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

July 11th 2017

Pierre FEILLET – IBM ODM Product Architect

Nigel Crowther – IBM ODM Senior Consultant

Business Rules and Big Data

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

Big Data, Business Rules and ML

Big Decision

Map/reduce
Cluster engine
Analytical algorithms
Machine Learning

Agility for Business Users
Governance
Transparency
Reasoning

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

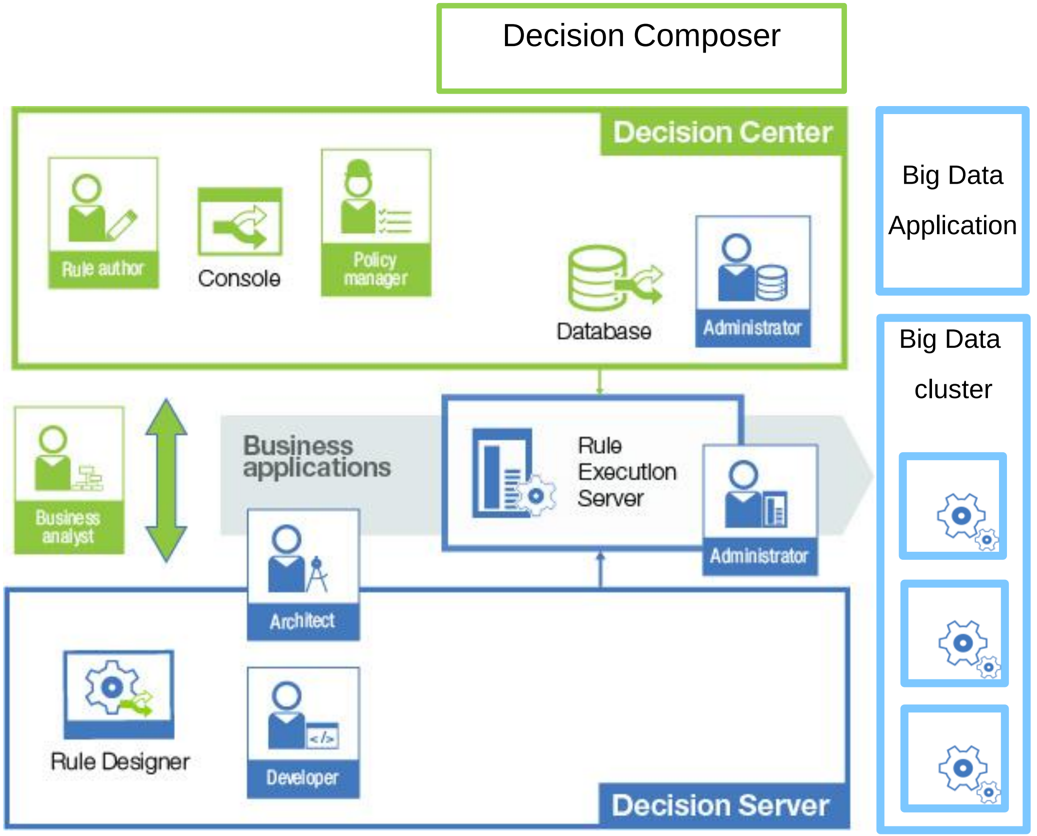
- Automate massive decision making batches
- Running business policies simulations on large historical dataset
- Detect situations on data lakes
- Invent new algorithm combinations to solve new classes of enterprise problems at scale

Enterprise Use cases

- A bank simulates new mortgage segmentation policies against ten million customers in under 3 minutes
- A credit/debit card tests new fraud detection rules on hundreds of millions of past transactions
- Recommend products based on regulations and client history
- 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 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

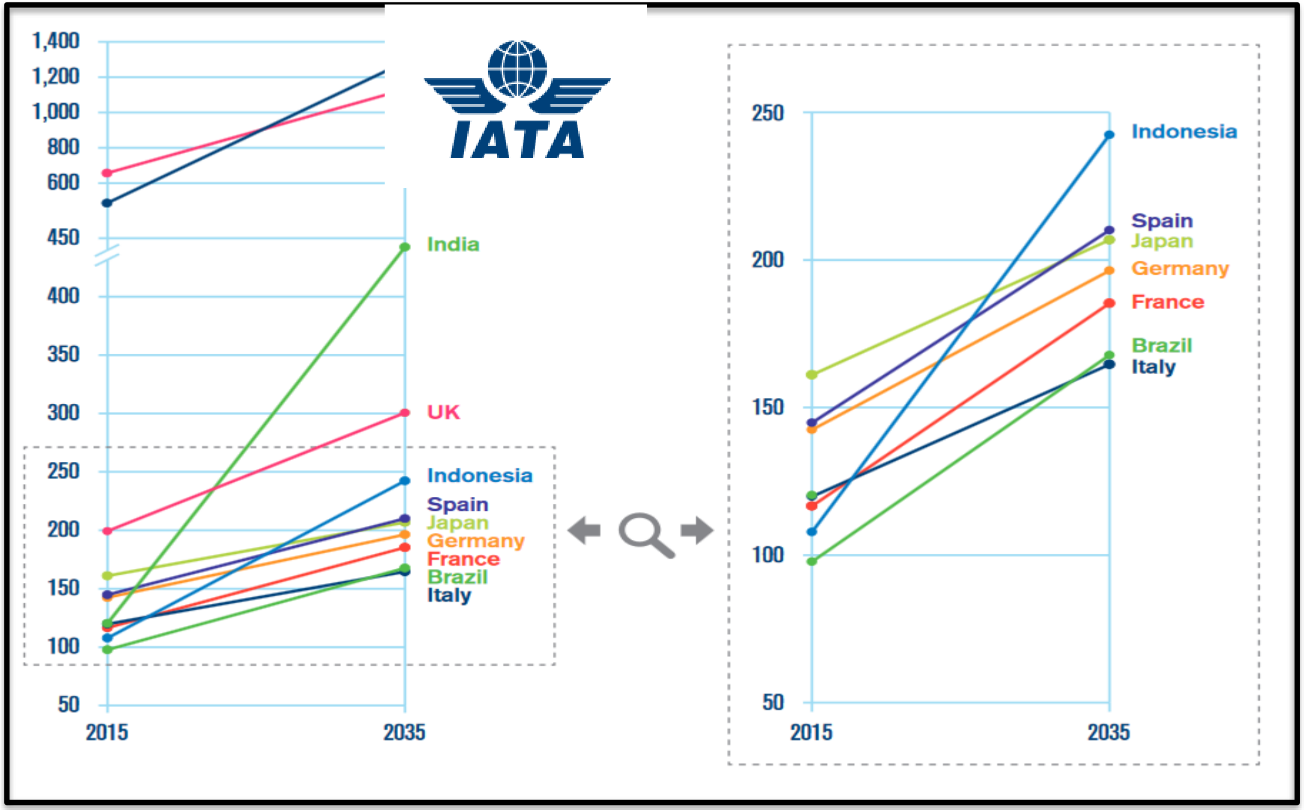
- Where Business Rules fit in the World of Big Data
- Think Big Use Case - Border Control
- Business Rules Blueprints
  - Generalities
  - in Hadoop MapReduce
  - in Apache Spark
- Rule coverage, Analytics and ML

