Overview

- Introduction
- Target
- Requirements
- Development
- Deployment

www.big-data-europe.eu
4 Vs of Big Data

- **Volume**: Quantity of data
- **Velocity**: Speed at which data is provided
- **Variety**: Different formats/models in which data is provided
- **Veracity**: Accuracy/truthfulness of the data

www.big-data-europe.eu
Lambda Architecture
• Answer some queries directly
• Perform batch analytics
• Programmer is part of the game
Pipeline requirements

- Questions to be answered
- # queries & variety
- # data

www.big-data-europe.eu
Developing a pipeline
Developing a pipeline
Running a pipeline

- Deploying
- Monitoring
- Logging

www.big-data-europe.eu
More information

- **Big Data Integrator Deployment & Component Interface Specification**
  http://www.big-data-europe.eu/wp-content/uploads/D3.3-
  Big_Data_Integrator_Deployment_and_Component_Interface_Specification.pdf

- **Big Data: Principles and best practices of scalable real-time data systems**
  by Nathan Marz and James Warren

- **Code repository**
  https://github.com/big-data-europe