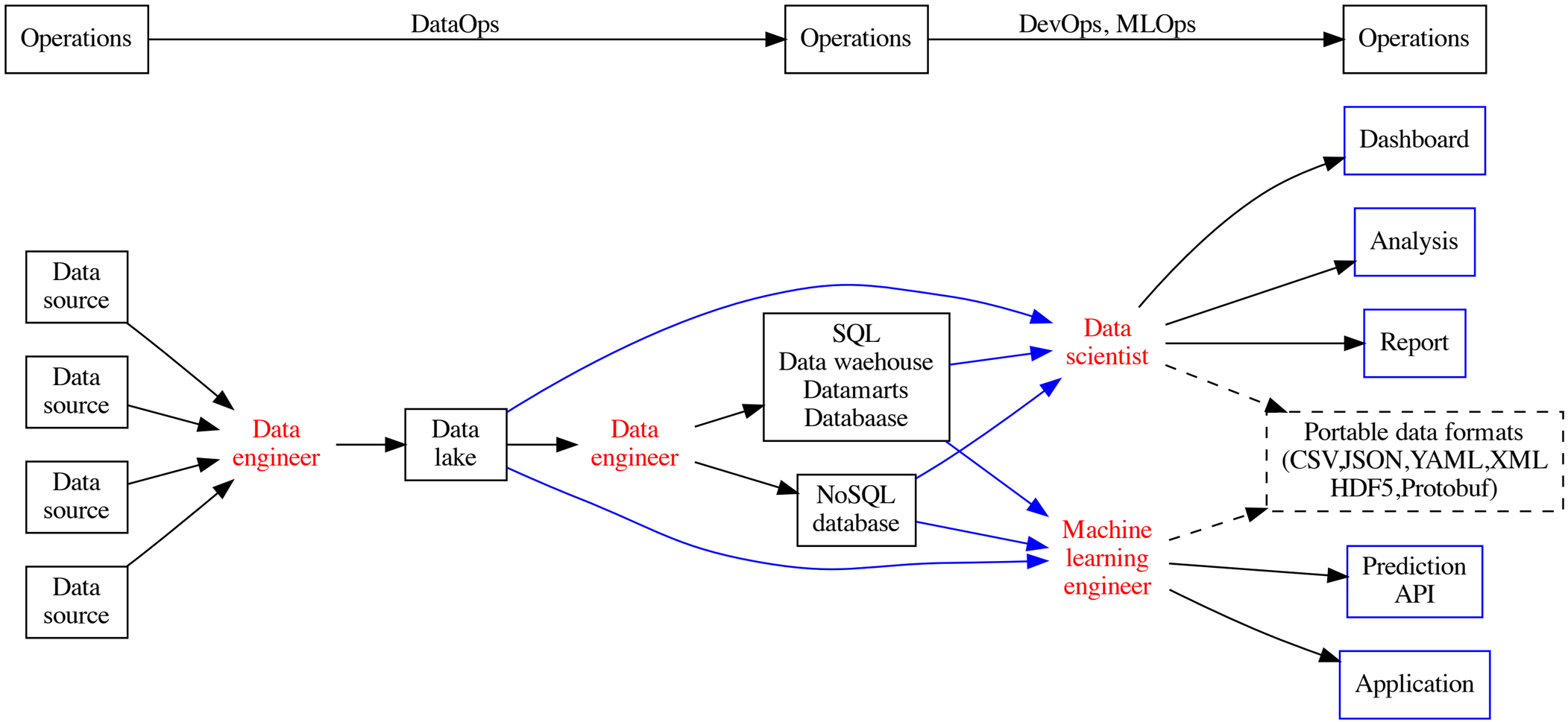
Review of learning objectives

What do data scientists do?


Operations
DataOps
Operations
DevOps, MLOps
Operations
Data source
Data source
Data source
Data source
Data engineer
Data lake
Data engineer
SQL
Data waehouse
Datamarts
Databaase
NoSQL database
Data scientist
Machine learning engineer
Dashboard
Analysis
Report
Portable data formats
(CSV,JSON,YAML,XML
HDF5,Protobuf)
Prediction API
Application

Foundational skill sets

Domain knowledge

- Asking the right questions
- Framing problems
- Collaborating with stakeholders

