



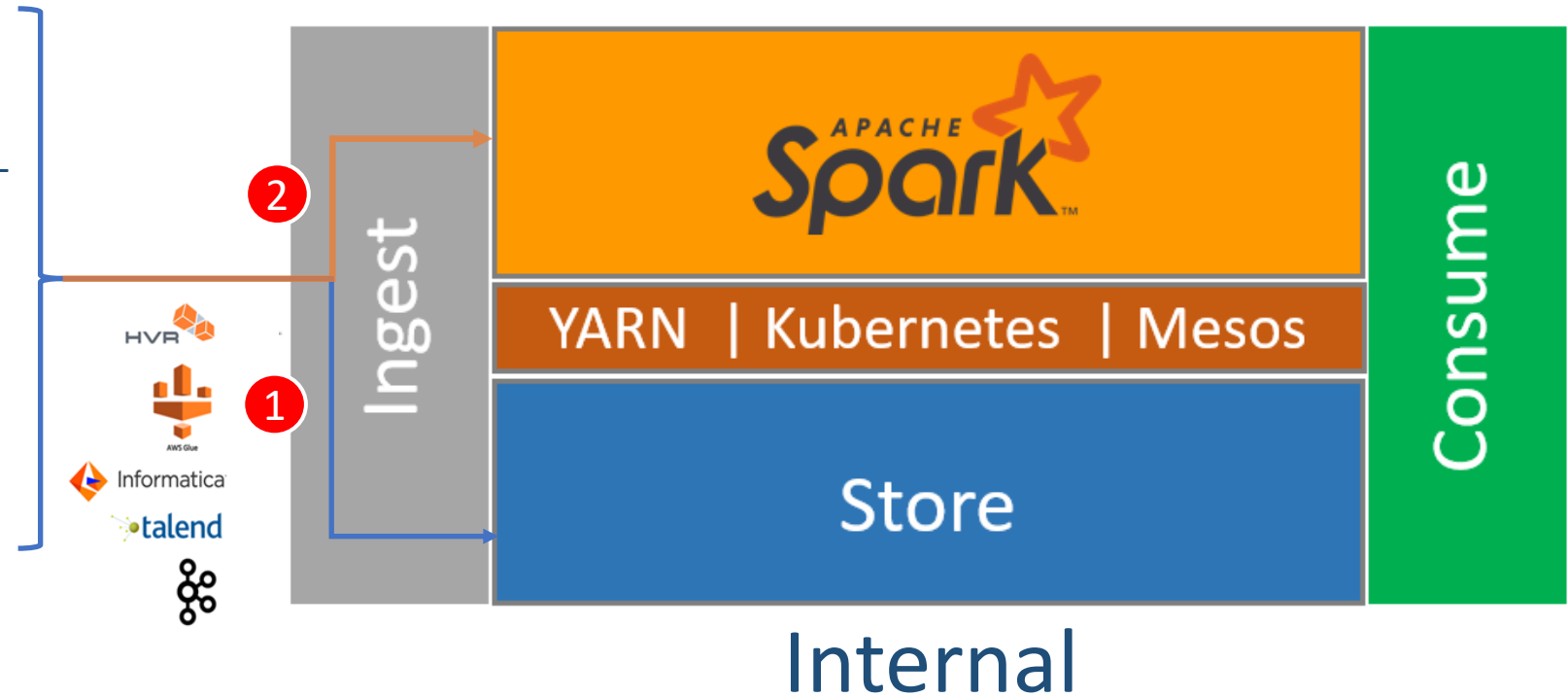
Apache Spark

Spark data sources/sinks

Data Lake

External

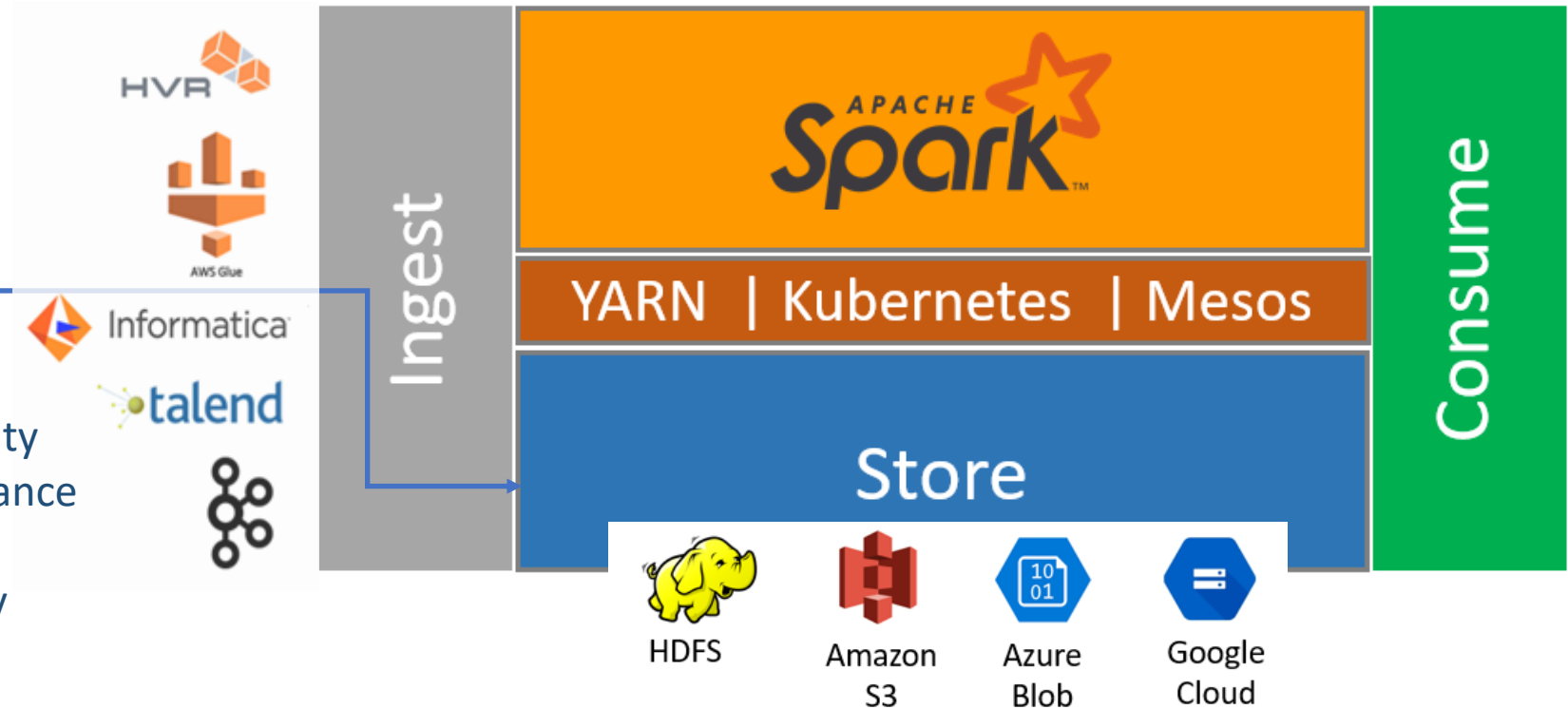
1. **JDBC Data Sources**
Oracle, SQL Server, PostgreSQL
2. **No SQL Data Systems**
Cassandra, MongoDB
3. **Cloud Data Warehouses**
Snowflake, Redshift
4. **Stream Integrators**
Kafka, Kinesis



