

Vestas[®]

Wind. It means the world to us.[™]

Nacelle

Blades

Hub

Tower

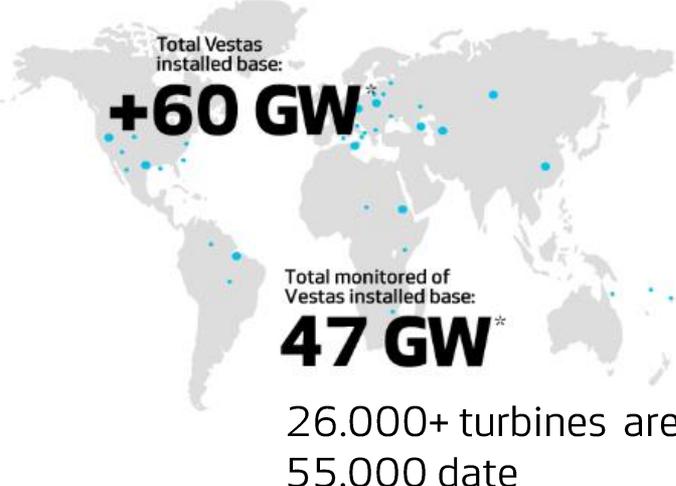
Industrial Big Data Analytics for Wind Turbines

UseR! 2015 Ålborg

Sven Jesper Knudsen - sjknu@vestas.com

Kim-Emil Andersen & Martin Qvist

Vestas Performance & Diagnostic Centre (VPDC)



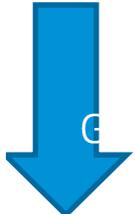
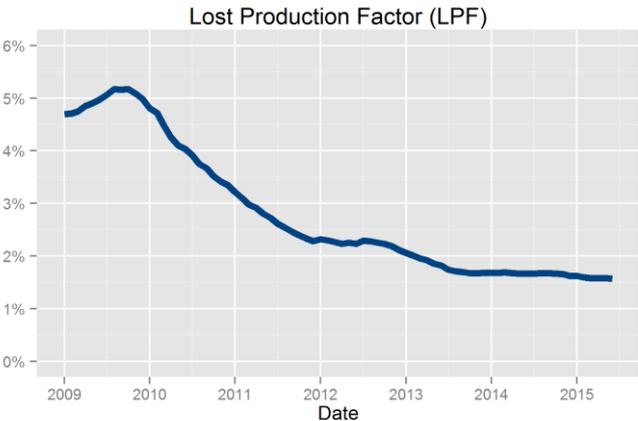
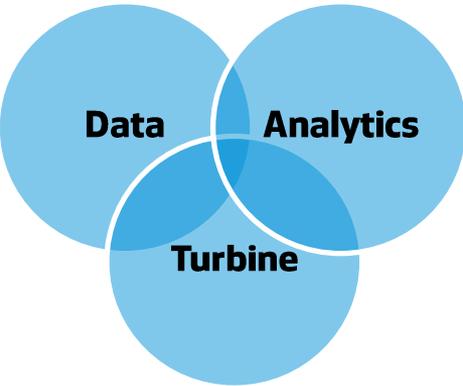
Vestas Installed base: **+60 GW***

Worldwide installed wind capacity: **+318 GW****

*Vestas Performance & Diagnostics Centre, 31/12/13
**Global Wind Statistics 2013 (Year-end), GWEC

Vestas Wind Turbine Data Science

- A Decade of data driven Asset Management
- Predictive Maintenance
- Condition Based Monitoring



Vestas UseR!

Desktop

RStudio is on our Vestas Software portal

Markdown reports to clients

TIBCO Spotfire + R Apps users

SQL, Py, .NET

Vestas' HPC, FireStorm

15000 cores, 2.5+ petabyte, 40Gb infiniband interconnected

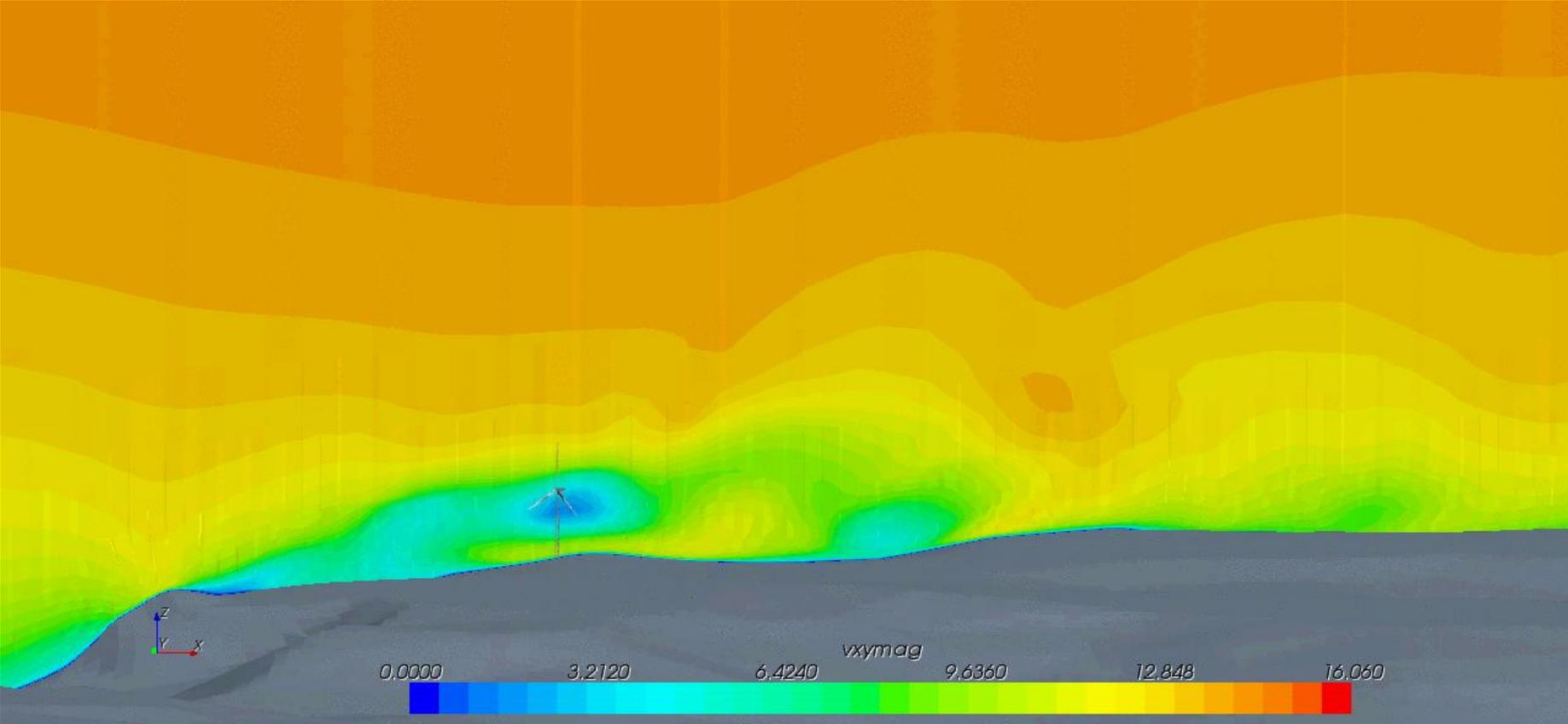
mesoscale climate model, petabyte storage

