

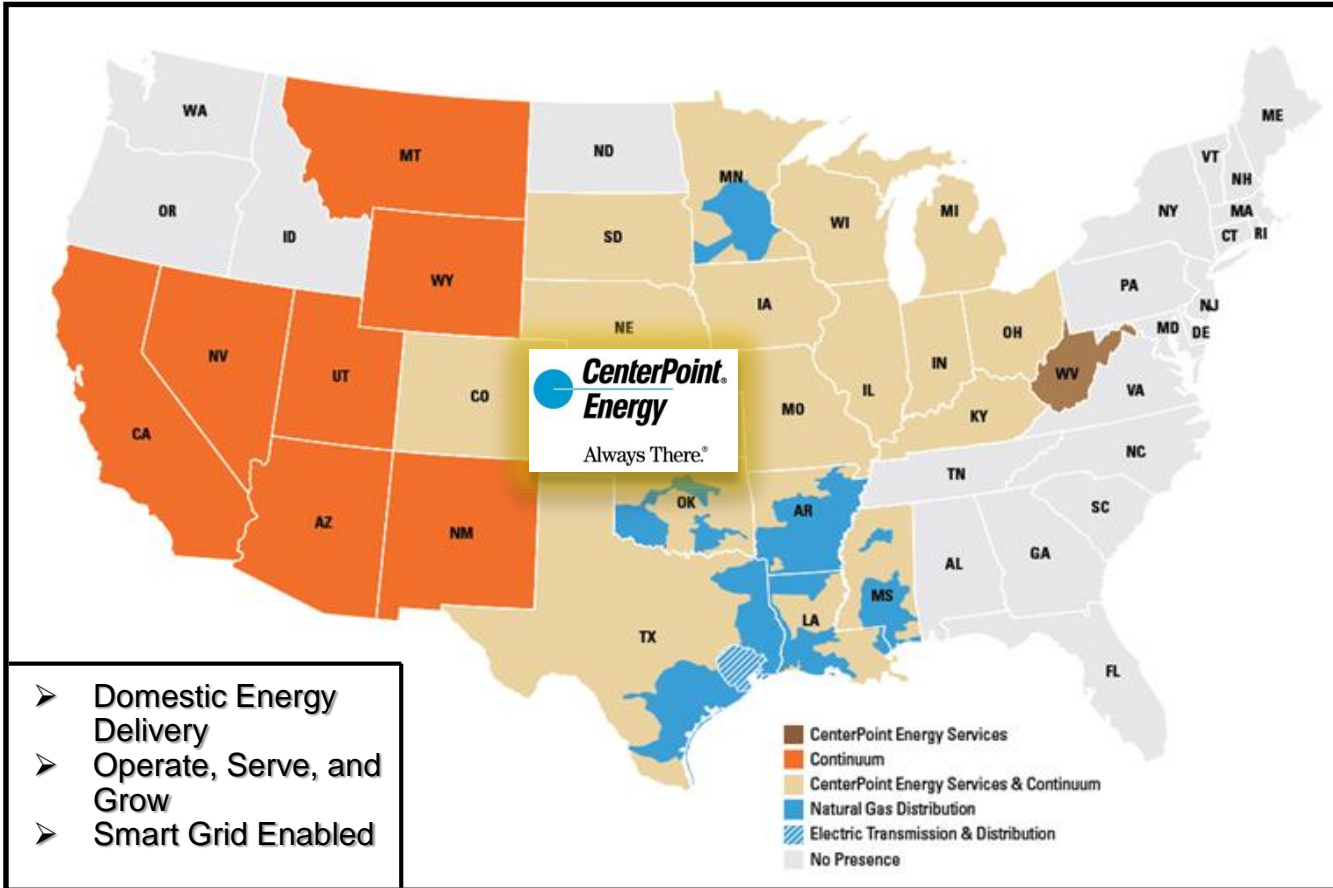
# Analyze Big Data Faster and Store It Cheaper

Dr. Steve Pratt, CenterPoint  
Russell Hull, SAP

Public



# About CenterPoint Energy, Inc.



- Publicly traded on New York Stock Exchange
- Headquartered in Houston, Texas
- Over 5000 square *miles* of electric transmission and distribution service area
- Assets total more than \$22 billion
- Over 8,700 plus employees
- CNP & its predecessor companies in business for over 130 years

- Twenty-Eight State Geography
- Over Five Million Metered Customers
- 2.3 million Smart Meters
- 4000 Miles of Transmission
- 47,000 Miles of Distribution

- Electric Transmission & Distribution
- Natural Gas Distribution
- Competitive Natural Gas Sales and Services



# 5 year Goal / Motivation and Scope



## E2E Insight Management Platform

- 12TB Operation DW and 120TB Smart Meter DW will be end of life by 2016.
- Current Mainframe solution is complex, and expensive to maintain.

