Javeriana, por concepto de inscripción del curso Ciudadano de Datos (Citizen Data Scientist).

(Para generar el reporte el proponente debe ingresar al formulario del SIGP, hacer clic en el menú "Reportes", opción Reportes y seleccionar la casilla identificada como "Reporte de contrapartidas por entidad". Finalmente debe hacer clic en el botón "Generar reporte" e imprimirlo para adjuntarlo al presente documento).

2. De la aceptación de las condiciones y términos de referencia que establece COLCIENCIAS:

En virtud de la presente CARTA DE COMPROMISO, en el marco de ejecución de la convocatoria para la formación de ciudadanos en ciencia de datos, EL BENEFICIARIO declara y acepta que:

1. Tengo poder y/o representación legal para firmar y presentar la propuesta.
2. La información suministrada es veraz y no fija condiciones artificiales.
3. Aceptan y reconocen que cualquier omisión o inconsistencia en la que hayan podido incurrir y que pueda influir en la propuesta, no les eximirá de la obligación de asumir las responsabilidades que les llegue a corresponder como futuros contratistas y renuncian a cualquier reclamación, reembolso o ajuste de cualquier naturaleza, por cualquier situación que surja y no haya sido contemplada en razón de la falta de diligencia en la obtención de la información.
4. No se encuentran incursos en ninguna de las causales de inhabilidad y/o incompatibilidad establecidas en el Estatuto General de Contratación y demás normas legales pertinentes.
5. Aceptan y autorizan a COLCIENCIAS para que verifique la información aportada en la propuesta.
6. Se encuentran al día con las obligaciones y compromisos adquiridos con Colciencias.
7. Asistir a los cursos, en las fechas y horas programadas, las cuales dependerán del número mínimo de inscritos y podrán estar sujetas a modificaciones.
8. El participante deberá suscribir un pagaré a favor de la Pontificia Universidad en representación del Centro de Excelencia en Big Data y Data Analytics (CAOBA) por un valor de cuatro millones de pesos (\$4.000.000), el cual se hará efectivo en caso de no lograr las certificaciones de asistencia y suficiencia.
9. Presentar el examen de certificación del curso ciudadano de datos en las fechas estipuladas.

Además, **ACEPTAMOS** expresa e irrevocablemente que conocemos detalladamente las características, requisitos y condiciones de la convocatoria (nombre de la convocatoria), de manera que nos sometemos a lo establecido en los Términos de Referencia y los anexos determinados por COLCIENCIAS para el desarrollo de la misma y para la entrega del recurso en caso que la propuesta resulte financiable.

Declaramos que la información suministrada es veraz, corresponde a la realidad y es coherente con lo consignado en el Sistema Integral de Gestión de Proyectos - SIGP. En caso de encontrarse alguna incoherencia o inconsistencia en la información o documentación suministrada, COLCIENCIAS podrá en cualquier momento, rechazar esta propuesta o finiquitar el beneficio, sin perjuicio de las acciones legales correspondientes”.

Atentamente,

EL BENEFICIARIO

FIRMA

NOMBRE PERSONA NATURAL

CC

DIRECCION

TELÉFONO