Predict, Prescribe, Perform: integrating traditional and new data sources to enable Smart Herd Management

S. van der Beek, H.M. Knijn, D. Zouari

Speaker: Sijne van der Beek
Predict – Prescribe – Perform

Integrating traditional and new data sources to enable Smart Herd Management

Sijne van der Beek, Hiemke Knijn, Dean Zouari – ICAR 2017 – June 15
WE FACE SEVERAL CHALLENGES
DATA FORTRESS
WE LIVE IN THE AGE OF PERSONALIZATION
WE LIVE IN THE AGE OF PERSONALIZATION
Useful deadlines

Some useful deadlines for the upcoming August 2017 runtime:

Please note that new MACE data should always be uploaded before new GMACE data. Therefore:

IDEA will be open for receiving MACE and GMACE data starting from Tuesday, July 4th.

Deadline for uploading of pedigrees for MACE, GMACE and Intergenomics is set for Tuesday, July 18th at 5 pm CET.

Deadline for uploading and submitting MACE data into IDEA is set for Tuesday, July 18th at 5 pm CET.

Deadline for sending Intergenomics data is set for Tuesday, July 18th at 5 pm CET, data should be placed on Interbull ftp server under upload directory.

Deadline for uploading and submitting GMACE data is set for Tuesday, July 25th at 5pm CET.

Deadline for uploading file733 and file734 is set for Tuesday, July 25th at 5pm CET.

Pre-release of MACE results is scheduled for Thursday, July 27th.

Pre-release of GMACE results is scheduled for Wednesday, August 2nd.

Official release of MACE, GMACE & InterGenomics results is scheduled for August 8th, 2017.
WE LIVE IN THE AGE OF REAL TIME, BUT STILL HAVE A COMPUTING 3 TIMES A YEAR RHYTM
Decision support based on

- Old algorithms not updated with recent data
- Derived from small data sets
- Hard coded into management support software
STAGE – GATE – DEVELOPMENT
lacks speed and customer input
WHAT WE WANT

BETTER COWS | BETTER LIFE
FIRST BRING THE PEOPLE TOGETHER THAT HAVE TO BRING THE PIECES TOGETHER
BRINGING THE PIECES TOGETHER

<table>
<thead>
<tr>
<th>GOVERNANCE INTEGRATION</th>
<th>TOOLS</th>
<th>SECURITY</th>
<th>OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Lifecycle &amp; Governance</td>
<td>Zeppelin</td>
<td>Administration</td>
<td>Provisioning, Managing, &amp; Monitoring</td>
</tr>
<tr>
<td>Falcon</td>
<td>Ambari User Views</td>
<td>Authentication</td>
<td>Ambari</td>
</tr>
<tr>
<td>Atlas</td>
<td></td>
<td>Authorization</td>
<td>Cloudbreak</td>
</tr>
<tr>
<td>Data workflow</td>
<td></td>
<td>Auditing Data Protection</td>
<td>Atlas</td>
</tr>
<tr>
<td>Sqoop</td>
<td></td>
<td>HAWQ Partners</td>
<td>HDFS Encryption</td>
</tr>
<tr>
<td>Flume</td>
<td></td>
<td></td>
<td>Ranger</td>
</tr>
<tr>
<td>Kafka</td>
<td></td>
<td></td>
<td>Knox</td>
</tr>
<tr>
<td>NFS</td>
<td></td>
<td></td>
<td>ZooKeeper</td>
</tr>
<tr>
<td>WebHDFS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DATA ACCESS

- Batch (Map Reduce)
- Script
- SQL (Hive)
- NoSQL
- HBase
- Accumulo
- Phoenix
- Stream (Storm)
- Search (Solr)
- In-Mem (Spark)
- Others
- HAWQ Partners

YARN: Data Operating System

HDFS: Hadoop Distributed File System

DATA MANAGEMENT
BRINGING THE PIECES TOGETHER

APPLICATIONS
- SAP
- Tableau
- SAS
- MicroStrategy
- Splunk
- Excel
- HP
- Platform
- Datameer
- Revolution Analytics

DEV & DATA TOOLS
- .NET
- Visual Studio
- Informatica
- Spring
- Java
- Talend
- IBM

OPERATIONAL TOOLS
- System Center 2012
- Teradata
- OpenStack

INFRASTRUCTURE
- Red Hat
- Cisco
- Rackspace
- HP
- NetApp

DATA SYSTEM
- SQL Server
- Vertica
- SAP HANA
- Oracle
- Teradata

Hortonworks HDP 2.2
- Governance & Integration
- Data Access
- Data Management
- YARN
- Security
- Operations

SOURCES
- Existing Systems
- Clickstream
- Web & Social
- Geolocation
- Sensor & Machine
- Server Logs
- Unstructured
Transition application

- Period around calving
- Most diseases occur
- Cost a lot of money through loss of milk
Data to use

- Transition diseases registration
- Milk recording data
- Breeding values
- Milk production from automatic milking system
- Rumination data from sensor
WE HAVE TO ADD CREATIVITY TO THE MIX
PROTOTYPING
PREDICT – PRESCRIBE – PERFORM
In the end it is about enabling Smart Dairy Farming