Computational Fluid Dynamics, high performance

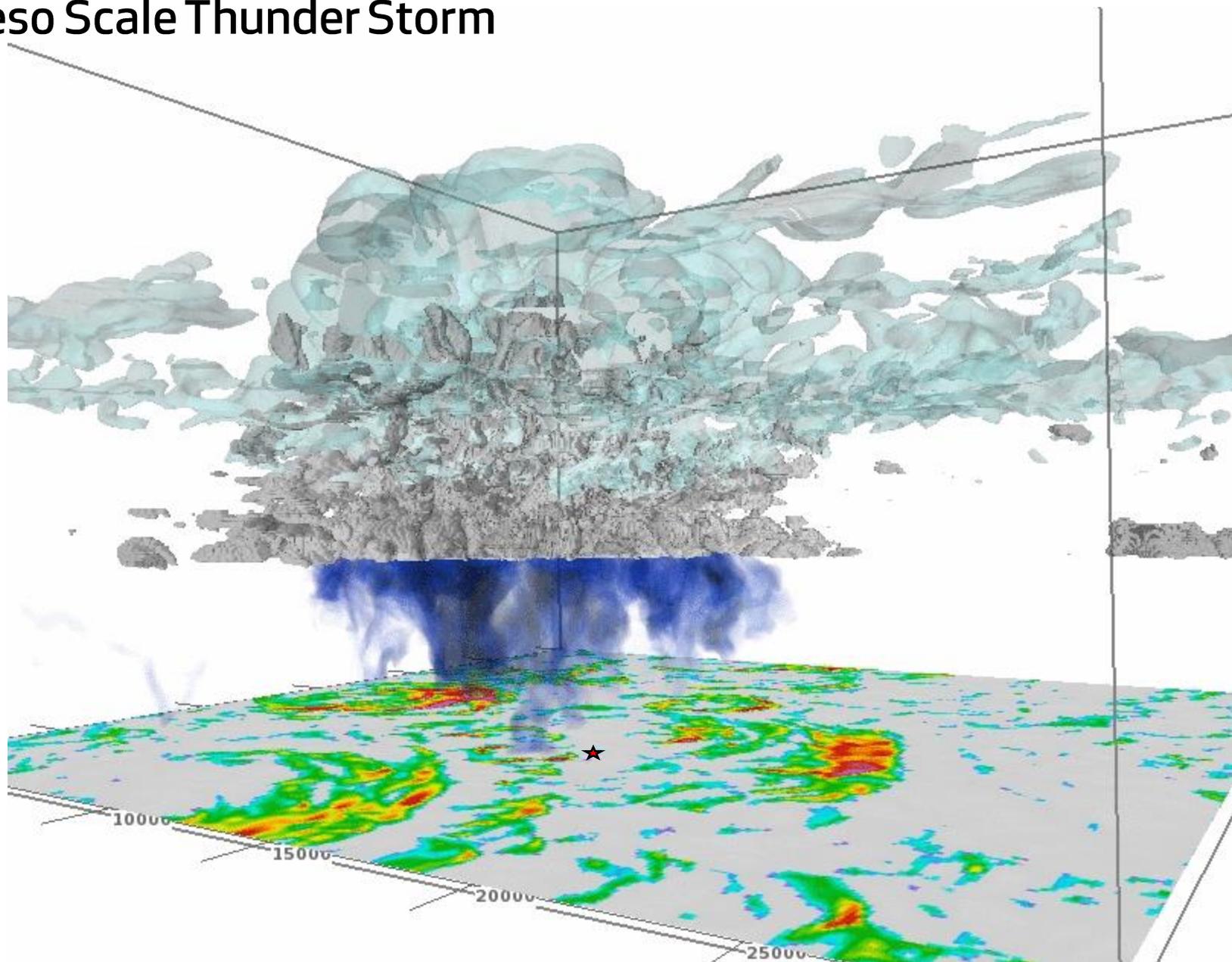
Today focus on HADOOP+HIVE+R

Open Source and Open Standards

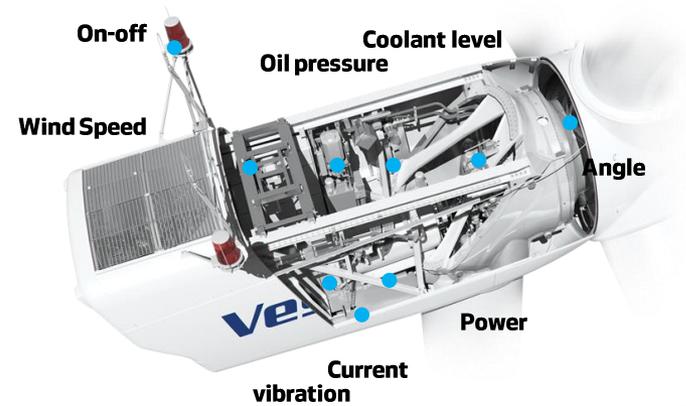
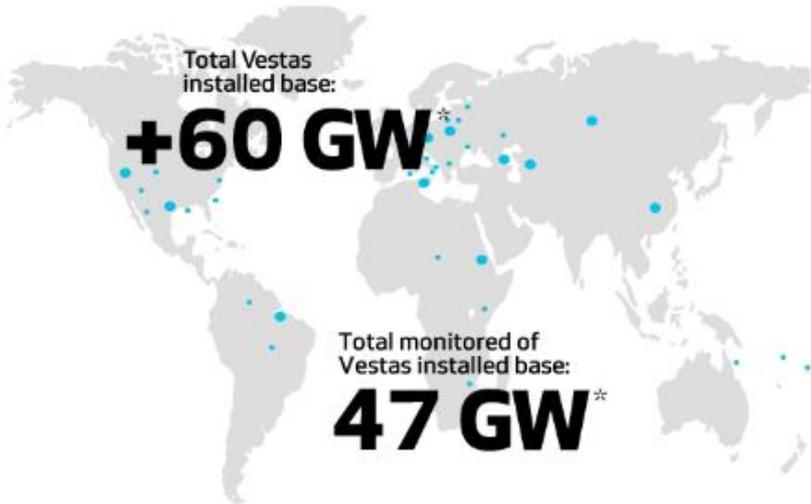
CFD Wind Flow



Meso Scale Thunder Storm



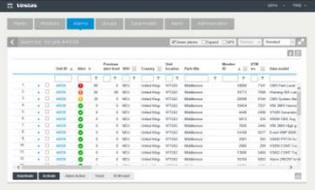
Industrial Big Data at Vestas



- Volume** 50.000 turbines, ca 150terabyte
Vestas Climate Lib, ca petabyte
- Variety** 500+ sensors, vibration 'audio', event logs
