The Modern Analytics Experience
A joint SAS | Cloudera | Intel roadshow

Building a Data Platform
Kunal Taneja, Systems Engineer

#SASanz
Building A Data Platform
with Hadoop and an Enterprise Data Hub

Kunal Taneja – Lead Solution Engineer ANZ
Data Changes How We Work

Instrumentation
Everything that can be measured will be measured.

Consumerization
Employees and customers expect more personal interactions, but not at the cost of their privacy.

Experimentation
The most innovative companies embrace experimentation and agility.
Traditional Enterprise Data Platform

BI Abstraction & Reporting Layer
- Dashboard & Reports
- Ad-hoc Analysis
- Mobile

Enterprise Data Warehouse
- Performance & Access
  - Transformation & Calculation
  - Performance & Access
- ATOMIC Layer
  - Normalisation & Storage
- Data Acquisition Layer
  - Extraction & Staging
  - Cleansing

Source Systems

© Cloudera, Inc. All rights reserved.
Limited Insights
Power users struggle with data. Many users have no data.

Compliance and Privacy
More data, more users, and more tools create complexity. Need to balance business agility with security and governance.

Limited Data
Not efficient to keep existing data, let alone handle new data sources. Time consuming to transform data for analysis in existing systems.
Time consuming transforms..

BI Abstraction & Reporting Layer
- Dashboard & Reports
- Ad-hoc Analysis
- Mobile

Enterprise Data Warehouse
- Performance & Access
  - Transformation & Calculation
  - Performance & Access
- ATOMIC Layer
  - Normalisation & Storage
- Data Acquisition Layer
  - Extraction & Staging
  - Cleansing

Source Systems

 Basel III Value at Risk (VaR): How far behind?

LDM

STAR

3NF

CDC

4 Hours

8 Hours

8 Hours

4 Hours

Data Modelling
Limited Data

Business: “Do we need to consider BlackBerry users part of our digital strategy?”
Limited Data

1. Find the data
2. Get access to data
3. Learn about the data
4. Move data to ADW and process data
5. Data Modeling
6. Model Deployment

Data Discovery
Limited Insights (Not Only SQL)
A new approach for building a Data Platform

What we do
Copy Data to Applications

What we should do
Bring Applications to Data

Process-centric businesses use:
- Structured data mainly
- Internal data only
- “Important” data only
- Multiple copies of data

Information-centric businesses use all Data:
- Multi-structured, Internal & external data of all types
Cloudera Enterprise Data Hub
*Powered by Apache Hadoop*

A new kind of data platform.
- One place for unlimited data
- Unified, multi-framework data access

Only with Cloudera:
- Leading performance
- Enterprise system and data management
- Fundamentally secure
- Open source, open standards
Hadoop and The Enterprise Data Hub

Open Source, Scalable, Flexible, and Cost-Effective

Unified and Managed

Open Architecture

Secure and Governed

CLOUDERA’S ENTERPRISE DATA HUB

- BATCH PROCESSING (MR, Hive, Pig)
- INTERACTIVE SQL (Impala)
- SEARCH ENGINE (SOLR)
- MACHINE LEARNING (SPARK)
- STREAM PROCESSING (SPARK)
- 3RD PARTY APPS (Many)

WORKLOAD MANAGEMENT (YARN)

STORAGE FOR ANY TYPE OF DATA
UNIFIED, ELASTIC, RESILIENT, SECURE (Sentry, Gazzang, Rhino)

FILESYSTEM (HDFS)

ONLINE NOSQL (HBASE)

DATA COLLECTION (Flume, Kafka, Sqoop, NFS)

DATA MANAGEMENT (Navigator)

SYSTEM MANAGEMENT (Cloudera Manager)

© Cloudera, Inc. All rights reserved.
The Most Complete Ecosystem

Applications

Operational Tools

Data Systems

Infrastructure

More than 1,400 partners ensure compatibility with existing investments, lower skill barriers, and help maximize value from your data.
Basel II CVA across limited asset classes in batch?

Basel III CVA across 90% instruments, end-of-day, intra-day, pre-deal?
What trade delays did I experience yesterday?

What trade delays did I experience in the last 60 seconds, what can I expect in the future?
A policy holder has reported a car accident, provide claim form?

Vet with black box whether the claimant meets all conditions?
What has the customer bought from us?

Are we about to lose this customer, what can we do to prevent this?
A customer has phoned in an internet failure, how to troubleshoot?

A network fault has been detected for an at risk customer, please call?
Why Cloudera?

☑ Enterprise-Grade Hadoop
   Differentiated performance, security, management, and governance.

☑ Expertise
   No one knows Hadoop better than Cloudera.

☑ Enablement
   Support, Training, and Professional Services enable and deliver success.

☑ Ecosystem
   Cloudera ensures that Hadoop works with the platforms, tools, and integrators you rely on.

☑ Sustainable Innovation
   Our hybrid open source model delivers the benefits of open source and what the enterprise requires, while enabling us to invest in the future for our customers.