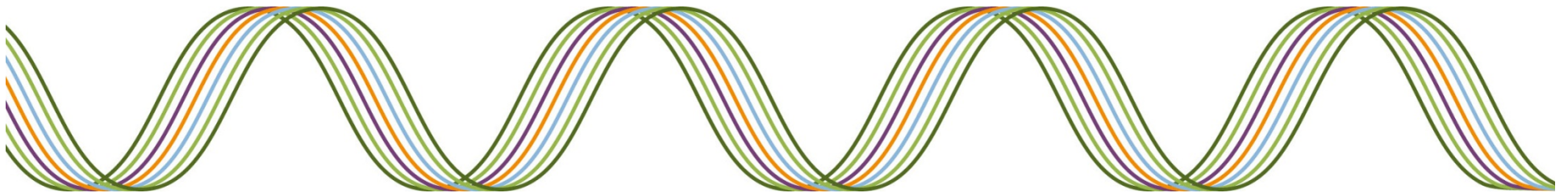


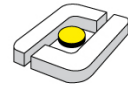
Hochschule Osnabrück
University of Applied Sciences

The Impact of Big Data on the Job Market – Results from a Job Mining Study

Prof. Dr. Frank Bensberg
f.bensberg@hs-osnabrueck.de

Gandalf Buscher
gandalf.buscher@hft-leipzig.de



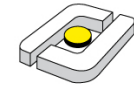


Motivation

Design of the Study

Big Data Job Profiles / Institutional Profiles

Conclusion

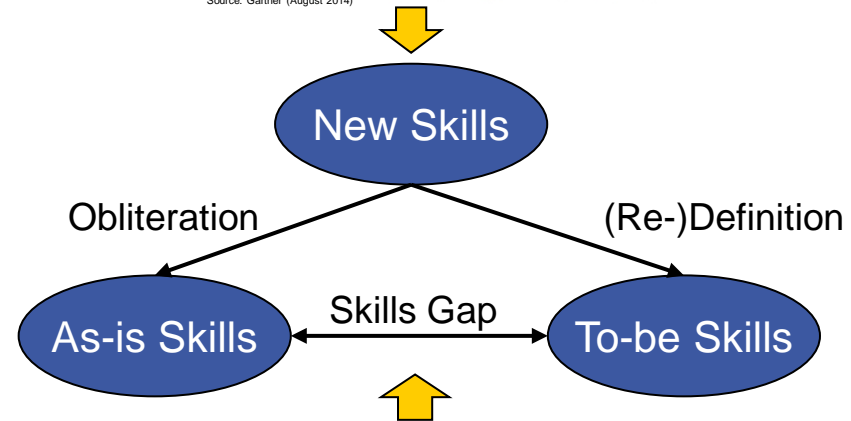
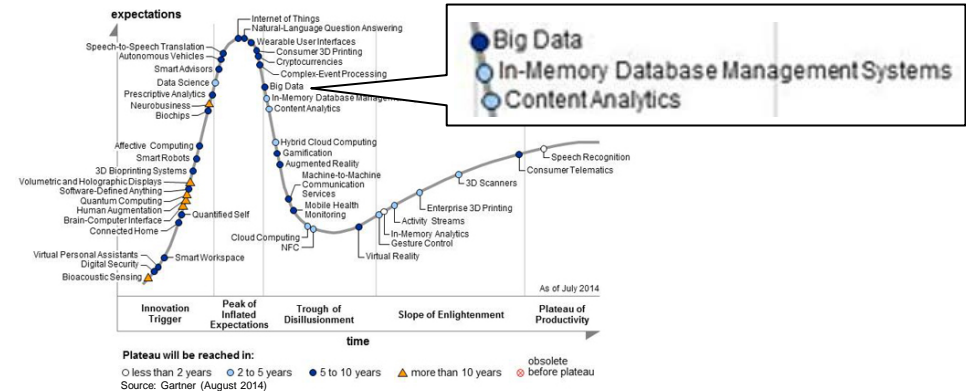


Motivation

- Evolution of new information technologies leads to erosion of knowledge
- New skills obliterate as-is skills and redefine competency requirements (to-be skills)
- Skills gaps reduce the capability to innovate and to develop new markets
- How does Big Data redefine competency requirements and shape the job market?

Evolution of New Technologies leads to IT Skills Gaps

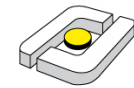
Gartner's 2014 Hype Cycle for Emerging Technologies



Business
Development

HR Recruitment &
Development

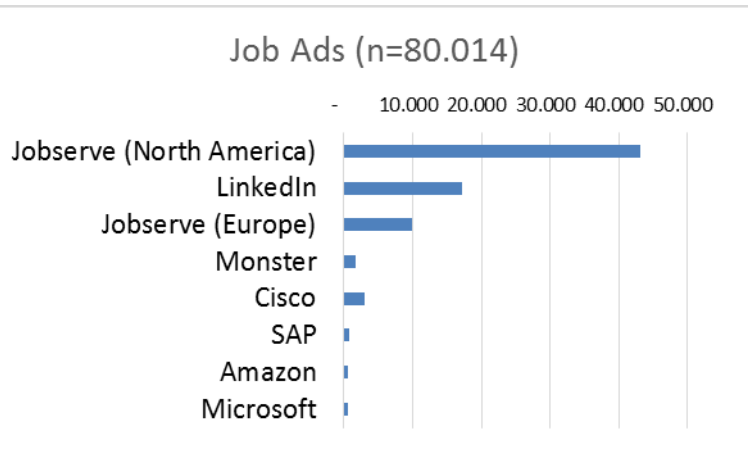
Education
Providers



Design of the Study

- Research question
 - Which skills and skill profiles are searched for in the labour market for the domain of Big Data?
- Research method
 - Job advertisement analysis in combination with Text Mining methods

