

November 2011

Apache Sqoop (Incubating)

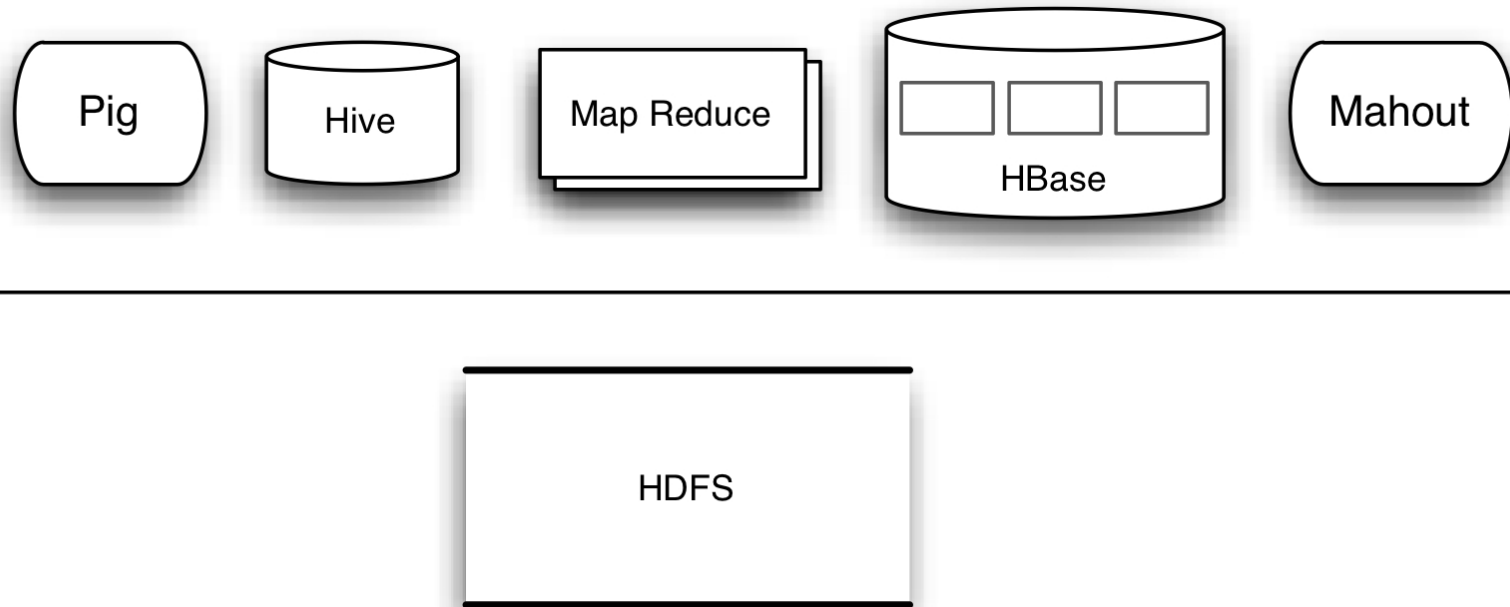
Integrating Hadoop with Enterprise RDBS – Part I

Arvind Prabhakar (arvind at apache dot org)

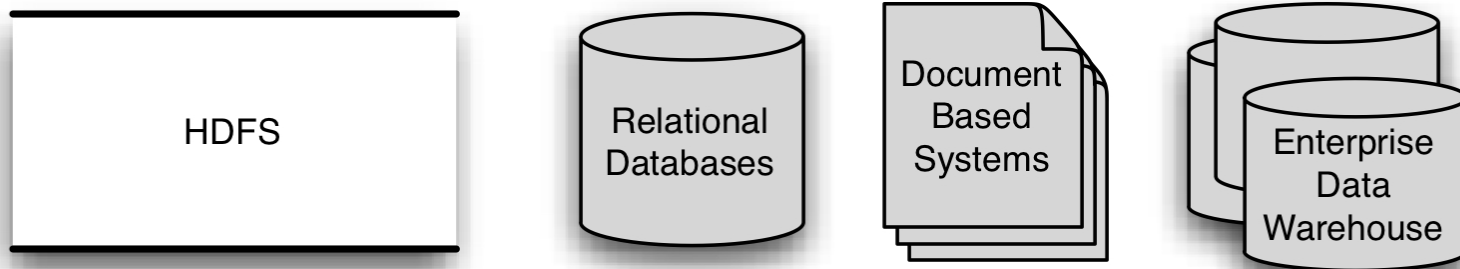
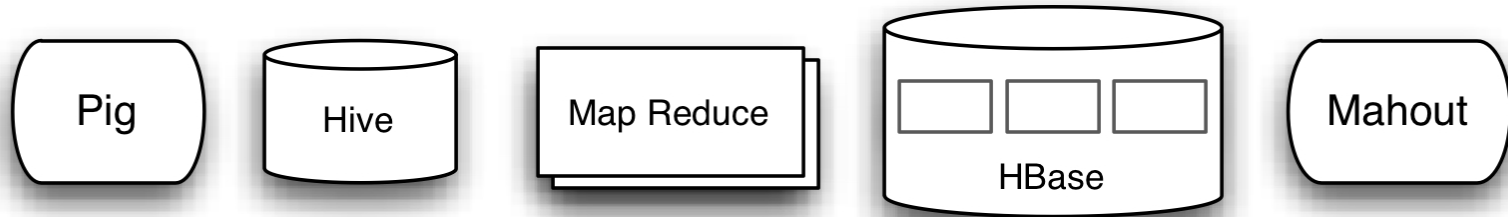
Apache Sqoop Committer and Software Engineer at Cloudera



