

# **SPARKLING WATER 2.0 PREVIEW**

Tom Kraljevic

tomk@h2o.ai

H2O Open Tour 2016, New York City

July 19, 2016

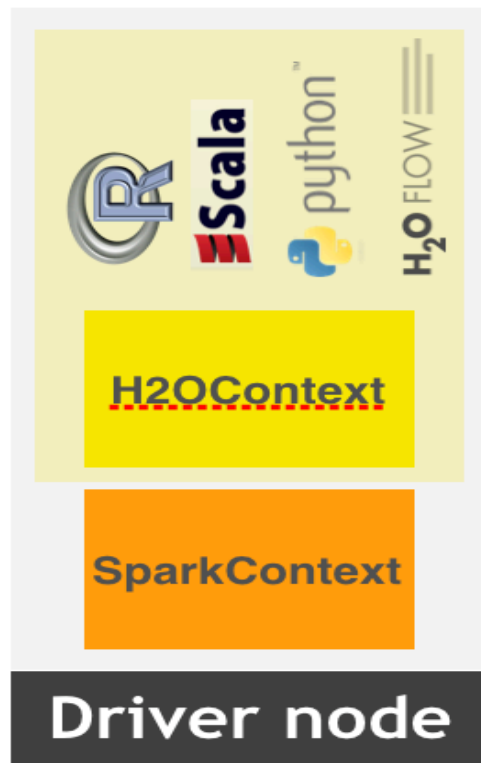
# **DISTRIBUTED SPARKLING TEAM**

- Michal - Mt. View, CA
- Kuba - Prague, CZ
- Mateusz – Tokyo, JP

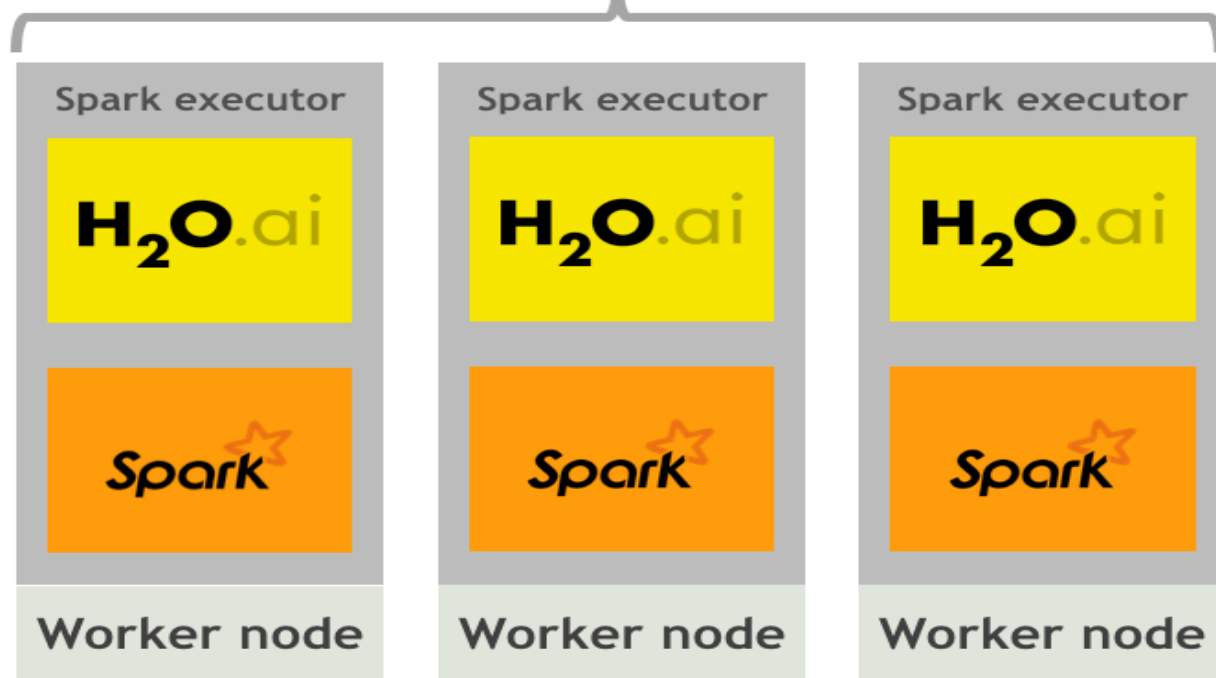
# INTRO TO SPARKLING WATER

- Sparkling water brings H2O's world class modeling algorithms seamlessly to Spark users
  - Transparent availability in Spark workflows
- Access MLlib and H2O algorithms side-by-side from a single Scala or Python driver program