Job Mining Process



- Job ads periodically collected
- from June 2014 until April 2015
- n=80.014 job ads selected (out of 2.8m)

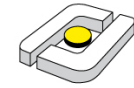
Data Preprocessing
& Data Cleansing

- Identification of job ads with genuine analytical focus

Discovery of
Big Data
Job Profiles

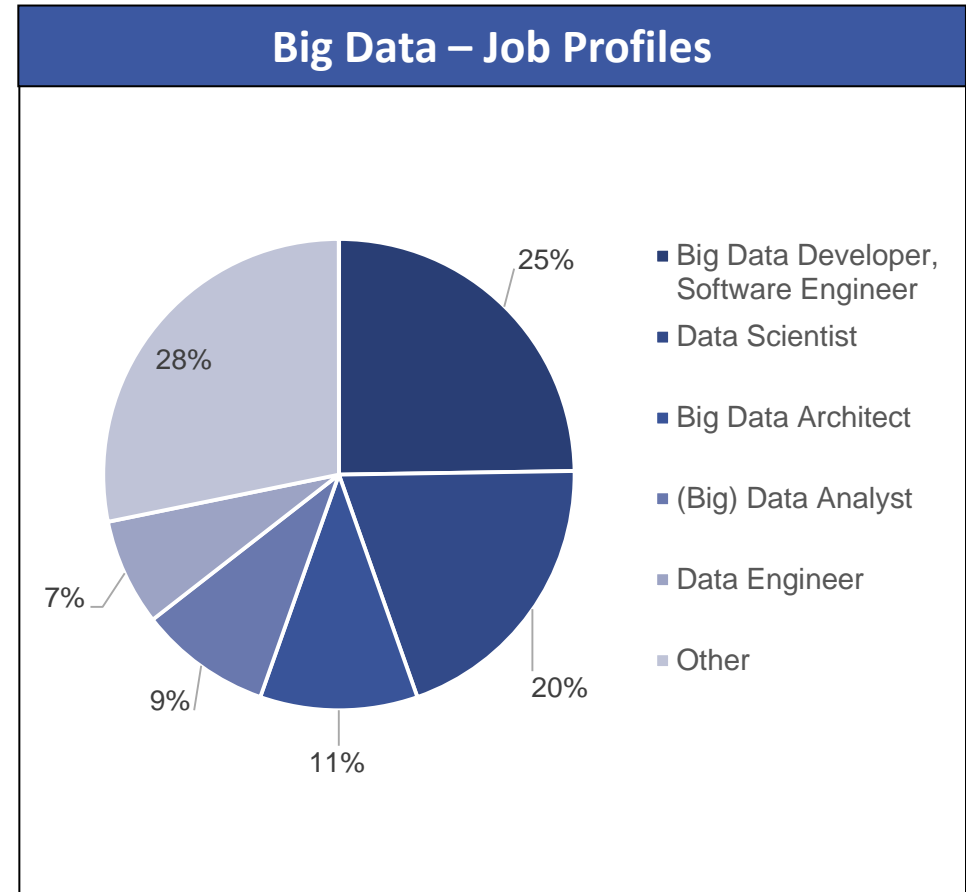
Analysis of Big Data Skills
for specific Sectors
and Institutions

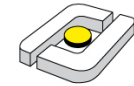
- Usage of Text Mining methods to discover job families and institutional profiles



Big Data Job Profiles

- About 70% of the job titles represent five job families
 - Big Data Developer/Software Engineer
 - Data Scientist
 - Big Data Architect
 - (Big) Data Analyst
 - Data Engineer
- Specific job profiles
 - Big Data Technology Researcher (e.g. SAP)
 - Data Quality Manager (e.g. Amazon)
 - Data Services Developer (e.g. Microsoft)
 - Data Visualization Engineer (e.g. BigR.io)
 - Decision Scientist (e.g. Microsoft)





Data Scientist – Job Profile

Job Ads

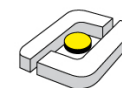
1.942

Organizations (Top 10)

Liberty Personnel Services; Opus Recruitment Solutions Ltd; BigR.io; Career Brokers; Next Ventures Ltd; ITBM Consulting Inc; A3Logics US; Engage3; All In Analytics; Alivia technology

