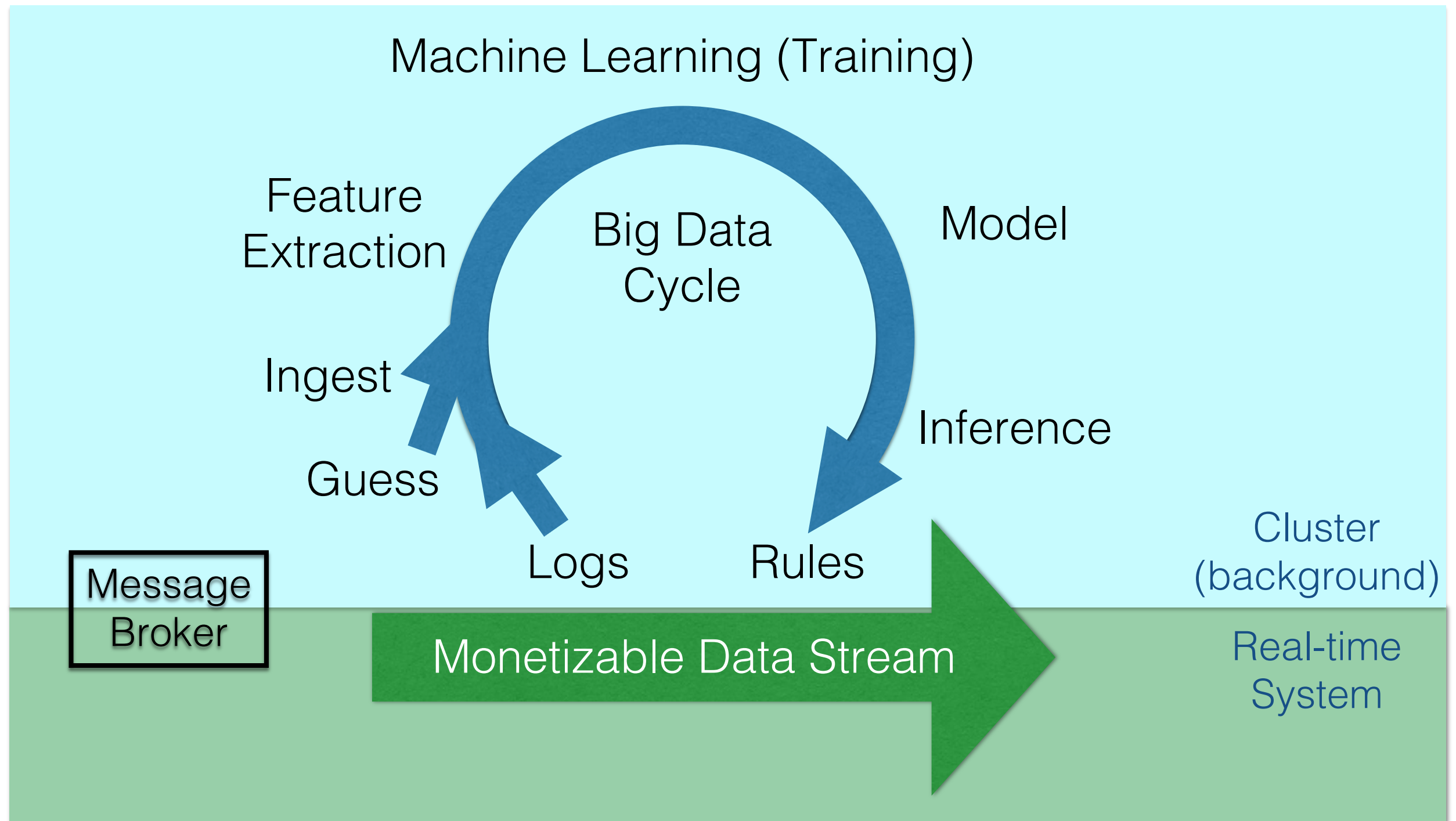


2.8 Big Data

