Getting Ready to Use Redis with Apache Spark

Tague Griffith – Head of Developer Advocacy
Redis Labs

Agenda

- Introductions
- Why Redis and Spark
- What is Redis
- Redis-ML module

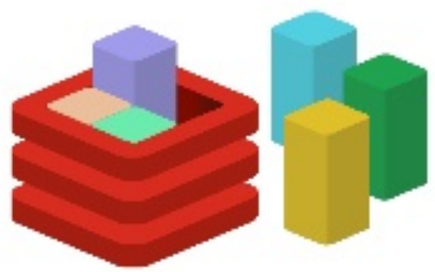
Redis Labs – Home of Redis

The commercial company behind Open Source Redis

Provider of the **Redis Enterprise (Redis^{e})** technology, platform and products

- *Founded in 2011*
- *HQ in Mountain View CA, R&D center in Tel-Aviv IL*

Who I am



- Head of Developer Advocacy
- Developer and architect turned Evangelist
- Infrastructure and Distributed Systems
- Large Scale Redis Systems
- Former: Apple, Netscape, Yahoo/Flickr, GoPro

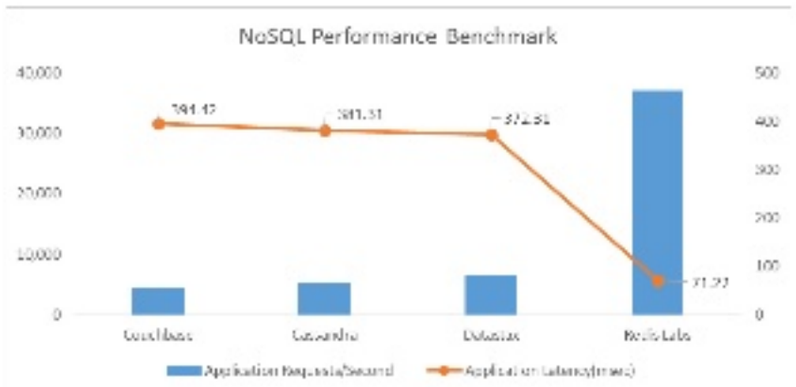
ML Models Challenges

- Models are becoming bigger and more complex
- Can be challenging to deploy
- Do not scale well, speed and size
- Reliable services are hard to do
- Can be very expensive

What is Redis

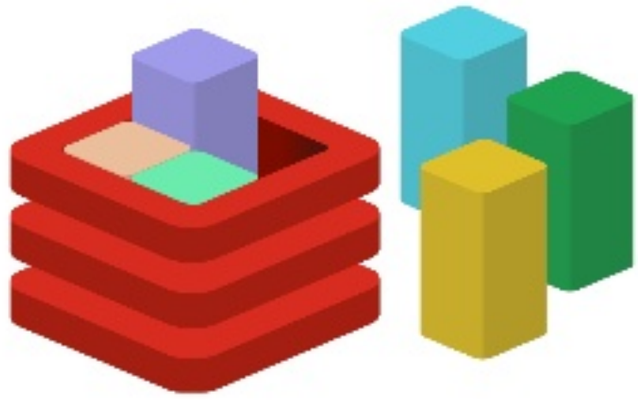
- NoSQL Database
- Open Source
- In-Memory
- Most operations performed < 1ms

