

Future Processing

Evolution from a traditional data warehouse to lakehouse.
Implementaton of lakehouse on a given example.

www.future-processing.com





Speakers



Tomasz Krawczyk, big data architect
data lake architecture ■
tkrawczyk@future-processing.com ■



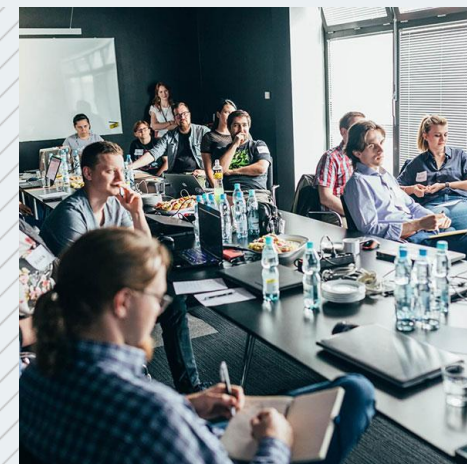
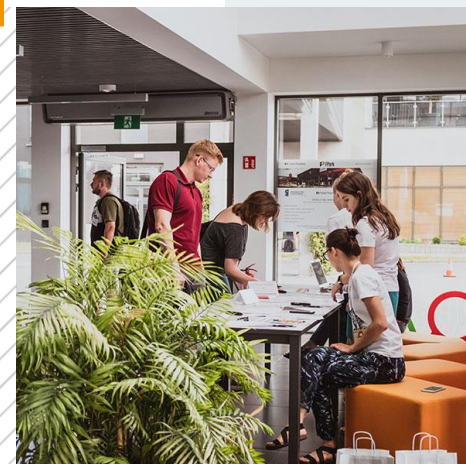
Michał Mierzwa, bi developer
■ reporting, statistical data analyses
■ mmierzwa@future-processing.com





Agenda

1. Outlook on a traditional data warehouse and data lake (~10m)
2. Comparing modern data warehouse to lakehouse (~10m)
3. Implementing lakehouse on a given example – demo (~30m)
4. Summary and useful links
5. Q&A (~10m)





