

Bacharelado em Ciência de Dados

Resumo

Este documento apresenta os referenciais de formação na área de Computação para os cursos de Bacharelado em Ciência da Dados (RF-CD-21). Estes Referenciais foram construídos em torno da noção de competência, em consonância com as competências definidas pela Força Tarefa em Ciência de Dados da Association for Computing Machinery (ACM) em 2021 (ACM Data Science Task Force (2021)). Assim como feito pela SBC na preparação de um Currículo de Referência para outras áreas da Computação, as 17 (dezesete) competências apontadas como necessárias estão resumidas em 8 (oito) eixos de formação, de forma a facilitar a construção de currículos nas Instituições de Ensino Superior (IES) brasileiras. Cada eixo de formação relaciona os conteúdos considerados úteis no desenvolvimento das competências necessárias. Por fim, este referencial busca nortear a construção de um Projeto Pedagógico de Curso (PPC) para cursos de graduação em Ciência de Dados pelas IES, proporcionando flexibilidade para que cada uma delas defina seus PPC conforme sua vocação e seus objetivos.

II.1. Apresentação

A Diretoria de Educação da Sociedade Brasileira de Computação (SBC), ciente da demanda por um Referencial de Formação para as áreas de Ciência de Dados, Inteligência Artificial e Cibersegurança, consultou as Comissões Especiais de Banco de Dados, de Inteligência Artificial, de Inteligência Computacional e de Segurança para estudar e propor Referenciais de Formação para três 3 novos curso de Bacharelado em temas relacionados à Computação: Ciência de Dados, Inteligência Artificial e Cibersegurança.

Com aprovação de sua Diretoria e Conselho, a SBC, por meio de sua Diretoria de Educação, designou uma comissão para elaborar um Referencial de Formação para o curso de Bacharelado em Ciência de Dados (RF-CD). A comissão foi composta pelos professores Angelo Roncalli Alencar Brayner (coordenador), do Departamento de Computação (DC) da Universidade Federal do Ceará (UFC), André Carlos Ponce de Leon Ferreira de Carvalho, do Instituto de Ciências Matemáticas e de Computação (ICMC) da Universidade de São Paulo (USP), Duncan Ruiz, da Escola Politécnica (EP) da Pontifícia Católica do Rio Grande do Sul (PUCRS) e Eduardo Ogasawara, do Departamento de Computação (DC) do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ).

O trabalho da comissão teve início em 2021 e foi orientado pelas seguintes diretrizes:

1. Estar alinhado com as Diretrizes Curriculares Nacionais (DCNs) (MEC, 2016).
2. Seguir um modelo baseado em competências (Referenciais de Formação para os Cursos de Graduação em Computação, 2017).

A metodologia adotada para a construção deste documento se baseia no paradigma de orientação a competências esperadas do egresso do curso. Mais especificamente, adotou-se o modelo de referência baseado na Taxonomia de Bloom Revisada (Ferraz e Belhot, 2010). Neste modelo, uma competência pode expressar o conhecimento, as habilidades ou as atitudes esperadas do egresso do curso, sob a perspectiva de objetivos de aprendizagem ("o que o aluno será capaz de"). Além disso, a articulação e estruturação das competências foram estabelecidas em eixos temáticos de formação (Anastasiou, 2010).

II.2. Breve histórico do curso

Em 2016, com base no Parecer CNE/CES N° 136/2012, foi definida a Resolução CNE/CES N° 5/2016 (MEC, 2016), que institui as DCNs para os cursos de graduação na área da Computação, abrangendo os cursos de Bacharelado em Ciência da Computação, em Sistemas de Informação, em Engenharia de

43 Computação e em Engenharia de Software e de Licenciatura em Computação, homologada e publicada no
44 D.O.U. de 17/11/2016, Seção 1, Pág. 22. Esta resolução e o correspondente Parecer CNE/CES Nº 136/2012
45 são conjuntamente referenciados como DCN16.

46 **II.3. Os benefícios do curso para a Sociedade**

47 A quantidade e complexidade de dados gerados por seres humanos e máquinas aumentam a uma velocidade
48 crescente. Esses dados contêm informações relevantes que podem resultar em avanços não apenas em
49 outras áreas do conhecimento, mas também na oferta de melhores serviços públicos, desenvolvimento
50 social, crescimento econômico e aumento da oferta de empregos. Os empregos gerados podem reduzir
51 riscos não só para os seres humanos, mas também para todo o planeta, facilitando a preservação e
52 recuperação do meio ambiente. Para que isso ocorra, é importante a formação de recursos humanos em toda
53 a cadeia de análise de dados com sólido conhecimento nas áreas de Computação, Estatística e Matemática.
54 Esta boa formação possibilitará o desenvolvimento de ferramentas computacionais capazes de explorar de
55 uma forma competente e eficiente os dados gerados e apoiar o processo de tomada de decisão nos eixos
56 academia, governo e empresas.

57 **II.4. Aspectos relacionados com a formação de um profissional em Ciência de Dados**

58 O Bacharel em Ciência de Dados deve ser capaz de compreender o funcionamento dos processos de coleta,
59 gerenciamento e análise eficientes de dados em larga escala, em diferentes formatos e de fontes
60 heterogêneas. Referente aos dados, deve estar apto a definir e executar procedimentos de gestão e uso,
61 desenvolver estratégias para formular e testar hipóteses, interpretar e extrair conhecimentos úteis e
62 significativos, desenvolver e validar algoritmos para analisá-los, interagir com profissionais de outras áreas
63 de conhecimento e trabalhar de forma ética e colaborativa.

64 O Bacharel em Ciência de Dados deve ter a capacidade de atuar em várias áreas, grande parte delas
65 de forma ética e colaborativa, com profissionais de outras áreas do conhecimento, que entendam o
66 significado dos dados analisados e possam validar os resultados obtidos. Assim, os Bacharéis em Ciência
67 de Dados estarão qualificados para analisar dados de áreas tão diversas como administração, agronomia,
68 ciências florestais, ciências sociais, economia, engenharia, geografia, história, medicina e veterinária, para
69 citar algumas.