## Solution & Benefits

- Consolidate Big Data DW onto SAP HANA platform:
  - Leverage SAP HANA compression
  - Data Tiering technology (Dynamic Tiering, Hadoop) to manage data size and growth.
- Reduce annual technology maintenance expense
- Additional saving (HW backup storage, synergy in support skillset).
- Avoid further capital spending on aging technology HW and SW.
- Leverage SAP HANA real-time reporting using in-memory capabilities.



# Business Challenge

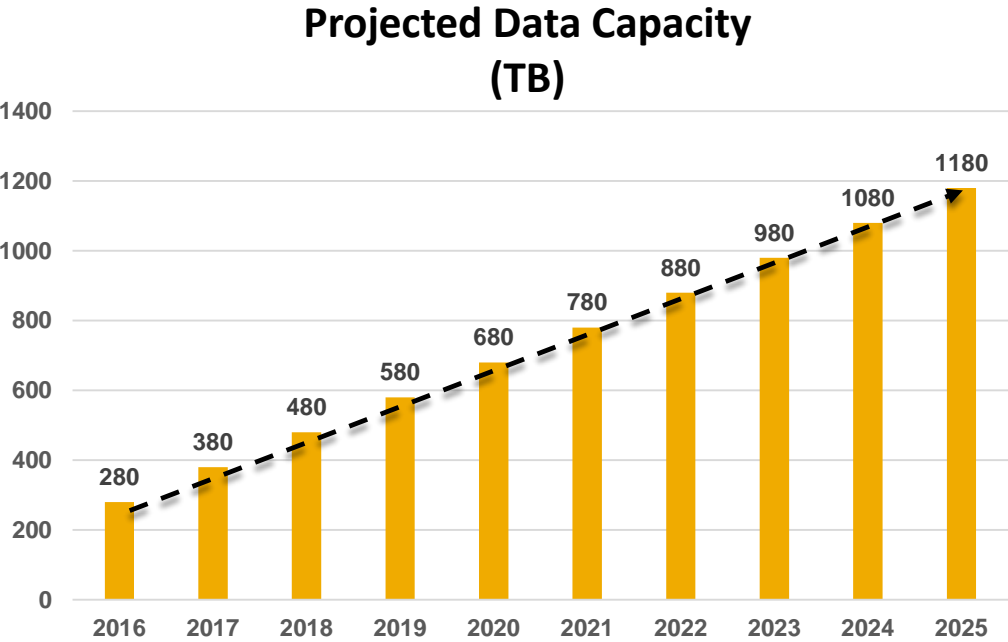


## 1+ PB of SmartMeter Data

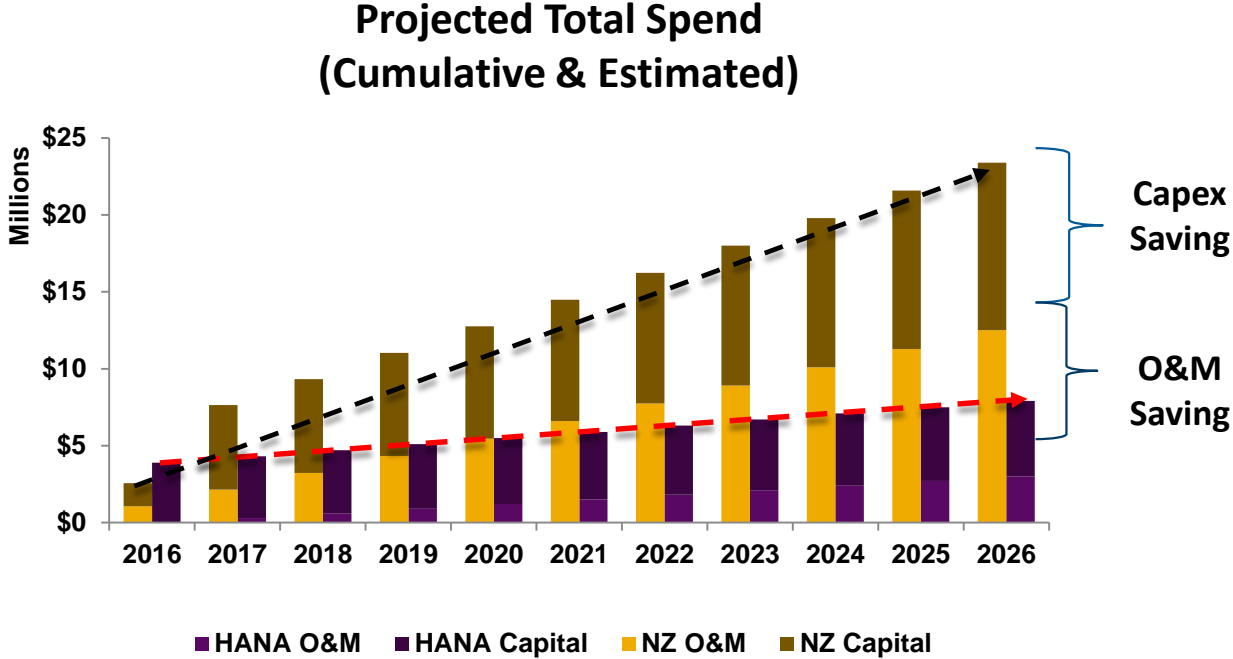
- **2.3MM** SmartMeters taking **readings every 15 minutes** creating **225MM Readings** per day, or over **800 Billion Readings** in a Year.
- Regulatory requirements require historical readings to be available for **10 years**.
- Uncompressed Data Growth of **8TB** per month and over **1PB** in a 10 year period.
- Current DW technology is approaching **End of Life**
- Massive amounts of data stored in prior Mainframe Solution was **hard to manage** and has a significantly high total cost of ownership.
- Need a **cost effective solution** for today's analytics, regulatory requirements and preparation for future use cases.



