



Introduction to Data Science

Khoá học: Data Science Essentials



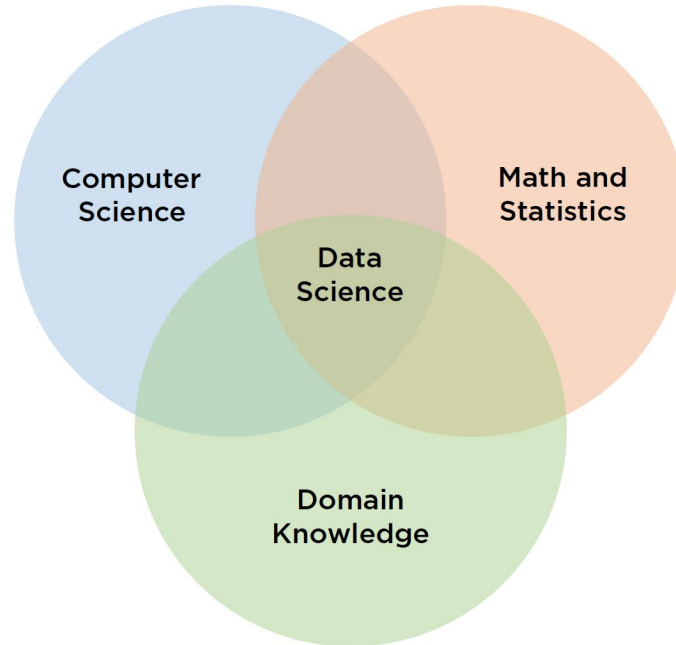
Mục tiêu

- Trình bày được khái niệm Data Science
- Trình bày được khái niệm Data Analytic Thinking
- Trình bày được Data Science Process
- Biết được các công cụ thông dụng hỗ trợ Data Science
- Sử dụng được Azure Learning Studio
- Tạo được Machine Learning Model



What is Data Science?

Data Science is the exploration and quantitative analysis of all available structured and unstructured data to develop understanding, extract knowledge, and formulate actionable results.