Data Lake

External

1. Modularity
2. Load Balance
3. Security
4. Flexibility



Internal

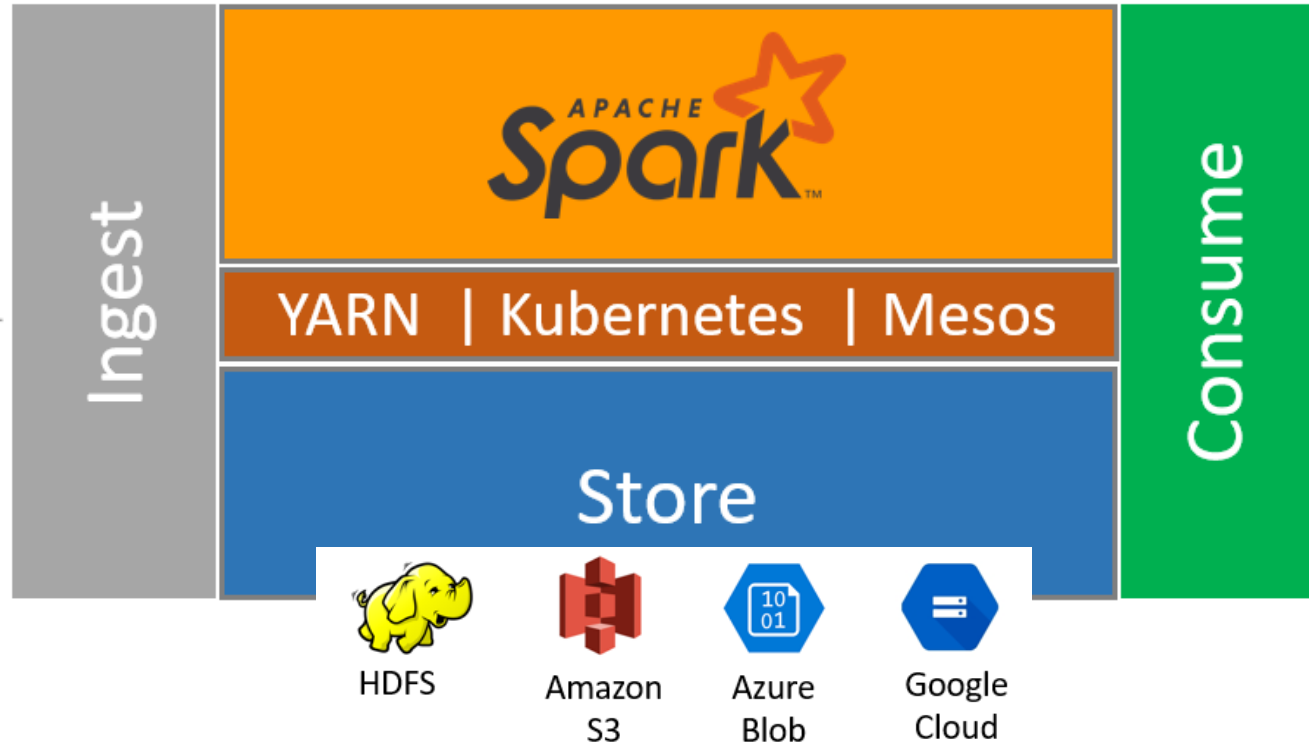
1. CSV
2. JSON
3. Parquet
4. AVRO
5. Plain Text

1. Spark SQL Tables
2. Delta Lake

Data Lake

External

1. JDBC Data Sources
2. NoSQL Data Systems
3. Cloud Data System
4. Stream Integrators



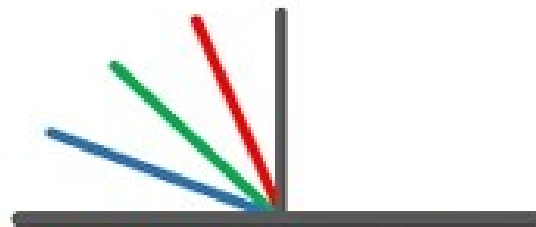
External

1. JDBC Data Sources
2. NoSQL Data Systems
3. Cloud Data System
4. Stream Integrators

Internal

1. CSV
2. JSON
3. Parquet
4. AVRO
5. Plain Text

1. Spark SQL Tables
2. Delta Lake



LEARNING JOURNAL

www.learningjournal.guru