



BIG DATA EUROPE

Empowering Communities
with Data Technologies



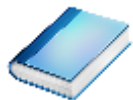
BDE-HEALTH WEBINAR

25-nov-2015

Technical overview



Overview



Introduction



Target



Requirements



Development



Deployment



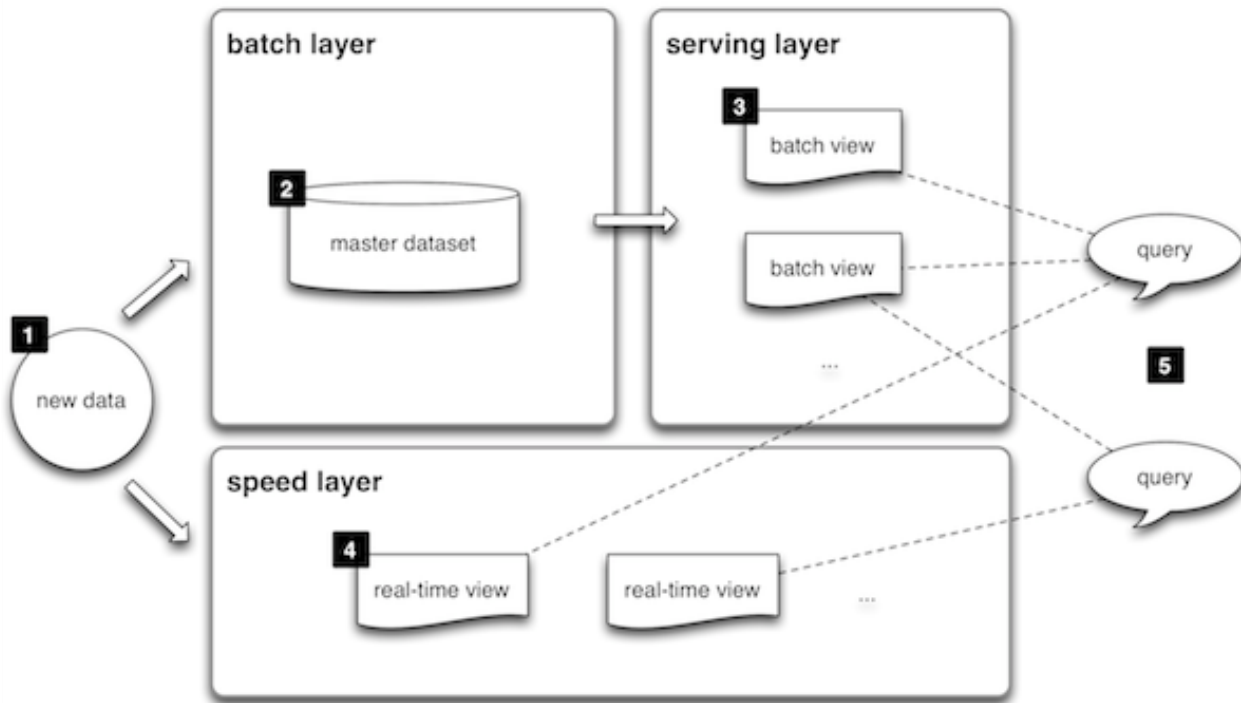
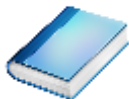
4 Vs of Big Data



- ① **Volume**
Quantity of data
- ① **Velocity**
Speed at which data is provided
- ① **Variety**
Different formats/models in which data is provided
- ① **Veracity**
Accuracy/truthfulness of the data



Lambda Architecture





- Answer some queries directly
- Perform batch analytics
- Programmer is part of the game



