



Introduction to Data Science

Venkat Reddy

Contents

- Introduction to Data Science
- Data Science Skills
- Data Science Techniques
- Data Science Tools
- Training Needs for a data scientist

How Target-Super Market Figured Out A Teen Girl Was Pregnant Before Her Father Did?

Data Science is everywhere

- Do you receive any marketing calls? Have you ever received any marketing call for Audi car? -Marketing analytics
- Have you ever wondered, why only you are getting promotional offers on cloths and accessories where as I am getting offers on apartments? - Retail sales analytics
- Why not the full talk time and message offers not same across all states & all networks ? Telecommunications
- How does a bank decide the potential fraud transactions from millions of credit card swipes? - Fraud analytics

What is data science?

- Data Driven decision making
- Building business strategies based on data analysis
- Removing the manual and judgemental error
- Making use of past experience(data) to make build future strategies

Applications of Data Science

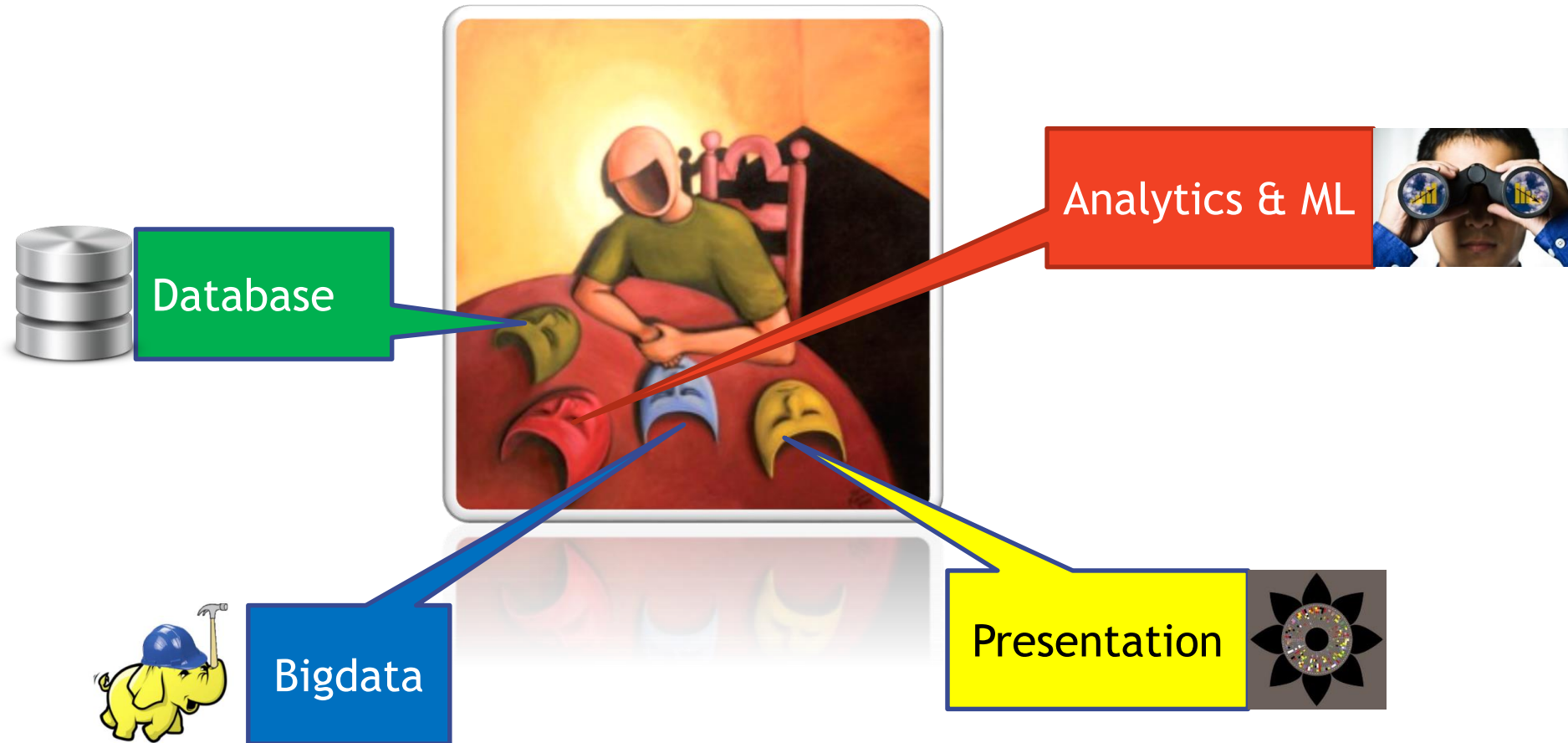
Data is growing at a rapid pace that lead to many application of data driven strategies

