

Leverage Big Data

Migrate
Optim Archived Data
to Hadoop

Hari Kasina
Optim Consultant



Hari Kasina

Optim Consultant

Also, supporter of Green Planet
&
Fight against Global Warming

Worked at IBM for 7+ years on Optim Technology as Technical Manager

Optim & Big Data Specialist

20+years in Information Technology Expertise

Speaker at: Oracle OAUG Conference, IBM IOD & Webinars

Working on US Design Patent for Archiving Data using Table Partitions

Why Archive Data

- **Cost savings:** By purging static data, the system can be freed which improves efficiency as well as operating costs can be reduced significantly.
- **Performance:** Data volume can affect the performance of the system. With real-time information, the system can be kept agile as well as seamless access can still be provided to historic data.
- **Compliance:** Data management is important because more data means more compliance risks. By purging data with archiving, new valuable data can come into the system, which will be more useful in the current scenario.

History of Optim Archived Data

- Organizations have been archiving data using Optim for 18+ years
- Archived data is an asset for preserving valuable company information
- There are terabytes TB of data within Organizations, archived and mostly under utilized

Industry Facts about Archived Data

Organizations are quickly learning the value of analyzing vast amounts of previously untapped archival data. Industry studies suggest that only about 20% of all digital data is ever accessed or used again after it is stored, underscoring the archival challenge. The need to effectively store, search for and retrieve enormous volumes of archival content is fueling new advancements in archive solutions

-- Source "The [Active Archive Alliance](#) State of the Industry 2018 report"

Time to utilize Archived Data in Hadoop

For Analytics, Machine Learning, Artificial Intelligence

- **Better real-time data-driven decisions:**

Incorporate emerging data formats (streaming audio, video, social media sentiment and clickstream data) along with semi- and unstructured data not traditionally used in a data warehouse. More comprehensive data provides more accurate analytic decisions in support of new technologies such as artificial intelligence (AI) and the Internet of Things (IoT).

- **Improved data access and analysis**

Hadoop helps drive real-time, self-service access for your data scientist, line of business (LOB) owners and developers. Hadoop is helping to fuel the future of data science, an interdisciplinary field that combines machine learning, statistics, advanced analysis and programming.

- **Data offload and consolidation**

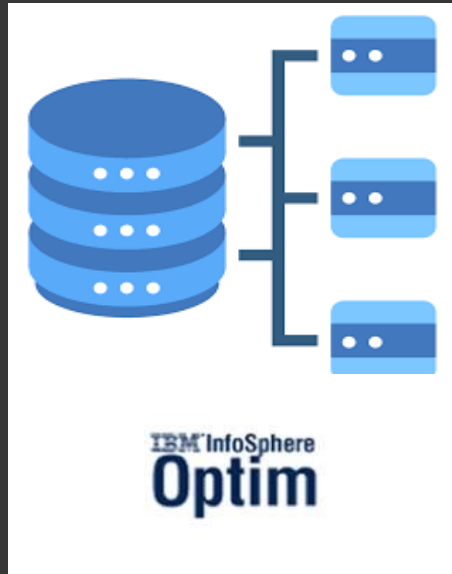
Optimize and streamline costs in your enterprise data warehouse by moving “cold” data not currently in use to a Hadoop-based distribution. Or consolidate data across the organization to increase accessibility, decrease cost and drive more accurate data-driven decisions.

-- Source “www.ibm.com - Apache Hadoop- think Digital Event Experience”

Benefits of Migrating Data to Hadoop

Archived Data in Hadoop	Data in Conventional Archive
Data is Active and Queries can return similar to original system form	Data is Passive and Queries can take a long time to return data
Data can be accessed by Hadoop Tools and third party BO Tools	Limited tools to access the data
Data available for Analytics like Spark ML, AI	Not ready for Analytics