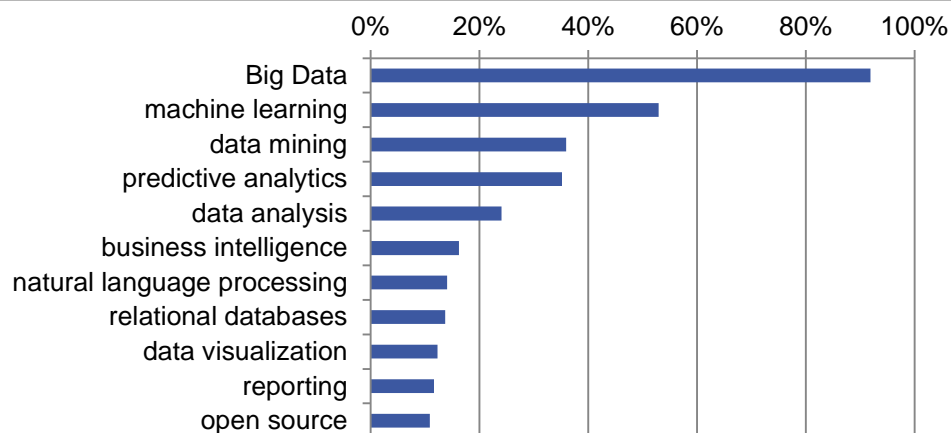
Sample Job Titles

Data Scientist
Big Data Scientist
Senior Data Scientist
Principal Data Scientist (R, Matlab, [...], Spark)
Senior Data Scientist (Python, Matlab, R, SQL)

