Faculty of Computer
and Information Sciences
Ain Shams University

كلية الحاسبات والمعلومات
جامعة عين شمس

Data Science

Code:

Instructor

Prof .Dr. Abeer M. Mahmoud

Professor of computer science-faculty of Computer and Information Sciences-
Ain Shams University
Abeer.mahmoud@cis.asu.edu.eg

Data Science and Big Data Analytics v2

Introduction to Big Data analytics

Lecture 2

Instructor

Dr. Abeer M. Mahmoud

Associate professor-faculty of Computer and Information Sciences- Ain Shams University

DELL Technologies

Topics : Data Science and Big Data Analytics Course

Introduction to Big Data Analytics + Data Analytics Lifecycle	Review of Basic Data Analytic Methods Using R	Advanced Analytics – Theory and Methods	Advanced Analytics - Technology and Tools	The Endgame, or Putting it All Together + Final Lab on Big Data Analytics
Big Data Overview State of the Practice in Analytics The Data Scientist Big Data Analytics in Industry Verticals Data Analytics Lifecycle	Using R to Look at Data - Introduction to R Analyzing and Exploring the Data Statistics for Model Building and Evaluation	K-means Clustering Association Rules Linear Regression Logistic Regression Naive Bayesian Classifier Decision Trees Time Series Analysis Text Analysis	Analytics for Unstructured Data (MapReduce and Hadoop) The Hadoop Ecosystem In-database Analytics – SQL Essentials Advanced SQL and MADlib for In-database Analytics	Operationalizing an Analytics Project Creating the Final Deliverables Data Visualization Techniques + Final Lab – Application of the Data Analytics Lifecycle to a Big Data Analytics Challenge

DELL Technologies

Lesson 2: Business value from Big Data

Lesson 2: Business value from Big Data

In this lesson we discuss:

- Business drivers for organizations to adopt Big Data analytics
- Business intelligence vs. data science
- Typical analytical architecture for business intelligence
- Considerations for Big Data analytics

Business drivers : are key inputs and activities that drive the operational and financial results of a business

Big Data analytics

- Organizations can use their Big Data to:
  - Uncover new emerging trends.
  - Identify potential business opportunities.
  - Discover new ways to gain competitive advantages.
- Big Data demands an approach to analytics that is **flexible**, **accessible**, and **fast**.
- **To maximize the value of Big Data, analysts**:
  - Leverage data lakes that can store a massive amount of data.
  - Apply statistical and machine learning techniques.
  - Collaborate and share insights (it’s a team sport).

Business drivers to adopt Big Data analytics

Business driver	Desired outcomes
1-Optimize business operations	Improve profitability and operating efficiency
2-Identify business risk	Reduce customer churn and fraud
3-Identify new business opportunities	Increase sales revenue—for example, upsell, cross-sell, and find new customer prospects
4-Stay informed of laws or regulatory requirements	Cost-effectively comply with industry regulations—anti-money laundering, Fair Lending, Basel II

Fraud: deception to obtain financial gains

Customer churn : customer turn over or loss of clients or customer un-satisfaction

Deriving business value with Big Data analytics—communication, media, and entertainment

- Ability to predict what customer wants by analyzing usage patterns
- Ad Targeting to provide personalized advertising at the right time and right place
- Increased customer acquisition and retention by analyzing their social media behavior
- Efficient allocation of capital, to drive growth and profitability
- Enhanced planning and optimization of network services according to trends and predictive analytics

DELL Technologies

Deriving business value with Big Data analytics—financial services

- Obtain a 360° view of customer to deliver better customer experience, improved branding, and increased revenues.
- Analyze call logs and social media activity to understand customer satisfaction levels and improve retention.
- Enhance Lender Risk management capability through behavioral analysis and understanding spending habits of customer.
- Use historical data to feed trading models and improve the performance of portfolio and revenue.
- Improve risk management by analyzing data from research, news, articles, social media, and so on.

