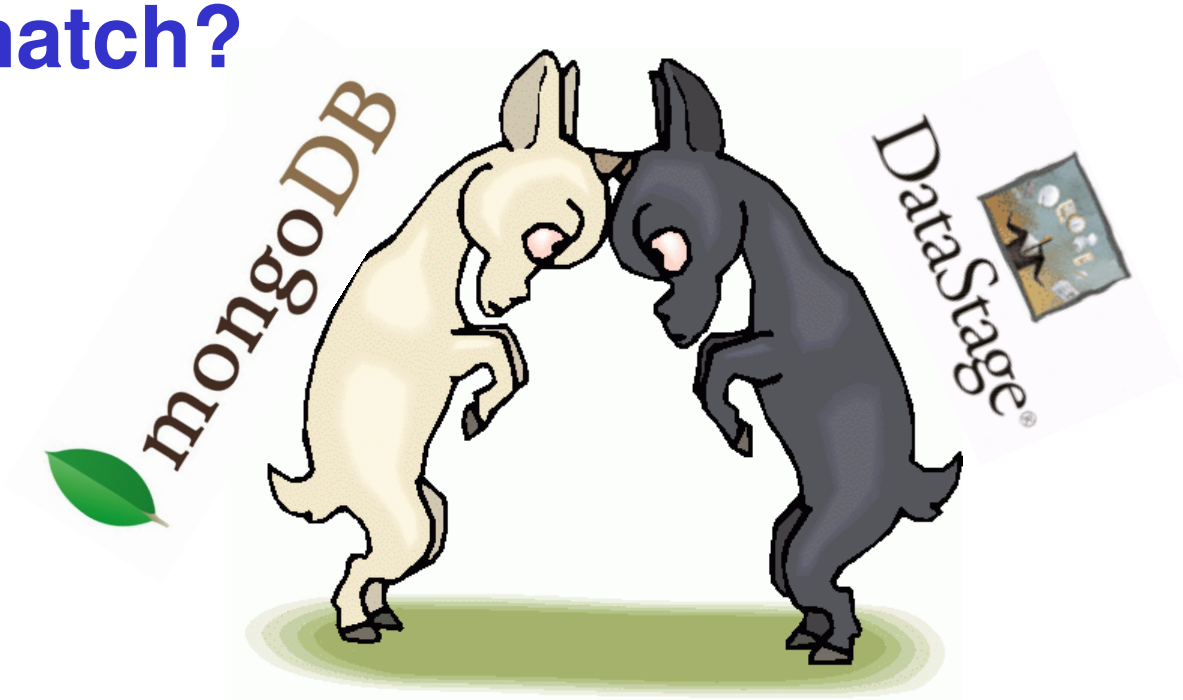
DataStage & MongoDB

NoSQL and ETL Collide!

Paul Stanley

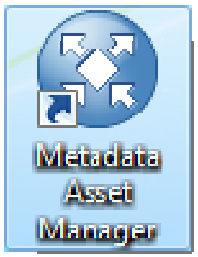
IBM Information Server Connectivity Architect

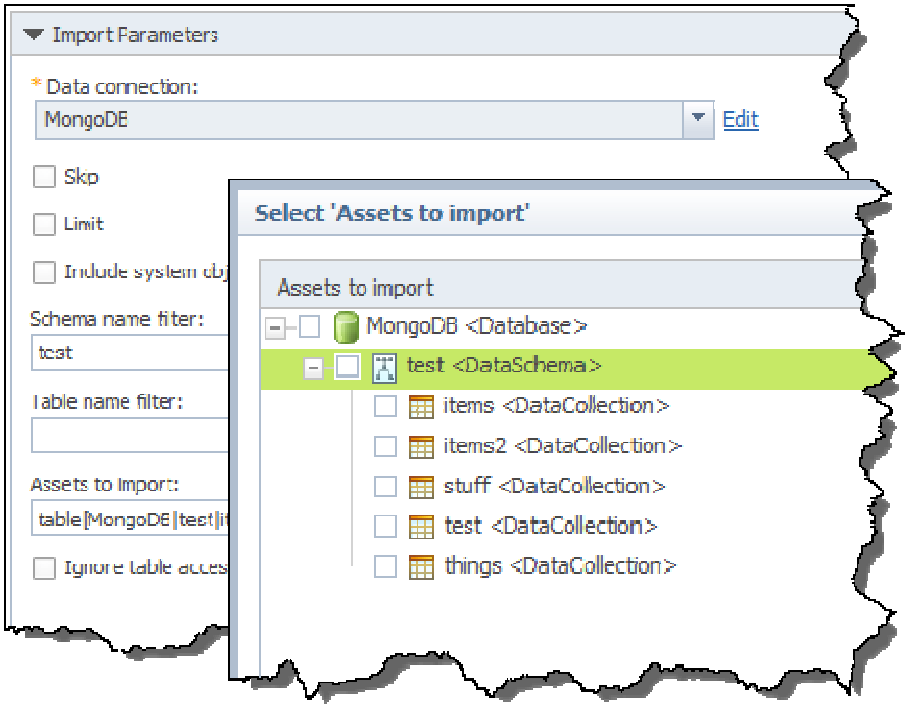
NoSQL and ETL : An Impedance Mismatch?