Why Use Redis



NoSQL Benchmark

Redis Data Structures



Redis Modules

Performance

Simplicity

Extensibility

SPARK SUMMIT 2017

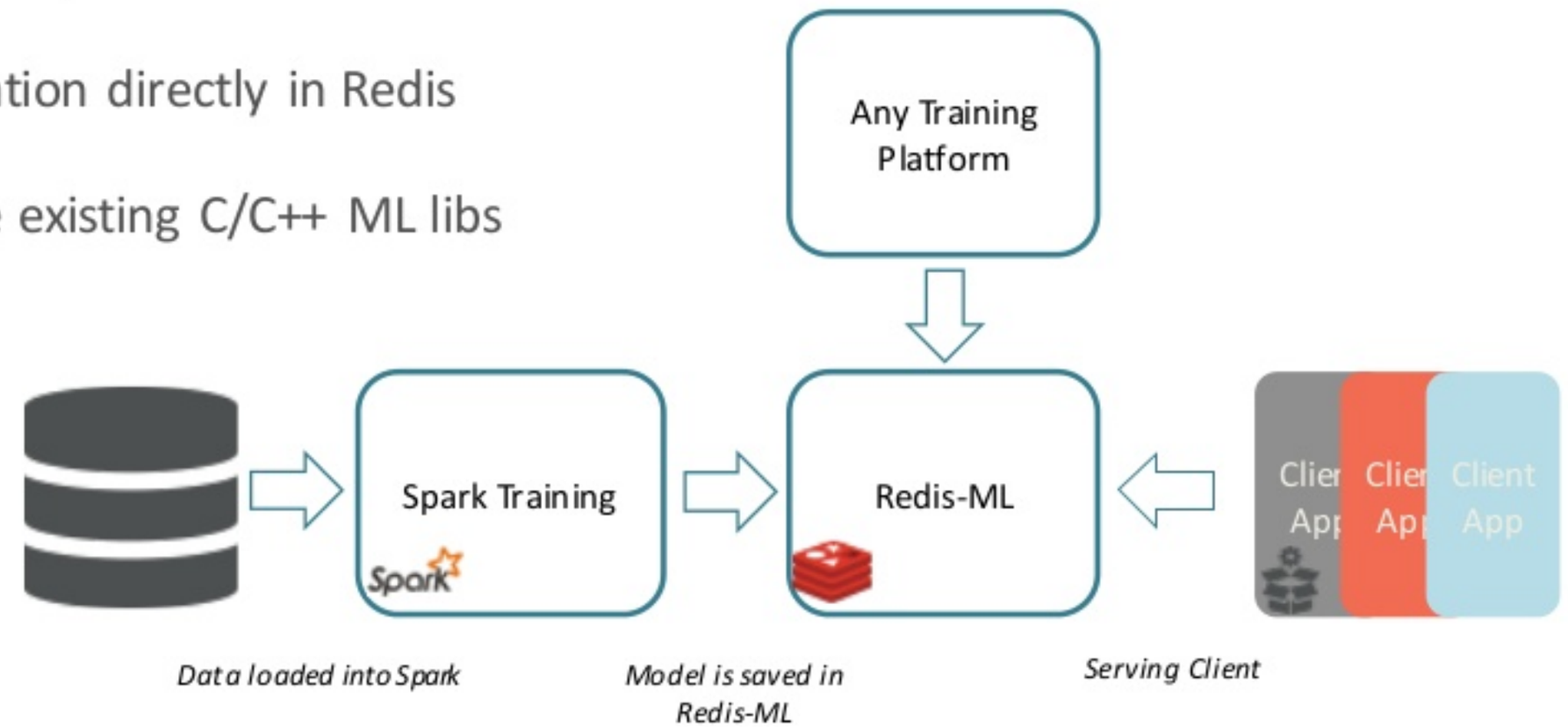
Redis Modules

- Any C/C++ program can now run on Redis
- Use existing or add new data-structures
- Enjoy simplicity, infinite scalability and high availability while keeping the native speed of Redis
- Can be created by anyone

SPARK SUMMIT 2017

Redis - ML

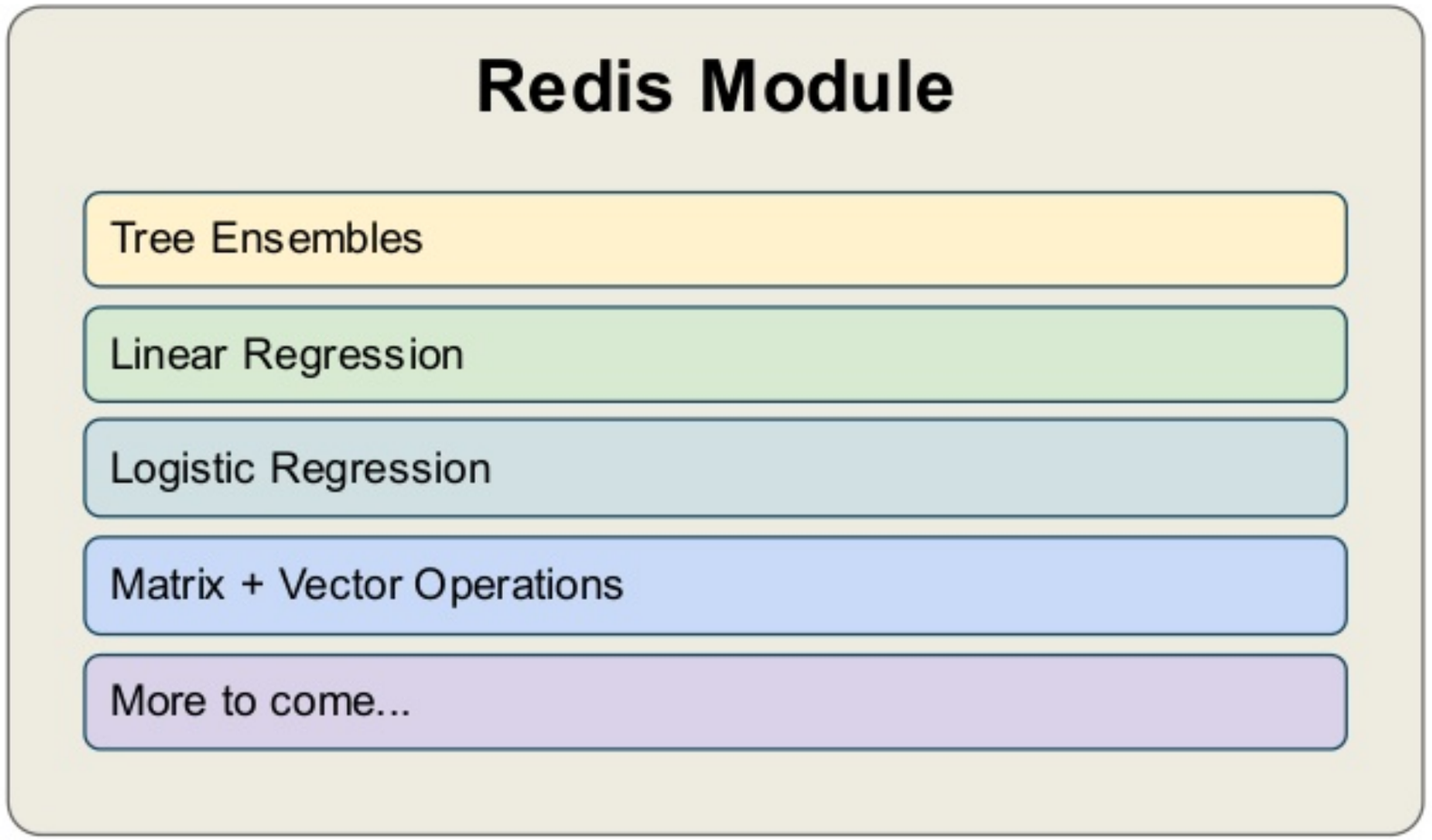
- Define data structures for models
- Store training output as “hot model”
- Perform evaluation directly in Redis
- Easily integrate existing C/C++ ML libs