- Recommendation systems
- Image recognition
- Speech recognition
- Fraud transaction identification
- Spam filtering
- Identify Sales leads
- Cross selling and upselling

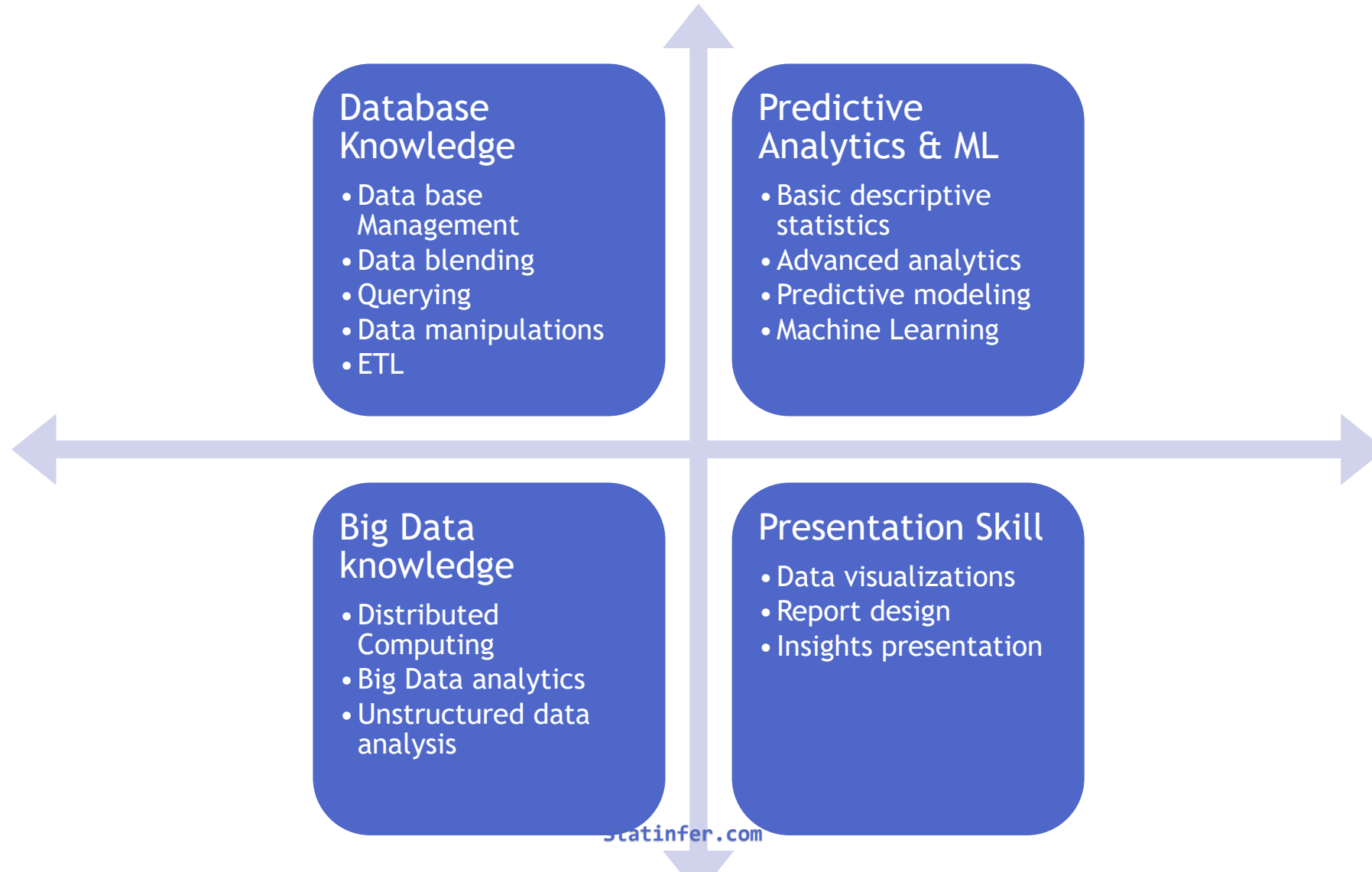
What is Data Science

- Data Science is a Fusion of many fields
