marionete
Turning data to value

FAST DATA ANALYTICS

Raúl Coutinho Garrido, CTO and Founder
David Navalho, Lead Big Data Architect, specialist on distributed and streaming architectures

WHAT BIG DATA IS NOT

CHALLENGE

ABOUT

Q and A

joinus@marionete.co.uk

THANK YOU!

Who We Are?

Marionete

Marionete is an IT specialised data-driven consultancy with London headquarters and operational offices in Lisbon. The only Portuguese based company, solely dedicated to Big Data Analytics.

Founded in July 2016 by Architects and Leaders in the British Financial Services with more than 6 years of experience in the Big Data Analytics domain.

marionete

Turning data to value

Technologies we work with...

Hadoop, Elastic, Cassandra, Kafka, Spark, Neo4J, MongoDB, AWS, Google, Arcadia Data, Jenkins, Kinetica, StarDog, TensorFlow, Dataiku, Python, Scala, Java, JavaScript, Linux, Kubernetes, Docker, Ansible, OpenShift, Terraform, Jupyter, IOTA, Hyperledger, EOS, Streaml.io, Blockchain, Etherium, and many other open source and data-driven technologies ...

Nope...

SQL Server, .Net, TFS, Salesforce, Oracle DB, Exadata, Teradata, Tableau, Qlik, SAP, SAS, Outsystems, UX, UI and many other proprietary software and non agile methodologies!
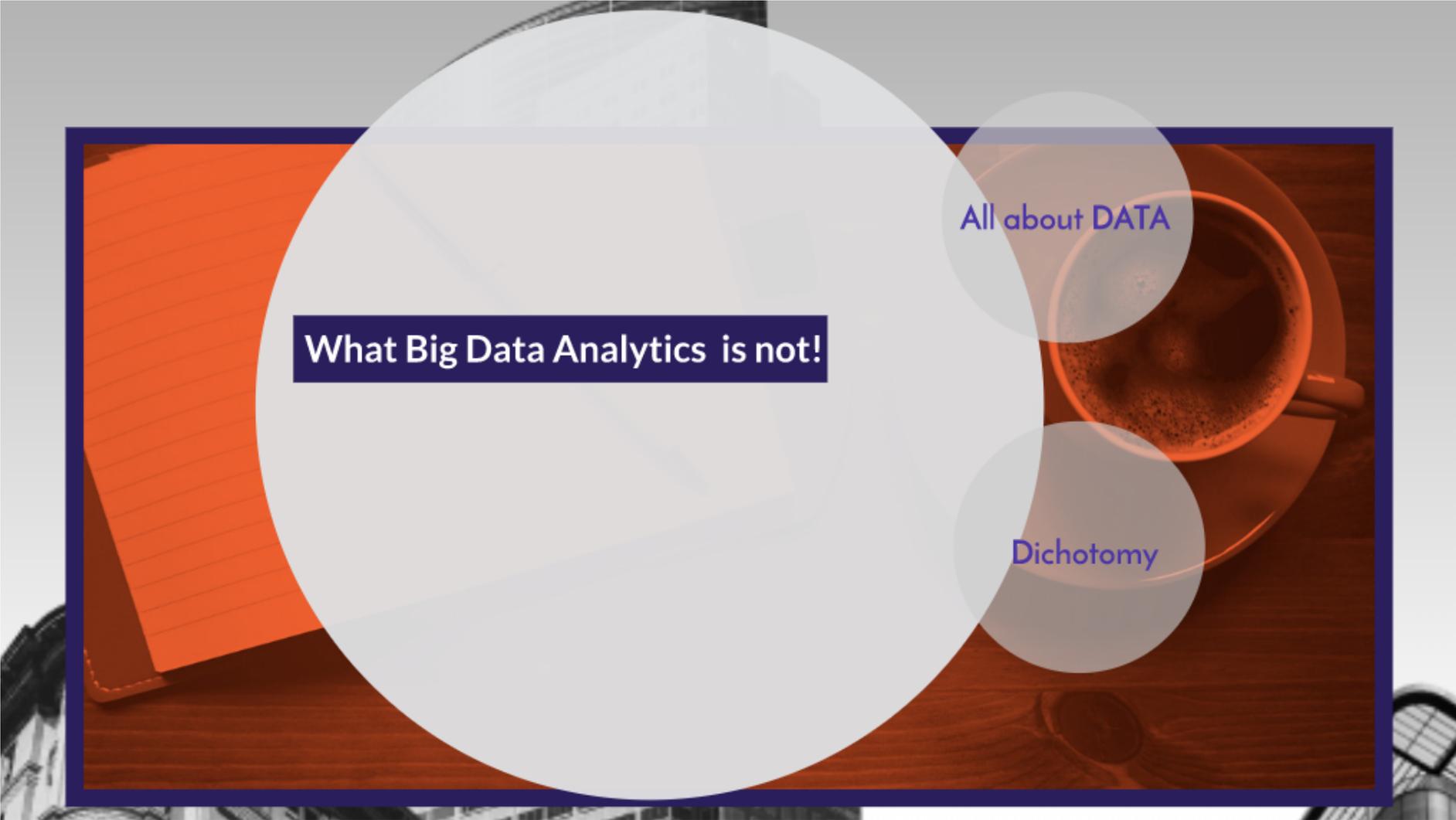
What Big Data Analytics is not!

All about DATA

Dichotomy

marionete
Turning data to value

Data

01. We are on the era of Data-driven economy where data is in perpetual movement and change

02. Whatever the industry, Data and its derivatives are the most valuable commodity, businesses can produce.

03. In the past companies were digging for commodities and prospecting for oil, today they should focus on digital extract, insights and monetisation.