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]



Millions of passenger movements Per Year

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

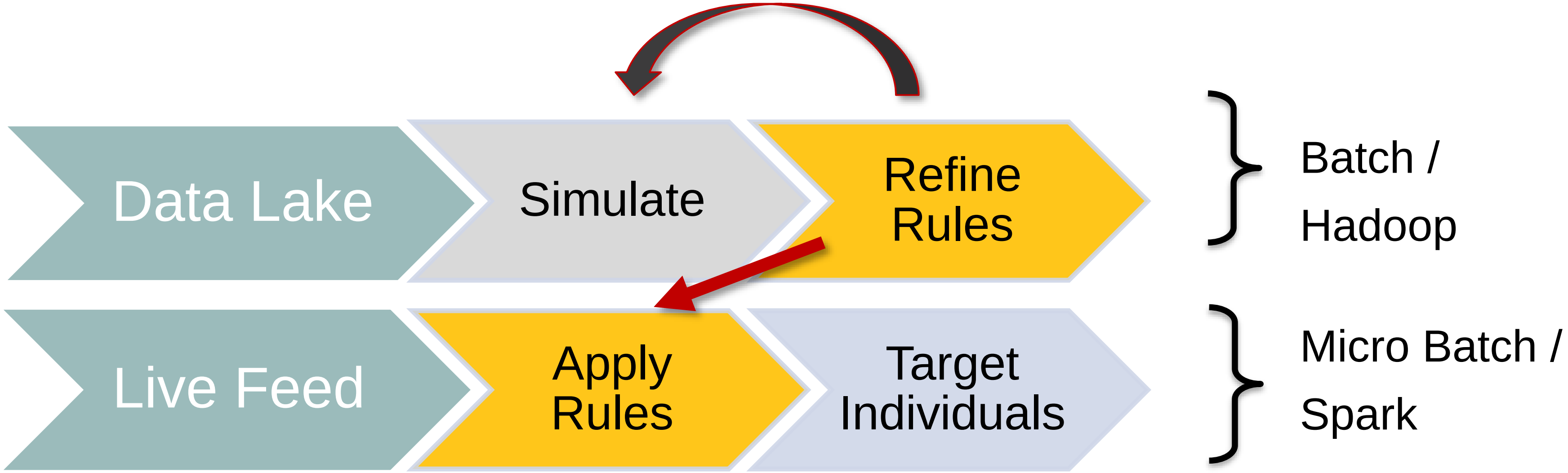
Check in

Risk Score

Social Media
Government feeds

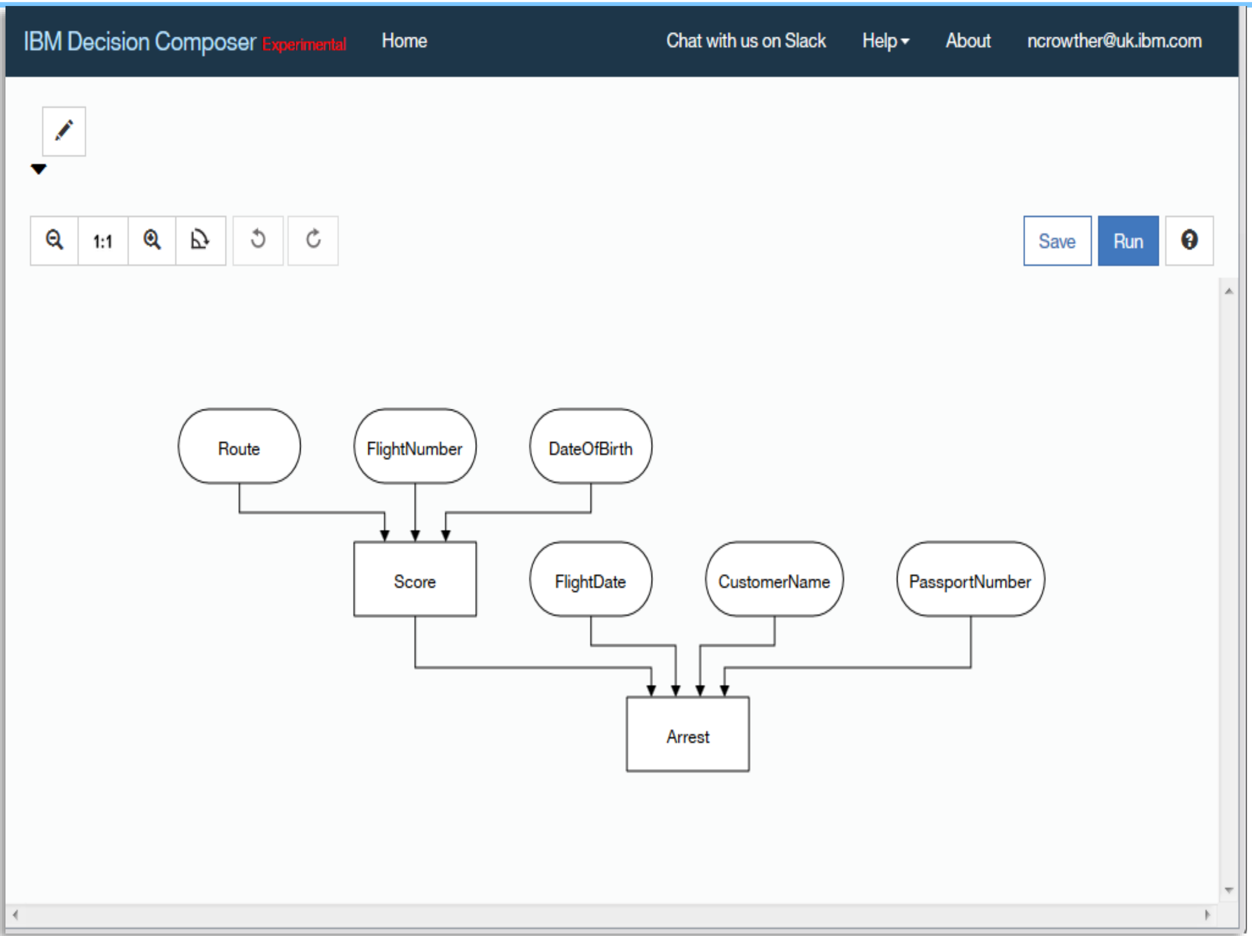
Hadoop Enhances Conventional Architecture

