Think Big: Scale Your Business Rules Solutions Up to the World of Big Data

2017

Nigel Crowther
Smarter Process Technical Lead
IBM UKI Cloud Client Technical Engagement

Decision Camp 2017

Please note

IBM’s statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM’s sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

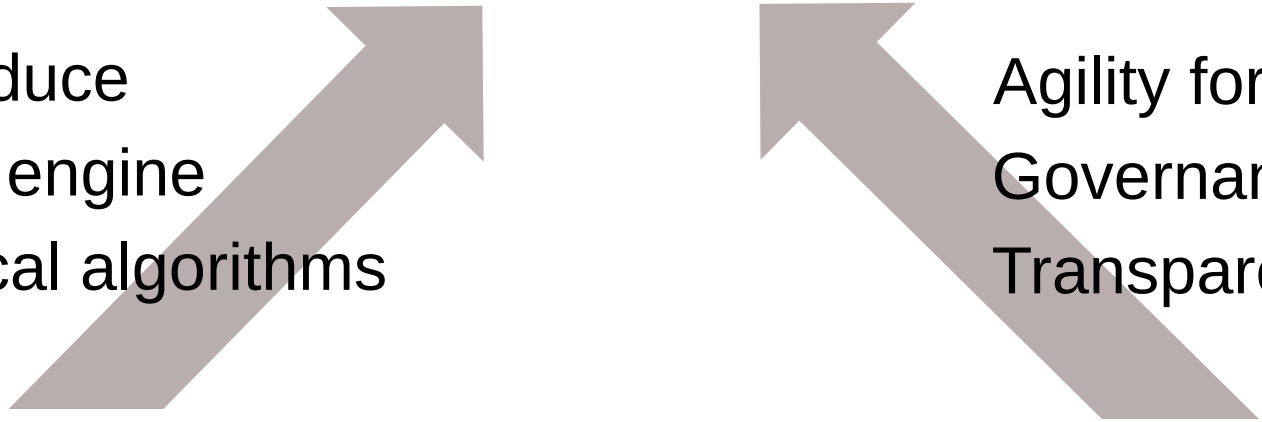
Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user’s job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

Business Rules and Big Data

- Where Business Rules fit in the World of Big Data
- Think Big Use Case - Border Control
- Business Rules Blueprints
  - Generalities
  - in Hadoop MapReduce
  - in Apache Spark
- Rule coverage, Analytics and ML

Big Data and Business Rules

Big Decision



Map/reduce

Cluster engine

Analytical algorithms

Agility for Business Users

Governance

Transparency

Big data is defined as extremely large data sets … analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior.

Google

A **BRMS** or Business Rule Management System is used to define, deploy, execute … decision logic

Wikipedia

Big Decision Use Cases at a glance

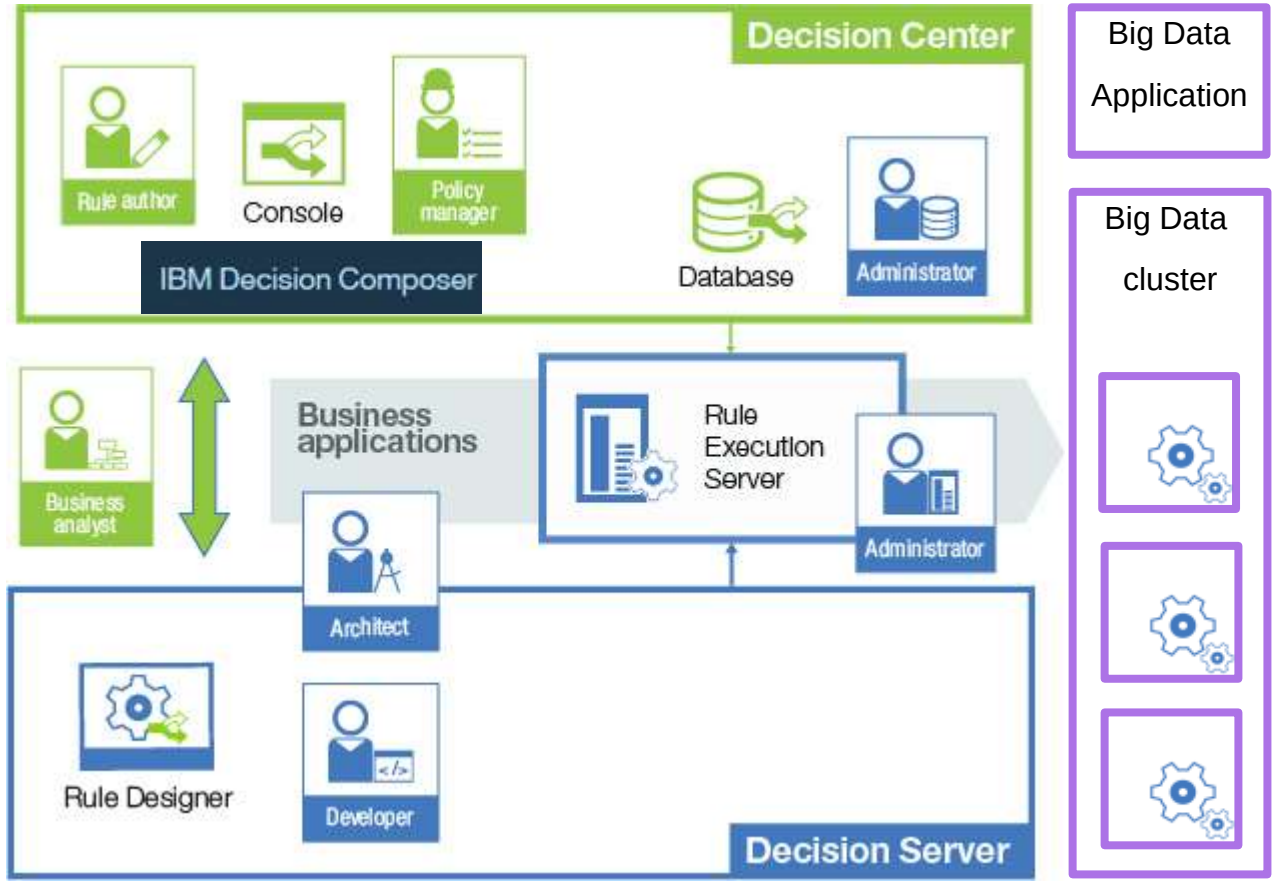
- Automate massive decision making batches
- Running business policies simulations on large historical dataset
- Detect situations on data lakes
- Invent new algorithm combinations to solve new classes of enterprise problems at scale

