

MAPR CONVERGED PLATFORM

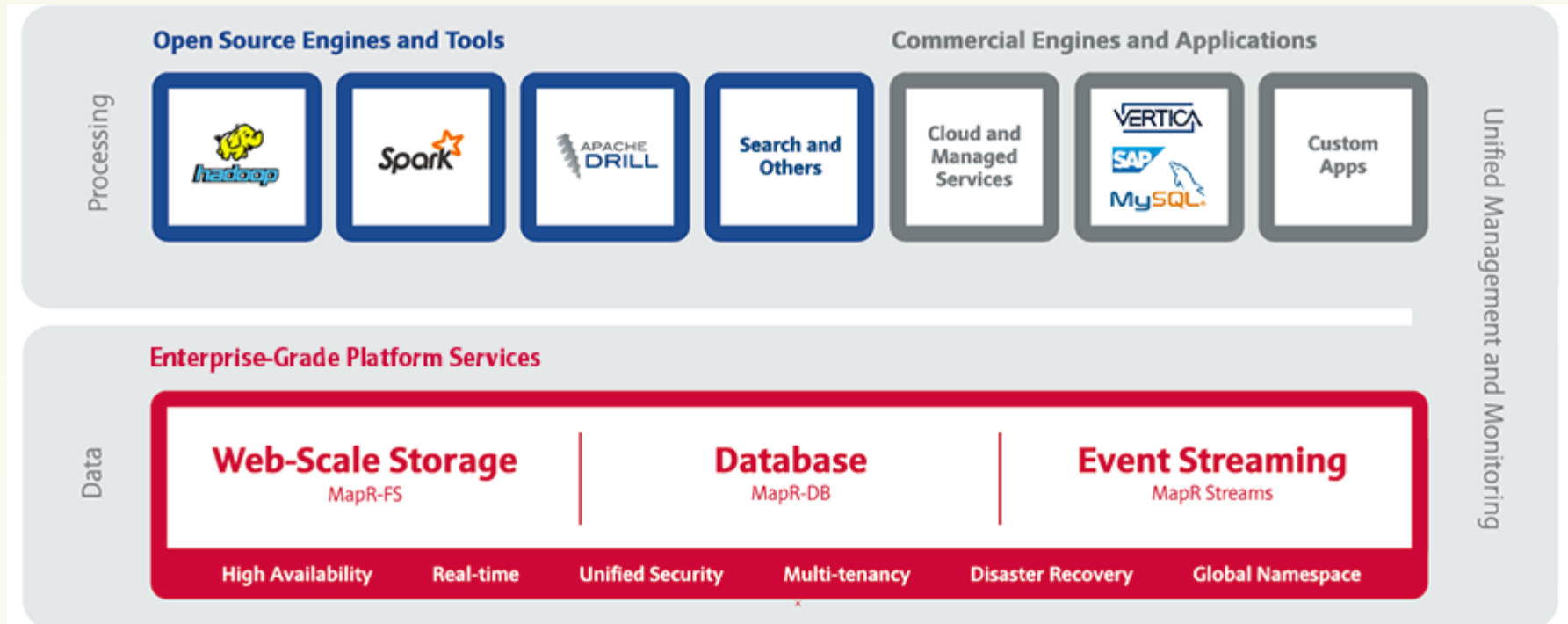
Abstract blue and cyan curved lines, resembling a stylized 'S' or a series of connected arcs, positioned on the right side of the slide.

APR. 2016

NOEL GAUR.
PRINCIPAL SOFTWARE ENGINEER

ADVANCED
ANALYTIC SERVICE

Introduction



- PROPRIETARY CLOSED SOURCE
- REWRITE OF CORE HADOOP TECHNOLOGIES IN SYSTEM LANGUAGE - C/C++
- API COMPATIBILITY WITH APACHE HADOOP.

MAPR FILESYSTEM

- FULL NFS SUPPORT
- HADOOP API COMPATIBILITY
- SUPPORT FOR VOLUMES AND NETWORK PARTITION ON CLUSTER.