image, lab test, service text
- Velocity** Real time, near time, batch
- Varacity** Sensitive and diverse sensors, parameters, humans ...

Data Platform

Vestas Apps



Internal App

Vestas Performance & Diagnostic Centre



Data Science

Vestas Data Centre RDBMS



Data Warehouse Application DB

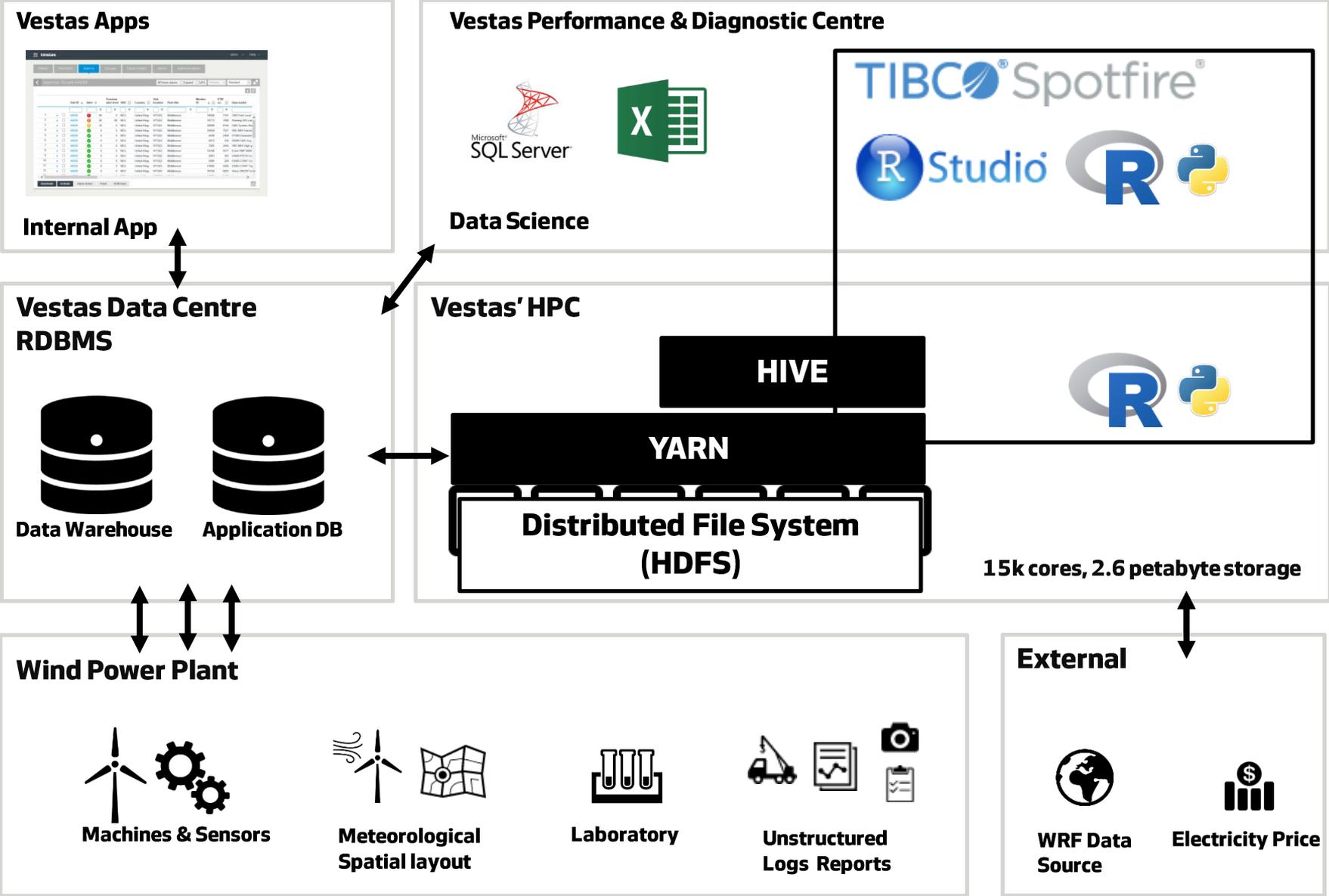
- MS SQL Server: 150tb SAN
- + It works
- + Business penetration
- + Transactional & relational
- Costly
- Staging areas, re computation
- Where is R

