Hive Client/Server Options

Prasad Mujumdar
(prasadm@cloudera.com)
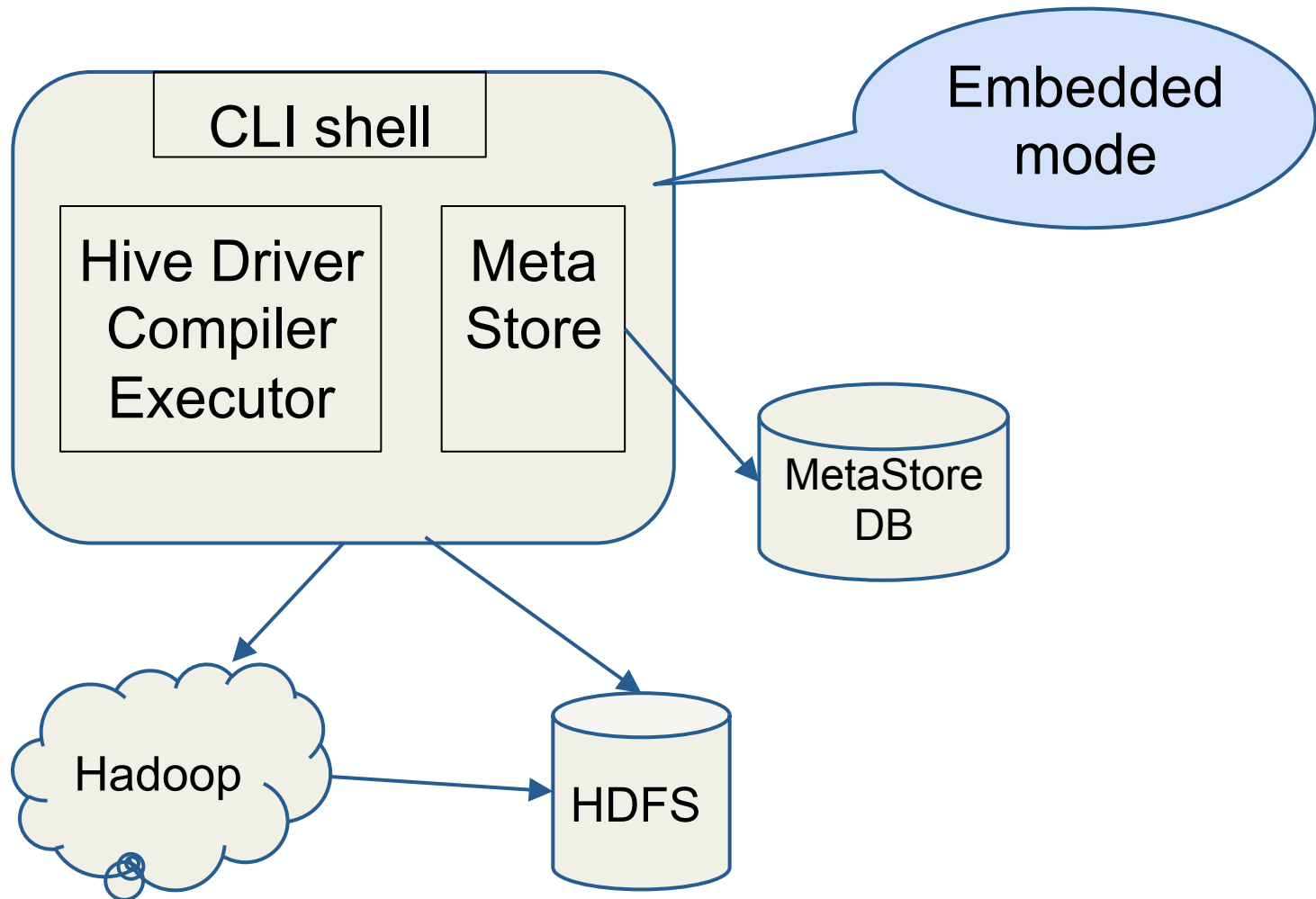
Hive Clients

- Command line shell (CLI)

- JDBC

- ODBC

- Custom thrift client
  - python, perl, C++, …

Hive Server processes

- Thrift based client/server messaging
- Hiveserver
    - Executing SQL and fetch data
- HiveServer2, improved thrift server
    - concurrency and authentication
- MetaStore server
    - Remote metadata access

