WATSON BRAZIL DEVELOPER SUMMIT 2017

BigData e Analytics

Ricardo Boaretto
rboarett@br.ibm.com

Luiz Godoy
lgodoy@br.ibm.com

In just two days

we now generate as much data as was generated in total through 2003

Over 1 billion tweets

are sent every 3 days

80% of all data

is **unstructured** and growing 15 times the rate of structured data

5 million trade events

are clocked every day

In just two days

we now generate as much data as was generated in total through 2003

Over 1 billion tweets

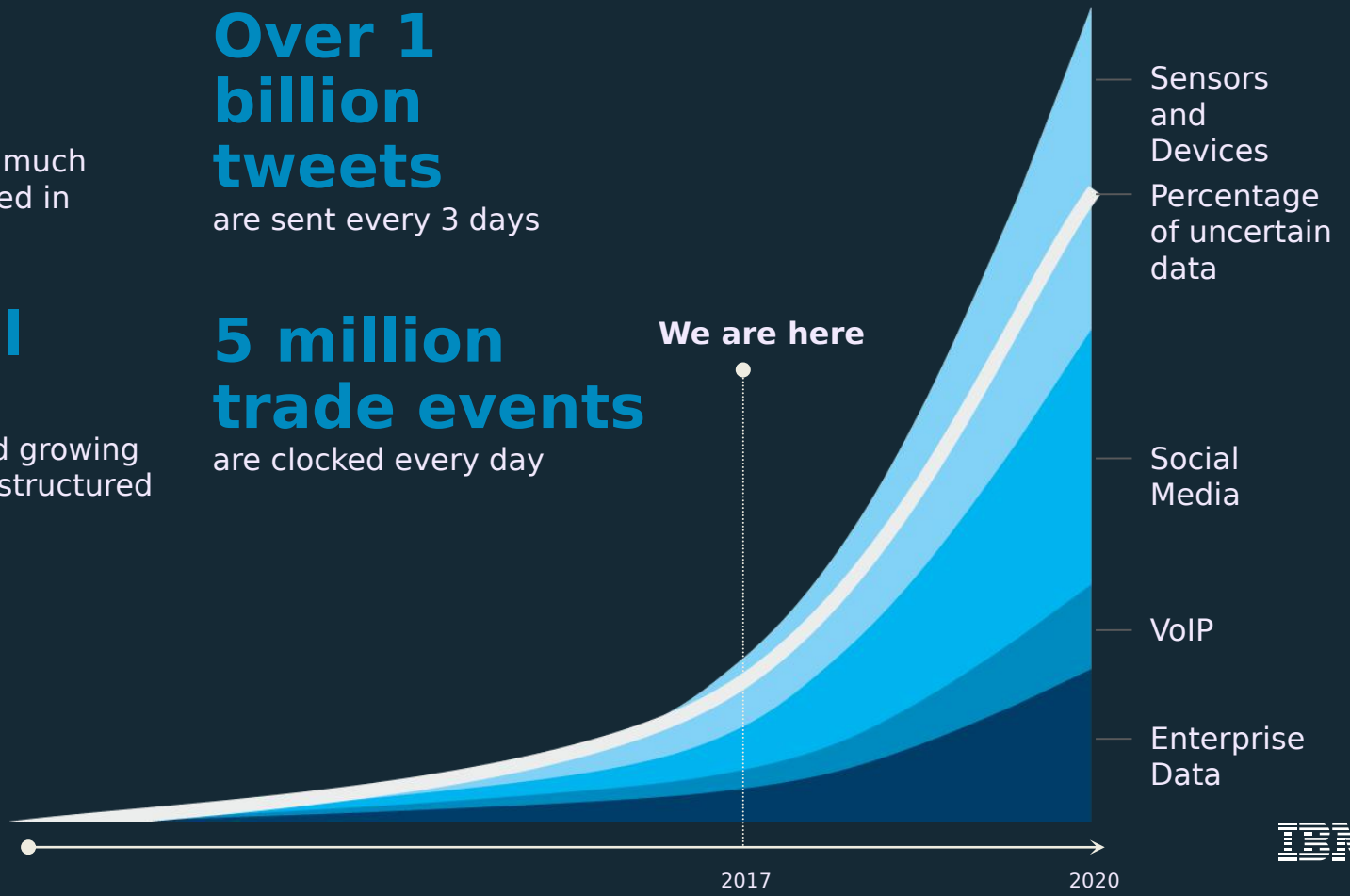
are sent every 3 days

80% of all data

is **unstructured** and growing 15 times the rate of structured data

5 million trade events

are clocked every day



Data is emerging as the world's newest resource for competitive advantage

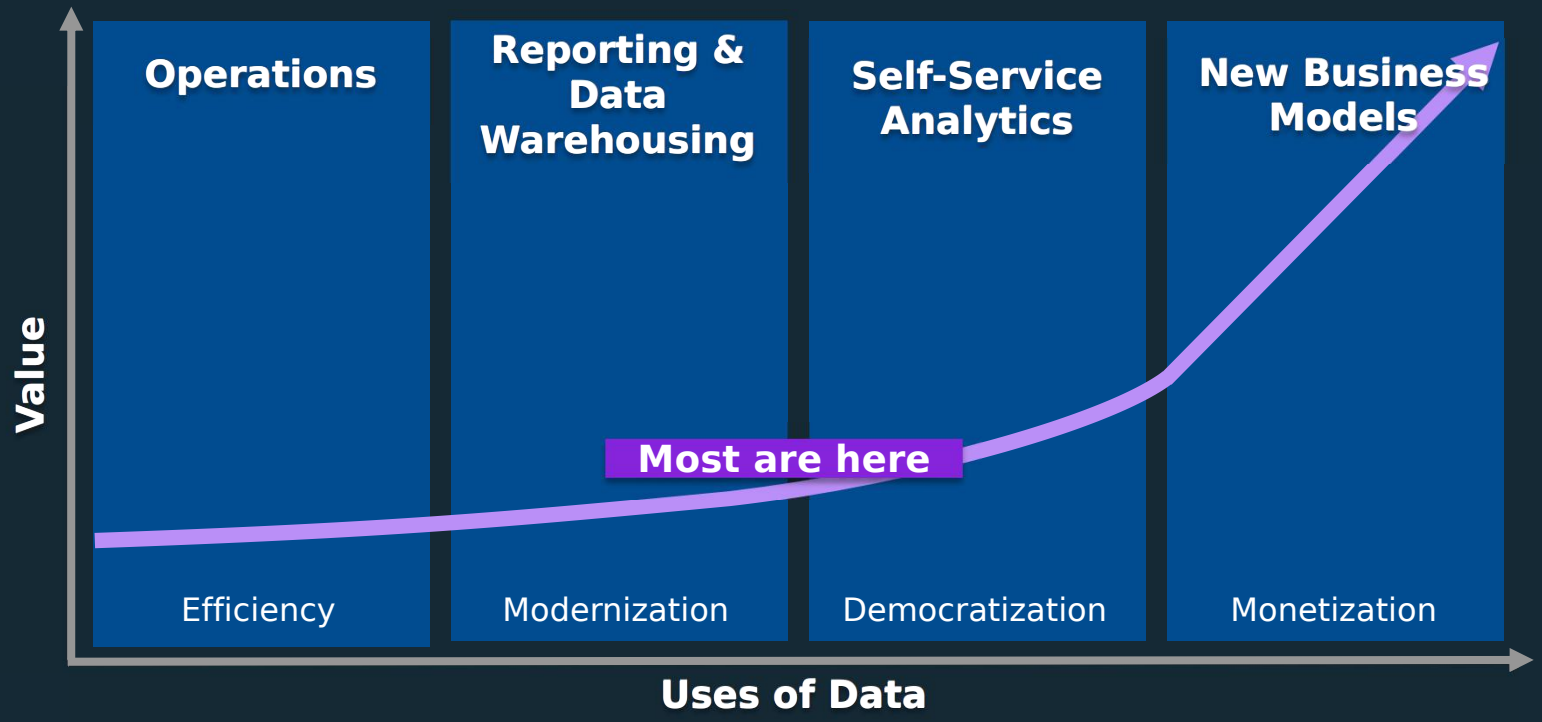
Data is emerging as the world’s newest resource for competitive advantage

Decision-making is moving from the elite few to the empowered many

Data is emerging as the world’s newest resource for competitive advantage

Decision-making is moving from the elite few to the empowered many

As the value of data continues to grow – current systems won’t keep pace


Operations
Reporting & Data Warehousing
Self-Service Analytics
New Business Models
Most are here
Value
Efficiency
Modernization
Democratization
Monetization
Uses of Data

Data

0010101101001100

Data

00101011010010110001010110
10010110001010110100101100
01010110100101100010101101
001011000101

Data

00101011010010110001010110
10010110001010110100101100
01010110100101100010101101
00101100010101101001011000
10101101001011000101011010
01011000101011010010110001
01011010010110001010110100

Data

00101011010010110001010110
10010110001010110100101100
01010110100101100010101101
00101100010101101001011000
10101101001011000101011010
01011000101011010010110001
01011010010110001010110100

Technology

Tools used to craft raw materials

Tools & Skills

Hadoop

MapReduce

```
val textFile = sc.textFile("hdfs://..
.")
val counts = textFile.flatMap(line =>
 line.split(" "))
                 .map(word => (word,
1))
                 .reduceByKey(_ + _)
counts.saveAsTextFile("hdfs://...")
```