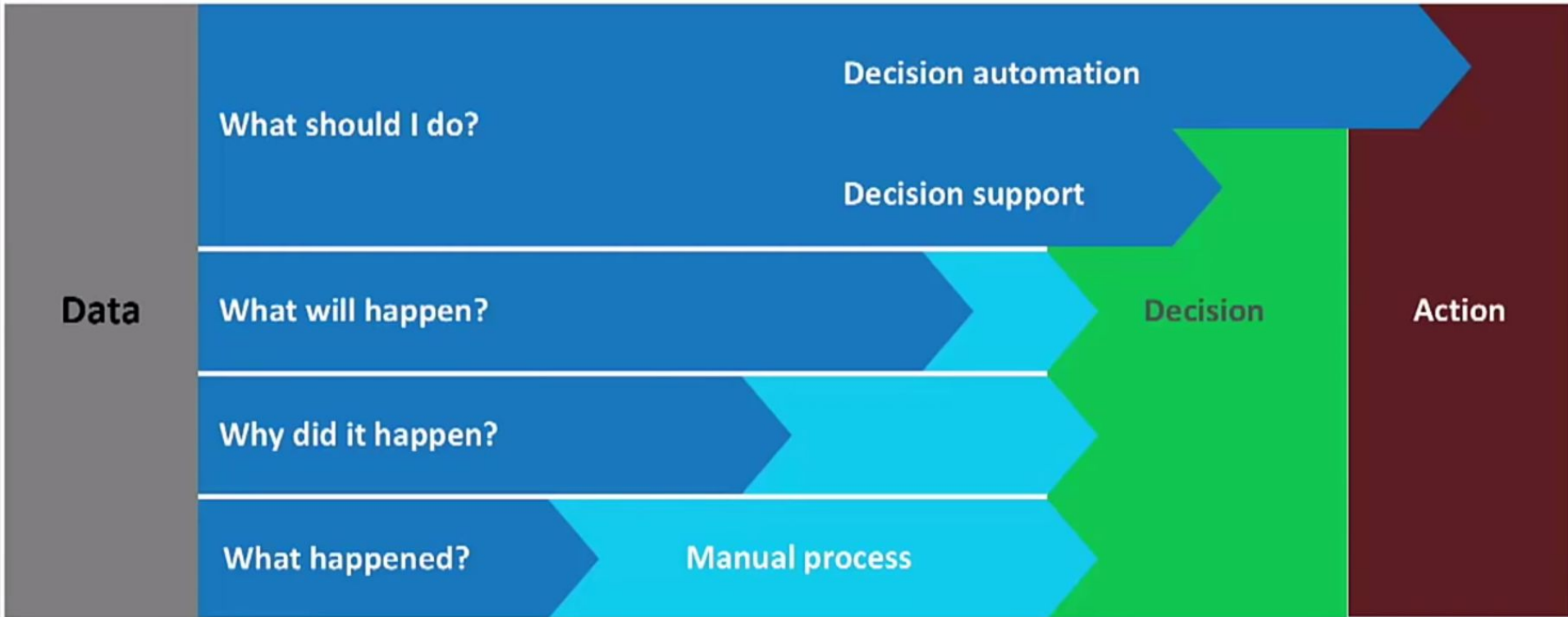


Data Analytic Thinking

- Nghĩ về dữ liệu, biết được tầm quan trọng của dữ liệu
- Thay đổi các việc làm hằng ngày
- Thay đổi cách làm việc dựa trên thói quen bằng cách làm việc dựa trên dữ liệu được phân tích
- Dữ liệu ngày càng nhiều, từ nhiều nguồn khác nhau
- Cần tổ chức tốt, phân tích tốt và sử dụng các công cụ để trích xuất và sử dụng tri thức để ra các quyết định và hành động
- Tăng tốc các quyết định và hành động hằng ngày