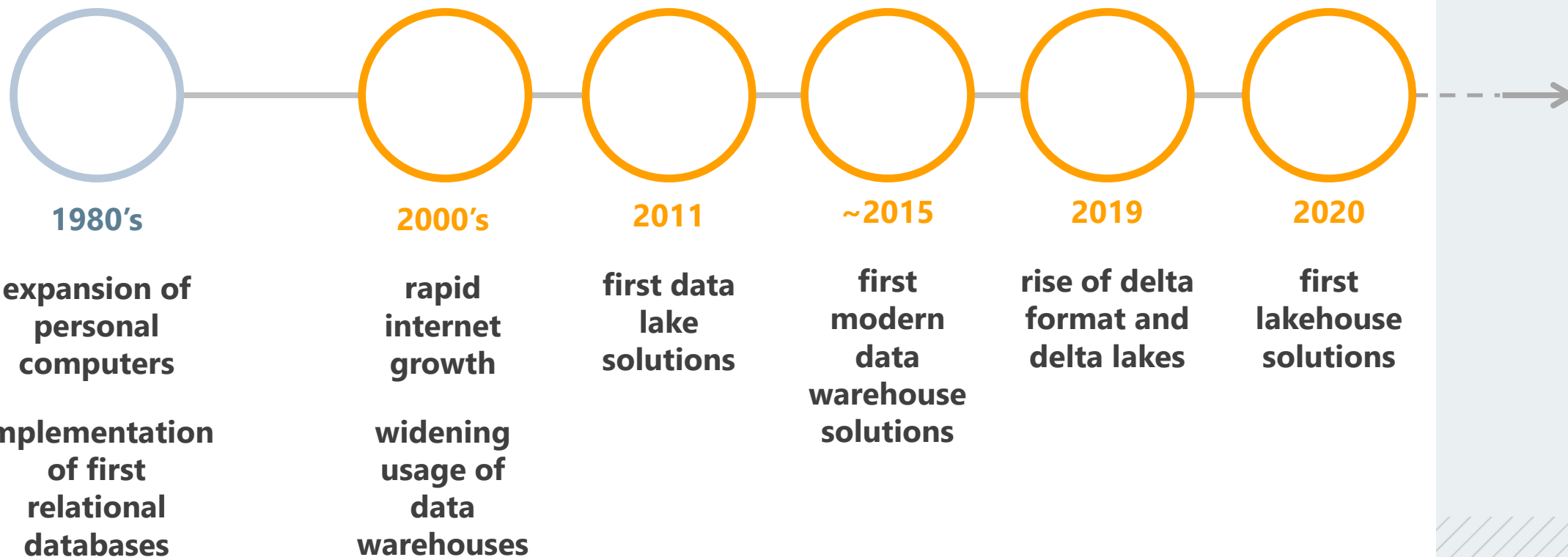
Key takeaways

1. Understanding concepts of a lakehouse and modern data warehouse.
