



Predicting Customer Behavior Using Big Data Analytics with MATLAB in the Cloud

Rachid el Mimouni
Head of Data Science
Finance
NLE
2017-06-20

Key Takeaways

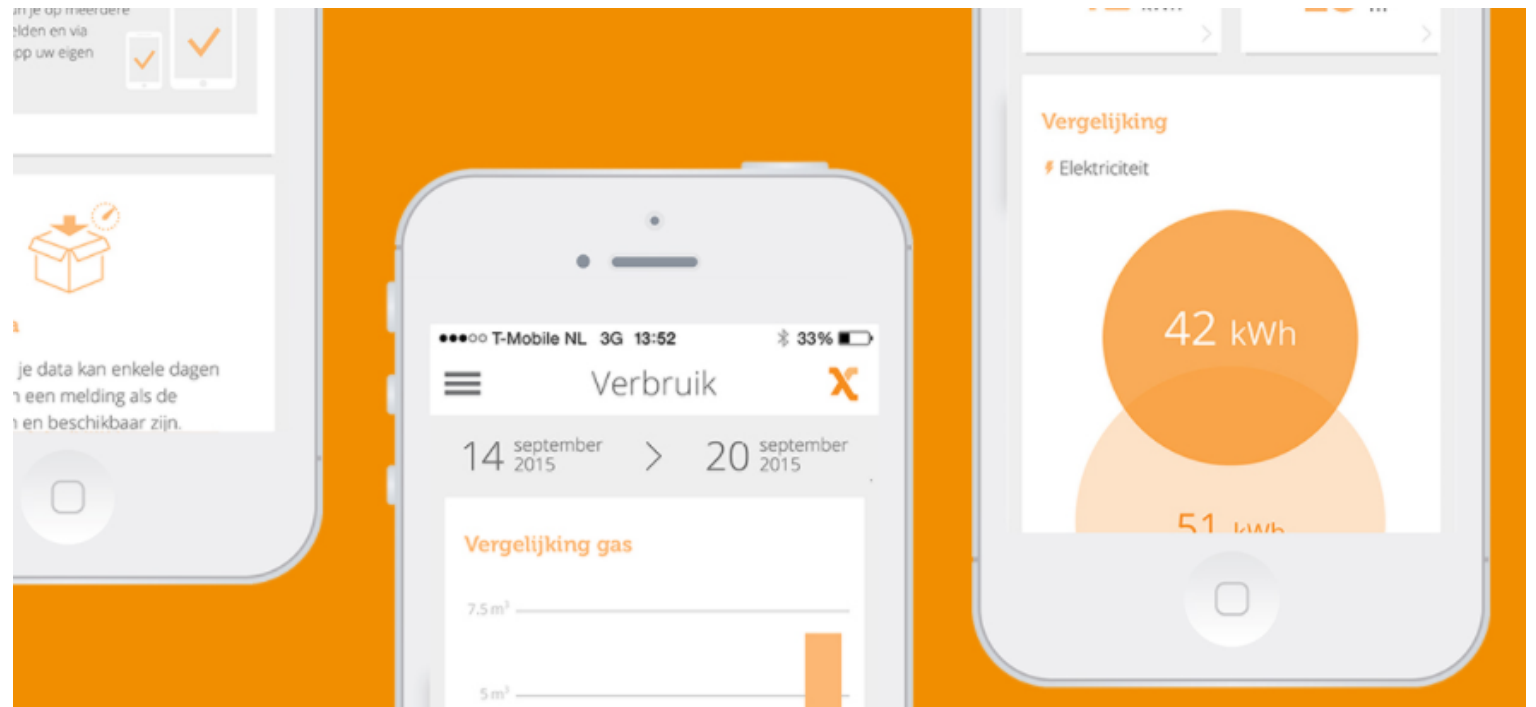
- **Fast scalability and broad community support**
- **Integration with other languages**
- **Calculations up to 20x faster**
- **Make the impossible possible**
- **Use a tool (including MATLAB) where it's good at!**

NLE (formerly known as: Nederlandse Energie Maatschappij)

- Largest independent energy and utility service provider in NL (Energy/Broadband/Boiler services, etc)
- Migrating to a multi-utility service provider

My role:

- Data Science department
- 5 people
- Small and lean organization



Innovation Challenges and Achievements

- *Goal: inference about customer behavior in response to actions by company*
 - *Predictions on customer level instead of group*
 - *Fast iterations for quick model optimization*
 - *Maximum level of granularity*
- Calculations became slow
- Could not forecast long periods due to memory limitations (on-disk caching makes things very slow)
- Achieved boundaries of local optimization (changed data-types where possible to lower precisions etc.)