DELL Technologies

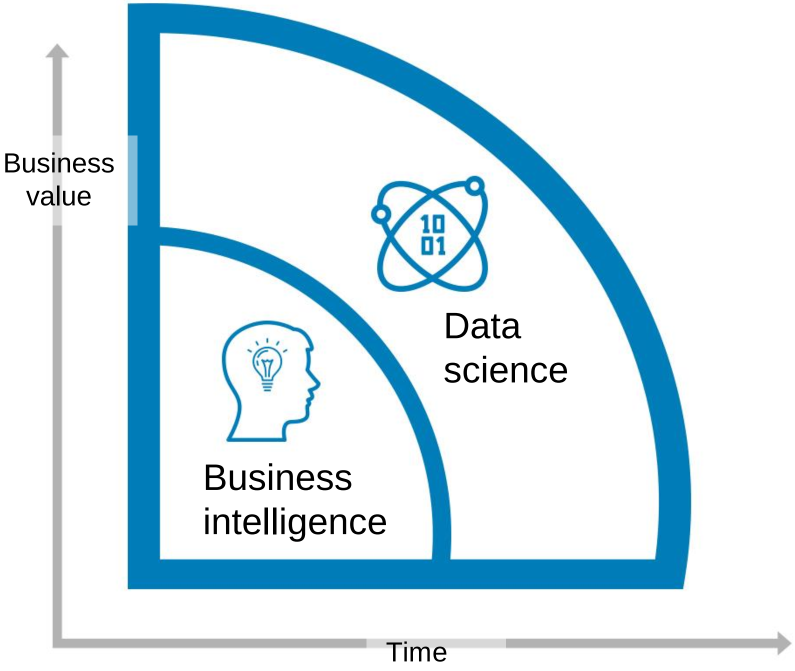
Deriving business value with Big Data analytics—healthcare

- Reduced hospital admissions and re-admissions by identifying high-risk patients ahead of time through data analytics
- Sensors embedded into every technology, creating streams of data that, when analyzed, provide insights into patient health and behavior
- Efficient allocation of capital between R&D, clinical trials, research, and so on
- Effective use of genome sequencing to make personalized medical suggestions

Data science—an emerging interdisciplinary field

- Data science combines several existing disciplines.
  - Statistics
  - Mathematics
  - Data visualization
  - Machine learning
  - Computer science
- This combination enables insights (data mining) as well as foresights (predictions).

Business intelligence versus data science



- Business intelligence reports on what has happened.
- Data science predicts what will happen.

DELL Technologies

Business intelligence versus data science, cont.

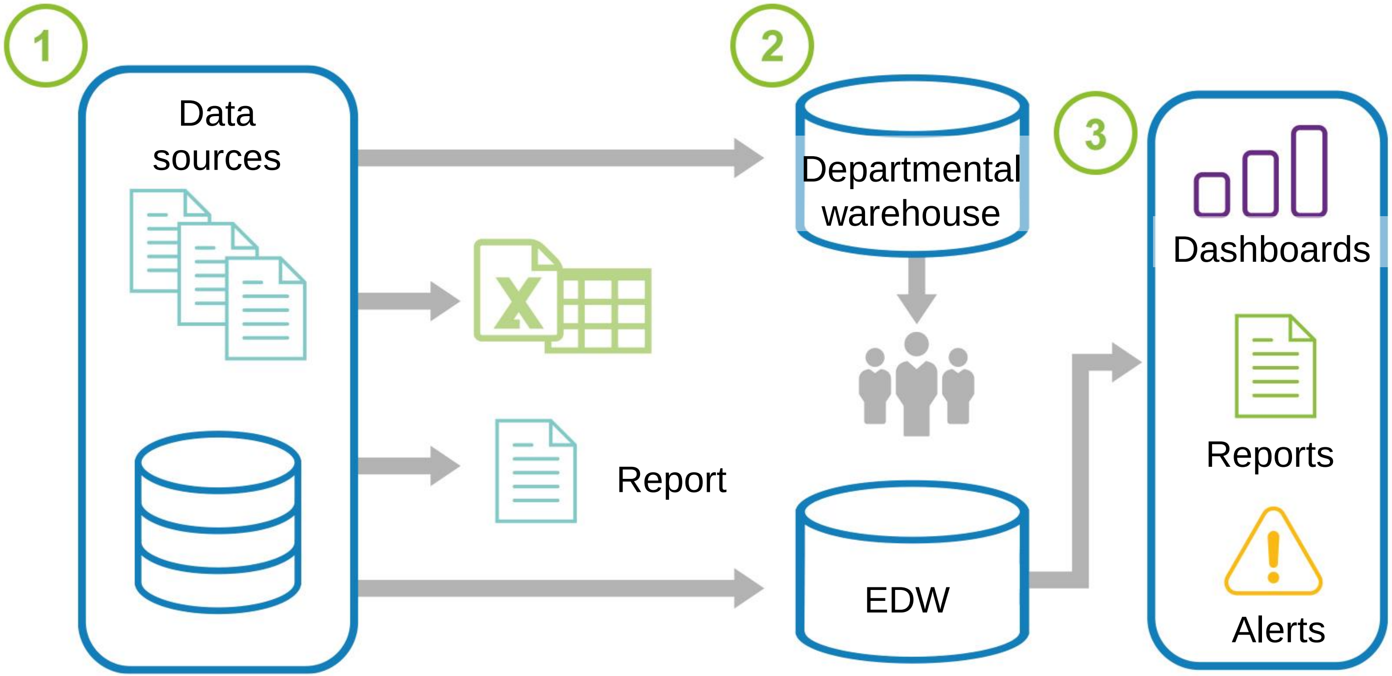
Business intelligence

- Typical techniques and data types
  - Standard and ad hoc reporting, dashboards, alerts, queries, details on demand
  - Structured data, traditional sources, manageable datasets
- Common questions
  - What happened last quarter?
  - How many did we sell?
  - Where is the problem? In which situations?