Wind Power Plant



Machines & Sensors Meteorological Spatial layout Laboratory Unstructured Logs Reports

Modern Data Platform

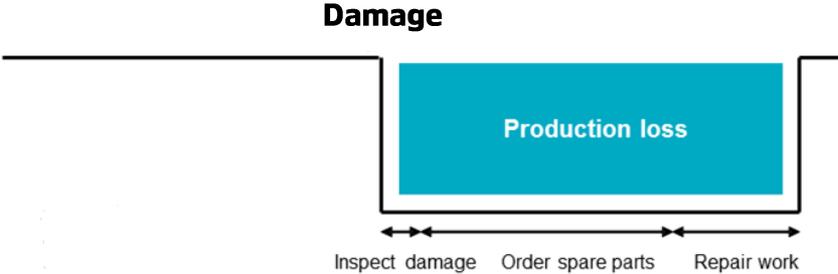


Value Driver

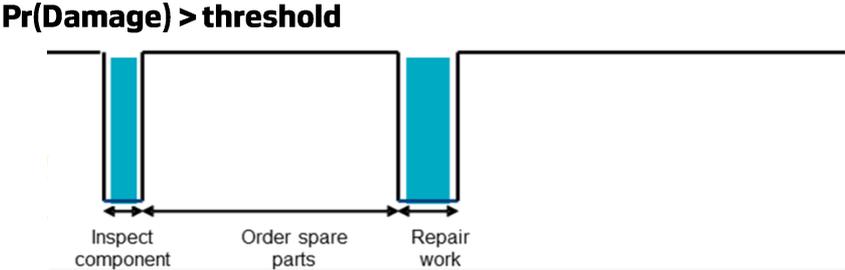
Time is Money – Downtime is Expensive

Downtime Lost Wind Energy Production

Corrective

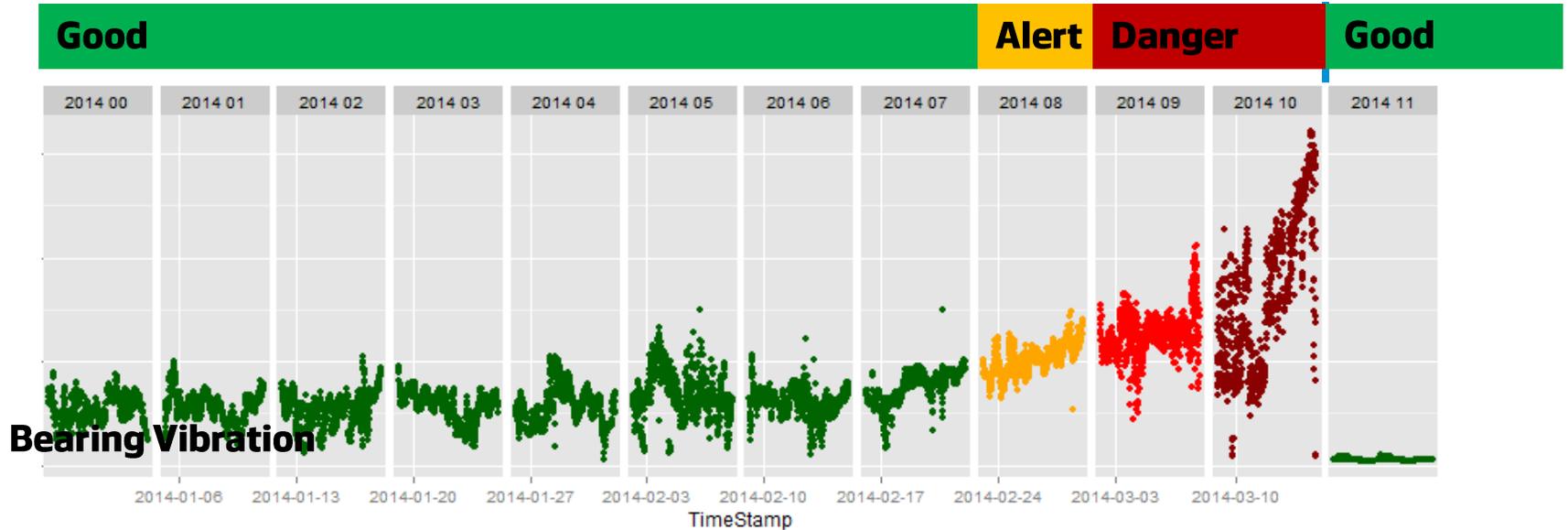


Predictive



Vestas Turbine Monitor

**Repair Work:
Bearing Replaced**



Predict generator bearing failure

- Training data :
Vibration 10min values, several parks
filter down to turbine in full operation
- Changepoint algorithm on raw or residuals of a machine learned model
- Tune using confusion matrix : best precession, accuracy, sensitivity
False Alert vs Wasted Service Visit

