The BigFoot Project: Selected Research Challenges

Big Data Management Challenges and Solutions in the Context of European Projects

EDBT/ICDT 2014, Athens

Web: www.bigfootproject.eu
GitHub: github.com/bigfootproject
BitBucket: bitbucket.org/bigfootproject

What is BigFoot?

A software stack:

- Data-Intensive Scalable Computing
- *In-situ* Interactive Data Analytics

… as a service

- Built for Private and Hybrid clouds

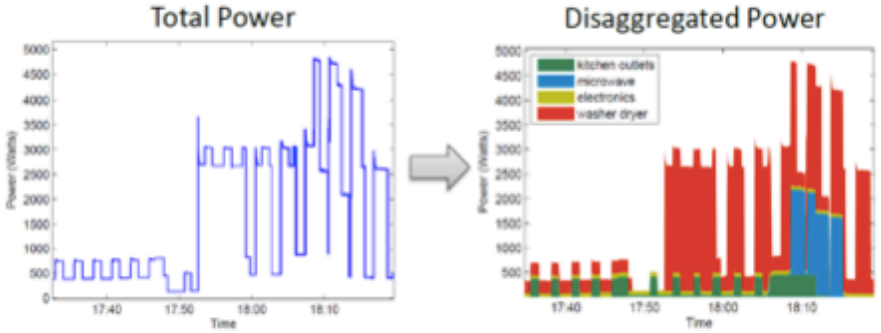
… with efficiency, failure tolerance and performance guarantees

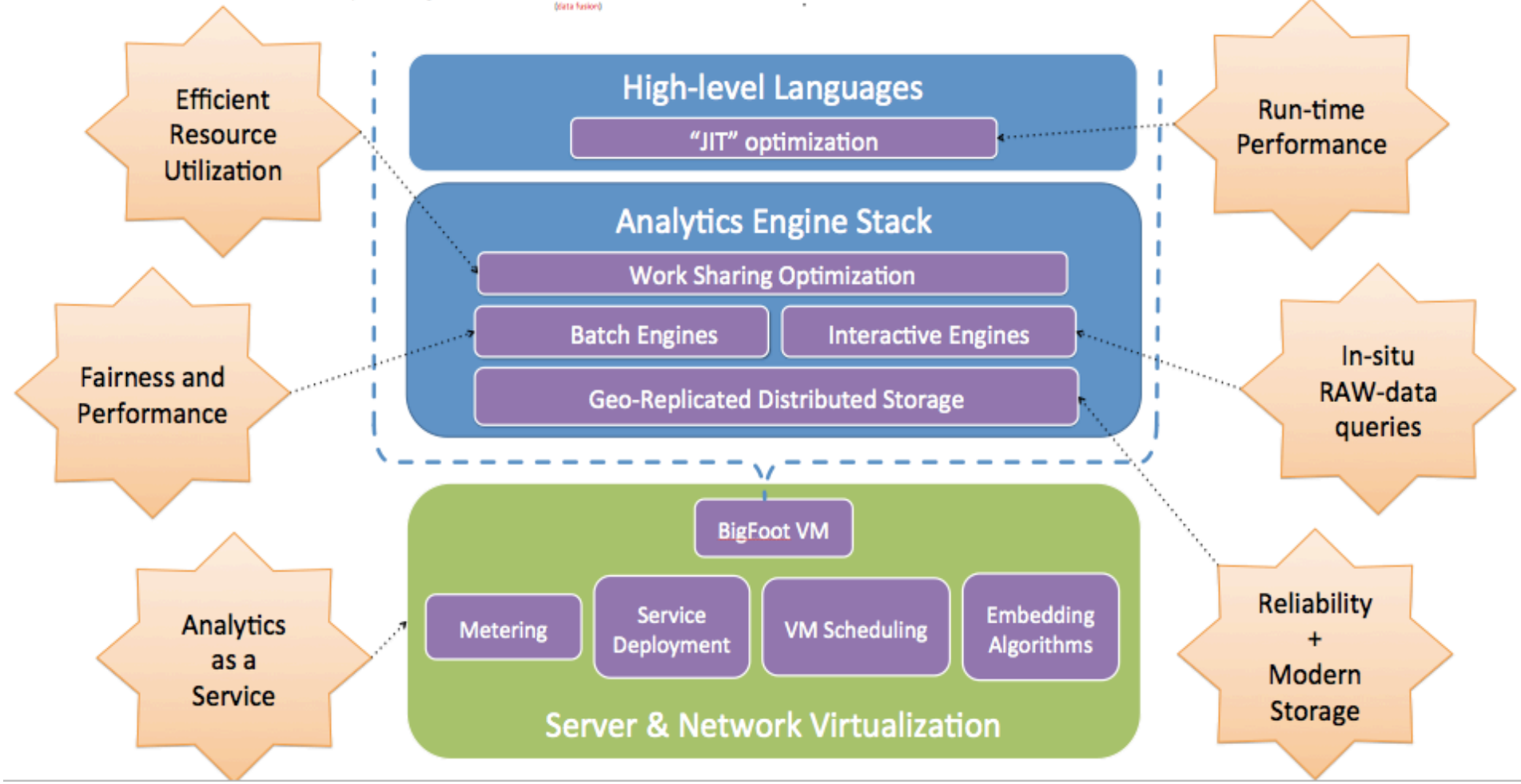
And its users: *ICT Security* and *Smart Grids*

- Scalable machine learning
- Time-series analytics

Cyber-Security **Applications**: attack classification

SmartGrid **Applications**: disaggregation





Interactive meets Batch

Why? Application use-case in BigFoot

- Machine learning algorithm tuning
- Temporary output, use it a few times, throw it away
- How do you query raw data?

State-of-the-art

- MapReduce (Hadoop / Spark) writes to HDFS
- Load output to a traditional DB, wait, then query
- Impala, Shark, BlinkDB are not suitable for this use-case

The BigFoot approach

- DiNoDB: distributed in-situ queries on raw data
- Collapses the L in ETL, generates positional maps and other indexes
- SQL queries dispatched to HDFS data nodes

Resource Allocation

Context

- Batch processing (MapReduce)
- Heterogeneous workloads: interactive + production jobs

State-of-the-art

- Performance OR fairness
- Tedious manual exercise to create pools

The BigFoot approach

- Size-based scheduling ➔ HFSP, Hadoop Fair Sojourn Protocol
- Coarse job runtime estimation
- OS-assisted preemption

Analytics as a Service

Context

- DISC services: Hadoop, Spark, …
- Ephemeral (a la AWS EMR) vs. Elastic
- Separation of storage and compute layers

The problem

- (Dynamic) Bin packing
- Objective functions hard to define

The BigFoot approach

- Distributed commit log to dispatch scheduling requests
- Measurement service provides system state
- OpenStack Sahara contributors

Thank you!!

Web: www.bigfootproject.eu
GitHub: github.com/bigfootproject
BitBucket: bitbucket.org/bigfootproject