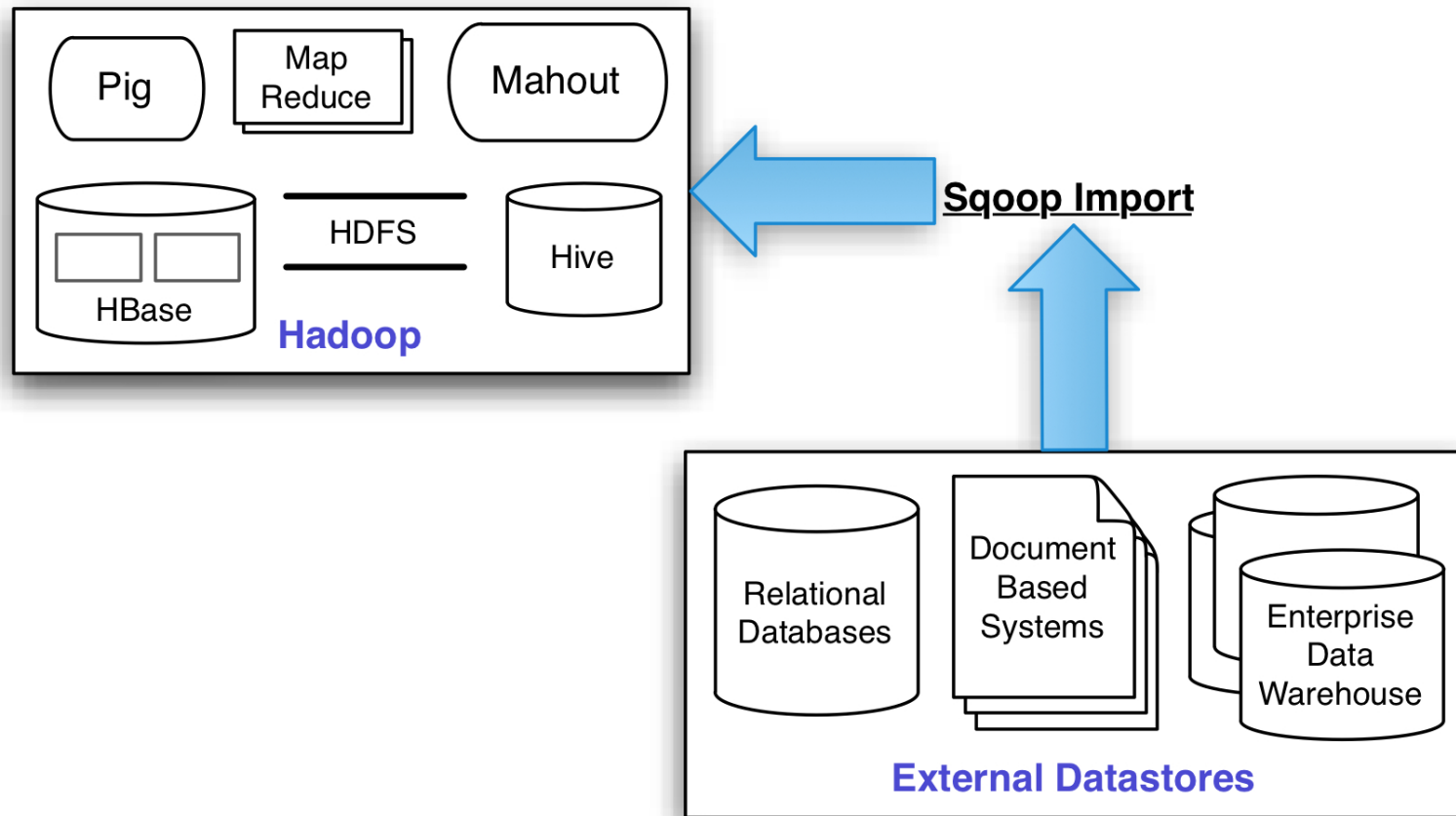
Hadoop Data Processing



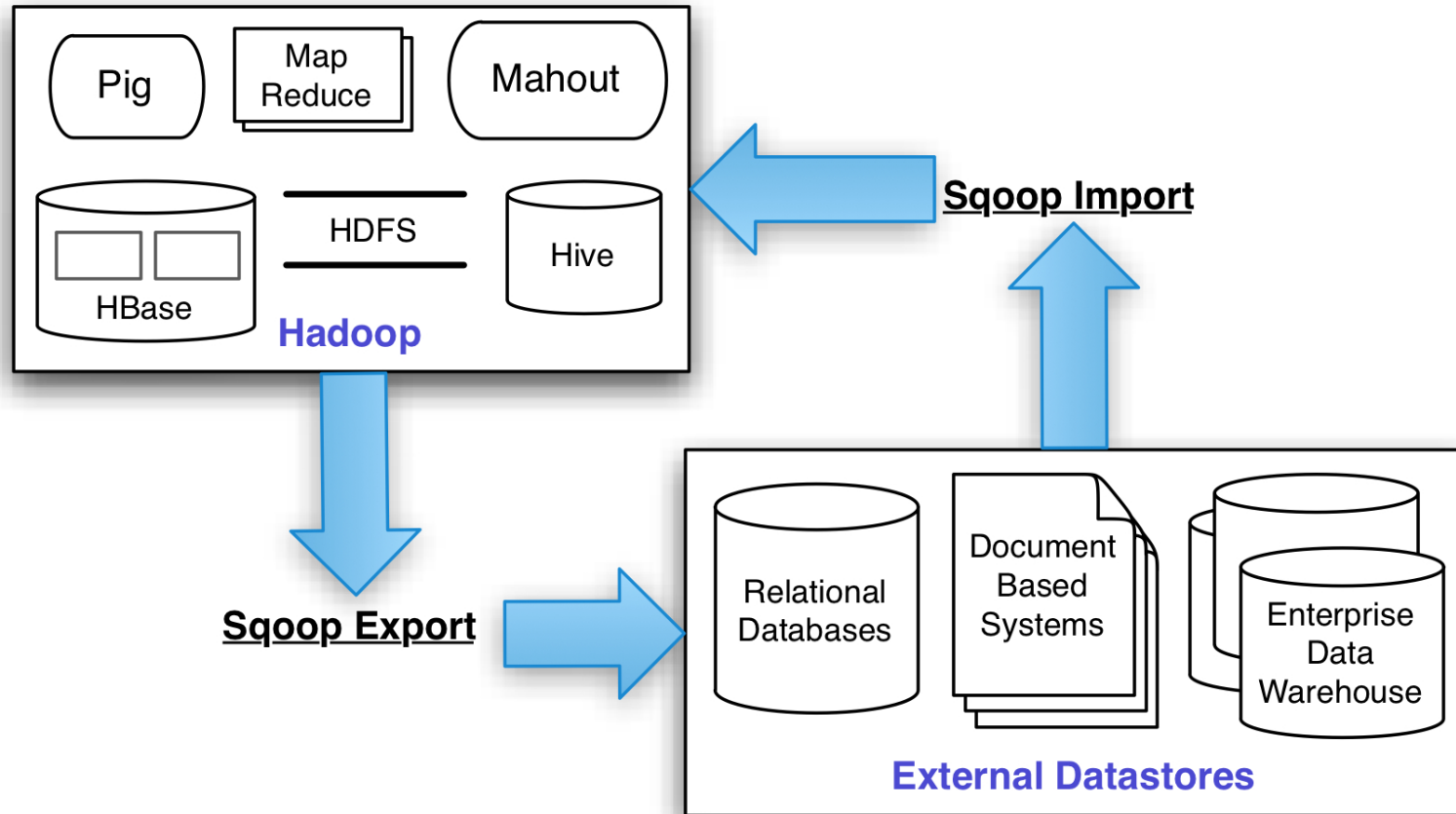
Hadoop Data Processing



Hadoop Data Processing



Hadoop Data Processing



In This Session...