70 **II.5. Perfil do egresso**

71 A seguir, são apresentados o perfil geral dos egressos de cursos de Bacharelado em Ciência de Dados. O
72 curso de Bacharelado em Ciência de Dados visa a formação profissionais capazes de “pensar com dados”,
73 com competência teórica, (técnica, e metodológica,) e experiência prática para lidar com as mais variadas
74 situações e domínios de aplicação. Em linhas gerais, o egresso deve ser capaz de (i) entender, formular e
75 refinar as questões apropriadas;, (ii) obter, modelar e explorar os dados relacionados;, (iii) processar os
76 dados e realizar as análises necessárias;, (iv) obter e comunicar o conhecimento relevante;, (v) apoiar o
77 desenvolvimento e implantação de soluções com base nos resultados atingidos; e, (vi) entender e atender
78 aspectos éticos e sociais relacionados à sua atuação. Para isso, deve apresentar as seguintes habilidades e
79 competências:

- 80 1. Ter sólida formação nas áreas de Computação, Matemática e Estatística, que permita a aplicação de
81 conceitos dessas áreas em tarefas de Ciência de Dados.
- 82 2. Utilizar efetivamente técnicas computacionais, matemáticas e estatísticas para, de forma analítica,
83 avaliar a factibilidade e, quando factível, extrair conhecimento dos dados disponíveis, que permita
84 realizar descobertas em diferentes domínios de aplicação, de forma a apoiar o processo de tomada
85 de decisão.
- 86 3. Empregar os princípios de Tecnologia de Informação e Comunicação (TIC) para pesquisar, projetar,
87 implementar e avaliar novas abordagens e técnicas para construção de ferramentas de análise de
88 dados.
- 89 4. Ser capaz de realizar experimentos utilizando diferentes infra estruturas, que apoiem a gestão e o

- manuseio eficiente de dados, estruturados e não estruturados, durante o ciclo de vida dos dados.
5. Definir e implementar estratégias de gestão de dados para curadoria, coleta, integração, armazenamento, visualização, preservação e disponibilização destes para futuro processamento.
 6. Gerenciar projetos interdisciplinares que incluam as diversas etapas do ciclo de vida dos dados.
 7. Identificar novos desafios, necessidades, oportunidades de negócios e desenvolver soluções inovadoras.
 8. Investigar, compreender e estruturar as características de domínios de aplicação em diversos contextos que levem em consideração questões ambientais, éticas, sociais, legais e econômicas;
 9. Trabalhar de forma individual e colaborativa, com profissionais da mesma área ou de diferentes áreas.
 10. Seguir os princípios de uma Ciência de Dados justa, transparente, sem viés e respeitando a privacidade, atendendo aos requisitos da legislação de proteção aos dados em vigor.
 11. Ter uma visão crítica e criativa na identificação e resolução de problemas contribuindo para o desenvolvimento da área.
 12. Atuar de forma empreendedora, abrangente e cooperativa no atendimento às demandas ambientais, sociais e econômicas da região onde atua, do Brasil e do mundo.
 13. Utilizar racionalmente os recursos disponíveis de forma transdisciplinar.
 14. Conseguir atuar em um mundo globalizado do trabalho, buscando o domínio de idiomas estrangeiros, em particular o idioma inglês.

Para atender essas habilidades e competências ao longo do curso, o egresso deverá passar por várias experiências de uso do conhecimento adquirido em múltiplos contextos organizacionais e sociais.

Com uma formação sólida, o profissional formado no curso está apto a trabalhar em empresas de setores diversos, em órgãos públicos e do terceiro setor, além de IES e em institutos de pesquisa.

II.6. Eixos de formação, competências e conteúdos

A estrutura do RF-CD-21 é baseada no modelo conceitual apresentado no Capítulo I dos referenciais de 2017. As 17 competências e habilidades, gerais e derivadas, definidas pelas DCN21 para egressos dos Cursos de Bacharelado em Ciência de Dados, foram agrupadas em 8 (oito) **eixos de formação**.

Cada eixo de formação do RF-CD-21 corresponde a uma macro competência e engloba um grupo de **competências derivadas**, que por sua vez possuirão uma correlação com as competências e habilidades que serão propostas para atualização da DCN16. Em conjunto, as competências associadas aos eixos de formação possibilitam ao egresso de um Bacharelado em Ciência de Dados a lidar com as várias facetas da computação. Os eixos de formação traduzem o entendimento de que tal formação deve levar em conta: a capacidade de atuar em todas as fases que envolvem a aplicação de Ciência de Dados em problemas diversos, desde a concepção de sistemas computacionais até a efetiva implementação de soluções adequadas; a capacidade de se reciclar e buscar novos conhecimentos; e a capacidade de seguir estudos avançados visando ao desenvolvimento da ciência e da tecnologia. Os eixos de formação são os seguintes:

1. Fundamentos de Matemática, Estatística e Computação para Ciência de Dados
2. Resolução de Problemas
3. Desenvolvimento de Sistemas
4. Engenharia e Exploração de Dados
5. Dados em Larga Escala
6. Mineração de Dados e Aprendizado de Máquina
7. Aprendizado Contínuo e Autônomo
8. Ciência, Tecnologia, Inovação e Empreendedorismo

Cada eixo de formação tem a seguinte estrutura:

- **Código:** algarismo indo-arábico que identifica o eixo de formação.
- **Título:** rótulo que identifica o eixo de formação.
- **Descrição:** pequeno texto que contextualiza a competência associada ao eixo de formação.
- **Competência de eixo:** descrição da competência associada ao eixo de formação.
- **Competências derivadas:** lista de competências, oriundas das 48 competências e habilidades, gerais e específicas, definidas pelas DCN21, necessárias para construir a competência de eixo. As competências gerais das DCN21 são indicadas pelo identificador CG e as específicas do curso de Bacharelado em Ciência de Dados, pelo identificador CE. Por sua vez, cada competência derivada é constituída dos seguintes subcampos:
 - **Código:** é formado pela junção da letra C (inicial da palavra “competência”), do código do eixo (1 a 8) e de um número indo-arábico que ordena sequencialmente a competência derivada no contexto do eixo de formação.
 - **Classificação:** um dos seis níveis do processo cognitivo da Taxonomia de Bloom Revisada (Ferraz e Belhot, 2010).
 - **Conteúdo:** lista de conhecimentos que devem ser trabalhados para desenvolver a competência derivada. Cada conteúdo é definido simplesmente por um título, oriundo da listagem constante nas seções 3.1 e 3.2 do Parecer CNE/CSE 136/2012 (MEC, 2012). O detalhamento dos conteúdos, em sua maioria, pode ser obtido nos Currículos de Referência da SBC de 1999 a 2005.

