Public

Big Data in ASML

Durk van der Ploeg

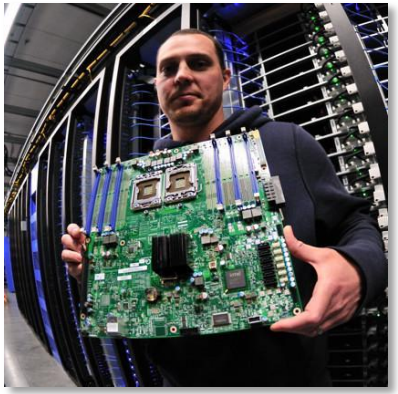
ASML System Engineering – Product Industrialization,

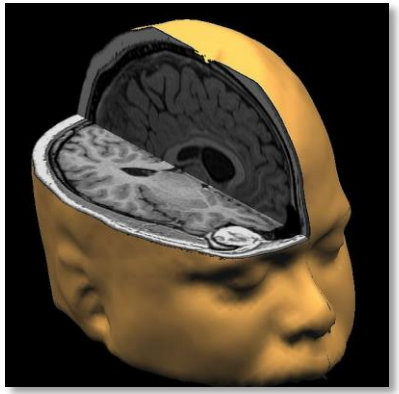
October 7, 2014 | SASG @ NTS Eindhoven

- ASML Company
- (BIG) Machine Data in ASML
- Management and infrastructure for big data
- Exploit the big data value
- Conclusion

ASML Company

It’s hard to imagine a world without chips





ASML makes the machines for making those chips

- Lithography is the critical tool for producing chips
- All of the world's top chip makers are our customers
- 2013 sales: €5.2 bln
- Payroll: ~13,600 FTEs

ASML makes the machines for making those chips

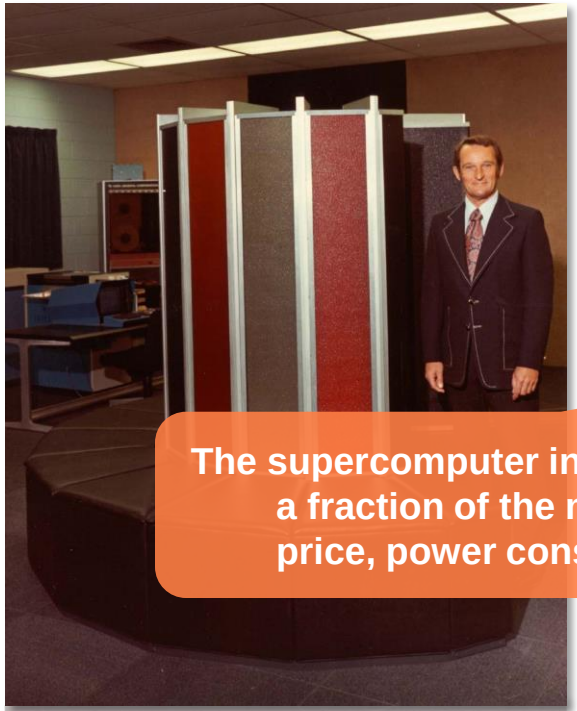
Moore’s Law means doing more with less

1976

Cray 1: The first supercomputer

- 8 megabytes of memory
- Weight of an elephant (5.5 tons)
- 150 kilowatt power supply (enough for 50 households)
- Freon cooling system
- $8.8 million ($30 million in today’s dollars)

Moore’s Law means doing more with less



1976

**The supercomputer in your pocket:
a fraction of the materials,
price, power consumption**

2014

High R&D spending to sustain technology leadership

1980s:

PAS 2000/5000

R&D: 50 mln €

1990s:

PAS 5500

R&D: 400 mln €

2000s:

TWINSCAN

R&D: 1500 mln €

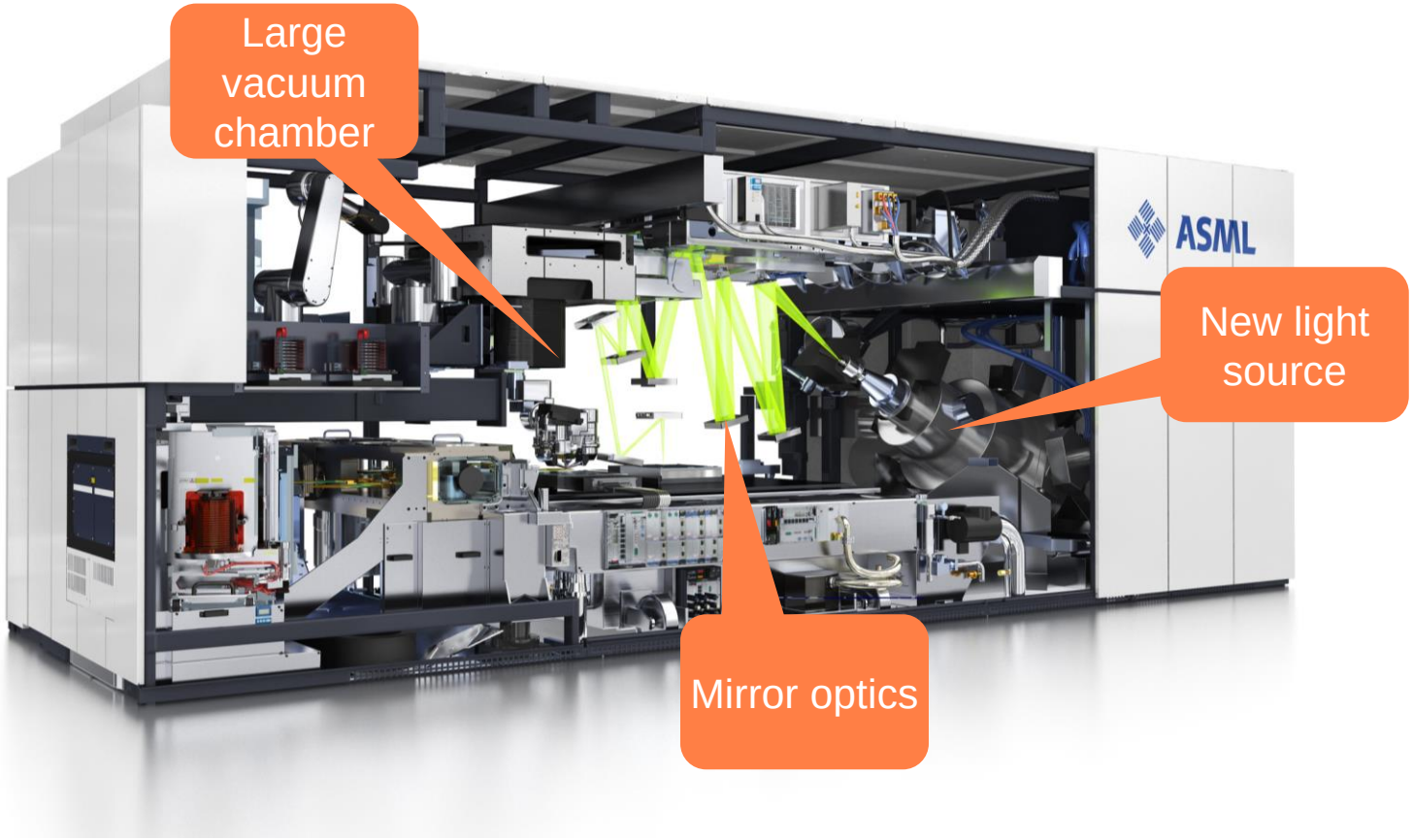
2010s:

NXE EUV

R&D: 2000 mln €

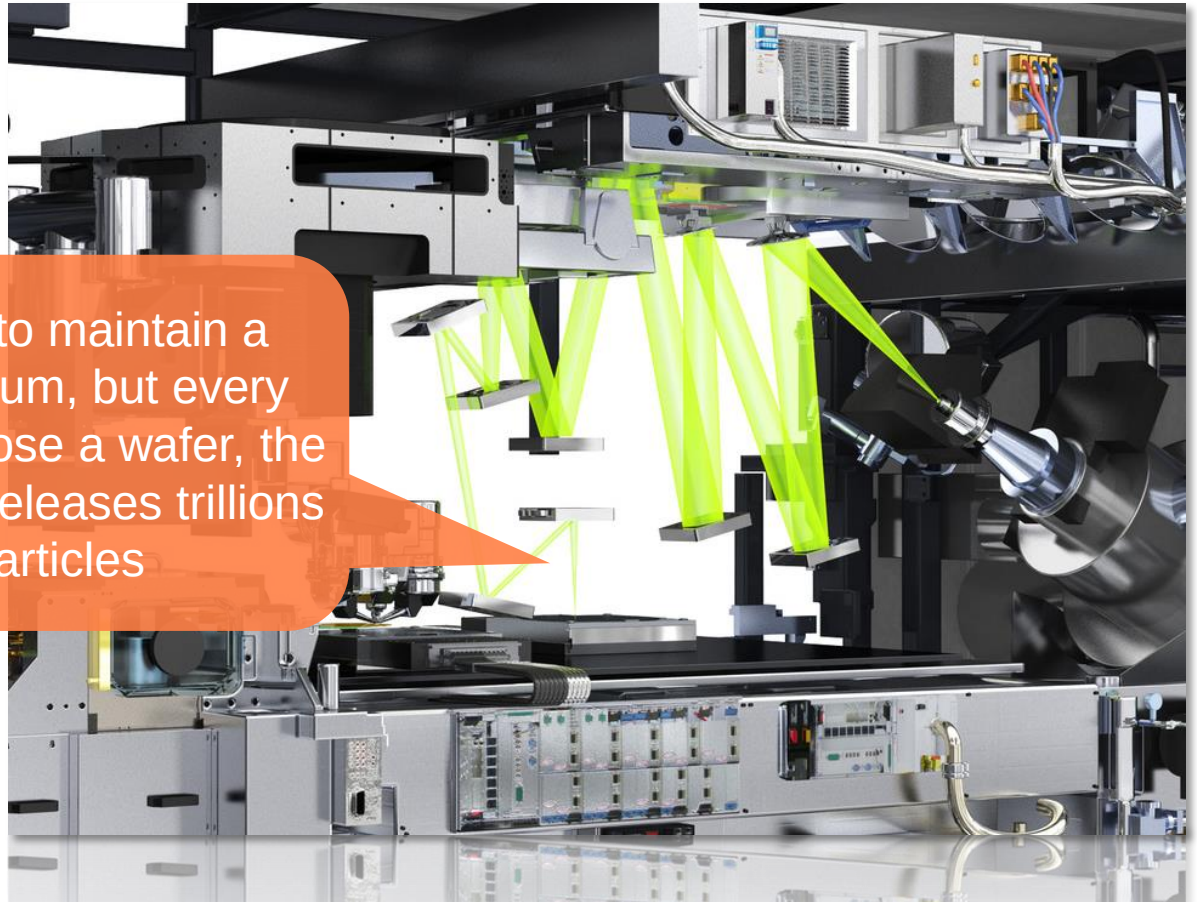
Increasing Complexity, Number Sensors / Actuators
Increasing machine data volumes

The future of lithography: EUV



Maintaining a clean vacuum

We need to maintain a clean vacuum, but every time we expose a wafer, the photoresist releases trillions of particles



Mirrors: Polished to sub-nanometer accuracy

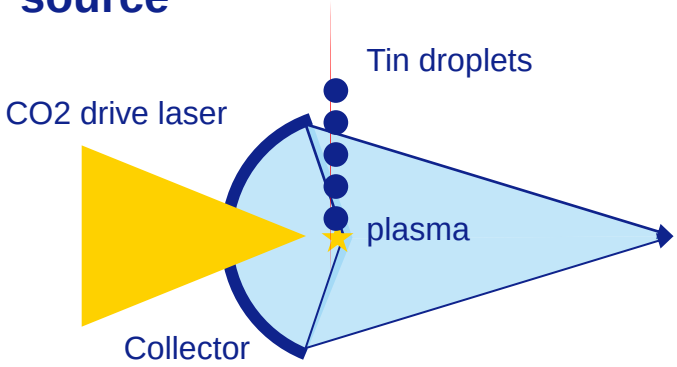
EUV mirrors are polished to an accuracy of ~50 picometers – less than the diameter of a silicon atom.