Enterprise Use cases

- A bank simulates new mortgage segmentation policies against ten million customers in under 30 seconds
- A credit/debit card tests new fraud detection rules on hundreds of millions of past transactions
- A financial service company brings together data science and operational decision teams to build an end to end practice and platform
- A border control agency simulates and applies profiling rules on international travelers to detect terrorists

Concept of Operations of ODM Rules in Big Data

- Rules are authored in Decision Composer, Decision Center or Rule Designer.
- Rules are versioned and deployed over HTTP(S) to a Rule Execution Server
- Big Data App fetches the latest deployed decision service
- At runtime the Big Data App applies the Decision Service against a large data set executing in parallel



Business Rules and Big Data

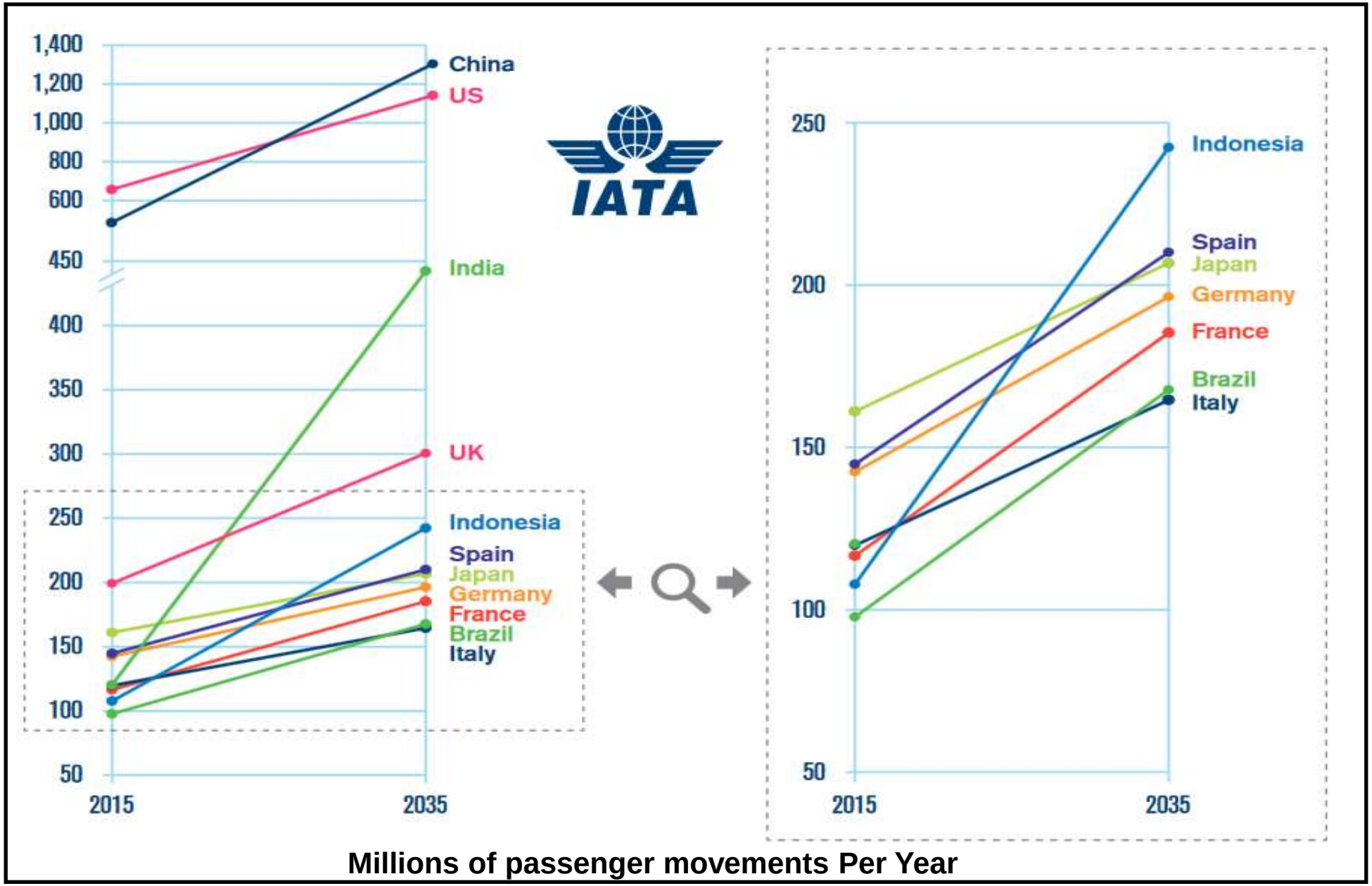
- Where ODM fits in the World of Big Data
- A Border Control use case
- Business Rules Blueprints
  - Generalities
  - in Hadoop MapReduce
  - in Apache Spark
- Rule coverage, Analytics and ML

Think Big Use Case - Border Control

Passenger travel will double from 3.8 billion to 7.2 billion in 2035. 20 Million per day.

Source: International Air Transport Association (IATA)

[IBM Case Study: The European Passenger Name Record Directive White Paper]



Use Case - Border Control

By profiling passenger data, a tiny minority can be detected and prevented from flying

Passenger Profiling Existing World

National

< 1 M passengers per day

Passenger Profiling New World

Cross Border

20 Million passengers per day

Advanced profiling

THINK BIG USE CASE : BORDER CONTROL

Bag Weight
Seat no

Advance Passenger Information (API)

Flight Date
Destination
Passport no
DOB

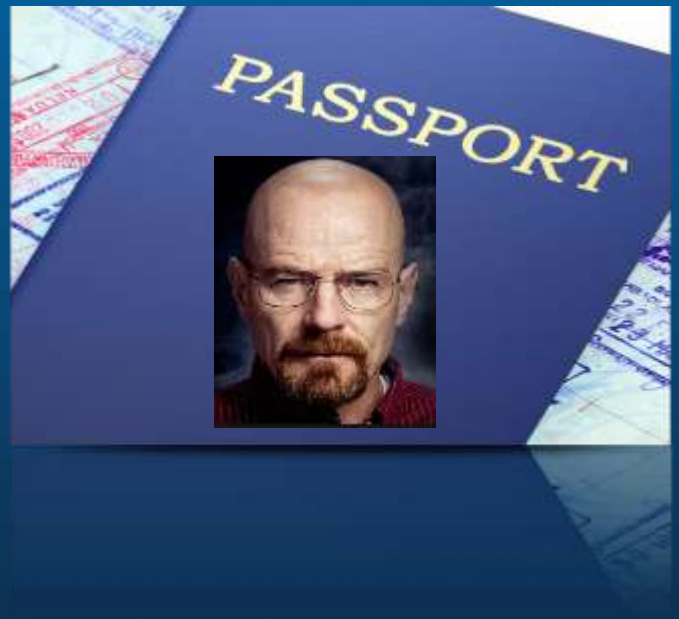
Passenger Booking Record (PNR)

