Doctusoft

Google Cloud Partner

BUILDING A **DATA WAREHOUSE** IN THE **GOOGLE CLOUD**

Kassai Csaba - Lead Data Architect

Merkur

Retail BI system

Processing and storing sales transactions in real time, in order to do:

- **Performance metrics**
- **Demand prediction**
- **Logistic optimization**
- **Collecting and selling insights**

Dagrofa

Merkur

Motivations

2 day latency

Expensive licences/infrastructure

Scalability problems

Cumbersome/slow development

Dagrofa

Merkur

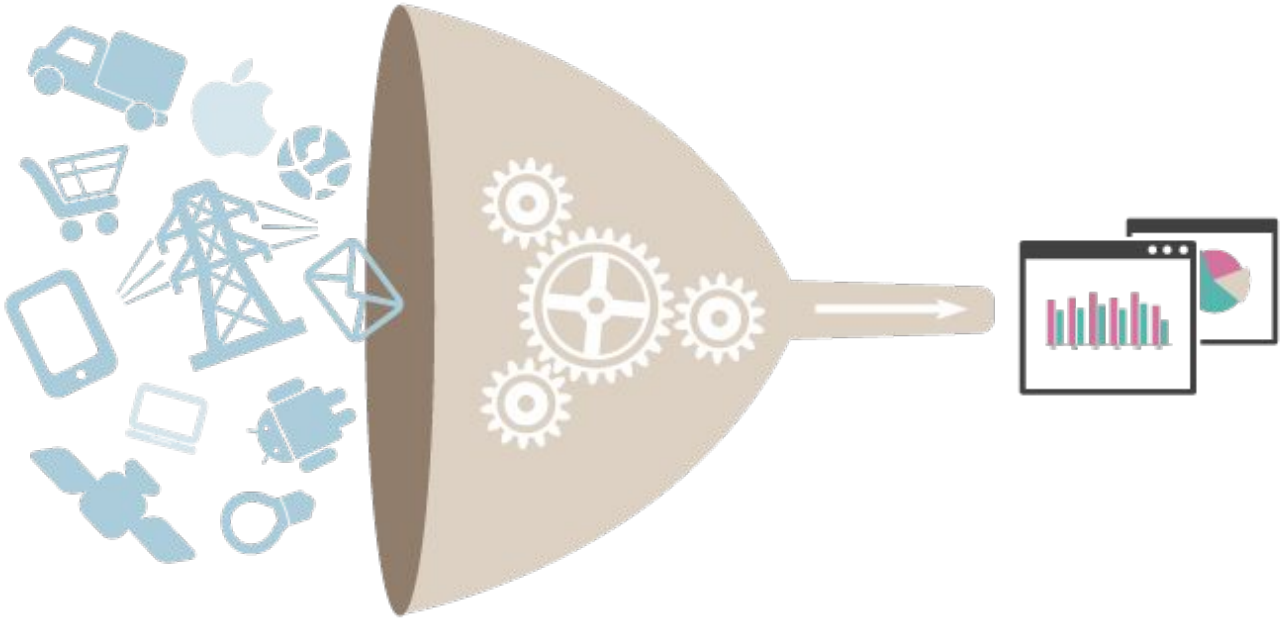
Users

- Stores - Performance
- In house users - Marketing
- Vendors - Market data

Dagrofa

Architecture

Data warehouse architecture



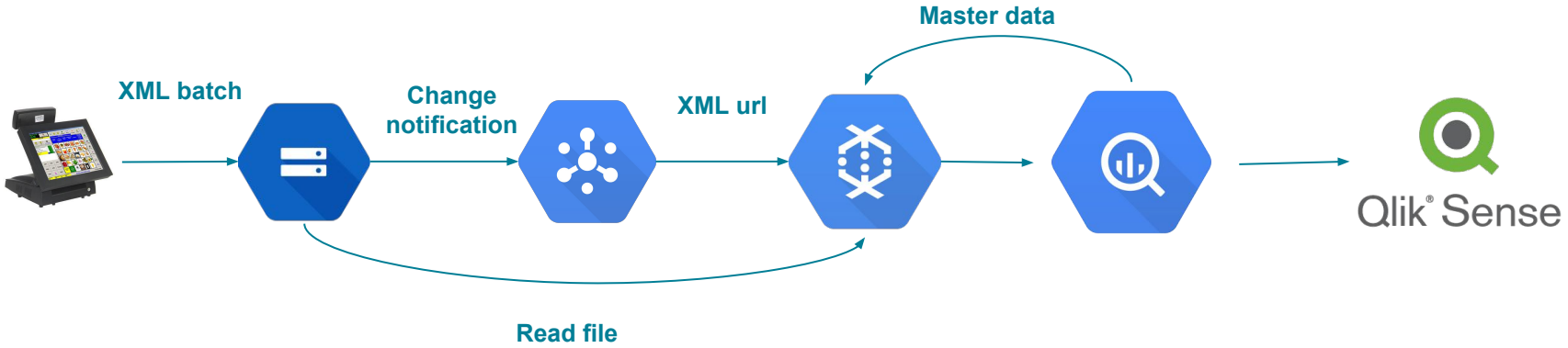
Merkur

POS data

~ 5 TB POS data in BigQuery

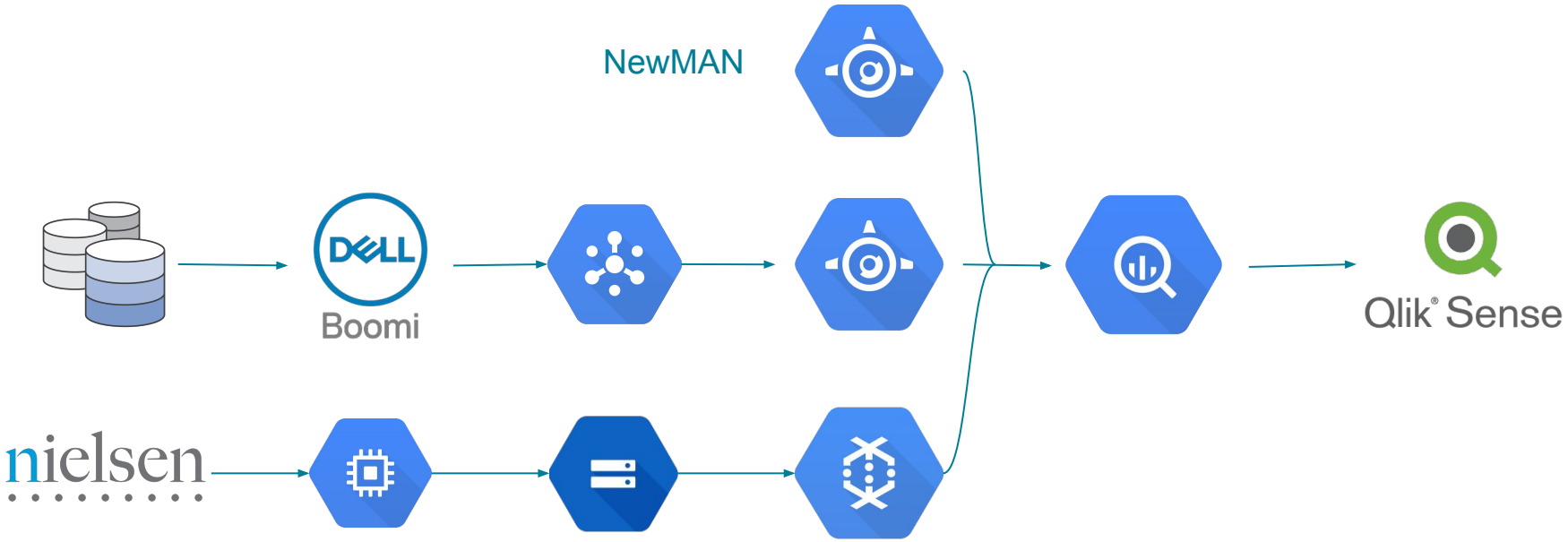
~ 6 million new transaction line daily