Hive CLI

Hive CLI

- Most commonly used with hive

- Thick client mode
  - Includes Hive execution engine
  - Command shell
  - Metastore

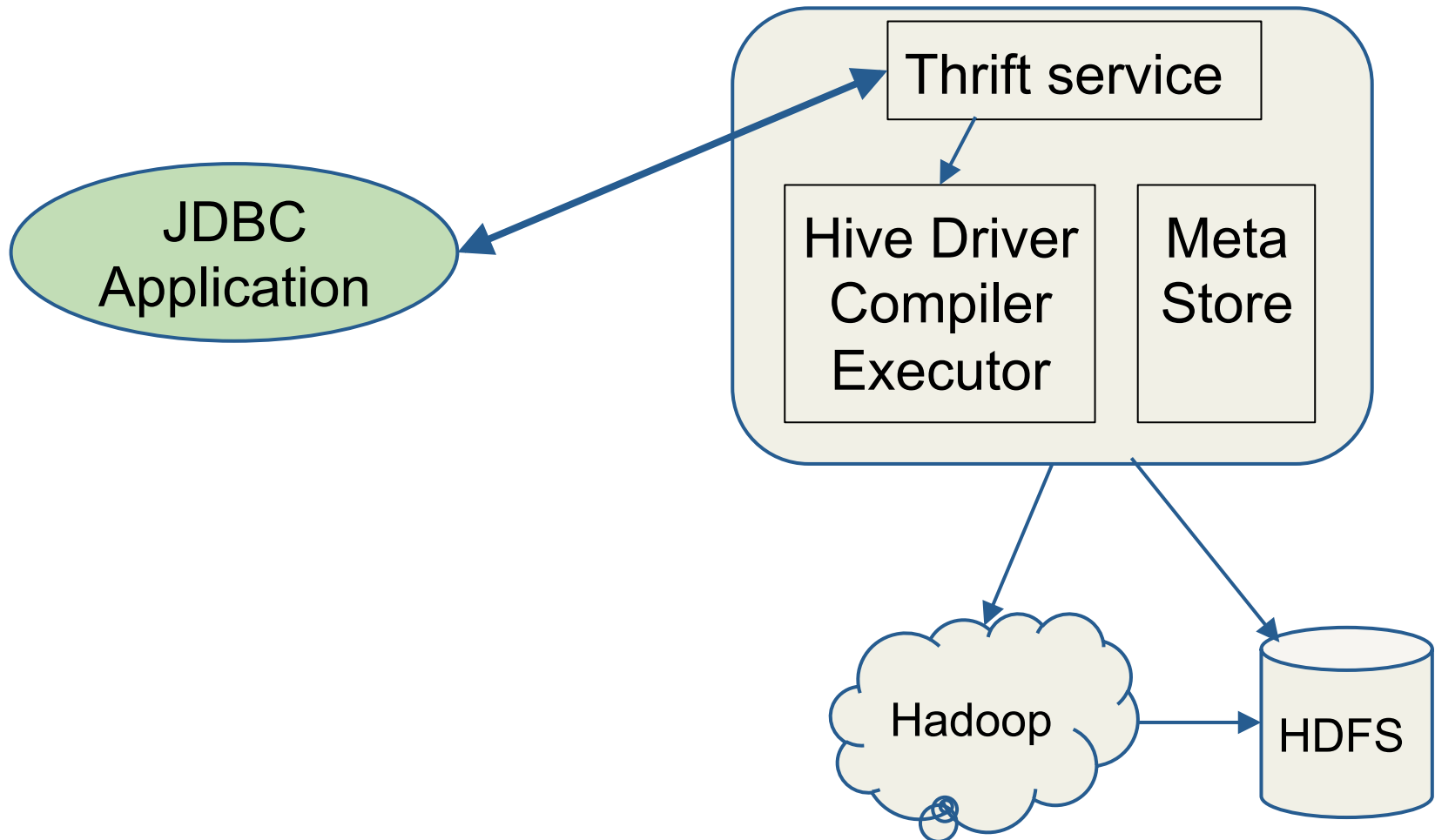- Supports a thin client mode with Thrift server