Data → Decisions → Actions



Type of Analytics

Retrospective analytics



Predictive analytics



Real-time analytics



Intelligent SaaS apps





Ví dụ ứng dụng phân tích dữ liệu

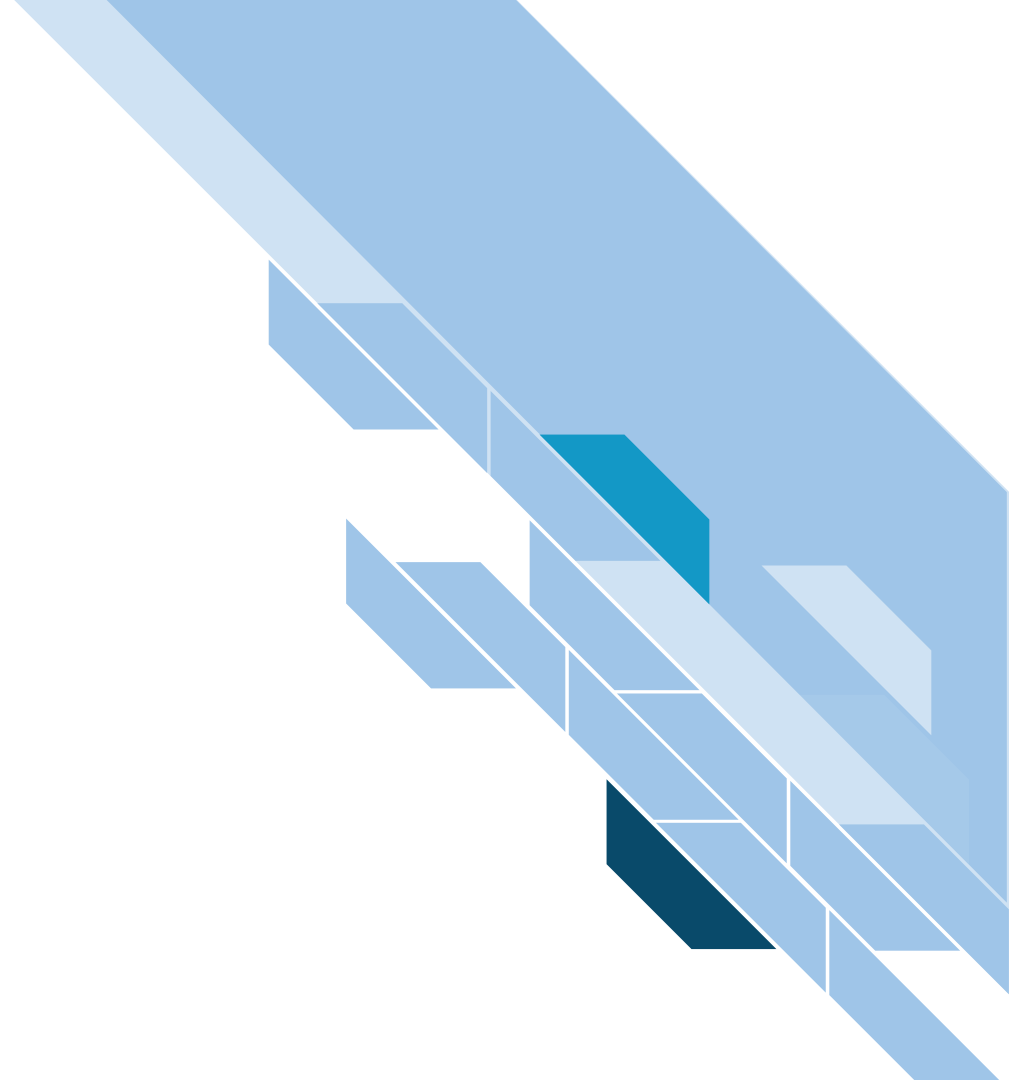
- Phát hiện gian lận trong các thanh toán bằng tài khoản ngân hàng
- Hỗ trợ điều trị y tế
- Bảo trì dự phòng (preventive maintenance)
 - Máy bay
 - Thang máy
- Dự đoán nhu cầu thuê xe đạp của một chuỗi cửa hàng
- Dự đoán nhu cầu sử dụng điện để bảo trì máy phát điện