- How Sqoop Works
- Roadmap

Data Import

CU_ORDERS
CUO_ID (PK)
CUO_CUSTOMER
CUO_SALESREP
...

3721	Acme	Jon	
1283	Xyz	Dan	



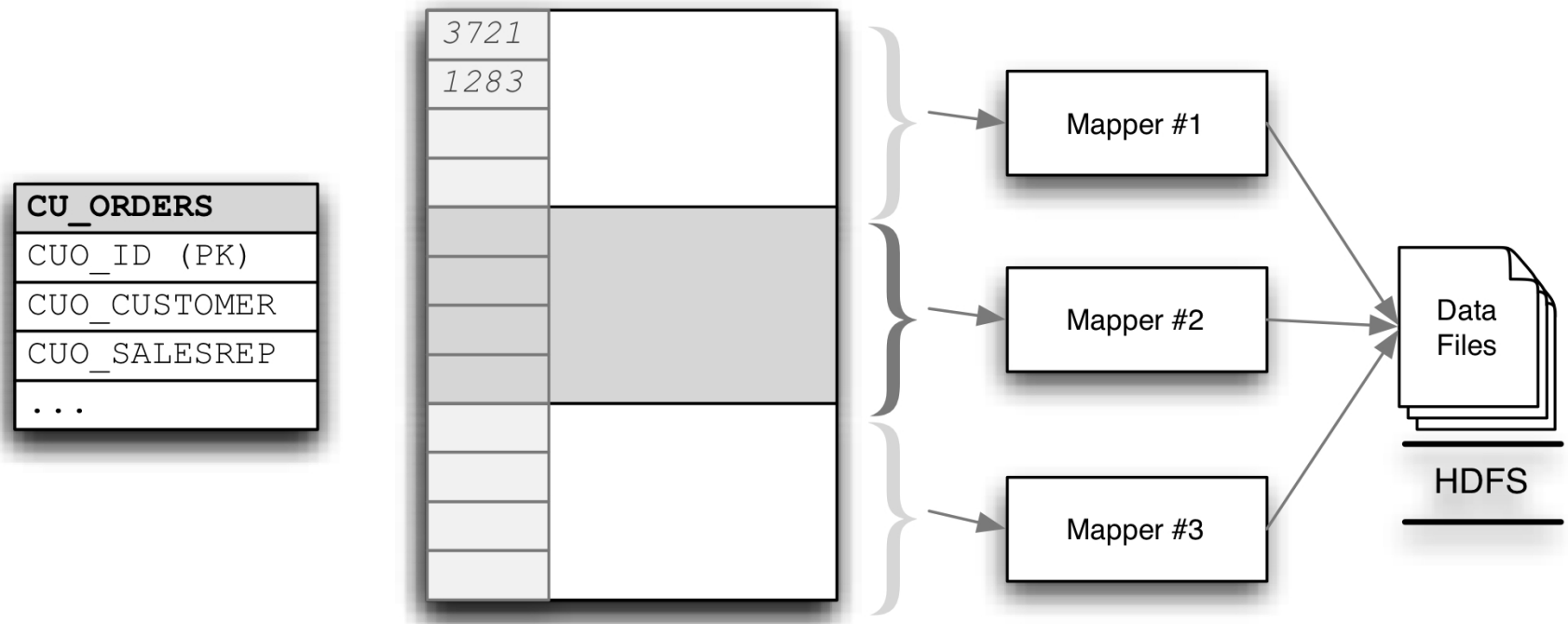
Data Import

CU_ORDERS
CUO_ID (PK)
CUO_CUSTOMER
CUO_SALESREP
...