Same rules used against data lake and live feed



The DMN Model - Decision Composer

- Decision Composer is an experimental tool to generate rule projects
- Uses DMN (Decision Modelling Notation) to model decisions
- Build and deploy directly to Bluemix runtime
- Good for rapid prototyping and simple rulesets



Example of Stateless Rules


```
if
    tweet contains "#DRUGS"
    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 of Stateful Rules

```
F607631362 Walter Black 16/08/1959 MA652 01/01/2017 LAXMEX
F607631362 Walter Black 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

Business Rules

HBase

NoSQL Database

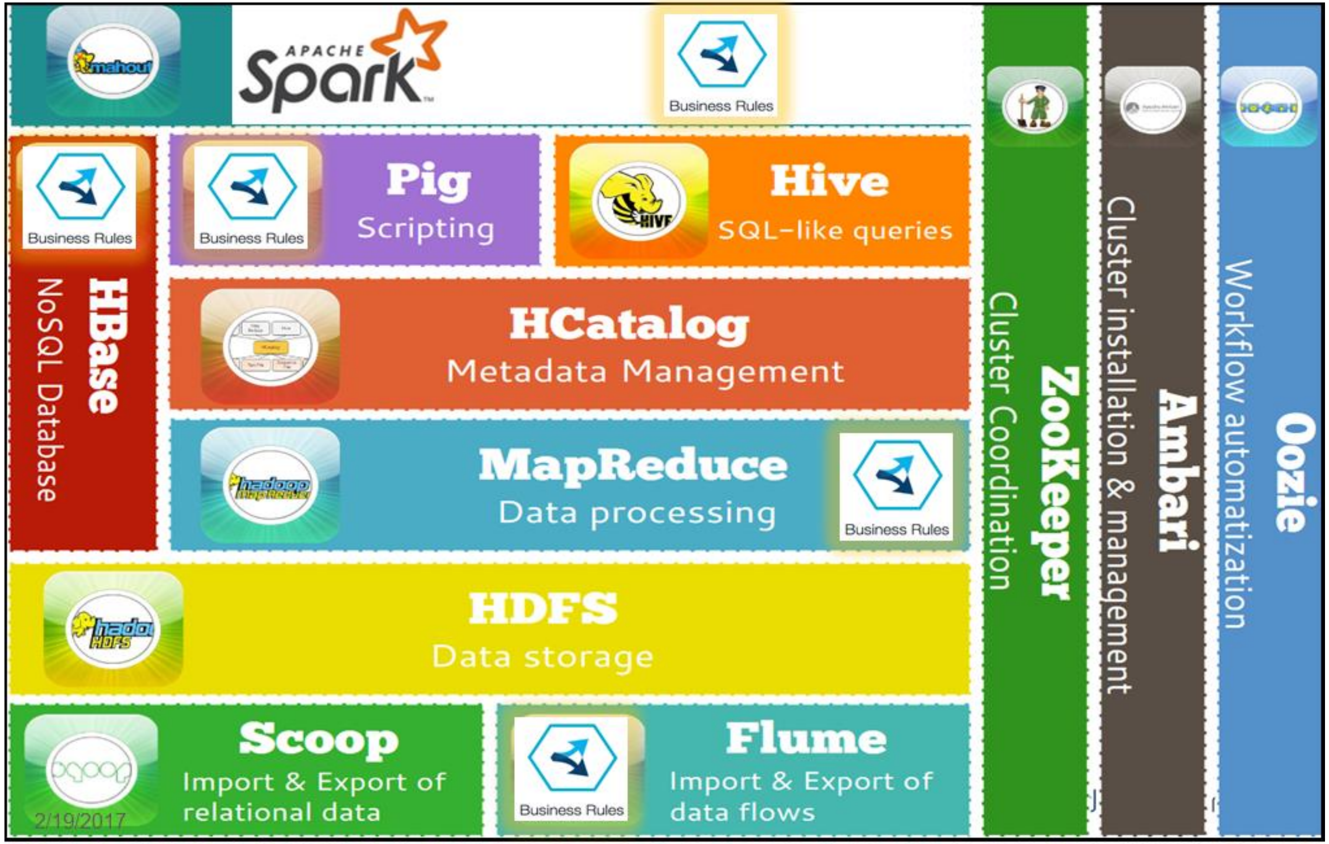
Business Rules and Big Data

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

Landscape: Where ODM Rules fits in Big Data

Run ODM within:

- Apache Spark
- Hadoop map/reduce
- Flume

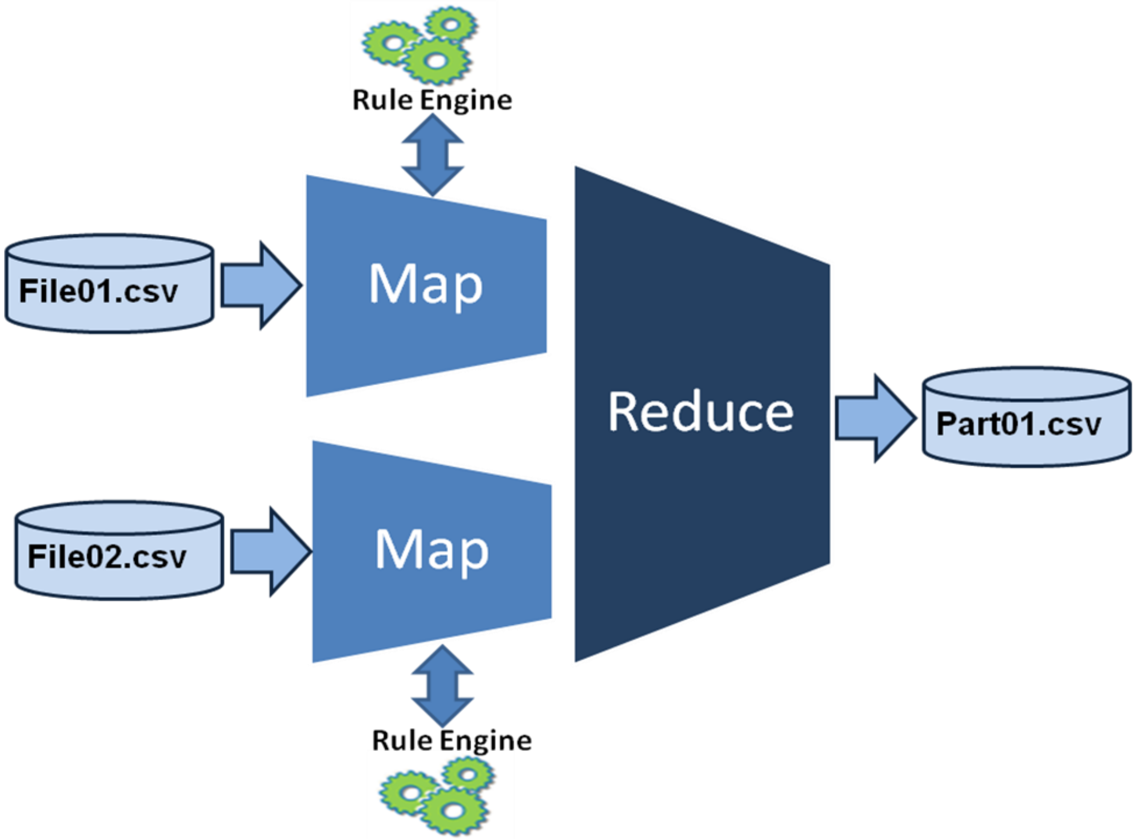


Business Rules and Big Data

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

Rule Engine Integration 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.



Execute with a local Rule Engine, Remote RES

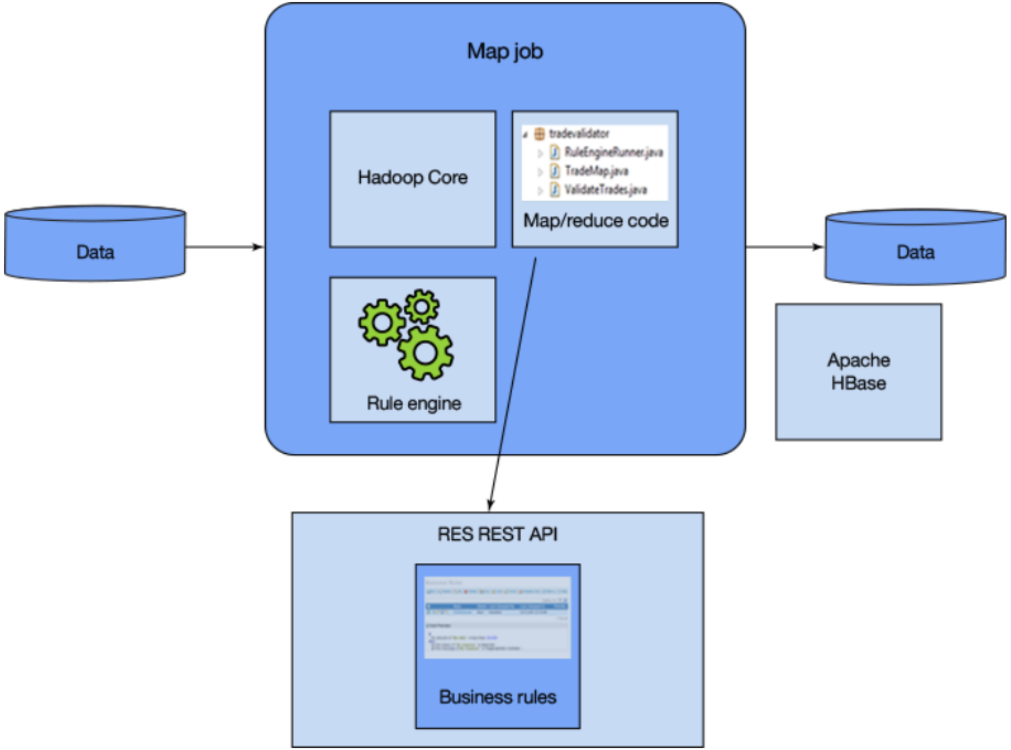
REST API extracts latest ruleset from RES. Ruleset executed against embedded engine in Map Job.

Advantages:

- Versioning of rules within RES
- Avoid rebuilding Hadoop executable
- Embedded engine gives high performance
- Leverage full Hadoop stack – e.g. Hbase

Disadvantages:

Need to manage engine license costs

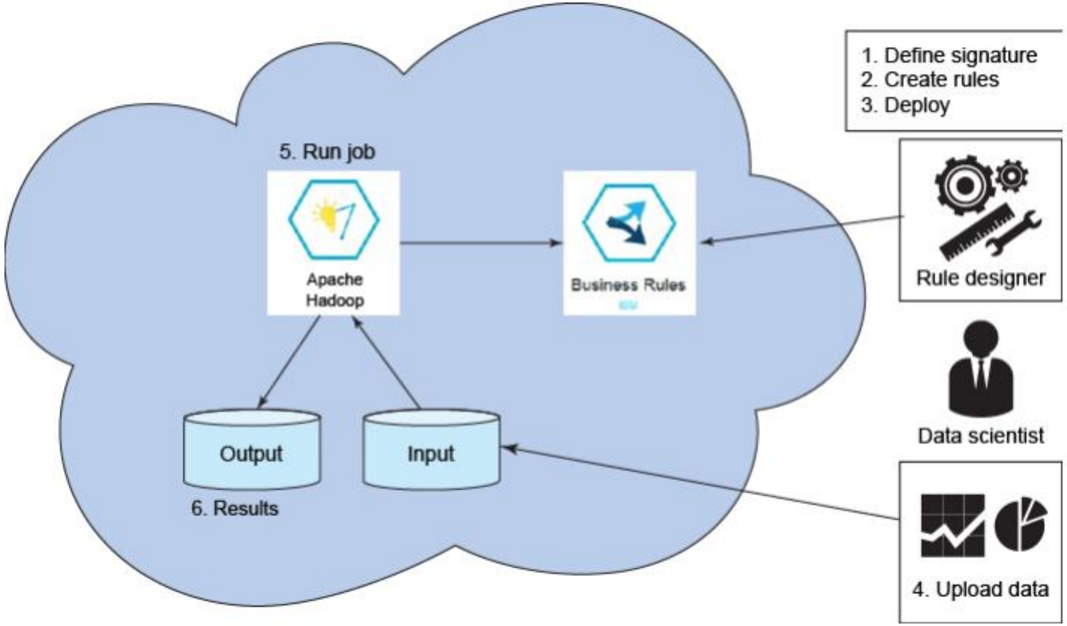


ODM/Hadoop Asset

Integration of ODM and Hadoop provided as a free asset:

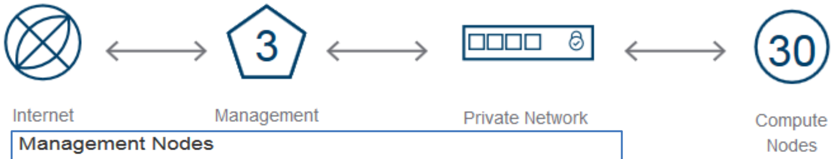
1. Define ruleset signature
2. Create rule service
3. Deploy
4. Upload data
5. Configure and run job
6. Examine results

Think Big! Developerworks article



Let's Create a Hadoop Super Computer on Bluemix!

Provisioning



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.

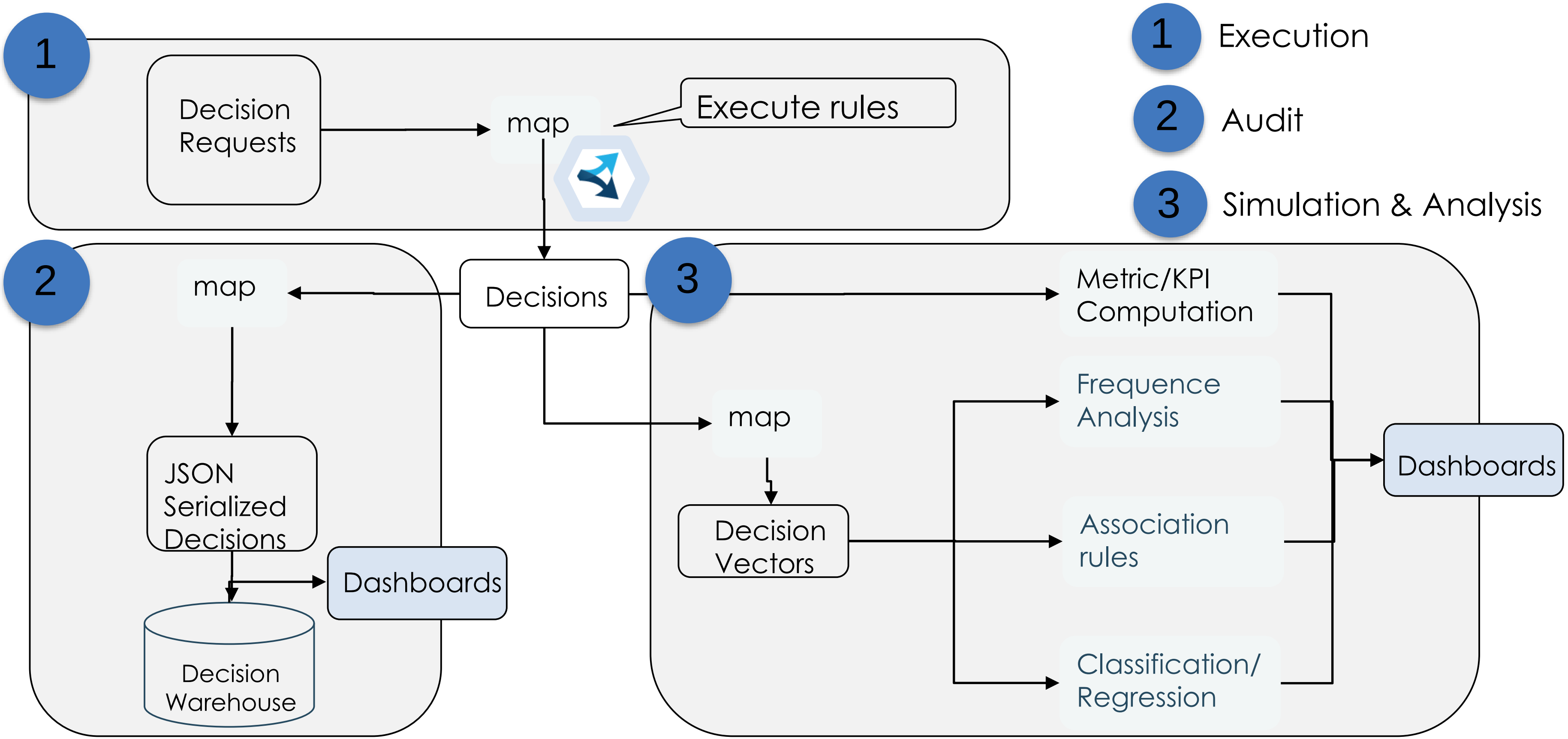
One Day, 20 Million PNRs :
- 3 compute nodes: *2min 46secs*
- **120,000 TPS**

