

Title: On-Demand Risk Management

Abstract:

Value at Risk (VaR) is a popular mechanism used to measure risk of an investment. On-Demand Risk Management solution gives the ability to run VaR on a portfolio for any given date. It also provides the ability to compare portfolio VaR values across different days.

Introduction:

Value at Risk measures the potential loss in value of a risky asset or portfolio over a defined period for a given confidence interval. Thus, if the VaR on an asset is \$1 million with a 95% confidence level, it means that there is only a 5% chance that the value of the asset will drop more than \$1 million over the time period. VaR can be calculated using several different mechanisms. Monte Carlo simulation is the most widely used mechanism to generate random paths which are needed to calculate VaR. It is also the mechanism mandated by regulatory authorities to measure risk in financial institutions.

Problem Definition:

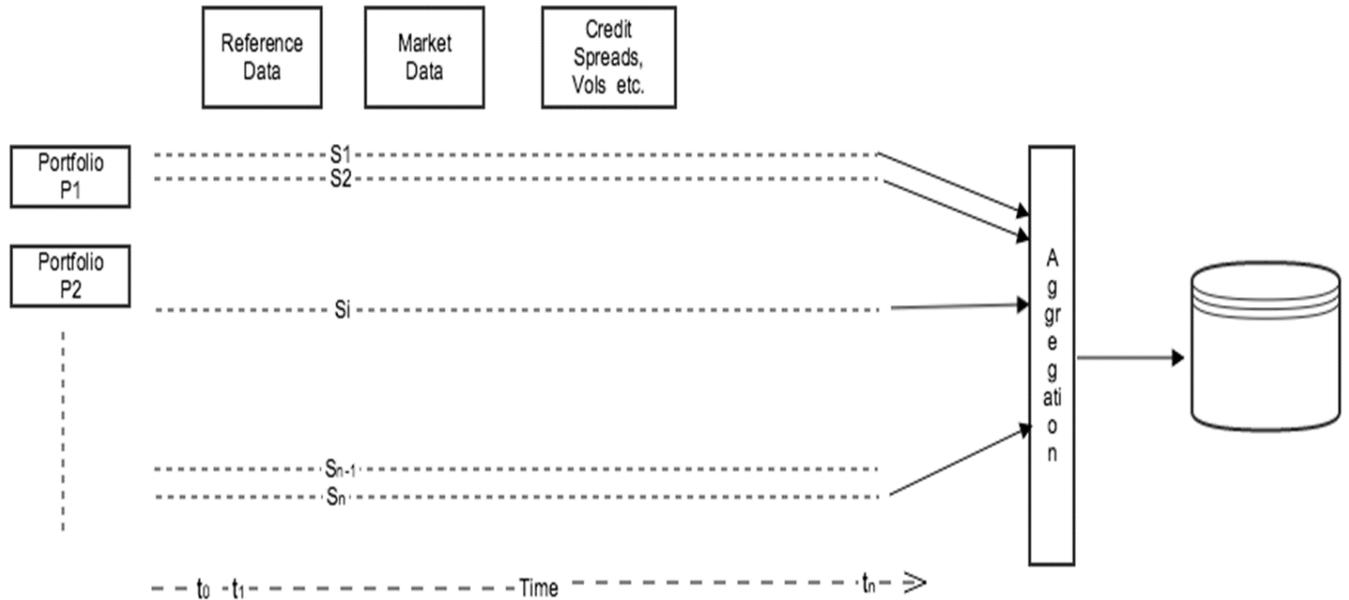
Calculating VaR using Monte Carlo simulation is a computationally intensive process. Monte Carlo simulation is used to generate several random paths or scenarios which the portfolio can take during the given time period. The return of the portfolio is calculated across all the random paths. VaR is then calculated by aggregating the values across all the scenarios. The value of VaR becomes more precise as the number of random paths simulated increases. On-Demand VaR application should provide users the ability to kick off VaR computation for any given date. It should also provide the users ability compare the results across dates and do a deep dive analysis of the results.

Solution Details:

The process of VaR calculation involves large amount of data. Some of the data that we are going to talk about in this solution are as below:

- Position Data - The set of trades or portfolios for which VaR is being calculated
- Reference Data – Data with information about the legal entities, instruments etc.
- Market Data – Data that represents market conditions of any given date
- Scenario Data – Simulation data that is generated using Monte Carlo simulation. Also called Exposure data.

