

CAPITOLUL 11 - BIG DATA

Introducere în
conceptele,
Instrumentele,
tehnologiile și
aplicațiile Big
Data

INTRODUCERE

- ▶ Big Data reprezintă seturi de date extrem de mari și complexe.
- ▶ Caracteristicile de bază: Volum, Viteză, Varietate.
- ▶ Necesitatea unor tehnologii avansate pentru procesare și analiză.

CARACTERISTICILE BIG DATA

- ▶ Volum: Cantitatea imensă de date generate.
- ▶ Viteză: Viteza rapidă de generare și procesare a datelor.
- ▶ Varietate: Diversitatea tipurilor și surselor de date.

TEHNOLOGII BIG DATA

- ▶ Hadoop: Framework open-source pentru stocare și procesare distribuită.
- ▶ Apache Spark: Motor de analiză a datelor în memorie.
- ▶ NoSQL: Baze de date pentru gestionarea datelor non-relaționale.

ARHITECTURA BIG DATA

- ▶ Arhitectura distribuită pentru scalabilitate și redundanță.
- ▶ Sistemul Hadoop Distributed File System (HDFS).
- ▶ Procesarea datelor cu MapReduce și Apache Spark.

PRE- PROCESAREA DATELOR

- ▶ Curățarea datelor:
Eliminarea erorilor și a
valorilor lipsă.
- ▶ Transformarea datelor:
Normalizarea și
codificarea variabilelor.
- ▶ Agregarea datelor:
Combinarea și
sintetizarea datelor.

ANALIZA DATELOR

- ▶ Învățare automată:
Algoritmi de predicție și clasificare.
- ▶ Analiza statistică: Teste de ipoteze și regresii.
- ▶ Analiza textuală:
Procesarea limbajului natural și clasificarea textului.

STOCAREA DATELOR

- ▶ Depozite de date distribuite: Scalabilitate și redundanță.
- ▶ Depozite de date în memoria principală: Viteză și timp real.
- ▶ Depozite de date bazate pe coloane: Eficiență în citirea selectivă.

SECURITATEA ȘI PROTECȚIA DATELOR

- ▶ Criptarea datelor:
Protecție împotriva
accesului neautorizat.
- ▶ Autentificare și autorizare:
Acces controlat la date.
- ▶ Monitorizare și raportare:
Detectarea activităților
suspecte.

APLICAȚII PRACTICE ALE BIG DATA

- ▶ Marketing: Personalizarea ofertelor și campaniilor publicitare.
- ▶ Sănătate: Prevenirea epidemiilor și îmbunătățirea tratamentelor.
- ▶ Finanțe: Identificarea fraudelor și optimizarea investițiilor.

PERSPECTIVE PRIVIND BIG DATA

- ▶ Analiza de date în cloud:
Procesarea fără
infrastructură hardware.
- ▶ Analiza de date în timp
real: Decizii rapide și
răspuns la schimbări.
- ▶ Creșterea utilizării IoT:
Explozie de date
generate de dispozitive
conectate.

REZUMAT

- ▶ Big Data este esențial pentru analiza și înțelegerea datelor complexe.
- ▶ Necesită tehnologii avansate pentru stocare, procesare și analiză.
- ▶ Aplicațiile Big Data sunt diverse, cu impact în multe domenii.

ÎNTREBĂRI DE DISCUȚIE

- ▶ Care sunt principalele caracteristici ale Big Data?
- ▶ Cum influențează Big Data deciziile în diferite industrii?
- ▶ Ce tehnologii sunt esențiale pentru gestionarea Big Data?