Salienta-se que conteúdo e disciplina não são sinônimos. De fato, a associação entre conteúdos e disciplinas é um dos grandes desafios na elaboração da matriz curricular de cada curso. Uma disciplina oferecida por uma IES em particular poderá abordar mais de um conteúdo elencado nestes referenciais, combinando-os a fim de tratar situações complexas. Ao mesmo tempo, certo conteúdo poderá ser abordado em mais de uma disciplina, evidenciando a sua aplicação em diferentes contextos, possivelmente com diferentes níveis de profundidade. Esse arranjo de conteúdos e disciplinas dependerá, essencialmente, da estratégia adotada por cada curso para o desenvolvimento de competências de seus estudantes.

Uma competência das DCN21 pode estar presente em mais de um eixo, sendo que o conteúdo é específico para cada relacionamento entre eixo de formação e competência das DCN21. Assim, uma competência DCN pode requerer diferentes conteúdos, dependendo do eixo. Da mesma forma, um conteúdo pode estar presente em mais de um eixo. E, ainda, um conteúdo pode estar presente em mais de uma competência das DCN21 de certo eixo.

Um curso pode usar uma estratégia para implementar sua matriz curricular tal que cada disciplina seja desenhada para desenvolver no estudante uma ou mais competências das DCN21, no contexto de um ou mais eixos de formação. Assim, cada disciplina deverá abordar (integral ou parcialmente) os conteúdos recomendados para as respectivas competências das DCN21, de acordo com eixos de formação em questão.

A seguir, cada eixo de formação é detalhado em termos de suas competências derivadas e conteúdos associados.

1. EIXO DE FORMAÇÃO: FUNDAMENTOS DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO PARA CIÊNCIA DE DADOS

COMPETÊNCIA: Entender teorias e princípios básicos das áreas de Computação, Estatística e Matemática; **aplicando** estas teorias e princípios para **resolver** problemas técnicos de Ciência de Dados, incluindo sistemas de aplicação específica.

Competências derivadas	Classificação	Conteúdos
C.1.1. Dominar os principais conceitos de programação.	Aplicar	Algoritmos
		Linguagem de Programação
		Estruturas de dados
C.1.2. Compreender os principais temas de teoria da computação.	Aplicar	Análise de Algoritmos
		Teoria dos Grafos
		Lógica Matemática
		Indução Matemática
		Linguagens Formais e Autômatos Finitos
C.1.3. Compreender os principais temas de estatística básica	Aplicar	Estatística Básica
		Planejamento de Experimentos
		Teste de Hipóteses
		Distribuições de Probabilidade
C.1.4. Compreender os principais temas de matemática básica	Aplicar	Análise Combinatória
		Matemática Discreta
		Cálculo Diferencial e Integral
		Geometria Analítica
		Álgebra Linear e Matricial
		Cálculo Numérico

177

2. EIXO DE FORMAÇÃO: RESOLUÇÃO DE PROBLEMAS

A resolução de problemas por meio da computação é possível com a execução de passos finitos e bem definidos. Nesse sentido, os egressos devem ser “capazes de criar soluções, individualmente ou em equipe, para problemas complexos caracterizados por relações entre domínios de conhecimento e de aplicação” [DCN16].

COMPETÊNCIA: Resolver problemas que tenham solução analítica, considerando os limites da Ciência de Dados, o que inclui:

- a identificação dos problemas que apresentem soluções algorítmicas viáveis;
- a seleção ou criação de algoritmos apropriados para situações particulares;
- a implementação das soluções usando o paradigma de programação adequado;
- o conhecimento da complexidade das soluções encontradas.

Competências derivadas	Classificação	Conteúdos
------------------------	---------------	-----------

C.2.1. <i>Conhecer a importância social da Ciência de Dados</i>	Aplicar	Sustentabilidade
		Ética em Ciência de Dados, Privacidade, Legislação e Direitos Civis
		Projetos Comunitários
C.2.2. <i>Conhecer a importância científica e econômica da Ciência de Dados</i>	Aplicar	Empreendedorismo
		Fundamentos de Administração
		Metodologia Científica
		Fundamentos de Economia
		Gerenciamento de Projetos
		Propriedade Intelectual
C.2.3. <i>Compreender os principais temas relacionados e relevantes para a Ciência de Dados</i>	Aplicar	Aprendizado de máquina
		Probabilidade e Estatística
		Engenharia e Exploração de Dados
		Pesquisa Operacional, Otimização e Meta-heurísticas
C.2.4. <i>Aplicar fundamentos e tecnologias para o projeto de soluções eficientes de Ciência de Dados</i>	Aplicar	Sistemas Paralelos e Distribuídos
		Computação de Alto Desempenho
		Segurança de Sistemas Computacionais
		Arquitetura e Organização de Computadores

3. EIXO DE FORMAÇÃO: DESENVOLVIMENTO DE SISTEMAS

O desenvolvimento de sistemas computacionais inclui tanto a criação de sistemas quanto a adaptação de sistemas existentes, em especial aqueles que incluem análise de dados. Tal desenvolvimento deve contemplar o levantamento de requisitos funcionais e não-funcionais, a sua análise, modelagem, projeto, implementação e teste. Em todo o processo de desenvolvimento dos sistemas computacionais deve-se empregar teorias, métodos, técnicas e ferramentas para garantia e controle de qualidade do processo e do produto e identificar potenciais vulnerabilidades. Este eixo é definido pelas DCN21 como: *"especificar, projetar, implementar, manter e avaliar sistemas de computação, empregando teorias, práticas e ferramentas adequadas"* (CE-V).

