Walmart Data Café Case Study 

## High lights:

* **Real time** analysis of over 200 billion rows of POS data, load 20 million rows/min
* **Scale:** 16+1 node cluster flexibility to grow and supports 1000+ concurrent users.
* **Compression:** 6 to 1compression ratio.
* **No pre-join, No pre-aggregation, dynamic** categorization on the fly
* **Performance:** Critical queries – under 2 seconds.

## Business requirements

Walmart’s business relies more and more on **real time** data analysis to help their daily operations. They want to be able to **drill down** to item level of Point of Sales (POS) data, enable more users to **concurrently** access the data with **high performance**, support **additional business requirements** such as business alerts. Even more critical, with the ever changing organizational structure and categorization of goods and products, they can’t get accurate analysis from pre-aggregates. They need the **flexibility** to dynamically calculate, aggregates dimensions on the fly.

## Technical situation

Walmart’s old implementation is based on Greenplum and Tableau with pre-aggregates. They can only support daily data load, limited number of users (3-5), drill down only to the pre-aggregated level and they have to develop very complex logic and operation procedures to generate and manage aggregates.

## HANA Solution and Benefits

With SAP HANA, we remove all the pre-aggregates and allow dynamic categorization of the entire 200B row of POS data, it not only meet the business requirement of flexibility but also significantly reduce the complexity which was required to maintain the aggregates. Our HANA solution can now support hourly data load (with the real bottle neck on the data gathering side), scalable to 1000+ users with fast response time. HANA Solution also provides significant data scalability through scale up and scale out approach. HANA provides dynamic queries with fast response time, for critical ones under 2 seconds. Instead of limited by aggregates, HANA provides drill down capability to item/row level and can now support non-cumulative metrics for business.

**Dynamic, Real-time, High Concurrency, High Performance** and **High Scalability** is what separate HANA from competitors and meet the growing business demands.