Scala/Py main  
program



Spark + H<sub>2</sub>O  
SPARKLING  
WATER



Cluster manager

# SCALA CODE EXAMPLE

```
// Perform Spark SQL Join on both tables
```

```
val bigTable = sqlContext.sql(
  """SELECT
    |f.Year,f.Month,f.DayofMonth,
    |f.CRSDepTime,f.CRSArrTime,f.CRSElapsedTime,
    |f.UniqueCarrier,f.FlightNum,f.TailNum,
    |f.Origin,f.Distance,
    |w.TmaxF,w.TminF,w.TmeanF,w.PrcpIn,w.SnowIn,w.CDD,w.HDD,w.GDD,
    |f.IsDepDelayed
  |FROM FlightsToORD f
  |JOIN WeatherORD w
  |ON f.Year=w.Year AND f.Month=w.Month AND f.DayofMonth=w.Day""".stripMargin)
```

```
// Train H2O Deep Learning model
```

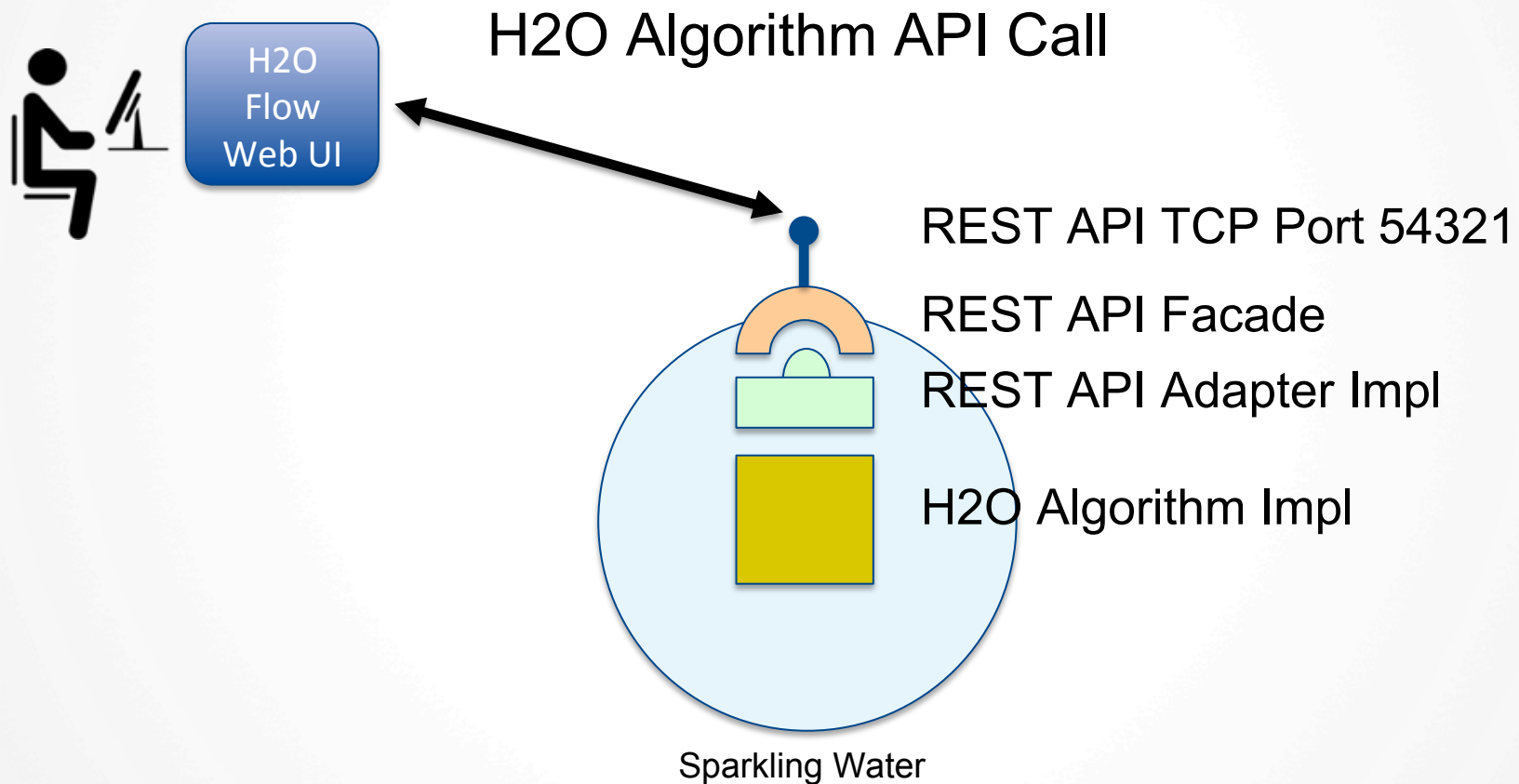
```
val trainFrame:H2OFrame = bigTable
trainFrame.replace(19, trainFrame.vec("IsDepDelayed").toCategoricalVec)
trainFrame.update()
```

```
val dlParams = new DeepLearningParameters()
dlParams._epochs = 100
dlParams._train = trainFrame
dlParams._response_column = 'IsDepDelayed
dlParams._variable_importances = true
val dl = new DeepLearning(dlParams, Key.make("dlModel.hex"))
val dlModel = dl.trainModel.get
```