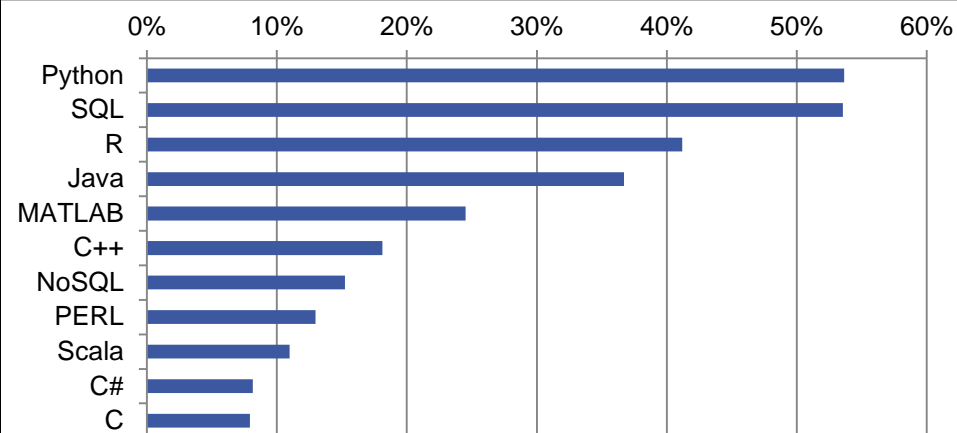


Data Scientist – Job Profile

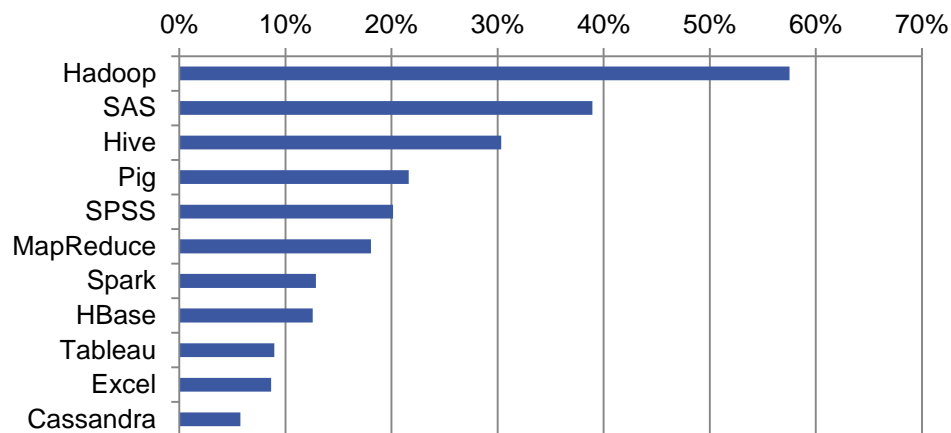
Concepts



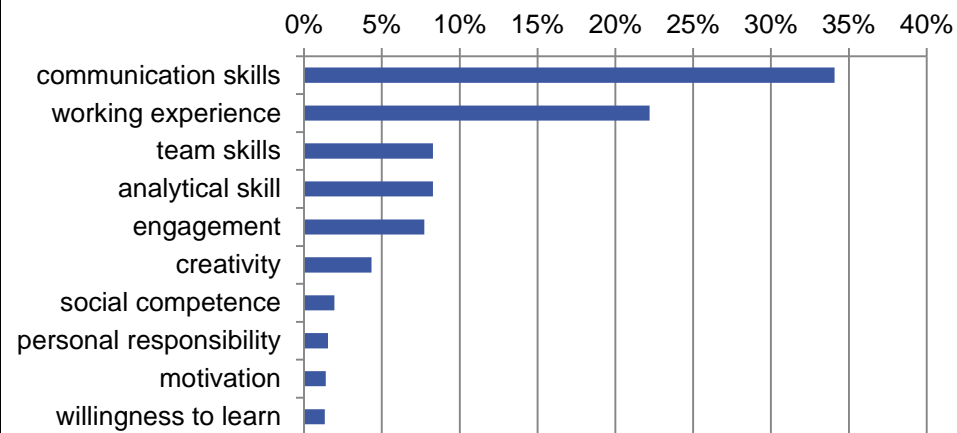
Programming Languages

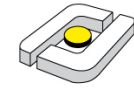


Products



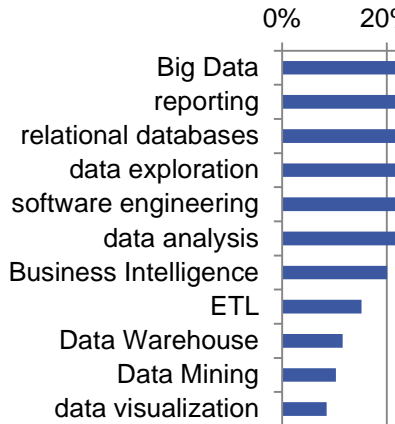
Soft Skills





Data Analyst – Job Profile

Concepts



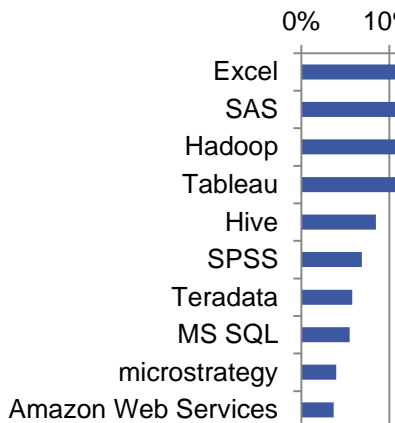
Job Ads

1.690

Organizations (Top 10)

MedeAnalytics, Liberty Personnel Services, FILD, TalentRISE, R Square, Libertyjobs.com, LS Direct Marketing, The Judge Group, UnitedHealth Group, Cognius

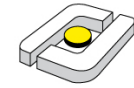
Programming Languages



Sample Job Titles

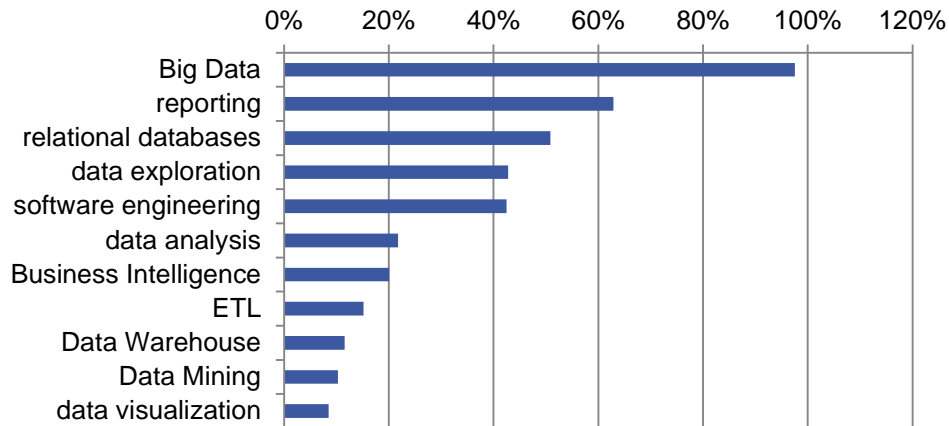
Business **Data Analyst** - Health Plan Solutions
Data Analyst
Senior **Data Analyst**
Big **Data Analyst**
Data Analyst (BIG DATA - Finance Analytics)

organizational skill]

