

2020

Informační technologie – Data velkého objemu – Přehled a slovník

ČSN
ISO/IEC 20546

36 9008

Information technology – Big data – Overview and vocabulary

Technologies de l'information – Mégadonnées – Vue d'ensemble et vocabulaire

Tato norma je českou verzí mezinárodní normy ISO/IEC 20546:2019. Překlad byl zajištěn Českou agenturou pro standardizaci. Má stejný status jako oficiální verze.

This standard is the Czech version of the International Standard ISO/IEC 20546:2019. It was translated by the Czech Standardization Agency. It has the same status as the official version.

Národní předmluva

Souvisící ČSN

ČSN ISO/IEC 17788:2017 (36 9865) Informační technologie – Cloud computing – Přehled a slovník

Upozornění na národní poznámky

Do kapitoly 3 byly k termínům 3.1.26 a 3.1.31 doplněny národní poznámky upozorňující na chybný odkaz.

Vysvětlivky k textu převzaté normy

Pro účely této normy byly použity následující anglické termíny v původní podobě, vzhledem k rozšíření těchto termínů v odborné komunitě a/nebo absenci českého ekvivalentu:

cloud computing, master-slave, MapReduce

Pro účely této normy byl termín „garbage-in, garbage-out“ použit v původní podobě, vzhledem ke specifickému významu tohoto spojení v IT technologiích (GIGO), které vyjadřuje způsob zacházení s daty v paměti počítače.

Vypracování normy

Zpracovatel: Ing. Alena Hönigová, IČO 61470716

Technická normalizační komise: TNK 42 Výměna dat

Pracovník České agentury pro standardizaci: Ing. Miroslav Škop

Česká agentura pro standardizaci je státní příspěvková organizace zřízená Úřadem pro technickou normalizaci, metrologii a státní zkušebnictví na základě ustanovení § 5 odst. 2 zákona č. 22/1997 Sb., o technických požadavcích na výrobky a o změně a doplnění některých zákonů, ve znění pozdějších předpisů.

ICS 01.040.35; 35.020

Obsah

Strana

Předmluva.....	5
Úvod.....	6
1..... Předmět normy.....	7
2..... Citované dokumenty.....	7
3..... Termíny, definice a zkrácené termíny.....	7
3.1..... Termíny a definice.....	7
3.2..... Zkratky.....	11
4..... Hlavní charakteristiky dat velkého objemu.....	11
4.1..... Obecně.....	11
4.2..... Hlavní charakteristiky dat.....	11
4.2.1... Objem dat.....	11

4.2.2... Rychlost	
dat.....	
.....	11
4.2.3... Rozmanitost	
dat.....	
.....	12
4.2.4... Variabilita	
dat.....	
.....	12
4.3..... Hlavní charakteristiky zpracování	
dat.....	12
4.3.1... Nauka	
o datech.....	
.....	12
4.3.2... Nestálost	
dat.....	
.....	12
4.3.3... Věrohodnost	
dat.....	
.....	12
4.3.4...	
Přínos.....	
.....	12
4.3.5... Vizualizace	
dat.....	
.....	12
4.3.6... Strukturovaná a nestrukturovaná	
data.....	12
4.3.7...	
Škálování.....	
.....	12
4.3.8... Distribuovaný systém	
souborů.....	
13	
4.3.9... Distribuované zpracování	
dat.....	13
4.3.10 Nerelační	
databáze.....	
.....	13

Příloha A (informativní) Křížové koncepty dat velkého objemu.....	14
---	----

Bibliografie.....	16
-------------------	----



DOKUMENT CHRÁNĚNÝ COPYRIGHTEM

© ISO/IEC 2019

Veškerá práva vyhrazena. Žádná část této publikace nesmí být, není-li specifikováno jinak nebo nepožaduje-li se to v souvislosti s její implementací, reprodukována nebo používána v jakékoliv formě nebo jakýmkoliv způsobem, elektronickým ani mechanickým, včetně pořizování fotokopii nebo zveřejňování na internetu nebo intranetu, bez předchozího písemného souhlasu. O souhlas lze požádat buď ISO na níže uvedené adrese, nebo členskou organizaci ISO v zemi žadatele.

ISO copyright office

CP 401 • Ch. de Blandonnet 8

CH-1214 Vernier, Geneva

Phone: +41 22 749 01 11

Fax: + 41 22 749 09 47

Email: copyright@iso.org

Website: www.iso.org

Publikováno ve Švýcarsku

Předmluva

ISO (Mezinárodní organizace pro normalizaci) a IEC (Mezinárodní elektrotechnická komise) tvoří specializovaný systém celosvětové normalizace. Národní orgány, které jsou členy ISO nebo IEC, se podílejí na vypracování mezi-národních norem prostřednictvím svých technických komisí ustavených příslušnými organizacemi pro jednotlivé obory technické činnosti. Technické komise ISO a IEC spolupracují v oborech společného zájmu. Práce se zúčastňují také další vládní i nevládní mezinárodní organizace, s nimiž ISO a IEC navázaly pracovní styk.