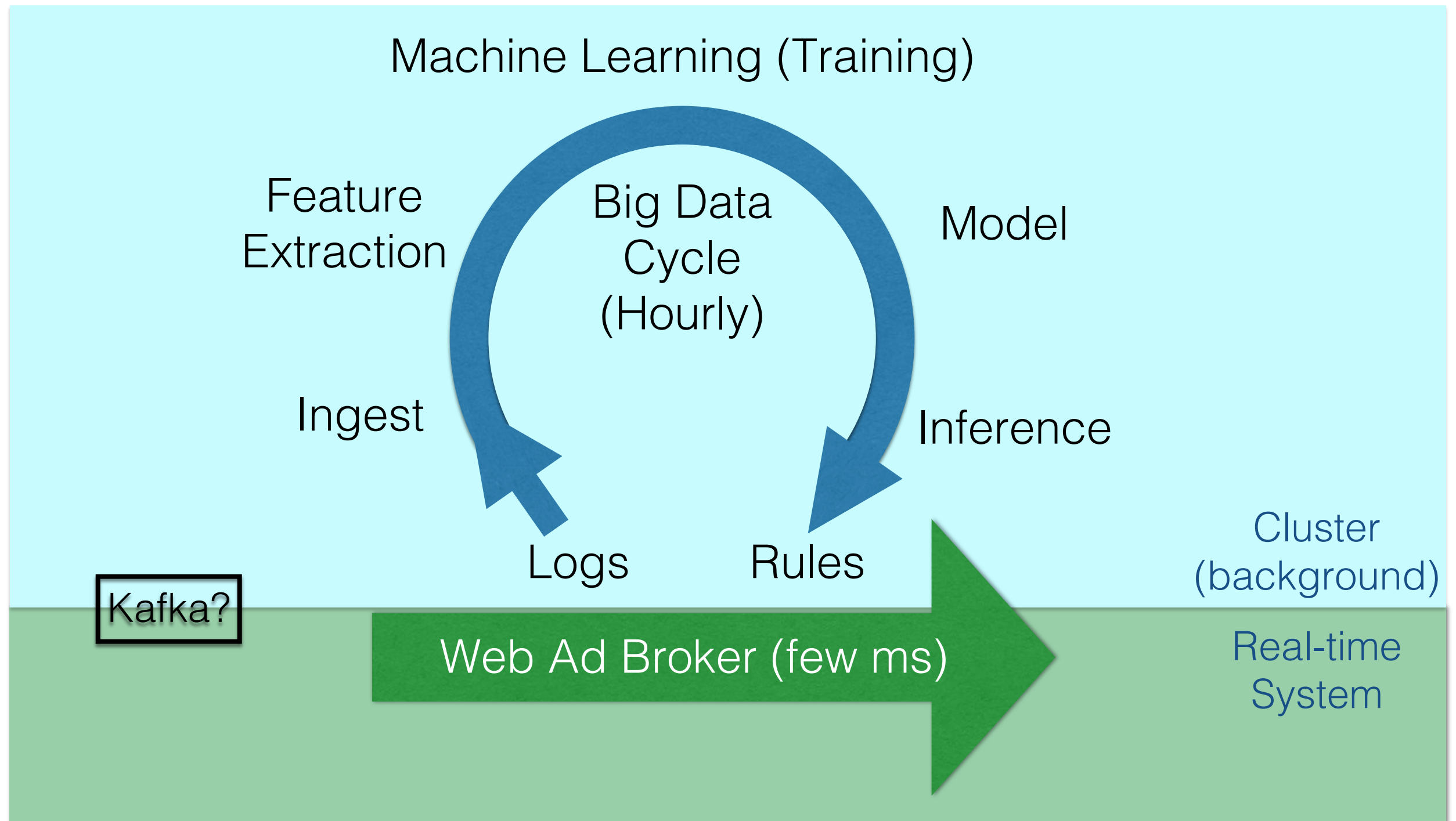
What is it all about?