The Data Science Process





Các thao tác trong Data Science

- Finding data
- Acquiring data
- Cleaning and transforming data
- Understanding relationships in data
- Delivering value from data

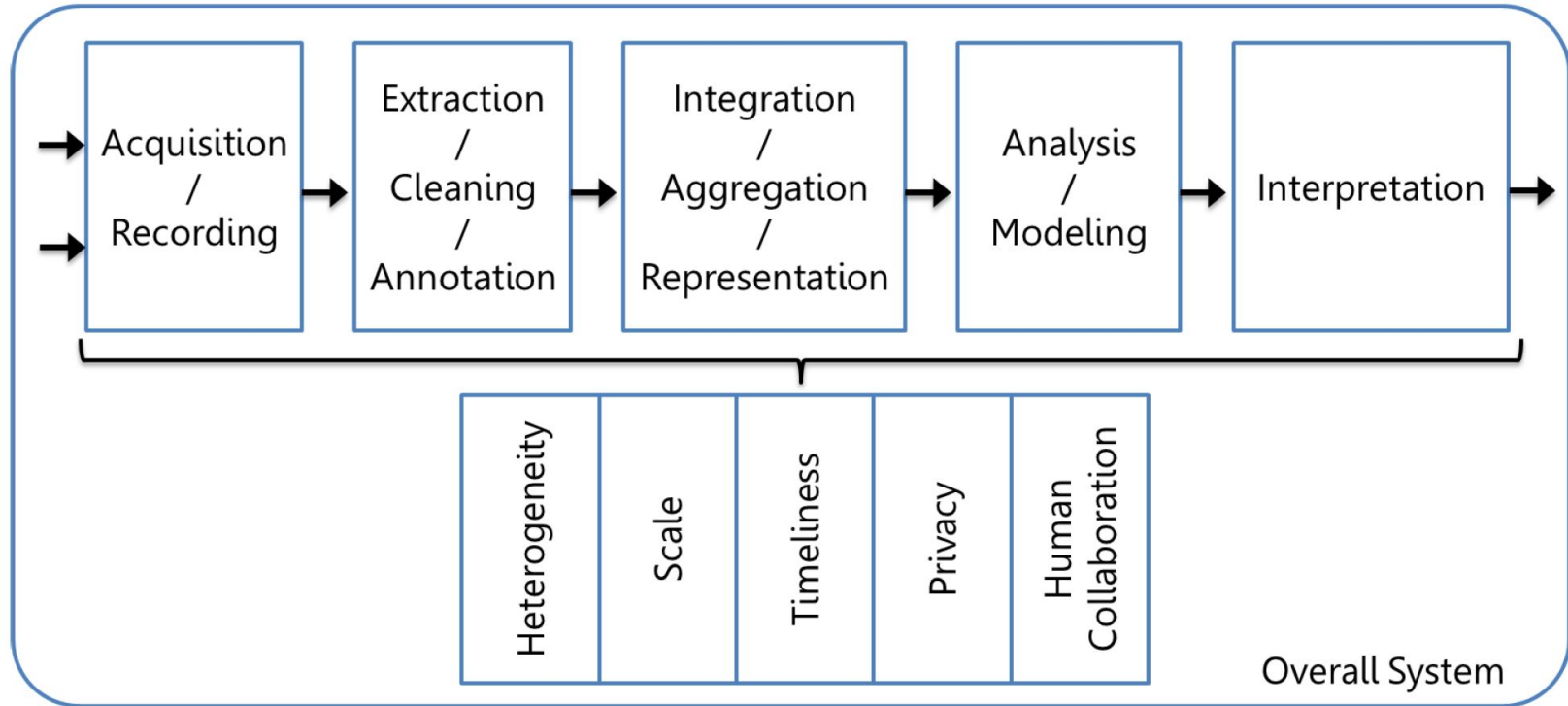


Các hoạt động trong một quy trình Data Science

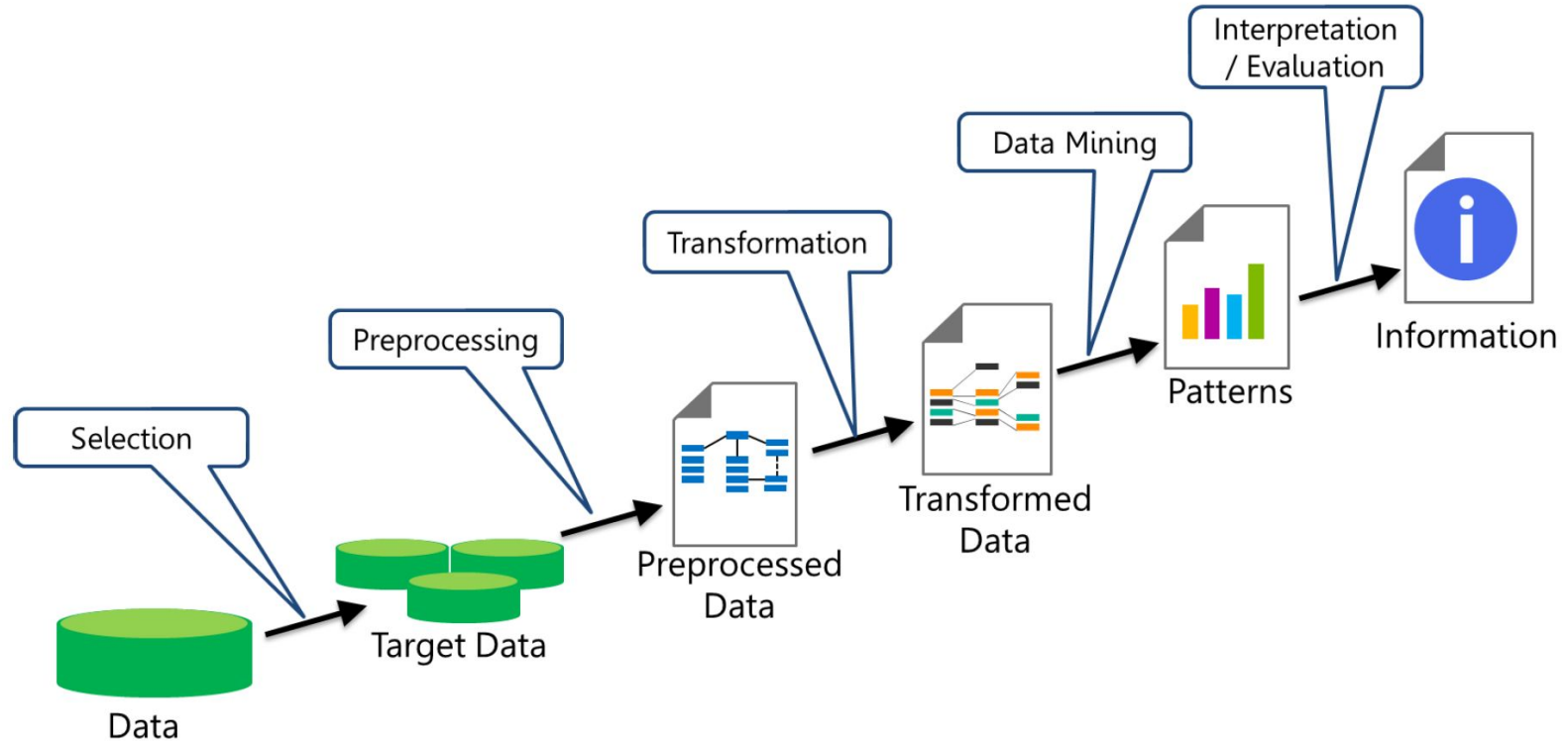
1. Chọn lọc dữ liệu
2. Tiền xử lý dữ liệu
3. Chuyển đổi dữ liệu
4. Khai phá dữ liệu
5. Diễn dịch và đánh giá