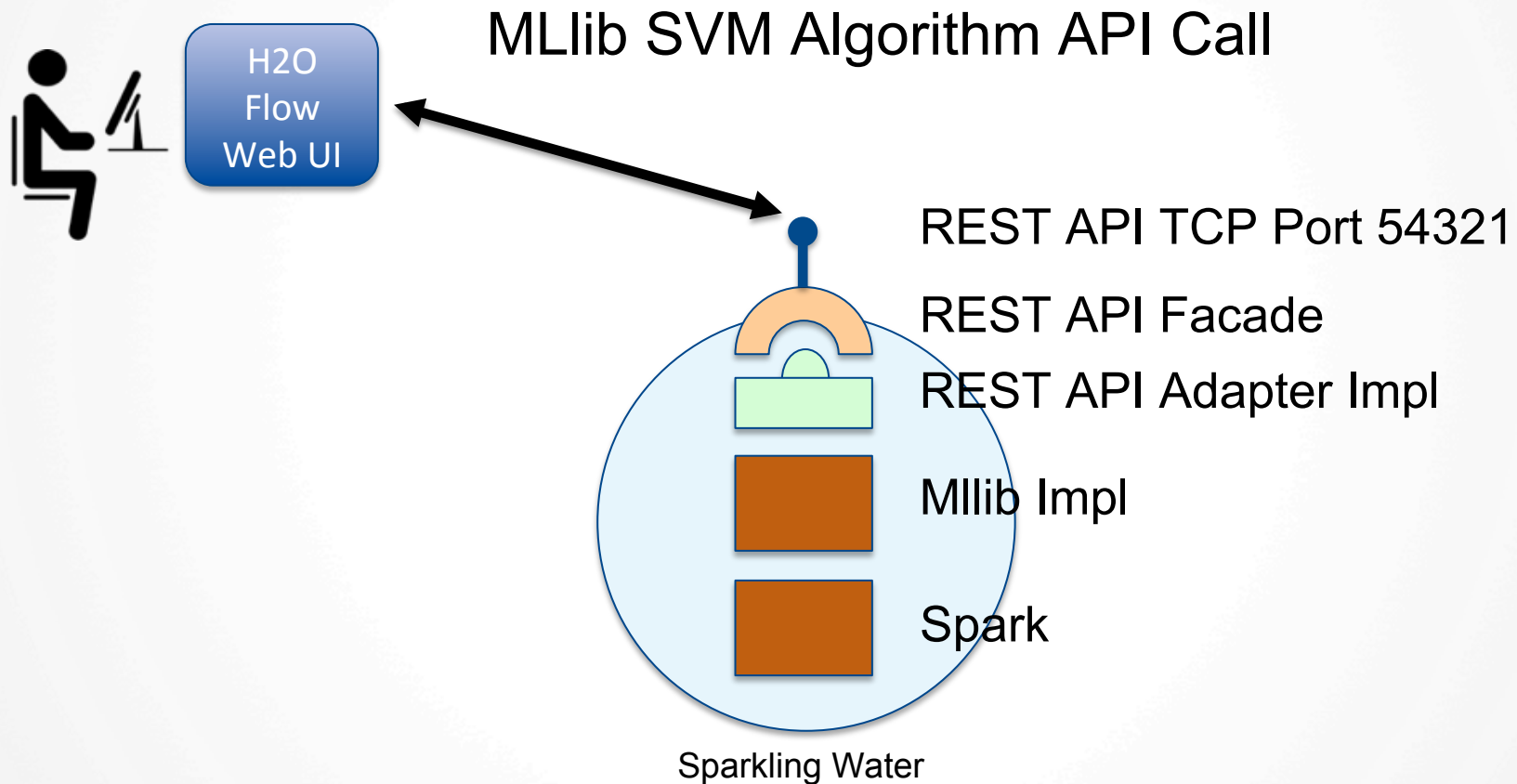
## NEW FOR SPARKLING WATER 2.0

- Enabling first pure MLlib algorithm from the H2O Flow Web UI
  - SVM
- Providing cluster stability during Spark HA events
  - (Optionally) separating the Spark and H2O clusters as two separate jobs

