Big Data within Automotive Testing
Bringing Analysis to the Data

ASAM US Workshop, Novi

Dr. Ralf Nörenberg
ralf.noerenberg@highqsoft.de

Bringing Analysis to the data
Content

Disclaimer
What is this presentation about?

An abstract view on big data technologies & objectives

A view on core objectives and limitations within automotive testing industries
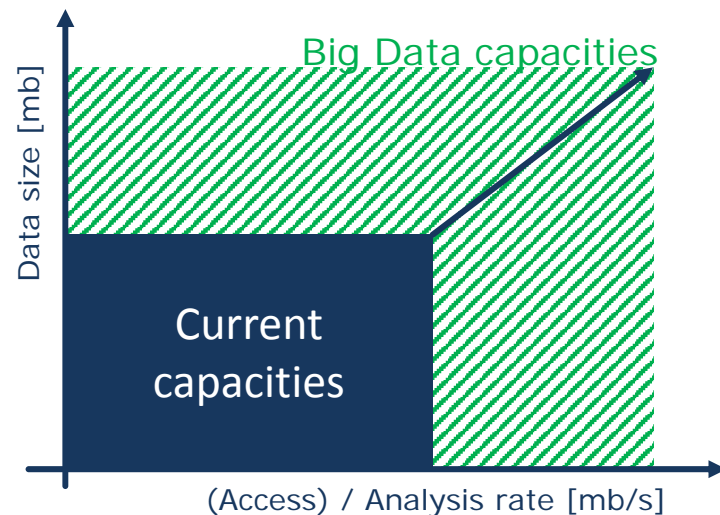
Current "big-data" industry use-cases

Approaches on technology transfer into the domain

Big Data Overview
What are the objectives of Big Data?

Objective: Capacities

Objectives: On-Demand Results

Big Data capacities

Data size [mb]

Current capacities

(Access) / Analysis rate [mb/s]

Big Data
1. Integration of various data (types)
2. Handling of variable data and content
3. Quick adaption to new scenarios and individual case analysis
4. Iterative and explorative analyses in reaction to missing or changing requirements
5. Optimized for flexibility

Big Data (Storage) enables Data Mining (Access and Analysis)
Explore huge amounts of data from various systems and domains to derive models / patterns / results. We speak of a transformation of data into information / knowledge.

Big Data Overview
What's the challenge with Big Data?

- Huge amount of data to be stored

- Exponential Growth Rate

- Challenges in Access and Analysis
  - Volume   (Scale of Data)
  - Velocity   (Analysis of Data)
  - Variety   (Different forms of data)
  - Veracity   (Uncertainty of data)

▸ How do we store the data (data structure / semantics)?

▸ How do we provide (enterprise) scalability?

      ▸ We've got dependencies to specific Use-Cases!