- ETL is metadata driven
  - Transformation, governance, lineage
- NoSQL is metadata-free!

Metadata Determined by Introspection

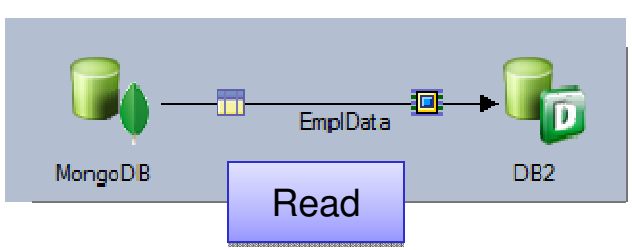


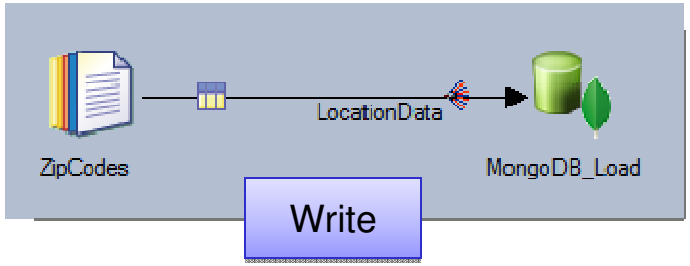


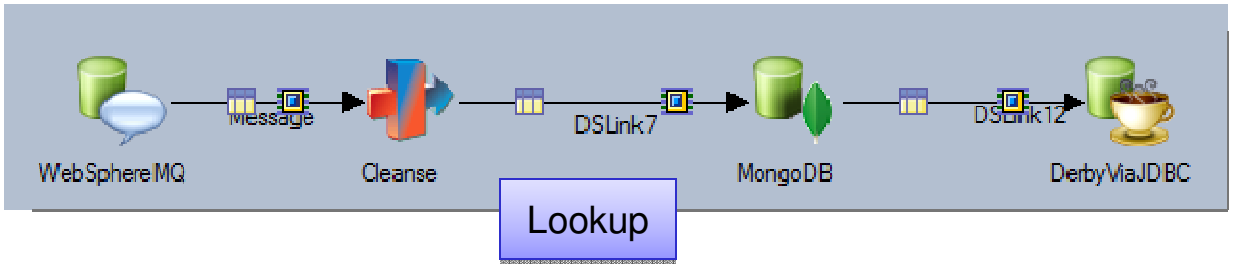
- Import metadata by examining a sample of documents
- Structure flattened to a relational view
- Metadata stored in Information Server repository