The Computing Community Consortium Big Data Whitepaper (2012)

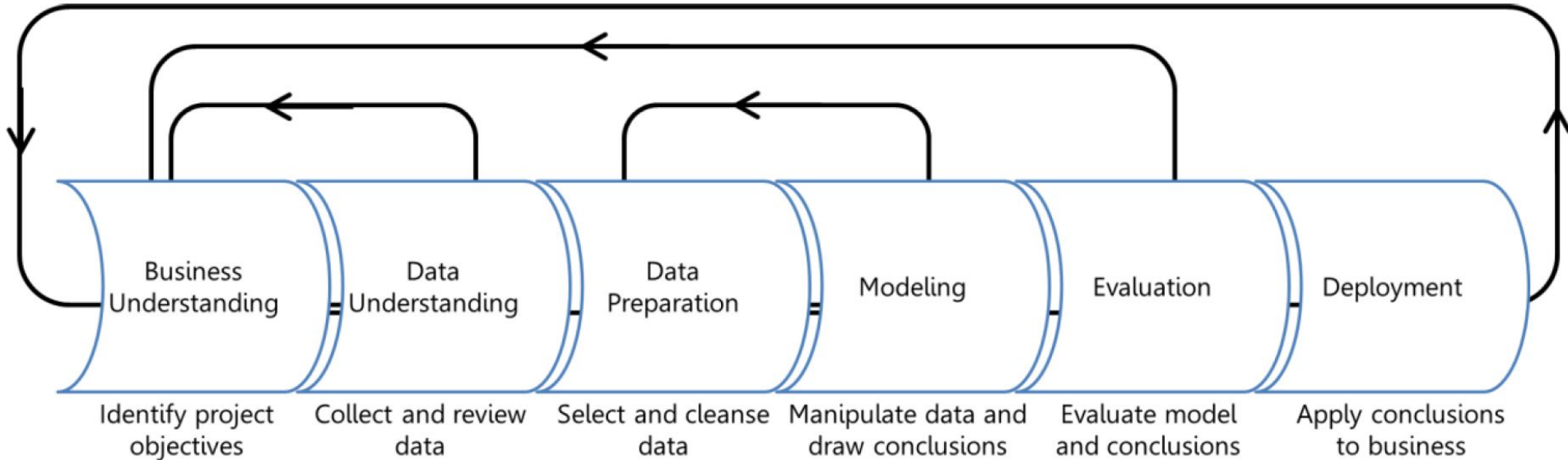


The Knowledge Discovery in Databases (KDD) process (1997)





The Cross Industry Standard Process for Data Mining (CRISP-DM) (2000)

