



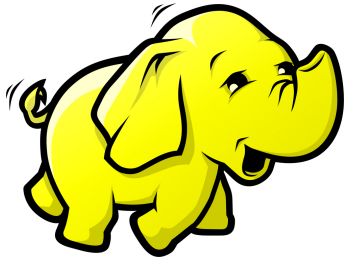
HCatalog

Extending Hive's Metadata to Pig and MR

Alan F. Gates
@alanfgates



Many Data Tools



MapReduce

- Early adopters
- Non-relational algorithms
- Performance sensitive applications



Pig

- ETL
- Data modeling
- Iterative algorithms



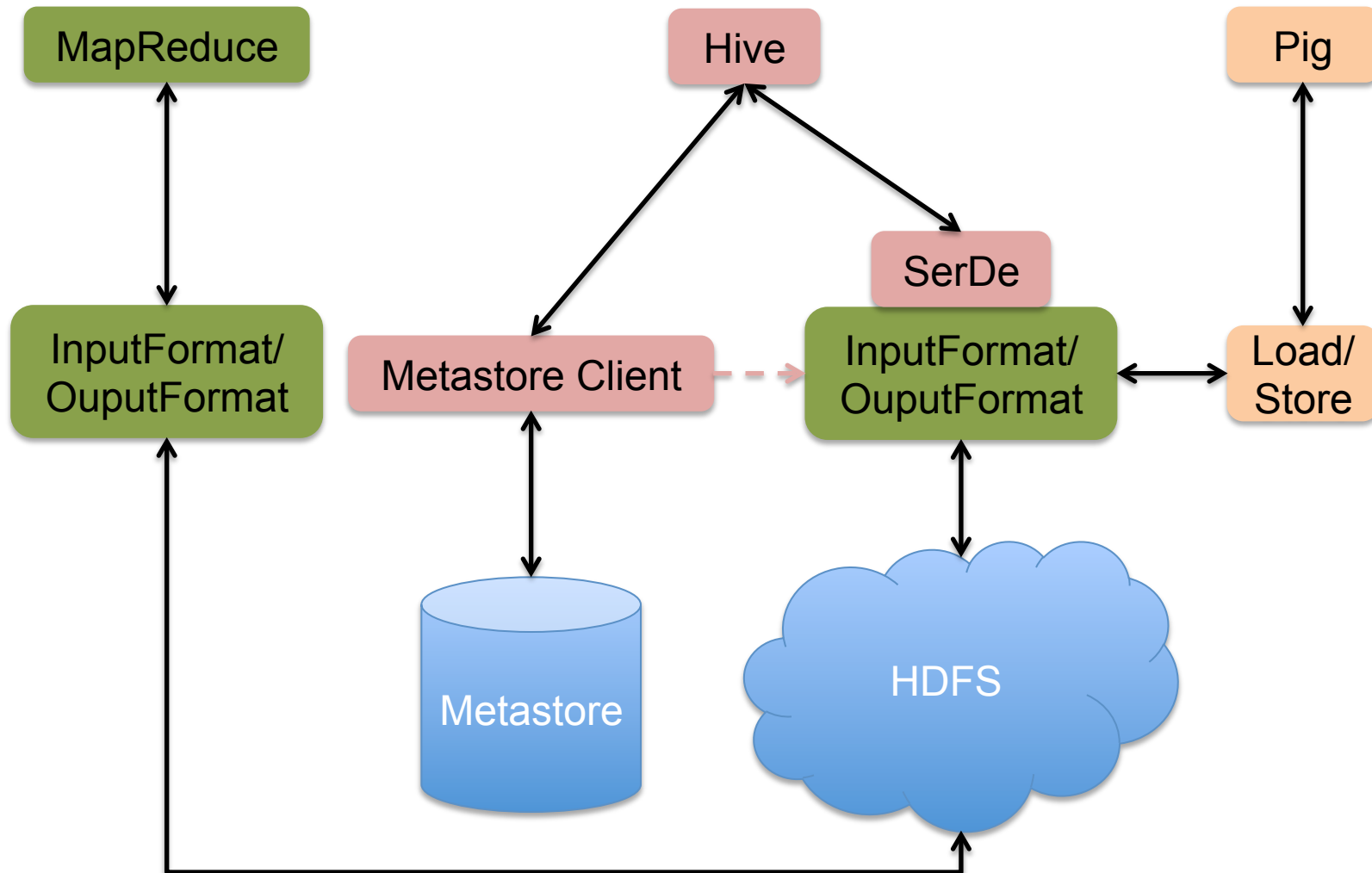
Hive

- Analysis
- Connectors to BI tools

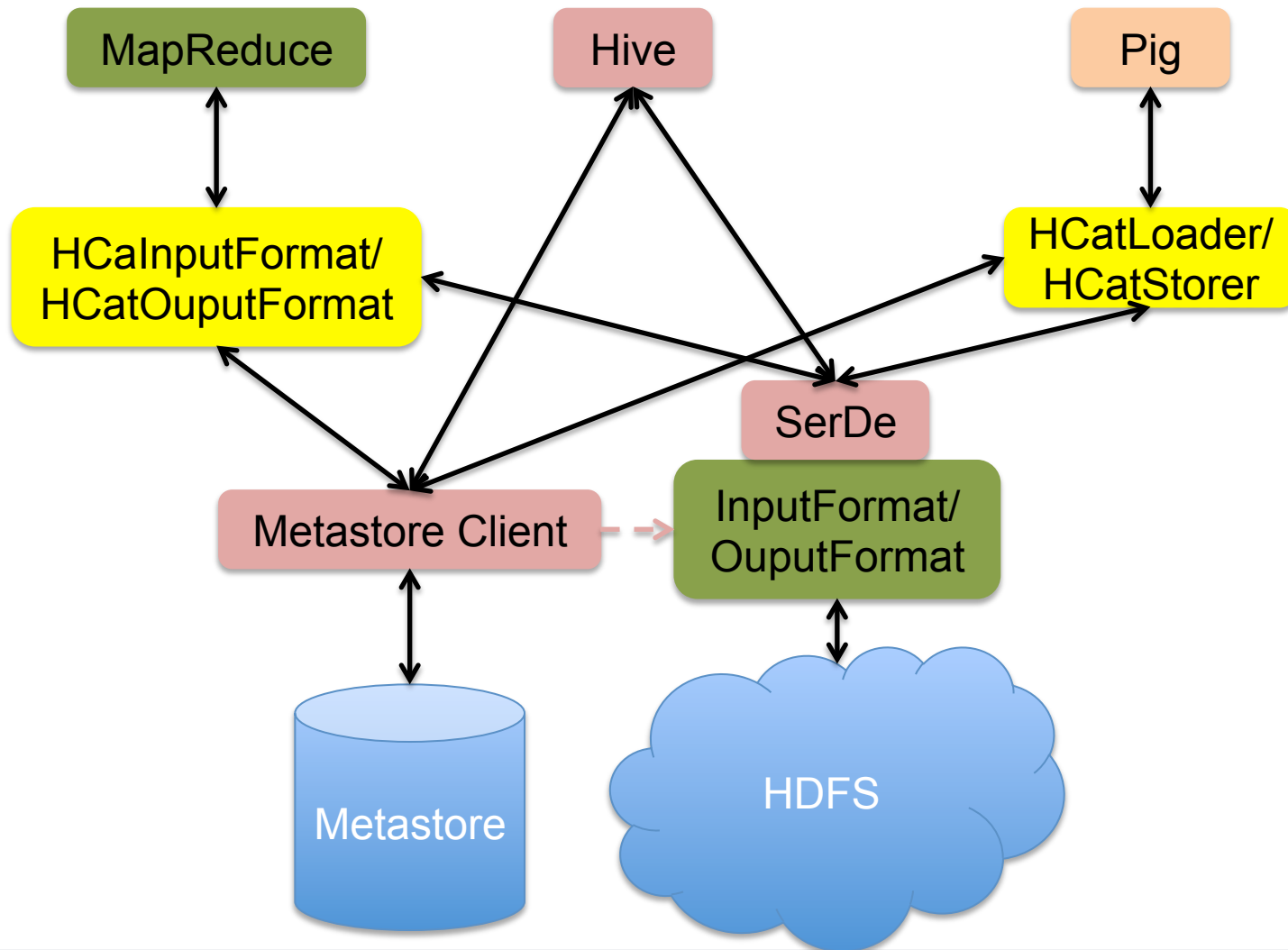
Strength: Pick the right tool for your application

Weakness: Hard for users to share their data

Hadoop Ecosystem



Opening up Metadata to MR & Pig



HCatalog Project

- HCatalog is an Apache Incubator project
- Version 0.4.0-incubating released May 2012
 - Hive/Pig/MapReduce integration
 - Support for any data format with a SerDe (Text, Sequence, RCFile, JSON SerDes included)
 - Notification via JMS
 - Initial HBase integration