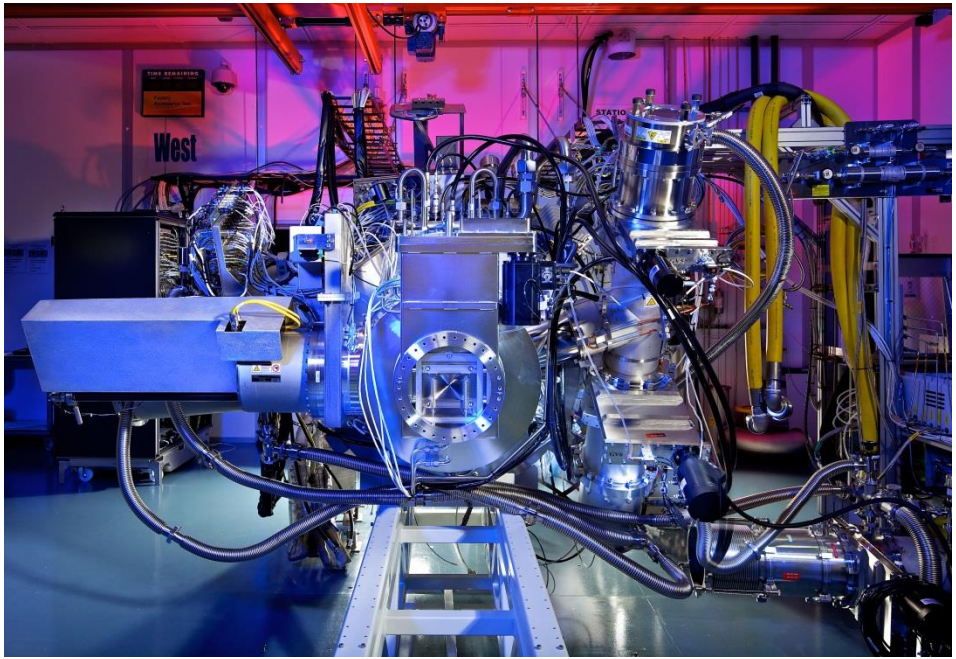
Blown up to the size of the Netherlands, the biggest difference in height would be less than a millimeter.

ZEISS CARL ZEISS SMT

Firing a laser on a tin droplet 40,000 times a second

Laser-Produced Plasma (LPP) source





(BIG) Machine Data in ASML

What is Big Data ?

Not something ASML is involved with….

A ‘general accepted’ definition:

- Big Data is typically a large set of unstructured data from various sources, combined and analyzed using mathematics in order to reveal statistical relevant relations.

  → Dear Customer, if you buy this book there is a big change you will like this one as well.

Following this definition ASML has no developments ongoing on Big Data.

- ASML systems generate a lot of data, a subset is centrally stored for analysis and reporting.
- These data volumes are growing rapidly, posing the next challenge on data management and offering new possibilities
- That’s Big Data for ASML (for now…)

Who is using data from ASM's systems for what ?

Customers

- Production and Process Control
- Process Optimization - Yield Control
  - Data of large amount of wafers combined over all equipment in a fab
  - This is real Big Data: customers use it…
  - ASML systems deliver data, data contains customer IP, ASML has no access

ASML

- Equipment and Process Maintenance
  - Regular Maintenance and Maintenance Escalations
- Product Development
- Business Reporting

Focus of this ppt

Data Challenge in Maintenance and Product Development

Data volumes are rapidly growing:

- Daily collected data for one EUV source is more then the amount of daily data collected for all NXT systems.
- Caused by early shipment and few test systems available at ASML