 - Database management
 - Data Analytics
 - Predictive modelling
 - Machine Learning
 - Big data Distributed computing
 - Coding
 - Data visualizations
 - Reporting

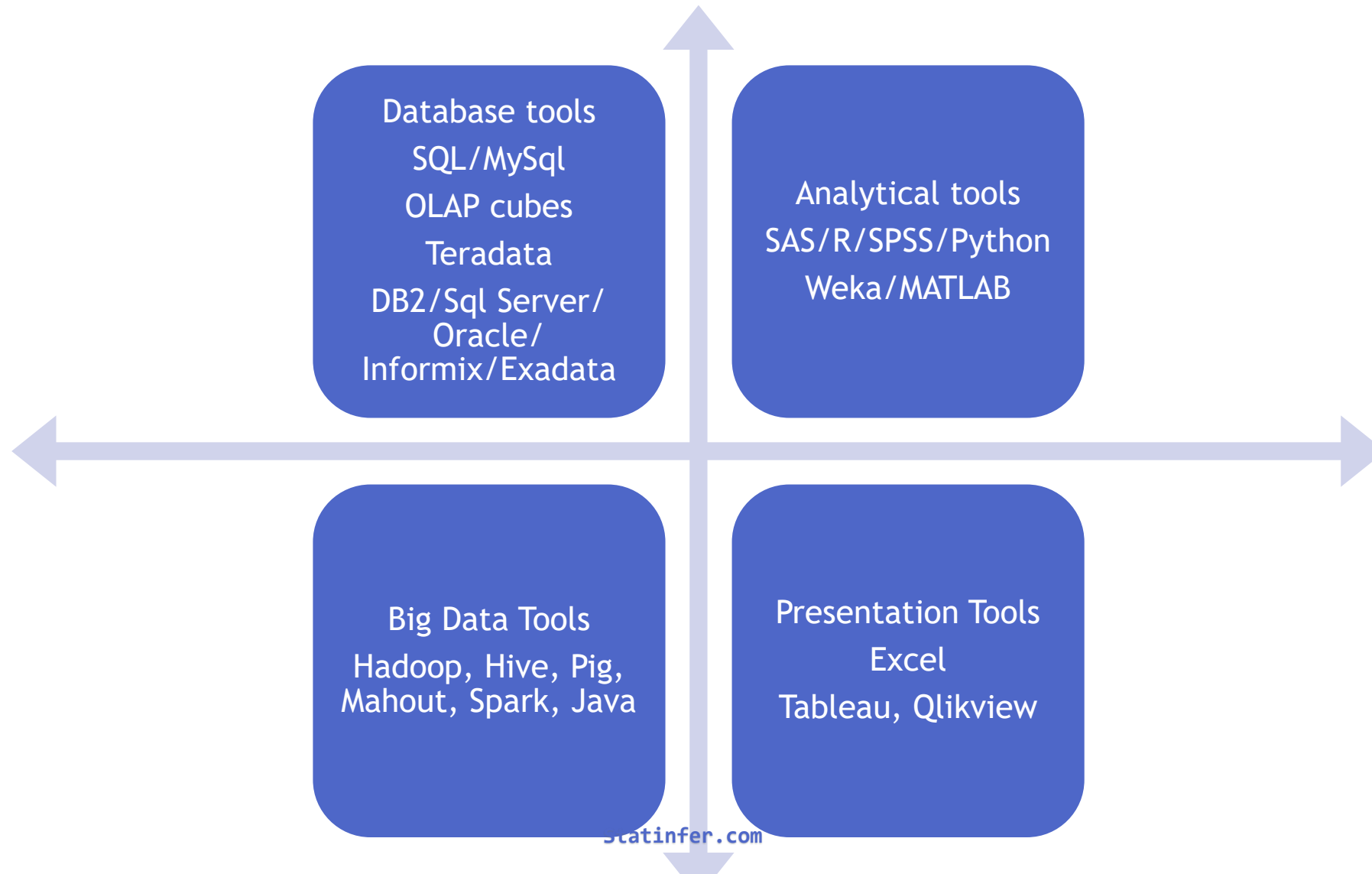
Data Science – Four Major Type of Skills



