CASE STUDY

WHICH TOOL IS RIGHT FOR YOUR BIG DATA?

EPAM HELPS TELEFÓNICA GET DATA RUNNING 10 TIMES FASTER

‹epam›

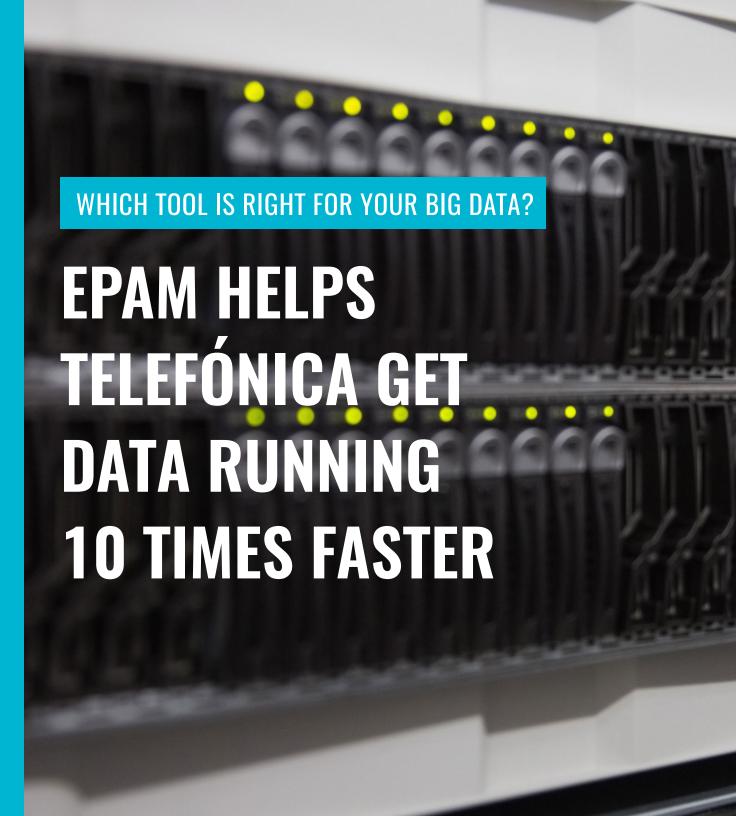## SOLUTION HIGHLIGHTS

• Implementations of the selected use cases done in three technology alternatives : Plain Java MapReduce, Apache Hive, Apache Spark

• Final implemented solution for Hive without using Oozie

• Full data access on Hadoop from Oracle

• Publishing to Oracle using the following methods:
  - Sqoop
  - Oracle Loader for Hadoop (OLH)
    • Using JDBC interface
    • Using OCI Direct Path interface

• Approaches to Cluster Management: user and resource management

TECHNOLOGIES AND TOOLS

• Hadoop 2.0.6
• Apache Hive, Apache Spark
• Oracle DB 11g, Oracle Loader for Hadoop
• Java, HiveQL, Scala
• Development environment: EPAM Cloud

CASE STUDY:

EPAM ACCELERATES TELEFÓNICA'S MOVE TO BIG DATA

BACKGROUND: BIG DATA IS GO, NOW WHAT?

Telefónica Germany is a subsidiary of Madrid-based Telefónica S.A., one of the largest telecommunications companies in the world with a significant presence in 21 countries and a customer base of more than 313 million. As part of the push to expand personalized offerings to private and business customers, as well as the need to more efficiently utilize investments into the Oracle Enterprise Data Warehouse platform, Telefónica BIC (Business Intelligence Center) decided to move forward with a new Big Data platform – one that enabled clustering with Hadoop.

The decision to push into Big Data was made, but the challenge was still ahead – namely, to find the best platform among a wide range of options, and the best path forward to enable growth, efficiency and insights for the future.

EPAM ADVANCED TECHNOLOGY MAPS OUT THE PATH

In response to the challenge posed by the client, EPAM specialists from the Advanced Technology Lab ran with three priority use cases for Big Data needs – relying heavily on the elaboration and testing of various platform and configuration options in a truly Agile BI Lab environment. The process, which was time-boxed at three months was concluded with a fully benchmarked analysis and a true platform 'bake-off'.

THE RESULT WAS A WAY FORWARD DERIVED FROM TRULY AGILE COLLABORATION

Working closely with Telefónica, the EPAM team was able to demonstrate several implemented solutions, in the end selecting the most successful one, which utilized EPAM's own Cloud infrastructure and obfuscated data. Past the initial proof of concept stage, the complete solutions were deployed and rigorously tested on Telefónica BIC's infrastructure.

CONDENSED DECISION MATRIX

FACTORS	HIVE	SPARK	PLAIN JAVA
Coding efforts	1 week/person + 1 week/person for "glue"	1.5 weeks/person	3 weeks/person
Lines of code (incl. DDL and glue)	676 + 662 = 1,338	1,600	2,500
Speed of data processing	90 min.	30 min.	70 min.
Potential performance improvement impact	Likely	Significant	Possible
"Glue" required	Yes	No	No
Scalability	No restrictions	Restricted by size of cluster memory	No restrictions
Unit testing	Difficult	Excellent (junit)	Good (mrunit)
Debugging	Logs only	Excellent	Good
Query optimizer	Available	Manual	Manual

EPAM ACCELERATES TELEFÓNICA'S MOVE TO BIG DATA

FUTURE-PROOFING BIG DATA

As the benchmarks were completed against the defined decision criteria, the developed solutions can allow for the processing of large amounts of unstructured data at speeds up to 10 times faster than previous alternatives. Beyond the benchmarks, the team's efforts enabled the creation of a BI blueprint that takes Telefónica forward towards a scalable technological base for future Business Intelligence solutions, including new Big Data tools and techniques.

QUESTIONS?
CONTACT US AT
SALES@EPAM.COM

(i) For more information,
PLEASE VISIT EPAM.COM

41 University Drive, Suite 202,
Newtown, PA 18940 USA
P: +1 267 759 9000 | F: +1 267 759 8989