Bringing Analysis to the data
Content

ODS (Open Data Service) Overview
What is ODS?
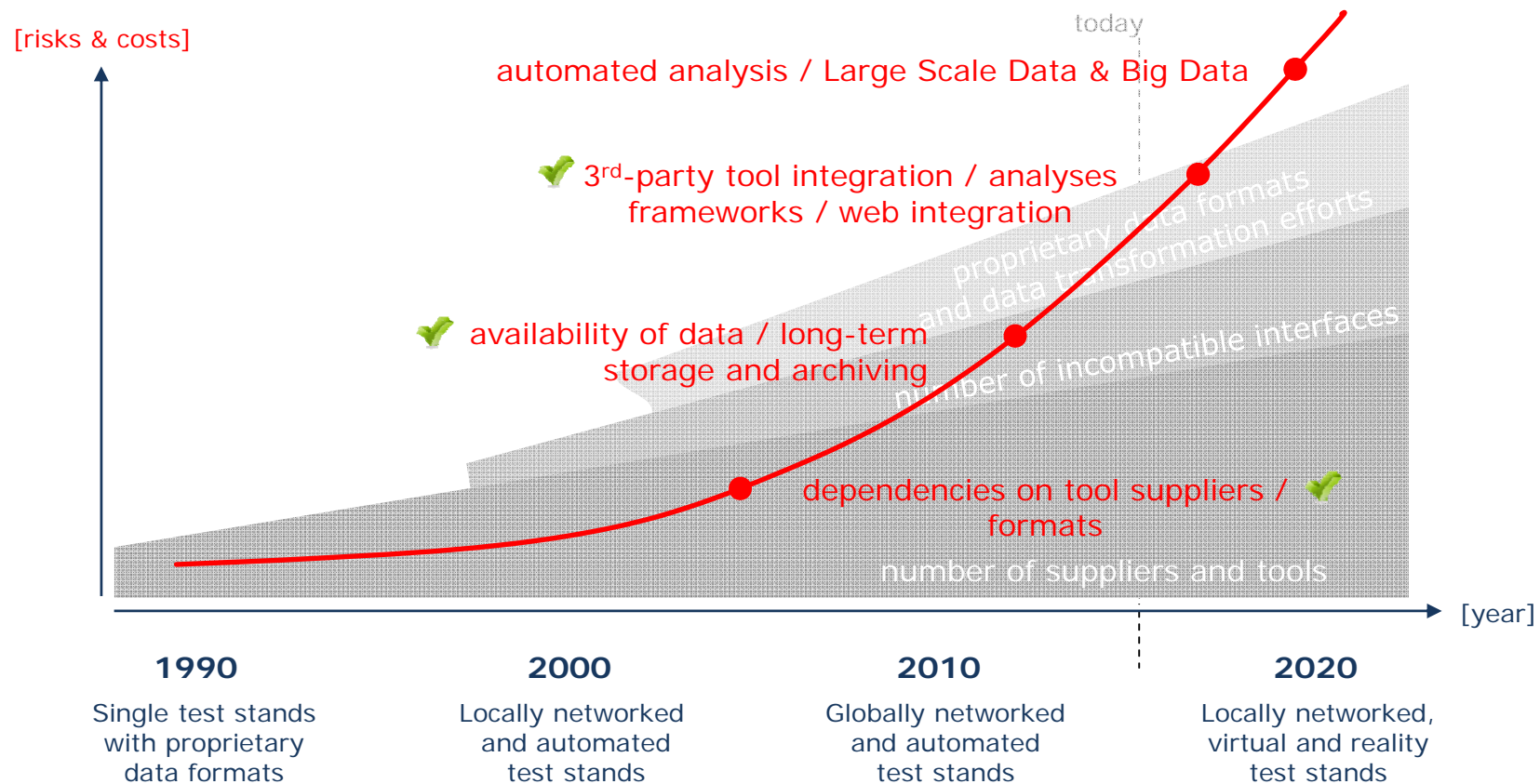
The Open Data Service (ODS) ….
- is a generic definition of measurement related objects and its relations
- covers all domains
  - Noise-Vibration-Harshness
  - Engine
  - Road-Load
  - Crash
  - …

1. Base Model: Standardized data
2. Application Model: specific derivations of the base model are allowed
3. ATFx Files: exchange data format for data exchange

ASAM ODS Base Model 5.3.0

Application Model	
Organization / Description	Test Test Object Test Environment Test Sequence
Measurement	Measurement MeasurementQuantity Submatrix LocalColumn **ExternalComponent** **AoFile**
Physics	Units Physical Dimensions Quantities
Flexibility	Any ParameterSet Parameter

ODS (Open Data Service) Overview
The challenges in ODS

[risks & costs]

automated analysis / Large Scale Data & Big Data

today

3rd-party tool integration / analyses frameworks / web integration

proprietary data formats and data transformation efforts

availability of data / long-term storage and archiving

number of incompatible interfaces

dependencies on tool suppliers / formats

number of suppliers and tools

[year]

1990
Single test stands with proprietary data formats

2000
Locally networked and automated test stands

2010
Globally networked and automated test stands

2020
Locally networked, virtual and reality test stands

Big Data vs. ODS
A comparison

Big Data	ODS
1. Un- / Semi-structured Data	1. Structured Data
2. Handling of variable data and content	2. Focus on "high value data"
3. Quick adaption to new scenarios and individual case analysis	3. Identification of repeatable operation and processes
4. Iterative and explorative analyses in reaction to missing or changing requirements	4. Definite requirements of specialists
5. Optimized for flexibility	5. Optimized for quick access and given analyses (which depend on data structure)

Find a BIG ODS solution technology-wise:

Big Data technologies are the wheels, we need a tires (includes use-cases).