One Year, 7.2 Billion PNRs :
- 30 compute nodes: *1.5 hours*
- **1.2M TPS**

Business Rules and Big Data

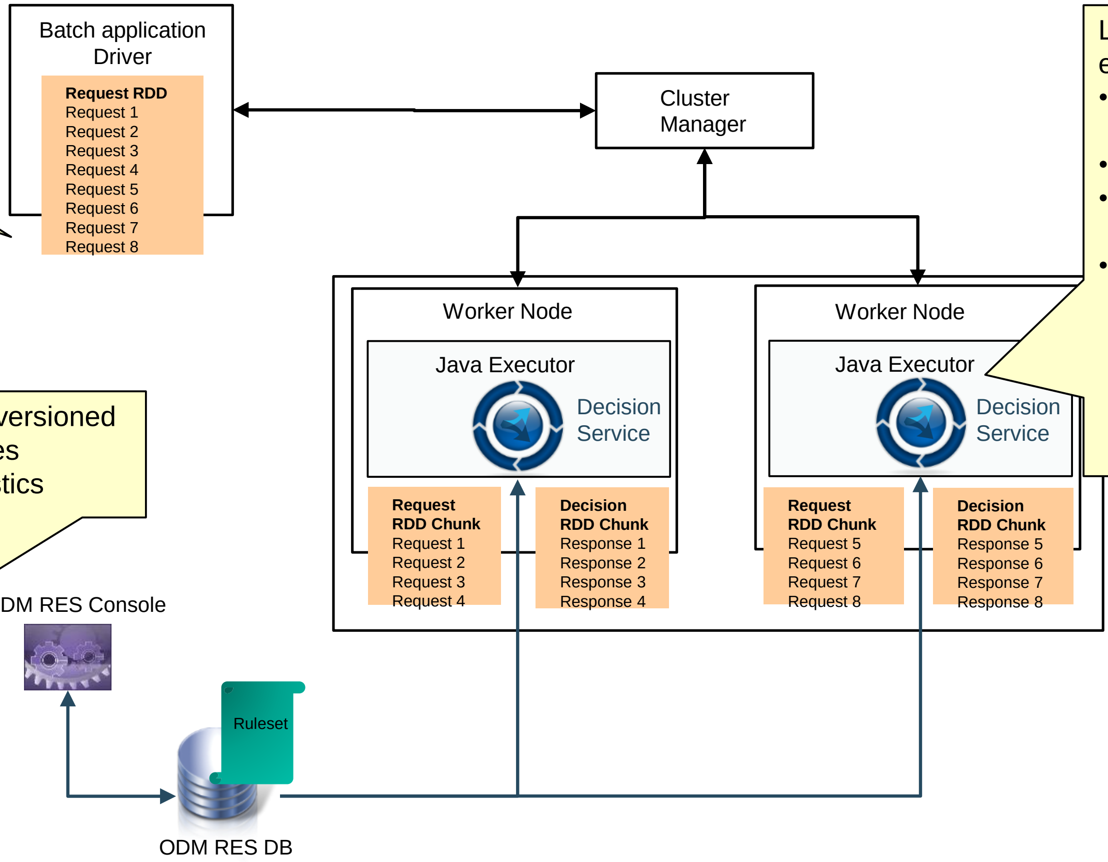
- Where Business Rules fit in the World of Big Data
- Think Big Use Case - Border Control
- Business Rules Blueprints
  - Landscape
  - in Hadoop MapReduce
  - in Apache Spark
- Rule coverage, Analytics and ML

Automate, Audit and Analyze your decision making in Apache Spark



Running a Decision service in an Apache Spark cluster

Map a Resilient Distributed Dataset (RDD) to the grid



Local Decision Service execution:

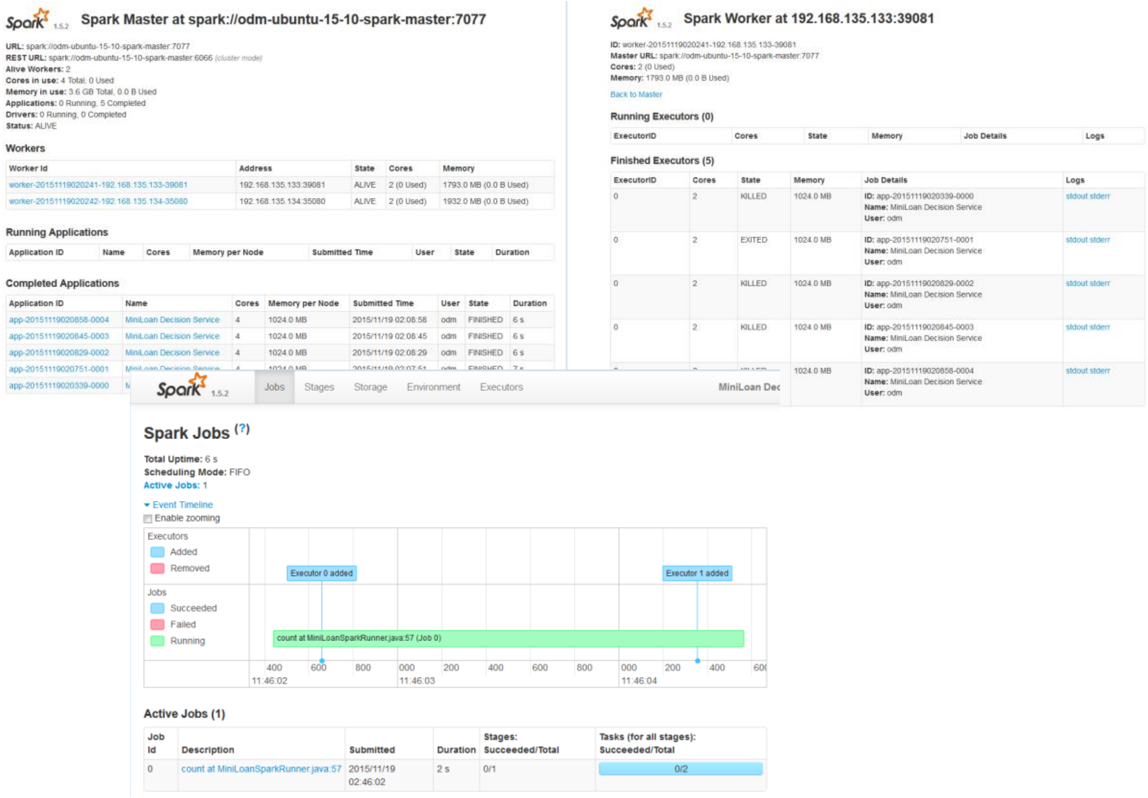
- Leverage Rule Execution Server J2SE eXecution Unit
- Pool of Decision Engines
- Multi-thread local rule execution
- TCP-IP notification when a new ruleset is deployed to refresh rules in executors