Copyright © Robin Hillyard, 2017

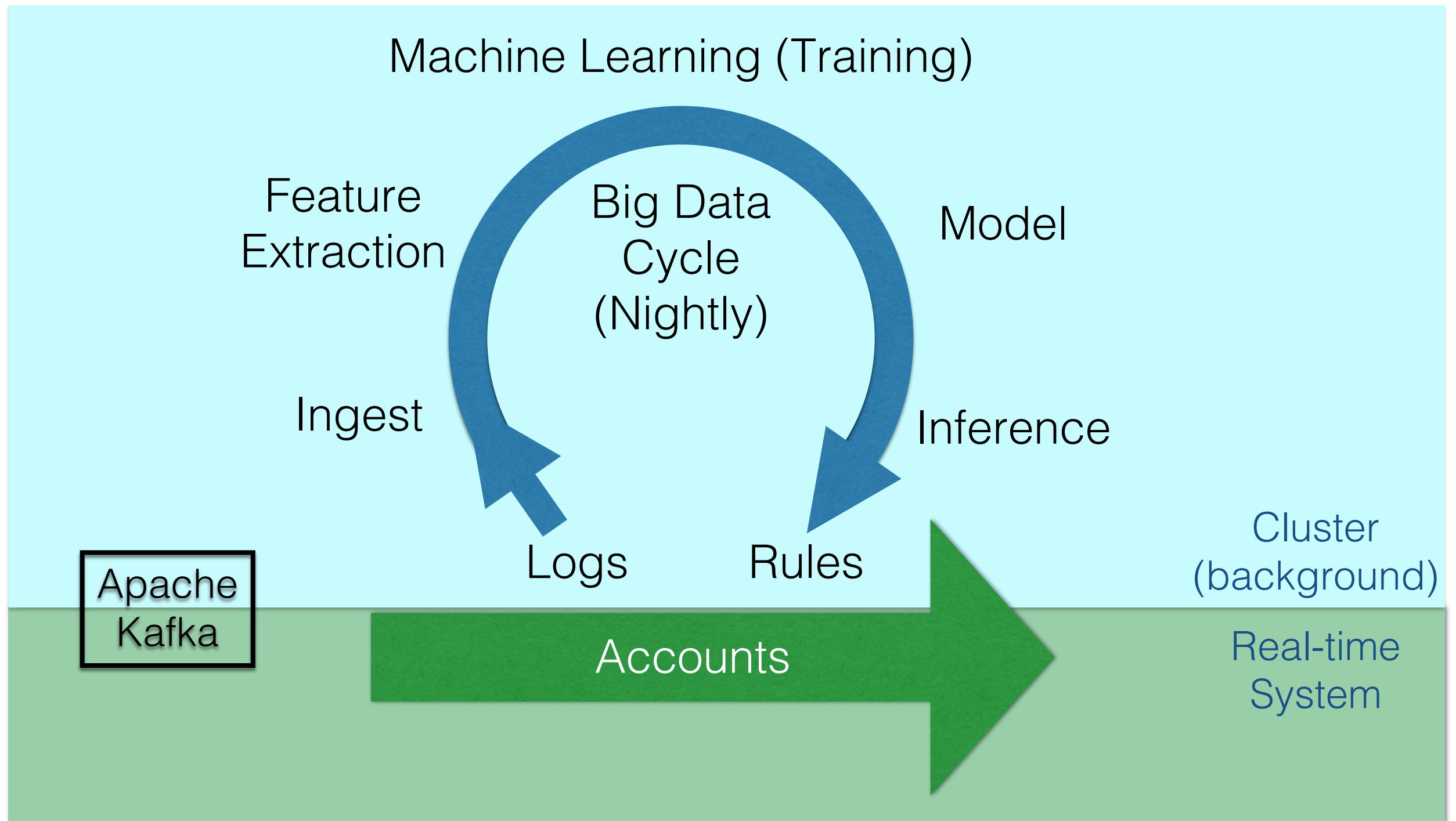
Big Data information cycle



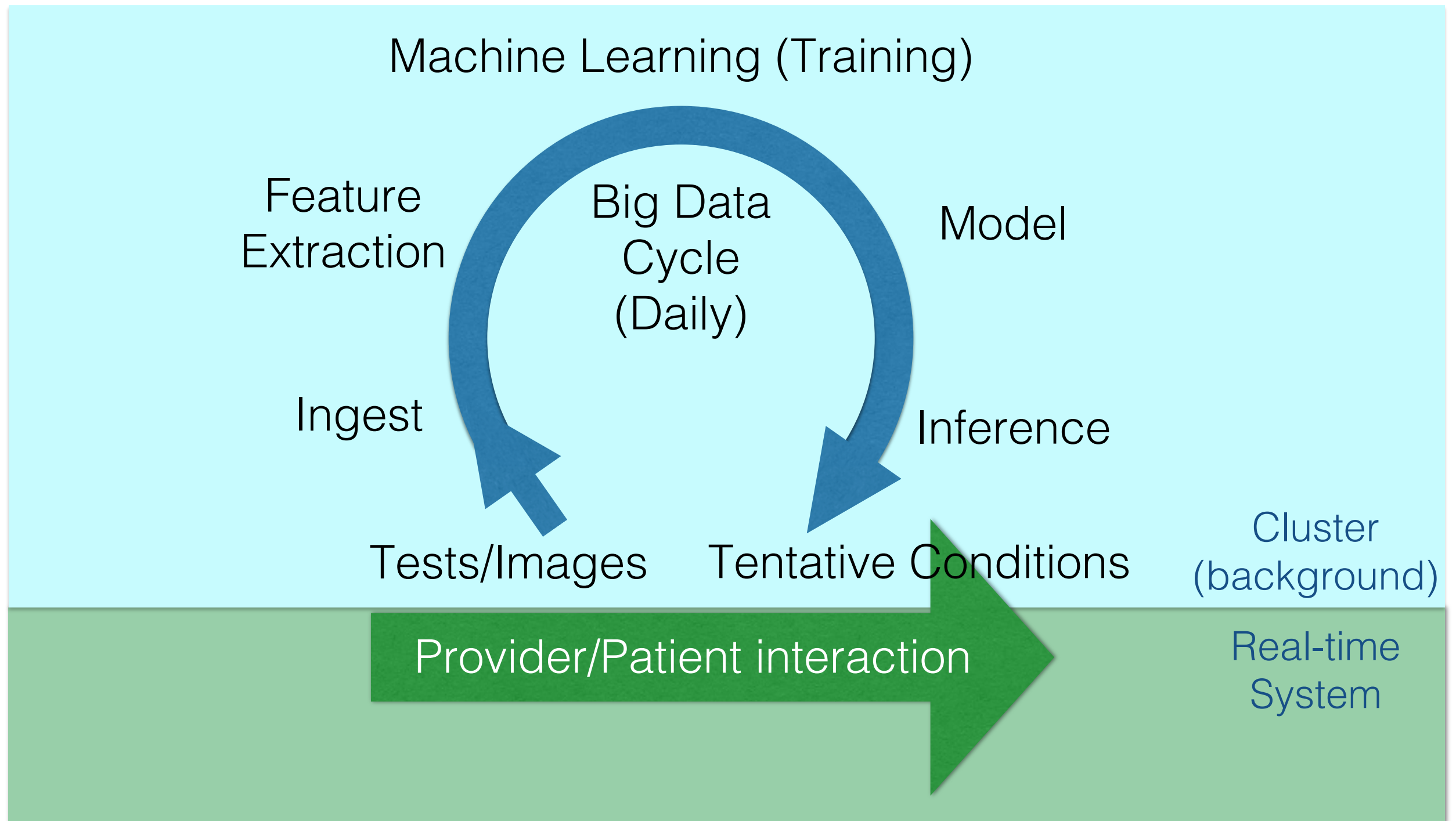
Big Data in web ad service



Big Data in Finance



Big Data in Healthcare



So, what's important?

- Ingest (Scala)
 - Whatever the source of information to be ingested, the ingest (typically, deserialization) process is *enormously* important.
 - We will learn about Domain-specific-languages (DSLs), Parsing, JSON, etc.
- Machine learning (more or less language-independent)
 - *Spark, PredictionIO, H2O, TensorFlow*, etc.
 - Feature extraction, training, validation, application, etc.
- Rules and/or other mutable state (Scala)
 - Actors to encapsulate mutable state;
 - Write to message broker to forward to monetizing stream.

Ingest, etc.

- Scala is very well suited to the tasks required for ingest.
- This is why we will be spending a lot of time in this class on ingest and related techniques.
- Especially, the assignments.