MongoDB Connector Highlights

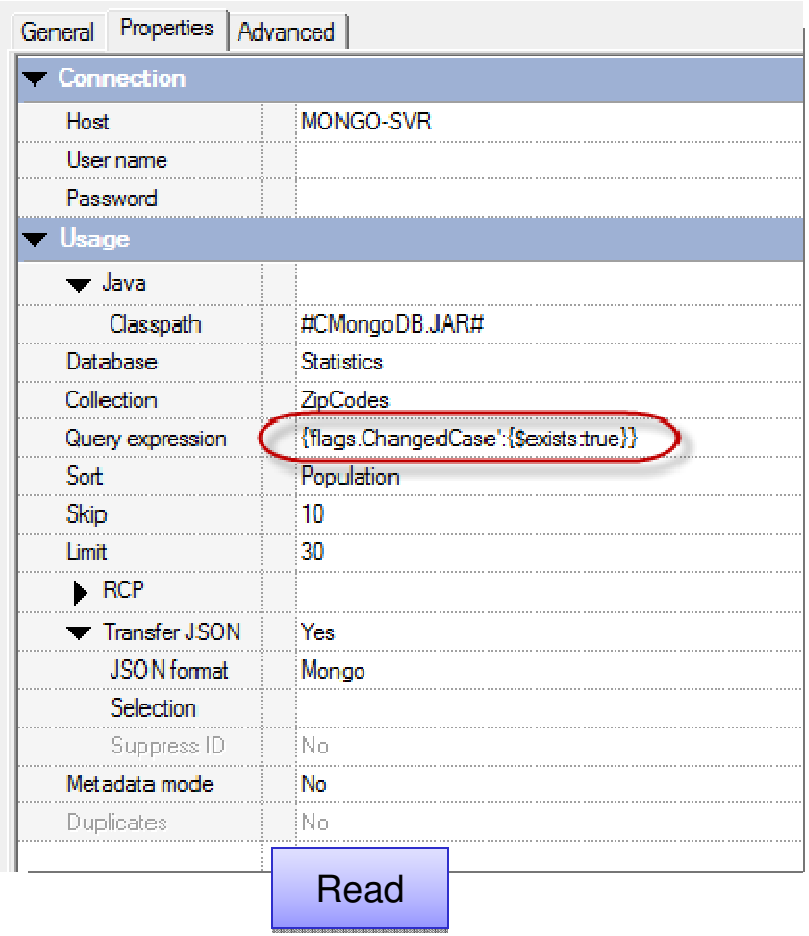
- Data transferred as “relational” or JSON, or a hybrid
- Create databases, create collections, create indexes
- Insert, Update, Replace or Delete
- Query using Mongo’s JSON syntax
- Sparse lookups
- Return metadata in jobs either as columns or JSON Schema
- Parallel reads

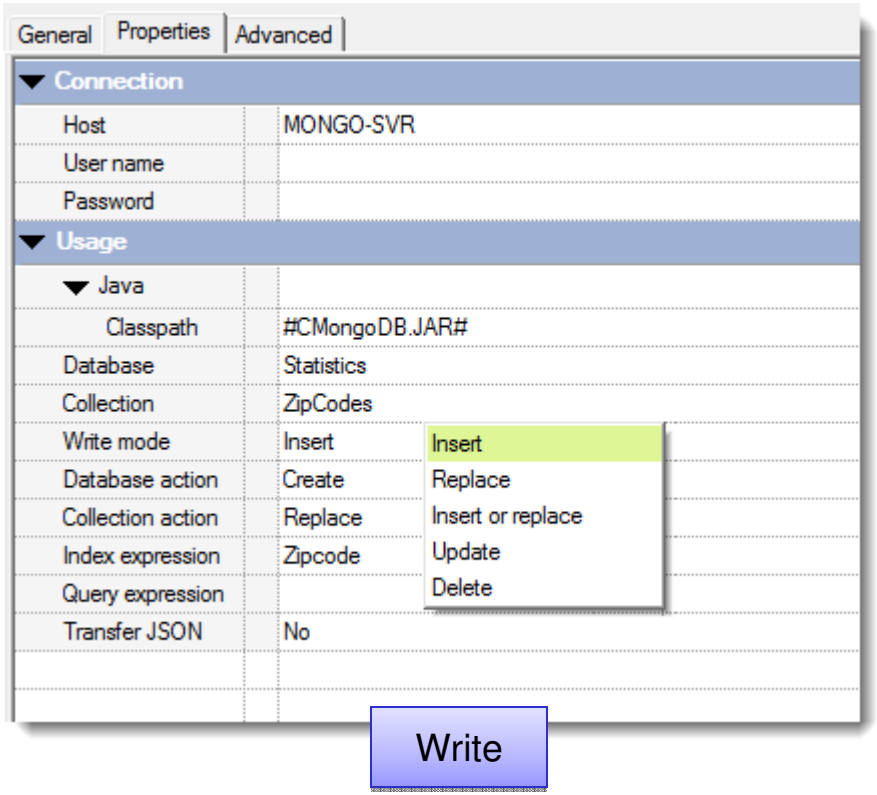






MongoDB Connector Configuration





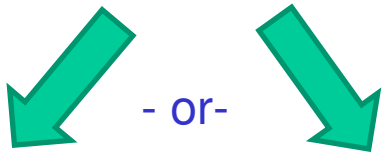
MongoDB Array Handling*

*** Coming soon…**

```
{
"type": "record",
"fields":
  [
    {"name": "a", "type": "int"},
    {"name": "b", "type": {"type":"array", "items":"int"}}
  ]
}
```

- Primitive arrays and arrays of records can be denormalized.
- Supports nested arrays