The cycle of "Wata"

Transforming
Cleaning
Enhancing
Merging
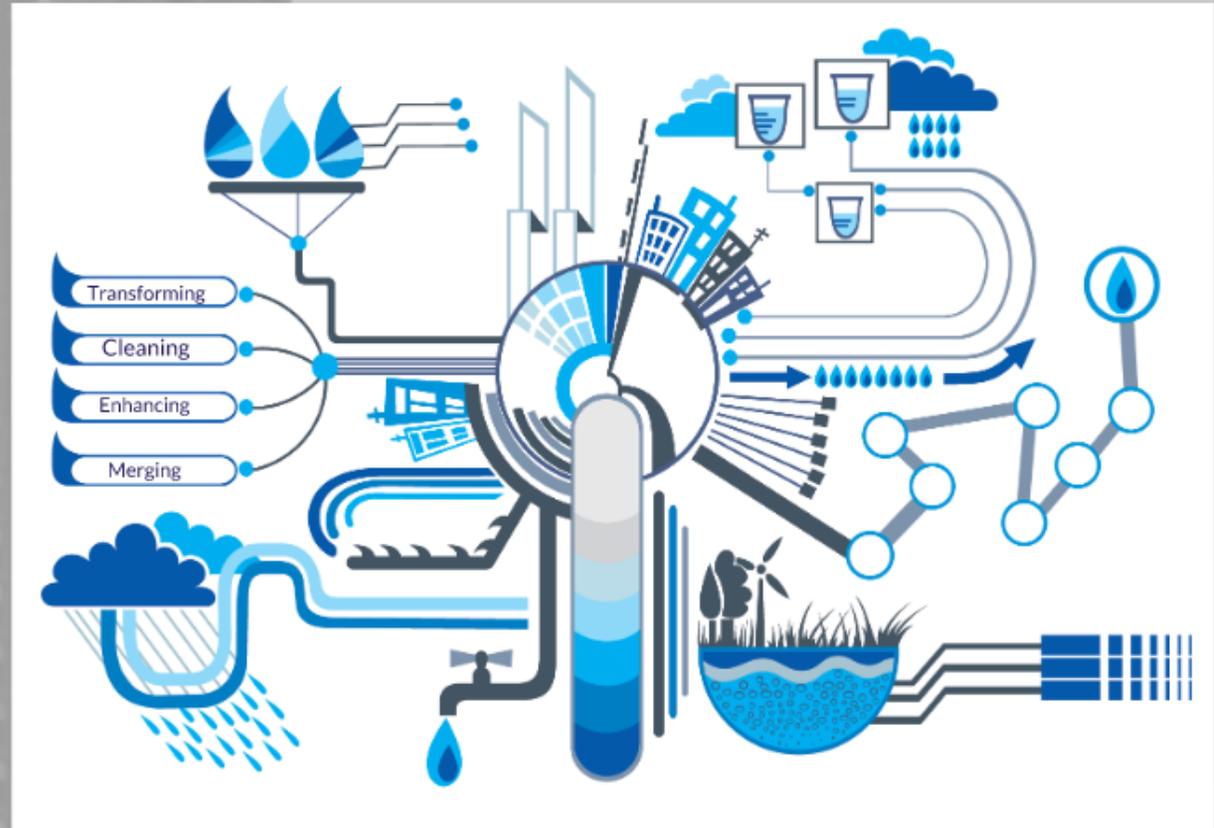
Dichotomy

- Data Lake
- Interface based Data Model, Materialised Views, Data Virtualisation
- Functional Reactive Programming, Events

- Lean, Agile, Automated testing
- DevOps, Continuous Integration, Continuous Delivery, Continuous Deployment, Automation
- ETL is dead, long live streams, Real-time data, Streaming
- Unstructured and Structured
- Open Source
- Shared resources, independent development units
- Microservices, APIs
- Decentralised, Distributed
- In-memory processing, persistence and transit
- Variety, Velocity, Volume, Data Locality, Processes on Data
- Cloudera, HortonWorks, Databricks, DataStax, Confluent

VS

- Data Warehouse, ROLAP, OLTP
- Canonical Data Model

- OOP, Procedural, DAG (Directed Acylic Graph)
- Waterfall, functional testing

- ETL
- RDBMS, MPP
- Proprietary Software
- Product based Infrastructure
- SOA and multi-tier architectures
- Centralised DBs and App Servers
- Process on middleware
- On-Disc
- SAP, SAS, Oracle, Microsoft, IBM, Teradata, Tableau, Qlik

Use Cases

01. ODB car data for driver behavioural, road surface, weather, air quality, smart city, fleet management, road traffic analitycs

02. Market and Price data for P&L Attribution and Risk Analytics, regulatory banking reporting

03. Cyber-Security, Forensic Analytics for anomaly and novelty detection to predict machinery breakdown and equipment failure, fatigue and wear

04. Streaming sports with real-time advert royalties analytics

marionete

Turning data to value

Challenges

01. The fusion of: IoT with Descentralised (Blockchain) and Distributed architectures

02. Crise Managment real-time Analytics for the detection of fake news by clustering sources, ranking and generate summary recomendations.

03. Determine teacher and student performance based on Machine Learning with Graph Topology.

04. State-based streaming computations: very low latency computations, requiring previous state information, with incremental modifications (avoid recomputations)

Q and A

joinus@marionete.co.uk