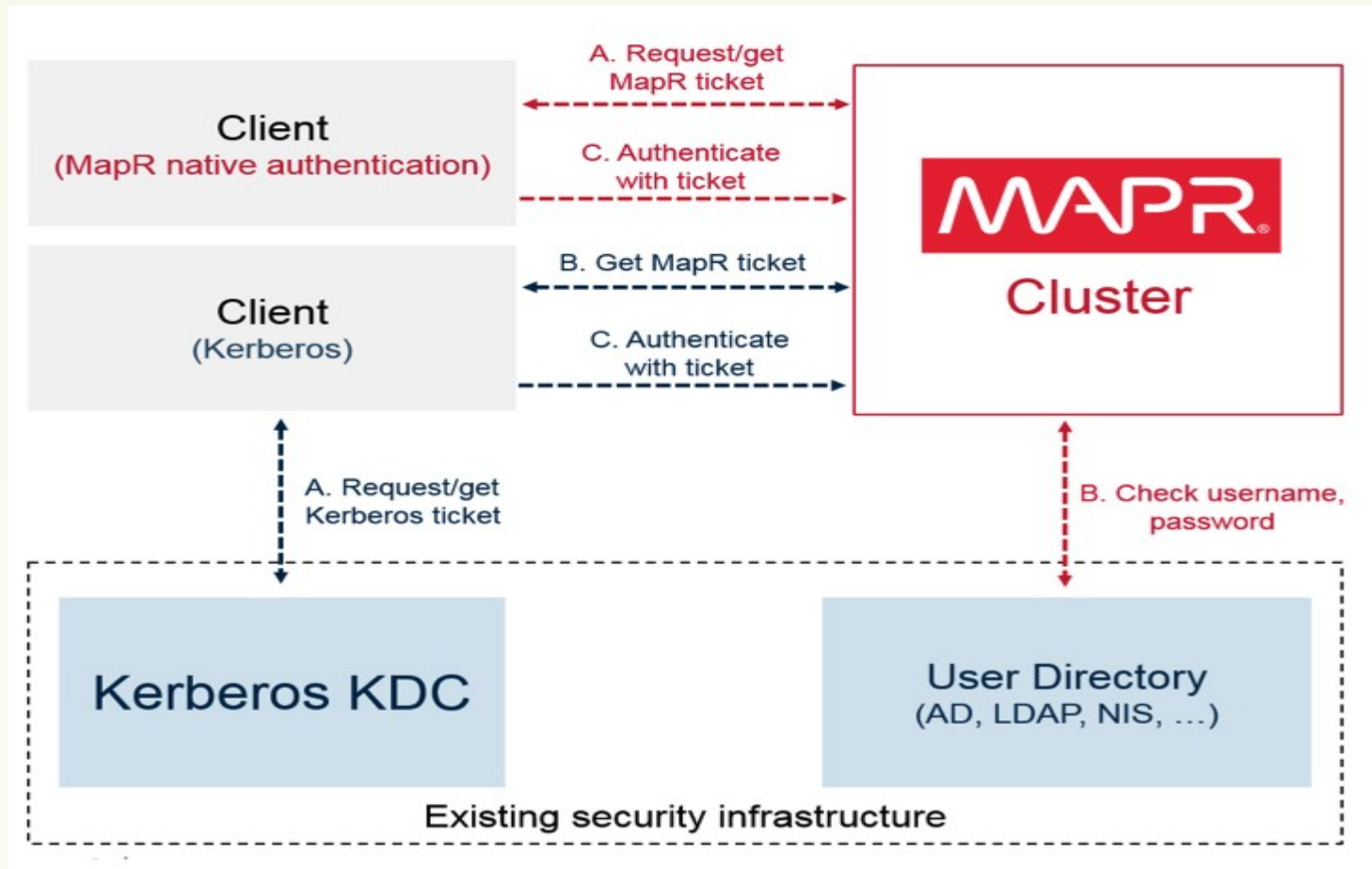
MAPR DB

- REAL-TIME OPERATIONAL ANALYTICS
- MINIMAL ADMINISTRATIVE OVERHEAD
- FINE GRAINED SECURITY
- OPTIMIZED FOR SSDs
- NATIVE JSON SUPPORT
- OJAI SUPPORT

MAPR STREAMS

- PERSISTENCE OF STREAMING DATA
- UNIFIED SECURITY FRAMEWORK
- UTILITY-GRADE RELIABILITY
- GEO-REPLICATION CAPABILITY
- KAFKA API

CENTRALIZED AUTH



BUILDING IOT SCALE – DATA INGESTION PLATFORM.

OUR CASE

- COLLECT WIFI ROUTER DATA THROUGH TCP/IP CHANNEL.
- 20,000 DEVICES.
- SHORT MESSAGES OF 0.1 – 10K KB
- SCALE OF 1 TB/DAY INGESTION VOLUME.