JDBC

- Java client package implements JDBC APIs

- Java Thrift client for Hive thrift servers

- Part of Apache Hive distribution

- Enables ETL/BI/Analytics tools and custom apps

ODBC

- C client library implements ODBC APIs

- C Thrift client for Hive thrift servers

- Multiple distros
  - Apache Hive
  - Cloudera
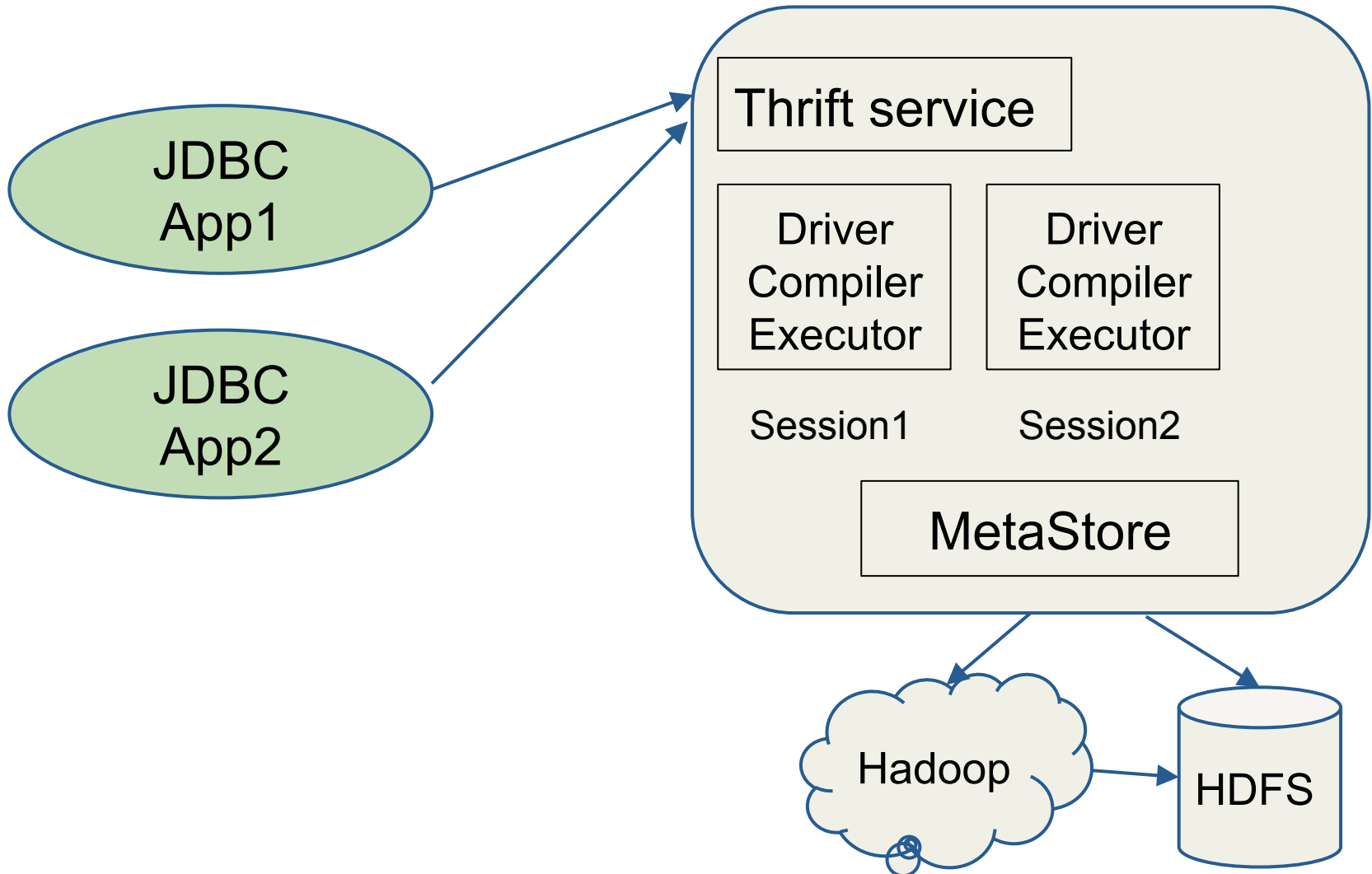 - Simba