COMPETÊNCIA: *Desenvolver sistemas computacionais que atendam qualidade de processo e de produto, considerando princípios e boas práticas de engenharia de sistemas e engenharia de software, incluindo:*

- a identificação, análise, especificação e validação de requisitos;
- o projeto de soluções computacionais em harmonia com o ambiente social e físico no seu entorno de aplicação;
- a implementação de sistemas computacionais utilizando ambientes de desenvolvimento apropriados;
- o teste e manutenção de sistemas computacionais;

<ul style="list-style-type: none"> a identificação e análise de potenciais vulnerabilidades. 		
Competências derivadas	Classificação	Conteúdos
C.3.1. <i>Resolver problemas usando ambientes de programação</i>	Aplicar	Algoritmos
		Programação
		Banco de Dados
		Interação Humano-Computador
		Processamento Paralelo e Distribuído
C.3.2. <i>Tomar decisões e inovar, com base no conhecimento do funcionamento e das características técnicas de hardware e da infraestrutura de software dos sistemas de computação, consciente dos aspectos éticos, legais e dos impactos ambientais decorrentes</i>	Aplicar	Arquitetura e Organização de Computadores
		Banco de Dados
		Segurança de Sistemas Computacionais
		Computação em Nuvem
		Legislação de Computação, Inteligência Artificial e de Proteção de Dados
		Engenharia de Software
		Sustentabilidade
		Ciência de Dados, Ética e Sociedade
C.3.3. <i>Avaliar projetos de sistemas de computação</i>	Avaliar	Engenharia de Software
		Métodos Quantitativos em Computação
		Avaliação de Desempenho de Sistemas
		Confiabilidade e Segurança de Sistemas

		Sistemas Paralelos e Distribuídos
C.3.4. <i>Realizar trabalho cooperativo e seus benefícios</i>	Aplicar	Gerenciamento de Projetos
		Comunicação Profissional
		Ciência de Dados, Ética e Sociedade
		Comportamento Humano nas Organizações
C.3.5. <i>Compreender os fatos essenciais, os conceitos, os princípios e as teorias relacionadas à Ciência da Computação para o desenvolvimento de software e hardware e suas aplicações</i>	Aplicar	Algoritmos
		Matemática Discreta
		Arquitetura e Organização de Computadores
		Teoria da Computação
		Teoria dos Grafos
C.3.6. <i>Identificar e analisar requisitos e especificações para problemas específicos e planejar estratégias para suas soluções</i>	Aplicar	Engenharia de Software
C.3.7. <i>Empregar metodologias que visem garantir critérios de qualidade ao longo de todas as etapas de desenvolvimento de uma solução computacional</i>	Aplicar	Engenharia de Software
C.3.8. <i>Analisar quanto um sistema baseado em computadores atende os critérios definidos para seu uso corrente e futuro (adequabilidade)</i>	Avaliar	Engenharia de Software
C.3.9. <i>Aplicar os princípios de gerência, organização e recuperação da informação, incluindo texto imagem som e vídeo</i>	Aplicar	Estruturas de Dados
		Banco de Dados
		Busca e Recuperação da Informação
		Inteligência Artificial
		Sistemas Multimídia
		Inteligência Computacional
		Processamento de Imagens
C.3.10. <i>Aplicar os princípios de interação humano-computador para avaliar e construir uma grande variedade de soluções incluindo interface do usuário, modelos cognitivos, sistemas de apoio a visualização, escalas, cores, tipos de gráficos, mapas, páginas WEB, sistemas multimídia e sistemas móveis</i>	Criar	Interação Humano-Computador
		Visualização de dados
		Sistemas Multimídia

4. EIXO DE FORMAÇÃO: ENGENHARIA E EXPLORAÇÃO DE DADOS

Na formação de um cientista de dados, a compreensão de como os dados podem ser produzidos, armazenados e gerenciados é de fundamental importância. Nesse sentido, o estudo de Sistemas Gerenciadores de Banco de Dados proverá ao egresso do curso de Ciência de Dados a fundamentação teórica e prática de como grandes volumes de dados são manipulados. Para isto, o entendimento do funcionamento de diversos sistemas de bancos de dados dará solidez à formação de cientistas de dados. A governança de dados deve embasar o egresso na habilidade de trabalhar aspectos de qualidade, privacidade e curadoria dos dados. Nessa linha de estudo, a proveniência de dados e LGPD devem estar presentes.

COMPETÊNCIA: Desenvolver e implementar projetos conceitual, lógico e físico de banco de dados, independente do modelo de dados, garantindo-se o nível de privacidade de dados esperado, bem como identificar gargalos e soluções para melhorar o acesso aos bancos de dados, quanto ao tempo de resposta, variedade de fontes e formatos, e aumento da vazão do sistema quanto ao volume de dados providos por unidade de tempo. Torna-se necessário, então, o desenvolvimento das seguintes competências específicas:

- a aplicação de conceitos, métodos e ferramentas de modelagem conceitual e física de bancos de dados, visualização e privacidade de dados, com o objetivo de garantir o cumprimento dos requisitos especificados quanto à qualidade, desempenho e nível de privacidade dos dados;
- a interação com pessoas de diferentes perfis, de diversas áreas do conhecimento, incluindo especialistas do domínio da aplicação, estatístico e, engenheiros de dados;
- a realização de ações de prospecção na busca de soluções mais eficazes, incluindo novas tecnologias, produtos e serviços;
- a adequação rápida às mudanças tecnológicas e aos novos ambientes de trabalho.

Competências derivadas	Classificação	Conteúdos
C.4.1. Compreender os requisitos de dados e construir modelos conceituais de dados.	Aplicar	Banco de Dados
		Sistemas de Gerência de Bancos de Dados
		Integração de Dados
C.4.2. Avaliar o desempenho de sistemas de dados	Aplicar	Sintonia de Desempenho de Bancos de Dados
		Sistemas de Dados
		Estrutura de Dados
		Sistemas Paralelos e Distribuídos
C.4.3. Aplicar técnicas de visualização de dados	Aplicar	Visualização de dados
		Privacidade de dados
C.4.4. Conhecer técnicas de privacidade de dado de acordo com a LGPD	Aplicar	Privacidade de dados
C.4.5. identificar soluções tecnológicas adequadas para o domínio da aplicação	Avaliar	Banco de Dados
		Sintonia de Desempenho de Bancos de Dados

		Integração de Dados
		Visualização de Dados
		Privacidade de dados
		Segurança de Sistemas Computacionais
		Sistemas Paralelos e Distribuídos

5. EIXO DE FORMAÇÃO: Dados em Larga Escala

Dados em Larga Escala se refere ao conjunto de métodos, técnicas e algoritmos capazes de coletar, manipular, gerir, consultar e analisar conjuntos de dados que se apresentam grande escala individualizada ou combinada de volume, variedade (apresentando diversidade e/ou multidimensionalidade), velocidade (vazão na produção), valor (capaz trazer valor ao negócio) e veracidade (de ser verdadeiro/fidedigno ou falso/fabricado). O estabelecimento de soluções capazes de gerenciar Dados em Larga Escala o objeto deste eixo, seja pelo desenvolvimento ou pela implantação de sistemas computacionais (desenvolvidos ou adquiridos) para este fim em algum ambiente computacional (centralizado, paralelo e distribuído, nuvem, dentre outros) Este desenvolvimento pode envolver a integração de sistemas computacionais, a adequação de infraestrutura, a garantia das regras de negócio (requisitos funcionais), das regras sistêmicas (desempenho, contingência, confiabilidade, segurança), e de políticas internas e externas legais.