Predictive analytics and data mining—data science

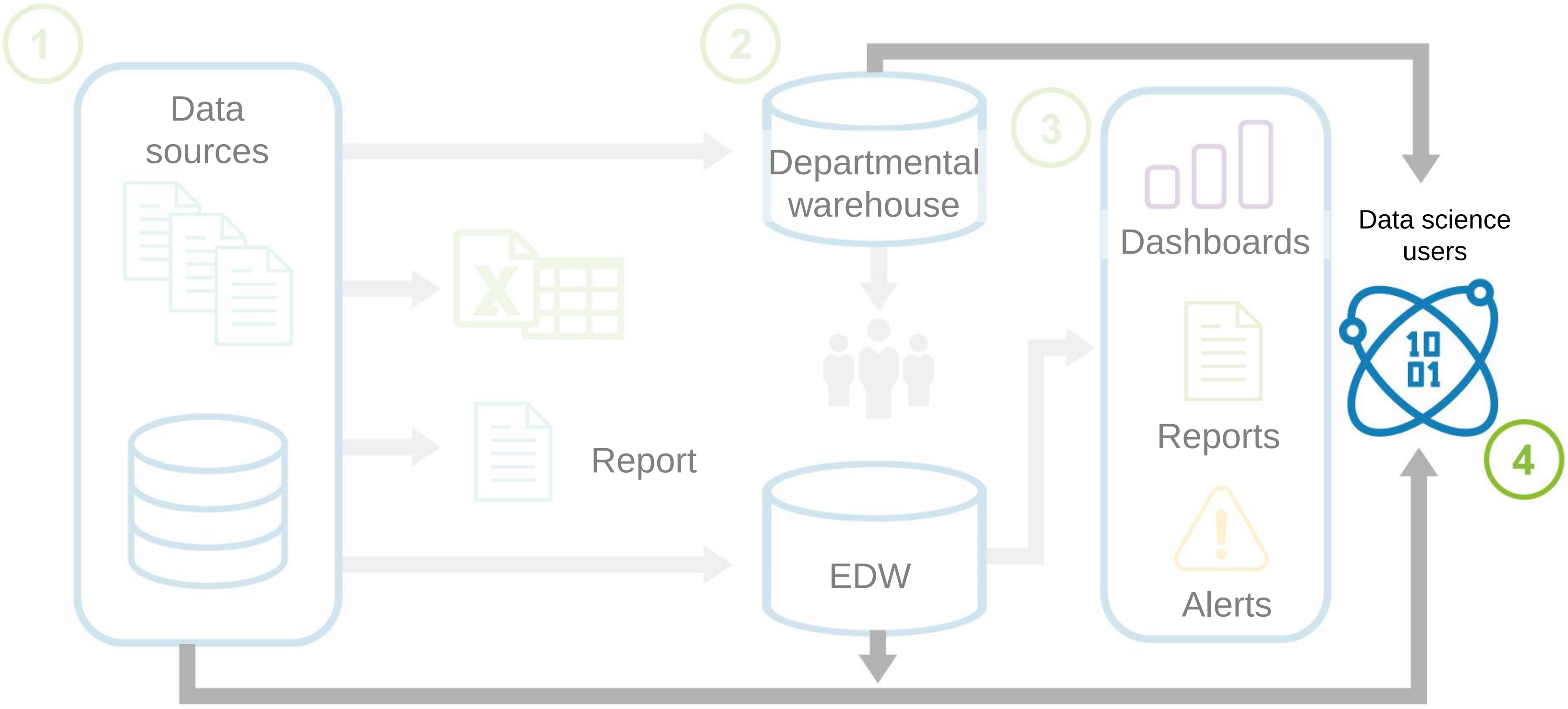
- Typical Techniques and Data Types
  - Optimization, predictive modeling, forecasting, statistical analysis
  - Structured/unstructured data, many types of sources, very large data sets
- Common Questions
  - What if…?
  - What's the optimal scenario for our business?
  - What will happen next? What if these trends continue? Why is this happening?