Bringing Analysis to the data
Content

1 Disclaimer and Big Data

2 ODS (Open Data Service) Overview

3 Prerequisites to ODS

4 State-of-Today: Big ODS Projects (Large Scale Data)

5 State-of-Tomorrow:
      Integration of Big Data into ODS (BIG ODS)

ODS System Solutions
General use-cases and facts (keep in mind)

1. ODS data is only beginning to have distributed locations
   - One location, (hopefully) minimal data redundancy

2. Test stands generate same data for a long period of time
   - Changes are of structural nature (new format)
   - Flexibility within data schema rarely required (e.g. new channels)

3. Established file formats like MDF4.x for storage and analysis
   - Data is itemized by ODS for access & search (meta-data)
   - File is generally required in whole (third party tool)

4. Analysis is done locally by local experts
   - very specific and complex (vs. string-based or analysis on semi-structured data)

5. ODS needs a single point of access
   - web-access primarily targets at meta-data in Oracle
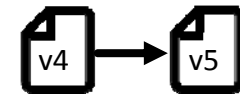
ODS System Solutions
Definining the use-case is essential

STEP 1: Business and Data Understanding
What do I want to do with my data?
Derive business cases and identify data available.

STEP 2: Data Organization & Preparation
How do I store my data?
Data Semantics, Validation, Data Organization / Structure

STEP 3: Data Analysis
Creation of models and analysis programs and its configuration.

STEP 4: Evaluation of Analysis Results
Evaluation, Comparison and Optimization of Analysis Results

STEP 5: Integration of all steps into an environment
Evaluation, Comparison and Optimization of Analysis Results
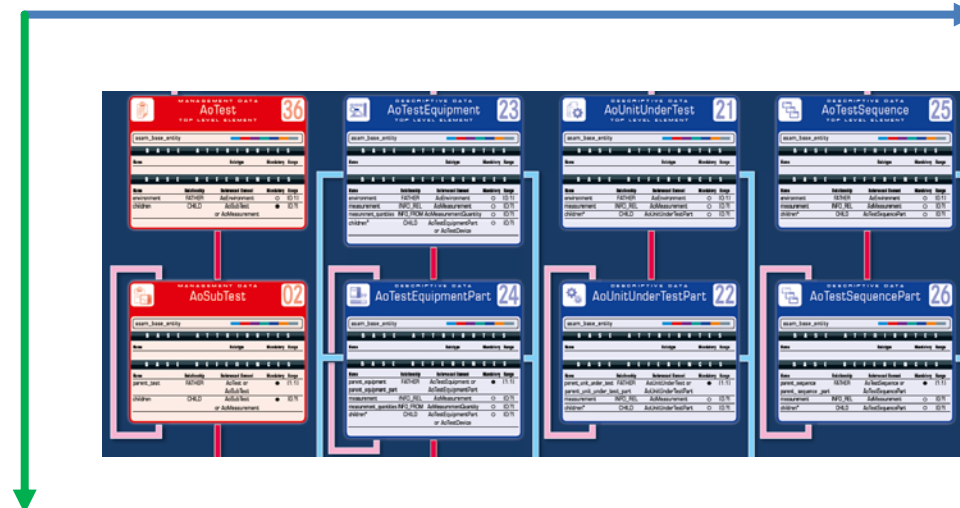
Iterative Process and an estimated

80%

of the whole project effort.

ODS System Solutions
Data Semantics

A broad structure in the application model...

...allows specific result-sets for searching (bound on use-case)

