

Autonomous Database and Data Lake, a match made in... the Cloud Dr.-Ing. Holger Friedrich





#### Introduction

### Conclusions



## Puddles

3

#### Lakes

### Cloud





- Implementation & consulting services in CH
- Experts for
  - Data Warehousing/Big Data,
  - Business Intelligence/Advanced Analytics
- Tech focus on
  - Oracle on-premises and
  - Oracle and Snowflake in the cloud
- Wide area of industries
- Multiple specialisations
- Motto: Get Value From Data
- Web site: www.sumit.ch (in German)

#### sumIT AG





Specialized Oracle Business Intelligence Foundation



Specialized Data Warehousing





### **Dr.-ing Dipl.-inform Holger Friedrich**

- Informatik diploma from Karlsruhe Institute of Technology (KIT)
- Ph.D. in Robotics and Machine Learning
- 20+ years of experience with database technology
- Expert for
  - Data Integration
  - Data Warehousing / Big Data,
  - Advanced Analytics and
  - Business Intelligence
- CTO of sumIT AG



Oracle ACE for Data Warehousing / Analytics / Business Intelligence

y (KIT) Learning atabase technology





### 

© 2020 SUMIT AG 4





/



A data lake is a system or <u>repository of data</u> stored in its natural/raw format,<sup>[1]</sup> usually object <u>blobs</u> or files. A data lake is usually a single store of all enterprise data including raw copies of source system data and transformed data used for tasks such as <u>reporting</u>, <u>visualization</u>, <u>advanced</u> <u>analytics</u> and <u>machine learning</u>. A data lake can include structured data from relational databases (rows and columns), semi-structured data (<u>CSV</u>, logs, <u>XML</u>, <u>JSON</u>), unstructured data (emails, documents, PDFs) and binary data (images, <u>audio</u>, video).<sup>[2]</sup> © WIKIPEDIA

#### **Data Lake Definition**







#### **Data Puddles**







### **SUM/IT** Traditional Information Management Architecture











#### • 9i — read-only

- 10g write-back
- 11.2 inline pre-processing

create table dummy cust id number, cust name varchar2(20), limit number(10)) organization external type oracle loader default directory dummy dir access parameters

fields terminated by ",")

location ('dummydata.txt'));

#### Improvement... External Tables

#### Non-Partitioned



Pre-processing, loading, parsing, and processing on-the fly













- shared-nothing
- massively parallel
- scalable
- cheap storage
- non-relational
- schema on read
- scores of new
  - -tools
  - formats
  - players



### **HADOOP Facilitating Data Lakes**









- Enhanced External Tables
- Oracle HADOOP connectors
  - directly on HDFS
  - via HIVE







All processing in the database as before

#### Making HADOOP Accessible to The RDBMS









- 12c (12.2 in 2016/17)
- partitioned XTs, saving I/O through
  - pruning
  - partition-wise joins





Making the database data-lake-ready on-premises

#### Improving External Tables for Data Lakes







### **Moving Computation out of RDBMS**

Further Enhanced External Tables

NoSQL

- Big Data SQL
  - directly on HDFS
  - -via HIVE
  - on Kafka
  - on NoSQL

kafka



#### SQL pushdown to HADOOP Cluster





### **Adapted Information Management Architecture**





e	
nce	
nent	
_	
_	
k	
_	
_	
_	
_	
on	
_	
_	
_	
_	
_	
_	
_	
10	
14	

- Virtualized infrastructure
- Platforms and Software as a Service
- Automation everywhere
- Cheap object storage

#### And Then... The Cloud

















e.

### **Cloud Data Lake Architecture**

© MICROSOFT







### **Oracle Cloud Warehousing**







### External Tables on Object Storage

- for multiple Clouds
- partitioned
- hybrid



### **External Tables to The Rescue... Again**







- Credentials
- XT on Data Lake
- Partitioned XT
- Hybrid XT
- Offloading

#### Demo

# 





 Persistent staging layer for cloud DWH SQL Data Hub on Data Lake Data Offloading





- Savings
- Simplicity
- Scalability
- Access
- Agility
- Performance
- Versatility

#### Advantages





- Cloud & Data Lake are natural matches
- DWH & Data Lake play well together
- External Tables are key concept for DWH/Data lake integration
- Partitioned & hybrid XTs
  - open up new use cases
  - provide secure, perfomant SQL access
  - combine best of both worlds



**Recommendation: have a** closer look & improve your **DWH** architecture

#### Conclusions