package: changepoint, caret

Vestas Turbine Monitor

Full Business Support for Process & Information Flow



Vestas Turbine Monitor

Alarm list for unit 44933

Unit ID	Alert	Previous alert	Unit	Country	Unit	Part title	Member ID	VM	State model
1	Alert	OK	8	NEP	Unknown	Middle	1800	7141	CMF Fuel Level
2	Alert	OK	8	NEP	Unknown	Middle	1873	7036	Sharing 185 Loc
3	Alert	OK	8	NEP	Unknown	Middle	2008	8144	CMF Sensor Act
4	Alert	OK	8	NEP	Unknown	Middle	1945	7037	VIB 3MPT Alarm
5	Alert	OK	8	NEP	Unknown	Middle	1440	2106	V1500 Converter
6	Alert	OK	8	NEP	Unknown	Middle	1813	126	W006 GCR Act
7	Alert	OK	8	NEP	Unknown	Middle	1606	2485	VIB 3MPT High P
8	Alert	OK	8	NEP	Unknown	Middle	1400	5077	Event PWR OFF
9	Alert	OK	8	NEP	Unknown	Middle	2001	380	V4000 PTCR Au
10	Alert	OK	8	NEP	Unknown	Middle	2003	389	V1500 CONTC Cnt
11	Alert	OK	8	NEP	Unknown	Middle	1300	5465	V3042 CONTC Tr
12	Alert	OK	8	NEP	Unknown	Middle	1074	6353	Alarm 200207 for V

Automatic Notification To Service Before Failure

Work Order (ERP)

Display Service order 5047350: Central Heater

Order: YW01 5047350 236 On Invoicing in Hold

Site Status: OK GPS: RACK: R057 PNC: S076

Order address: Operations Components: Check Partner: Additional Data: Location: Planning: Control

Send to party: [Field]

Invoice No. [Field] Location [Field] Telephone [Field] Partner/Emzone [Field]

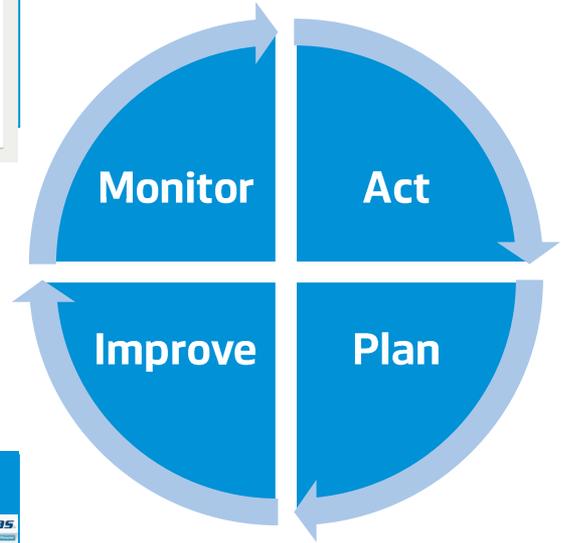
Service: ServiceNo. [Field] Quantity: 0.000 Per order: [Field] Billing form: [Field]

Partner responsible: Planning: 01/1/2041 Schedule: [Field] Order: [Field] User reason: [Field]

Mitigation & work instructions

Vestas Global Advisor

Alerts and work instructions



Plan Service using Vestas Forecasting™ to Minimize Production Losses Maximize Safety



Improve Mitigation Action, Monitor Accuracy

Maintenance Findings

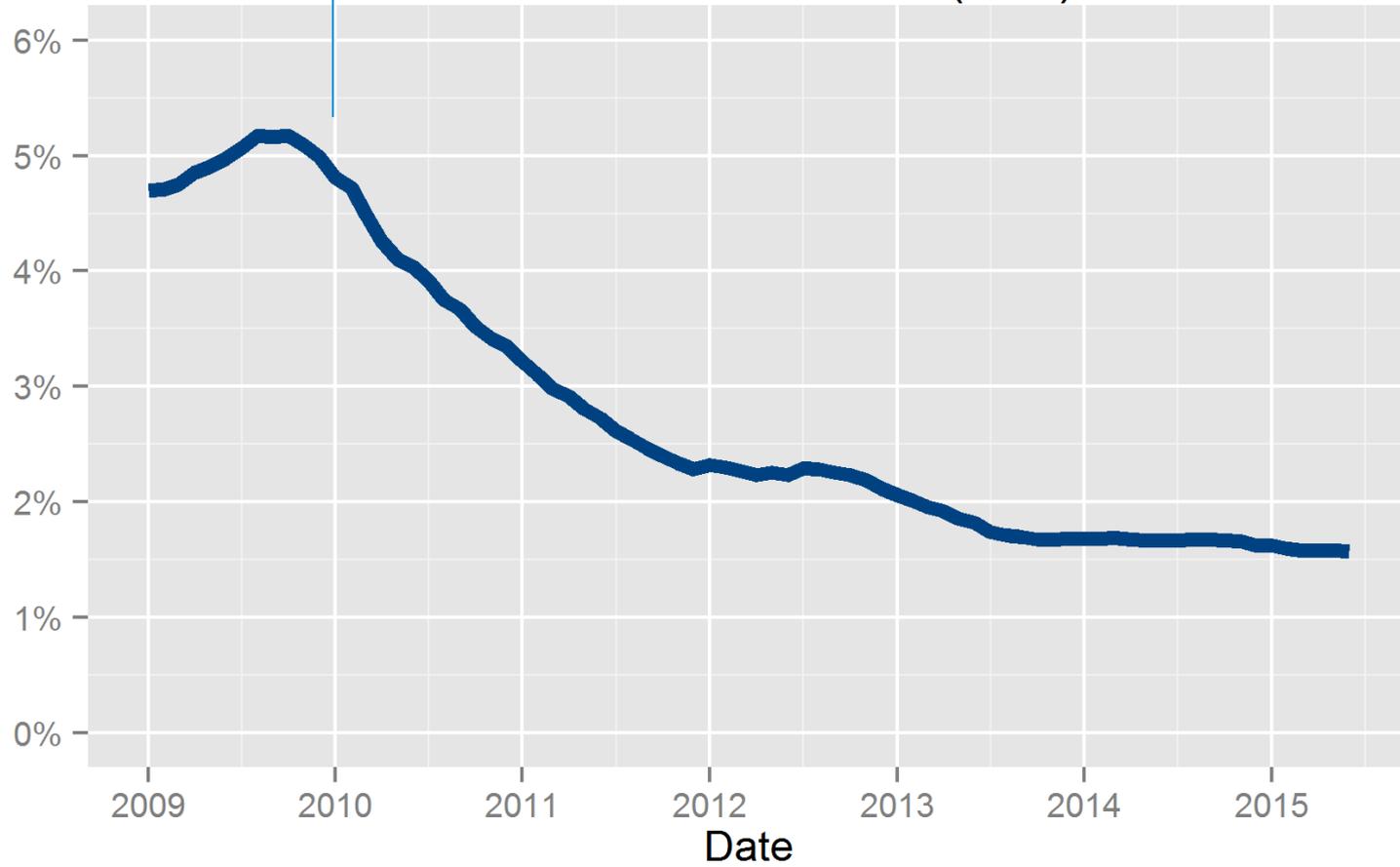
Display Service order 50462697: Component Overview