~ 5-10 minutes latency



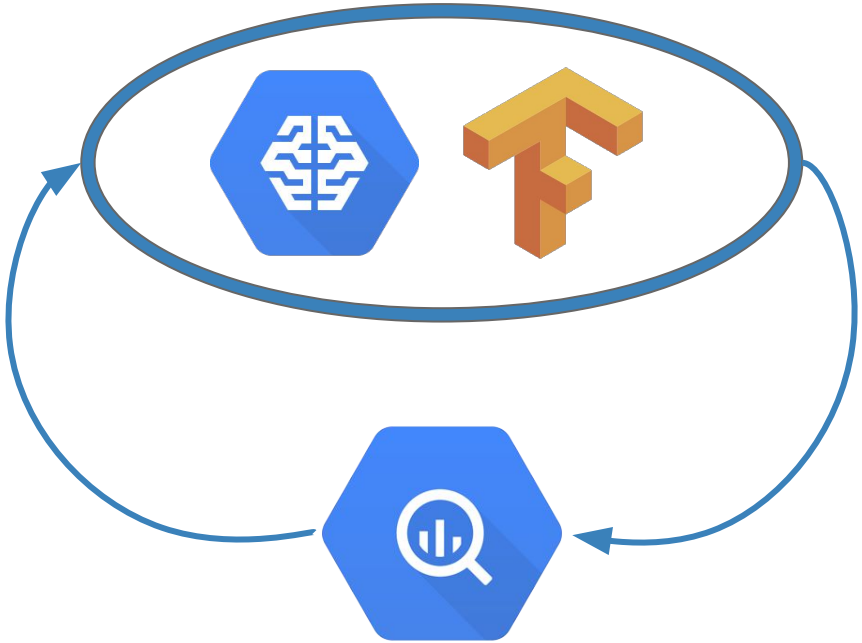
Merkur

Master data



Next steps - Predictive analytics

- **Demand prediction for stock optimization**
- **Personalized coupons in loyalty program**

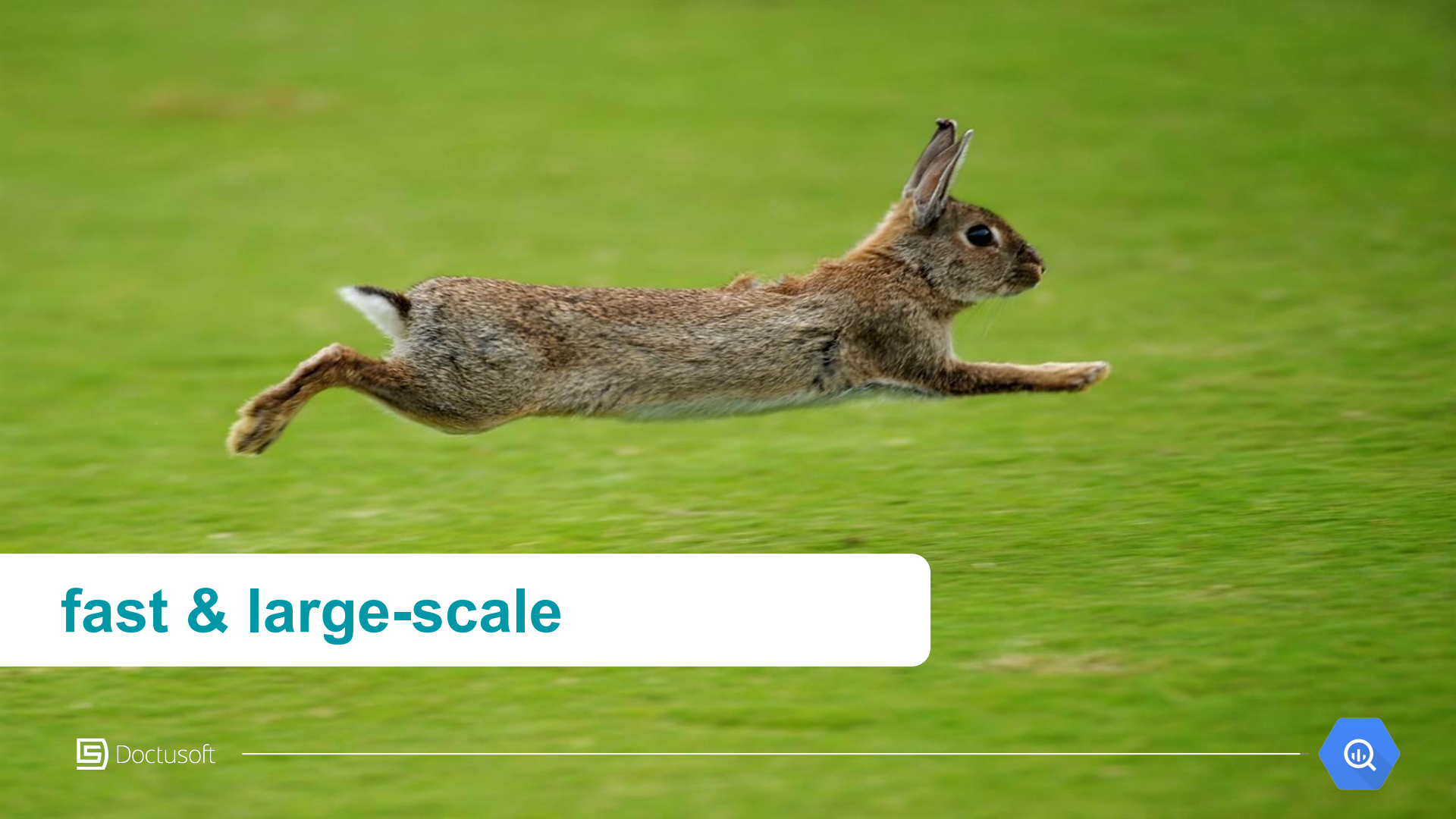



Questions?
csaba.kassai@doctusoft.com

A fast, economical and fully-managed enterprise data warehouse for large-scale data analytics