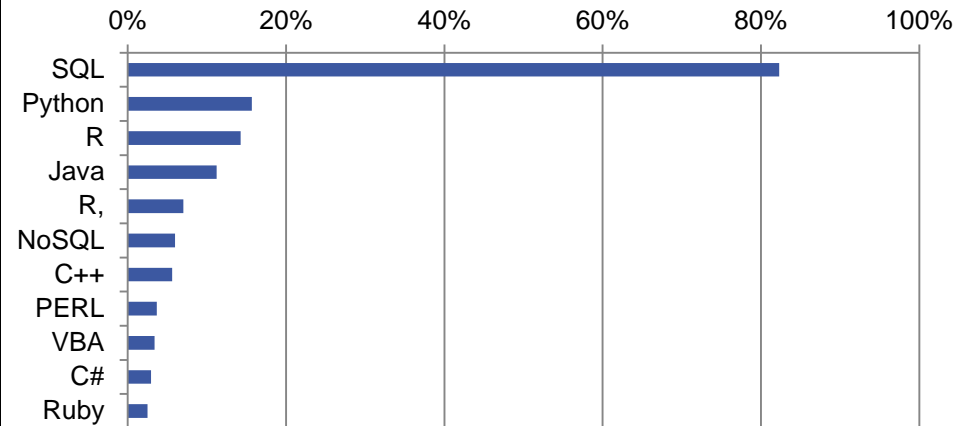


Data Analyst – Job Profile

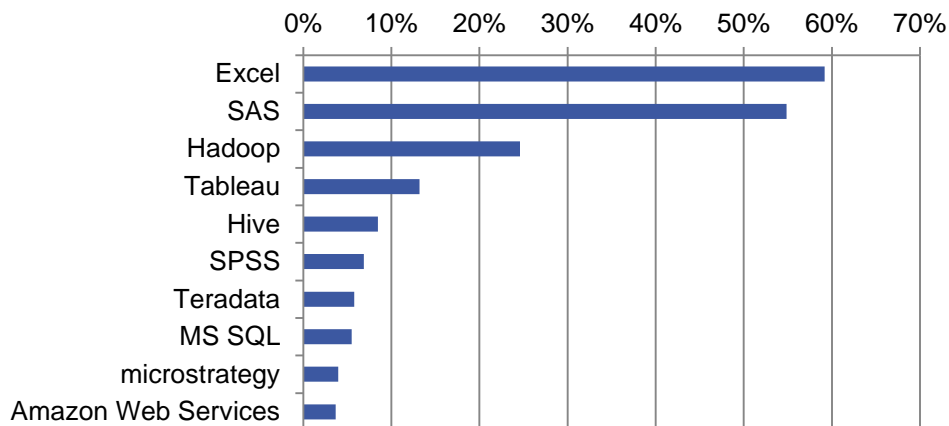
Concepts



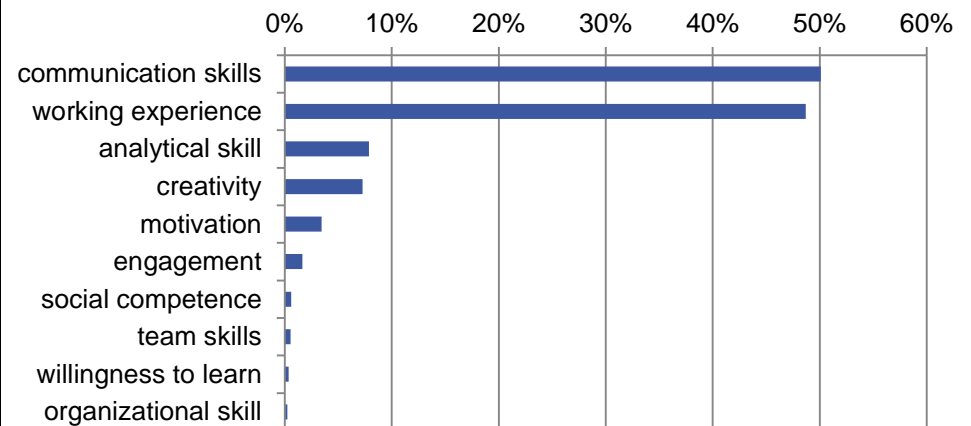
Programming Languages

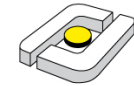