Order: YW01 50462697 189 Feedback = 0. Brake

Item	Component	Description	LT	Report On	Inv	Loc	Loc/Prm	Lead/Runch	Proc. Category
5013	158475	BATTERY LCR-12V 24H-LC-R12/72P	1	EA	L	1000	2050	3010	Reservation for Order
5020	180448	REPLACEMENT KIT FOR 180449	1	EA	L	1000	2050	3010	Reservation for Order
5030	158053	HYDR ELEC PRESSURE SENSOR 108HK	1	EA	L	2040	3010		Reservation for Order
5048	1893453	RELAY R45-1 FOR AD-4110	1	EA	L	2040	3010		Reservation for Order
5050	158045	ARC DETECTION SYSTEM 230V AC	1	EA	L	2040	3010		Reservation for Order
5058	158212	CSB MOUNTS 16.8 15A	1	EA	L	2040	3010		Reservation for Order
5073	158011	HYDR PRESSURE SWITCH DS-12-70-F	1	EA	L	2040	3010		Reservation for Order

Measurable Business

Lost Production Factor (LPF)



HIVE – Hadoop Interactive

Hadoop allows distributed processing of **large data sets across clusters**

Called Google's data model (Map Reduce)

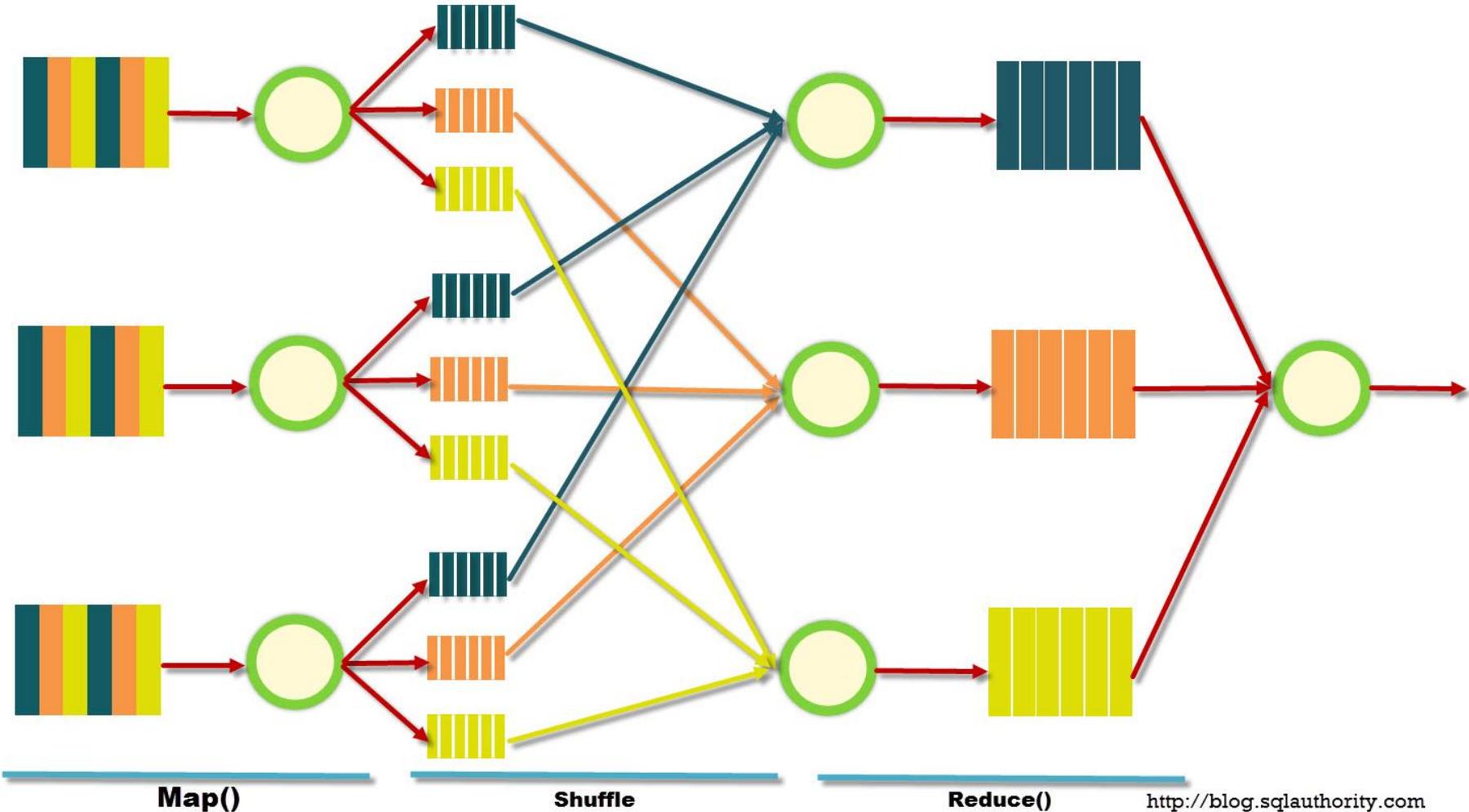
HIVE provides SQL like queries over **Hadoop**

```
Select
    turbineNumber,
    avg(WindSpeed)
from Turbine10minData
group by turbineNumber
```