ODM
Big Data

Risk Score

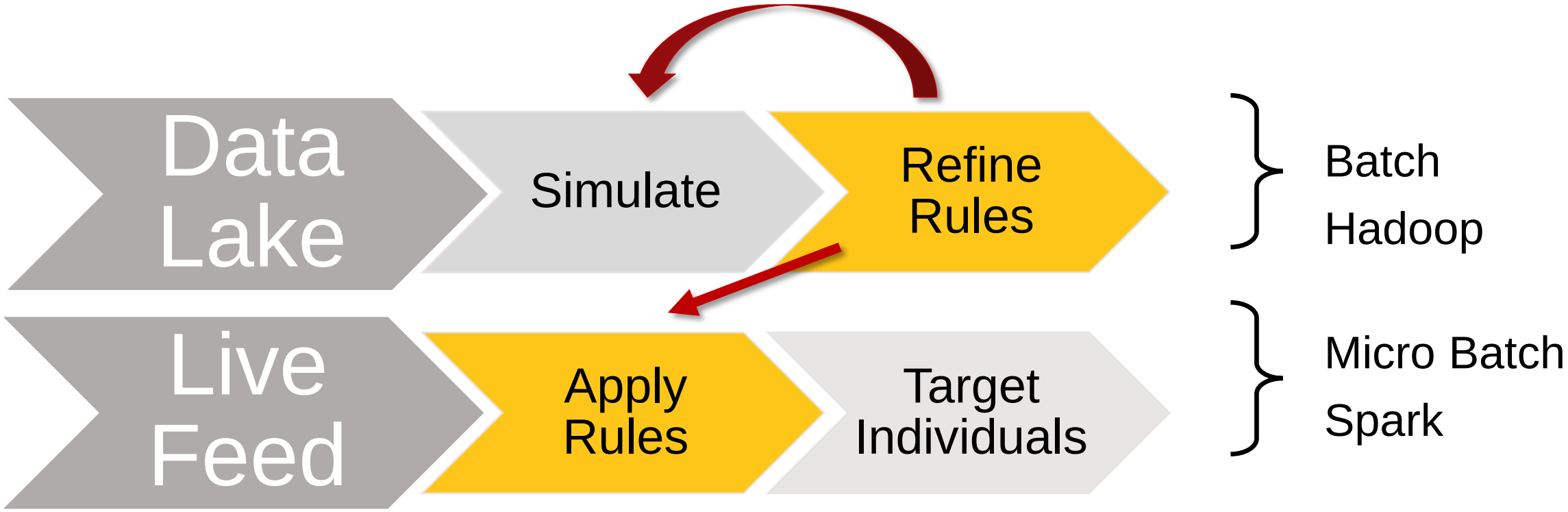
Social Media
Government feeds

Check in



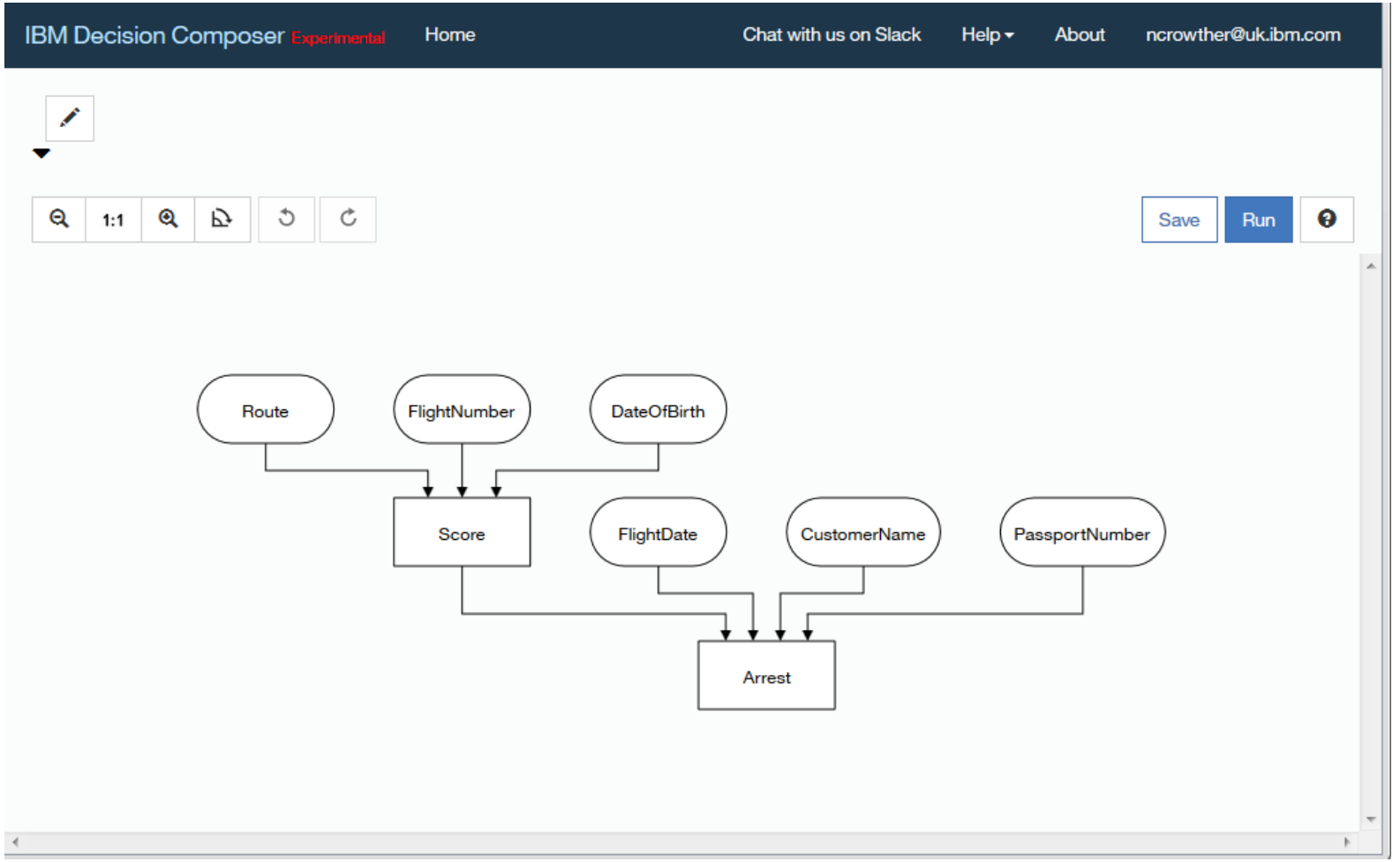
Hadoop Enhances Conventional Architecture