2. Experiencing end-to-end implementation of a lakehouse (demo).





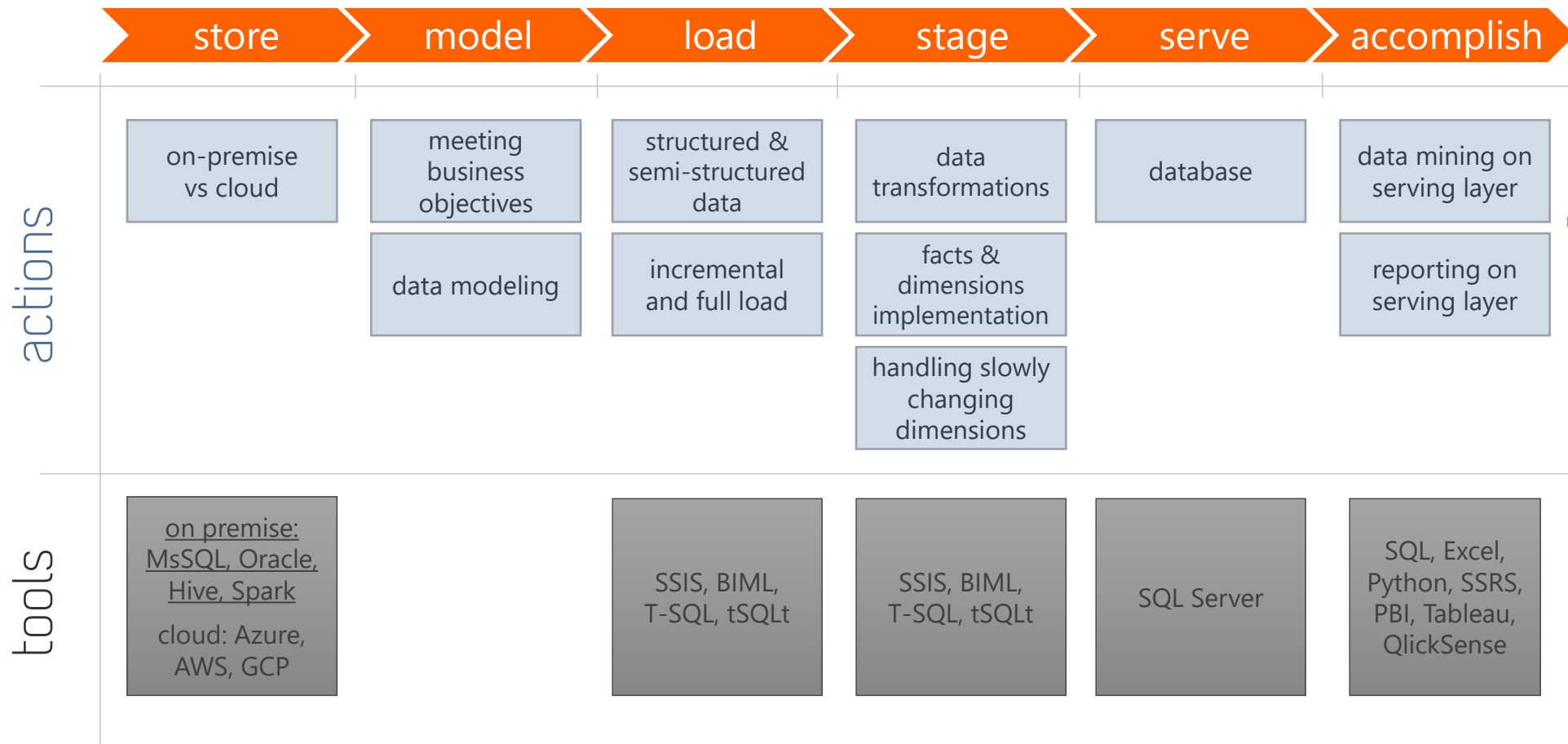
Brief history of data collection and processing