COMPETÊNCIA: *Implantar soluções de Dados em larga escala, considerando:* **planejar** e **executar** o processo de **implantação** de sistemas computacionais baseados no paradigma de programação e gerência de dados em larga escala; **identificar** os problemas que demandem soluções algorítmicas escaláveis; **selecionar** ou **criar** algoritmos escaláveis que abordem armazenamento de dados, computação de alto desempenho e teoria da complexidade; **prover capacitação** das pessoas envolvidas (técnicos e usuários), considerando a documentação e a operacionalização do sistema computacional; **garantir a consistência** da implementação com as normas legais e éticas da comunidade envolvida.

Competências derivadas	Classificação	Conteúdos
C.5.1. <i>Tomar decisões e inovar, com base no conhecimento do funcionamento e das características técnicas de hardware e da infraestrutura de software dos sistemas de computação consciente dos aspectos éticos, legais e dos impactos ambientais decorrentes (CG-IV)</i>	Aplicar	Arquitetura e Organização de Computadores
		Redes de Computadores
		Sistemas Operacionais
		Segurança de Sistemas Computacionais
		Banco de Dados
		Processamento Paralelo e Distribuído
		Privacidade de Dados
		Legislação de Computação, inteligência artificial e de proteção de dados

		Sustentabilidade
		Ética e Sociedade para Ciência de Dados
C.5.2. <i>Compreender e explicar as dimensões quantitativas de um problema (CG-V)</i>	Aplicar	Estatística Básica
		Métodos Quantitativos em Computação
C.5.3. <i>Conceber e preparar soluções para gerência de dados em larga escala (CG-VI)</i>	Aplicar	Arquitetura e Organização de Computadores
		Processamento Paralelo e Distribuído
		Computação em Nuvem
C.5.4. <i>Planejar soluções de gerência e armazenamento de dados distribuídos (CG-VI)</i>	Aplicar	Banco de Dados
		Sistemas de Dados
		Integração de Dados
C.5.5. <i>Conceber soluções de gerência de dados em larga escala a partir de decisões visando o equilíbrio de todos os fatores (CE-VI)</i>	Criar	Engenharia de Software
		Simulação de Sistemas
		Privacidade de Dados
		Sustentabilidade
		Empreendedorismo
C.5.6. <i>Empregar metodologias que visem garantir critérios de qualidade para o gerenciamento de dados em larga escala ao longo de todas as etapas de desenvolvimento de uma solução computacional (CE-VII)</i>	Aplicar	Engenharia de Software
		Gerência de Projetos
		Teste, Verificação e Validação de Software

6. EIXO DE FORMAÇÃO: MINERAÇÃO DE DADOS E APRENDIZADO DE MÁQUINA

Um dos aspectos centrais na formação de bons cientistas de dados é saber como desenvolver e empregar técnicas para extrair conhecimento útil de um conjunto de dados e como criar modelos que capturam as principais características presentes nos dados, sejam os modelos descritivos, preditivos ou prescritivos. Para isso, deverá conhecer e selecionar as diferentes técnicas e algoritmos utilizados nos processos de mineração de dados e de

aprendizado de máquina para um conjunto de dados.

COMPETÊNCIA: Analisar, compreender, aplicar e desenvolver técnicas e algoritmos para mineração de dados e aprendizado de máquina.

Competências derivadas	Classificação	Conteúdos
C.6.1. <i>Avaliar e tratar a qualidade dos dados</i>	Aplicar	Estatística
		Banco de Dados
		Visualização de Dados
		Pré-processamento de Dados
C.6.2. <i>Compreender as dimensões quantitativas de um problema</i>	Aplicar	Estatística
		Métodos Quantitativos em Computação
		Visualização de Dados
C.6.3. <i>Planejar experimentos</i>	Projetar	Estatística
		Fundamentos de Programação
		Visualização de Dados
C.6.4. <i>Avaliar necessidade e aplicar técnicas de pré-processamento de dados</i>	Aplicar	Estatística
		Fundamentos de Programação
		Introdução a Ciência de Dados
		Visualização de Dados
		Mineração de Dados
		Aprendizado de Máquina
		Álgebra Linear e Matricial
C.6.5. <i>Identificar e analisar requisitos e especificações para problemas específicos e planejar estratégias para suas soluções</i>	Criar	Introdução à Ciência de Dados
		Projeto de Sistemas Computacionais
C.6.6. <i>Especificar, projetar, implementar, manter e avaliar sistemas de computação, empregando teorias, práticas e ferramentas adequadas (CE-V)</i>	Criar	Projeto de Sistemas Computacionais
		Aprendizado de Máquina
		Avaliação de Desempenho de Sistemas

		Banco de Dados
		Processamento Paralelo e Distribuído
		Computação em Nuvem
		Redes Neurais
		Internet das Coisas (IoT)
C.6.7. Conceber soluções eticamente e socialmente responsáveis	Aplicar	Sustentabilidade
		Ciência de Dados e Sociedade
		Ética em Ciência de Dados
C.6.8. Aplicar os princípios de gerência, organização e recuperação da informação de vários tipos, incluindo texto, imagem, som, e vídeo (CE-XII)	Aplicar	Banco de Dados
		Recuperação da Informação
		Sistemas Multimídia
		Processamento da Língua Natural
		Processamento de Imagens
		Processamento Paralelo e Distribuído

182

7. EIXO DE FORMAÇÃO: APRENDIZADO CONTÍNUO E AUTÔNOMO

A teoria e prática da Ciência de Dados estão em permanente evolução, levando: (1) ao surgimento de novos instrumentos (processos, métodos e ferramentas) que visam à melhoria do gerenciamento e da qualidade dos dados, além das informações e conhecimentos extraídos deles; (2) ao aperfeiçoamento de instrumentos existentes; (3) ao surgimento de novas tecnologias de infraestrutura computacional; e, (4) à expansão dos domínios de aplicação da ciência de dados.