# Business Case – Capex & Opex Savings



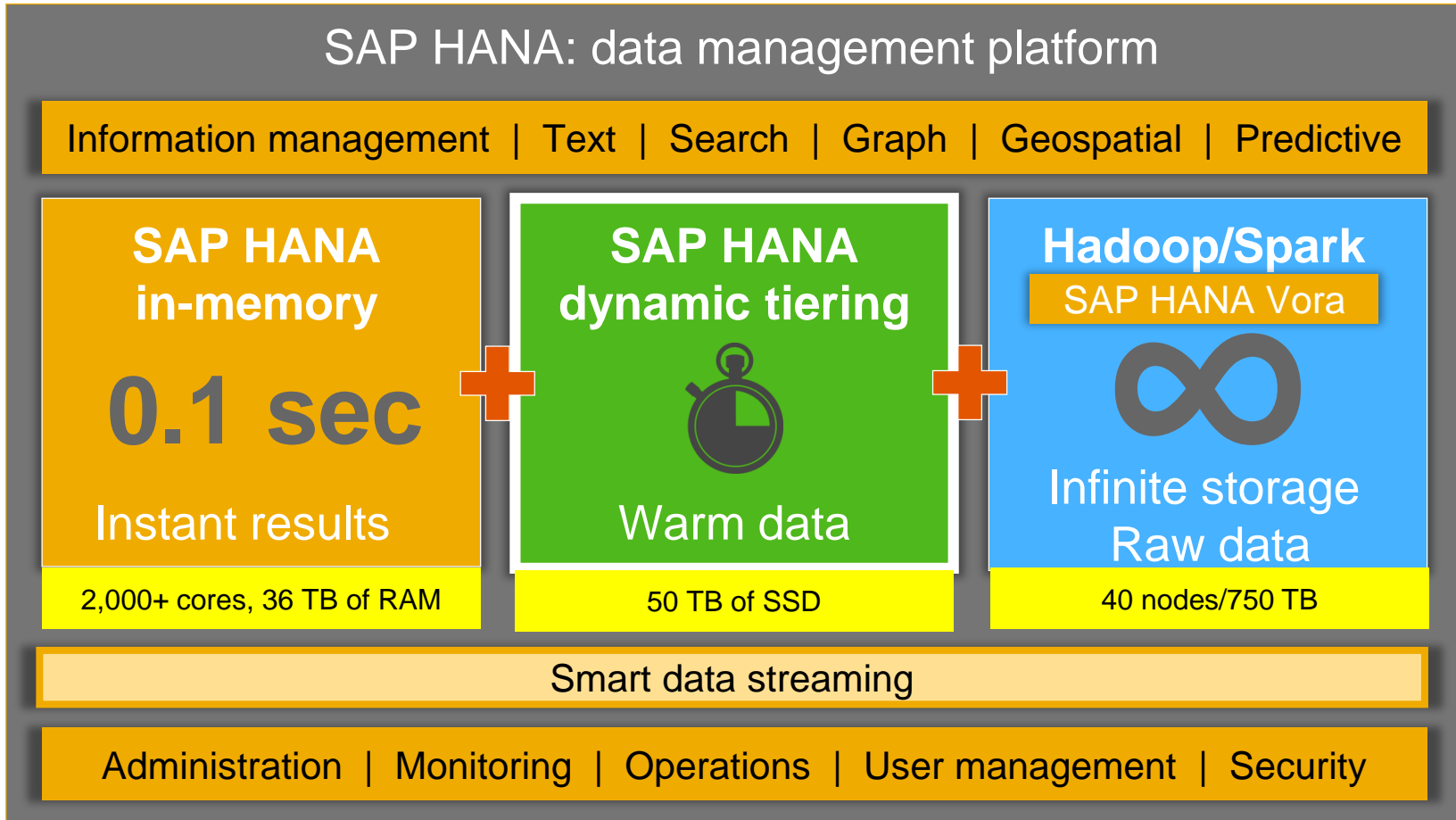
----- Projected Growth



Business as usual  
Move to SAP HANA/Hadoop  
Projected Savings  
75% Capex and Opex saving

Smart Meter Data grows more than 100TB/year, 1PB+ in 10 years

# Big Data platform



## SAP HANA and native Big Data

- Dynamic tiering
- Smart data streaming
- NoSQL, graph, geo, time series

## SAP HANA and Hadoop

- Smart data access → Hive, Spark
- MapReduce, HDFS
- Admin and monitoring
- User mgmt. security

## SAP HANA Vora and Hadoop extension

- SAP HANA Vora engine
- Integrated with SAP HANA and Hadoop
- Hierarchies

# Value

- Instant **real-time** analytics through SAP HANA
- **Ultimate flexibility** in choosing a storage tier based on the value of data
- SAP HANA allowing for more-precise **predictive** abilities for forecasting energy requirements; storage savings: **50%–75%**
- Compression with SAP HANA: Smart meter data **8:1 compression ratio**
- **Delayed training:** Using SAP HANA as the source does not require learning the Hadoop technology stack (Spark, MapReduce, etc.) to access data stored in Hadoop.
- Hadoop storage for **inexpensive** data storage for regulatory data
- Hadoop integration that allows for **data scientists** to use the Hadoop tool set