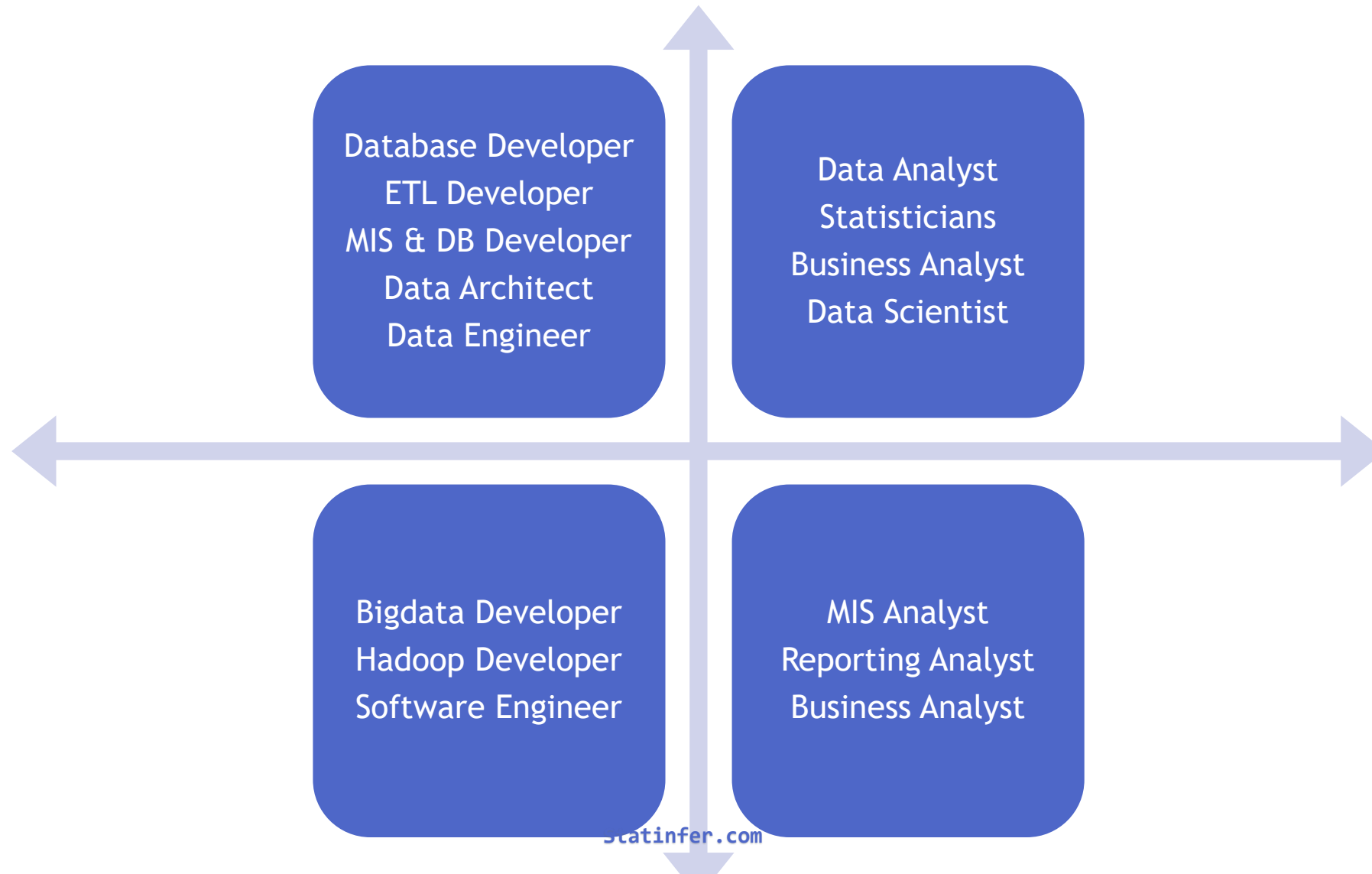
The Techniques you need to know



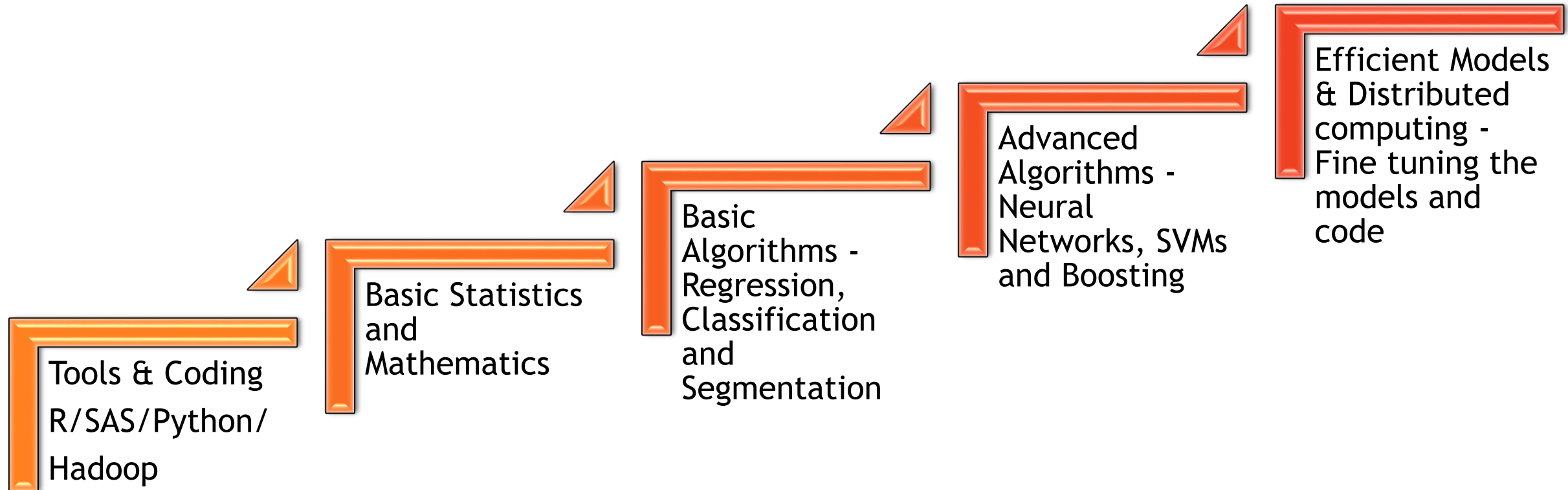
Data Science tools and software's



Data Science -Designations