Este eixo de formação em especial agrupa competências orientadas ao desenvolvimento pessoal (habilidades e atitudes), em vez da assimilação de conteúdos tradicionais. Por exemplo, é possível apresentar o conteúdo "auto-regulação da aprendizagem" em forma de palestra extra-curricular ou de aula em alguma disciplina da matriz curricular. Porém, essa habilidade será plenamente desenvolvida se for estimulada transversalmente à apresentação de conteúdos técnicos. Os docentes devem ser estimulados a adotarem abordagens pedagógicas para promover as competências deste eixo de formação, tais como (mas não exclusivamente), aprendizagem colaborativa, aprendizagem baseada em projetos (*project-based learning* - PBL), aprendizagem ativa (*active learning*), ensino híbrido (*blended*), entre outras.

COMPETÊNCIA: *Aprender contínua e autonomamente sobre métodos, instrumentos, tecnologias de ciência de dados e de seus domínios de aplicação, abrangendo: Desenvolver estudos para manter-se atualizado sobre a evolução da ciência de dados, além de desenvolver pesquisas que contribuam para essa*

<p>evolução. Essa atividade de estudo inclui o acompanhamento de publicações científicas e comerciais em periódicos e eventos especializados, bem como a participação em cursos de treinamento, extensão, especialização, mestrado e doutorado; Avaliar novos instrumentos e métodos, bem como a viabilidade de seu emprego no seu contexto de trabalho e, eventualmente, propor novos instrumentos e métodos a partir da sua própria experiência; Avaliar novas tecnologias computacionais e a sua adequação à ciência de dados; Adequar-se rapidamente às mudanças tecnológicas e aos novos ambientes de trabalho (CG-IX).</p>		
Competências derivadas	Classificação	Conteúdos
<p>C.7.1. <i>Gerir a sua própria aprendizagem e desenvolvimento, incluindo a gestão de tempo e competências organizacionais (CG-VI)</i></p>	<p>Aplicar</p>	Gestão de Tempo
		Gestão de Carreira Profissional
		Autorregulação da Aprendizagem
<p>C.7.2. <i>Realizar trabalho cooperativo e entender os benefícios que este pode produzir (CG-XII)</i></p>	<p>Aplicar</p>	Aplicável a todos os conteúdos, utilizando práticas pedagógicas colaborativas
<p>C.7.3. <i>Compreender os fatos essenciais, os conceitos, os princípios e as teorias relacionadas à Ciência de Dados e suas aplicações (CE-I)</i></p>	<p>Avaliar</p>	Algoritmos
		Aprendizado de Máquina
		Teoria da Computação
		Mineração de Dados
<p>C.7.4. <i>Analisar quanto um sistema baseado em computadores atende os critérios definidos para seu uso corrente e futuro (adequabilidade) (CE-VIII)</i></p>	<p>Avaliar</p>	Métodos Quantitativos em Computação
		Modelagem de Sistemas
		Banco de Dados
		Simulação de Sistemas

183

8. EIXO DE FORMAÇÃO: CIÊNCIA, TECNOLOGIA E INOVAÇÃO

Um curso de Ciência de Dados deve fornecer aos seus egressos uma base teórica sólida que os permita desenvolver estudos avançados e prepará-los para grandes desafios.

A inovação em ciência de dados exige conhecimentos científicos e tecnológicos que vão além dos necessários para suas aplicações tradicionais. Além disso, a formação do egresso deve levar em conta a cultura das pessoas envolvidas, as oportunidades do mercado e as necessidades da sociedade.

COMPETÊNCIA: *Desenvolver estudos avançados visando o desenvolvimento científico e tecnológico da ciência de dados e a criação de soluções computacionais inovadoras para problemas em qualquer domínio de conhecimento, abrangendo: **Compreender fundamentos científicos e tecnológicos** da computação, em profundidade, e associá-los à ciência de dados; **Dominar ferramentas matemáticas e estatísticas** necessárias para a pesquisa e desenvolvimento de produtos e processos em ciência de dados; **Adaptar-se a novos domínios de aplicação**, que envolvam contextos particulares e novas tecnologias; **Realizar ações inovadoras** na busca de soluções mais eficazes, incluindo novos produtos e processos (CG-IX); **Adequar-se rapidamente** às mudanças*

tecnológicas e aos novos ambientes de trabalho.		
Competências derivadas	Classificação	Conteúdos
C.8.1. <i>Identificar problemas que tenham solução algorítmica (CG-I)</i>	Aplicar	Algoritmos
		Teoria da Computação
		Complexidade de Algoritmos
		Estatística
C.8.2. <i>Tomar decisões e inovar, com base no conhecimento do funcionamento e das características técnicas de Ciência de Dados, consciente dos aspectos éticos, legais e dos impactos ambientais decorrentes (CG-IV)</i>	Aplicar	Privacidade de Dados
		Legislação de Computação, inteligência artificial e de proteção de dados
		Sustentabilidade
		Ciência de Dados, Ética e Sociedade
C.8.3. <i>Conceber soluções computacionais capazes de trazer inovação a um produto ou processo (CG-V)</i>	Aplicar	Empreendedorismo
		Interação Humano-Computador
		Metodologia Científica
		Gerência de Projeto
C.8.4. <i>Identificar e analisar requisitos e especificações para problemas específicos e planejar estratégias para suas soluções de Ciência de Dados (CE-IV)</i>	Aplicar	Engenharia de Software
		Teste, Verificação e Validação de Software

184 II.7. Relação com as Diretrizes Curriculares Nacionais

185 Esta seção apresenta em quais eixos de formação estão mapeadas cada uma das competências e
 186 habilidades definidas pelas DCN21. Dessa maneira, a tabela a seguir explicita resumidamente a relação
 187 entre as DCN21 e os RF-CC-17.