Data engineering

- Establish and maintain data lake
- Database construction
- ETL operations

Data analysis

- Wrangling
- Analysis
- Visualization

Machine learning

- Building
- Training
- Evaluating
- Logging
- Deployment

Operations

- Automation
- CI/CD
- Continuous delivery
- Scaling
- Monitoring

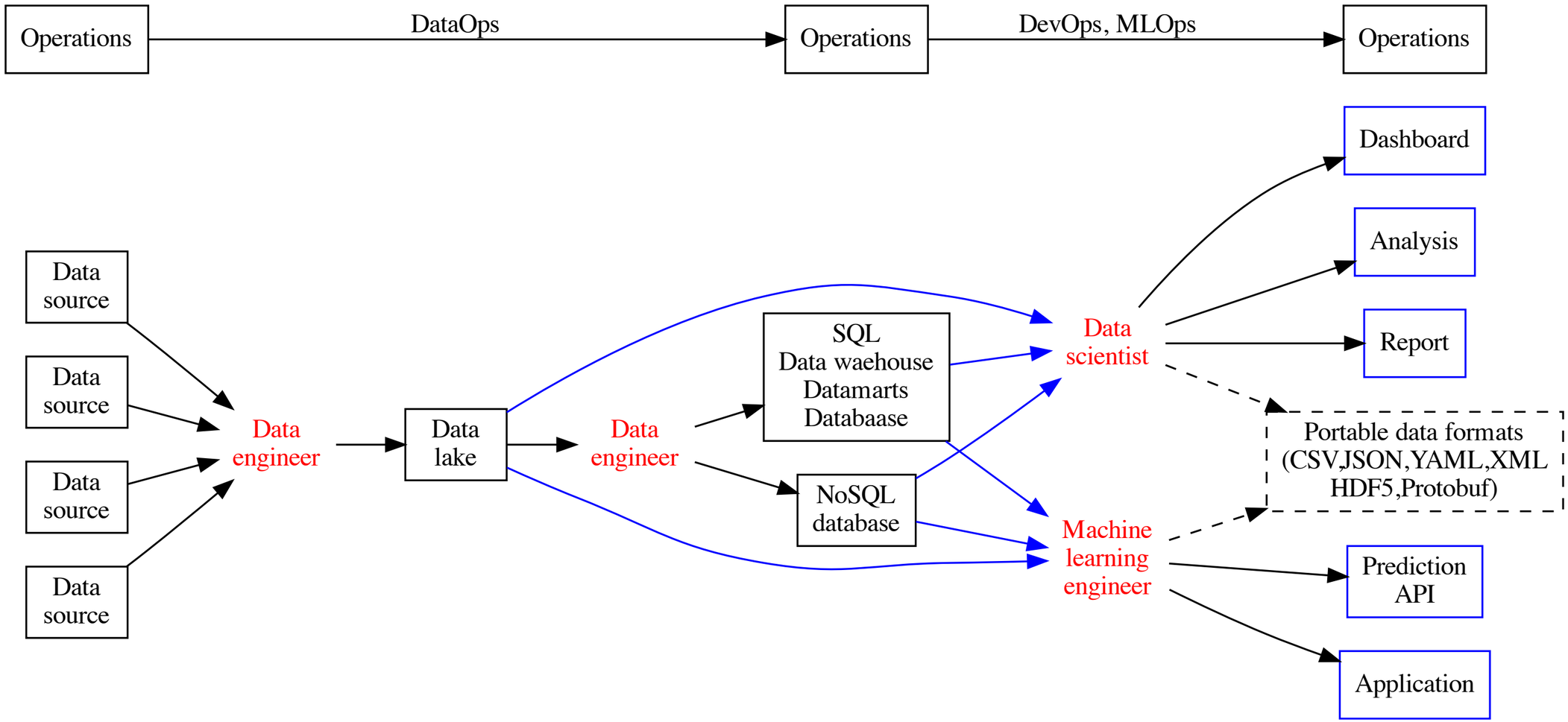
Organization

Core

- Data engineer
- Data scientist
- ML engineer
- MLOps / DevOps

Ancillary

- Domain expert
- Statistician
- ML researcher
- Software engineer


Operations
DataOps
Operations
DevOps, MLOps
Operations
Data source
Data source
Data source
Data source
Data engineer
Data lake
Data engineer
SQL
Data waehouse
Datamarts
Databaase
NoSQL database
Data scientist
Machine learning engineer
Dashboard
Analysis
Report
Portable data formats
(CSV,JSON,YAML,XML
HDF5,Protobuf)
Prediction API
Application

Milestones

Domain knowledge

- What data scientist do
- How to teach yourself

Development practices

- Jupyter
- Markdown
- Literate programming
- Git and version control

Programming

- Fluent Python
- numpy
- pandas

Professional development

- Jekyll
- Hugo / Pelican
- GitHub-pages
- Data science portfolio

Remote learning

- Ask questions on chat
- Ask questions in-person
- Interacting with peers on Slack or Teams
- Present your blogs to the class

What's next

Continue to understand data

- Data standards
- storage
- sharing and serving
- configuration
- Data formats

Data ecosystem and DataOps

- Data lake
- Data warehouse
- Data mart
- Working databases

Databases

- Relational databases and SQL
- NoSQL databases

Course structure

Lectures

- Highlight tools and knowledge
- Just a guide
- Dig further
- Practice on your own

Quizzes

- Practice basic skills
- Essential for the exam

Exam

- Closed book
- Time constrained
- Simulate skills interview

Homework assignments

- Encourage you to try out other tools
- Assemble your data science portfolio

Final project

- Messy data challenge
- Simulates data science team
- Capstone - integrates multiple skill groups