...allows all-in-depth analysis functionality

This topic is controversially discussed in the Europe / US. Data models here tend to store more information (than needed) into AoAny / AoParameterSets / AoParameter. This results in a loss of data semantics and access capabilities (leads to the necessity of an indexer).
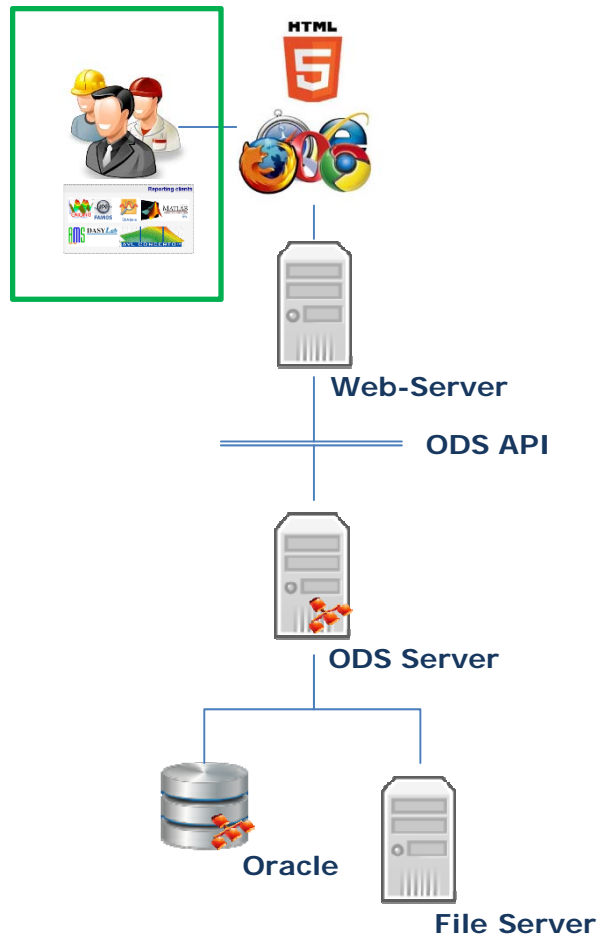
Bringing Analysis to the data
Content

State-of-Today: Big ODS Projects (I)
Integration of an Indexer (Elastic Search)

Web-Server

ODS API

ODS Server

Oracle

File Server

The project stores data of a number of test stands.

The project is based on a regular ODS setup:
- centralized Web-Server access for users
- Oracle and ODS Server on same hardware
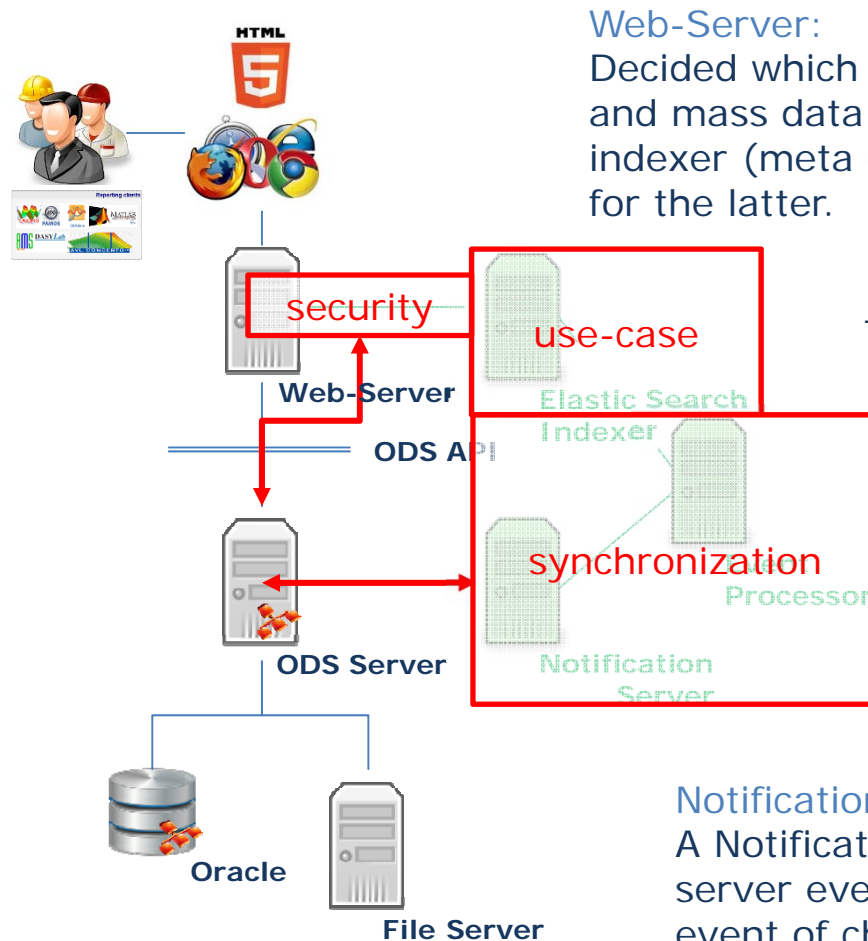- File Servers linked by "FILE_MODE=MULTI_VOLUME"

A web interface allows the user to look up measurements and add information to ODS according to the project use-cases.