Faster than indexed MS SQL Server

Much faster if grouped by non indexed column, like temperature bin

HADOOP Map Reduce (MR)

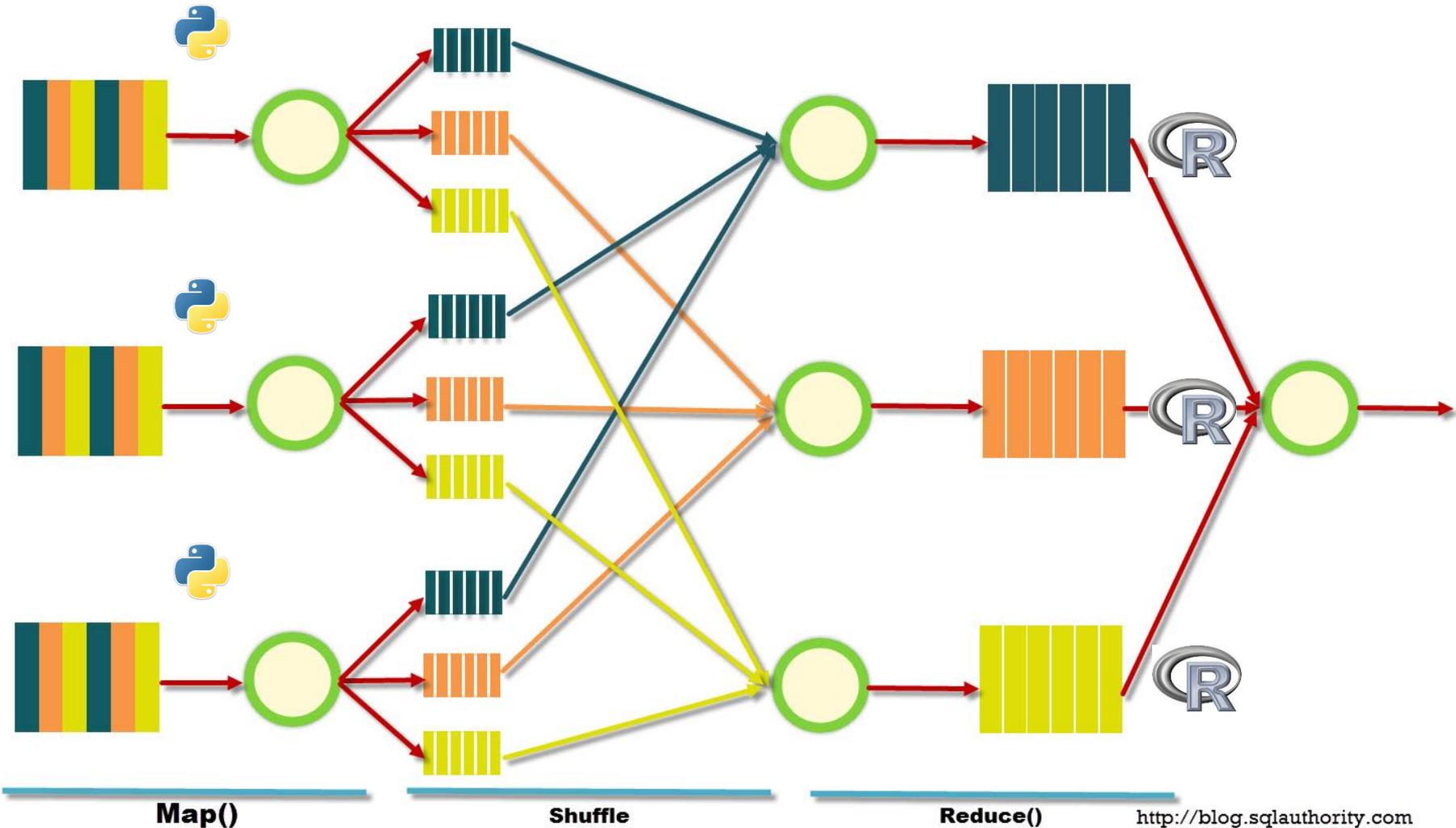


Distributed Data

Are collected and sorted

Distributed Compute

HIVE Custom Map Reduce - inject R , Py, Octave ...



<http://blog.sqlauthority.com>

Distributed Data

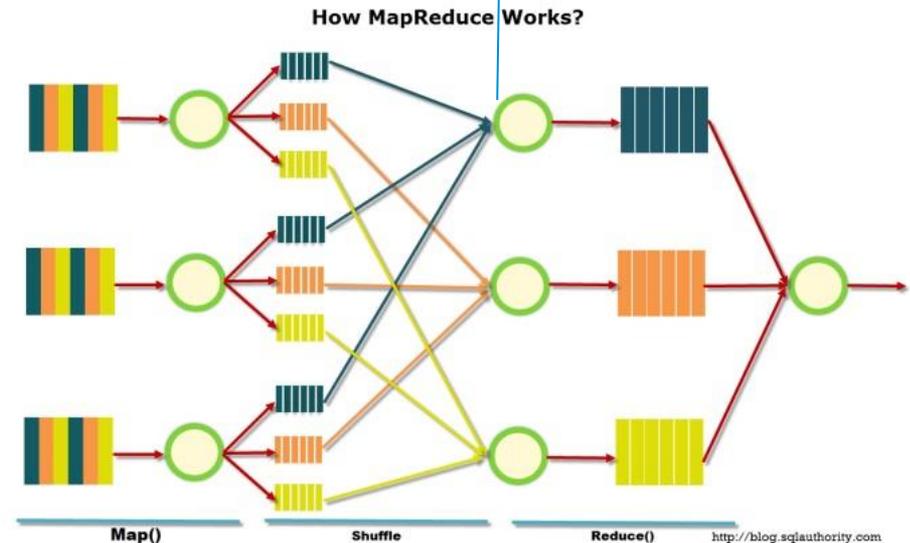
Are collected and sorted

Distributed Compute

Inject scripts using HIVE Transform

```
FROM (  
  SELECT * FROM TurbineData  
  DISTRIBUTE BY turbineId  
) as map  
SELECT  
  TRANSFORM ( * ) USING 'wrapper.sh someAlgorithm.R'  
  as (turbineId,ttimeStamp, propbilityOfFailure);
```