BEFORE

- AKKA HTTP FRAMEWORK TO RECEIVE TCP/IP BOUND MESSAGES.
- APACHE KAFKA – DISTRIBUTED MESSAGE QUEUE.
- CASSANDRA – DATA STORAGE, ANALYTIC REPORTING.
- APACHE SPARK – COMPUTATION PLATFORM.
- DISTRIBUTION – CDH – 5.4

CHALLENGES

- KAFKA PARTITIONS OCCASIONALLY WENT DOWN.
- KAFKA SECURITY – THIS IS BEFORE 0.9
- CASSANDRA COMPACTION CYCLE – EXCESS LOAD ON SERVER.
- SPARK STREAMING JOBS OCCASIONALLY WENT DOWN.

STILL BEFORE

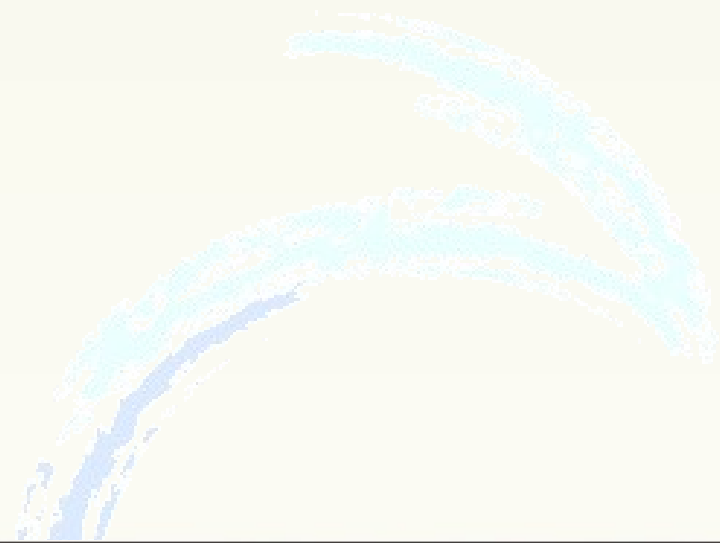
- LIBUV DEVELOPED MULTI-THREADED C DAEMONS.
- MIGRATE AKKA RECEIVER PROGRAMS TO C.
- MIGRATE SPARK STREAMING ESSENTIAL JOBS TO C.
- LIBRDKAFKA & LIBCASSANDRA ARE EXCELLENT LIBRARIES.
- LIBCASSANDRA IS CASSANDRA EQUIVALENT OF ASYNCHBASE ONLY BETTER.

AFTER

- C CUSTOM DAEMONS TO RECEIVE TCP/IP INBOUND SENSOR DATA.
- MAPR STREAMS – DISTRIBUTED MESSAGE QUEUE.
- MAPR DB – OPERATIONAL STORAGE.
- MYSQL CLUSTER – FOR PARTNER'S NEED FOR ACCESSING DATA.
- SPARK STREAMING – COMPUTATION & AGGREGATION
- MAPR DB/DOCUMENT STORAGE – REAL TIME REPORTING NEEDS.
- MAPRFS – ARCHIVING

| MISS..

- CASSANDRA



Stockholm syndrome

| GAIN...

- PERFORMANCE AND RELIABILITY
- ZERO DOWNING OF DAEMONS.
- LESSER CLUSTER SPIKES.
- MORE CLUSTER RESOURCES TO DO COMPUTATION TASKS.
- LESS MAINTENANCE HEADACHES

LESSON LEARNED

- ITS A BI*** DEVELOPING THINGS ON C BUT ITS WORTH IT.
- MAINTAINING “STATE” IN SPARK STREAMING IS LIMITED.
- MAPR STREAMS C API IS KNOWN TO HAVE DEADLOCK IN SUBSCRIBER API.
- NAT IS NOT SUPPORTED BY MAPR .

Look Ma no hands..

MAPR COMPATIBILITY

- SPARK/STREAMING PROGRAMS CAN SEAMLESSLY MIGRATE MAPR PLATFORM WITHOUT ANY CODE CHANGE.
- MAPR C PROGRAMS CAN BE BUILD USING LIBHDFS, LIBHBASE OR LIBMAPRCLIENT

THANK YOU!

CONTACT:

NOEL GAUR
PRINCIPAL SOFTWARE ENGINEER
ANALYTIC SERVICE LTD.
(+86) 1872 1780 560
NOEL.GAUR@ANALYTICSERVICE.NET