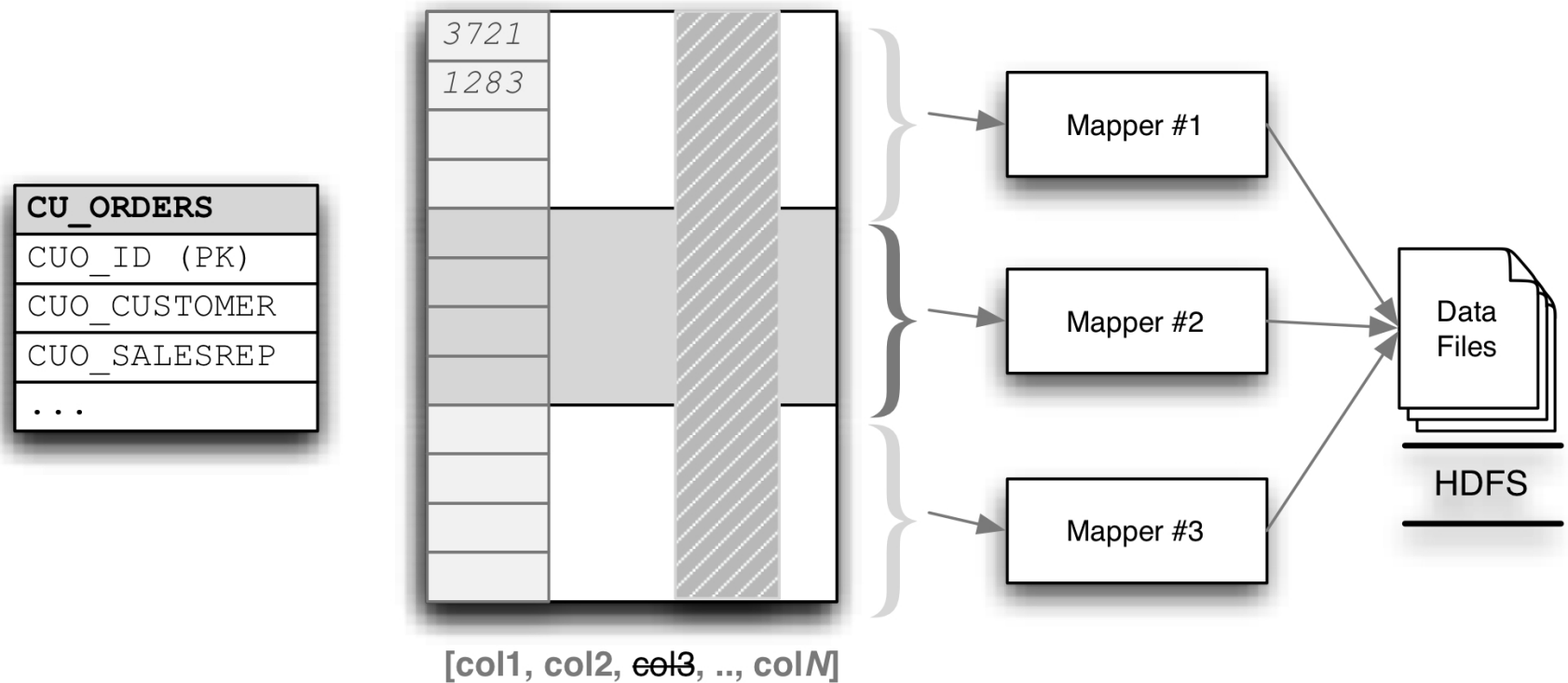
3721	
1283	



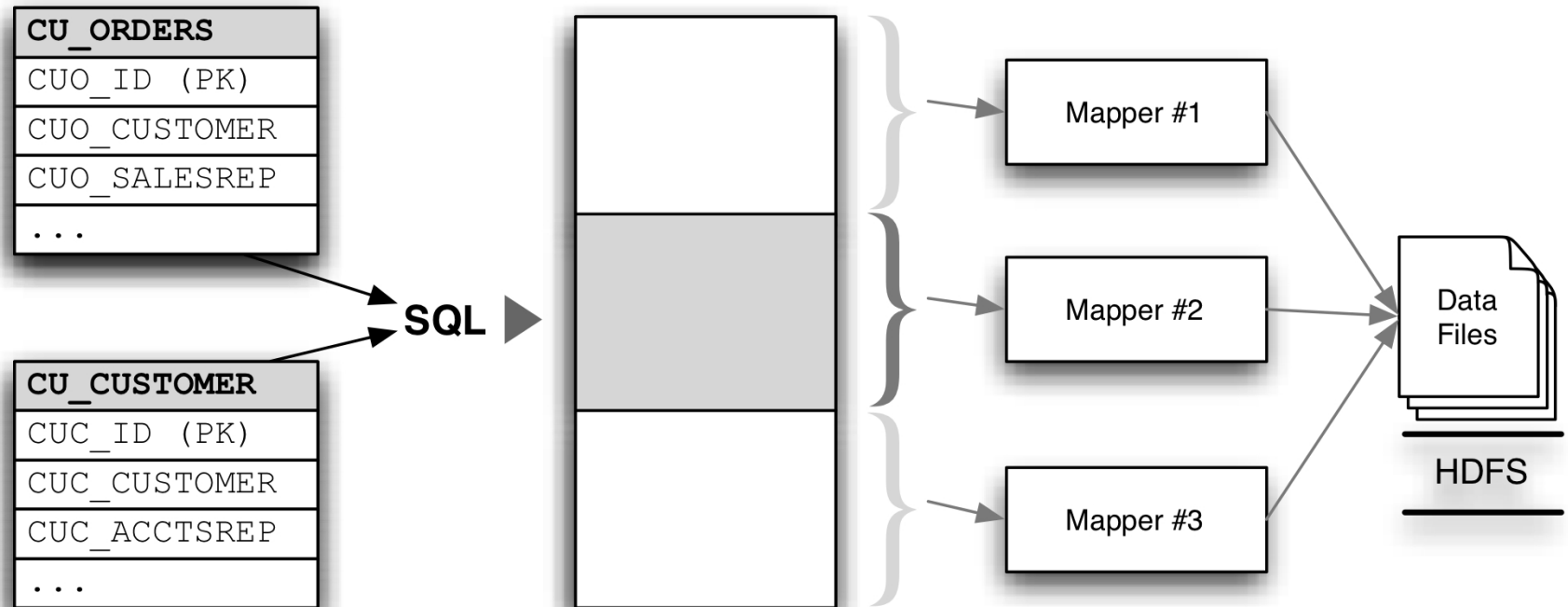
Data Import



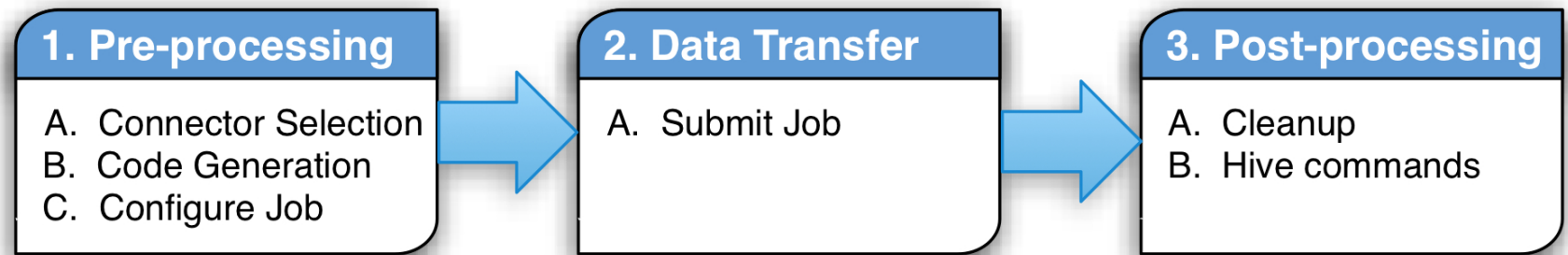
Data Import



Data Import