Building data warehouse

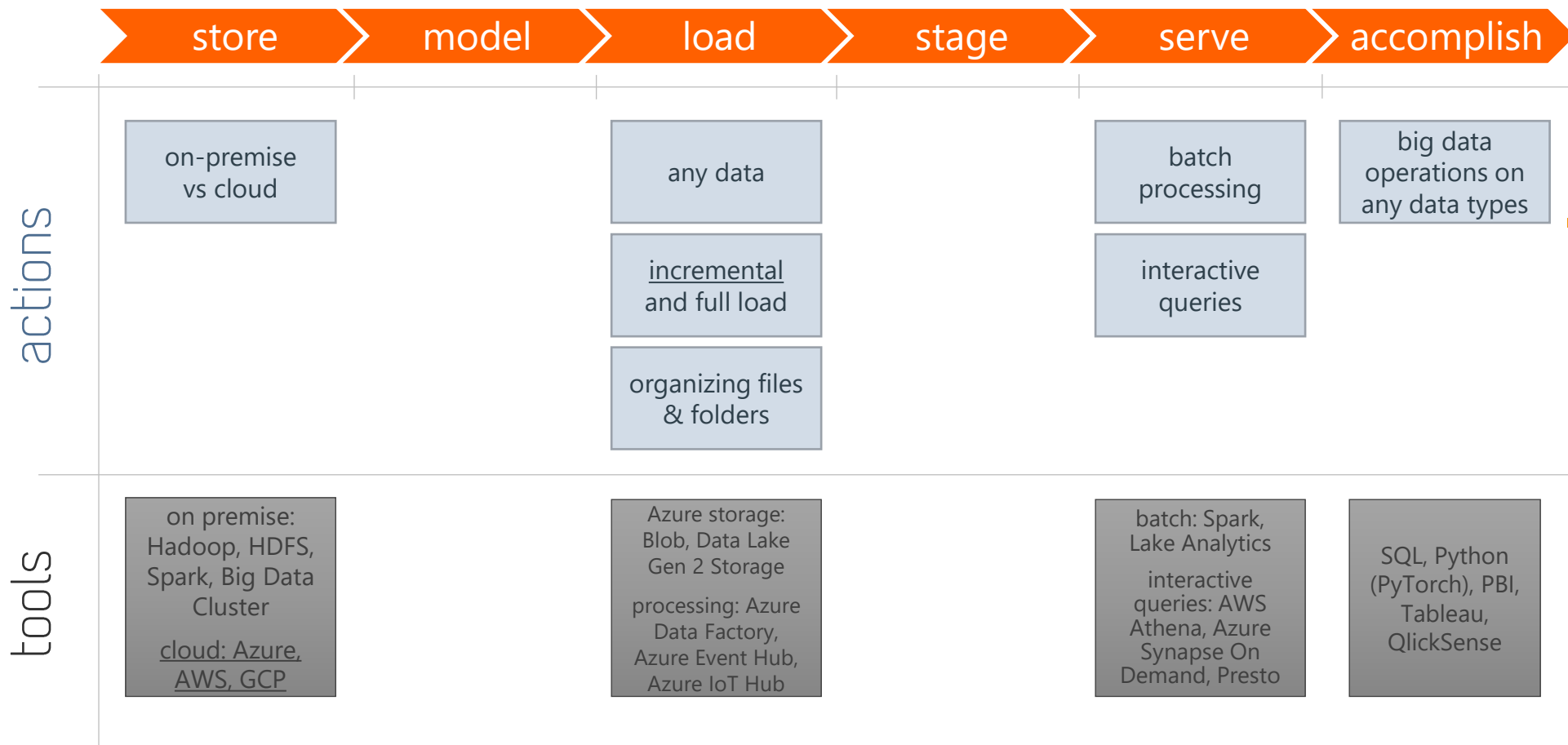
warehouse aggregates data into structured form to support business intelligence tools