Products

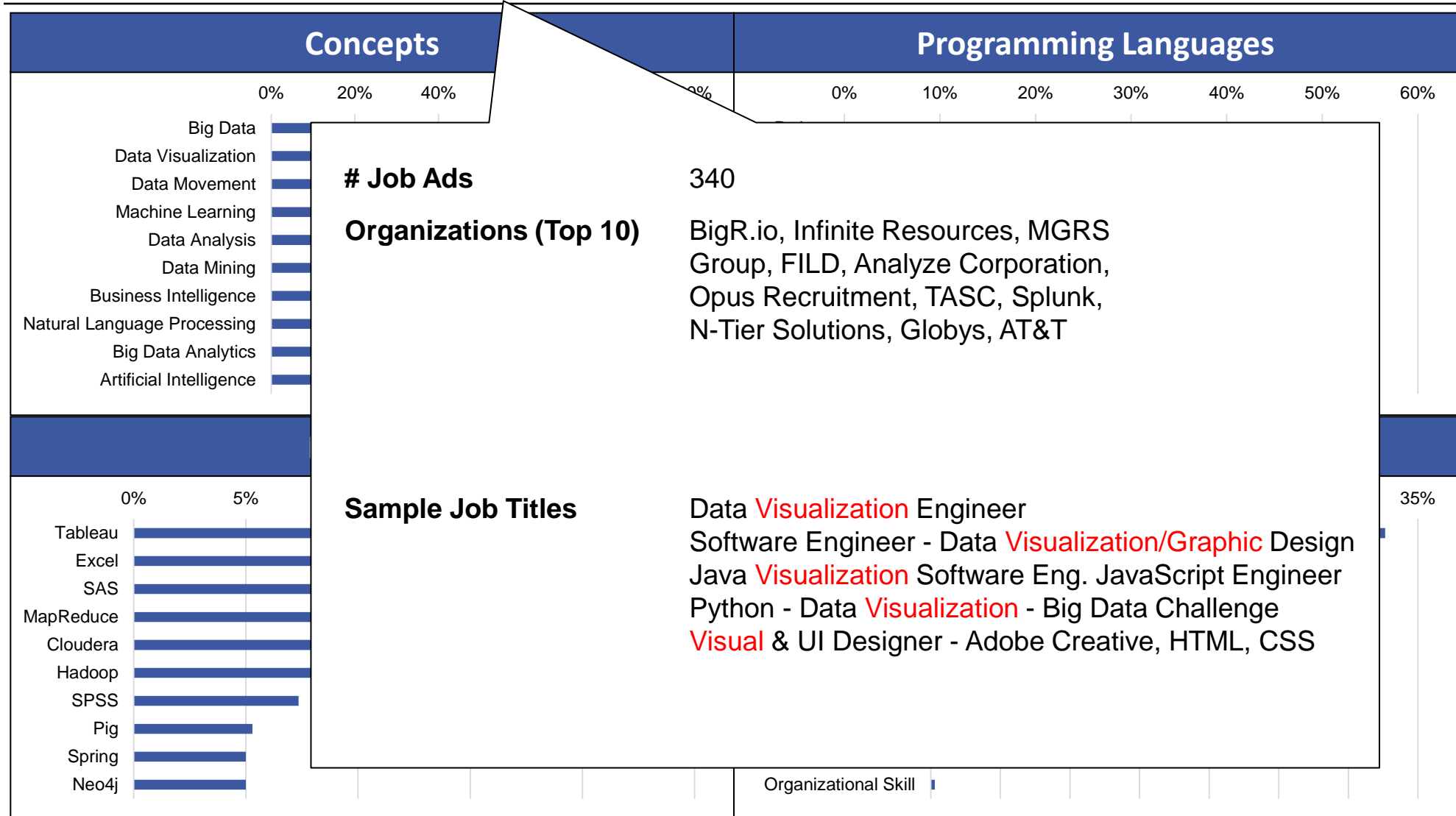


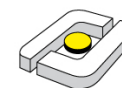
Soft Skills



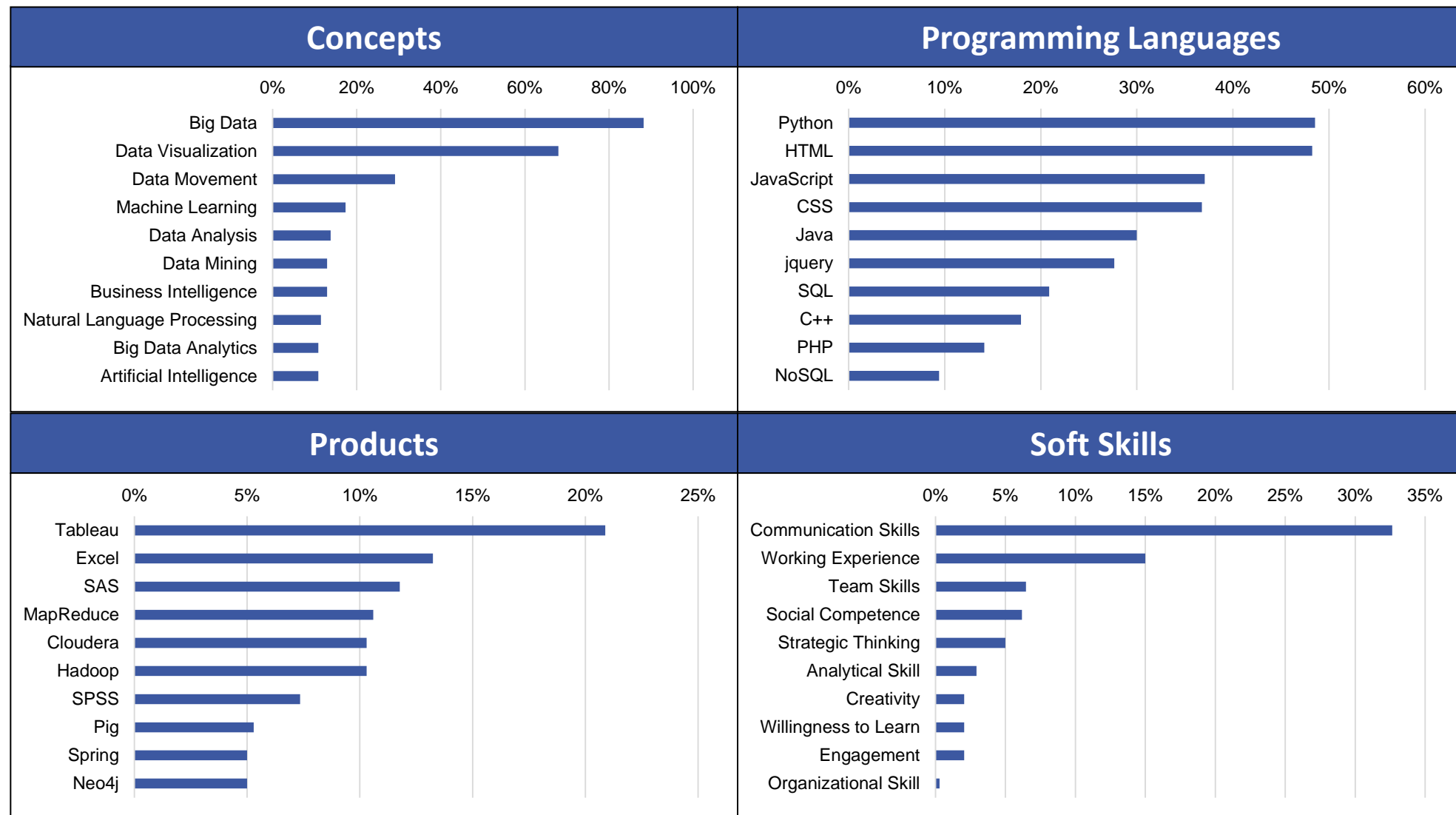


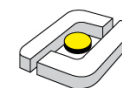
Data Visualization Engineer – Job Profile