Big Data Use-Case (Access):
- On-the-fly meta data search (auto-complete)
- Full-text meta data search
- ("Live Mass-Data Search")

State-of-Today: Big ODS Projects (I)
Integration of an Indexer (Elastic Search)

Web-Server:
Decided which calls are directed to the server (write and mass data access) and which calls go to the indexer (meta data read access) and handles security for the latter.

Elastic Search Indexer:
The indexer allows on demand response for the structure used (company use-cases). The indexer is synchronized with ODS.

Event Processor:
The event processor acknowledges notifications, schedules and executes the task of applying the change to Elastic Search.

Notification Server:
A Notification is generated based on import / any server event / user action / ..., therefore covers any event of change.

security

use-case

synchronization

Web-Server

ODS API

Elastic Search
Indexer

Processor

Notification
Server

ODS Server

Oracle

File Server

State-of-Today: Big ODS Projects (I)
In Numbers

- Project Data (processed)
  - Oracle   ~10 GB
  - Files   1.216.610 Files
  - Data-Root Mixed-Mode (BTF)   1.700 GB
- Files on File Server (last 1-2 years)
  - Files   43.000.000 Files
  - Volume   4.000 GB
- Files on tape archive (5 years before that)
  - Files   60.000.000 Files
  - Volume   6.000 GB

▸ Data in Oracle rather small, mass-data is stored externally

▸ Data Volume strongly increasing

▸ 1 ODS server instance and adequate IT infrastructure (file server/network)

▸ Analysis is mostly done manually (per file download)

▸ Currently the project will be enhanced to <u>automated analysis functionality</u>.

State-of-Today: Big ODS Projects (II)
Vehicle Fleet Testing (planned tests)

This **road-load data project** is set up in an environment of a car manufacture who has all vehicle fleet testing data (and most activities) centralized within one project.

The scope of this projects contains
- many customers (all internal development departments)
- all passenger vehicle types and almost all ECUs available
- all domains and data types
- over **500 testing vehicles** to be managed

Within ODS, this is the upmost state-of-the-art system to manage the test execution flow and standardized measurement data.

State-of-Today: Big ODS Projects (II)
Integration of Automated Analysis
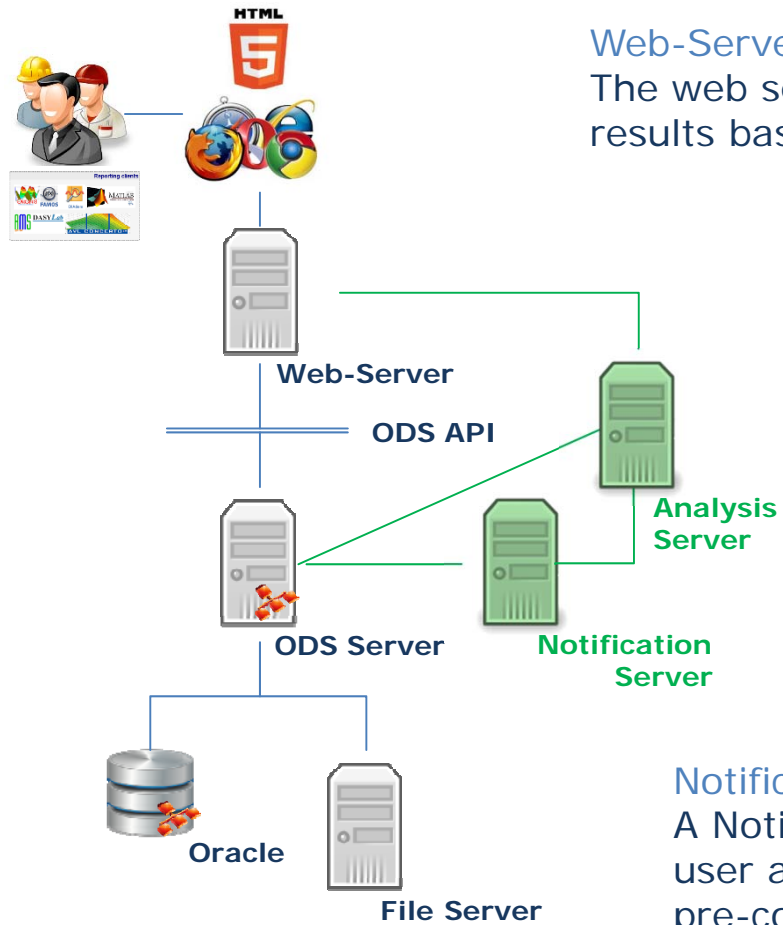
Web-Server:
The web server allows access to the data and analysis results based on the role of the employee.