Building data lake

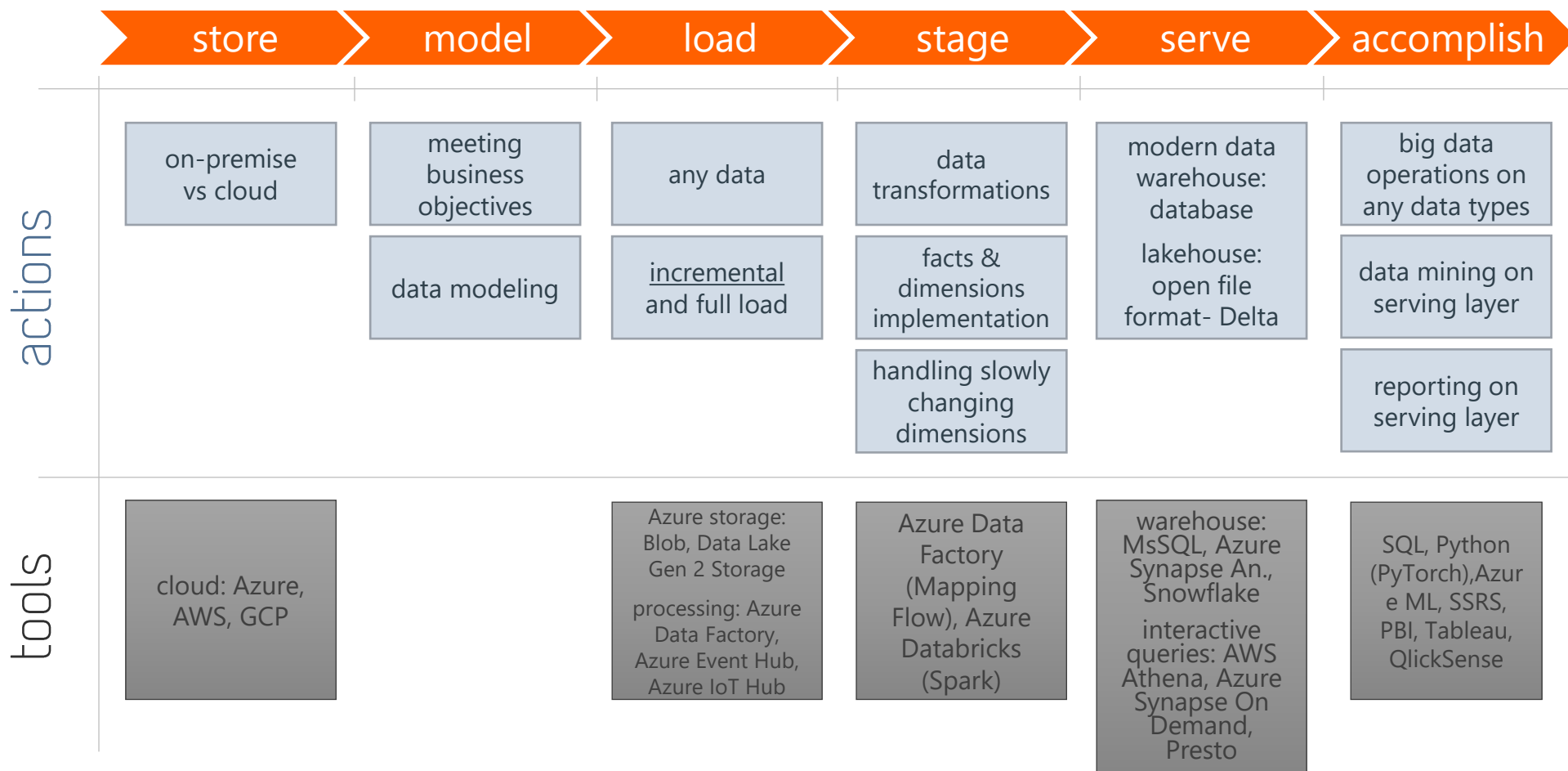
„If you think of a datamart (a subset of a data warehouse) as a store of bottled water – cleansed and packaged and structured for easy consumption – the data lake is a large body of water in a more natural state.” Pentaho CTO James Dixon (2011)