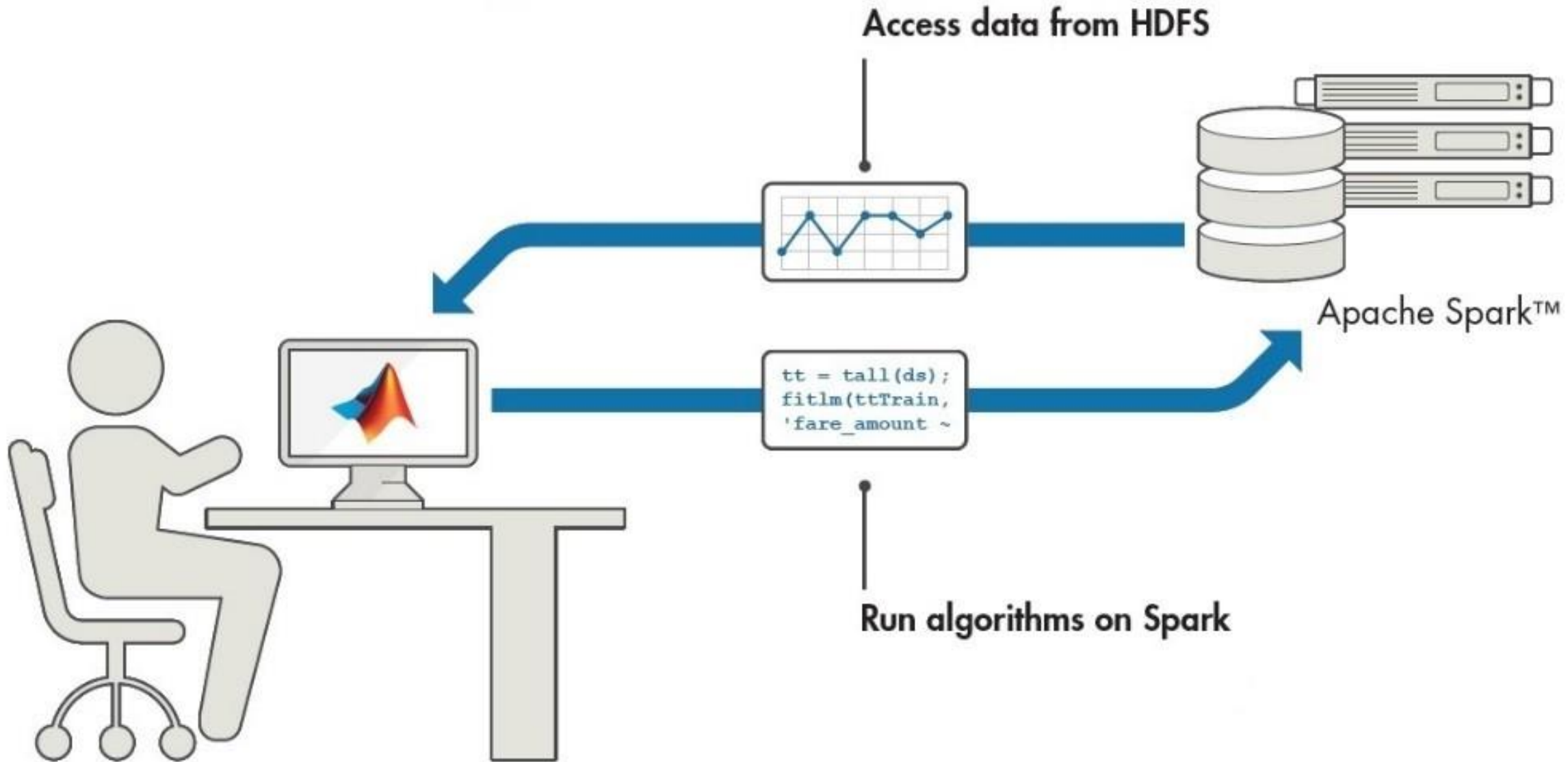
Result:

- We can now run the calculations in 10-15 minutes instead of the 5-8 hours we had before

How did we get there and leverage MathWorks

- **Maximum parallelization and unlimited scalability**
- **Low cost due to the usage of EC2**
- **Only pay for what we use to keep costs low but performance high**

How did we get there and leverage MathWorks



How did we get there and leverage MathWorks

- **Maximum parallelization and unlimited scalability**
- **Low cost due to the usage of EC2**
- **Only pay for what we use to keep costs low but performance high**
- **MATLAB spark integration very easy**
- **Approximately 20x faster runs**
- **Converting to Spark took 3 days thanks to MATLAB toolkit**

Lessons Learned

- **Make sure to read the documentation thoroughly before starting**
- **Spark configuration can be quite difficult, make sure to understand the concepts**
- **First optimize locally then parallelize in Spark**
- **Only use Spark when you really need it, MATLAB out of the box performance is enough for most usage**

Concluding Remarks

Tips

- Start off by trying some small things
- Read the documentation thoroughly
- Understand the spark concepts

Future plans

- MATLAB full PySpark (Python-Spark) integration for full leverage of spark platform. (f.e. Spark SQL to MATLAB)

The background is a solid orange color. On the right side, there is a large, stylized graphic consisting of several concentric, overlapping circles and arcs. The innermost part is a small circle, followed by a larger ring, and then a very large, thin ring that almost fills the right half of the frame. The overall effect is reminiscent of an eye or a target. The text "Questions?" is positioned in the lower-left quadrant of the image.

Questions?