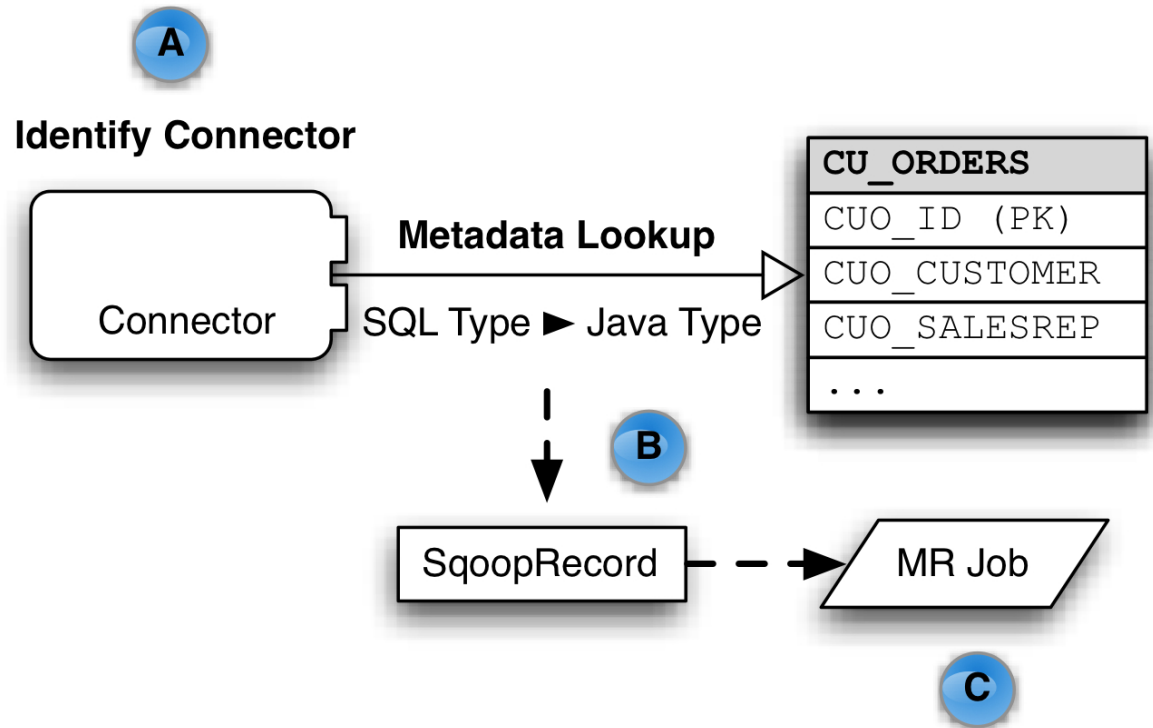
Sqoop Overview



Pre-processing

1. Pre-processing

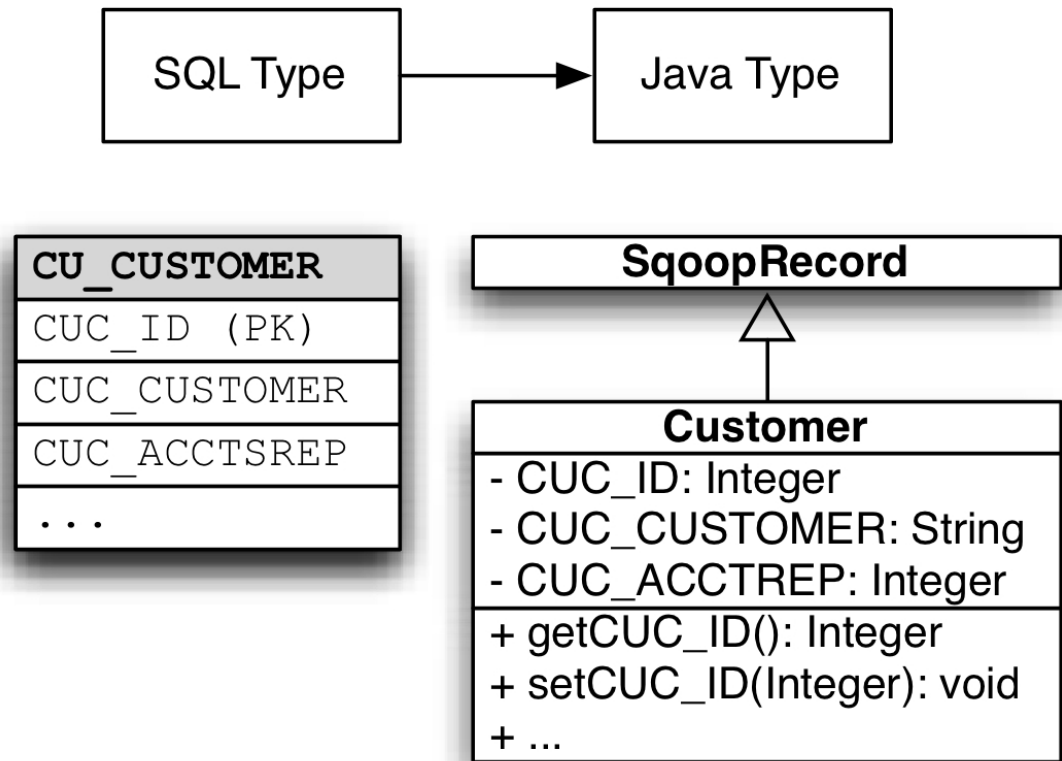
- A. Connector Selection
- B. Code Generation
- C. Configure Job