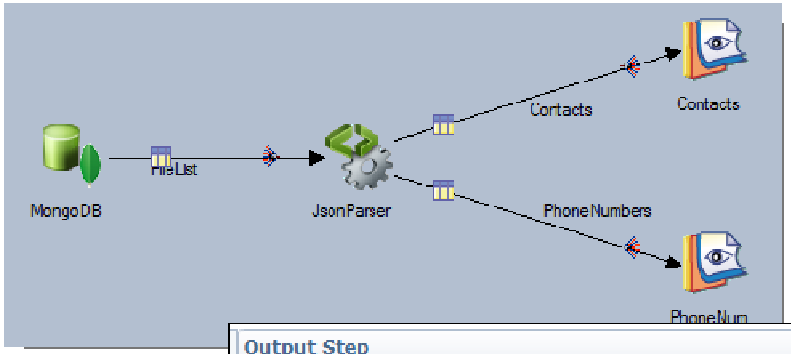
```
[{ "a": 100, "b":[1001,1002]}, { "a": 200, "b":[2001,2002]}]
```



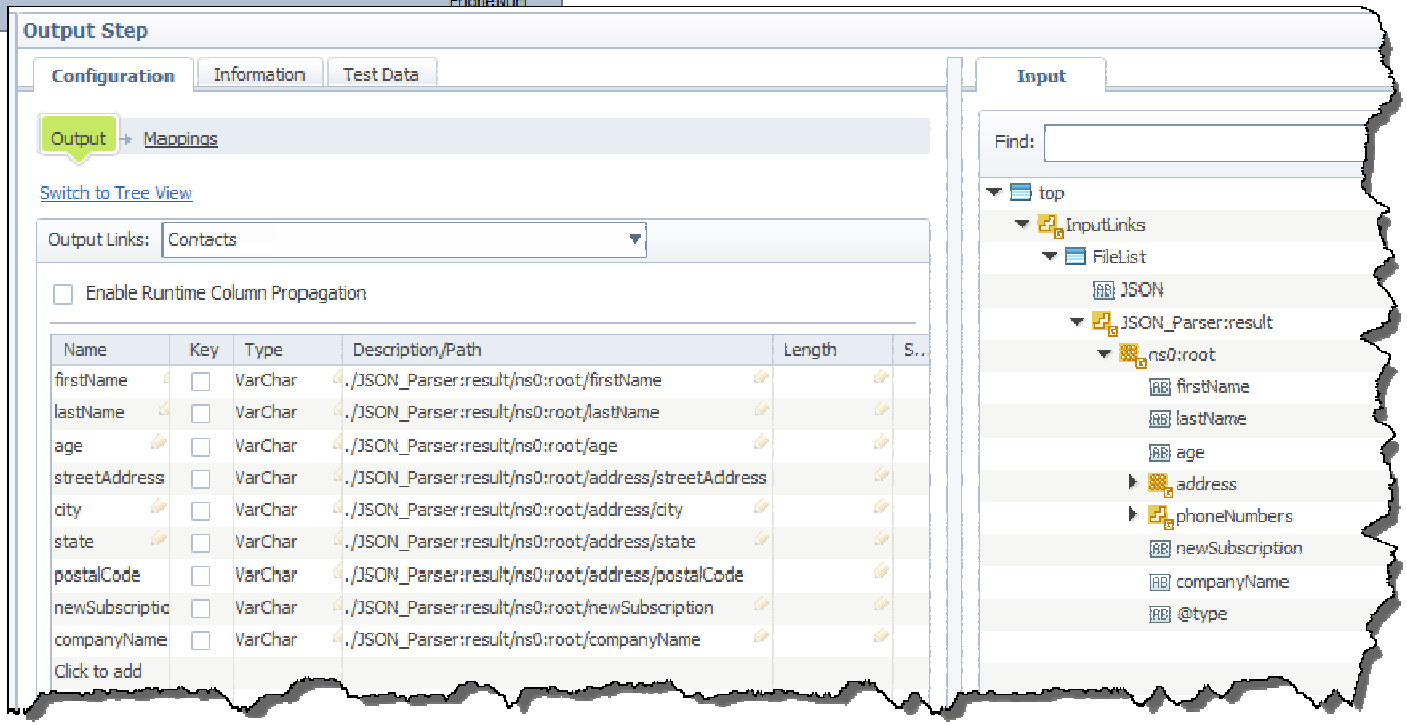

```
100 | [1001,1002]
200 | [2001,2002]
```

```
100 | 1001
100 | 1002
200 | 2001
200 | 2002
```


MongoDB Connector + JSON Stage



- JSON Stage can be used to parse or compose JSON documents.
- Use with MongoDB Connector for complex document handling.




Thank
You