Postupy použité při tvorbě tohoto dokumentu a postupy určené pro jeho další udržování jsou popsány ve směrnících ISO/IEC, část 1. Zejména se má věnovat pozornost rozdílným schvalovacím kritériím potřebným pro různé druhy dokumentů. Tento dokument byl vypracován v souladu s redakčními pravidly uvedenými ve směrnících ISO/IEC, část 2 (viz www.iso.org/directives).

Upozorňuje se na možnost, že některé prvky tohoto dokumentu mohou být předmětem patentových práv. ISO a IEC nelze činit odpovědnými za identifikaci jakéhokoliv nebo všech patentových práv. Podrobnosti o jakýchkoliv patentových právech identifikovaných během přípravy tohoto dokumentu budou uvedeny v úvodu a/nebo v seznamu patentových prohlášení obdržených ISO (viz www.iso.org/patents) nebo v seznamu patentových prohlášení obdržených IEC (viz <http://patents.iec.ch>).

Jakýkoliv obchodní název použitý v tomto dokumentu se uvádí jako informace pro usnadnění práce uživatelů a neznamena schválení.

Vysvětlení nezávazného charakteru norem, významu specifických termínů a výrazů ISO, které se vztahují k posuzování shody, jakož i informace o tom, jak ISO dodržuje principy WTO týkající se technických překážek obchodu (TBT), jsou uvedeny na www.iso.org/iso/foreword.html.

Tento dokument vypracovala společná technická komise ISO/IEC JTC 1 *Informační technologie*, subkomise SC 42 *Umělá inteligence*.

Jakákoliv zpětná vazba nebo otázky týkající se tohoto dokumentu by měly být směřovány na národní normalizační orgán uživatele. Úplný seznam těchto orgánů lze nalézt na www.iso.org/members.html.

Úvod

Paradigma dat velkého objemu představuje rychle se měnící oblast s rychle se měnícími technologiemi.

Termín data velkého objemu představuje datové soubory, velké, pokud jde o objem, rychlost, rozmanitost a/nebo variabilitu. Tento termín však nepředstavuje data, která jsou prostě větší, než byla předtím, protože se to dělo pravidelně po desetiletích. Konkrétní výskyt, který vedl k rozšířenému použití tohoto termínu, je v polovině 20. století, kdy rozsáhlé datové soubory nebylo možné zvládnout pomocí velkých datových systémů. Techniky dat velkého objemu představovaly v té době posun k používání řízení distribuovaných dat a zpracování s použitím horizontálního škálování, aby se dosáhlo požadované výkonnosti za dostupnou cenu.

Během rozvoje systémů zpracování dat se mnohokrát stalo, že si potřeba účinné, cenově efektivní analýzy dat vynutila změnu existujících technologií. K přesunu k relačnímu modelu například došlo, když metody ke spolehlivému zvládnutí změn strukturovaných dat vedly v 80. letech k přesunu k relačním databázím, které vytvořily model relační algebry. To byl základní posun v zacházení s daty. Revoluce v technologiích, o kterých se mluví jako o datech velkého objemu, nastala, protože relační model již dále nemohl efektivně řešit všechny potřeby týkající se analýzy velkých a často nestrukturovaných datových souborů. Není to tím, že data mají větší objem než dříve, protože data se plynule zvětšovala po desetiletí. Revoluce týkající se dat velkého objemu je místo toho jednorázový zásadní posun v architektuře směrem k paralelizaci, stejně jako byl jednorázovým posunem posun k relačnímu modelu. Stejně jako se relační databáze v průběhu dekad stávaly efektivnější, i technologie dat velkého objemu budou pokračovat ve vývoji. Mnohé podpůrné koncepty dat velkého objemu existovaly po celá léta, ale v letech od poloviny 20. století nastal prudký nárůst škálovacích technologií a jejich vyžívání a aplikace na systémy škálovaných dat.

Termín data velkého objemu je v běžném použití nadužíván a je používán k reprezentaci mnoha souvisejících konceptů, zejména protože několik odlišných dimenzí systémů je trvale ve vzájemné interakci. Abychom porozuměli této revoluci, je třeba vzít v úvahu souhrn následujících aspektů: data a charakteristiky zpracování datových souborů, analýzy datových souborů, výkonnost systémů, které zacházejí s daty, podnikatelské zvážení efektivnosti nákladů, a nové inženýrství a techniky analýz distribuovaného zpracování dat s použitím horizontálního škálování.

V příloze A je uveden přehled několika konceptů obecnější výpočetní oblasti, které jsou průřezové vzhledem k datům velkého objemu.

1 Předmět normy

Tento dokument poskytuje soubor termínů a definic, potřebných k podpoře zlepšené komunikace a porozumění této oblasti. Poskytuje terminologický základ norem týkajících se dat velkého objemu.

Tento dokument poskytuje koncepční přehled v oblasti dat velkého objemu, jejich vztahu ke snahám dalších technických oblastí a norem, a koncepty přisuzované datům velkého objemu, která nejsou pro data velkého objemu nová.

Konec náhledu - text dále pokračuje v placené verzi ČSN.