188 II.8. Estágios, TCC e atividades complementares

189 De acordo com as DCN21, cabe às Instituições de Educação Superior estabelecerem a
 190 obrigatoriedade ou não do Estágio Supervisionado ou do Trabalho de Curso e a definição dos respectivos
 191 regulamentos.

192 II.8.1. Estágio Supervisionado

193 O estágio supervisionado é um ato escolar educativo desenvolvido no ambiente de trabalho. A atividade

194 não cria vínculo empregatício e seu objetivo é preparar o estudante para o mercado de trabalho, por meio
195 do desenvolvimento de competências inerentes à atividade profissional e da contextualização curricular. O
196 estágio, bem como seu regulamento, deve fazer parte do projeto pedagógico de curso.

197 O estágio supervisionado pode ser obrigatório ou não-obrigatório. No primeiro caso, a sua carga horária é
198 requisito para aprovação no curso e obtenção do diploma, enquanto que, no segundo caso, trata-se de
199 atividade opcional com carga horária acrescida à carga horária regular e obrigatória. Compete às IES, por
200 meio dos Núcleos Docentes Estruturantes (NDE), determinar a obrigatoriedade ou não do estágio
201 supervisionado para os cursos de bacharelado em Ciência de Dados.

202 De acordo com as DCN16, os cursos podem envolver a formação para a realização de: (i) atividades
203 orientadas à realização de processos; e (ii) atividades orientadas à transformação de processos, com
204 desenvolvimento de novas tecnologias. Nos cursos orientados à realização, ou reprodução, de processos,
205 há forte recomendação de estágio para os alunos. Por meio de estágio, os estudantes podem conhecer
206 previamente o ambiente onde são desenvolvidas as atividades de trabalho para as quais eles estão sendo
207 preparados, como forma de iniciação à profissionalização.

208 Como o estágio se trata de uma atividade em que o estudante replica conhecimentos para atividades de
209 realização de processos, é recomendável que o NDE do curso defina no regulamento de estágio o momento,
210 no que concerne períodos, ou uma formação mínima, em se tratando de conclusão de componentes
211 curriculares, para que o estudante possa iniciar o estágio supervisionado.

212 **II.8.2. Trabalho de Conclusão de Curso**

213 O Trabalho de Conclusão de Curso (TCC) é uma atividade acadêmica curricular cujo objetivo é desenvolver
214 e verificar as habilidades cognitivas de compreensão, aplicação, análise, avaliação e criação acerca dos
215 conhecimentos científicos, técnicos e culturais produzidos ao longo do curso.

216 Compete às IES, por meio dos Núcleos Docentes Estruturantes (NDE), determinar a obrigatoriedade ou
217 não do TCC para os cursos de bacharelado em Ciência da Dados.

218 De acordo com as DCN16, os cursos podem envolver a formação para a realização de: (i) atividades
219 orientadas à realização de processos; e (ii) atividades orientadas à transformação de processos, com
220 desenvolvimento de novas tecnologias. Nos cursos orientados à transformação de processos, ou criação de
221 novas tecnologias, há forte recomendação de realização de TCC para os alunos.

222 Por meio de TCC, os estudantes devem aplicar conhecimentos de vanguarda na produção de aplicações
223 científicas, tecnológicas ou de inovações.

224 Compete às IES regulamentar as normas e os procedimentos no seu regimento e dar conhecimento ao aluno
225 acerca do TCC. O regulamento deve especificar os critérios, procedimentos e mecanismos de avaliação,
226 assim como as diretrizes e técnicas relacionadas à sua elaboração. Deverá constar no regulamento quando
227 o estudante poderá iniciar o TCC.

II.8.3. Atividades complementares

De acordo com as DCN16, as Atividades Complementares são componentes curriculares enriquecedores e implementadores do próprio perfil do formando e deverão possibilitar o desenvolvimento de habilidades, conhecimentos, competências e atitudes do aluno, inclusive as adquiridas fora do ambiente acadêmico, que serão reconhecidas mediante processo de avaliação.

As Atividades Complementares podem incluir atividades desenvolvidas na própria instituição ou em outras instituições e variados ambientes sociais, técnico-científicos ou profissionais de formação profissional, incluindo:

- experiências de trabalho;
- estágios não obrigatório;
- extensão universitária
- iniciação científica;
- participação em eventos técnico-científicos;
- publicações científicas;
- programas de monitoria e tutoria;
- disciplinas de outras áreas;
- representação discente em comissões e comitês;
- participação em empresas juniores;
- incubadoras de empresas;
- atividades de empreendedorismo e inovação.

II.9. Metodologia de ensino

De acordo com as DCN16, a metodologia de ensino deve ser centrada no aluno como sujeito da aprendizagem e apoiada no professor como facilitador do processo de ensino-aprendizagem. O trabalho extraclasse deve ser empregado de forma que o aluno aprenda a resolver problemas e seja estimulado a aprender a aprender, tornando-se independente e criativo. O professor deve apresentar as aplicações dos conteúdos teóricos, ser um mediador, estimular a competição, a comunicação, provocar a realização de trabalho em equipe, motivar os alunos para os estudos e orientar o raciocínio e desenvolver as capacidades de comunicação e de negociação. Quando aplicável, deve-se empregar metodologias ativas, de forma que o aluno passe mais tempo em atividades nas quais seja protagonista no processo de ensino e aprendizagem.

Considerando os cenários atuais da educação, e uma nova sociedade tecnológica, o projeto pedagógico deve prever o emprego de metodologias de ensino e aprendizagem que promovam a explicitação das relações entre os conteúdos abordados e as competências previstas para o egresso do cursos, além da inserção de novos paradigmas educacionais que aborçam novas prática curriculares e de metodologias inovadoras. Novas mídias e novas tecnologias da informação e da comunicação são inseridas no meio acadêmico levando a uma reflexão sobre as práticas pedagógicas utilizadas e a efetividade das mesmas no processo de ensino e aprendizagem.

As metodologias de ensino e paradigmas atualmente empregadas são constantemente abordadas nos eventos do CSBC (Curso de Qualidade, WEI, SECOMU, Desafie!, JAI) e eventos temáticos organizados pelas Comissões Especiais.