Data Science Learning Path



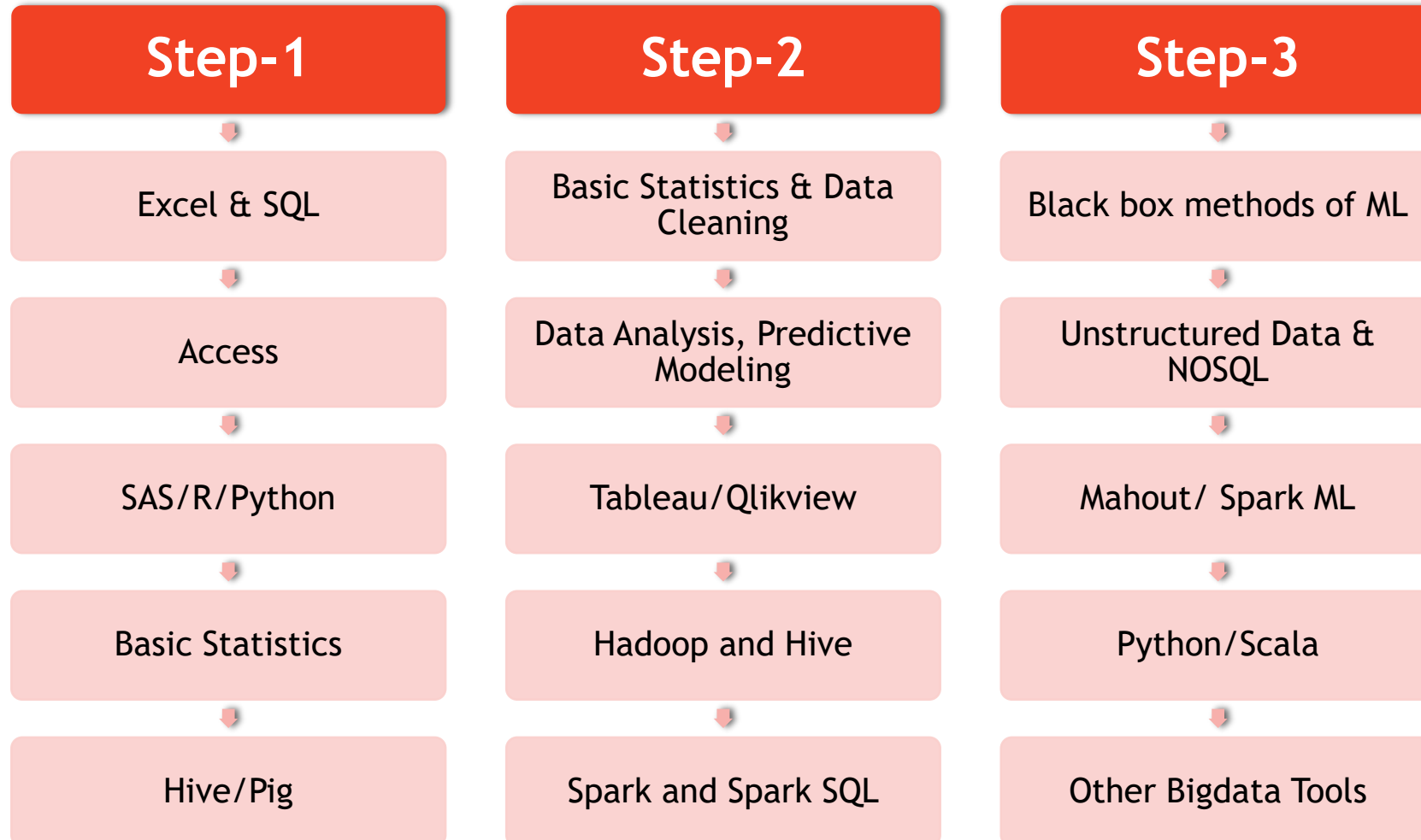


What training should I take

FAQ by Data Science Aspirants

- I want to be data scientist what training should I take?
- I already have knowledge on few tools, what are my next steps?
- What skill should I add to my profile to make it to next level?
- I am new to data science, where can I start ?