- Deployment of versioned decision services
- Execution statistics

Apply on distributed and z/OS

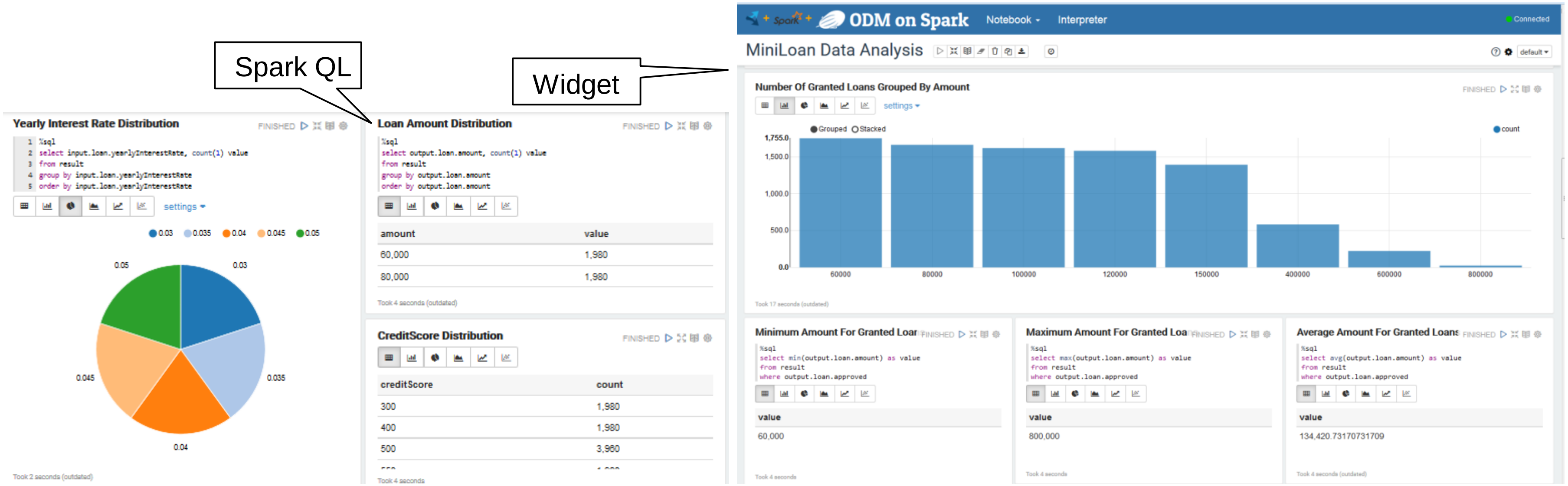
Apache Spark Monitoring

- A Spark job is deployed on the master.
- Each cluster member starts an executor (JVM)
- Each JVM runs in multiple threads a rule engine of different chunks of the decision request dataset.



Interactive Analytic Notebooks

- Interactive development with Zeppelin or Jupiter
- Manipulation of large dataset typically decisions
- Decision Service invocation possible through a Scala application
- Exploit Data results. (Metric & KPI computation & representation with Spark SQL)



Business Rules and Big Data

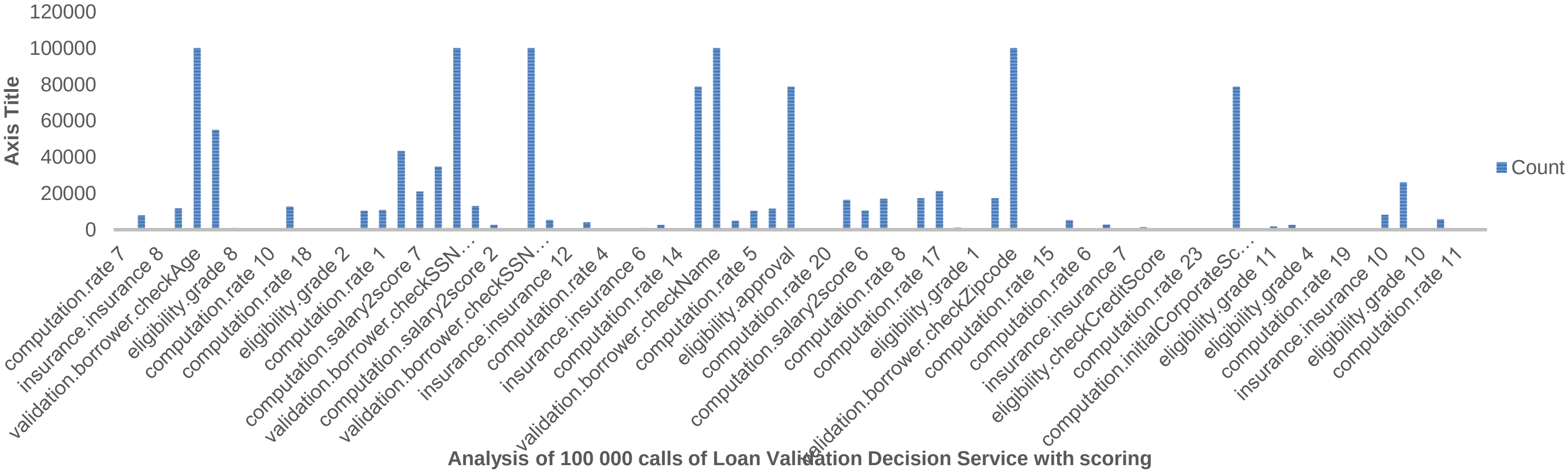
- Where Business Rules fit in the World of Big Data
- Think Big Use Case - Border Control
- Business Rules Blueprints
  - Landscape
  - in Hadoop MapReduce
  - in Apache Spark
- Rule coverage, Analytics and ML

Rule coverage for a Decision set

- What is the rule distribution when running my test campaign? In my production?
- What is my data distribution for corresponding requests & answers?
- Are there rules never fired? Always fired?
- What are all the decisions that fired a particular rule?
- Can I check that an exception handling rule is only called with the expected data?
- How does data fit with my rules to achieve my business goals?