Code Generation

1. Pre-processing

- A. Connector Selection
- B. Code Generation
- C. Configure Job



Type Mapping

1. Pre-processing

- A. Connector Selection
- B. Code Generation
- C. Configure Job

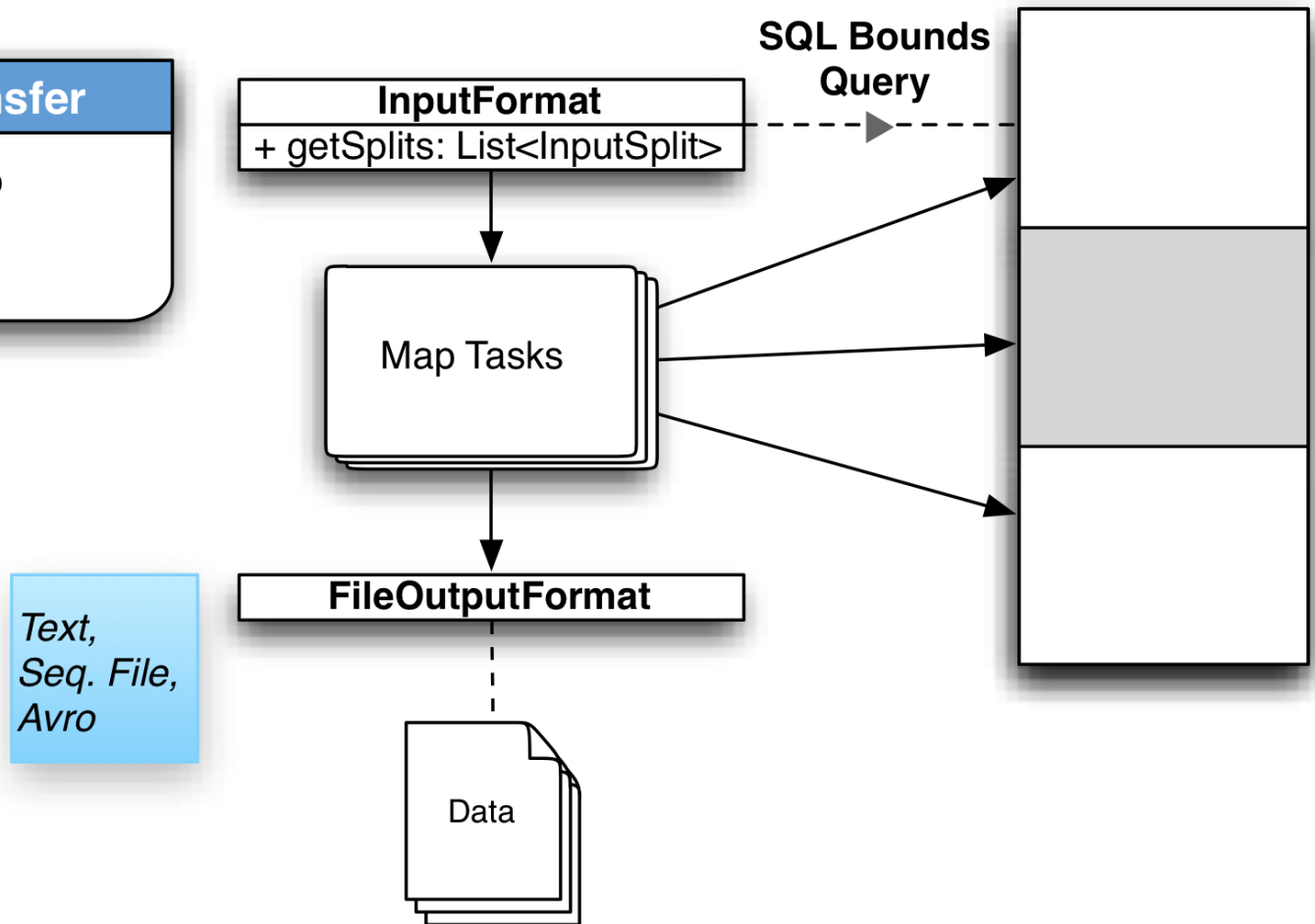
SQL Type	Java Type
INTEGER	<code>java.lang.Integer</code>
VARCHAR	<code>java.lang.String</code>
LONGVARCHAR	<code>java.lang.String</code>
NCHAR	<code>java.lang.String</code>
NUMERIC	<code>java.math.BigDecimal</code>
DECIMAL	<code>java.math.BigDecimal</code>
BOOLEAN	<code>java.lang.Boolean</code>
DATE	<code>java.sql.Date</code>
...	...