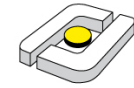
Data Visualization Engineer – Job Profile





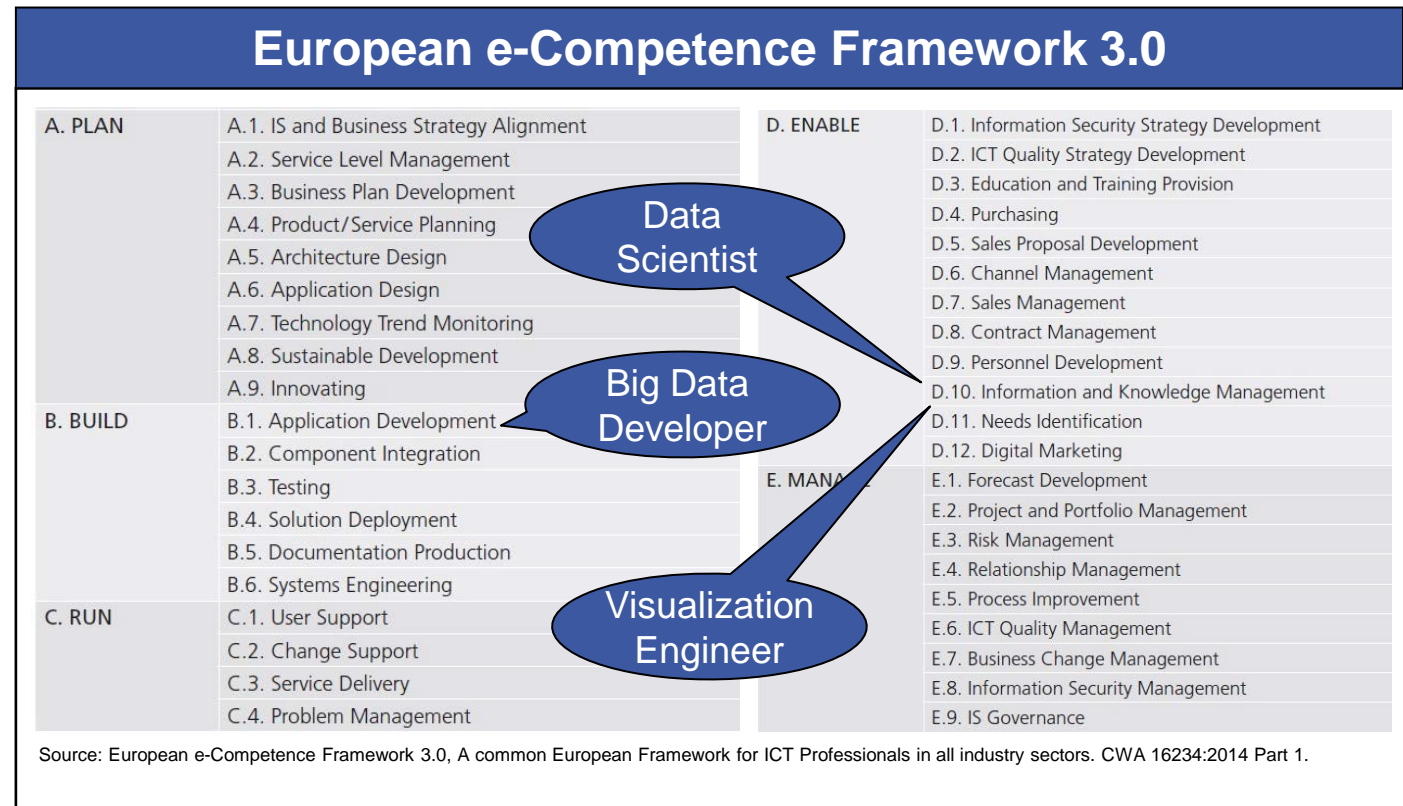
Institutional Profile – BigR.io

| Base Data | | Concepts | Programming Languages | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|--|-------------|-----------------------|----------|------|---------------|------|-----------------------|------|--------------------|------|---------------|------|---------------|------|---------|------|----------------------|------|---------------------|------|-----------|------|--|----------|------------|----------------------|------|--------------------|------|-------------------|------|-------------------|------|------|------|--------|------|------|------|------|------|---|------|---|------|
| Data Source <u>Public Job Board</u> Corporate Job Board First Ad 06/25/2014 Last Ad 04/18/2015 Duration 10 months # Job Ads Total 4.026 Net 475 Locations USA, Boston (188), USA, Waltham (55), USA, Cambridge (51) | <table border="1"> <caption>Concepts</caption> <thead> <tr> <th>Concept</th> <th>Percentage</th> </tr> </thead> <tbody> <tr><td>Big Data</td><td>~85%</td></tr> <tr><td>Data Movement</td><td>~85%</td></tr> <tr><td>Business Intelligence</td><td>~15%</td></tr> <tr><td>Data Visualization</td><td>~15%</td></tr> <tr><td>Data Modeling</td><td>~15%</td></tr> <tr><td>Data Analysis</td><td>~15%</td></tr> <tr><td>ETL</td><td>~15%</td></tr> <tr><td>Predictive Analytics</td><td>~15%</td></tr> <tr><td>Distributed Systems</td><td>~15%</td></tr> <tr><td>Dashboard</td><td>~15%</td></tr> </tbody> </table> | Concept | Percentage | Big Data | ~85% | Data Movement | ~85% | Business Intelligence | ~15% | Data Visualization | ~15% | Data Modeling | ~15% | Data Analysis | ~15% | ETL | ~15% | Predictive Analytics | ~15% | Distributed Systems | ~15% | Dashboard | ~15% | <table border="1"> <caption>Programming Languages</caption> <thead> <tr> <th>Language</th> <th>Percentage</th> </tr> </thead> <tbody> <tr><td>SQL</td><td>~55%</td></tr> <tr><td>Python</td><td>~35%</td></tr> <tr><td>CSS</td><td>~20%</td></tr> <tr><td>NoSQL</td><td>~20%</td></tr> <tr><td>HTML</td><td>~20%</td></tr> <tr><td>jquery</td><td>~20%</td></tr> <tr><td>Java</td><td>~15%</td></tr> <tr><td>.Net</td><td>~15%</td></tr> <tr><td>C</td><td>~15%</td></tr> <tr><td>R</td><td>~15%</td></tr> </tbody> </table> | Language | Percentage | SQL | ~55% | Python | ~35% | CSS | ~20% | NoSQL | ~20% | HTML | ~20% | jquery | ~20% | Java | ~15% | .Net | ~15% | C | ~15% | R | ~15% |
| Concept | Percentage | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Big Data | ~85% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data Movement | ~85% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Business Intelligence | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data Visualization | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data Modeling | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data Analysis | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ETL | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Predictive Analytics | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Distributed Systems | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Dashboard | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Language | Percentage | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| SQL | ~55% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Python | ~35% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CSS | ~20% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| NoSQL | ~20% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HTML | ~20% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| jquery | ~20% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Java | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| .Net | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| C | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| R | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Interpretation | Products | Soft Skills | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <ul style="list-style-type: none"> ■ High volume of vacancies for Big Data and associated technologies ■ Data Movement as a new topic ■ High diversity of products ■ Offerings for Data Visualization Engineers | <table border="1"> <caption>Products</caption> <thead> <tr> <th>Product</th> <th>Percentage</th> </tr> </thead> <tbody> <tr><td>Hadoop</td><td>~35%</td></tr> <tr><td>Hive</td><td>~25%</td></tr> <tr><td>HP Vertica</td><td>~20%</td></tr> <tr><td>Microstrategy</td><td>~20%</td></tr> <tr><td>SAS</td><td>~20%</td></tr> <tr><td>Spark</td><td>~20%</td></tr> <tr><td>Tableau</td><td>~20%</td></tr> <tr><td>MS SQL</td><td>~15%</td></tr> <tr><td>Solr</td><td>~15%</td></tr> <tr><td>MapReduce</td><td>~15%</td></tr> </tbody> </table> | Product | Percentage | Hadoop | ~35% | Hive | ~25% | HP Vertica | ~20% | Microstrategy | ~20% | SAS | ~20% | Spark | ~20% | Tableau | ~20% | MS SQL | ~15% | Solr | ~15% | MapReduce | ~15% | <table border="1"> <caption>Soft Skills</caption> <thead> <tr> <th>Skill</th> <th>Percentage</th> </tr> </thead> <tbody> <tr><td>Communication Skills</td><td>~35%</td></tr> <tr><td>Working Experience</td><td>~25%</td></tr> <tr><td>Analytical Skills</td><td>~10%</td></tr> <tr><td>Social Competence</td><td>~5%</td></tr> </tbody> </table> | Skill | Percentage | Communication Skills | ~35% | Working Experience | ~25% | Analytical Skills | ~10% | Social Competence | ~5% | | | | | | | | | | | | |
| Product | Percentage | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hadoop | ~35% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hive | ~25% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HP Vertica | ~20% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Microstrategy | ~20% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| SAS | ~20% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Spark | ~20% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Tableau | ~20% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| MS SQL | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Solr | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| MapReduce | ~15% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Skill | Percentage | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Communication Skills | ~35% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Working Experience | ~25% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Analytical Skills | ~10% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Social Competence | ~5% | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |



Conclusion

- Big Data contributes to the evolution of new job profiles in the labour market
- Educational programmes and ICT curricula should consider this evolving skill demand
- Established competency frameworks should be adapted to support HR management



THANK YOU
FOR YOUR ATTENTION!