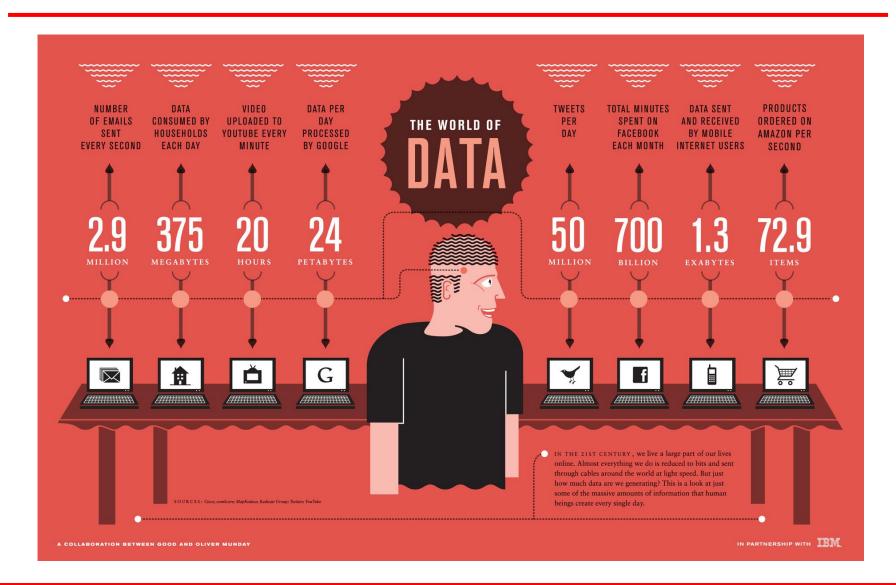
A Data Based World

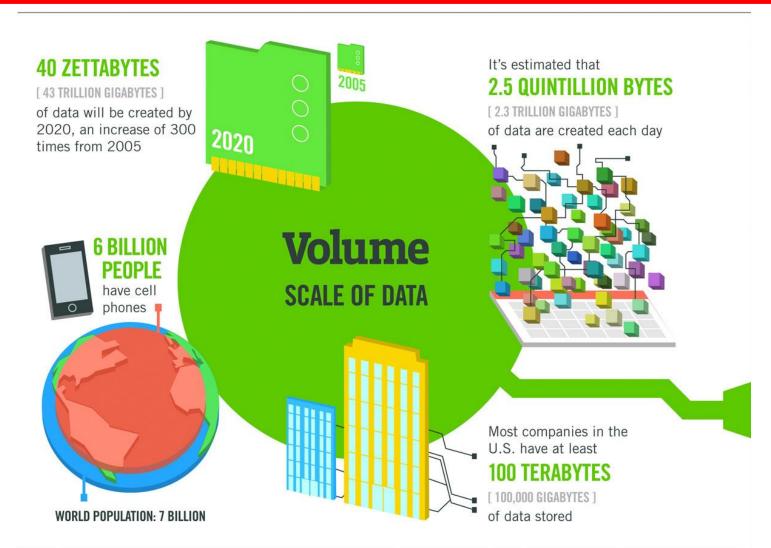
```
11010
             10x 016101 ±0101011010000000
               1010101
                    10110101011010101
                                                        .510001
         J1100_C 0_070101
                        10000000000
                                                    11/201010101010101111
10101101120116.30101e1J1 J1010161
                                                  ._\0\1011011101110110001001001010101010111
                         010110101ბ
110101010110101011010101010
                                       '10h
                         0101011
11011101110110001001001010101\u00e4101
                         1.100
                               1101
                                      (01010100011011010101017
                                      101
                         101
      1001010101010101
                                      10100
                                                                       _U1
                                      )1'
       101010101101000c 000001c
                                  10101010110101010101
                                   )1100101010101010171
                                  1017 + 23 2010101010 0101010101011010101012017
         71010101101000000
                                   710101000110110
                                  ``000001010111
           2010101°11 2
                                  10101001010010010
           1,7101
                                 .0101010101011010101010101
                                 .11010101011001010 \101010
                                                    10101 10101
             1016-01
              <sup>1</sup>∩1.0±ι
                                40001101101010101101.01°
                                                     101
                                                          10101
                 16 ....010.
                                 9101010110101010101110.
                                                     0.7
                                                          1710
                   1001001
                                                                .11
                                  10010101010101010110
                                                             ±01
                   101010101
                                       10101010101010
                  010101010101
                                                           16 01
                                       11111110101
                                                           310
                  701010101101010
                                       010101101
                   10110001001007
                                       001010101
                   711011010101010
                                       .101010101 1
                                                                J101 1v
                    2101010101010
                                       0101011ቦ
                                              10
                                                               ...01110116
                    01011111111
                                              0.7
                                       )101010:
                                                              υ10100011011
                    1010110
                                        010101
                                                              0101101010001
                    0101010
                                        71010
                                                              )101101010101
                                        1000
                    L010101
                                                                   °0101′
                    1011^7
                                                                    111
                    1011
                                                                     7
                    011
                    10
```

 $\mathfrak{O}^{\mathbf{Z}}$

Data growth expectations (by 2010)



Data growth expectations (by 2020)



taken from http://www.ibmbigdatahub.com/infographic/four-vs-big-data

Why is data so desirable?