Analysis Server:
The analysis server is a framework to execute a <u>measurement</u> with an <u>analysis program</u> (JAVA, Matlab, DIAdem, ...) to produce a <u>result</u> (measurement, report, email, ...). The analysis programs are defined by specialists and may be put into an order.

Web-Server

ODS API

Analysis Server

ODS Server

Notification Server

Oracle

File Server

Notification Server:
A Notification is generated based on import event / user action and reports measurements/tests based on pre-configured <u>MIME-types</u> and analysis programs.

State-of-Today: Big ODS Projects (II)
In Numbers

- Data Base Size (Oracle)   900 GB
- Files on File Server(s)
  - Files   39.346.529 Files
  - Volume   5.400 GB
  - Absolute growth   1.5-1.8 GB per day
- Files in tape archive
  - Files   58.462.941 Files
  - Volume   7.700 GB
- Analysis Servers
  - No. of Servers   1 in total
  - No. of Analyses   **210.000** per month
- ODS Servers
  - Productive ODS Servers   3 in total

▸ Data volume in Oracle and on file server is very extensive.

▸ In terms of data volume, analysis rate and enterprise setup, this is one of the most "beautiful" ODS projects

   ▸ Project is state-of-the-art and meets performance requirements

Bringing Analysis to the data
Content

1 Disclaimer and Big Data

2 ODS (Open Data Service) Overview

3 Prerequisites to ODS

4 State-of-Today: Big ODS Projects (Large Scale Data)

5 State-of-Tomorrow:
Integration of Big Data into ODS (BIG ODS)
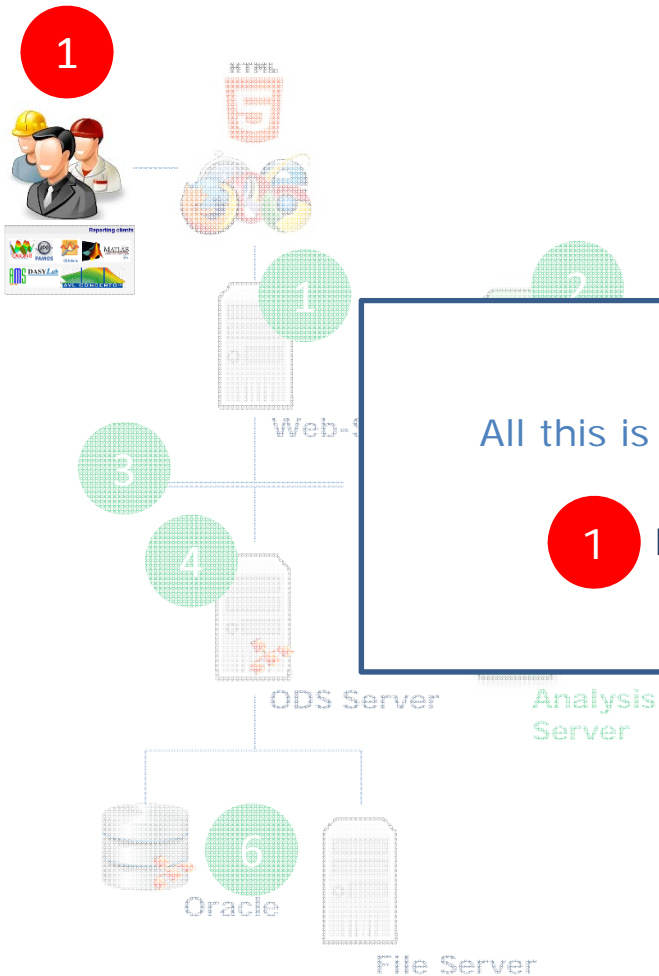
Big ODS Systems
State-of-Tomorrow
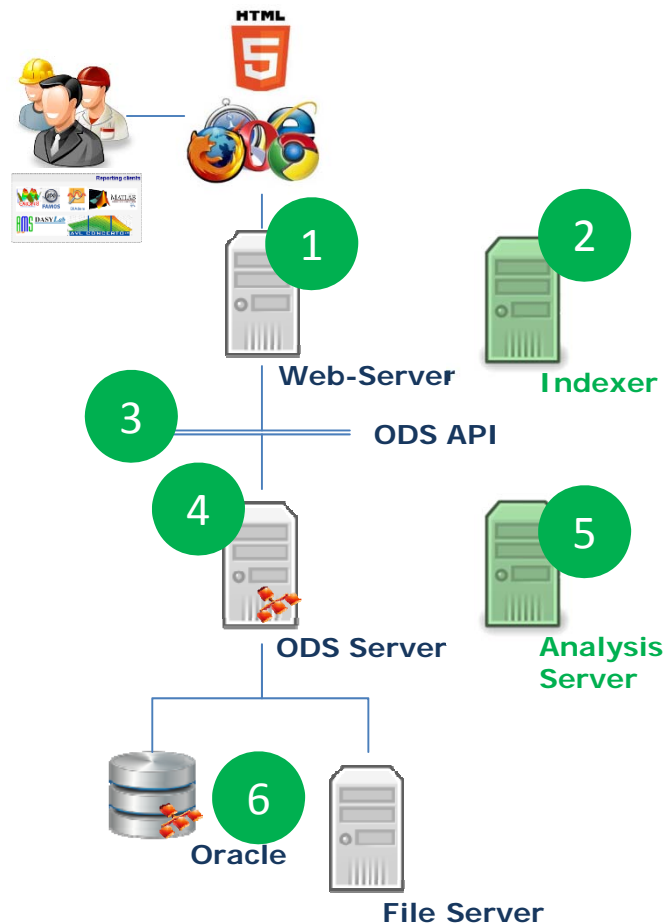
Current ODS projects are manageable and performant…and there's still room for improvement!