II.10. Requisitos legais

Os RF-CC-17 foram elaborados em consonância com as Diretrizes Nacionais Curriculares para os cursos de graduação na área de Computação, aprovada no dia 09 de março de 2012 e homologada oficialmente no DOU do dia 28 de outubro de 2016. Espera-se que os RF-CC-17 auxiliem as instituições e coordenadores

de curso na elaboração e revisão de seus Projetos Pedagógicos de Curso (PPCs).

Os RF-CC-17 não são um PPC para os cursos de Bacharelado em Ciência de Dados. Na elaboração do PPC de seu curso, as instituições, coordenadores e Núcleos Docentes Estruturantes (NDEs) podem usar este documento como guia para decidir quanto a eixos, competências e conteúdos de formação para seus cursos, mas precisam considerar diversas questões, tais como: o contexto regional onde atua, suas estratégias de formação, sua capacidade de formação, a estratégia institucional para formação de egressos, entre vários outros aspectos.

Deve-se observar também que outros requisitos legais e normativos são exigidos pelas instituições para elaboração de seus Projetos Pedagógicos de Curso (PPCs). Tais requisitos legais e normativos devem ser cuidadosamente analisados e interpretados pelos NDEs antes de se atualizar ou propor um novo PPC. É necessário destacar também a necessidade de adequar o PPC para a modalidade do curso, haja vista que os cursos à distância exigem o cumprimento de requisitos legais e normativos específicos para esse tipo de modalidade, além daqueles exigidos para os cursos presenciais.

O Ministério da Educação é responsável por organizar o sistema de ensino no país e definir as políticas e diretrizes educacionais nacionais. Isso envolve dar publicidade e zelar pelo cumprimento das leis, decretos, portarias, pareceres, resoluções, e demais instrumentos legais e normativos federais para o ensino, inclusive superior. Em consonância com a esfera federal, as Secretarias de Educação, estaduais, municipais e do Distrito Federal são responsáveis por organizar o sistema de ensino e definir as políticas e diretrizes educacionais; porém, dentro dos seus respectivos domínios: estadual, municipal e distrital.

Normalmente, as Instituições de Ensino Superior (IES) elaboram os seus próprios regulamentos e regimentos internos, bem como Planos de Desenvolvimento Institucional e outros instrumentos. Assim, é relevante também observar as questões internas das IES; porém, respeitando a hierarquia legal.

É de grande importância o acompanhamento da evolução legal e normativa pertinentes à educação superior, no país, estados, municípios e na própria IES, haja vista que leis, decretos e outros instrumentos estão sempre sujeitos a mudanças.

II.11. Agradecimentos

Referências

ACM/IEEE (2013). Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science. Final Report. ACM, New York, NY, USA. 2013. Disponível em: <http://dx.doi.org/10.1145/2534860>. Último acesso em: 27/03/2017.

Ferraz, A. P. C. M., Belhot, R. V. (2010). Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. Gest. Prod., São Carlos, 17(2), 421-431.

Le Boterf, G. (2006) Avaliar a competência de um profissional: três dimensões a explorar. Reflexão RH, 1.1: 60-3.

Sacristán, J. G.; Gómez, Á. I. P.; Rodríguez, J. B. M.; Santomé, J. T.; Rasco, F. A.; Méndez, J. M. Á. (2016). Educar por competências: O que há de novo?. Artmed Editora.

SBC (1991). Currículo de Referência da SBC para cursos de Graduação Plena em Computação (CR91). <http://homepages.dcc.ufmg.br/~bigonha/Cr/cr91.html>. Último acesso em: 27/03/2017.

SBC (1996). Currículo de Referência da SBC para cursos de Graduação Plena em Computação (CR96). <http://homepages.dcc.ufmg.br/~bigonha/Cr/cr.html>. Último acesso em: 27/03/2017.

- 318 SBC (1999). Currículo de Referência da SBC para cursos de Graduação em Computação (CR99).
319 <http://lad.dsc.ufcg.edu.br/ec/cr99.pdf>. Último acesso em: 27/03/2017.
- 320 SBC (2003). Currículo de Referência da SBC para cursos de Graduação em Computação e Informática
321 (CR03). <https://pt.slideshare.net/derciolr/curriculo-de-referencia-sbc>. Último acesso em: 27/03/2017.
- 322 SBC (2005). Currículo de Referência da SBC para cursos de Graduação em Bacharelado em Ciência da
323 Computação e Engenharia da Computação (CR05).
324 <https://drive.google.com/drive/folders/0B2FfrKdMF9EuOVFGWUJES3M2QUE>. Último acesso em:
325 27/03/2017.
- 326 Scallon, G. (2015) Avaliação da aprendizagem numa abordagem por competências. Curitiba: PUCPress.
- 327 MEC (1999). Diretrizes Curriculares Nacionais de cursos da área de Computação e Informática (DCN99).
328 Disponível em: <http://www.inf.ufrgs.br/ecp/docs/diretriz.pdf>. Último acesso em: 27/03/2017.
- 329 MEC (2012). Proposta de Diretrizes Curriculares Nacionais para os cursos de graduação em Computação.
330 Disponível em: [http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=11205-pces136-11-pdf&category_slug=julho-2012-](http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=11205-pces136-11-pdf&category_slug=julho-2012-pdf&Itemid=30192)
331 [pdf&Itemid=30192](http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=11205-pces136-11-pdf&category_slug=julho-2012-pdf&Itemid=30192). Parecer CNE/CES nº 136/2012, aprovado em 8 de março de 2012. Último acesso
332 em: 27/03/2017.
- 334 MEC (2016). Diretrizes Curriculares Nacionais para os Cursos de Graduação em Computação (DCN16).
335 Disponível em: [http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=52101-rces005-16-pdf&category_slug=novembro-2016-](http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=52101-rces005-16-pdf&category_slug=novembro-2016-pdf&Itemid=30192)
336 [pdf&Itemid=30192](http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=52101-rces005-16-pdf&category_slug=novembro-2016-pdf&Itemid=30192). Resolução CNE/CES nº 5, de 16 de novembro de 2016. Último acesso:
337 27/03/2017.
- 339 MEC (2021). Diretrizes Curriculares Nacionais para os Cursos de Graduação em Computação
340 (DCN21). Em elaboração.