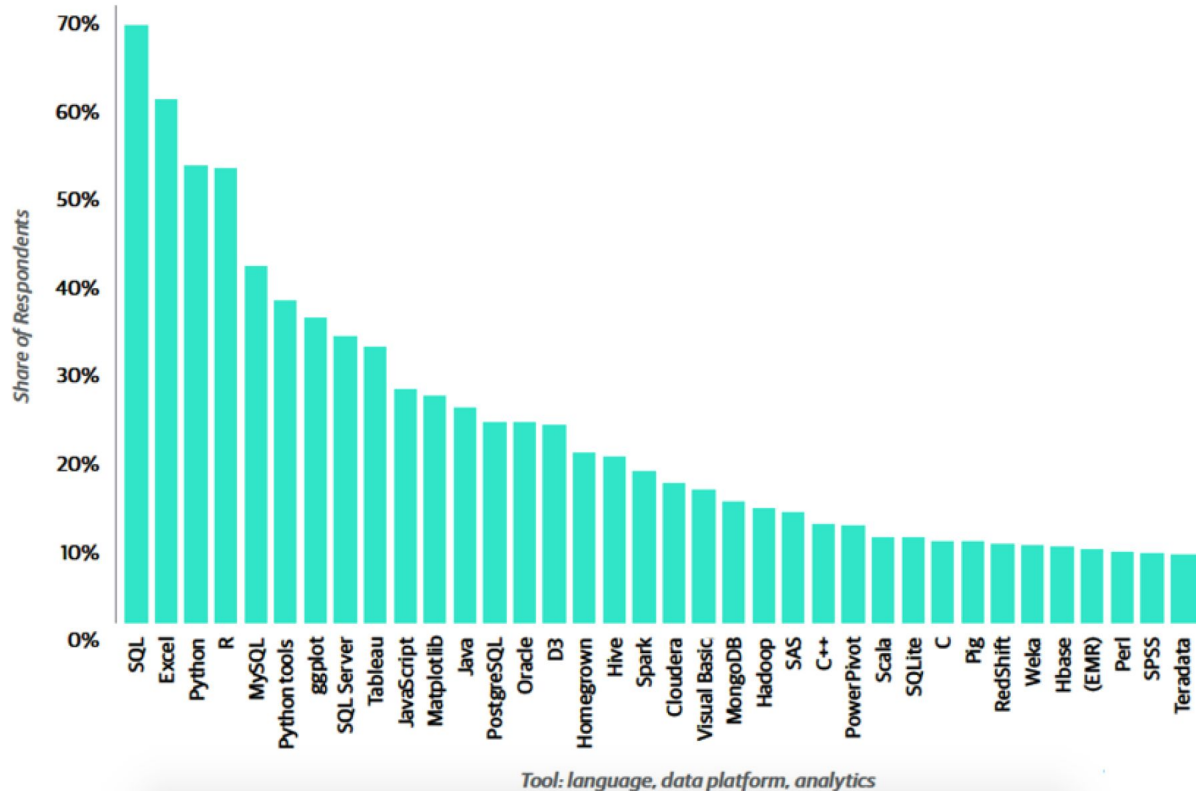




Introduction to Data Science Technologies

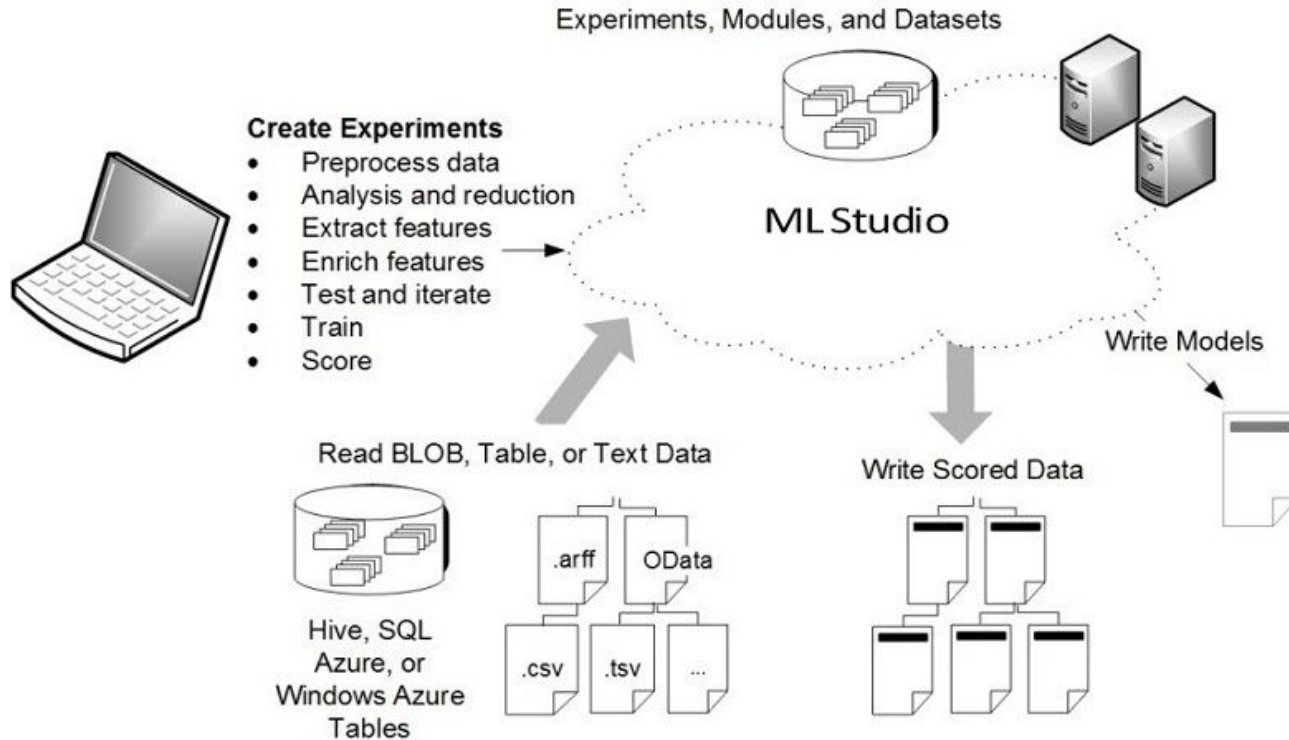


Tools for Data Science










Source: O'Reilly 2015 Data Science Salary Survey

Azure Machine Learning Studio



Azure Machine Learning Studio

Microsoft Azure Machine Learning Studio

-  PROJECTS
-  EXPERIMENTS
-  WEB SERVICES
-  NOTEBOOKS
-  DATASETS
-  TRAINED MODELS