1. Integration of **Web-Services** for single-point of access (generic, specific work-flows)

2. Integration of Indexer Technology for quick data access (RESTfull user interface)

3. Integration
   • of Hadoop database interface
   • substitute transport protocols (Corba)

4. Integration of ODS server distributor

5. Integration of Analysis Server Frameworks (distributor?)

6. Integration of Big Data Storage (e.g. Hadoop)

Big ODS Systems
State-of-tomorrow: Conclusion

1

Current ODS projects are manageable and performant...and there's still room for improvement!

1. Integration of **Web-Services** for single-point of access (generic, specific work-flows)

2. ...chnology for quick ...r interface)

All this is great, but we do need:

1   BIG ODS use-cases needed!

...terface ...rotocols (Corba)

... distributor

5. Integration of (multiple) Analysis Server Frameworks (distributor?)

6. Oracle Adapter with Mass Data in Hadoop / Hadoop

Web-S...

ODS Server

Analysis Server

Oracle

File Server

Big ODS Systems: A vision
Road-Load Data from Vehicle Field Testing (unplanned)

Car drives full day with random scenarios.

| ODS based (meta data) |
| HDFS (mass data) |

Data is kept in ODS or ODS based big data technology as focus is on analysis.

"Everything" is recorded (time and senor) and directly stored in binary data.

Map-Reduce Jobs as analyses to allow

| Fleet Analysis, e.g. none-periodic events (particle filter burst) | Identification of planned driving scenarios described and tested before | ??? |

Thank you!

Any questions, suggestions and ideas?

HighQSoft GmbH
ralf.noerenberg@highqsoft.de