Same rules used against data lake and live feed



The DMN Model - Decision Composer

- Decision Composer is a new experimental tool to create rules
- Uses DMN (Decision Modelling Notation) to design and model your decisions
- Build and deploy from the tool directly to Bluemix runtime
- Good for rapid prototyping and simple rulesets



Example Stateless Rules

```
if
    tweet contains "crystal meth"
    and the age from 'date of birth' is between 18 and 30
then
    set score to score + 1 ;
```

Ben Smith LASMEX 16/08/1990 ME652 22/03/2017 F607631362

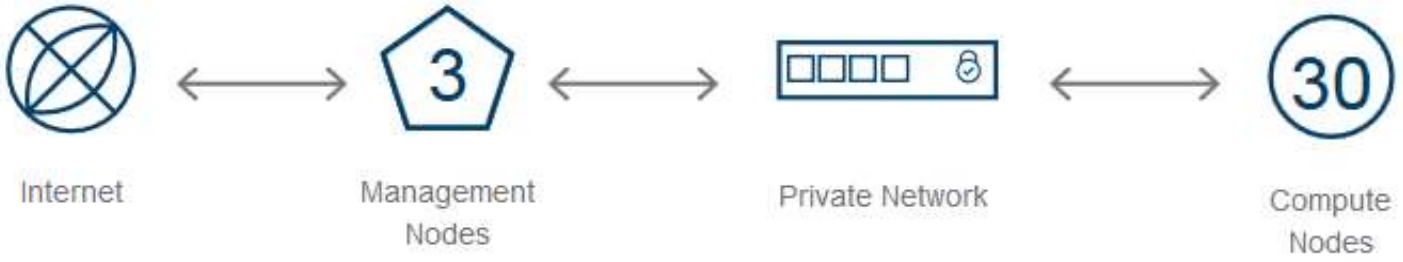
```
if
    'passport number' is one of {"U468924610", "F607631362"}
then
    set response to response + ",Passenger: " +
        'customer name' + " on watch list.  Flight " +
        'flight number' + " flying at " + 'flight date';
```

Example Stateful Rules

```
F607631362 Walter White 16/08/1959 MA652 01/01/2017 LAXMEX
F607631362 Walter White 16/08/1959 MA445 22/03/2017 MEXLAX
```

	Age		Months Away	Score	Route	Response Code
	min	max				
1	16	60	≥ 1	> 5	LAXMEXMEXLAX	D254
2	12	50	≥ 2	> 10	LHRDAMDAMLHR	T023
3	16	60	≥ 1	> 5	LHRAMSAMSLHR	D345

Let’s Create a Hadoop Super Computer on Bluemix!



Management Nodes	
CPU (# of cores)	**24 Cores** 2 X 12 = 24 Cores (2690v3)
RAM	**256 GB** 16 X 16 GB = 256 GB
OS disk	**8 TB** 4 X 4 TB = 16 TB (RAID 10)
Network	**10 GB**
Compute Nodes	
CPU (# of cores)	**24 Cores** 2 X 12 = 24 Cores (2690v3)
RAM	**256 GB** 16 X 16 GB = 256 GB
Data disk	**32 TB** 8 X 4 TB = 32 TB
OS disk	**8 TB** 4 X 4 TB = 16 TB (RAID 10)
Network	**10 GB**

Performance

PNR Validation on BigInsights Apache Hadoop on Bluemix.

One Day, 20 Million PNRs :

- 3 compute nodes: *2min 46secs* **(120,000 per second)**

One Year, 7.2 Billion PNRs :

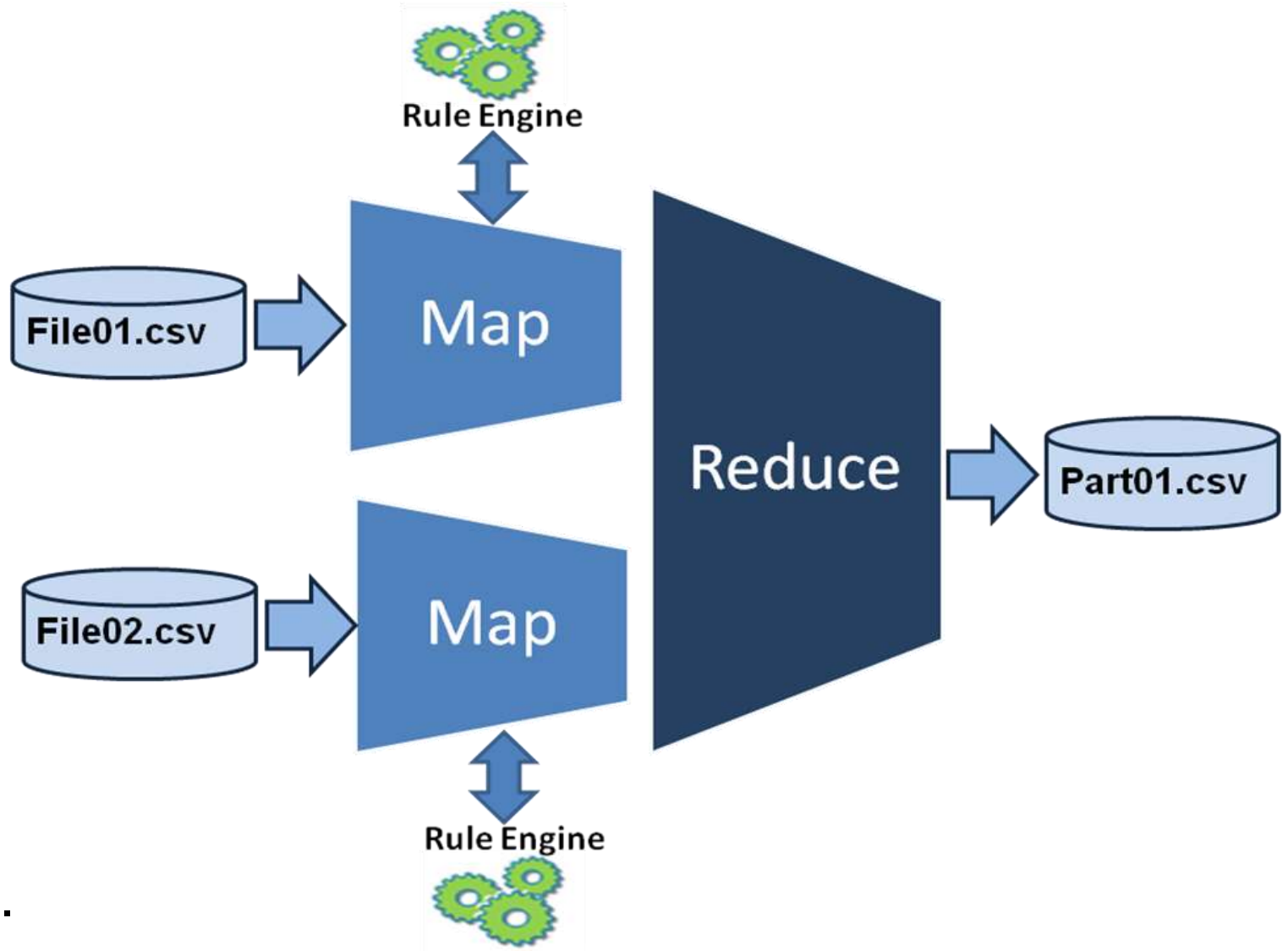
- 30 compute nodes: *1.5 hours* **(1.2M per second)**