Python

ETL

Hadoop

MapReduce

```
val textFile = sc.textFile("hdfs://..
.")
val counts = textFile.flatMap(line =>
 line.split(" "))
                 .map(word => (word,
1))
                 .reduceByKey(_ + _)
counts.saveAsTextFile("hdfs://...")
```

Python

ETL

Linear regression

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}). - y^{(i)})^2$$

Gradient Descent

Classification

Hadoop

MapReduce

```
val textFile = sc.textFile("hdfs://..
.")
val counts = textFile.flatMap(line =>
 line.split(" "))
                 .map(word => (word,
1))
                 .reduceByKey(_ + _)
counts.saveAsTextFile("hdfs://...")
```

Python

ETL

Linear regression

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}). - y^{(i)})^2$$

Gradient Descent

Classification

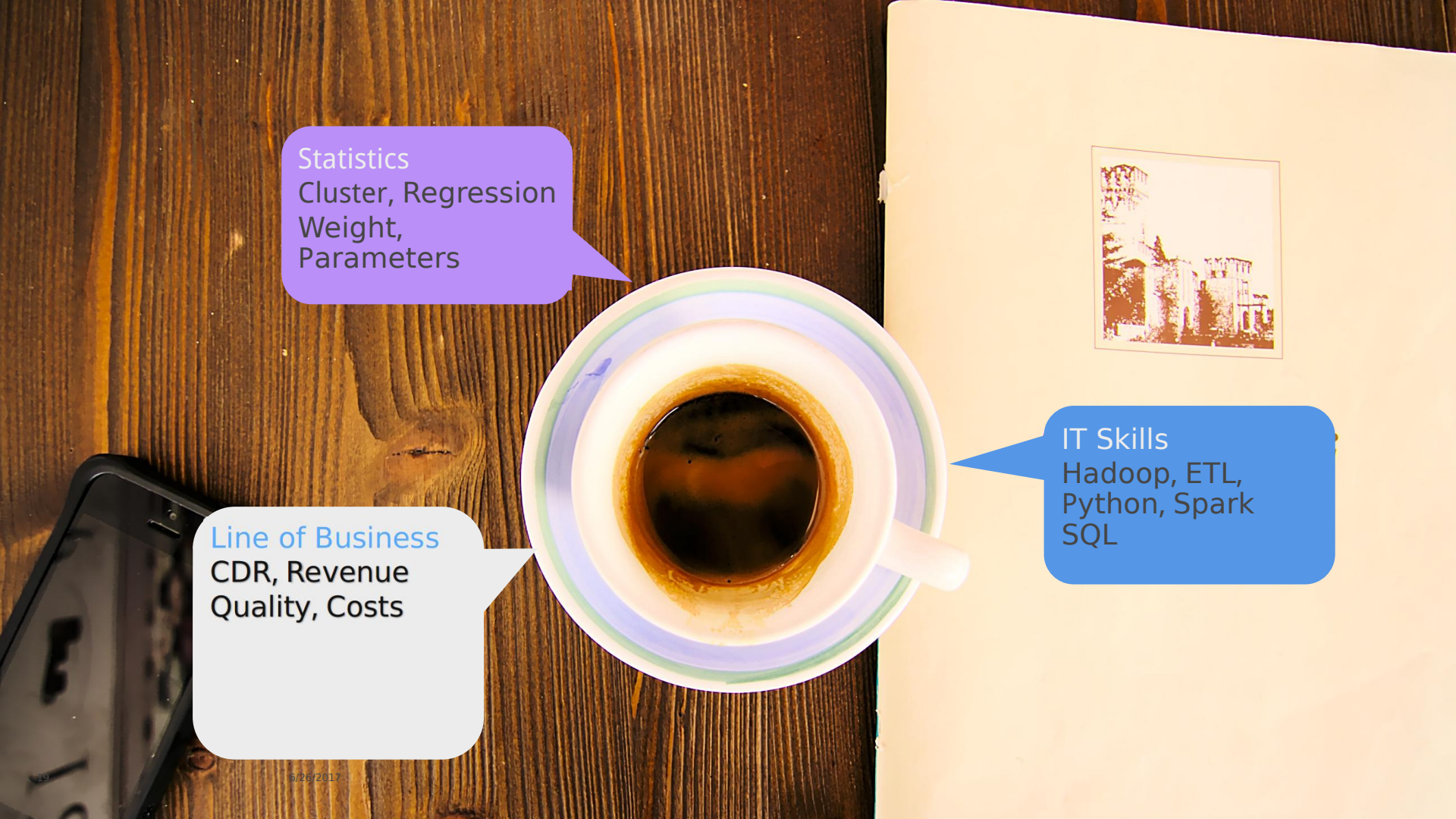
Taxes

Revenue

CDR

Line of Business
CDR, Revenue
Quality, Costs

Statistics
Cluster, Regression
Weight,
Parameters
Line of Business
CDR, Revenue
Quality, Costs


Statistics
Cluster, Regression
Weight,
Parameters
IT Skills
Hadoop, ETL,
Python, Spark
SQL
Line of Business
CDR, Revenue
Quality, Costs

Vamos criar aplicações melhores, mais adaptáveis, mais inteligentes.
Para nossos clientes.
Para nós mesmos.