- Enables ETL/BI/Analytics tools and custom apps

Hive Server

Hive Trift Server

- Thrift based server for data access

- Work with CLI, JDBC and ODBC clients

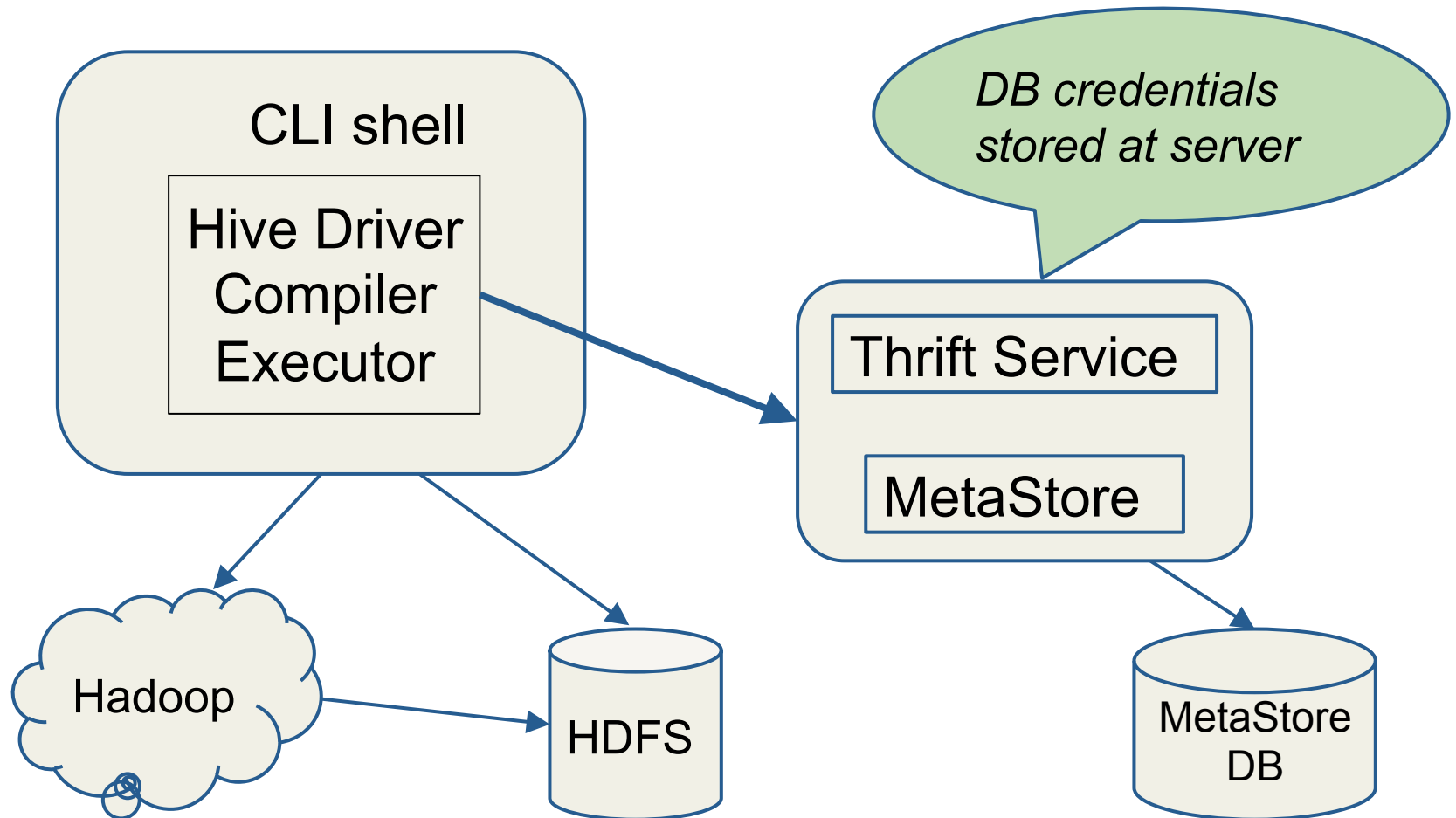- Doesn't support concurrent client access

- No authentication

HiveServer2

HiveServer2

- Improved Thrift server for data access

- Supports concurrent clients

- Supports authentication

  - Kerberos

  - Pass through LDAP

- Includes new command line tool 'beeline'