Business Rules and Big Data

- Where ODM fits in the World of Big Data
- Think Big Use Case - Border Control
- Big Data integration blueprints
  - landscape
  - in Hadoop MapReduce
  - in Apache Spark
- Rule coverage, Analytics and ML

Call a Local Rule Engine in Hadoop

- Each Map job is given a part of the data (the split)
- The Map sends the split to an instance of the rule engine where it is processed.
- The Rule Engine can either be embedded within the Map job, or called externally.
- Data created by the rules are combined by the Reduce jobs.



Calling the Bluemix Business Rules Service from Hadoop

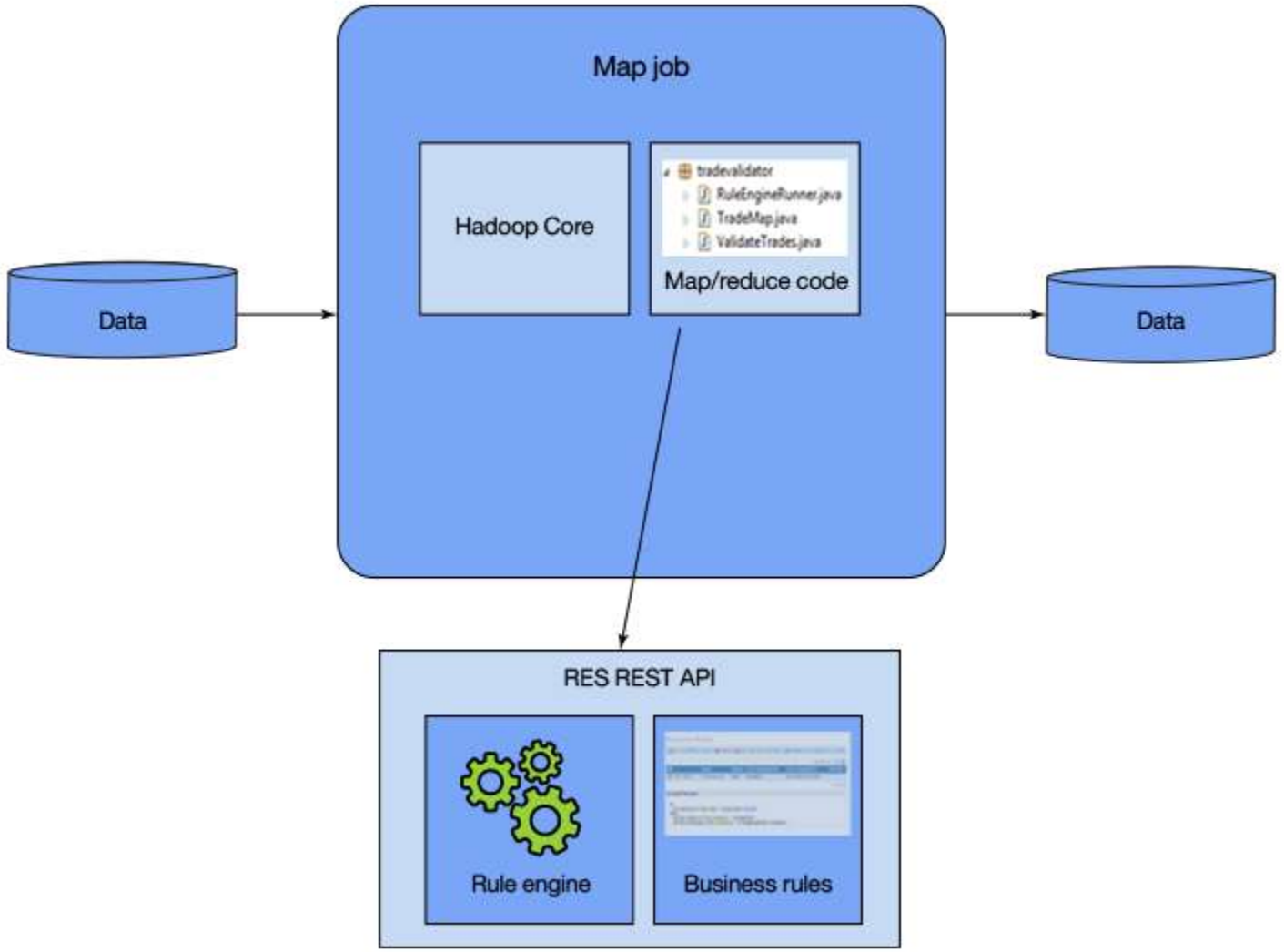
The Rule Engine is executed via a REST API external to the Map Job.

Advantages:

- Unleashes multi-threading capability of RES to handle parallel invocations from multiple map jobs
- No need to rebuild Hadoop job for each rule change
- Versioning and management of rules managed within RES
- Licencing managed by RES
- Works well with Bluemix and cloud solutions.

