SAP Digital Interconnect Services is a global leader in mobile interconnection and mobile consumer engagement services. It provides mobile operators with capabilities in global messaging interconnect, data roaming, and an array of IPX-based services and enables enterprises to engage with their consumers through innovative mobile marketing and engagement solutions. SAP Digital Interconnect helps businesses process more than 1 billion messages per day, reaching 1,100 operators and 6.1 billion subscribers across 220 countries.

**THE CHALLENGE**

Mobile devices have become an extension of our personalities and offer a unique view into consumer behavior, but filtering out the noise that comes with data tracking is difficult. SAP Digital Interconnect provides anonymized data on 50 billion events, roughly 40-50 terabytes of uncompressed data, each day. That’s a yearly data volume of 1.3PB at a 1:10 compression ratio. This data needs to be cleaned up and normalized in order to provide customers with meaningful insights. To sustain current growth rates, SAP Digital Interconnect needed a cost effective storage solution that met its high SLAs. "We need to have the appropriate resources to store large volumes of data, process that information, and be able to load it without losing any time. Our SLAs are very high if we want to sustain our current growth," says Steven Garcia, Head of Engineering and Operations for Cloud Solution Services.

**MAPR SOLUTION**

SAP Digital Interconnect chose the MapR-XD Cloud-Scale Data Store because they required a modular, componentized architecture with implicit Hadoop features and enhanced performance that could run on commodity hardware or in the cloud. Cost was an important factor. "MapR-XD provided us with a huge amount of storage for a fairly low cost, which was ultra-critical for us," says Joe Love, Senior Staff Infrastructure Engineer. "Storage was our biggest challenge because it’s the largest consumer of resources. MapR-XD was the only choice. No other file system allowed us the flexibility to grow at the rates we needed or gave us tunable performance without having to get into huge engagements of changing out components of massive traditional enterprise solutions."
BENEFITS

SAP Digital Interconnect chose MapR-XD as the underlying storage layer with SAP IQ because of its scalability, flexibility, speed, and cost.

SCALABILITY AND FLEXIBILITY. The MapR-XD architecture is designed to allow for work on individual nodes while the system is running without any performance issues. In testing, SAP Digital Interconnect found they could take down multiple nodes with no performance degradation, and the system did not fail until half the installed nodes were taken out.

MapR-XD can be scaled up or down as necessary, which SAP has done quite a bit in the last year. In addition, other system components can be upgraded or retrofitted without bringing the system down. When additional storage is needed, more nodes are simply added.

“Once one of our carriers saw that we could handle a certain amount of data with no issue, we negotiated to handle much more of their data,” says Garcia. “As a result, we have experienced a data explosion of immense proportions over the past 15 months with growth rates five to seven times larger than our projections. The way we handled it was to throw more storage at it and let MapR-XD handle it for us.”

SPEED. Speed was an important factor when choosing MapR-XD. SAP Digital Interconnect needed a solution that could store, process, and load data without losing any time. In production, the majority of their volumes experience extremely high activity with hundreds of thousands of files in some directories, tens of thousands of which are open at any time. “Even when searching against one trillion records, ten minutes is too long. With MapR, reports are happening in a fraction of that time,” says Garcia.

HARDWARE EFFICIENCY. Using MapR software on commodity hardware, Outbrain increased their data center capacity while using significantly less hardware. “With the new hardware configuration and improved technology, we realized a dramatic reduction in the number of nodes per cluster resulting in one-third the footprint within the datacenters,” explains Yaron. Jobs migrated to ORC show a 10-20% improvement for the cluster daily workload. The top line of the graph below shows a reduction of 50,000,000 CPU seconds per day.

COST. SAP Digital Interconnect estimates MapR-XD costs one-tenth that of commodity arrays. They plan to convert other IQ implementations to MapR-XD storage for data servers as well as application related file systems. “We plan to get out of any high commodity arrays and run completely on MapR by the end of 2017,” says Garcia. “For most things, it just works and is highly durable.”

“WE PLAN TO GET OUT OF ANY HIGH COMMODITY ARRAYS AND RUN COMPLETELY ON MAPR BY THE END OF 2017.

FOR MOST THINGS, IT JUST WORKS AND IS HIGHLY DURABLE.”

Steven Garcia
Operations for Cloud Solution Services, SAP