- Foundation for solutions for future analytics and questions that have **not yet been asked**
- End-user reporting that accesses all data from a **single location**
- **Acceleration of SAP HANA Vora** for Hadoop (hierarchies)

## Use cases

- Extraction of data to SAP HANA, running queries and measuring performance, and testing on all three tiers (SAP HANA, DT, Hadoop)
- Moving data from SAP HANA to extended storage/Hadoop
- Customer bill and correspondence storage and retrieval (100 million documents, will grow to 200 million documents before archiving, currently stored in DB2, 4-TB storage, PDF documents, HTML, text files)
- 15-minute interval data stored for regulatory reasons
- Involved parties:
  - CenterPoint
  - SAP (CoE, PE, Global ITP)
  - HP (hardware)
  - Hortonworks (Hadoop)

# POC test results

## Hadoop

HDP customer bill store and retrieval

→ **40-ms** response time to search and display a document from **19 million** PDFs

HDP batch data load via Sqoop into Hadoop

→ **4 min 24 sec** to load **2.5 million** records (single thread); **1 min 10 sec** (10 threads)

Data load from SAP HANA to HDP Hadoop via SAP HANA Vora

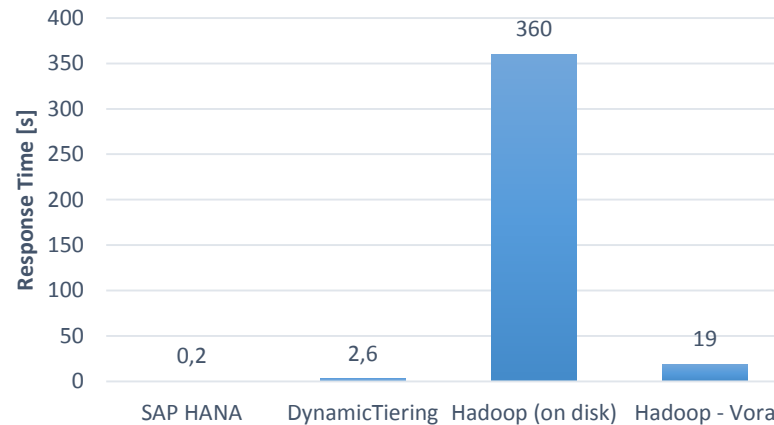
→ Total 6.2-GB ORC files stored in HDFS against original size of 172 GB.