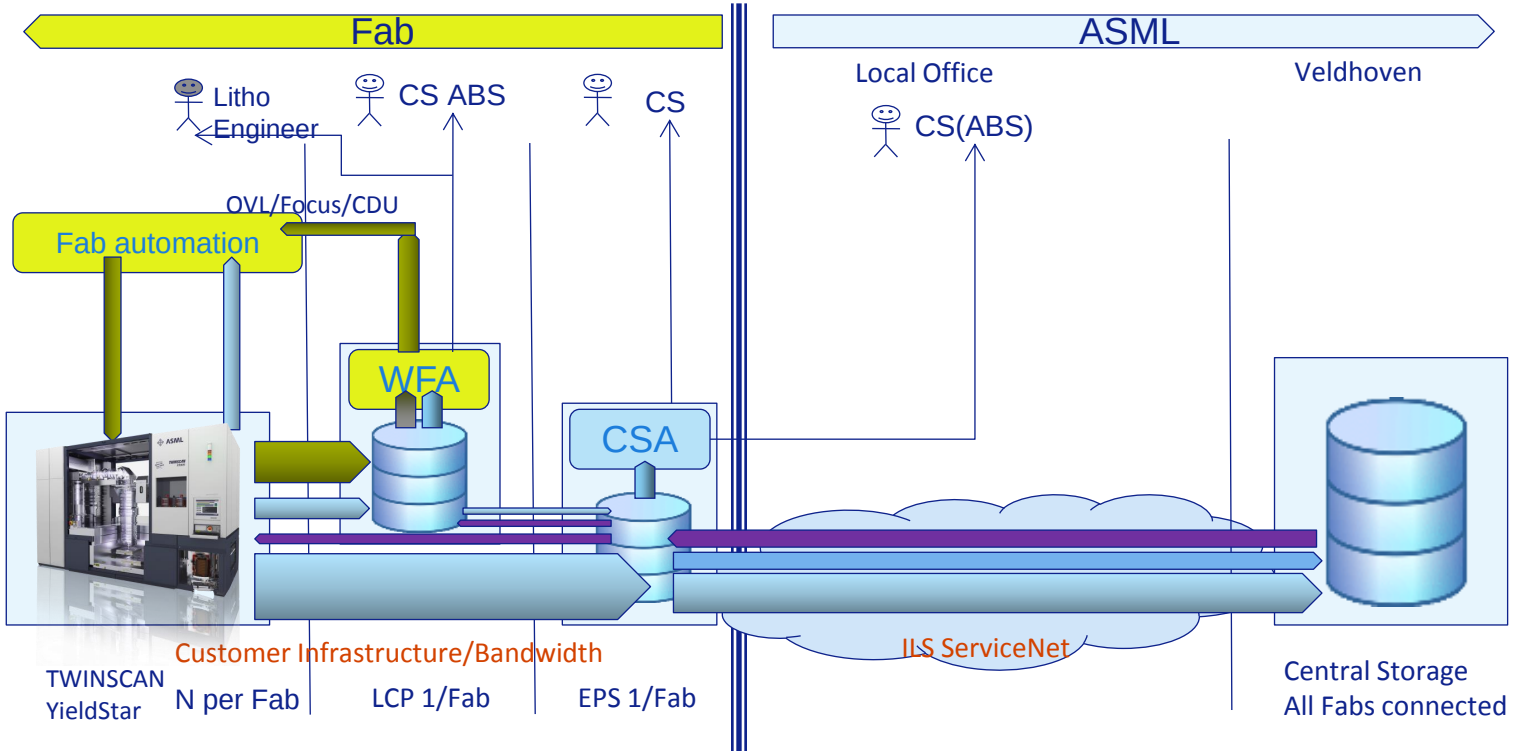
Challenges involved:

- Managing, accessing and using this data volume
- Exploit the potential value of this data to improve equipment maintenance

