Hot, Hotter, Hyper
How to Handle Big Data

Marius Kaiser
Solution Consultant
Tableau

Erwin van Laar
Team Lead, Product Consultancy
Tableau
Session Overview

Hot, hotter, Hyper | How to handle big data

Is your organization drowning in the data lake too? Leading big data organizations are leveraging a multi-tiered approach to data for consumption in Tableau. In this session, plan to discuss the challenges of visualizing big data, establish best practices for leveraging big data solutions directly, talk about our new in-memory data engine Hyper for fast visual analysis in Tableau, and learn how to use Tableau Prep to structure data within your big data strategy.

Speaker(s):

  Marius Kaiser, Tableau
  Erwin van Laar, Tableau

Content Type: Breakout Session

Level: Advanced

Track: Data
Hot, Hotter, Hyper
How to Handle Big Data

Marius Kaiser
Solution Consultant
Tableau

Erwin van Laar
Team Lead, Product Consulting
Tableau
Hot, Hotter, Hyper
How to Handle Big Data

Marius Kaiser
Solution Consultant
Tableau

Erwin van Laar
Team Lead, Product Consulting
Tableau
Introduction
What is Big Data anyway?

Volume

Velocity

Variety
What is Big Data?
What is Big Data?
Agenda

• Tableau’s Big Data Focus
• Hyper
• Big Data Framework
• Practical Example
• Summary
Tableau’s Big Data Focus
Big Data Focus

Connectivity
Access to all data

Performance
Fast interaction with all data

Discovery
Finding the right data
Hybrid Data Strategy

- A combination of in-memory and live query engines
- Designed for adaptability and flexibility
- Unified user experience for full range of data stores
Live Query Engine

- Supports the world’s largest databases
- Attaches directly to enterprise data stores
- Pushes down workload
- Leverage your Big Data investments
In-Memory Engine

• Extremely fast computational ability
• Fast performance with large data sets
• Optionally aggregate data before extracting

Data Extract API & Web Data Connector

Hadoop

Fast databases

Data warehouses
Hyper
Fast analysis on data of all sizes

*Hyper replaces TDE in all of our products*
A High Performance Next Generation Database System

One System

One State

No Tradeoffs

No Delays
Scalability

Hyper’s parallelism fully utilizes large amounts of cores (> 120)
Hyper

Key benefits

• Up to $5x$ faster query performance with large or complex datasets

• Fresher data with up to $3x$ faster extract creation

• Support for larger extracts
Caution!

• It’s not the WIN Button!
Data Strategy

• Selection of Technology

• Align Technology to Business Use and Data

• What questions do we want to answer?
Big Data Framework
Hot, Warm, Cold
Let me introduce Air Tableau.
Air Tableau

• New, technology driven airline
• Data of hundreds of millions of flights and their performance

Goals
• Analyse the data and identify gaps
• Monitor Flights performance going forward
• -> Performant Dashboards for everyone!
Cold - Store
Flights Performance Data in the Data Lake

Amazon S3
Amazon S3

STORE

ANALYZE

SPEED UP

CONSUME

Analyst

Data Scientist
The Data Lake

Files

SQL

HDFS

JSON
Hive is too slow for Interactivity
Warm - Analyze
Faster ways to query the Data Lake
Hive is too slow for Interactivity
Hot - Let’s Speed It Up
Leverage Tableau

Secret weapons

Aggregated Extracts

Prep
Pre-aggregate
Optimize

Advantage:
Same results, vastly improved performance
SW#2: Prep

Tableau Prep - Hot - Warm - Cold Flow

Connections:
- cdhprastoldap.test:ts... Parse
- TC London GA - Flights ...
- Tableau extract
- L_UNIQUE_CARRIERS_c... Table file

Search:

Tables:
- L_UNIQUE_CARRIERS

Diagram:
- Flights RAW
- Join AIRLINE ID
- Clean
- Clean City Name...
- Remove Fields
- Output
- HOT
- HOT-Clean
- HOT-Output
- WARM
- WARM-Clean
- WARM-Output
Human Scale of Data
Consume
Amazon S3

Amazon Athena

presto

hadoop

HIVE

Amazon Redshift

AT SCALE

Reporting

Data Science
Leverage Tableau

Secret weapons:

- Aggregated Extracts
- Prep
- Cross-Datasource Actions
- Hybrid Data Strategy
Leverage Tableau

Secret weapons:

- Aggregated Extracts
- Prep

Cross-Datasource Actions

Hybrid Data Strategy
Human Scale of Data
Questions?
Summary
Summary

• Big Data needs to be for the people
• Tableau enables people to use Big Data
• Tableau is a Leader in Big Data Analytics
• Tableau’s Hybrid Data Engine is a secret weapon!
• Tableau enables Big Data agility
• Use the Hot, Warm, Cold framework
Over to you!

- Use the Hot, Warm, Cold framework

- Remember Tableau’s Secret Weapons
  - Aggregated Extracts
  - Prep
  - Cross-Data Source Actions
  - Hybrid Data Strategy
Conclusion

Tableau is great for Big Data

Use the right technology for the right use case
Advancing your big data strategy

Expedia | Big data self-service analytics with Tableau and AWS
Please complete the session survey from the Session Details screen in your TC18 app
Thank you!

Marius Kaiser
mkaiser@tableau.com

Erwin van Laar
evanlaar@tableau.com