Disadvantages:

- Serialization and remoting penalty



Execute with a local Rule Engine

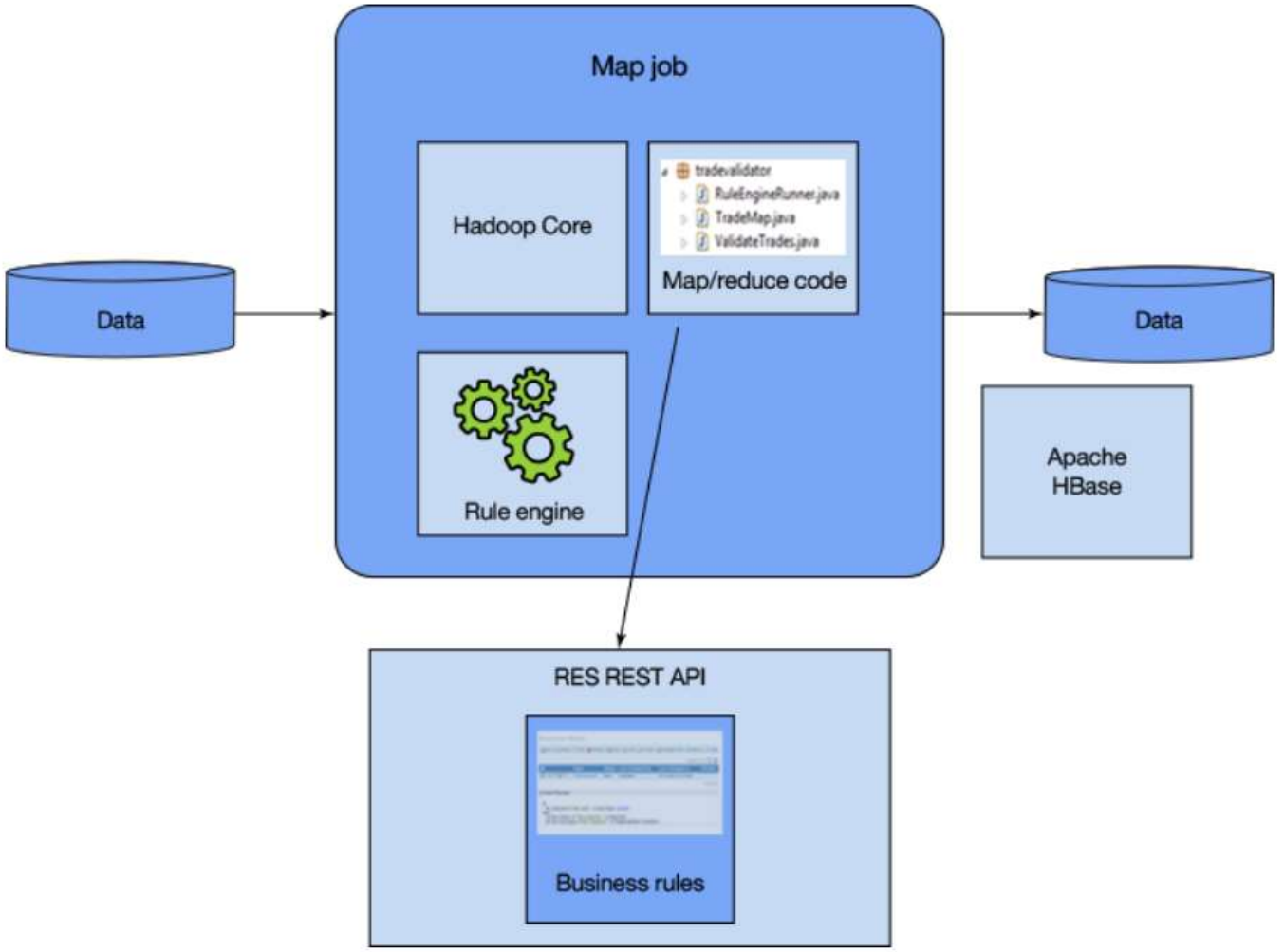
The REST API extracts the latest version of the ruleset from the RES. The ruleset is executed against an embedded engine in the Map Job.

Advantages:

- Versioning and management of rules within RES
- No need to rebuild Hadoop executable for each rule change
- Embedded engine gives high performance
- Can leverage full Hadoop stack – e.g. Hbase

Disadvantages:

Embedded engine for each Hadoop job requires careful management of PVU costs.