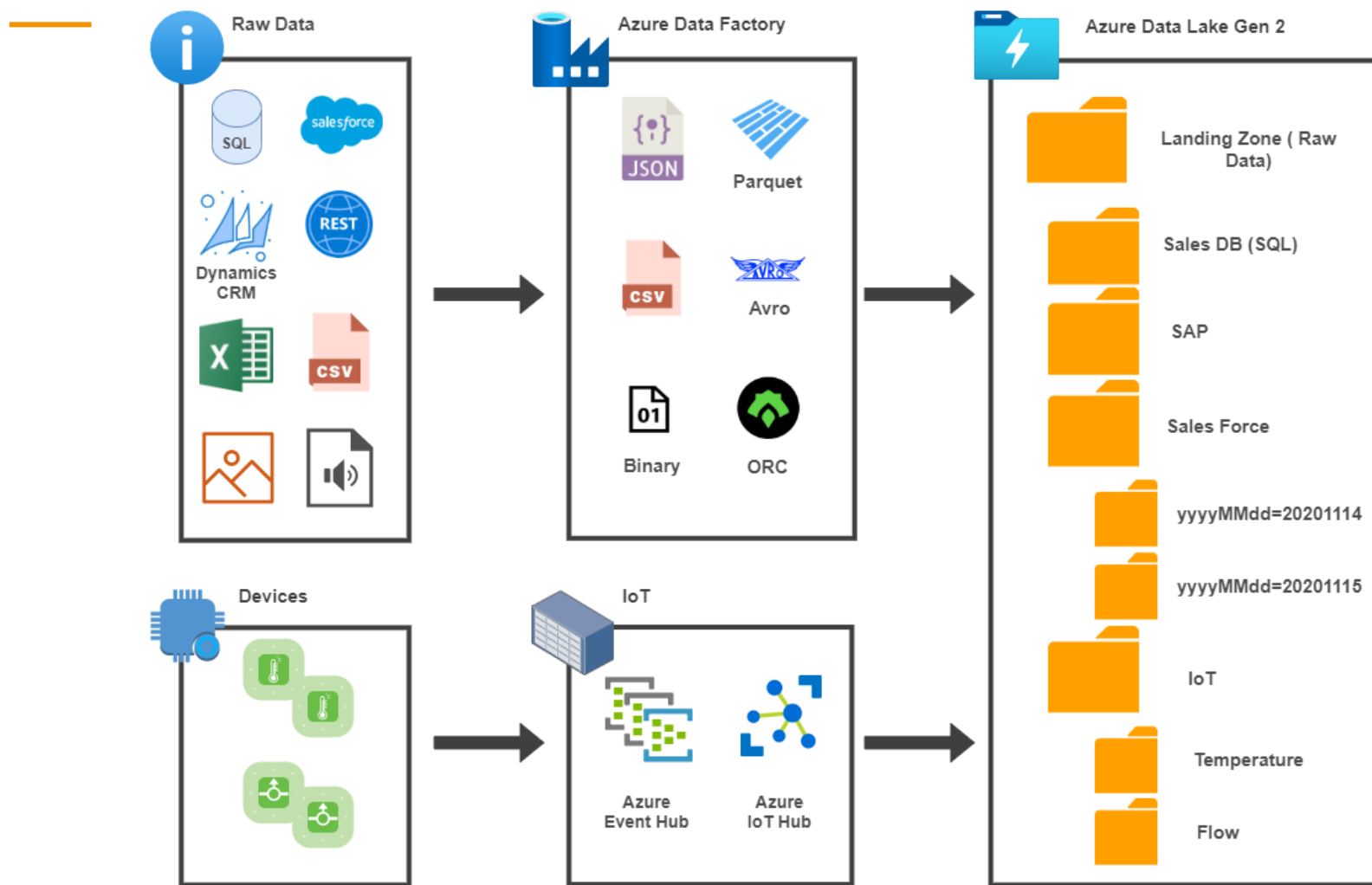
Building modern data warehouse / lakehouse