Business Rule coverage

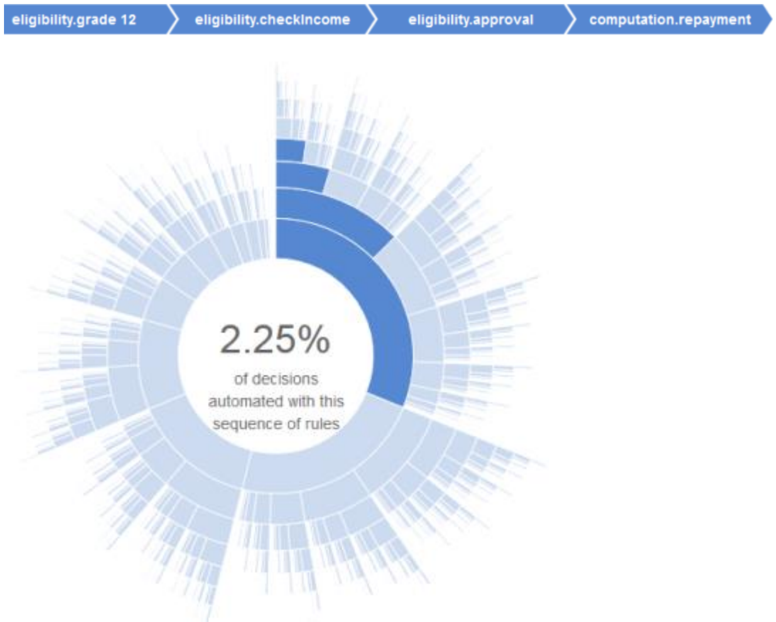
Analysis of 100 000 Loan validation decision set



26 rules & Decision Table lines on a 74 total have no execution in the decision set

Analyzing your automated decisions

- Frequency analysis
  - View in large your sequences of executed rules in your decision set



- Conditional Rule appearance
  - Likelyhood that a rule comes executed after a specific sequence
  - Indicate that some rules always go together and that their logic may be merged

Executed Rule sequence	=>		%
validation. borrower. checkZipcode		validation.borrower.checkAge	1.0
		eligibility.approval	1.0
		eligibility.checkIncome	0.54
		approved=false	0.76

Loan validation associated rules generated on the 100 000 execution traces – Processed in Spark ML

Serialize your rule based Decisions for Audit and Analytics

- Decision model
  - Decision = Request + Trace + response
  - Serializable in JSON, CSV or XML
  - Serializable into a numeric Vector


```
{"decision":{
  "id":"231422841",
    "request":{
      "borrower":{
        "name":"John Doe",
        "creditScore":810,
        "yearlyIncome":12000},
    "loan":{
      "amount":510000,
      "duration":60,
      "yearlyInterestRate":0.03,
      "yearlyRepayment":10781"}},
  "response":{
    "loan":{"amount":50000,
      "duration":60,
      "yearlyInterestRate":0.0255,
      "yearlyRepayment":10781,
      "approved":false}}
  }
}
```


- From an object model to numeric vectors
- A past decision is a vector of features
- A feature has a numeric value
- For each feature vector you associate a label that is the expected variable value
- Value scaling to help numeric computation

eligibility is the boolean label

`1 [1:810 2:120 3:510 4:60 5:255]`

CreditScore

Yearly Income scaled in K

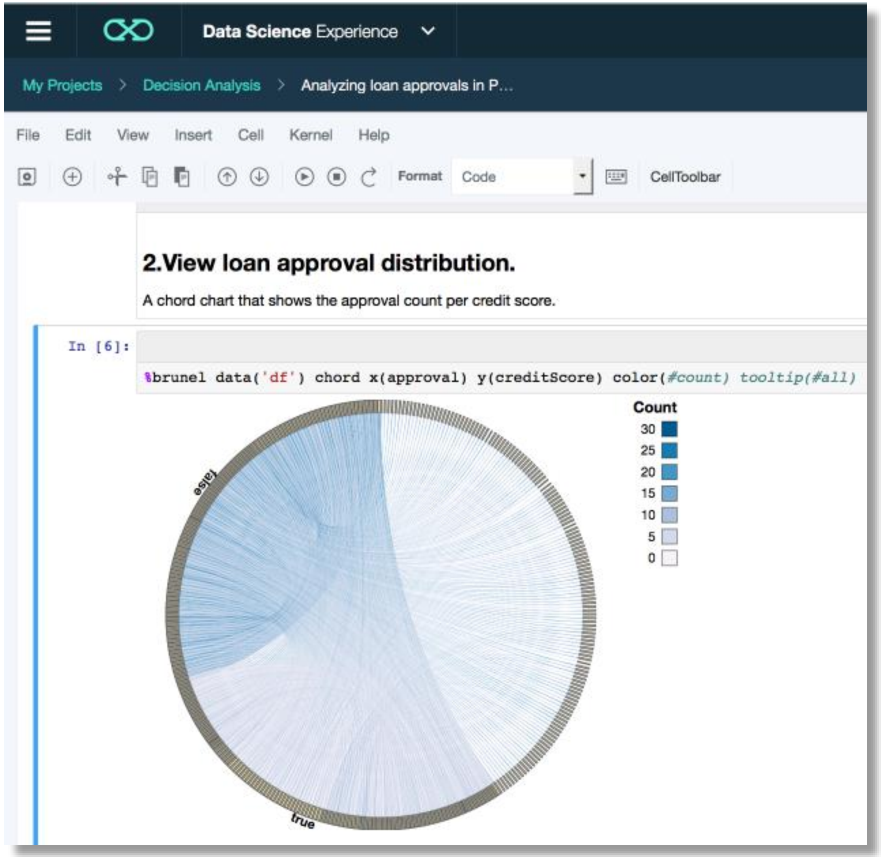
Loan amount scaled in K

Duration in months

Interest rate scaled by 100

Business Rules in Data Science Experience

- Execute business rules in a Scala notebook
- Mix business rules with predictive model invocations
- Store serialized rule based decisions (JSON, CSV, XML)
- Query and analyze in your online decision warehouse
  - In Scala with Spark QL
  - In Python Panda or SparkSession dataframes
  - In R
- Visualize your decision insights in collaborative notebooks hosted in the IBM Cloud
- Publish your notebooks on Github and contribute to a community



Rules & Machine Learning in Artificial Intelligence

Business Rules

Machine Learning

AI

- Symbolic Artificial Intelligence
- Structured data
- Formalized model with facts
- Causality
- Mainly Boolean logic
- Reasoning

- Non symbolic AI
- Unstructured data
- Signal processing
- Correlation
- Dealing with uncertainty
- Perception, Classification, Regression

Wrap up

- Combine Today business rules and Big Data into Big Decision in Hadoop and Apache Spark
- Bridge Predictive and Operation decision management technologies & teams together
- Automate massive decision making in standard OSS compute grids
- Running business policies simulations on large historical dataset with parallel metric and KPI computation
- Store your automated decisions and Analyze their efficiency against your business objectives
- Optimize the fitness between your decision logic and dataset
- Detect situations on data lakes
- Invent new algorithm combinations to solve new classes of enterprise AI at scale

Wrap up

- 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
  - Decision Composer

- Data Science Experience
  - http://datascience.ibm.com/


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