Data Transfer

2. Data Transfer

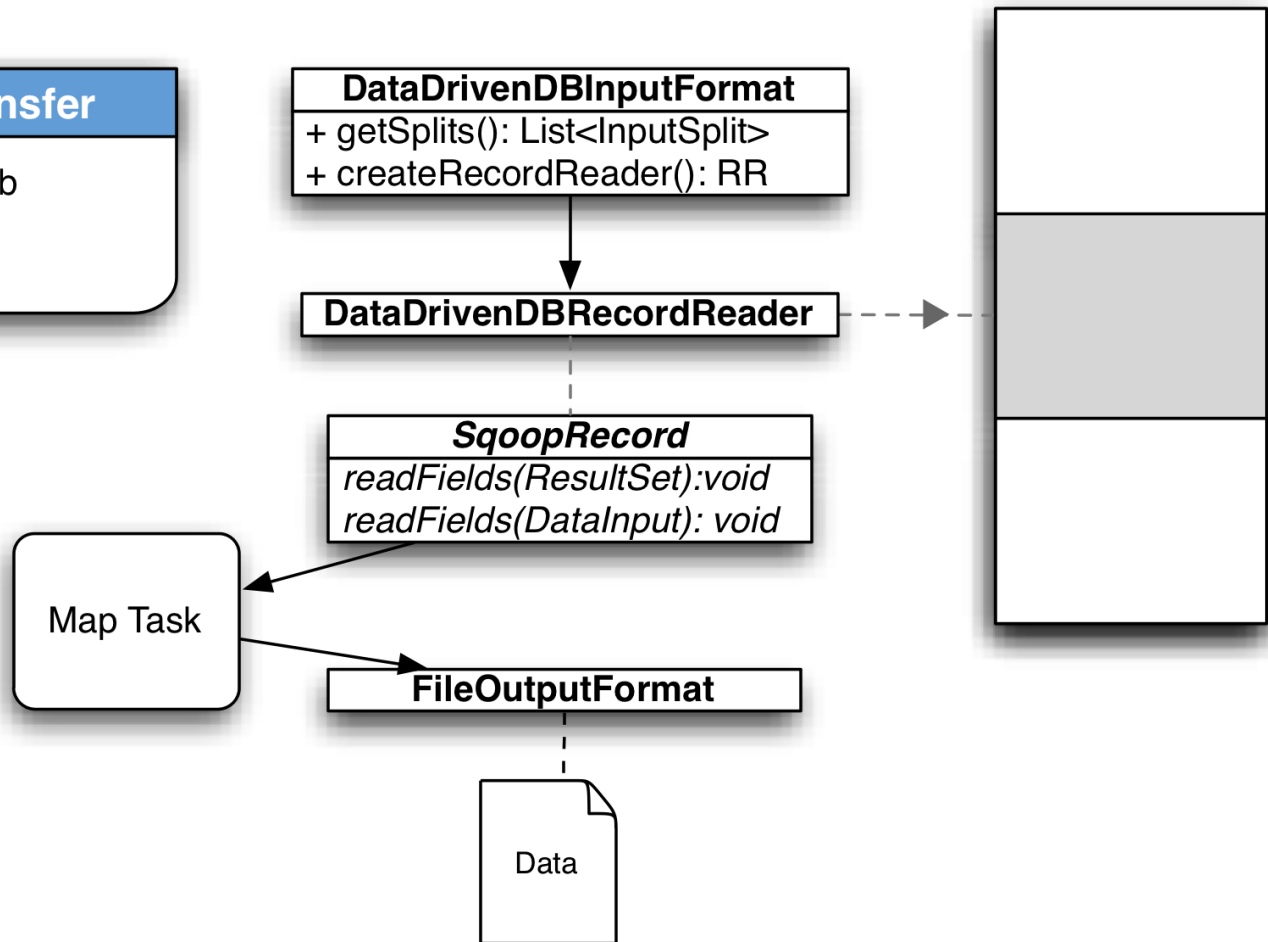
A. Submit Job



Data Transfer

2. Data Transfer

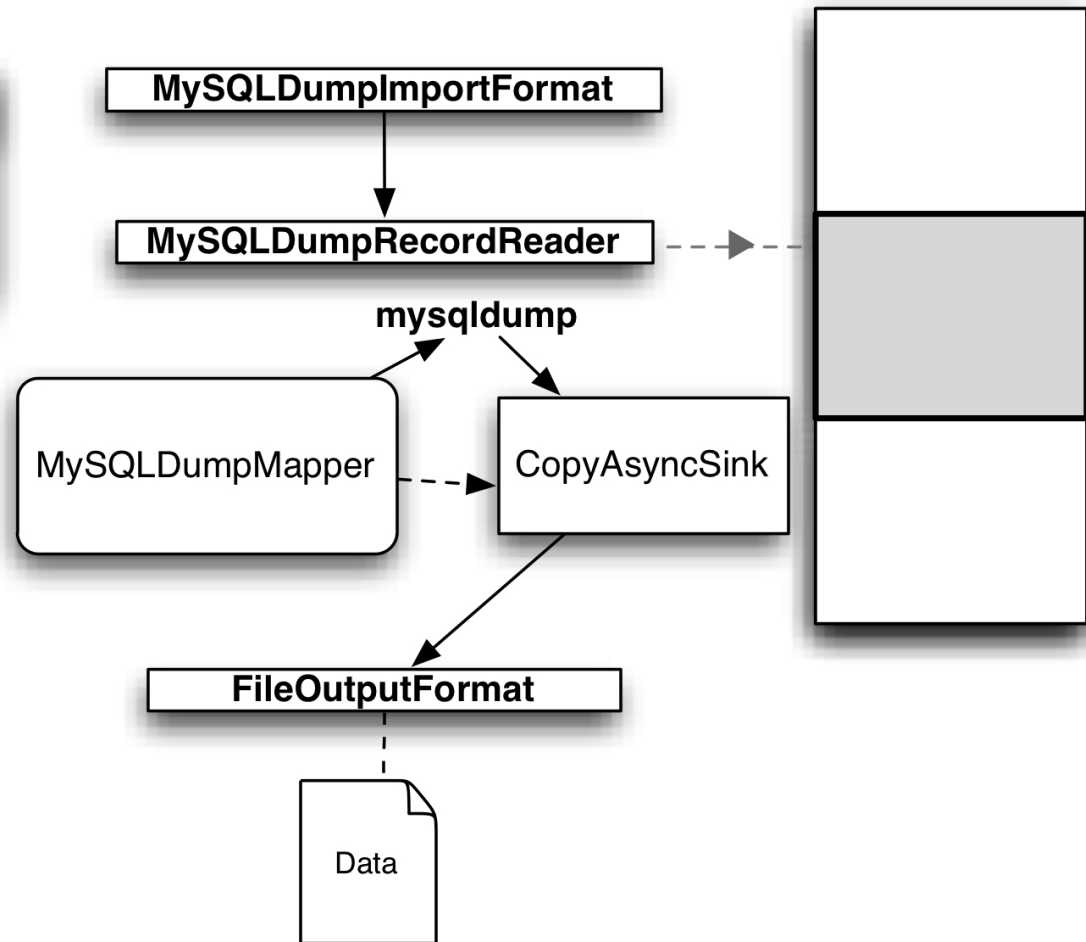
A. Submit Job



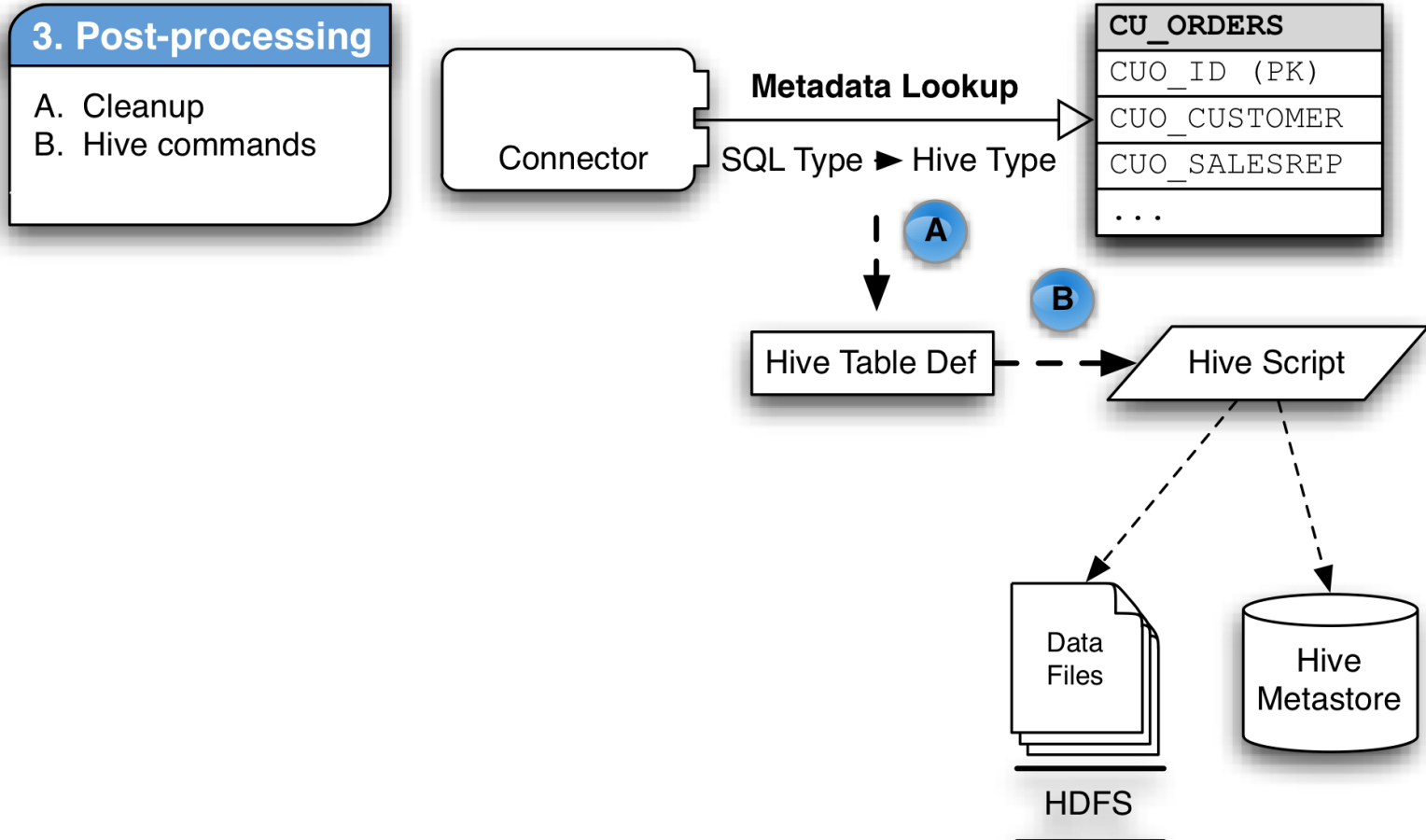
Data Transfer

2. Data Transfer

A. Submit Job



Post-Processing



Sqoop Connectors

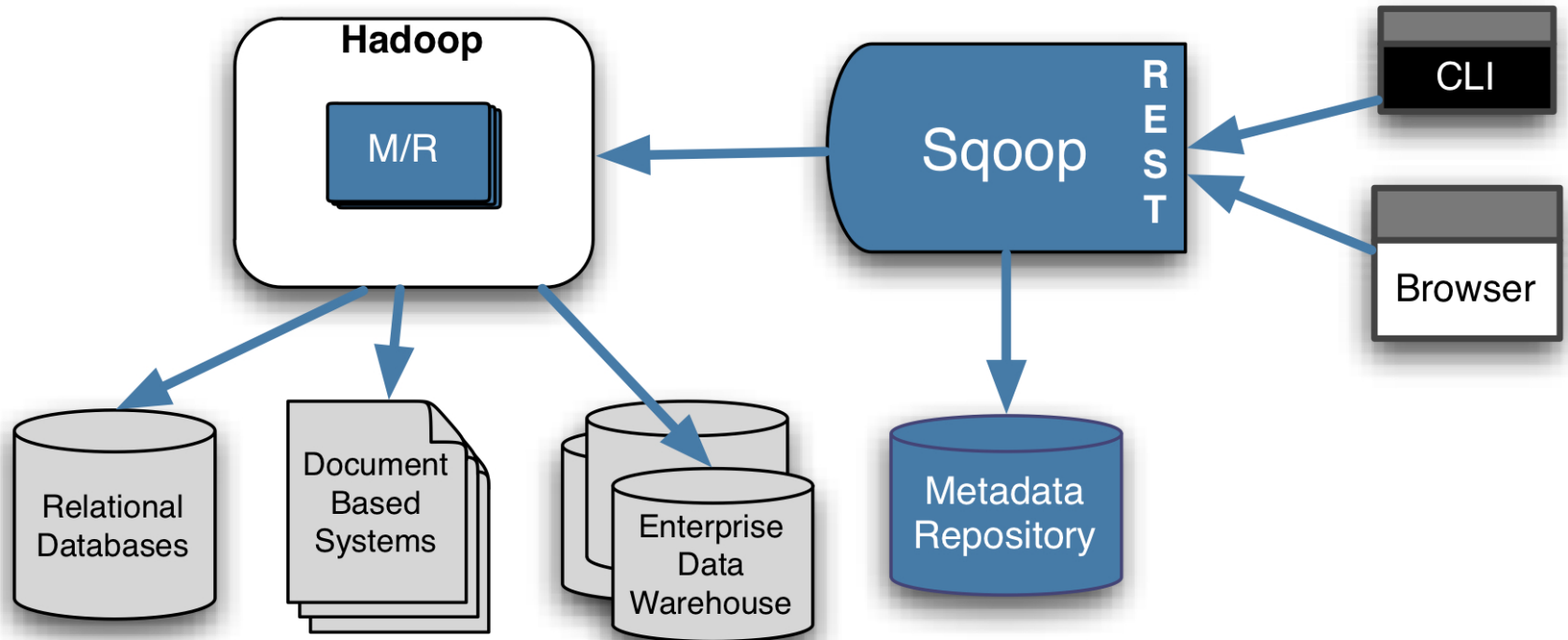
- Oracle – Developed by Quest Software
- Couchbase – Developed by Couchbase
- Netezza – Developed by Cloudera
- Teradata – Developed by Cloudera
- SQL Server – Developed by Microsoft
- Microsoft PDW – Developed by Microsoft
- Volt DB – Developed by Volt DB



Sqoop Roadmap

- SQOOP-365: *Proposal for Sqoop 2.0*
 - <https://issues.apache.org/jira/browse/SQOOP-365>
- Highlights
 - Sqoop as a Service
 - Connections as First Class Objects
 - Role based Security

Sqoop 2 Architecture (proposed)



For More Information

- Website:

<http://incubator.apache.org/sqoop/>

- Mailing Lists:

incubator-sqoop-user-subscribe@apache.org

incubator-sqoop-dev-subscribe@apache.org

- Issue Tracker:

<http://issues.apache.org/jira/browse/SQOOP>

Thank You!

Q & A will be after part II of this session.