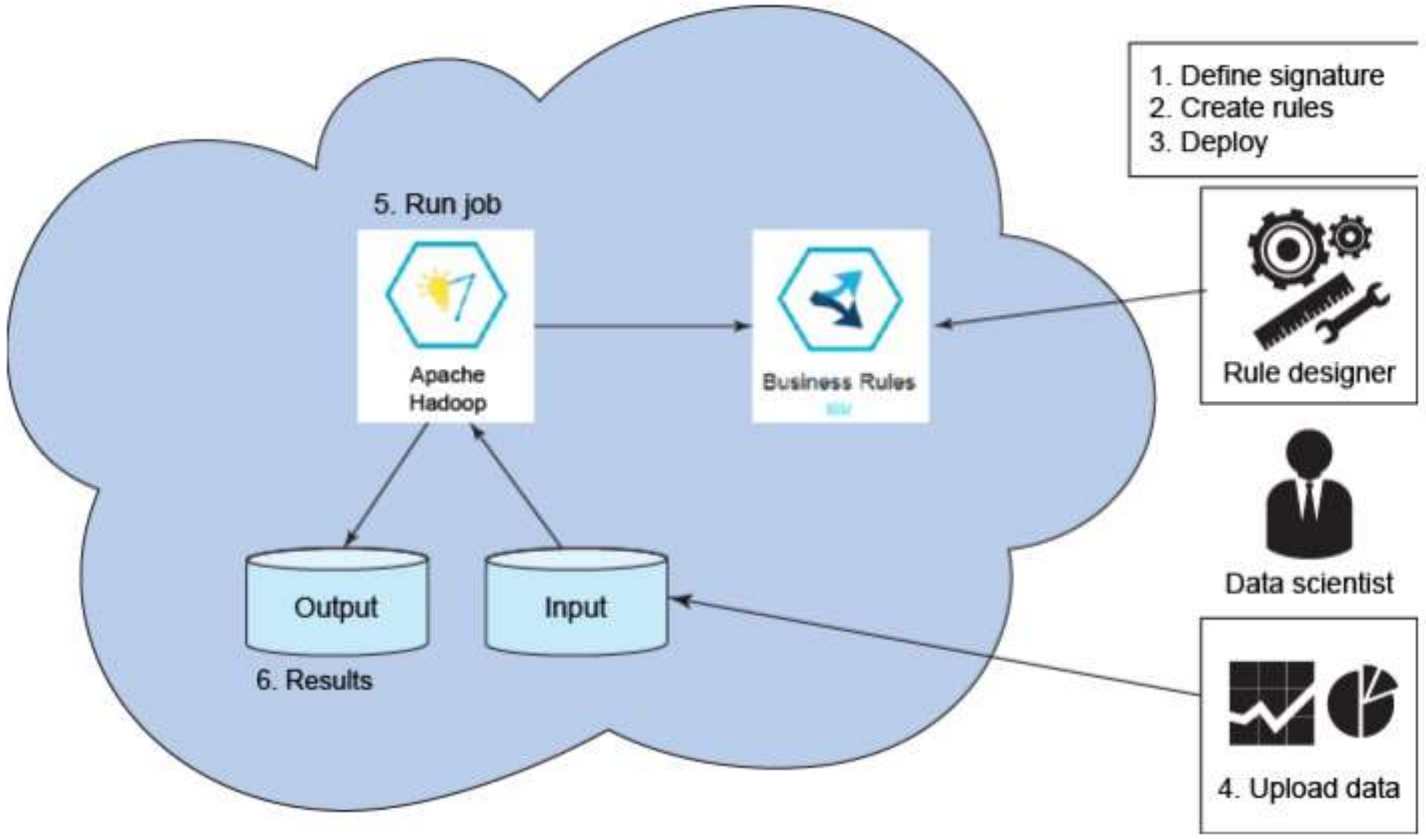


ODM/Hadoop Asset

Integration of ODM and Hadoop provided as a free asset:

1. Define ruleset signature
2. Create rule service
3. Deploy
4. Upload data
5. Configure and run job
6. Examine results

[Think Big! Developerworks article]



Think Big! Developer works Article

Think big! Scale your business rules solutions up to the world of big data

Build an app that uses Business Rules and Apache Hadoop services on IBM Bluemix

Nigel Crowther

Published on November 12, 2014 / *Updated: March 16, 2017*

0

±

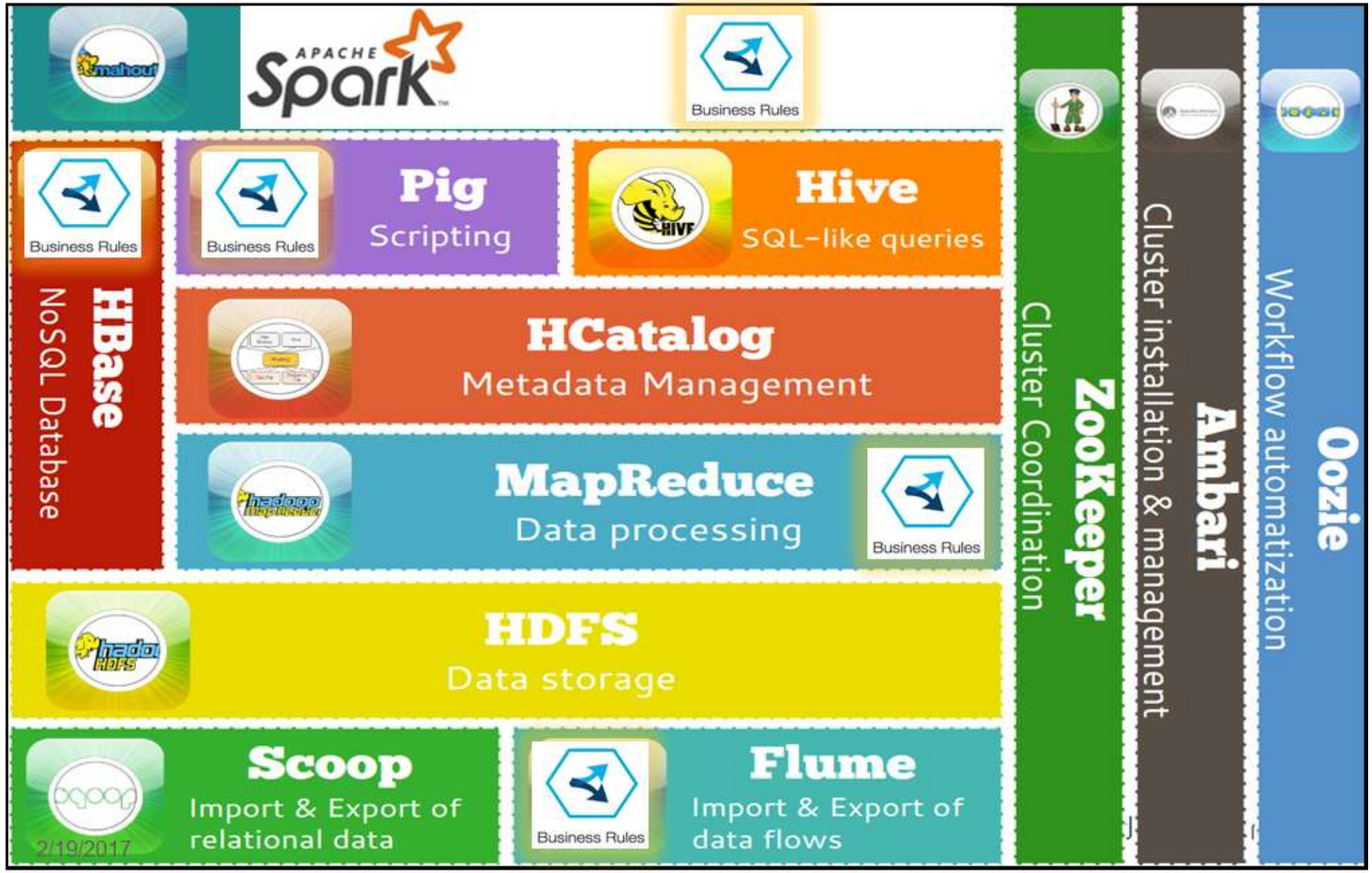
Business Rules and Big Data

- Where ODM fits in the World of Big Data
- Think Big Use Case - Border Control
- Big Data integration blueprints
  - landscape
  - in Hadoop MapReduce
  - in Apache Spark
- Rule coverage, Analytics and ML