# H2O ALGORITHM FROM FLOW

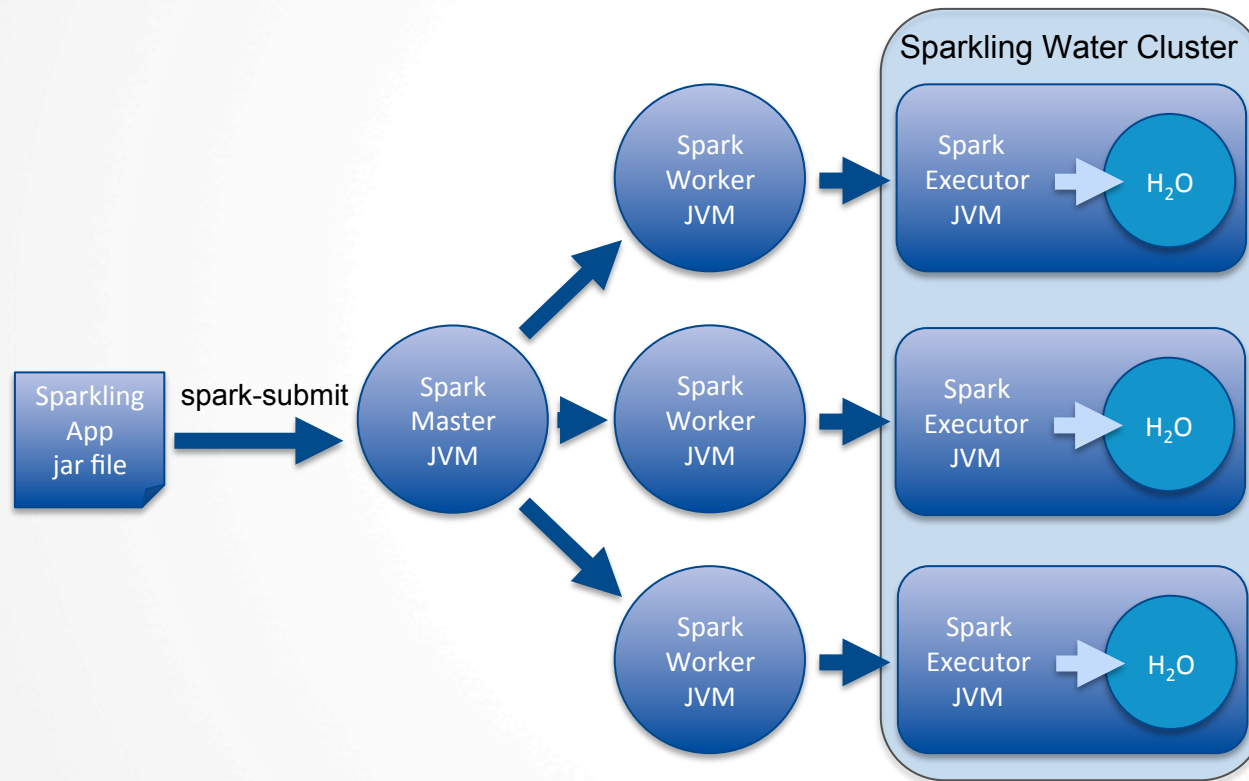


# PURE MLLIB ALGORITHM FROM FLOW

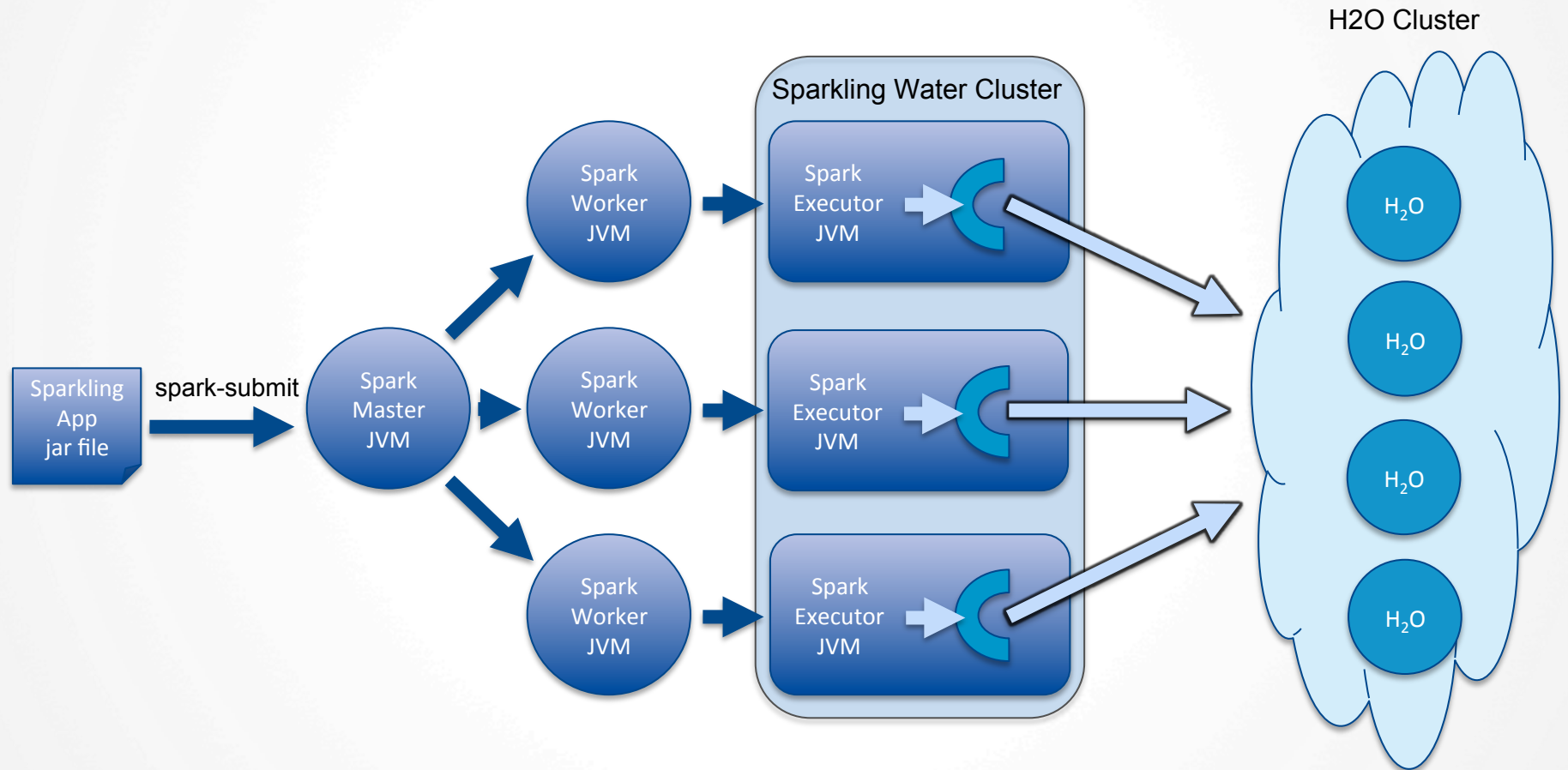




# SPARKLING WATER PRE-2.0



# SPARKLING WATER 2.0 (OPTIONAL)



# WHEN?

- Dependent on the release of Spark 2.0
  - So... Not a concrete answer...
  - BUT... First hand-held preview trial will start Aug. 1
- Also: keep aware of the numbering scheme
  - The things I just talked about will likely also be available in (for example) Sparkling Water 1.6.x

## Q & A

- Thanks for attending!
- Send follow up questions to:

h2ostream

(H2O Open Source Community Forum)