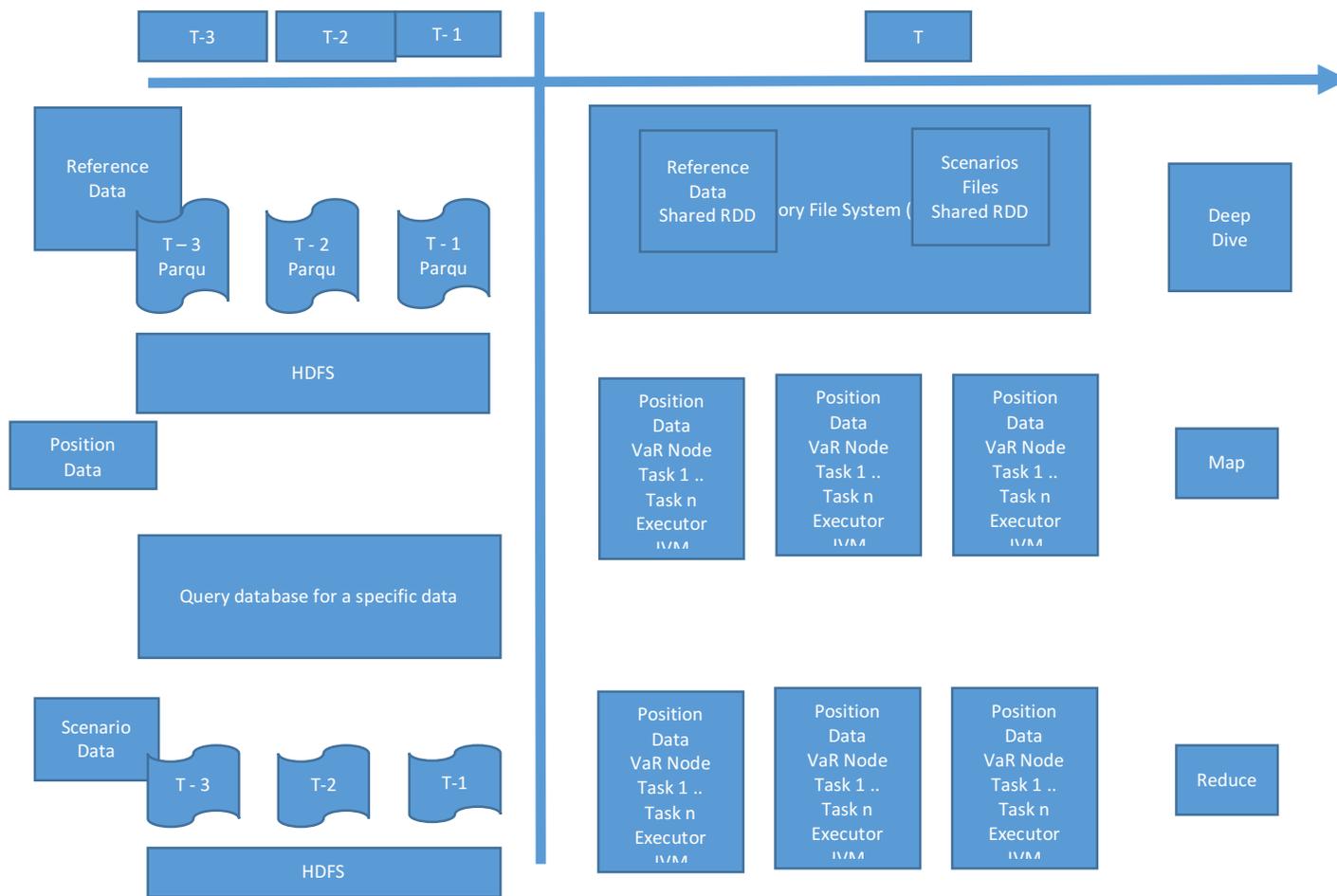
Hadoop distributed file system is chosen to store the vast amount of data that is generated as part of the VaR calculation process as it provides an easy and flexible storage platform.



Initially, market features that determine the value of a portfolio are extracted. Then, we use Monte Carlo simulation to randomly generate the market features across a fixed number of trials. Monte Carlo simulation relies on the assumption that each of the paths is independent and identically distributed. Because of this assumption, the entire process of simulation and calculation of portfolio returns across all the paths becomes an embarrassingly parallel problem. As shown in the diagram above, all the paths can be generated independently and in a distributed fashion. The computations done in these steps can be treated as map tasks. We have to do few aggregation tasks after all the simulations and valuations across all the simulated paths are completed. These aggregation operations can be treated as reduce tasks. From the diagram, we can see that the entire solution of calculating VaR is very similar to map reduce paradigm. This sums up the process of VaR calculation.

As part of On-demand Risk management application, VaR has to be calculated on the fly for a date given by the user. The solution should also have the ability to store the results and offer a way to compare the results across days. It should also provide the ability to do a deep dive analysis of the results if needed. In order to achieve this, we came up with the below architecture for the application.

Apache Spark was chosen as the distributed computation framework to implement VaR solution over the Hadoop file system. We are developing the application in Scala to get the maximum utilization of multi-core CPUs. We store Reference Data for the current date as a cached RDD so that it can be used for multiple runs that could be launched by different users. We use in-memory file system from Apache Ignite and store it as a shared RDD. We also cache the scenario files generated for previous runs. We store any scenario files generated for other dates on HDFS in parquet format. We will broadcast position data to all the nodes. By doing this, we aim to reduce the network I/O for lookups and joins on the position data.



We store Reference Data and Position Data in HDFS for earlier dates. This will enable us to quickly load the data if we have to run the VaR computation for any of the earlier dates. We also store the results from previous computations on the disk

in Parquet files. This would allow us to quickly search and retrieve the required data from the results. The worse performing scenarios and portfolios are often analyzed further in order to gain more insights and do better risk management. Storing the results on the disk would expedite the analysis process. Spark provides the ability to use other Spark libraries and packages to deep dive into the data. We store the logs for all the computations in HBase so that it is easier to retrieve the logs based on the date and portfolio.

Benefits of the Proposed Solution:

- Ability to run quick comparison of VaR results across different dates and portfolios
- Easy to run other VaR measures like CVaR, SVaR etc. by reusing the simulation data.
- Faster execution of computations by using Spark
- Broadcast position data to all the nodes in the cluster reduce network I/O
- Reduced Network I/O by using in-memory file system by Apache Ignite
- Reduced number of intermediary files generated
- Efficient storage of data on HDFS by using Parquet format

Conclusion:

We have developed On-Demand Risk management solution using Apache Spark framework. We used other open source technologies such as Apache Ignite to leverage its in-memory file system and integrated it with Spark framework. Although VaR calculations are computationally intensive, we have shown that the computation time can be improved significantly by using Apache Spark. We have also implemented several optimization techniques in Apache Spark to expedite the computations. In summary, the On-Demand Risk Management solution developed by us not only provides the ability to run VaR on demand for any given date but also gives the capability to do deep dive analysis into the results and run comparisons of results across dates.