Where ODM Rules fits in Big Data OSS ecosystem

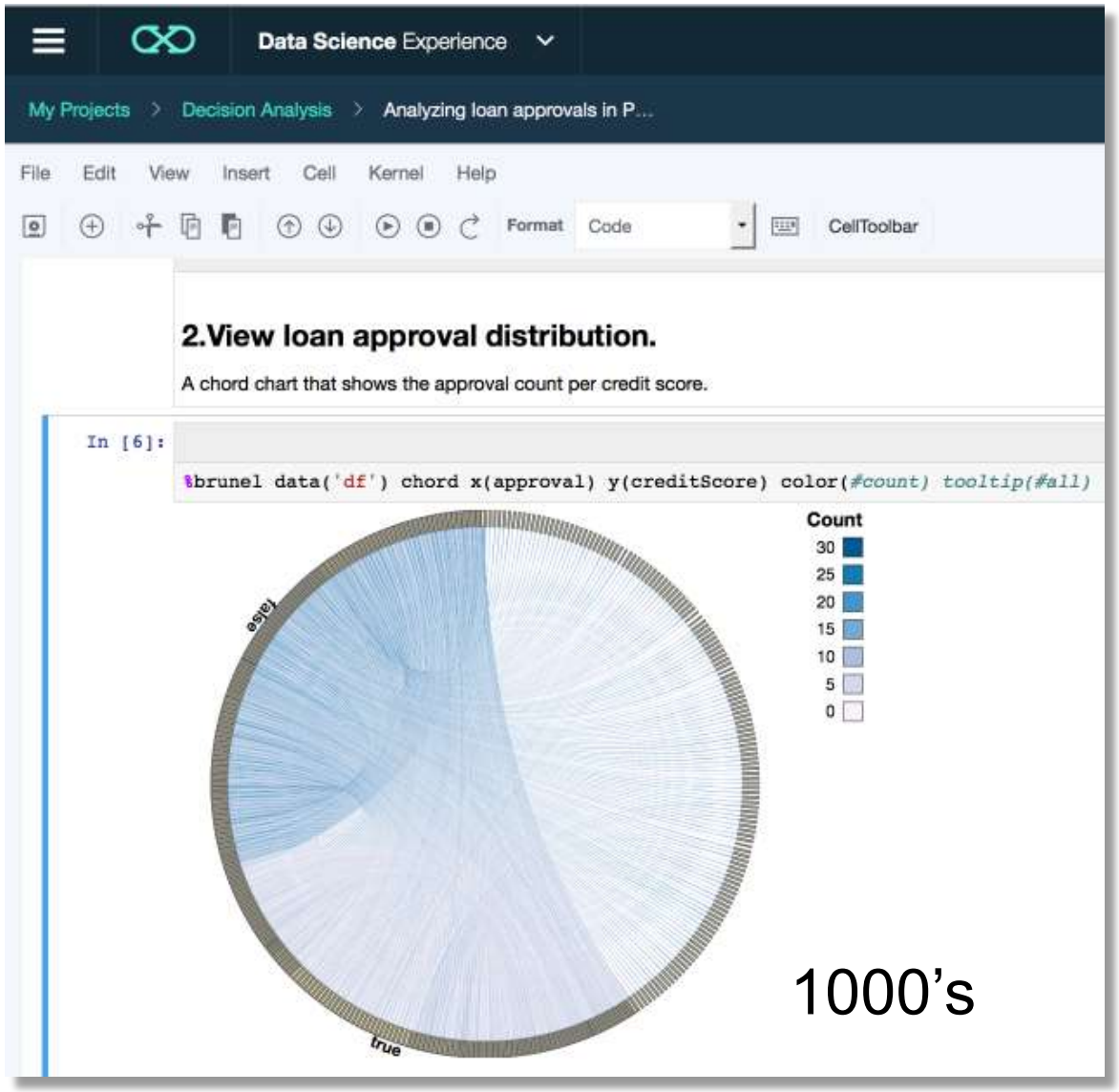
Run ODM within:

- Apache Spark
- Hadoop map/reduce
- Flume



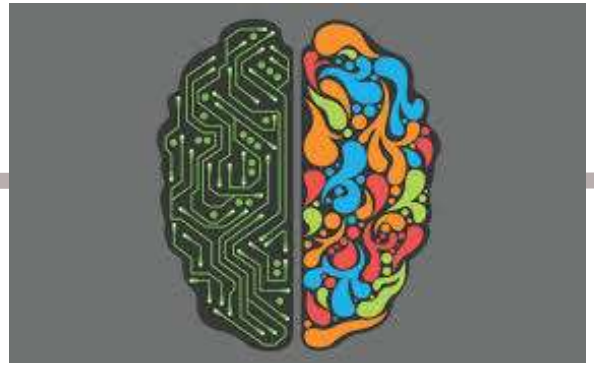
Business Rules in Data Science Experience

- Apply Big Data and Business rules to determine loan approval rates



Rules & Machine Learning

Business Rules



IBM Watson

AI

- Big Data aggregation
- ‘Business Logic’ Intelligence
- Structured data
- Formalized model with facts
- Causality
- Mainly Boolean logic
- Reasoning

- Fuzzy logic based on reasoning
- Unstructured data
- Signal processing
- Correlation
- Dealing with uncertainty
- Perception, Classification, Regression

Example: Fraud detection in filed company financial reports

Wrap up

- Combine Today business rules and Big Data into Big Decision in Hadoop and Apache Spark
- Detect situations on data lakes
- Join Data Scientists and decision management teams together
- Automate massive decision making in standard compute grids
- Running simulations on large historical dataset with parallel metric and KPI computation
- Invent new business rule algorithm combinations to solve new classes of enterprise AI at scale

References

- ODM on Hadoop
  - https://www.ibm.com/developerworks/bpm/library/techarticles/1411_crowther-bluemix/1411_crowther.html
- ODM on Spark article
  - https://developer.ibm.com/odm/docs/solutions/odm-and-analytics/odm-business-rules-with-apache-spark-batch-operations/
- Bluemix
  - https://console.ng.bluemix.net/catalog/services/apache-spark
  - https://console.ng.bluemix.net/catalog/services/biginsights-for-apache-hadoop
- Data Science Experience
  - http://datascience.ibm.com/

Q & A

Notices and disclaimers

U.S. Government Users Restricted Rights — use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted according to the terms and conditions of the agreements under which they are provided.

IBM products are manufactured from new parts or new and used parts. In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer is in compliance with any law.

Notices and disclaimers continued

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a particular**, **purpose.**

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.