Typical analytical architecture for business intelligence

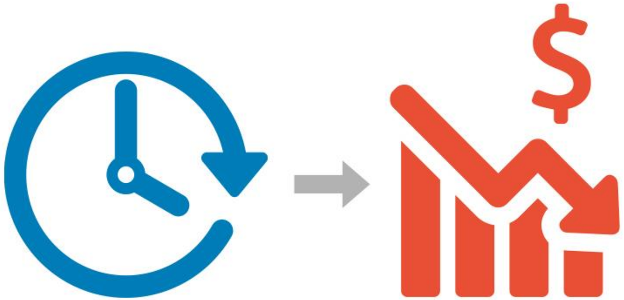


DELL Technologies

Typical analytical architecture for business intelligence, cont.



BI analytical architectures are <u>not suitable</u> for data science



- High-value data is hard to reach and leverage.
- Data is moved from EDW to <u>local analytical tools.</u>
  - EDW may have masked/hidden meaningful data.
  - Sensitive data is stored on PCs.
  - This data is difficult to share and collaborate on.
- <u>Isolated, ad hoc analytic projects</u>, rather than centrally-managed harnessing of analytics.
- Preferred state: Analytic sandbox with access to the raw data.

Mini-case study—reducing customer churn at SuperMom&PopShop

- Historical approach
  - Review quarterly reports generated on individual customer purchases.
  - In-home promotions are mailed to customers who have not made a purchase in a specified period of time.
- Big Data analytic approach
  - <u>Build an analytical model</u> to predict the likelihood of an individual customer churning.
  - Include new data sources and inputs.
  - <u>Analyze previous types of purchases</u> (home goods, tools, clothing, and so on).
  - Consider tendency to shop in-store, online, or both.
  - Consider distance from home to store locations.
  - Analyze customer demographics.

Test your self

- For a communication company suffer from customer churn explain in two to three point what traditional and data science will excute

Considerations for data science and Big Data analytics

- **Analysis flexibility**
  Where can the team explore and experiment with the data?
  - Data silos vs. analytic sandboxes
  - Analyst or IT owned
- **Decision making**
  How quickly must data-driven business decisions be made?
  - Batch processing
  - Real-time processing
- **Skills**
  Does the existing team have the necessary skills?
  - In-house expertise vs. outsourcing
  - Data science experts (aka data scientist)

Check your knowledge

Which emerging discipline provides insights (data mining) and foresights (predictions) from data?

A. Topology

C. Complex analysis

B. Business intelligence

D. Data science

DELL Technologies

Lesson 3: Data scientist

Lesson: Data scientist

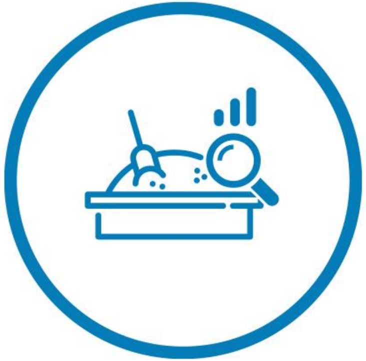
This lesson covers:

- Key roles for the new Big Data ecosystem.
- Responsibilities of a data scientist.
- Profile of a data scientist.

Key roles for the new Big Data ecosystems

Deep analytic
talent

Data savvy
professionals



Technology and
data enablers

Key roles for the new Big Data ecosystem—deep analytical talent

- Are tech-savvy with strong analytical skills
- Are able to build models and derive insights from data
- Are able to work with sandboxes
- Shortfall of some 250,000 data scientists
- Examples:
  - Statisticians
  - Data scientists