-  SETTINGS

projects **preview**

NAME	AUTH
------	------

Create project





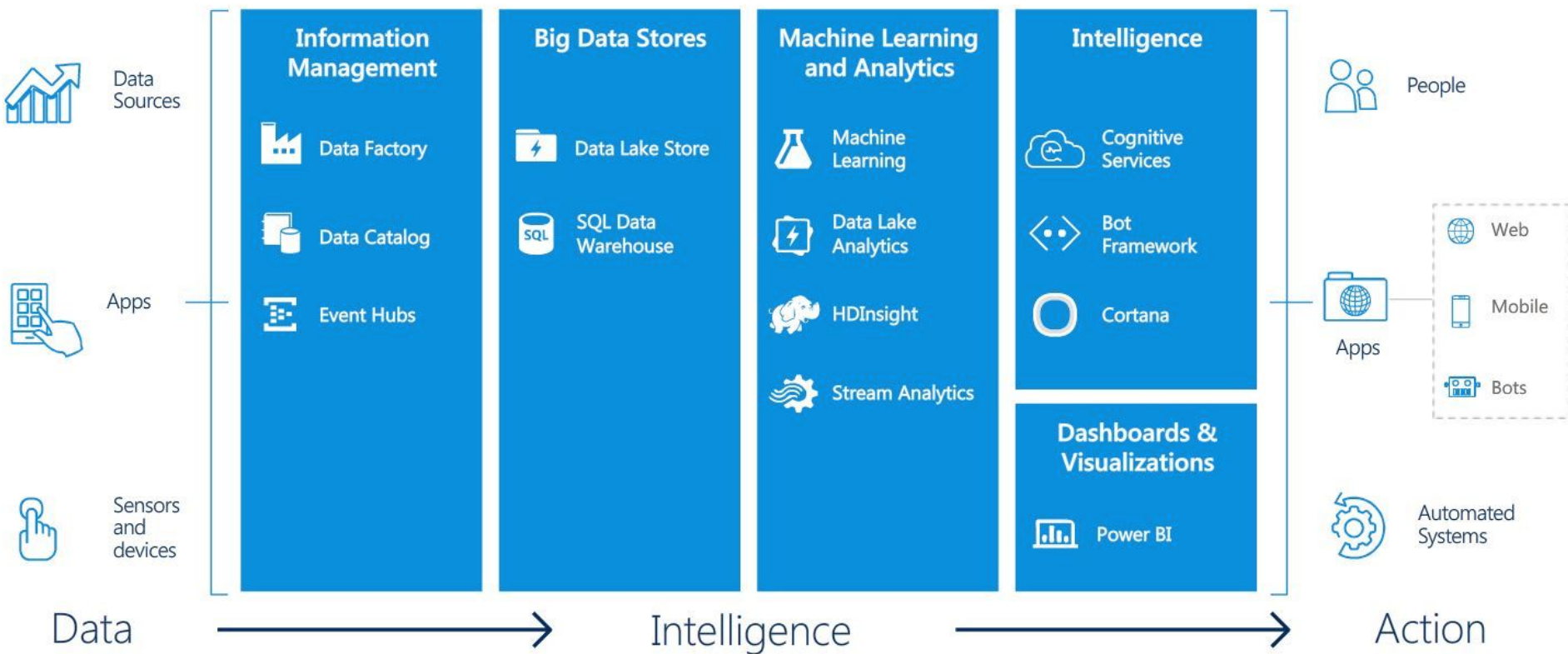
Jupyter Notebook

- Open-source web application
- Create and share documents
 - Live code
 - Equations
 - Visualizations
 - Narrative text
- Uses include
 - Data cleaning
 - Data transformation
 - Numerical simulation
 - Statistical modeling
 - Data visualization
 - Machine learning
 - ...