enterprise data warehouse

enterprise data warehouse



fast & large-scale

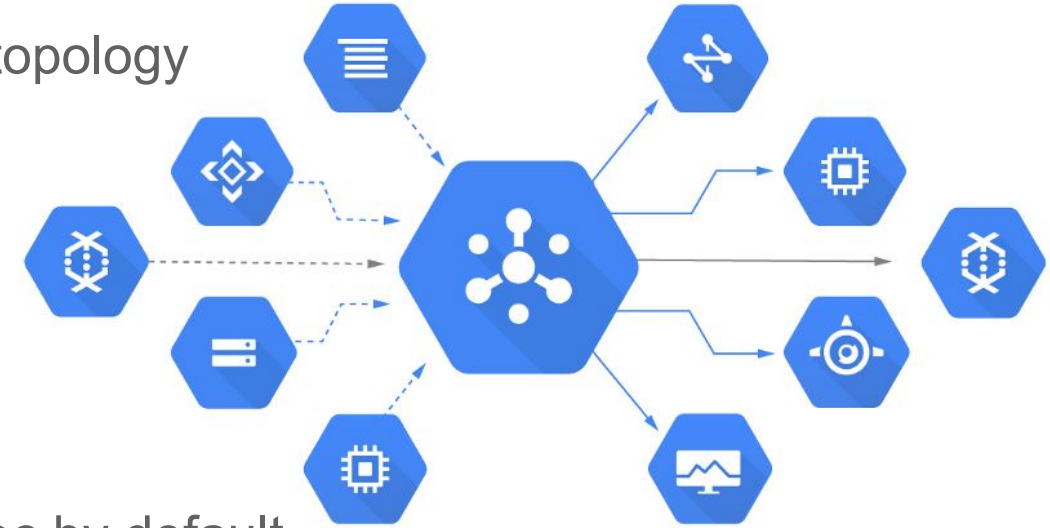

fully-managed

economical

Ingest

Google Cloud Pub/Sub

- Messaging - Many-to-many topology
- Topic - subscription model
- No-ops
- At-least one delivery
- Rest API
- Scalable: 10000 message/sec by default



Transform

Google Cloud Dataflow

- Apache Beam - Unified programming model for batch and stream processing
- Fully-managed
- Scalable

Sources

Ideal world

Real world

Data model

Slowly changing dimensions

New Store

Store

code	name	manager	validFrom	validTo
5	data store	K. CS.	2016.01.23.	null

Data model

Manager K. CS. -> T. T.

Store

code	name	manager	validFrom	validTo
5	data store	K. CS.	2016.01.23.	**2016.05.23.**
5	**data store**	**T. T.**	**2016.05.24.**	**null**

Store closing

Store

code	name	manager	validFrom	validTo
5	data store	K. CS.	2016.01.23.	2016.05.23.
5	data store	T. T.	2016.05.24.	**2017.06.02.**

New Store

Store

code	name	manager	validFrom	deleted
5	data store	K. CS.	2016.01.23.	false

Manager K. CS. -> T. T.

Store

code	name	manager	validFrom	deleted
5	data store	K. CS.	2016.01.23.	false
5	data store	T. T.	2016.05.24.	false

Store closing

Store

code	name	manager	validFrom	deleted
5	data store	K. CS.	2016.01.23.	false
5	data store	T. T.	2016.05.24.	false
5	data store	T. T.	2017.06.02.	true

Data model

View for standard type-2 history

```
#standardSQL
SELECT * FROM (
     SELECT *,
            LEAD(valid_from) OVER (PARTITION BY code ORDER BY valid_from) AS valid_to
     FROM store )
WHERE item_category_deleted = FALSE
```

code	name	manager	validFrom	validTo
5	data store	K. CS.	2016.01.23.	2016.05.23.
5	data store	T. T.	2016.05.24.	2017.06.02.