data lake utilizing most of data warehouse features





Building data lake

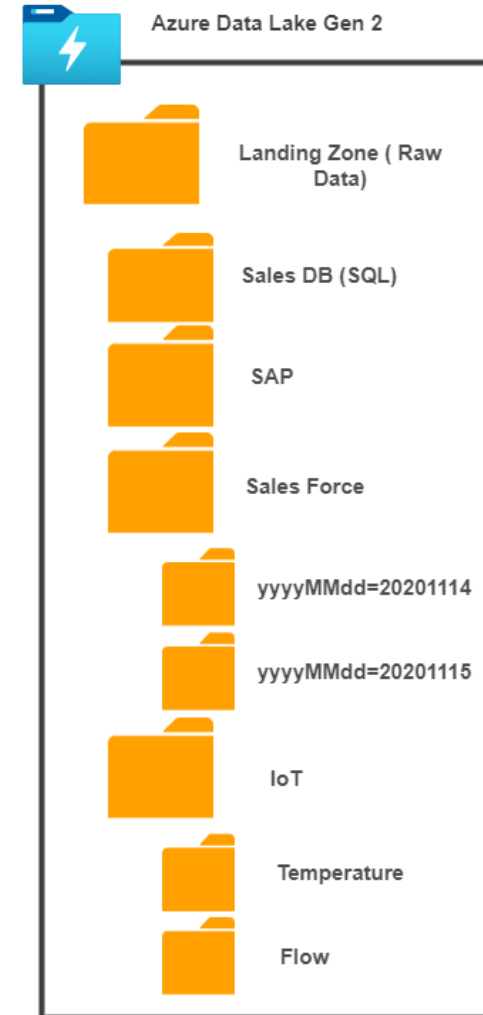




Building data lake

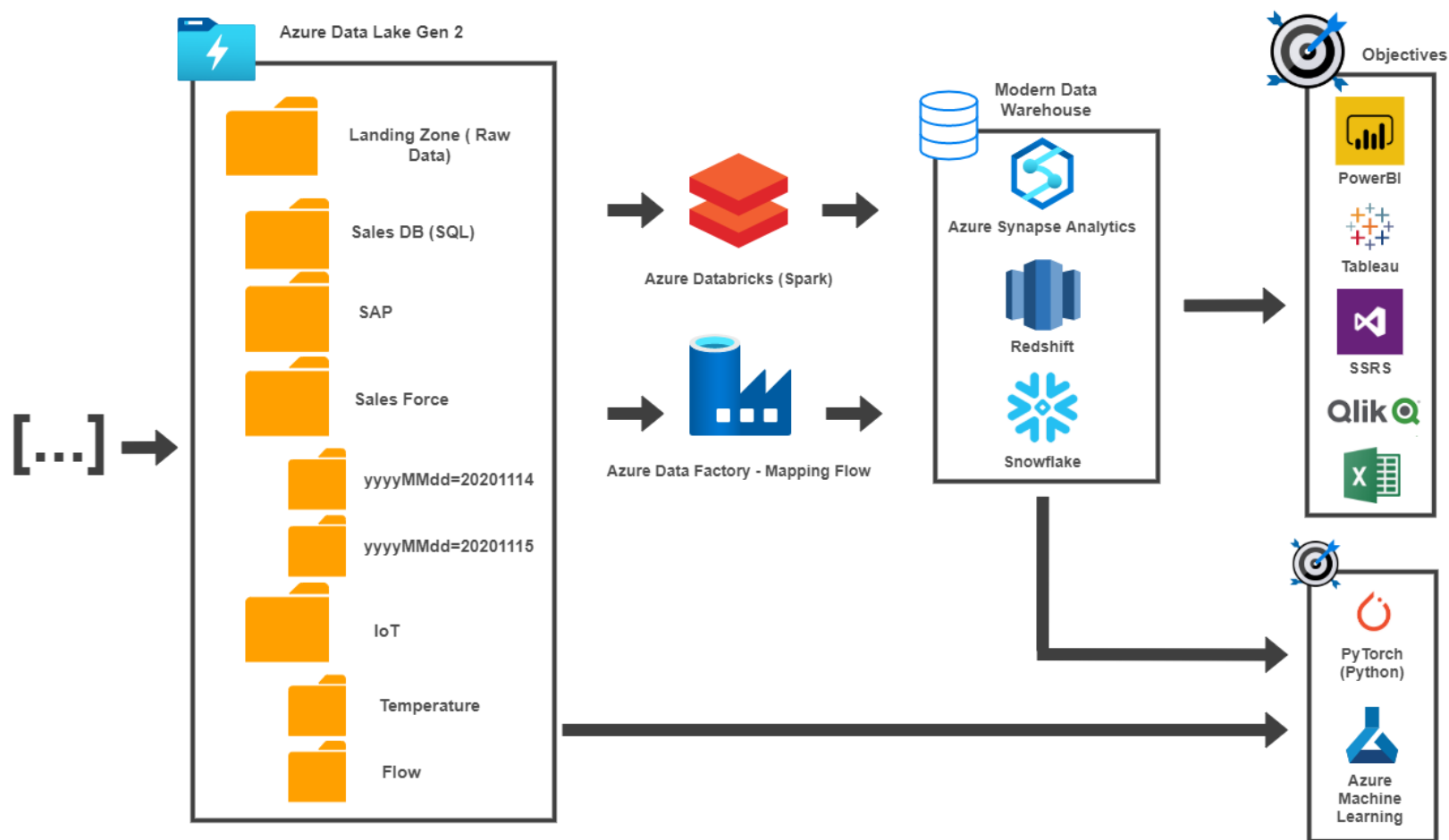


[...]