DELL Technologies

Key roles for the new Big Data ecosystem—data-savvy professionals

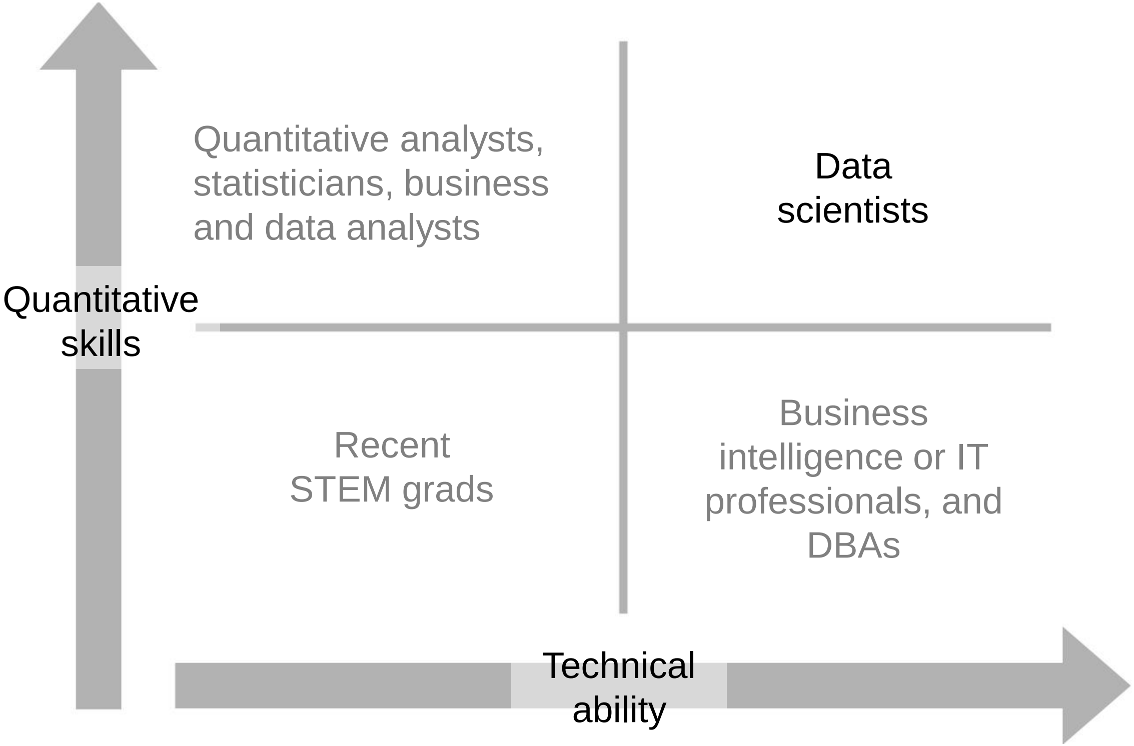
- Are equipped with domain knowledge
- Have a basic knowledge of statistics or machine learning
- Are able to appreciate models built by data scientists
- Examples:
  - Financial analysts
  - Market research analysts
  - Business or functional manager

DELL Technologies

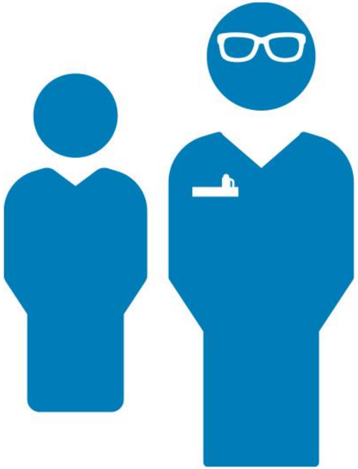
Key roles for the new Big Data ecosystem—technology and data enablers

- Provision and administer analytical sandbox
- Manage large-scale data architecture
- Examples:
  - Database administrator
  - Programmer

Roles by technical and quantitative skills



Data scientist—an emerging career



Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

Profile of a data scientist

Quantitative skills

- Expertise in mathematics and statistics

Technical aptitude

- Proficient programming skills
- Strong IT background

Skeptical and critical thinking

- Examine the work in a non-biased manner

Curious and creative

- Passionate about data
- Find novel ways to solve problems

Communicative and collaborative

- Articulate the business value in a clear way
- Collaboratively work with other groups

Test yourself

Which is an appropriate skill that a data scientist must have?

A. People management

B. Financial management

C. Skeptical and critical thinking

D. Graphics design

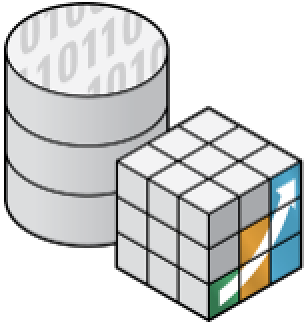
Module Summary

Key topics covered in this module were:

- Big Data and its characteristics
- Sources of Big Data
- Evolving analytical architecture
- The role of data scientist

Thanks
for your attention

data marts tend to be a single subject area data repository and/or linked to a specific corporate application (such as finance, HR, CRM, ERP, logistics, sales tracking etc). The source data is pushed to a specific line of business for analysis. The push and loading processes implements all the necessary data cleansing and transformation routines so the data arrives into its destination schema ready for use. Most importantly, the data push happens on a regular basis and is driven by the needs of the business.



A "sandbox" is generally meant as a non-operational environment where business analysts and data scientists can test ideas, manipulate data and model "what if" scenarios without placing an excessive computational load on the core operational processes. It has a finite life expectancy so that when timer runs out the sandbox is deleted and the associated discoveries are either incorporated into the enterprise warehouse