Optim to Hadoop



Optim
Connect



Hadoop Eco System

Optim Archived Oracle EBS Data in Hadoop HIVE

HIVE

QUERY JOBS **TABLES** SAVED QUERIES UDFs SETTINGS

DATABASE
Select or search database/schema

TABLES | 14 **TABLE > GL_JE_LINES**

Search COLUMNS DDL STORAGE INFORMATION DETAILED INFORMATION

INCLUDE COLUMNS

TABLE STATISTICS

STATS NAME	VALUE
Number of Files	4
Number of Rows	20+ Million <input type="text" value="20157940"/>
Raw Data Size	9899671786
Total Size	9919829726

gl_balances
gl_budget_versions
gl_budgets
gl_code_combinations
gl_daily_balances
gl_je_batches
gl_je_categories_tl
gl_je_headers

Report Performance Test Case

General Ledger Transactions Report

248,968
records



**8 Minutes
13 Seconds**



1,589,643
records

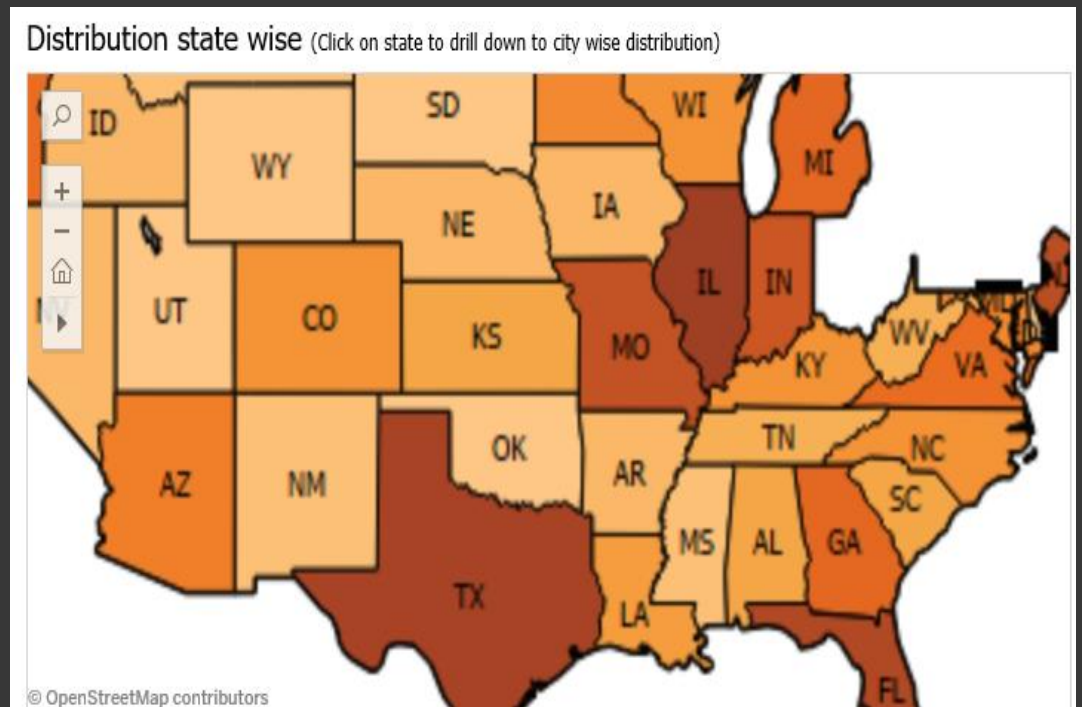


59 Seconds

Tableau Dashboard using HIVE

Hadoop Analysis: Example of how Tableau can help you analyze big data for Analyzing large sets of data.

The data has been ingested from Oracle Inventory data Optim Archive File(s)









Archived Data for AI & ML



CONNECT



 **Apache Hadoop Ecosystem**

 Zookeeper Coordination	 Oozie Workflow	 Pig Scripting	 Mahout Machine Learning	R Connectors Statistics	 Hive SQL Query	APACHE HBASE
-------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------	-----------------------------------	-----------------------------------------------------------------------------------------------------------------	-------------------------





DEMO:

**Optim Sample Tables Archive File(s)
Migrated to Hadoop Hive Tables**

Run Queries on Hadoop Hive

Run Reports on Hadoop Hive



Q & A

Hari Kasina

Optim Consultant

hari_kasina@yahoo.com