Data => Information => Knowledge

Information: the crucial stage

Data Quality Management

- Data Quality Analysis
- · Data Matching & Cleansing
- Data Conversion
- Continuous Monitoring & Reporting

Data Governance

- · Data Ownership & Protection
- Master Data & Metadata Management
- Data Dictionary & Standards Maintenance
- Logical Layer Security & Audit Validation
- Data Lifecycle Management
- Compliance & Privacy
 Management

Data Warehousing & Business Intelligence

- Relational & Dimensional Data Modeling
- Extract, Transform & Load (ETL) Development
- ETL Quality Assurance
- · Business Rules Development
- Online Analytical Cube (OLAP) Development

INFORMATION

Unstructured Data Management

Data Architecture & Design

Value Chain Analysis

• Enterprise Data Modeling

· Enterprise Data Integration

- · Content Management & Discovery
- · Retrieval, Searching & Indexing
- Security & Protection
- Storage & Retention
- Backup & Recovery

Online Transaction Processing (OLTP) Development

- · Relational Database Modeling
- Stored Procedure Design, Development & Quality Assurance
- Query Optimization

Information Presentation

- Dashboard & Interactive Reports
- Software Application Components
- Precision Printed Output
- Historical Reporting & Predictive Analytics
- Data Mining & Mobile Access

Database Administration

- Database Setup, Mirroring & Sharding
- Performance Analysis & Tuning
- Physical Layer Security & Audit Validation
- Retention & Capacity Planning
- Disaster Recovery & Availability Planning
- DBMS Release Migration
- Data Encryption & Data Masking

taken from Data Science Central®, 2018

Data related challenges

store but also process
all the available data

democratize but also secure the access to data

merge heterogeneous data but validate its origin

. . .

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21th century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- Supervised learning: decision trees, random forests. logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- Strategic, proactive, creative, innovative and collaborative



PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ★ Statistical computing packages, e.g., R
- □ Databases: SOL and NoSOL
- ☆ Relational algebra
- Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ★ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization: customer tracking and on-site analytics; predictive analytics and econometrics; data warehousing and big data systems; marketing channel insights in Paid Search, SEO, Social, CRM and brand.

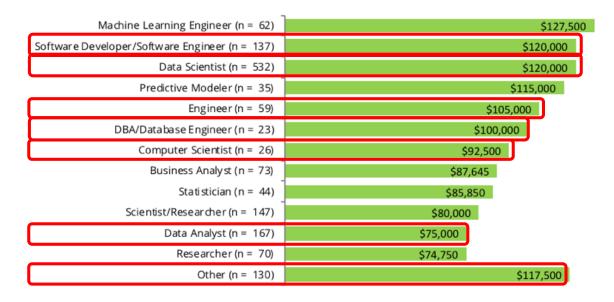


Data related challenges

Areas requiring strong data related skills

Data related jobs

Median Annual Compensation by Job Title for United States Data Professionals



Respondents are from the United States. Values are in US Dollars.

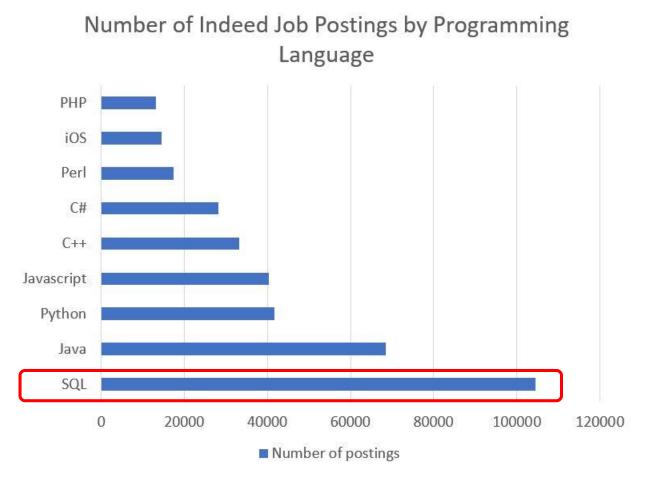
Data are from the Kaggle 2017 The State of Data Science and Machine Learning study. You can learn more about the study and download the data here: https://www.kaggle.com/surveys/2017.

Job Titles are ranked by median annual salary. Only job titles with ample sample size (n > 20) are presented.



Copyright 2018 Business Over Broadway

Data and Programming Languages



taken from Coding Dojo Blog®, 2017