Map step: DISTRIBUTE data to compute nodes BY turbine



<http://hortonworks.com/blog/using-r-and-other-non-java-languages-in-mapreduce-and-hive/>

HIVE: Inject script based computations

```
FROM (  
  SELECT * FROM TurbineData  
  DISTRIBUTE BY turbineId  
) as map  
SELECT  
  TRANSFORM ( * ) USING 'wrapper.sh someAlgorithm.R'  
  as (turbineId,ttimeStamp, propbilityOfFailure);
```

wrapper.sh loads R and #!Rscript
Data is piped to R using stdin
Data piped back using stdout

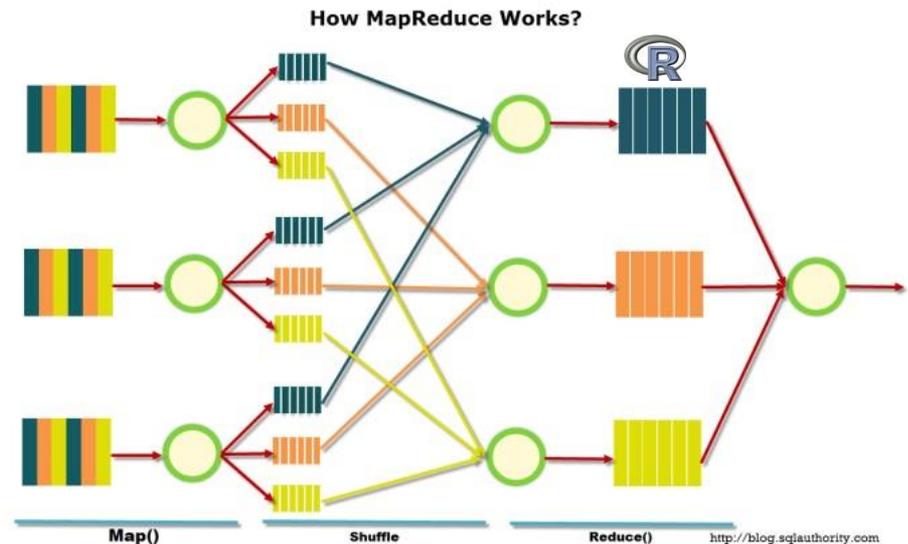
Hive gives total control to R – please mind

- Compute speed
- Memory consumption
- Error handling

Useful HIVE resource settings:

```
set mapred.reduce.tasks=200;  
set mapreduce.reduce.memory.mb=10240;
```

Executes on mapped data



```
1 #!/usr/bin/env Rscript
2
3 .libPaths('/gpfs01/R/packages/bmk/v1.0.0')
4 suppressMessages( library(data.table) )
5 suppressMessages( library(fasttime) )
6 suppressMessages( library(vbmk) )
7
8 outDt <- tryCatch( {
9
10   #fast read of piped stdin
11   inDt <- fread('file:///dev/stdin', sep = "\t", showProgress=FALSE)
12
13   #fast convert to POSIXct
14   inDt <- inDt[,timestamp:=fastPOSIXct(ttimestamp, tz = "UTC")]
15
16   #do your thing
17   bmkProcess(inDt)
18
19 },
20 error = bmkErrorHandler ) |
21
22 if (!is.null(odata)) {
23   write.table(odata, quote=FALSE, sep="\t",
24               col.names=FALSE, na = " ", row.names=FALSE)
25
26 }
```

Summary

HIVE for large batch / interactive jobs

- Proven more reliable with version 0.14

- Spark is lurking for interactive analysis

Familiar to users and cost efficient – no brainer

- Need to understand how to MapReduce the problem

Need to manage your **R**esources

- Fast readers (fread) make a difference

- tweek HIVE resources settings to get job completion

ValueBrings more data to table

References

Vestas and Big Data

<http://www.ibmbigdatahub.com/video/ibm-helps-vestas-turn-climate-big-data-capital>

R:

<http://www.r-project.org/>

<http://www.revolutionanalytics.com/what-r>

Python:

<https://www.python.org/>

Apache HADOOP

<http://hadoop.apache.org/>

Apache HIVE

<https://hive.apache.org/>

Apache Spark

<https://spark.apache.org/>

The Vestas logo is positioned in the top left corner of the image. It consists of the word "Vestas" in a bold, italicized, white sans-serif font, followed by a registered trademark symbol (®). The background of the top half of the image is a vibrant blue sky with wispy white clouds, suggesting a clear, sunny day.

Wind. It means the world to us.™

Thank you for your attention

Copyright Notice

The documents are created by Vestas Wind Systems A/S and contain copyrighted material, trademarks, and other proprietary information. All rights reserved. No part of the documents may be reproduced or copied in any form or by any means - such as graphic, electronic, or mechanical, including photocopying, taping, or information storage and retrieval systems without the prior written permission of Vestas Wind Systems A/S. The use of these documents by you, or anyone else authorized by you, is prohibited unless specifically permitted by Vestas Wind Systems A/S. You may not alter or remove any trademark, copyright or other notice from the documents. The documents are provided "as is" and Vestas Wind Systems A/S shall not have any responsibility or liability whatsoever for the results of use of the documents by you.

Hive setup

- Distribution
 - Hortonworks
 - HDP 2.2
 - Hive 0.14
 - 20 nodes
- Traditional setup runs 1/10GbE
- Disks on each node, HDFS running on top

- Our setup runs on Vestas' HPC
- Has its own parallel filesystem (GPFS)
- Fast (40Gb) Infiniband interconnect
- Very high performance with few nodes
- Very high utilization per node