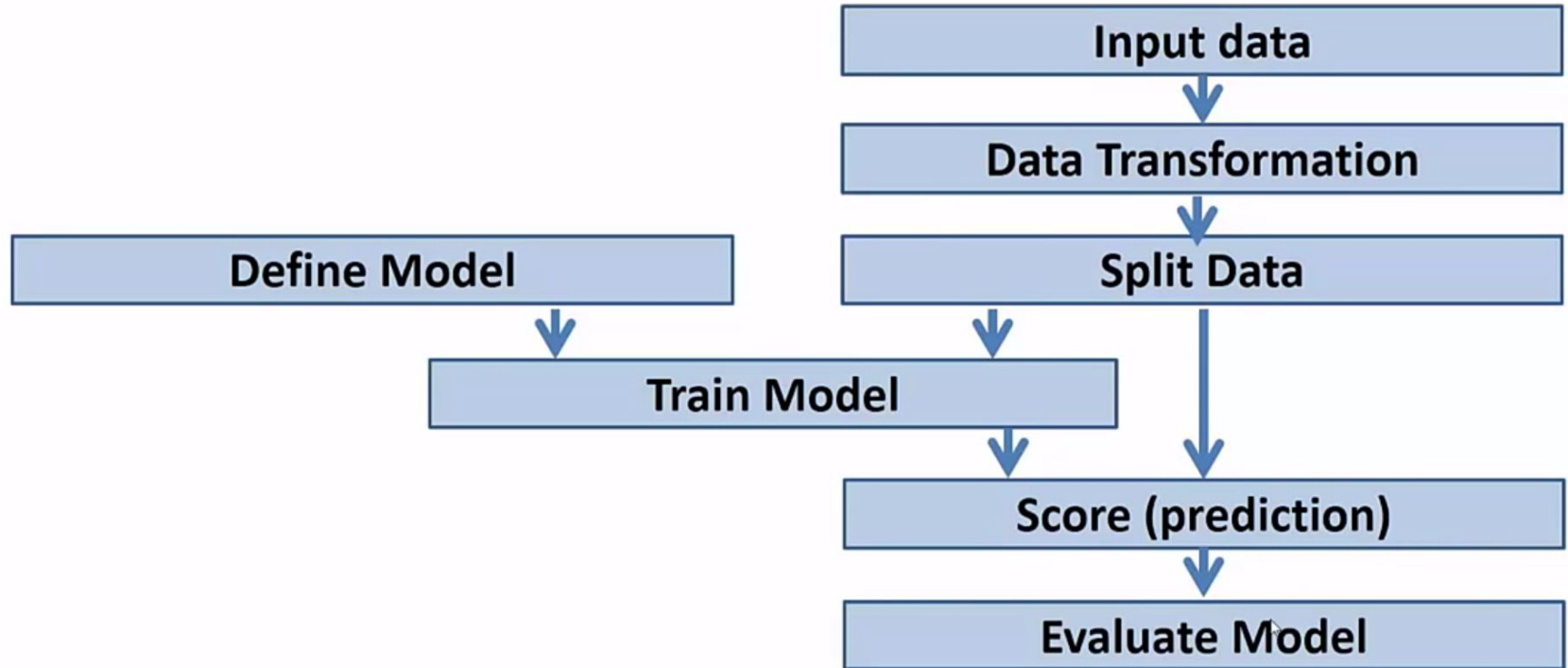
Cortana Intelligence Suite

Transform data into intelligent action





Machine Learning Workflow





Các module thông dụng

- Data Input and Output
- Data Format
- Conversions
- Data Transformation
- Feature Selection
- Machine Learning
- Python Language
- R Language
- Statistical Functions
- Text Analytics
- Time Series
- Web Services