Structured Learning Path to be a successful Data Scientist



Three Major Categories of Profiles

- You need training based on your skill level
 - Based on skill set we can divide the whole data science aspirants into four categories
1. Beginner - Completely new to Data Analytics
 2. Intermediate - MIS and Reporting Analyst
 3. Advanced - Data Analyst and Predictive Modeler
 4. Complete Data Scientist - ML, Hadoop, R, Python, Spark

If you are a beginner

Your Characteristics

1. No hands-on experience on Data Analytics projects
2. No hands-on experience on Data Analytics tools like SAS & R
3. No hands-on experience on databases and SQL
4. No hands-on experience on excel
5. No idea on automation
6. No idea on data exploration and validation
7. Never worked on reporting projects
8. Very limited relevant business experience
9. Want to get started with Analytics / Data Science

Your Training needs

- Get trained on one topic at least from each of these categories
 - 1
 - Excel
 - Sql
 - 2
 - SAS programming
 - R programming
 - SPSS programming
 - Python
 - 3
 - MS Access
 - Excel VBA
 - Hive
 - Pig
 - 4
 - Basics of Statistics & Data Cleaning

Suitable job profiles for beginner

1. Reporting Analyst
2. Business Analyst
3. Data Analyst
4. Analyst Associate
5. Junior data scientist
6. Hadoop Developer
7. Tableau Developer
8. Database Developer
9. ETL Developer
10. MIS Analyst

If you are a MIS and Reporting Analyst

Your Characteristics

1. Knows excel
2. Knows SQL
3. Knows Basics of SAS / R Coding
4. Relevant business experience & good domain knowledge
5. Experience in data exploration, validation and cleaning
6. Worked on lot of MIS and reporting projects, created lot of dash-boards
7. Limited experience in predictive modeling techniques like logistic regression, decision trees, time series
8. Worked on few automation projects
9. Want to get into core analytics, predictive modeling and model building

Your Training needs

- Get trained on one topic at least from each of these categories
 - 1
 - Basic Descriptive Statistics & Data Cleaning
 - 2
 - Predictive modeling techniques
 - 3
 - Visualizations using Tableau
 - Visualizations using Qlikview
 - 4
 - Hadoop Introduction & Hive
 - Spark Introduction & Spark Sql

Suitable job profiles for MIS and Reporting Analyst

1. Data Analyst
2. Business Analyst
3. Junior data scientist
4. Data Scientist
5. Hadoop Developer
6. Predictive Modeler

Course Contents

1. R programming and Basic Stats
2. Predictive modelling in R
3. Machine Learning in R
4. Machine Learning Projects
5. Bigdata & Hadoop
6. Hive and Pig
7. Python Programming
8. Machine Learning on Python
9. Machine Learning Projects in Python
10. Bigdata Projects
11. Spark and Scala(Optional)
12. Tableau (Optional)

If you are a Data Analyst

Your Characteristics

1. Knows SAS/R/Python
2. Already worked on few analytics projects
3. Already worked on predictive modeling techniques like logistic regression, decision trees, time series
4. Have very good business knowledge
5. Never worked on bigdata
6. No hands on experience on Hadoop, Hive or Spark
7. Very limited experience in machine learning techniques like neural networks, svm, bagging and boosting
8. Want to get a clear idea on end to end predictive modeling
9. Want to get into data science by learning big data and machine learning

Your Training needs

- Get trained on one topic at least from each of these three categories
 - 1
 - Black box methods of ML
 - Ensemble models
 - 2
 - Mahout
 - Spark ML
 - 3
 - Python
 - Scala

Suitable job profiles for Data Analyst

1. Data Scientist



Thank you
