“Hadoop is one way of using a set of cheap computers to store an enormous amount of data and then to process that data in parallel."
WHAT IS HADOOP?

AS A DATA PLATFORM LOWER STORAGE COSTS ARE MUCH LOWER...

<table>
<thead>
<tr>
<th>Number of Gigabytes</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$0,00</td>
</tr>
<tr>
<td>10</td>
<td>$2,000,000,00</td>
</tr>
<tr>
<td>100</td>
<td>$4,000,000,00</td>
</tr>
<tr>
<td>1000</td>
<td>$6,000,000,00</td>
</tr>
<tr>
<td>1000</td>
<td>$8,000,000,00</td>
</tr>
<tr>
<td>1000</td>
<td>$10,000,000,00</td>
</tr>
<tr>
<td>1000</td>
<td>$12,000,000,00</td>
</tr>
<tr>
<td>1000</td>
<td>$14,000,000,00</td>
</tr>
<tr>
<td>1000</td>
<td>$16,000,000,00</td>
</tr>
<tr>
<td>1000</td>
<td>$18,000,000,00</td>
</tr>
</tbody>
</table>

- Hadoop
- Teradata Warehouse Appliance
- Oracle Exadata
- IBM Netezza
WHY SHOULD YOU CARE?
THE TREND IS FOR INCREASING USAGE!
IT IS “THE” TECHNOLOGICAL MODERNIZATION TOPIC!

WHY SHOULD YOU CARE?

THE TREND IS FOR INCREASING USAGE! IT IS “THE” TECHNOLOGICAL MODERNIZATION TOPIC!

Hadoop is muscling into Traditional Data Warehouse market
MAJOR ADOPTIONS OF HADOOP

NOT MUTUALLY EXCLUSIVE
... BUT OFTEN NOT SEEN TOGETHER!

Hadoop as a Data Platform
(standalone or as part of a broader ecosystem)

.. to support an IT Transformation

Hadoop as a core component of the next generation of BI and Analytics

.. to support innovative business usage
Unstructured, Semi-structured and Streaming data (i.e. sensor data) handled often outside the Warehouse flow.
WHERE DOES HADOOP FIT?

HADOOP AS A “NEW DATA” STORE

Operational Data Sources

EDW

Data Mart

Analytic Mart

BI and Analytics
WHERE DOES HADOOP FIT?

HADOOP AS AN ADDITIONAL INPUT TO THE EDW

Operational Data Sources

EDW

Data Mart

Analytic Mart

BI and Analytics
WHERE DOES HADOOP FIT?

HADOOP DATA PLATFORM AS A BASIS FOR BI AND ANALYTICS

Operational Data Sources

EDW

BI and Analytics

Data Mart

Analytic Mart

Data Mart

Analytic Mart

Analytic Mart
WHERE DOES HADOOP FIT?

HADOOP DATA PLATFORM AS A “STAGING LAYER” AS PART OF A “DATA LAKE” – *Downstream stores could be Hadoop, data appliances or an RDBMS*
SAS/Access to Hadoop or Impala - Push *some* of SAS’ processing to Hadoop

SAS/Access to Hadoop
SAS/Access to Cloudera Impala
SAS Embedded Process - Push SAS processing to Hadoop with Map Reduce

SAS/Scoring Accelerator for Hadoop (Nov 2014)
SAS/Code Accelerator for Hadoop (Nov 2014)
SAS/Data Quality Accelerator for Hadoop (Nov 2014)
SAS/Data Loader for Hadoop (Nov 2014)

```
proc ds2;
/* thread ~ equiv to a mapper */
  thread map_program;
  method run(); set dbmslib.intab;
  /* program statements */
  end; endthread; run;
/* program wrapper */
data hdf.data_reduced;
  dcl thread map_program map_pgm; method run();
  set from map_pgm threads=N;
  /* reduce steps */ end; enddata;
run; quit;
```
SAS / HIGH PERFORMANCE ANALYTICS

SAS/High Performance Analytics – HP Enabled SAS Procedures

SAS High-Performance Statistics
SAS High-Performance Data Mining
SAS High-Performance Text Mining
SAS High-Performance Econometrics
SAS High-Performance Forecasting
SAS High-Performance Optimization
In-Memory Analytics – Process in Memory, use Hadoop for Storage persistence and commodity computing

WEB CLIENTS

APPLICATIONS

Data Director*

Visual Analytics

Visual Statistics

In-Memory Statistics

Visual Scenario Designer

SAS® LASR ANALYTIC SERVER

SAS® IN-MEMORY

HADOOP

SAS® IN-MEMORY

SAS® IN-MEMORY

SAS® IN-MEMORY

SAS® IN-MEMORY

Images

Audio and Video

Machine Logs

Text

Web and Social

ERP

SCM

CRM
Our strategy is to enable the entire lifecycle around Hadoop.
BIG DATA ANALYTICS

HADOOP SCENARIOS WE SEE MOVING NOW

Exploritative Analytics:
- Data Discovery Platform
- Next generation Analytical Platform
- A need to analyze ALL data including unstructured data

Digital data:
- A need to gain insight of digital data
- Store and analyze large amounts of digital data to understand consumers in a new digital era.

EDW cost take out / modernization:
- Current EDW too expensive
- Data growing at high rate
- A need to store all data for a longer period
- Agility

R&D and innovation programs:
- Customer business programs and investments around R&D and innovations
- Time to market by using more data sources for analytics
TAKK FOR MEG!
HENRIK.SLETTENE@SAS.COM