Management and Infrastructure for Big Data volumes

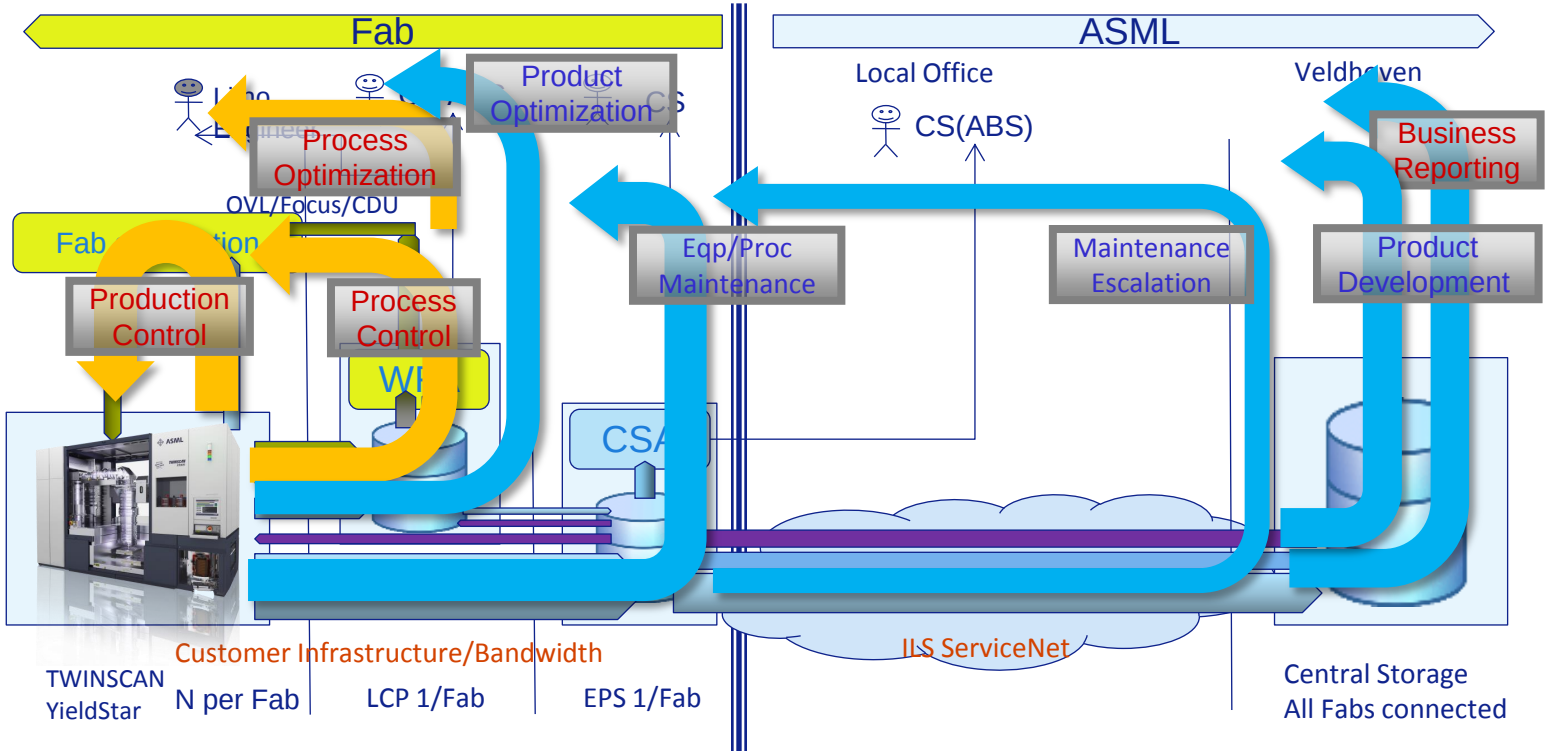
Data Management Reference Architecture

Most data will never leave the fab due to Customer IP restrictions



Data Management Reference Architecture

Main business processes in- and outside the customer fab



ASML Data Classification

As base for data management

Regular Data for Equipment Maintenance

- Limited, controlled and structured set of data, for use in FabTeam tools
- Regular data, available for standard maintenance process

Ad-Hoc Data for Maintenance Escalations

- Unstructured and uncontrolled, variable subset of ‘all’ available data
- Ad-hoc data, file-based, required for machine under maintenance, for use by D&E

On-Demand Data for Product Development

- Unstructured and uncontrolled, variable subset ‘all’ available data
  - Machine can generate too much data to handle, a selection needs to be made
- On-demand data, file based, required for set of machines for longer period, for use by D&E.

Data Infrastructure

All available machine data (sensor, actuator, metrology) need to made easily available for DE on-demand usage

- Internal machine software facilities and external facilities to support this
- The set of on-demand data is volatile, storage technology needs to support this.

