Ex:- Fraud Transaction : banks

Client Retention : business

Programming Languages	Softwares
• devise programs that can execute specific operations	Python + Java + R + Scala + SQL
• programs can be reused	• Smaller Scale
→ R Python	• easier to learn
• suitable for mathematical & statistical computations	→ Excel relatively complex computations & good visualizations quickly
• adaptable.	
• not able to address problems specific to some domains	→ IBM SPSS statistical analysis
→ SQL : RDBMS	→ Apache Hadoop, Apache Hbase
→ Matlab	Mongo DB
• ideal for mathematical functions / matrix manipulations.	→ Hadoop: software framework
	→ Business Intelligence : SAS, Power BI Qlik q , Tableau
• paid service	→ EViews: econometric time series Model

→ Java   Scala   → Stata : academic statistical &
- Combining data from   econometric research.
  multiple sources

business & theoretical knowledge	learning a software tool	learning a programming lang.
Relevant	• easy	• not easy
strong	• sufficient	• gives freedom
	• not enough	

Data Architect ⎫
Big Data Architect ⎪ • crusial titles in market
Data Engineer ⎪
Big Data Engineer ⎭

DATA

Data Architect:
- designs the way data will be retrieved, processed & consumed.

Data Engineer:
- process the obtained data so that it is ready for analysis

Database Administrator:
- handles control of data
- mainly works with traditional data (management of Big data is usually automated)

BI

BI Analyst:
- performs analysis & reporting of post historical data

BI Consultant:
- external BI Analyst

BI Developer: (uses python + SQL)
- performs analysis specifically designed for company