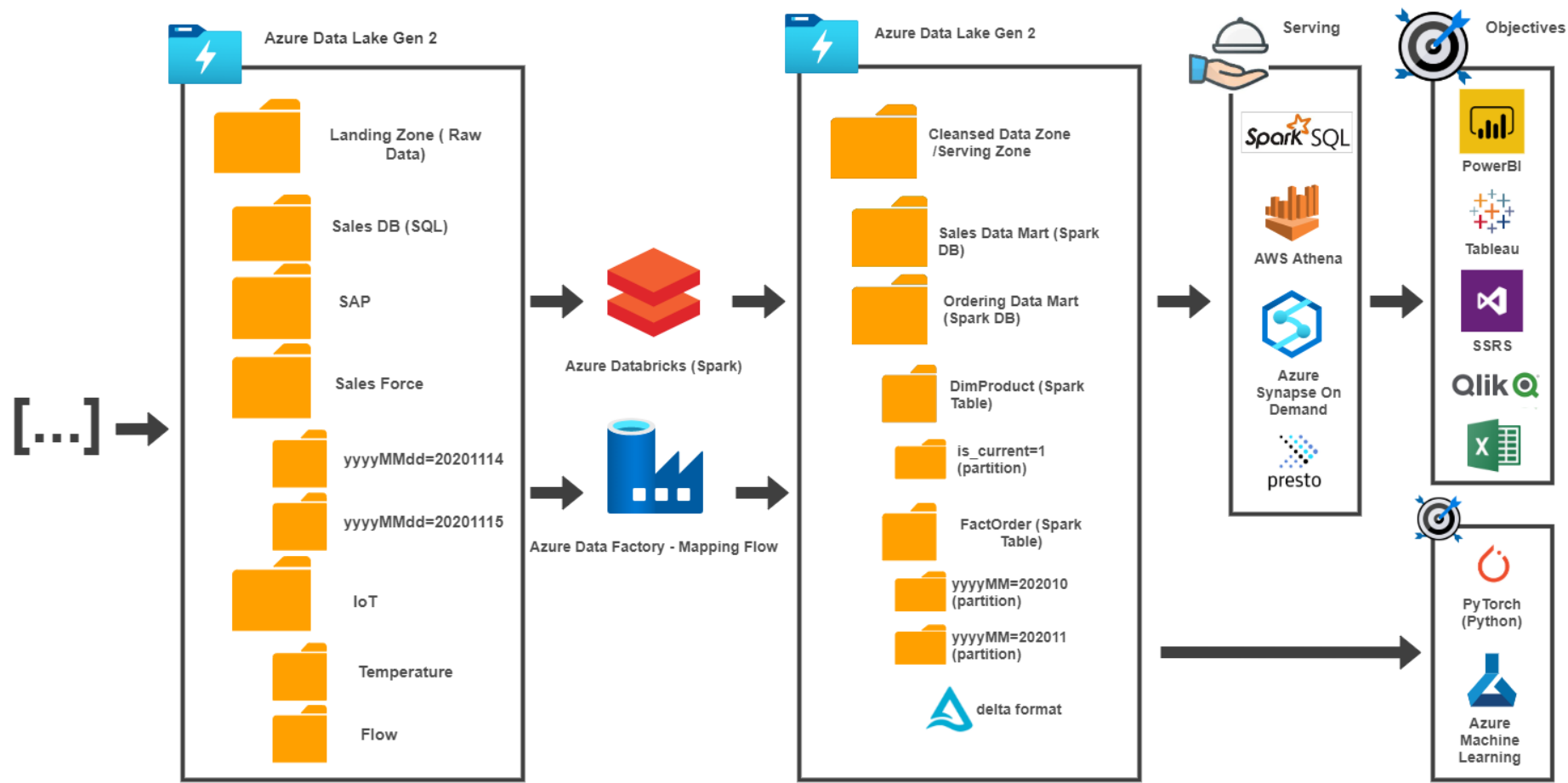


Building modern data warehouse





Building lakehouse





Why lakehouse approach

Enables a single combined cloud data platform



Increases data team efficiency



Unifies data warehousing and machine learning



Enables direct data access



Connects directly to BI tools





demo





Q&A





Summary

Over the last years volume of kept data has inflated, techniques of statistical analysis evolved and ways of storing information has developed. In order to maintain competitive advantage, commercial businesses need to reach out for cutting edge solutions.

Is lakehouse a state of the art platform? It surely unifies data formats, separates compute from calculation layers and offers most functionalities of its predecessors. Nonetheless, decision of choosing right architectural solution always require in-depth analysis.



Useful links

Future-Processing GitHub repository:

<https://github.com/FP-DataSolutions/SQLDay2020>

Other useful links:

<https://github.com/FP-DataSolutions/DeltaWarehouse>

<https://delta.io/>

<https://databricks.com/product/sql-analytics#overview>

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/on-demand-workspace-overview>

<https://github.com/microsoft/hyperspace>

<https://docs.databricks.com/security/access-control/table-acls/object-privileges.html>



Thank you
for your attention

 Future Processing

ul. Bojkowska 37A

+44 845 805 7479

www.future-processing.pl