→ **Compression rate: 9 (3 copies in HDFS)**

## SAP HANA, DT, Spark, SAP HANA Vora

Run aggregation query across SAP HANA, HDP Hadoop, and DT (~4 billion records):

**Query Response Time [s]**



## DLM

Move data from SAP HANA to DT

→ **289 million** records moved from SAP HANA to DT

→ **670K** records per minute  
Move data from SAP HANA to Hadoop via SAP HANA Vora into HDFS

→ **1.57 billion** records moved from SAP HANA to Hadoop

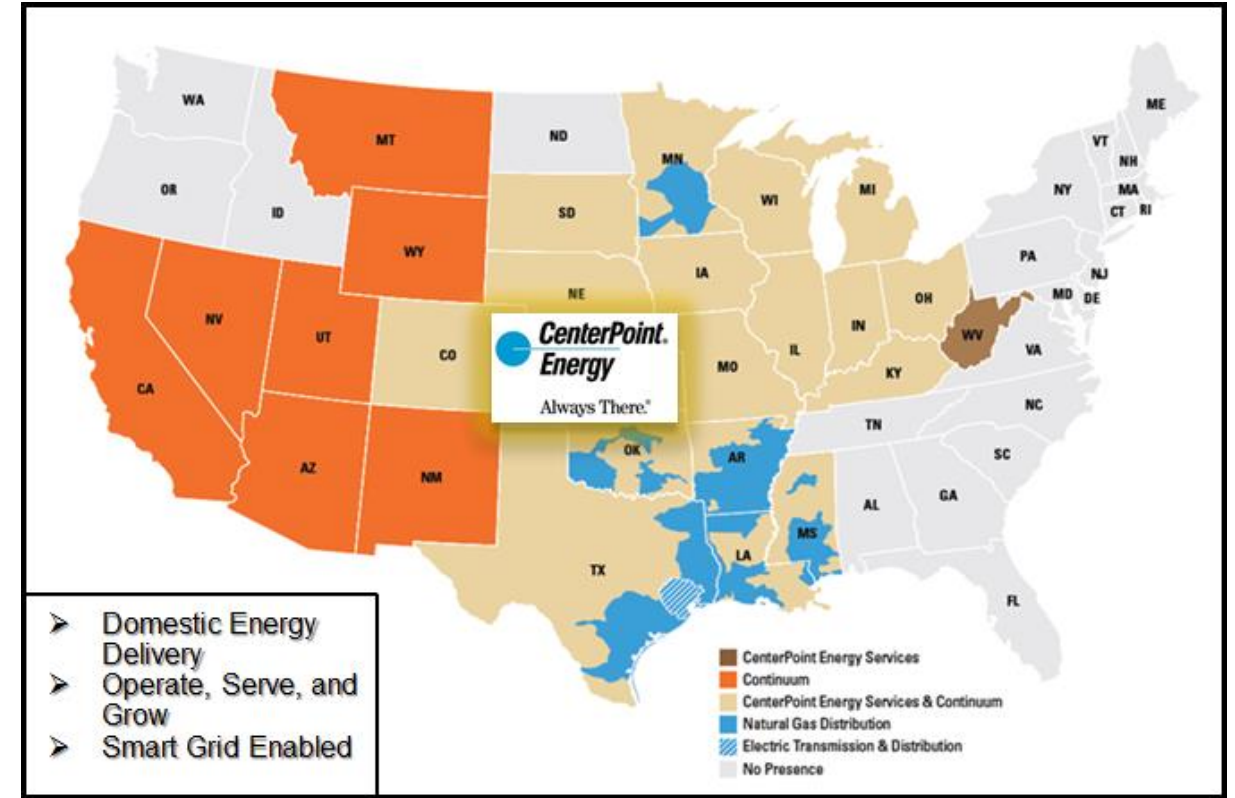
→ **22 million** records per minute



# Strategic Realization



- Big Data transitions from “cost liability” to “value asset”.
- Data Interrogation and analysis is platform based.
- Data Management is automated and optimized.
- Data Resolutions are real-time and decision oriented.
- Data sources are consolidated and integrated.



CenterPoint Energy Proprietary and Confidential

# © 2015 SAP SE or an SAP affiliate company. All rights reserved.

---

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. Please see <http://global12.sap.com/corporate-en/legal/copyright/index.epx> for additional trademark information and notices.

Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors.

National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP SE or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP SE or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.