Redis ML

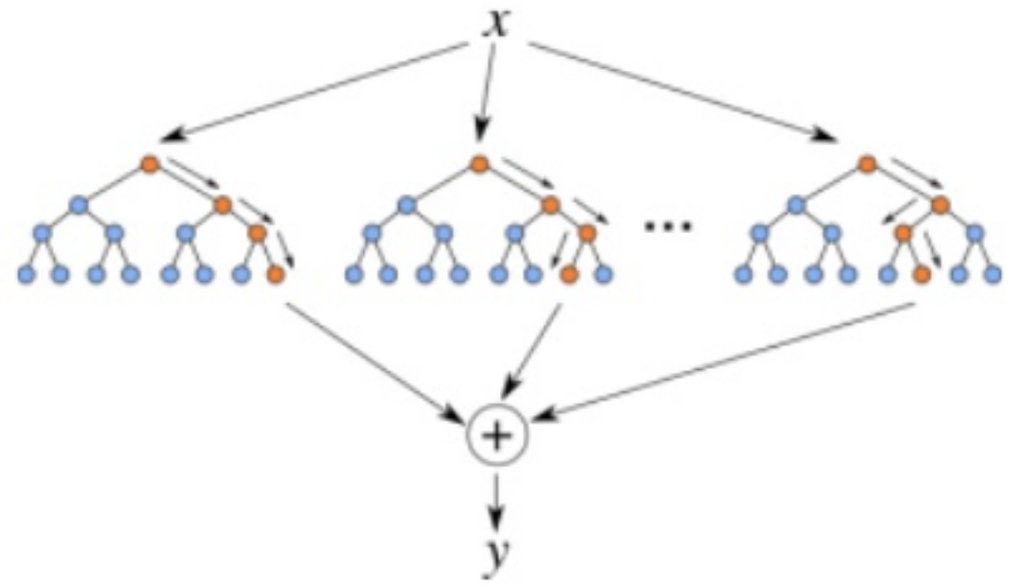
SPARK SUMMIT 2017

Redis ML Module



Random Forest Model

- A collection of decision trees
- Supports classification & regression
- Splitter Node can be:
  - Categorical (e.g. day == “Sunday”)
  - Numerical (e.g. age < 43)
- Decision is taken by the majority of decision trees



Forest Data Type API

Add nodes to a tree in a forest:

```
ML.FOREST.ADD <forestId> <treeId> <path>
    [ [NUMERIC|CATEGORIC] <splitterAttr> <splitterVal> ] |
    [LEAF] <predVal>
```

Perform classification/regression of a feature vector:

```
ML.FOREST.RUN <forestId> <features>
    [CLASSIFICATION|REGRESSION]
```

Forest Data Type Example

```
> MODULE LOAD "./redis-ml.so"
OK
```

Forest Data Type Example

```
> MODULE LOAD "./redis-ml.so"
OK
> ML.FOREST.ADD myforest 0  . CATEGORIC sex “male” .L
  LEAF 1 .R LEAF 0
OK
```

Forest Data Type Example

```
> MODULE LOAD "./redis-ml.so"
OK
> ML.FOREST.ADD myforest 0  . CATEGORIC sex “male” .L
  LEAF 1 .R LEAF 0
OK
> ML.FOREST.RUN myforest sex:male
"1"
```

Forest Data Type Example

```
> MODULE LOAD "./redis-ml.so"
OK
> ML.FOREST.ADD myforest 0  . CATEGORIC sex “male” .L
  LEAF 1 .R LEAF 0
OK
> ML.FOREST.RUN myforest sex:male
"1"
> ML.FOREST.RUN myforest sex:something_else
"0"
```

Redis Spark-ML Module

SPARK SUMMIT 2017

Resources

- Redis-ML: https://github.com/RedisLabsModules/redis-ml
- Spark-Redis-ML: https://github.com/RedisLabs/spark-redis-ml
- Online Demo: http://bit.ly/sparkredisml

Thank You.

tague@redislabs.com

@tague on Twitter