Distributed infrastructure to support in- and out-fab data

- Storage of data at ‘best’ location. Tools located at data. Tool access via web / network technology
  - Best location depends on data value, transport cost and IP-restrictions
- Paradigm shift: tools at data iso. data copied towards tools

Data Management

Management process to control the size of ad-hoc and on-demand data

- On-demand data either results in lessons learned by DE, resulting in aggregate regular data (eg. high level KPIs' calculated on the machine) or is not generated/collected anymore.
  - On-demand data only available on specific request for limited period which will only be extended of the data is actively used in a DE study and a plan is available for follow-up
- Ad-hoc data should be collected only once
  - All ad-hoc data requests are handled centrally and only executed if the data is not available yet as regular or on-demand data.

Exploit the big data value

Proactive Maintenance

Pro-Active Maintenance

Using On-Demand data (big data in the central storage) to construct models to predict equipment failure.

Complex models predicting large complicated assembly failure

- Using various data source and multiple data sets
- No predictive value for specific machine with sufficient confidence
  - → Eg. Dear Customer, We think that one of your machine may fail soon…

- ASML's decentralized machine data architecture and management (data definition / collection / correlation / enrichment / storage and data source knowledge sharing) restricts the possibilities, effectiveness and efficiency of complex pro-active model development.

Simple models based on single parameter monitoring

- Showed to have predictive value with high confidence levels

Lessons learned on complex models

No predictive value, difficult and expensive

Complex models are “complex”

- Requires a integral combination of
  - Statistical modeling expertise
  - ASML technical domain & expert knowledge
  - ASML data source insight and data quality & availability
- Integrating all this in one team was not successful
  - “Mental adjustments” required as the *data-mind* of (external) model experts and *physics-mind* of ASML experts showed little overlap
  - ASML data source insight, data quality and availability has proven to be the real challenge

ASML data source insight, data quality and availability

Definition of the events should be stable along the time and unambiguous

- Data sources are ambiguous and human interpretation and interaction impact the interpretation

Required parameters should be stored continuously and with the right detailed level

- Important parameters are stored only when the control threshold is passed, reducing visibility of condition evolution

Historical data horizon should be sufficient and aligned

- Each database has different historical ranges, the shortest ranges determines the model development and validation horizon

Data set should include all data on actions that influence the event being monitored

Conclusion

Conclusion

New competences in DE needed, data analysts with machine knowledge

- Comparable with embedded software engineers: computer specialists with machine knowledge

Just collecting data does not do the trick

- New big data volume require new data collection, storage and data access approach: keep data where it is used and keep tools at data
- Data management process needed to control amount of data
- Data quality is key

Value of ‘real’ big data infra structure unclear and needs to be evaluated

- Mindset change required: DE community used to working with files, not with data queries etc..

ASML

Big Data

Uit Wikipedia, de vrije encyclopedie

Men spreekt van big data wanneer men werkt met een of meer datasets die te groot zijn om met reguliere databasemanagementsystemen onderhouden te worden. De definitie van big data is niet altijd duidelijk. Volgens Gartner gaat het in elk geval om drie factoren: de hoeveelheid data, de snelheid waarmee de data binnenkomen en opgevraagd worden, en de diversiteit van de data. Daarnaast speelt ook de mogelijkheid tot statistische analyse een belangrijke rol.

Toepassingen[bewerken]

De Large Hadron Collider heeft 150 miljoen sensoren, die samen zo'n 40 miljoen metingen per seconde doen. Het verwerken van deze grote hoeveelheid metingen en het trekken van conclusies daaruit vereisen veel rekenkracht.

Het Amerikaanse bedrijf Walmart verwerkt meer dan een miljoen transacties per uur, die op een slimme manier in databases opgeslagen moeten worden. Uit deze grote berg gegevens kunnen verbanden tussen verschillende producten gedestilleerd worden (bijvoorbeeld hamburgers en broodjes, maar vaak zijn de verbanden minder voor de hand liggend). Ook kunnen op basis van aankoopgedrag klanten ingedeeld worden op basis van diverse kenmerken. Door geavanceerde statistische analyses uit te voeren kan Walmart niet alleen vrij nauwkeurig inschatten wat het geslacht, inkomen, de sociale klasse en de gezinssituatie van een klant is maar kunnen ook levensgebeurtenissen zoals huwelijk, geboorte, echtscheiding en overlijden met vrij grote precisie gedetecteerd worden aan de hand van veranderingen in aankoopgedrag.