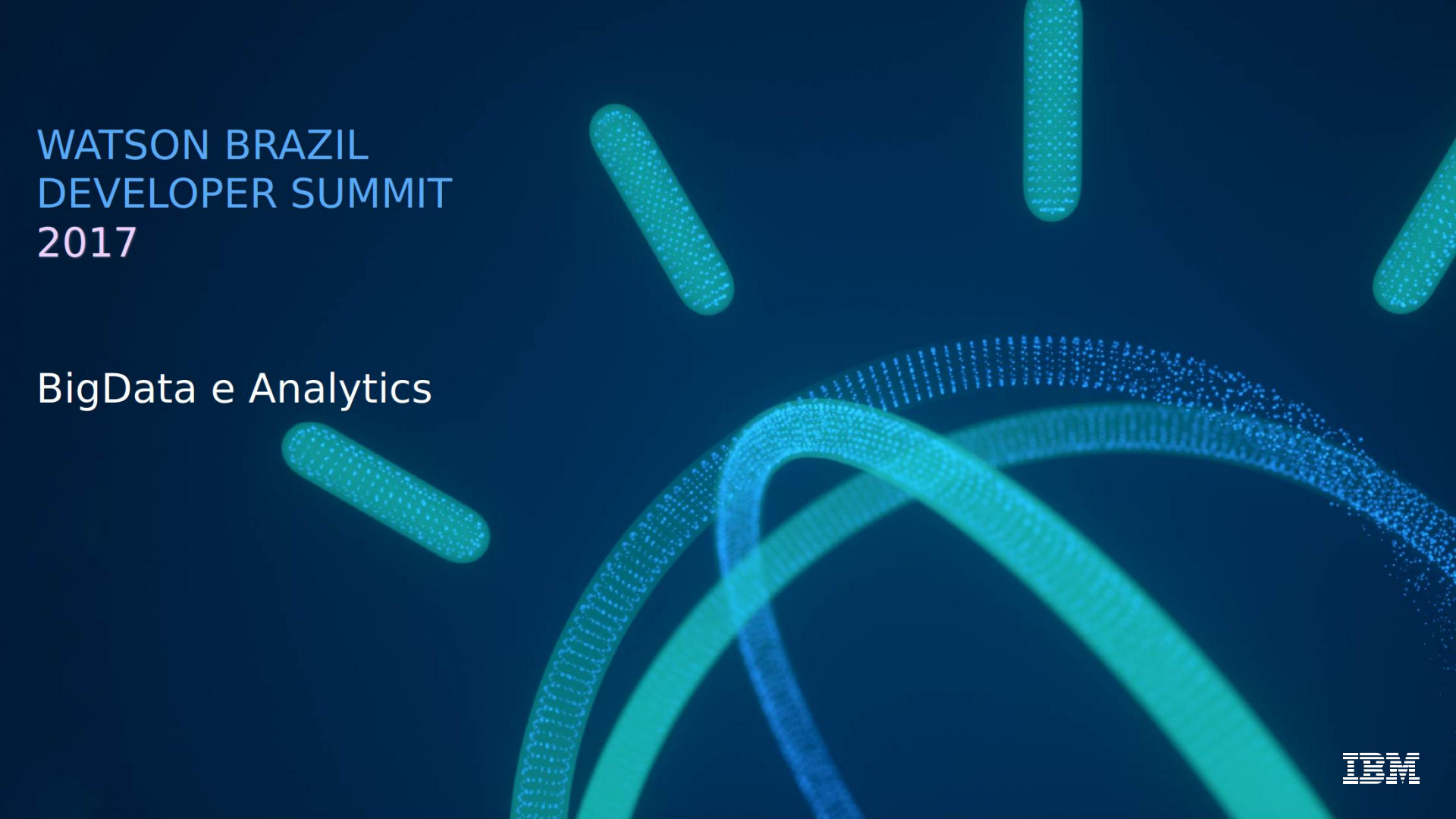
Ricado Boaretto
rboarett@br.ibm.com

Luiz Godoy
lgodoy@br.ibm.com

Laboratório
https://ibm.box.com/v/watsondevelopersummit

ricardo boaretto
rboarett@br.ibm.com

luiz godoy
lgodoy@br.ibm.com


WATSON BRAZIL
DEVELOPER SUMMIT
2017
BigData e Analytics
IBM