- Better support for JDBC/ODBC drivers
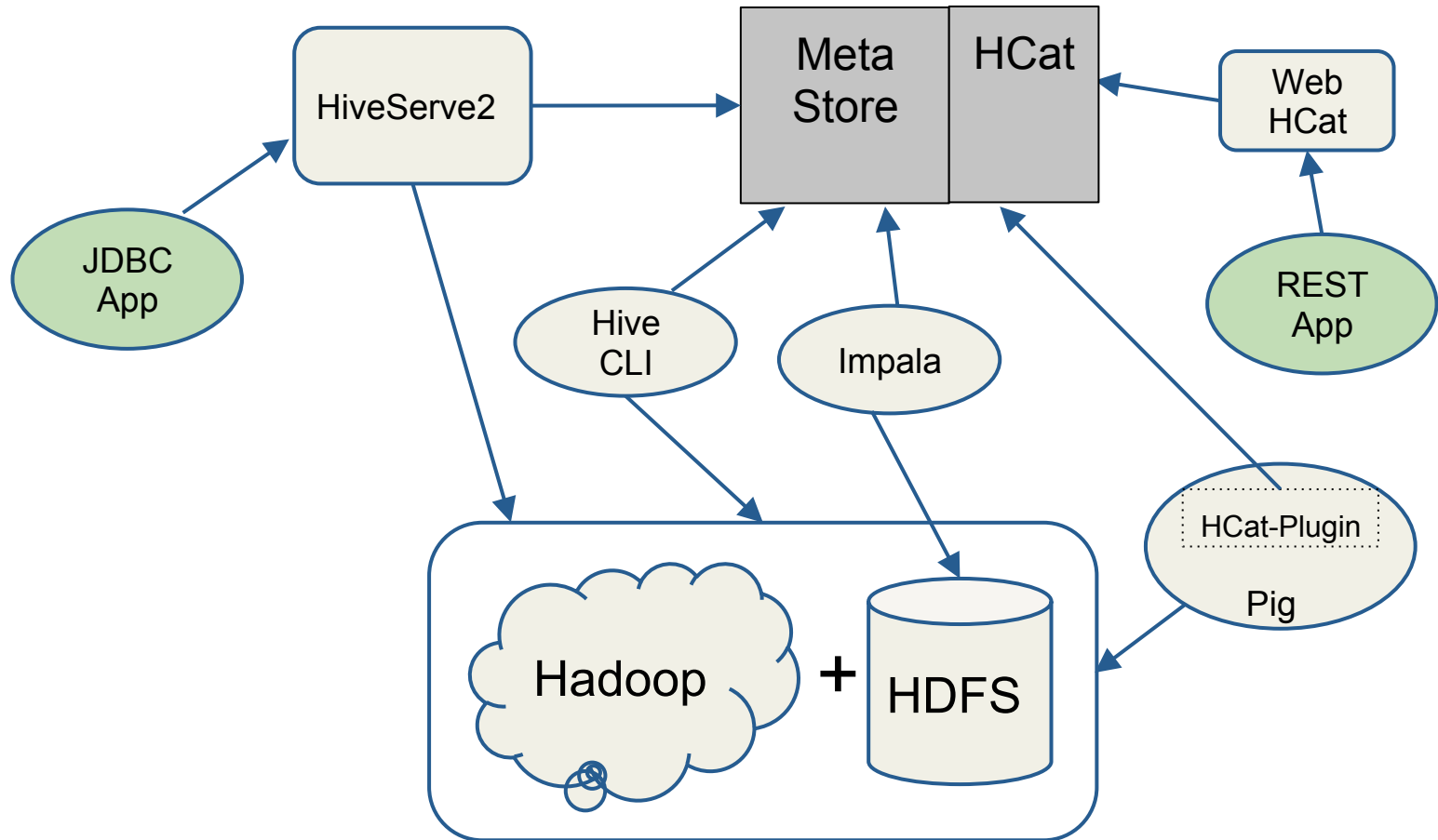
- Works with secure Hadoop

Metastore Server

Remote metastore

- Thrift server for metadata access

- Open hive metadata to external clients

- Consolidate BD credentials

- Secure access

  - Kerberos authentication

  - Hadoop user impersonation

Multi client Remote metastore

Questions ???