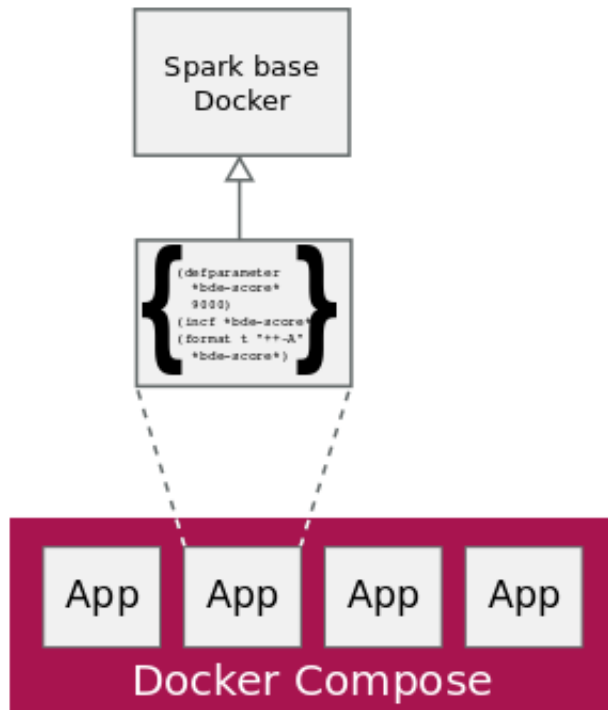
Pipeline requirements



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

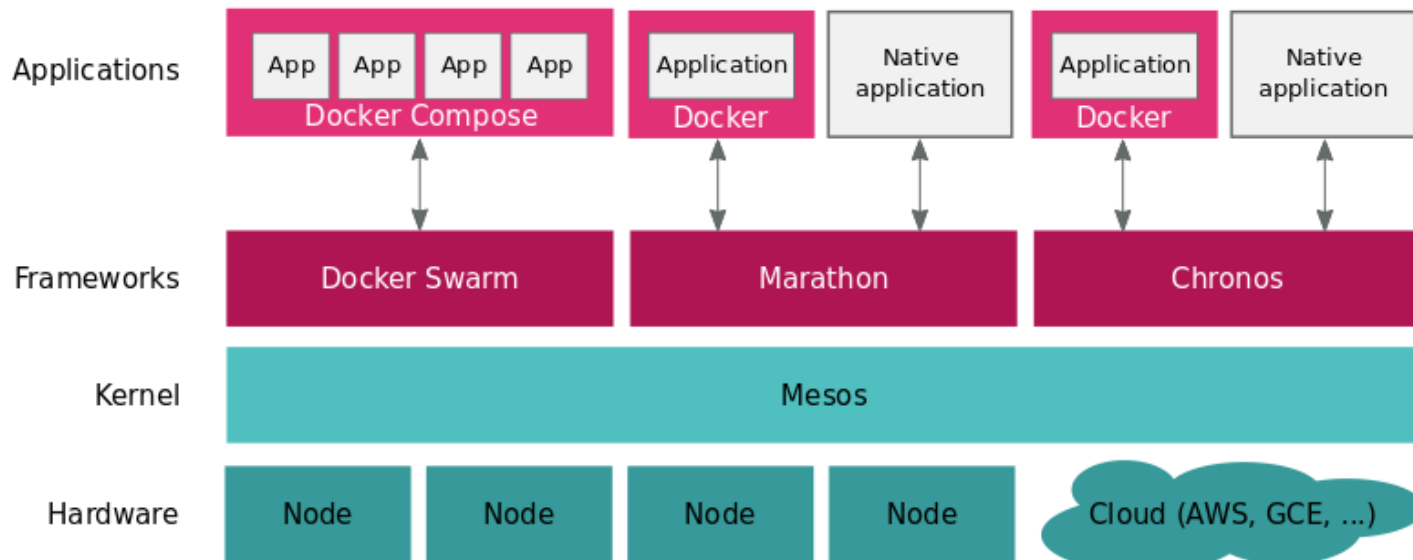


Developing a pipeline





Developing a pipeline





Running a pipeline



- Deploying
- Monitoring
- Logging



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>